

# On Multihop Distances in Wireless Sensor Networks with Random Node Locations

Serdar Vural, *Member, IEEE*, and Eylem Ekici, *Member, IEEE*

**Abstract**—Location and intersensor distance estimations are important functions for the operation of wireless sensor networks, especially when protocols can benefit from the distance information prior to network deployment. The maximum multihop distance that can be covered in a given number of hops in a sensor network is one such parameter related with coverage area, delay, and minimal multihop transmission energy consumption estimations. In randomly deployed sensor networks, intersensor distances are random variables. Hence, their evaluations require probabilistic methods, and distance models should involve investigation of distance distribution functions. Current literature on analytical modeling of the maximum distance distribution is limited to 1D networks using the Gaussian pdf. However, determination of the maximum multihop distance distribution in 2D networks is a quite complex problem. Furthermore, distance distributions in 2D networks are not accurately modeled by the Gaussian pdf. Hence, we propose a greedy method of distance maximization and evaluate the distribution of the obtained multihop distance through analytical approximations and simulations.

**Index Terms**—Wireless sensor networks, multihop distance, analysis, simulation, estimation.

## 1 INTRODUCTION

METHODS providing information on the distribution of the euclidean distance between nodes in wireless sensor networks (WSN) are regarded as versatile tools for protocol parameter tuning and performance modeling. Therefore, internode distance estimation is of profound importance for various WSN applications. For instance, when urgent data are disseminated from a source node via broadcast, it is critical to estimate the covered distance as a result of a sequence of data forwards. Similarly, estimation of the hop distance between two network locations is equivalent to estimating the minimum number of hops, which leads to maximization of the distance covered in multihop paths. Furthermore, hop distance estimation is closely related with transmission delay estimation and minimization of multihop energy consumption.

The current literature on the maximization of the multihop distance is limited; however, there are several studies that focus on estimation of single-hop or multihop distances. Methods in [1] and [2] estimate the distances to designated anchor nodes using optimization algorithms. There are also analytical methods that address the probabilistic evaluation of the euclidean distances such as that given in [3] and [4]. In [3], the probability distribution of the single-hop distance between two randomly chosen neighbors is investigated. In

[4], the distribution of the remaining distance in multihop greedy forwarding to a destination node is derived. Furthermore, in [5], the distribution of euclidean distances to  $n$ th neighbor in a Poisson point process is analyzed.

Hop distance estimation between two locations is studied analytically in [6] and [7]. In [6], connectivity probability in one or two hops is derived and connectivity in multiple hops is studied with analytical bounds. Furthermore, analytical bounds of the expected hop distance are derived and supported by simulation results. In [7], the expected number  $n$  of relay nodes between two randomly located sensors is analytically computed via iteration based on expressions for connectivity in one or two hops. In [8], the distribution of hop distance and its expected value are analyzed with simulations. It is shown that beamforming antennas significantly reduce the hop distance compared to omnidirectional antennas for medium and large networks with random node locations.

The maximum euclidean distance metric is first investigated in [9] for 1D networks. The distribution of the maximum distance taken in a single hop is derived analytically. In our previous study [10], we showed that the distribution of maximum multihop euclidean distance in a 1D network has an increasingly Gaussian character for increasing hop distance values. This result is used in [11] to develop a Bayesian decision mechanism to determine the number of hops for a given euclidean distance. An application of [10] to planar sensor networks is [12] which uses the Gaussian pdf to verify sensor positions. In [12], 1D network results are mapped to planar networks by defining a thin rectangular corridor between two locations. Since the definition and calculation of multihop path distances in a planar network are considerably complex compared to a 1D network, the distribution model is not highly accurate. Hence, results of [10] are largely limited to 1D networks.

In this paper, we propose a greedy method which is aimed at maximizing the multihop distance in 2D networks

• S. Vural is with the Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, and the Department of Computer Science and Engineering, University of California, Engineering Building Unit 2, 900 University Ave., Riverside, CA 92521. E-mail: vurals@ece.osu.edu.

• E. Ekici is with the Department of Electrical and Computer Engineering, The Ohio State University, 205 Dreese Laboratory, 2015 Neil Avenue, Columbus, OH 43210. E-mail: ekici@ece.osu.edu.

Manuscript received 8 Jan. 2008; revised 15 Sept. 2008; accepted 4 Aug. 2009; published online 12 Aug. 2009.

For information on obtaining reprints of this article, please send e-mail to: [tmc@computer.org](mailto:tmc@computer.org), and reference IEEECS Log Number TMC-2008-01-0005. Digital Object Identifier no. 10.1109/TMC.2009.151.

with random node locations. Our method is based on restricting the propagation direction outward from the propagation source in each hop and greedily searching the furthest neighbor each time. This is similar to the approach in [10]; however, we demonstrate that Gaussian pdf, which is shown to accurately model distance distributions in 1D networks [10], cannot represent distance distributions in 2D networks. The accuracy of the Gaussian pdf model depends on the number of hops and the chosen parameters, which affect the obtained error ranges. We show that a transformation of the well-known Gamma distribution effectively approximates the obtained maximum distances in a 2D network. For the same range of number of hops, uniform accuracy is provided by the proposed method in this paper. Compared to the Gaussian pdf model in [10], a much more consistent and accurate representation of maximal multihop distance pdf as well as hop distance pmf in 2D networks is obtained. Using the pdf obtained by the transformation of Gamma distribution, we derive the expected value and standard deviation of the euclidean multihop distance and calculate the parameters of the Gamma distribution. The accuracy of the Gamma-approximated distribution is demonstrated through comparisons with simulation results.

The rest of the paper is organized as follows: In Section 2, multihop distance in data dissemination scenarios is defined and its similarity to well-known distributions is statistically inspected. In Section 3, multihop distance is analyzed and approximation expressions of its standard deviation and expected value are derived. Section 4 presents an analytical way of estimating the hop distance for a given euclidean distance using the results of Section 3. In Section 5, simulation and numerical results are provided. Finally, Section 6 concludes the paper.

## 2 GREEDY MAXIMIZATION OF MULTIHOP EUCLIDEAN DISTANCE

### 2.1 Network Model

WSNs considered in this study consist of sensors with circular communication ranges of radius  $R$ . Node locations are static and uniformly distributed with a planar density  $\sigma$ . Hence, the number of points  $n$  that can be found in a given area  $A$  has a Poisson distribution given by  $p_n(n) = \frac{(\sigma A)^n e^{-\sigma A}}{n!}$ . It is assumed that nodes can receive every packet transmitted within their communication ranges.

### 2.2 Definition of Maximum Multihop Euclidean Distance

In a 2D network of randomly located nodes, there exists only one node with the maximum distance that can be reached in a given number of hops. Due to spatially random node locations, the position of this node as well as its distance to the source node is random. However, analytical computation of the maximum euclidean distance in  $N$  hops involves considering all nodes in the previous hops and recursively reaches the source node of the multihop propagation, which becomes intractable even for small instances. Hence, we propose a greedy method of maximization of the euclidean multihop distance. By selecting locally maximally distant nodes, the multihop propagation intends to reach further distances to the source node. After

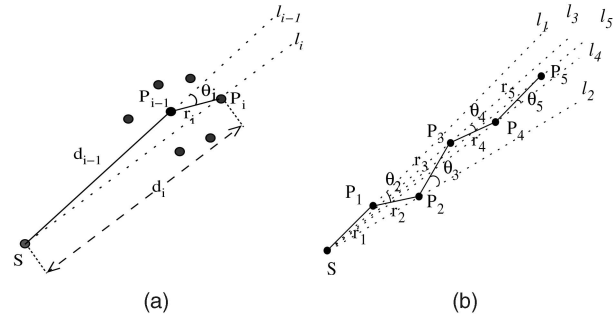


Fig. 1. Iterative definition of maximum multihop euclidean distance. (a) Definition of  $d_i$  and (b) multihop path of five hops.

selecting an initial propagation direction, the following iterative definition is provided for our greedy distance maximization scheme. It should be noted that our method does not locate the node with maximum distance to the source node for a given number of hops, yet it maximizes the euclidean distance toward a chosen initial direction greedily. For the definition of this distance, consider Fig. 1a for subsequent descriptions.

Let  $P_{i-1}$  be the node with the maximum euclidean distance, selected by the greedy selection, to a source sensor  $S$  in  $i-1$  hops. Furthermore, let the line  $l_{i-1}$  originating from  $S$  and passing through  $P_{i-1}$  act as a direction line which defines the outward propagation from  $S$ . Then, in the  $i$ th hop of the multihop path, node  $P_{i-1}$  chooses the neighbor node  $P_i$  whose euclidean distance to  $S$  is the largest among all neighbors of  $P_{i-1}$ . Such a node  $P_i$  may not be located on the line  $l_{i-1}$  since node density is finite. In fact, the line segment between  $P_{i-1}$  and  $P_i$  has an angle of deviation from  $l_{i-1}$  denoted by  $\theta_i$ . The length of this line segment is the distance of the  $i$ th hop denoted by  $r_i$ . Furthermore, we limit the search area of the new node  $P_i$  to a maximum angular deviation such that  $\theta_i$  is upper bounded. Finally, the euclidean distance of  $P_i$  to the source sensor  $S$  is denoted by  $d_i$  and represents the euclidean distance in a multihop path of  $i$  hops in an outward direction from  $S$ .

Fig. 1b demonstrates an example of a multihop path generated with this definition. After locating the maximally distant node  $P_i$  to  $S$  in each hop  $i$ , the propagation direction is changed and becomes the line  $l_i$  passing through  $S$  and the new node  $P_i$ . For a different initial direction, other distances can be obtained. Hence, our method maximizes the distance greedily while fixing the direction of propagation outward from the source node. Due to the uniformly distributed node locations, the distribution of  $d_i$  is identical for all propagation directions and denoted by  $p_{d_i}(d_i)$ .

### 2.3 The Probability Distribution of Multihop Euclidean Distance

In our previous study [10], the distribution of the maximum distance for a given number of hops in linear networks is modeled by the Gaussian pdf. The mean  $\bar{r}$  of the maximum distance of a single hop is numerically found by the implicit equation  $\ln(1 - \frac{\lambda \bar{r}}{\lambda R - \lambda \bar{r} - 1}) = \lambda \bar{r}$ , where  $\lambda$  is the linear node density and  $R$  is the sensor communication range. The expected value of the multihop maximum distance  $d_N$  of  $N$  hops is then found by  $E[d_N] = N\bar{r}$  and the variance of  $d_N$

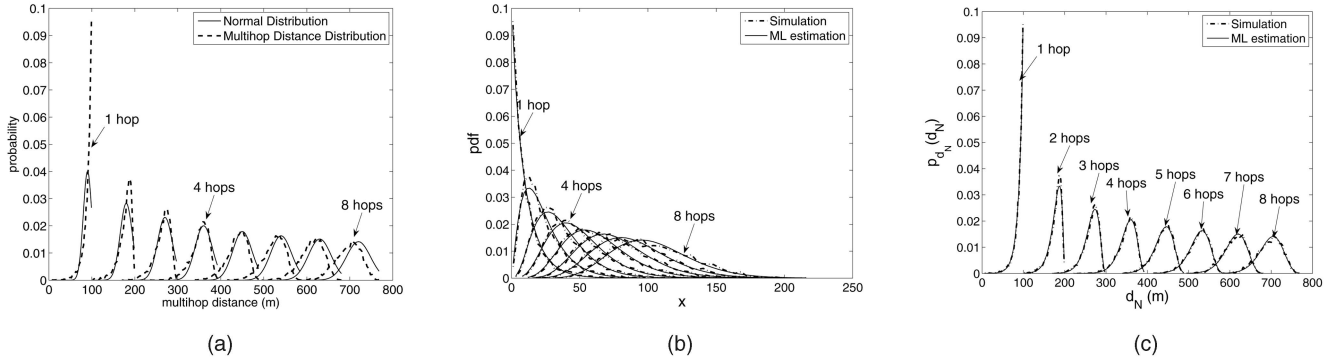


Fig. 2. Distribution estimations using 1D Model with Gaussian pdf and ML Estimation with Gamma pdf,  $\sigma = 0.001$  nodes/m<sup>2</sup>. (a)  $p_{d_i}(d_i)$  estimated by the 1D Model, (b)  $p_x(x)$  and ML estimation, and (c)  $p_{d_i}(d_i)$  and ML estimation.

is calculated by  $\sigma_{d_N}^2 \cong E[(\sum_{i=1}^N r_i)^2] - N^2\bar{r}^2$ , which involves a recursive computation. The results in [10] illustrate that the distribution of the multihop maximum distance in 1D networks can be modeled by the Gaussian distribution with high accuracy. This model is referred as the “1D Model” throughout the rest of this paper.

In contrast to 1D networks, a multihop path in a planar network is generally not located on a line. However, one can define a thin rectangular corridor with a small width  $W$  between two locations and search the multihop path within this corridor. Using this mapping, nodes can be assumed to be in a linear formation. Hence, the equivalent 1D network density becomes  $\lambda = W\sigma$ , where  $\sigma$  is the actual node density on the 2D network. With this approximation, the 1D Model can be applied to planar networks to represent the maximum multihop distance distribution with a Gaussian pdf [12].

The simulation results for multihop maximum distance distribution  $p_{d_i}(d_i)$  for  $i \leq 8$  and the Gaussian pdf found by the 1D Model are illustrated in Fig. 2a with  $W = R$ . Although Gaussian pdf accurately models 1D network distance distributions [10], it fails to represent the distributions in planar networks. However, since applications such as distance estimation and localization are sensitive to even small amounts of errors in intersensor distances, a more accurate model that can better represent distance distributions is required.

In a 2D network with random node locations, a path with  $N$  hops can have a length of at most  $NR$ , with a sensor communication range of  $R$ . However, due to the finite node density  $\sigma$ , the actual maximum distance  $d_N$  is smaller; hence,  $NR$  is an upper bound for  $d_N$ . If we define this difference as

$x = NR - d_N$ , then  $x$  is a random variable. To determine the similarity of its distribution to well-known distribution types, Kolmogorov-Smirnov (KS) Test is applied to a simulation data set of the random variable  $x$ . The data set is collected from 10,000 different networks with uniformly distributed node locations and a sensor communication range of  $R = 100$  m. The KS test suggests that a data set is similar to a given distribution type if the KS statistic is lower than the critical value of the test for the chosen significance level. The critical value of the KS Test for a significance level of 0.05 is determined and then compared with each KS statistic that is obtained by  $x$  values in simulations corresponding to different hop distance values.

Fig. 3 demonstrates the test results for two node density values,  $\sigma = 0.001$  nodes/m<sup>2</sup> and  $\sigma = 0.002$  nodes/m<sup>2</sup>, and for various distribution types. The highest similarity results are obtained for Gamma distribution. For  $\sigma = 0.002$  nodes/m<sup>2</sup>, the test results for the Gamma distribution are consistently below the critical value computed by the KS Test. This suggests that the distribution of  $x$  is similar to Gamma distribution. The Maximum Likelihood (ML) Estimation of the distribution of  $x$  with the Gamma pdf, as illustrated in Fig. 2b for a node density  $\sigma = 0.001$  nodes/m<sup>2</sup>, also clearly demonstrates this similarity.

By modeling the distribution of  $x$  using Gamma distribution, the distribution of the maximum euclidean distance  $d_N$  can be defined since  $d_i = iR - x$ . When the distribution of  $x$  obtained by ML estimation is transformed with  $d_i = iR - x$ , the ML estimate of the maximum distance distribution  $d_i$  is obtained, which is shown in Fig. 2c for  $\sigma = 0.001$  nodes/m<sup>2</sup>. This figure suggests that the pdf of  $d_i$  can be obtained through a simple linear

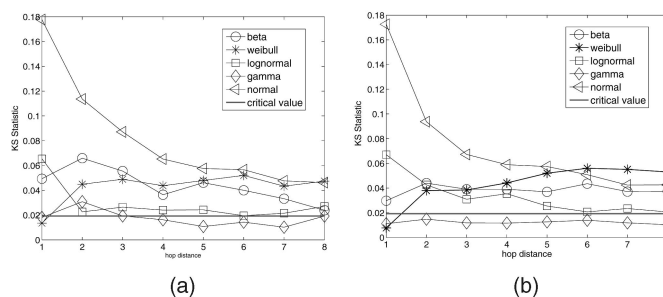


Fig. 3. Kolmogorov-Smirnov Test applied to  $x$ . (a)  $\sigma = 0.001$  nodes/m<sup>2</sup>. (b)  $\sigma = 0.002$  nodes/m<sup>2</sup>.

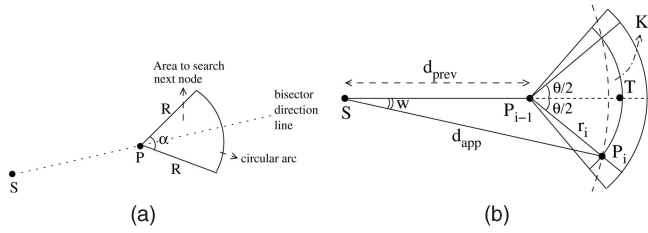


Fig. 4. Angular range model and model-related approximation. (a) Angular model and (b) model approximation.

transformation applied to a Gamma pdf. The Gamma pdf of the random variable  $x$  is given by  $p_x(x) = x^{a-1} \frac{b^a e^{-bx}}{\Gamma(a)}$ , where  $\Gamma(\cdot)$  is the Gamma function  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ , and  $a$  and  $b$  are shape and scale parameters, respectively. Furthermore, the mean and variance of the Gamma distribution in terms of  $a$  and  $b$  are given by  $E[g(x, a, b)] = ab$  and  $var(g(x, a, b)) = ab^2$ , respectively.

The maximum likelihood estimation provides the parameters  $a$  and  $b$  which uniquely define the Gamma distribution. However, ML estimation requires simulation data to estimate the pdf. *In contrast, our aim is to avoid the use of simulation data, and compute the parameters  $a$  and  $b$  analytically.* Hence, the analysis in this paper aims to analytically compute  $E[d_i]$  and  $\sigma_{d_i}$  and models the distribution of  $d_N$  using the Gamma distribution with parameters  $a$  and  $b$  that are calculated using  $E[d_i]$  and  $\sigma_{d_i}$ , i.e.,

$$b = \frac{\sigma_{d_i}^2}{E[d_i]} \quad \text{and} \quad a = \frac{E[d_i]^2}{\sigma_{d_i}^2}.$$

### 3 ANALYSIS

#### 3.1 Greedy Maximization of Multihop Distance

In Fig. 4a, the search area for a next hop sensor is modeled by an angular communication range with a “radial range”  $R$ , which is equal to the radius of the sensor communication range, and an “angular range”  $\alpha < \pi$ . The bisector direction line represents the outward direction from the source sensor  $S$ .

Maximization of multihop distance, as defined in Section 2.2, is obtained by choosing the next sensor node within the angular communication range such that its distance to the source sensor  $S$  is maximum. In Figs. 5a and 5b, no nodes are found in regions  $A_1$  and  $A_2$ , and  $d_a$  and  $d_b$  constitute the maximum distance to  $S$ .  $r_a$ ,  $r_b$ , and  $r_c$  denote the radial distance of the selected next node to  $P$ . However, the shapes of the regions created by this choice depend on the location of the node and distances  $d_a$  and  $d_b$  which complicate the analysis.

On the other hand, in Fig. 5c, the node with the maximum distance to node  $P$  is chosen as the next node.

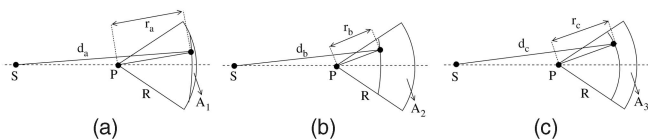


Fig. 5. Choice of next node in angular range.

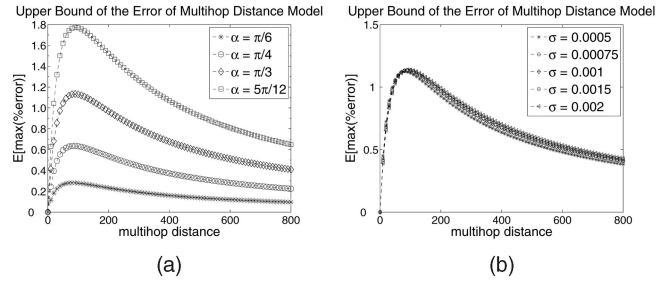


Fig. 6. Average upper bound of model-based error,  $R = 100$  m. (a) Different angles  $\alpha$  and (b) different density values  $\sigma$ .

Note that the region  $A_3$  can easily be defined by the difference between two arcs centered at  $P$ . However, this is an approximation since any node in region  $K$  in Fig. 4b is more distant to  $S$  compared to the radially most distant node to  $P_{i-1}$ . In fact, the largest amount of error occurs if a point exists in location  $T$ , as shown in Fig. 4b. The distance of  $T$  to  $S$  is  $d_{max} = d_{prev} + r_i - \varepsilon$ , where  $r_i$  is the distance of the radially distant node  $P_i$  to node  $P_{i-1}$  and  $\varepsilon$  is a negligible small number. However, the distance of  $P_i$  to  $S$  is

$$d_{app} = \sqrt{d_{prev}^2 + r_i^2 + 2r_i d_{prev} \cos(\theta/2)}.$$

The maximum percent error in the approximation is then

$$\begin{aligned} \gamma_{max}(r_i, \theta) &= \frac{\left| d_{prev} + r_i - \varepsilon - \sqrt{d_{prev}^2 + r_i^2 + 2r_i d_{prev} \cos(\theta/2)} \right| \times 100}{d_{prev} + r_i - \varepsilon}. \end{aligned} \quad (1)$$

Denoting the probability that the maximum single-hop distance is  $r_i = r_i$  as  $p_{r_i}(r_i)$ , the expected value of the upper bound of percentage error can be calculated by evaluating (1) for  $\theta \in [0, \alpha]$  and  $r_i \in [0, R]$ , as in (2). The derivation of  $p_{r_i}(r_i)$  is provided in Section 3.3.2

$$E[\gamma_{max}(r_i, \theta)] = 100 \int_0^\alpha \int_0^R \gamma_{max}(r_i, \theta) p_{r_i}(r_i) \frac{1}{\alpha} dr_i d\theta. \quad (2)$$

Simulation results demonstrate that the amount of error in the approximation is negligible and upper bounded. In Fig. 6, the percentage errors are all lower than 2 percent for all values of  $\alpha$ . Fig. 6a illustrates the expected value of the upper bound of the percentage error with respect to multihop distance for different values of the angular range  $\alpha$  in case of  $\sigma = 0.001$  nodes/m<sup>2</sup>. Furthermore, Fig. 6b shows the expected value of the upper bound of the percentage error for different values of the node density  $\sigma$  with respect to multihop distance for  $\alpha = \pi/3$ .

#### 3.2 Maximum Multihop Distance Approximated as a Sequence of Single-Hop Maximum Distances

To maximize the multihop distance, the maximum distances of individual single hops are considered within their angular communication ranges. Note that this is an approximation as mentioned in Section 3.1. In Fig. 7, node  $P_{i-1}$  forwards the packet to the radially furthest node  $P_i$ . Furthermore, it is known that no sensor node is located in the shaded area  $A_i$  (hence, the area  $V - i$ ) since otherwise

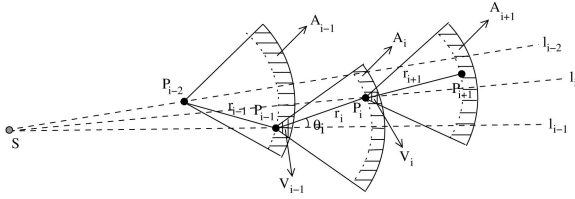


Fig. 7. Multihop distance formation.

such a node would be the furthest node and  $r_i$  would be larger. In the rest of the paper, areas which contain no sensor nodes are referred to as “vacant area.”

In Fig. 7, the next nodes of the multihop transmission at individual hops are searched within angular communication ranges that are pointed outward from the source node  $S$ . The lines  $l_{i-2}$ ,  $l_{i-1}$ , and  $l_i$  designate the bisectors of the angular slices at hops  $i-1$ ,  $i$ , and  $i+1$ , respectively. Hence, the multihop distance is established by a sequence of single-hop distances. The next section outlines the calculation of the maximum single-hop distance.

### 3.3 Analysis of Single-Hop Maximum Distance

During the analysis of single-hop maximum distance  $r_i$ , the following facts should be considered:

- The vacant region  $V_{i-1}$  is known to contain no sensors.
- $r_i$  is a random variable whose range is upper bounded by  $R$  and is independent of the angular deviation  $\theta_i$ .
- The angular deviation  $\theta_i$  is a random variable uniformly distributed over a range  $[-\frac{\alpha}{2}, \frac{\alpha}{2}]$  and is independent of  $r_i$ .

The analysis of single-hop maximum distance is different for the first hop from intermediate nodes since the first hop does not consider a vacant region  $V_0$  which does not exist by definition. However, the introduction of the vacant region  $V_{i-1}$  into the analysis of  $r_i$  is inevitable for intermediate nodes and requires some additional calculations due to the shape of  $V_{i-1}$ . Therefore, the single-hop maximum distance  $r_1$  is analyzed first, followed by the analysis of  $r_i$  for  $i = 2, 3, \dots$

#### 3.3.1 First-Hop Maximum Distance

The analysis of first-hop maximum distance  $r_1$  is the starting point of our analysis of the more general intermediate-hop distance  $r_i$ . In Fig. 8, the location of the source sensor  $S$ , which is known deterministically, is the apex of the angular slice of the first hop. The following three conditions should be satisfied for the furthest distance  $r_1$  to be equal to a specific value  $r_1$ : First, there should be a point on the circular arc  $a_1$  on which every point is  $r_1$  away from  $S$ . Second, there should be no sensors within the region

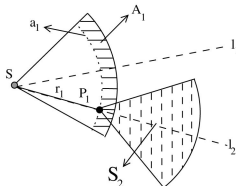
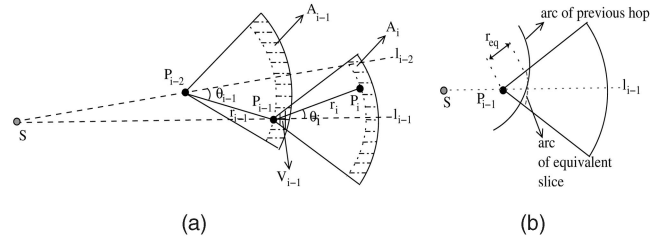


Fig. 8. Analysis of first-hop maximum distance.

Fig. 9. Analysis of intermediate-hop maximum distance. (a) Vacant region  $V_{i-1}$  and (b)  $r_{eq}$ .

designated as  $A_1$ . Third, there should be at least one point in the area  $S_2$ , as shown in Fig. 8, which is in the angular communication range of the second hop. The last condition should be met as otherwise connectivity is lost.

Since point locations are uniformly distributed, the number of points that can be found within a given area has a Poisson distribution. Furthermore, the number of points that are found in nonintersecting areas are independent random variables. Considering these facts, the probability that the three conditions are met can be defined as a function  $g = Prob(A \text{ sensor on arc } a_1) \cdot Prob(\text{No sensors in } A_1) \cdot Prob(\text{At least one sensor in } S_2)$ . For an infinitesimal radial width  $dr_1$ ,  $Prob(A \text{ sensor on arc } a_1) = \sigma \alpha r_1 dr_1$ .

Using the definition of Poisson distribution and by the approximation  $Area(S_2) \cong \frac{\alpha(R^2 - (R-r_1)^2)}{2}$ ,  $g$  becomes a function of  $r_1$  given by

$$g(r_1) = \sigma \alpha r_1 e^{-\frac{\alpha \sigma (R^2 - r_1^2)}{2}} \left( 1 - e^{-\frac{\alpha \sigma (R^2 - (R-r_1)^2)}{2}} \right) dr_1. \quad (3)$$

To find the probability distribution function of  $r_1$ , the function  $g(r_1)$  should be normalized over the range of values that  $r_1$  can take. The pdf of  $r_1$  is then given by

$$p_{r_1}(r_1) = g(r_1) / \int_0^R g(r_1) dr_1.$$

Then, the expected value  $E[r_1]$  of the first-hop maximum distance  $r_1$  can be found by the following expression:

$$E[r_1] = \int_0^R r_1 p_{r_1}(r_1) dr_1. \quad (4)$$

The variance of  $r_1$  can be calculated as  $\sigma_{r_1}^2 = E[r_1^2] - E[r_1]^2$ . Here, the second moment  $E[r_1^2]$  of  $r_1$  is determined using the following equation:

$$E[r_1^2] = \int_0^R r_1^2 p_{r_1}(r_1) dr_1. \quad (5)$$

#### 3.3.2 Intermediate-Hop Maximum Distance

The analysis of the maximum distance  $r_i$  of intermediate hops is similar to the analysis of  $r_1$ . However, the vacant regions  $V_i$  need also be considered.

In Fig. 9a, the lower bound of the range of values that  $r_i$  can take is determined by the vacant region  $V_{i-1}$ . However, the shape of  $V_{i-1}$  depends on the previous distance  $r_{i-1}$  and the previous angle of deviation  $\theta_{i-1}$ . Therefore, it is not possible to define an exact lower bound on the range values that  $r_i$  can take for given values of  $r_{i-1}$  and  $\theta_{i-1}$ . Hence, an approximate lower bound which takes the area of  $V_{i-1}$  into

account should be determined since this area contains no sensors. This is shown in Fig. 9b, where  $r_{eq}$ , called *equivalent radius*, is the radius of angular slice with the same area as  $V_{i-1}$ . The Appendix presents the calculation of  $r_{eq}$  for given values of  $r_{i-1}$  and  $\theta_{i-1}$ .

#### Calculation of intermediate-hop maximum distance.

$r_{eq}$  is the lower limit of the intermediate-hop maximum distance  $\mathbf{r}_{i-1}$  and is a function given by  $r_{eq}(r_{i-1}, \theta_{i-1})$ . Denoting  $\overline{r_{i-1}} = E[\mathbf{r}_{i-1}]$ , the pdf of  $\mathbf{r}_i$  is given as

$$g(r_i) = \sigma\alpha r_i e^{-\frac{\sigma\alpha(R^2-r_i^2)}{2}} \left(1 - e^{-\frac{\sigma\alpha(R^2-(R-r_i)^2)}{2}}\right),$$

$$p_{\mathbf{r}_i}(r_i) = \frac{g(r_i)}{\int_{r_{eq}(\overline{r_{i-1}}, \theta_{i-1})}^R g(r_i) dr_i}. \quad (6)$$

Hence, the expected value  $E[\mathbf{r}_i]$  and the second moment  $E[\mathbf{r}_i^2]$  of  $\mathbf{r}_i$  are calculated using (7) and (8), respectively. The variance of  $\mathbf{r}_i$  can be calculated with  $\sigma_{\mathbf{r}_i}^2 = E[\mathbf{r}_i^2] - E[\mathbf{r}_i]^2$ , where

$$E[\mathbf{r}_i] = \frac{1}{\alpha} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \int_{r_{eq}(\overline{r_{i-1}}, \theta_{i-1})}^R r_i p_{\mathbf{r}_i}(r_i) dr_i d\theta_{i-1}, \quad (7)$$

$$E[\mathbf{r}_i^2] = \frac{1}{\alpha} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \int_{r_{eq}(\overline{r_{i-1}}, \theta_{i-1})}^R r_i^2 p_{\mathbf{r}_i}(r_i) dr_i d\theta_{i-1}. \quad (8)$$

### 3.3.3 Last-Hop Maximum Distance

Since there is no condition that the next hop should exist for the last hop  $N$ , the pdf of the last-hop maximum distance  $\mathbf{r}_N$  is obtained as follows:

$$g(r_N) = \sigma\alpha r_N e^{-\frac{\sigma\alpha(R^2-r_N^2)}{2}},$$

$$p_{\mathbf{r}_N}(r_N) = \frac{g(r_N)}{\int_{r_{eq}(\overline{r_{N-1}}, \theta_{N-1})}^R g(r_N) dr_N}. \quad (9)$$

Using (9), the expected value and the second moment of the last hop are found, as given by (10) and (11). Note that the variance of the last-hop distance is  $\sigma_{\mathbf{r}_N}^2 = E[\mathbf{r}_N^2] - E[\mathbf{r}_N]^2$ , where

$$E[\mathbf{r}_N] = \frac{1}{\alpha} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \int_{r_{eq}(\overline{r_{N-1}}, \theta_{N-1})}^R r_N p_{\mathbf{r}_N}(r_N) dr_N d\theta_{N-1}, \quad (10)$$

$$E[\mathbf{r}_N^2] = \frac{1}{\alpha} \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \int_{r_{eq}(\overline{r_{N-1}}, \theta_{N-1})}^R r_N^2 p_{\mathbf{r}_N}(r_N) dr_N d\theta_{N-1}. \quad (11)$$

## 3.4 Analysis of Multihop Maximum Distance

This section outlines how the mean  $E[\mathbf{d}_N]$  and standard deviation  $\sigma_{\mathbf{d}_N}$  of the maximum multihop distance distribution  $p_{\mathbf{d}_N}(d_N)$  are calculated.

### 3.4.1 Expected Multihop Maximum Distance $E[\mathbf{d}_N]$

Fig. 10 illustrates the relation between the single-hop distance  $\mathbf{r}_i$  and the multihop distance  $\mathbf{d}_i$  in hop  $i$  and the multihop distance  $\mathbf{d}_{i-1}$  of the previous hop  $i-1$ . Denoting  $\theta = \{\theta_1, \theta_2, \theta_3, \dots\}$  and  $\mathbf{r} = \{\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots\}$ , we have

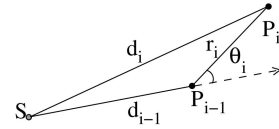


Fig. 10. Calculation of  $E[\mathbf{d}_N]$ .

$$E[\mathbf{d}_i] = E\left[\sqrt{\mathbf{d}_{i-1}^2 + \mathbf{r}_i^2 + 2\mathbf{r}_i\mathbf{d}_{i-1} \cos \theta_i}\right] = E[g(\mathbf{r}, \theta)],$$

where  $g$  is a function of  $\mathbf{r}$  and  $\theta$ . Hence,

$$E[\mathbf{d}_i] = \int_{r_{eq}(\overline{r_{i-1}}, \theta_{i-1})}^R \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \dots \int_{r_{eq}(\overline{r_1}, \theta_1)}^R \int_{-\frac{\alpha}{2}}^{\frac{\alpha}{2}} \mathbf{f}(\mathbf{r}, \theta) d\theta_1 dr_1 \dots d\theta_i dr_i, \quad (12)$$

where  $\mathbf{f}(\mathbf{r}, \theta) = g(\mathbf{r}, \theta) \cdot p_{\mathbf{r}, \theta}(\mathbf{r}, \theta)$  and  $p_{\mathbf{r}, \theta}(\mathbf{r}, \theta)$  is the joint probability distribution function.

The computation of the right-hand side of (12) is costly and not scalable in number of hops. One possible approximation is the representation of  $E[\mathbf{d}_i]$  as an iterative relation dependent on the expected value  $E[\mathbf{d}_{i-1}] = \overline{d_{i-1}}$  of the previous hop. This representation is given by

$$E[\mathbf{d}_i] = E\left[\sqrt{\overline{d_{i-1}}^2 + \mathbf{r}_i^2 + 2\mathbf{r}_i\overline{d_{i-1}} \cos \theta_i}\right].$$

Denoting

$$\mathbf{s}(\overline{d_{i-1}}, r_i, \theta_i) = p_{\mathbf{r}_i}(r_i) \sqrt{\overline{d_{i-1}}^2 + r_i^2 + 2r_i\overline{d_{i-1}} \cos \theta_i},$$

the computation of  $E[\mathbf{d}_N]$  is outlined iteratively as follows:

1.  $\overline{d_1} = \overline{r_1} = E[\mathbf{r}_1]$ , as given by (4).
2.  $E[\mathbf{r}_i] = \overline{r_i}$  by (7) for  $1 < i < N$ .
3. For  $1 < i < N$ ,

$$\overline{d_i} = \frac{1}{\alpha^2} \int_{-\alpha/2}^{\alpha/2} \int_{r_{eq}(\overline{r_{i-1}}, \theta_{i-1})}^R \int_{-\alpha/2}^{\alpha/2} \mathbf{s}(\overline{d_{i-1}}, \mathbf{r}_i, \theta_i) d\theta_i dr_i d\theta_{i-1}.$$

4. Expected multihop distance is then found by

$$\overline{d_N} = \frac{1}{\alpha^2} \int_{-\alpha/2}^{\alpha/2} \int_{r_{eq}(\overline{r_{N-1}}, \theta_{N-1})}^R \int_{-\alpha/2}^{\alpha/2} \mathbf{s}(\overline{d_{N-1}}, \mathbf{r}_N, \theta_N) d\theta_N dr_N d\theta_{N-1}.$$

### 3.4.2 Standard Deviation $\sigma_{\mathbf{d}_N}$ of Multihop Maximum Distance

To compute the variance  $\sigma_{\mathbf{d}_i}^2 = E[\mathbf{d}_i^2] - E[\mathbf{d}_i]^2$ , the moment  $E[\mathbf{d}_i^2]$  can be found by

$$E[\mathbf{d}_i^2] = E[\mathbf{d}_{i-1}^2] + E[\mathbf{r}_i^2] + 2E[\mathbf{r}_i\mathbf{d}_{i-1} \cos \theta_i].$$

Since  $\theta_i$  is independent of  $\mathbf{d}_{i-1}$  and  $\mathbf{r}_i$ ,  $E[\mathbf{r}_i\mathbf{d}_{i-1} \cos \theta_i] = E[\mathbf{r}_i\mathbf{d}_{i-1}] \overline{\cos \theta_i}$ , where  $\overline{\cos \theta_i}$  is the expected value of  $\cos \theta_i$  calculated over the range  $\theta_i \in [-\alpha/2, \alpha/2]$  and is equal to  $\overline{\cos \theta_i} = \frac{2}{\alpha} \sin \frac{\alpha}{2}$ . Hence, we obtain

$$E[\mathbf{d}_i^2] = E[\mathbf{d}_{i-1}^2] + E[\mathbf{r}_i^2] + 2E[\mathbf{r}_i\mathbf{d}_{i-1}] \frac{2}{\alpha} \sin \frac{\alpha}{2}, \quad (13)$$

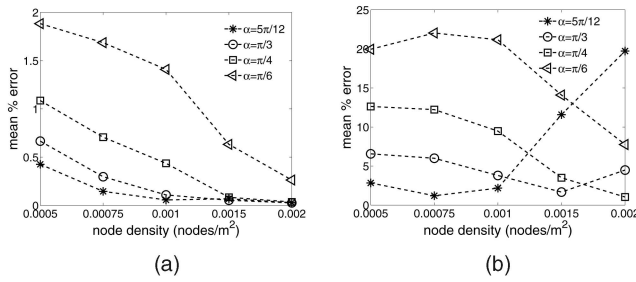


Fig. 11. Average percent error in approximation of  $E[d_N]$  and  $\sigma_{d_N}$ . (a) Percent error in  $d_N$  and (b) percent error in  $\sigma_{d_N}$ .

where  $E[r_i d_{i-1}] = Cov(r_i, d_{i-1}) + E[r_i]E[d_{i-1}]$ . Since the correlation of  $r_i$  with  $r_j$  for  $j < i - 1$  is negligible compared to the correlation of  $r_i$ , we have

$$E[d_i^2] = E[d_{i-1}^2] + E[r_i^2] + 2(Cov(r_i, r_{i-1}) + E[r_i]E[d_{i-1}]) \frac{2}{\alpha} \sin \frac{\alpha}{2}. \quad (14)$$

Finally, the iterative computation of  $E[d_N^2]$  is outlined as follows:

1.  $E[r_1^2] = E[d_1^2]$  by (5).
2.  $E[r_i^2]$  is found by (8).
3.  $E[d_i^2]$  is calculated using (14).
4.  $E[r_N^2]$  is found by (11).
5.  $E[d_N^2]$  is computed by substituting  $E[r_N^2]$  and  $E[d_{N-1}^2]$  in (14).

### 3.5 Approximation Accuracy for Different Angles $\alpha$

Before moving to the performance analysis of the derived approximation expressions, the effect of  $\alpha$  on the accuracy of the derived expressions is evaluated. Fig. 11a illustrates the average percent error in  $E[d_N]$ . For each value of  $\alpha$ , the error of  $E[d_N]$  approximation decreases with increasing node density and settles to a value less than 0.2 percent, except for  $\alpha = \pi/6$  which requires a larger node density to stabilize. One obvious observation is the decreasing amount of percent error in  $E[d_N]$  approximation for larger angles, which shows that our approximation is more accurate when single hops cover larger angular ranges.

The choice of a large angle, however, causes instability in the approximation of the standard deviations shown in Fig. 11b. Despite the decrease in the percent error, the percentage error of  $\sigma_{d_N}$  has a noticeable incline when  $\alpha = 5\pi/12$  for node densities larger than  $\sigma = 0.001$  nodes/m<sup>2</sup>. The increasing error for large angles is due to the multiplication of the two approximated values of  $E[d_{i-1}r_i]$  and  $\cos\theta_i$  in (13). Although these two terms are approximated with high accuracies of the order of 2.5 and 0.62 percent, respectively, their multiplication creates significant terms when  $\alpha$  and  $\sigma$  are large.

With a less than 2 percent error in approximated  $E[d_N]$  and a reasonably low error of 5 percent in  $\sigma_{d_N}$  estimation, we choose the slice angle as  $\alpha = \pi/3$  for the practical range of node density values used in our numerical studies. The subsequent performance analysis of the approximations are based on the multihop distances with  $\alpha = \pi/3$ .

### 3.6 Multihop Distances with Shadow Fading Model

The hop distances to a broadcast source in a planar sensor network depends on the positions of the sensors as well as

their transmission and sensing range patterns. The unit disk model has been widely used in the current literature for modeling wireless sensor networks and designing communication protocols [4], [6], [3], [16]. The unit disk model defines the communication range as the minimum radius of a circular reception area within which all transmissions are successfully received if no interference or packet collisions exist. In the event that the wireless medium is subject to the effects of fading, the reception power at receiver nodes is affected by the distance to the transmitter and decays exponentially with distance. Furthermore, with the presence of Gaussian noise, the received power becomes a random variable. This makes the reception of a packet a probabilistic event dependent on the distance to the transmitter node, the statistical characteristics of the channel noise, transmission power, and the threshold of reception power. Using the Shadow Fading Model with  $n$  denoting the path loss component, the path loss  $PL(d)$  at a distance  $d$  from the transmitter is given by [15]

$$PL(d) = PL(d_0) + 10.n.\log\left(\frac{d}{d_0}\right) + X_\sigma, \quad (15)$$

where  $d_0$  is the reference distance and  $X_\sigma$  is a zero-mean Gaussian random variable. Considering this path loss and denoting the received power at distance  $d$  as  $P_R(d)$ , the probability that a packet is successfully received by a sensor at a distance  $d$  away from the transmitter node,  $Pr(d)$ , is given by

$$Q(z) = \frac{1}{2\pi} \int_z^\infty \exp\left(-\frac{x^2}{2}\right) dx, \quad (16)$$

$$Q(z) = 1 - Q(-z),$$

$$Pr(d) = Pr[P_R(d) > \gamma] = Q\left[\frac{\gamma - Pr(d)}{\sigma}\right], \quad (17)$$

where  $\gamma$  is the receive threshold and (16) is the Q-function. When the packet transmissions occur in an environment subject to random path loss, hop distance patterns are also affected by this change. Since now the nodes may not receive a packet with insufficient reception power, the maximum multihop distance that a packet can traverse in a given number of hops is effectively reduced. Hence, the greedy multihop maximization model, as explained in Section 3, has to consider this change. This can be achieved by including the probability of packet reception given by (17) in (3), (6), and (9). Hence, (3) now becomes:

$$g(r_1) = Pr(r_1) \cdot \sigma \alpha r_1 e^{-\frac{\sigma \alpha (R^2 - r_1^2)}{2}} \left(1 - e^{-\frac{\sigma \alpha (R^2 - (R - r_1)^2)}{2}}\right) dr_1, \quad (18)$$

where  $Pr(r_1)$  is found by (17). Note that  $Pr(d)$  is, in fact, the probability of reception as provided by *any* fading model which is a function of communication radius, not necessarily the Shadow Fading Model. Hence, our model captures the effect of a given fading model by a multiplicative probability factor in distance probability functions in (3), (6), and (9).

## 4 ESTIMATION OF HOP DISTANCE

Apart from the estimation of coverage area boundaries as shown in previous section, an interesting application of the

maximum euclidean distance distribution is the estimation of hop distance between two locations. In fact, hop distance is usually a more popular metric in WSN research since parameters like multihop end-to-end delay and total transmission power can be estimated by hop distance. In this section, hop distance  $\mathbf{N}$  for a given euclidean distance  $D$  is estimated using the Gamma pdf Model of maximum multihop distance distribution  $d_i$ , where  $i = 1, 2, 3, \dots$

The posterior distribution of  $\mathbf{N}$  for a given  $D$  is

$$P(\mathbf{N}|D) = \frac{P(\mathbf{D}|\mathbf{N}) \cdot P(\mathbf{N})}{\sum_{i=1}^{N_{max}} P(\mathbf{D}|i) \cdot P(i)}, \quad (19)$$

where  $P(i)$  is the probability that a randomly selected node is  $i$  hops away from the source node.

The maximum a posteriori (MAP) estimate of  $N$  is

$$\hat{N}_{MAP} = \arg \max_N \left( \frac{P(\mathbf{D}|\mathbf{N}) \cdot P(\mathbf{N})}{\sum_{i=1}^{N_{max}} P(\mathbf{D}|i) \cdot P(i)} \right). \quad (20)$$

The probability  $P(i)$  can be approximated by calculating the area where nodes with hop distance  $i$  are located and then taking the ratio of this area to the area of the whole network. Assuming that a maximum hop distance of  $N_{max}$  is created as a result of the broadcast, we have

$$P(i) \cong \frac{\pi(d_i^2 - d_{i-1}^2)}{\pi d_{N_{max}}^2}. \quad (21)$$

Hence, (20) is modified to obtain:

$$\hat{N}_{MAP} = \arg \max_N \left( \frac{P(\mathbf{D}|\mathbf{N}) \cdot (d_N^2 - d_{N-1}^2)}{\sum_{i=1}^{N_{max}} P(\mathbf{D}|i) \cdot (d_i^2 - d_{i-1}^2)} \right). \quad (22)$$

Note that the value of  $N_{max}$  depends on factors like topology size, node locations, sensor communication range, and node density.  $N_{max}$  is lower bounded by  $\lfloor D_M/R \rfloor$ , where  $D_M$ , say, is the maximum distance to the source sensor found in the topology, which is possible in case of infinite node density. Since node density is finite and connectivity patterns vary for different topologies, an exact value of  $N_{max}$  is difficult to estimate. However, in (22),  $N_{max}$  is canceled out when taking the ratio of  $P(N)$  and  $P(i)$ , yet it still appears as an upper bound of the summation. The pdf  $P(d_i)$  has low tails as observed in Fig. 2c, and the proportion of the pdf terms for higher hop distances in the summation diminishes quickly and become negligibly small as  $i$  increases. Hence, for computational purposes,  $N_{max}$  in (22) can be chosen as an arbitrarily large number.

Equation (22) provides the MAP estimation of the hop distance for a given euclidean distance  $D$ . However, the posterior probability  $P(\mathbf{D}|\mathbf{N})$  is unknown for individual values of hop distance  $i$ . On the other hand, the distribution of  $d_i$  for  $i = 1, 2, 3, \dots$  is modeled by the Gamma pdf and can be used to calculate  $P(\mathbf{D}|\mathbf{N})$  by meeting the following requirements [13]:

1.  $D - R \leq d_{N-1} < D$ .
2.  $0 < d_j < D - R, \forall j < N - 1$ .
3. There must be at least one point in the area  $A$ , as shown in Fig. 12.

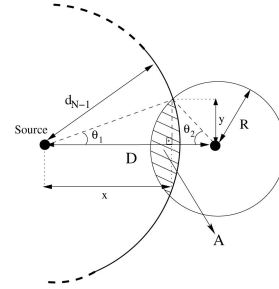


Fig. 12. Area A should contain a node.

Hence,  $P(\mathbf{D}|\mathbf{N})$  is equal to

$$P(\mathbf{D}|\mathbf{N}) = \text{Prob}\{d - R \leq d_{N-1} < D\} \cdot \text{Prob}\{0 < d_j < D - R\} \cdot (1 - e^{-\sigma A}). \quad (23)$$

Area  $A$  is calculated using the following equations:

$$\begin{aligned} x &= \frac{D^2 + d_{i-1}^2 - R^2}{2D}, \\ y &= \sqrt{d_{i-1}^2 - x^2}, \\ \theta_1 &= \left| \arctan \left( \frac{y}{D-x} \right) \right|, \\ \theta_2 &= \left| \arctan \left( \frac{y}{x} \right) \right|, \\ A &= \theta_1(R^2) - y(D-x) + \theta_2 d_{i-1}^2 - yx. \end{aligned} \quad (24)$$

Hence,  $P(\mathbf{D}|\mathbf{N})$  is found by

$$P(\mathbf{D}|\mathbf{N}) = \int_{d-R}^D p_{d_{N-1}}(d_{N-1}) dd_{N-1} \cdot \left( \prod_{j=1}^{N-2} \int_0^{D-R} p_{d_j}(d_j) dd_j \right) \cdot (1 - e^{-\sigma A}). \quad (25)$$

Finally, (25) is plugged in (22) to calculate the MAP estimation of hop distance for a given euclidean distance  $D$ .

## 5 NUMERICAL RESULTS

In this section, the results of our analytical approximations for the expected value and standard deviation of the maximum euclidean distance  $d_N$  are presented and compared with the simulation results. Furthermore, the distribution of  $d_N$  is approximated using a transformed Gamma distribution for various node densities. The simulation results are the averages of 10,000 independent topologies for all values of the node density,  $\sigma = 0.0005, 0.00075, 0.001, 0.0015$ , and  $0.002$  nodes/m<sup>2</sup>. In each simulation, the source sensor is placed in the middle of a topology of size  $1,600 \text{ m} \times 1,600 \text{ m}$ . The communication range of sensor nodes is  $R = 100 \text{ m}$  and the angular range is  $\alpha = \pi/3$ . The simulations are carried out by a broadcast from the source sensor to the network and determining the hop distances of individual nodes to the source. For each hop distance, the greedy propagation locates the node with the maximum distance to the previously selected node toward the direction of propagation as outlined in Section 2.2. The separation between the source sensor and this new node is recorded as the multihop distance for that particular hop distance in that simulation.



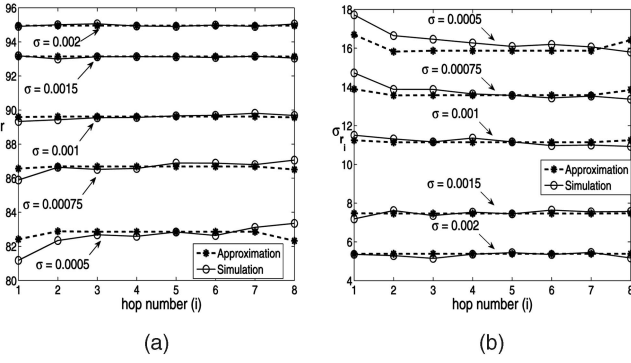


Fig. 13. Approximation of  $E[r_i]$  and  $\sigma_{r_i}$ . (a) Approximation of  $E[r_i]$  and (b) approximation of  $\sigma_{r_i}$ .

### 5.1 Mean and Standard Deviation of $r_i$

Fig. 13a illustrates the comparison between the simulated and analytical expected single-hop maximum distance  $E[r_i]$ . The increase in the accuracy of the approximations with increasing node density can be observed. Furthermore,  $E[r_i]$  becomes larger with the increase in node density. The approximated single-hop distances are close to the simulation results and the accuracy is enhanced for higher node densities.

The approximation result of the single-hop standard deviation  $\sigma_{r_i}$  is shown in Fig. 13b. With larger node density, the accuracy of the approximation of  $\sigma_{r_i}$  increases while the magnitude of  $\sigma_{r_i}$  decreases.

### 5.2 Mean and Standard Deviation of $d_N$

Fig. 14a illustrates the comparison between analytical and simulated  $E[d_N]$  in case of different node density values  $\sigma$ . It can be observed that  $E[d_N]$  monotonically increases for increasing node density. Furthermore, the approximations closely match the simulation values with negligible errors which become even smaller for higher node densities.

The standard deviation of  $d_N$  is shown in Fig. 14b as an error bar around the expected value. In this figure, only the highest and the lowest density values used in the simulations are shown for better visualization of the error bars. The accuracy of the approximated  $\sigma_{d_N}$  is higher for the higher node density. Furthermore, the increase in standard deviation for increasing hop distance  $N$  and lower node density can be observed.

### 5.3 Approximation of the pdf of $d_N$

The pdf of the maximum euclidean distance  $d_N$  corresponding to a hop distance  $N$  is approximated by a “transformed”

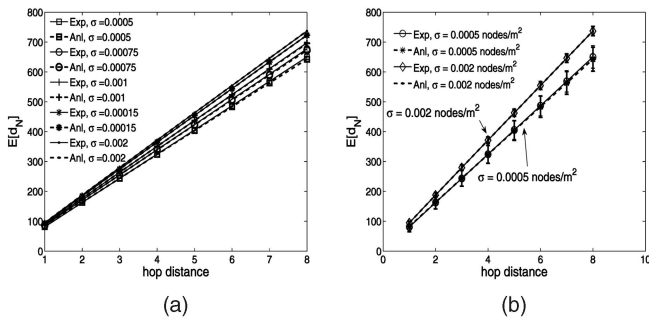


Fig. 14. Approximation of  $E[d_N]$  and  $\sigma_{d_N}$ . (a) Approximation of  $E[d_N]$  and (b) approximation of  $\sigma_{d_N}$ .

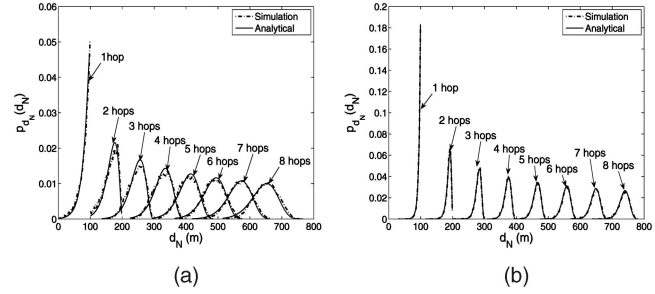


Fig. 15. Approximation of  $p_{d_N}(d_N)$  for low and high density. (a)  $\sigma = 0.0005$  nodes/m<sup>2</sup> and (b)  $\sigma = 0.002$  nodes/m<sup>2</sup>.

Gamma distribution, as explained in Section 2.3. After calculating  $E[d_N]$  and  $\sigma_{d_N}$ , the parameters  $a$  and  $b$  of the Gamma distribution of the remaining distance  $x$  are calculated. Then, transformation  $d_i = iR - x$  is applied to get the analytical approximation of the distribution of  $d_N$ .

In Fig. 15, our approximation of multihop distance distribution is illustrated for low and high node density values  $\sigma = 0.0005$  nodes/m<sup>2</sup> and  $\sigma = 0.002$  nodes/m<sup>2</sup>, respectively. The distribution of  $d_N$  becomes more peaked with increasing node density and the pdf curves corresponding to different values of hop distance  $N$  get more separated from each other due to the decrease in  $\sigma_{d_N}$  caused by larger node density.

The comparison of the 1D Model [10], [12] and the proposed 2D Model is demonstrated in Fig. 16. In this figure, the root-mean-square error (RMSE) between the model pdf and the distribution obtained by simulations is shown. Different corridor widths ( $W$ ) are evaluated for the 1D Model. In most cases, the least RMSE is obtained for  $W = R$ . It is observed that the RMSE of the 2D Model is less than the RMSE of the 1D Model regardless of the corridor widths  $W$ . Furthermore, the RMSE of the 2D Model is more invariant to changing hop distance values when compared with the RMSE of 1D Model. In summary, Fig. 16 clearly illustrates that our proposed 2D model is a

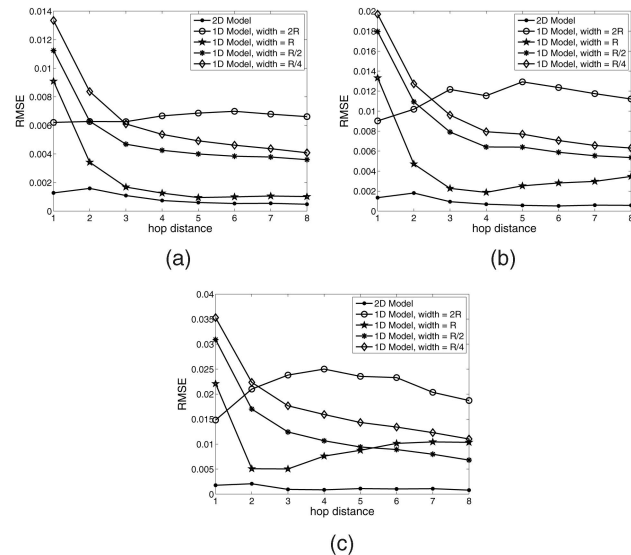


Fig. 16. Root-mean-square error of 1D and 2D Models. (a)  $\sigma = 0.0005$  nodes/m<sup>2</sup>, (b)  $\sigma = 0.001$  nodes/m<sup>2</sup>, and (c)  $\sigma = 0.002$  nodes/m<sup>2</sup>.

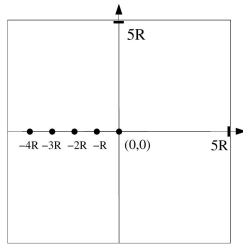


Fig. 17. Topology shape and source locations.

more accurate representation of the planar multihop maximum distance distribution.

### 5.4 Edge Effects

The analysis in Section 2 reveals the relationship between hop distance and euclidean distance in the greedy multihop forwarding mechanism, as outlined in Section 2.2. In this section, we study the effect of topology borders on the proposed estimation method.

Topology borders limit the distance that can be taken in a multihop path. For an  $n$ -hop path (if it exists) toward a border, the search area of the final hop (hop  $n$ ) may not be sufficiently large to form an angular slice, as defined in Section 2.2. This is an *edge effect* which reduces the multihop path distance.

The edge effect is not observed on an  $n$ -hop path when the distance to the border is sufficiently large. For a communication range of  $R$ , a distance of  $nR$  from the source node to the border ensures that the angular slice search area of each hop of the greedy multihop forwarding can be formed without being limited by the border. Hence, scenarios with source node located at the center of topologies of size  $2nR \times 2nR$  avoid the edge effect on paths of length  $n$  and less.

The average multihop distance is determined via simulations for each hop distance value  $n$ . To find the average distance in  $n$  hops for a selected source node location, 2,000 independent topologies are formed, and for each topology, a single sample multihop path of  $n$  hops is located. This path is determined by choosing a random initial direction (as outlined in Section 2.2. Note that if this random initial direction does not produce a multihop path of length  $n$ , we aggressively scan the whole area for an angle that produces an  $n$ -hop path for that particular simulation topology. With this setting, some of the  $n$ -hop paths experience distance limitation by the border since the distance of the source node to the border is not guaranteed to be sufficiently large to avoid the edge effect. Furthermore, note that with a random choice of propagation direction, not all the  $n$ -hop paths experience a distance reduction. However, on average, the effect is a decrease in the expected multihop euclidean distance of a randomly chosen  $n$ -hop path. In the simulations, it is observed that the reduction in the multihop euclidean distance is largely caused by the decrease in the distance taken in the final hop under the edge effects.

In Fig. 18, we study the effect of edges on the multihop distance estimation in terms of the average of *percent error* defined by  $\% \text{ error} = 100 \times \frac{|a-e|}{a}$ , where  $a$  and  $e$  denote an

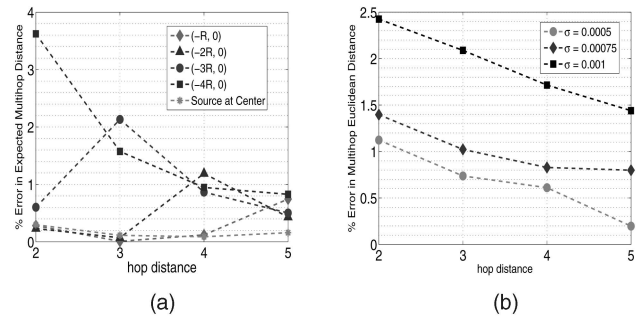


Fig. 18. Percent error in expected multihop distance estimation due to the edge effect. Topology size:  $5R \times 5R$ . Different source locations. (a) Source node approaching an edge and (b) random source locations, different node density values.

analytical and a simulation result for a multihop distance value, respectively. First, the source sensor is placed away from the topology center and at increasingly closer locations toward one of the edges. For a topology of size  $5R \times 5R$ , these locations are arbitrarily chosen to be at coordinates  $(-4R, 0)$ ,  $(-3R, 0)$ ,  $(-2R, 0)$ , and  $(-R, 0)$ , where  $(0, 0)$  is the topology center. Fig. 17 demonstrates the topology shape and the chosen source node locations. Fig. 18a shows the average percent error in multihop distance estimation with this set of source locations at topologies of node density  $\sigma = 0.001$  nodes/m<sup>2</sup>. Note that for each source location value, a new set of 2,000 independent topologies of random node locations are formed. For each topology, a single-sample multihop path is selected for each hop distance  $n$ . Second, we place the source node at randomly selected locations and vary the node density. Similarly, we form 2,000 independent topologies for each node density value. The average percent error results are illustrated in Fig. 18b.

In Fig. 18a, when the source node is sufficiently close to the edge, the percent error graphs make a peak. For instance, the graph for source location  $(-2R, 0)$  has a peak at hop distance 4. Note that the shortest distance of this location to the edge is  $3R$ . Paths of length 3 and less are not affected by the edge and have comparable percent error with those of topologies with source node at the center. However, for 4-hop paths, we observe a peak in percent error as some of these paths experience a distance reduction in their final (4th) hops. On the other hand, it is less likely that paths of higher hop distance are formed toward the edge at this proximity to the edge; hence, longer paths can be established in directions away from the edge. Therefore, the average of the percent error in distance estimation is less for higher values of  $n$ . Similar trends are observed at other source locations, although the peak of the graph is observed at a smaller hop distance value for locations closer to the edge. Another observation in Fig. 18a is that the average percent error value of the peak location is higher when the source is closer to the edge. As mentioned before, the peak location is found at a smaller hop distance when the source node is closer to the edge. Furthermore, the edge effect is predominantly on the final hop a multihop path and reduces its span, with an amount of reduction comparable for all values of hop distance. Hence, for a smaller hop distance, the corresponding euclidean distance is smaller and the ratio between the distance reduction and the path distance gets

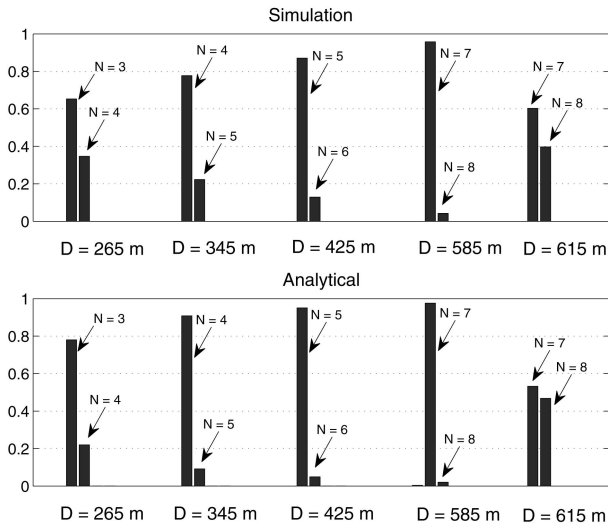


Fig. 19. Estimation of hop distance pmf for  $\sigma = 0.001$  nodes/m<sup>2</sup>.

higher, producing a higher percent error in distance estimation. (Recall that the estimated multihop distance is close to the multihop value as obtained without the edge effects.)

In Fig. 18b, three different node density values are tested to study the edge effect on distance estimation. The results demonstrate that for a smaller node density, the edge effect is less pronounced. This is an expected result since the edge effect reduces the final hop distance of a multihop path, which has a stronger limitation on higher densities with larger single-hop spans. As the node density gets smaller, the node with the maximum distance in the final hop is located closer to the most recently selected node (as outlined in Section 2.2) and its location is limited less frequently by the topology border. The diminishing character in the average percent error values is caused by the decrease in the ratio between the amount of distance in the final hop to the multihop path distance as the hop distance increases.

### 5.5 Approximation of Hop Distance pmf, $P(N|D)$

The MAP estimator of hop distance  $N$  for given euclidean distance  $D$  presented in Section 4 and given by (25) is studied for randomly selected values of  $D$ . Figs. 19 and 20 illustrate the probability mass function obtained by simulations and (19) for randomly deployed networks with uniform node density values  $\sigma = 0.001$  nodes/m<sup>2</sup> and  $\sigma = 0.002$  nodes/m<sup>2</sup>, respectively. The results show that the most likely hop distance is estimated accurately for all euclidean distances and for both node density values. Furthermore, there are two major hop distances that are observed for each euclidean distance  $D$  when node density is  $\sigma = 0.001$  nodes/m<sup>2</sup>, whereas only a single-hop distance has the dominant probability value in the pmf when the node density is increased to  $\sigma = 0.002$  nodes/m<sup>2</sup> indicating a more deterministic distribution.

## 6 CONCLUSION

Determination of the maximum multihop euclidean distance corresponding to a given hop distance in a 2D network is a complex problem. In this paper, we propose a

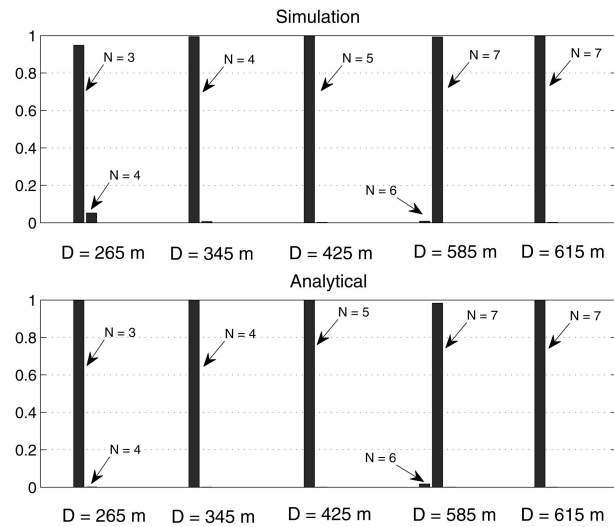


Fig. 20. Estimation of hop distance pmf for  $\sigma = 0.002$  nodes/m<sup>2</sup>.

greedy distance maximization model which approximates the maximum multihop euclidean distance in planar networks. We demonstrated that the Gaussian pdf used to model the maximum multihop distance distribution in 1D networks does not accurately represent the planar distance distribution. Using maximum likelihood estimation of distance distributions, we conjectured that the maximum distance can be accurately modeled by a transformation of the Gamma distribution. Furthermore, we provided expressions for the expected value and standard deviation of this distribution, which can be used to compute the parameters of the distribution and defines it uniquely. In order to investigate the accuracy of the model, extensive simulations are made. Distribution results are also used to estimate hop distance in a planar network using an MAP estimator.

## APPENDIX

### CALCULATION OF $r_{eq}$

Derivation of the equivalent radius  $r_{eq}$  at a hop  $i$  is carried out according to Fig. 21 in Cartesian coordinates and

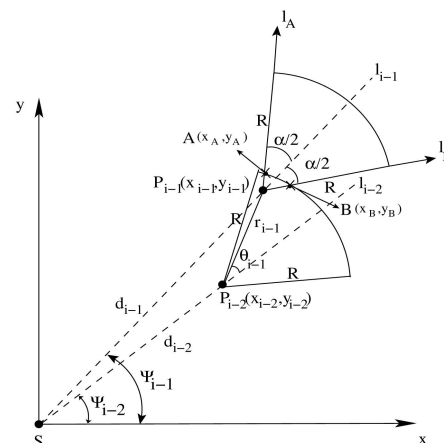


Fig. 21. Calculation of equivalent radius  $r_{eq}$ .

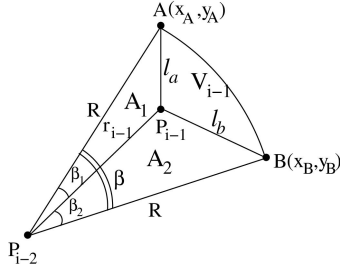


Fig. 22. Calculation of the area of the vacant region  $V_{i-1}$ .

requires determining the area of the vacant region  $V_{i-1}$  for the given values of  $r_{i-1}$  and  $\theta_{i-1}$ .

The  $x$ - and  $y$ -axes lines in (21) are any two lines that are perpendicular to each other passing through the source sensor node  $S$  which is the origin of the Cartesian coordinate system.  $l_{i-1}$  and  $l_{i-2}$  are the bisector lines of the angular communication ranges of nodes  $P_{i-1}$  and  $P_{i-2}$ , respectively. Furthermore, the lines  $l_A$  and  $l_B$  designate the angular range of the sensor  $P_{i-1}$ . Points  $A$  and  $B$  are the intersection points of the circular arc of the communication range of  $P_{i-2}$  with  $l_A$  and  $l_B$ , respectively.

As can be observed in Fig. 21, the area of the vacant region is defined by the circular arc of the communication range of  $P_{i-2}$  and the points  $A(x_A, y_A)$ ,  $B(x_B, y_B)$ , and  $P_{i-1}(x_{i-1}, y_{i-1})$ . Hence, first, the Cartesian coordinates of the point  $P_{i-1}(x_{i-1}, y_{i-1})$  are calculated by  $x_{i-1} = x_{i-2} + r_{i-1} \cos(\Psi_{i-2} + \theta_{i-1})$ , and  $y_{i-1} = y_{i-2} + r_{i-1} \sin(\Psi_{i-2} + \theta_{i-1})$ , where  $\Psi_{i-2} = \arctan(\frac{y_{i-2}}{x_{i-2}})$ . This requires the coordinate information  $P_{i-2}(x_{i-2}, y_{i-2})$ . Note that the choice of the  $x$ - and  $y$ -axes lines only makes a rotation of all sensor locations around the origin  $S$  without changing their relative positions. Hence, as long as the distance  $d_{i-2}$  between  $P_{i-2}$  and  $S$  is preserved, the choice of the axis lines determines the Cartesian coordinates  $(x_{i-2}, y_{i-2})$  and vice versa. Therefore, we can choose any pair of coordinates  $(x_{i-2}, y_{i-2})$  that satisfies the distance expression  $d_{i-2} = \sqrt{x_{i-2}^2 + y_{i-2}^2}$ .

Second, the coordinates  $(x_A, y_A)$  and  $(x_B, y_B)$  are calculated using the equations of the lines  $l_A$  and  $l_B$  and the circular communication range of the sensor  $P_{i-2}$ . This calculation requires the coordinates of  $P_{i-1}(x_{i-1}, y_{i-1})$  as well as the angle  $\Psi_{i-1}$  given by  $\Psi_{i-1} = \Psi_{i-2} + \arccos(\frac{d_{i-2} + r_{i-1} \cos \theta_{i-1}}{d_{i-1}})$ , where distance  $d_{i-1}$  is calculated by

$$d_{i-1} = \sqrt{d_{i-2}^2 + r_{i-1}^2 + 2r_{i-1}d_{i-2} \cos \theta_{i-1}}.$$

Fig. 22 outlines the calculation of the area of the vacant region  $V_{i-1}$  using the coordinates of the points  $A(x_A, y_A)$ ,  $B(x_B, y_B)$ ,  $P_{i-1}(x_{i-1}, y_{i-1})$ , and  $P_{i-2}(x_{i-2}, y_{i-2})$ . As can be seen in this figure, areas of  $A_1$  and  $A_2$ , and the area of the region  $V_{i-1}$  add up to the area of the angular slice defined by the points  $P_{i-2}$ ,  $A$ , and  $B$ . Hence, the areas  $A_1$  and  $A_2$  are first determined and then their sum is subtracted from the area of the angular slice to find the area  $V_{i-1}$ . First, the lengths  $l_a$  and  $l_b$  are calculated. Then, using the definition of the semimeter  $s$  of a triangle and Heron's formula [14], the areas  $A_1$  and  $A_2$  are calculated by

$$Area(A_1) = \sqrt{s_A(s_A - R)(s_A - r_{i-1})(s_A - l_a)}$$

and

$$Area(A_2) = \sqrt{s_B(s_B - R)(s_B - r_{i-1})(s_B - l_b)},$$

where  $s_A = \frac{1}{2}(l_a + R + r_{i-1})$  and  $s_B = \frac{1}{2}(l_b + R + r_{i-1})$ . The area of the angular slice defined by points  $P_{i-2}$ ,  $A$ , and  $B$  in Fig. 22 is equal to  $Area(Slice(P_{i-2}, A, B)) = \frac{\beta R^2}{2}$ , where

$$\beta = \beta_1 + \beta_2 = \arccos\left(\frac{R^2 + r_{i-1}^2 - l_a^2}{2Rr_{i-1}}\right) + \arccos\left(\frac{R^2 + r_{i-1}^2 - l_b^2}{2Rr_{i-1}}\right).$$

Hence, the area of the vacant region  $V_{i-1}$  and the equivalent radius  $r_{eq}$  is given by

$$r_{eq} = \sqrt{\frac{2Area(V_{i-1})}{\alpha}}$$

and  $Area(V_{i-1})$  is given as

$$Area(V_{i-1}) = \frac{\beta R^2}{2} - \sqrt{s_A(s_A - R)(s_A - r_{i-1})(s_A - l_a)} - \sqrt{s_B(s_B - R)(s_B - r_{i-1})(s_B - l_b)}.$$

## REFERENCES

- [1] D. Niculescu and B. Nath, "DV Based Positioning in Ad Hoc Networks," *Telecomm. Systems*, vol. 22, nos. 1-4, pp. 267-280, 2003.
- [2] C. Savarese, J. Rabay, and K. Langendoen, "Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks," *Proc. USENIX Ann. Technical Conf.*, pp. 317-327, 2002.
- [3] L.E. Miller, "Distribution of Link Distances in a Wireless Network," *J. Research of the Nat'l Inst. of Standards and Technology*, vol. 106, pp. 401-412, 2001.
- [4] S. De, A. Caruso, T. Chaira, and S. Chessa, "Bounds on Hop Distance in Greedy Routing Approach in Wireless Ad Hoc Networks," *Int'l J. Wireless and Mobile Computing*, vol. 1, pp. 131-140, 2006.
- [5] M. Haenggi, "On Distances in Uniformly Random Networks," *IEEE Trans. Information Theory*, vol. 51, no. 10, pp. 3584-3586, Oct. 2005.
- [6] C. Bettstetter and J. Eberspächer, "Hop Distances in Homogeneous Ad Hoc Networks," *Proc. IEEE Vehicular Technology Conf. (VTC '03) Spring*, pp. 2286-2290, Apr. 2003.
- [7] S.A.G. Chandler, "Calculation of Number of Relay Hops Required in Randomly Located Radio Network," *Electronics Letters*, vol. 25, pp. 770-777, July 1989.
- [8] R. Vilzmann, C. Bettstetter, D. Medina, and C. Hartmann, "Hop Distances and Flooding in Wireless Multihop Networks with Randomized Beamforming," *Proc. Int'l Workshop Modeling Analysis and Simulation of Wireless and Mobile System (MSWiM)*, pp. 20-27, 2005.
- [9] Y. Cheng and T. Robertazzi, "Critical Connectivity Phenomena in Multihop Radio Models," *IEEE Trans. Comm.*, vol. 37, no. 7, pp. 770-777, July 1989.
- [10] S. Vural and E. Ekici, "Analysis of Hop-Distance Relationship in Spatially Random Sensor Networks," *Proc. ACM MobiHoc*, pp. 320-331, Mar. 2005.
- [11] L. Zhao and Q. Liang, "Hop-Distance Estimation in Wireless Sensor Networks with Applications to Resources Allocation," *EURASIP J. Wireless Comm. and Networking*, 2007.
- [12] E. Ekici, J. McNair, and D. Al Abri, "A Probabilistic Approach to Location Verification in Wireless Sensor Networks," *Proc. IEEE Int'l Conf. Comm. (ICC '06)*, pp. 3485-3490, June 2006.
- [13] S. Vural and E. Ekici, "Probability Distribution of Multi-Hop-Distance in One-Dimensional Sensor Networks," *Computer Networks: The Int'l J. Computer and Telecomm. Networking*, pp. 3727-3749, Sept. 2007.

- [14] E. Maor, *The Pythagorean Theorem: A 4,000-Year History*, pp. 131-132. Princeton Univ. Press, 2007.
- [15] T.S. Rappaport, *Wireless Communications Principles and Practice*. Prentice Hall, 2002.
- [16] L.E. Miller, "Probability of a Two-Hop Connection in a Random Mobile Network," *Proc. Conf. Information Sciences and Systems*, Mar. 2001.
- [17] S. Vural and E. Ekici, "Wave Addressing for Dense Sensor Networks," *Proc. Second Int'l Workshop Sensor and Actor Network Protocols and Applications (SANPA '04)*, pp. 56-66, Aug. 2004.



**Serdar Vural** received the BS degree from the Electrical Engineering Department, Boğaziçi University, Istanbul, Turkey, in July 2003, and the MS and PhD degrees from the Electrical and Computer Engineering Department, The Ohio State University in March 2005 and December 2007, respectively. He is currently a postdoctoral researcher at the Computer Science and Engineering Department, University of California, Riverside. His research interests include wire-

less sensor networks, wireless ad hoc networks, and wireless mesh networks, specifically on medium access control, routing, and security. He is a member of the IEEE.



**Eylem Ekici** received the BS and MS degrees in computer engineering from Boğaziçi University, Istanbul, Turkey, in 1997 and 1998, respectively, and the PhD degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, in 2002. Currently, he is an associate professor with the Department of Electrical and Computer Engineering, The Ohio State University. His current research interests include wireless sensor networks, vehicular

communication systems, and cognitive radio networks, with a focus on routing and medium access control protocols, resource management, and analysis of network architectures and protocols. He is an associate editor of *Computer Networks Journal* (Elsevier) and the *ACM Mobile Computing and Communications Review*. He is a member of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at [www.computer.org/publications/dlib](http://www.computer.org/publications/dlib).**