Available online at www.sciencedirect.com
ScienceDirect
Computer Networks 51 (2007) 3727–3749
Computer Networks
www.elsevier.com/locate/comnet

Probability distribution of multi-hop-distance in one-dimensional sensor networks ☆

Serdar Vural *, Eylem Ekici

Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH 43210, USA

Received 6 December 2006; received in revised form 21 March 2007; accepted 24 March 2007
Available online 14 April 2007

Responsible Editor: B. Baykal

Abstract

In spatially random sensor networks, estimating the Euclidean distance covered by a packet in a given number of hops carries a high importance for various functions such as localization and distance estimations. The inaccuracies in such estimations motivate this study on the distribution of the Euclidean distance covered by a packet in spatially random sensor networks in a given number of hops. Although a closed-form expression of distance distribution cannot be obtained, highly accurate approximations are derived for this distribution in one-dimensional spatially random sensor networks. Using statistical measures and numerical examples, it is also shown that the presented distribution approximation yields very high accuracy even for small number of hops. Furthermore, the distribution of hop distance that covers a given Euclidean distance is also approximated.
Published by Elsevier B.V.

Keywords: Sensor networks; Localization; Probability and statistics

1. Introduction

The Euclidean distance between two nodes in a sensor network can be measured in various ways, such as using the coordinates of the nodes. When the coordinate information is not available or the coordinates cannot be computed, other methods are needed to estimate the distance. A candidate metric for indicating how far away one sensor is from another is the hop distance between the sensors. For networks in which sensor positions are deterministic, the relation between the hop distance and the Euclidean distance can be obtained using simple geometric and algebraic calculations. However, in the case of spatially random networks, the randomness of sensor positions creates random inter-sensor distances. Hence, the Euclidean distance of a hop becomes a random variable. Therefore, relating Euclidean distances with hop distances needs to be accomplished by means of a stochastic study.

☆ A preliminary version of this paper has appeared in the Proceedings of ACM MobiHoc 2005.

* Corresponding author. Tel.: +1 6142923430.
E-mail addresses: vurals@ece.osu.edu (S. Vural), ekici@ece.osu.edu (E. Ekici).

1389-1286/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.comnet.2007.03.008

Using the hop distance between sensors to obtain distance estimations is a technique applied to position estimation [5,6]. An average size of one hop is estimated using hop distances to a number of landmarks. The average hop distance is then used to estimate the Euclidean distances of sensors to the landmarks. In order to obtain approximate node locations, trilateration is performed using these distance estimates. Hop-TERRAIN algorithm in [7] finds the number of hops from a node to each of the anchors nodes in a network and then multiplies this hop count by a shared metric (average hop distance) to estimate the range between the node and each anchor. The known positions of anchor nodes and these computed ranges are used to perform a triangulation to obtain estimated node positions. The problematic issues in these schemes is their sensitivity to the accuracy of the initial position estimates, the magnitude of errors in the range estimates, and the fraction of anchor nodes.

Relating hop distance with Euclidean distance could be used by schemes that require distance estimations without localization of sensors [4,8]. The *maximum Euclidean distance* that can be attained for a given number of hops can be used to find approximate boundaries of regions where sensor nodes are estimated to exist. Nagpal et al. [4] use the hop distances of sensor nodes from one or more designated sources in order to obtain estimates of inter-sensor Euclidean distances which are used to locate sensor positions. The use of the maximum communication range of a sensor node as the expected distance of a single hop results in errors in estimations. This is due to the fact that the distance to the furthest node is not necessarily equal to communication range and changes according to node density.

The simulation study in [8] has shown that sensors that are at particular hop distances from two sources in a dense sensor network can be grouped according to these hop distances. Furthermore, the groups are found to be confined within well-defined regions. The boundaries between regions are formed by those sensors that are maximally distant from the sources. However, for a given hop distance from a source, the distance of the boundary sensors from the source node is not a scalar multiple of the communication range. Additionally, the locations of boundary sensors are not deterministic. Therefore, the distribution of the maximum distance for a given hop distance from sources must be used to determine possible boundary locations. As many sensor networks are randomly deployed, such an analysis involves probability density function of Euclidean distances.

The probabilistic evaluation of the distance between randomly deployed nodes is presented in [10,11], and [12] from different perspectives. In [10], the probability that two nodes at a certain Euclidean distance can communicate in one or two hops is derived. Furthermore, the expected hop distance is evaluated by analytical bounds and experimentations. However, an exact analytical formula for expected hop distance is not derived. In [12], the pdf of distance between two locations is investigated for uniform and Gaussian distributed node locations. Since the connectivity between selected locations is not considered, [12] does not cover the relation between hop distance and Euclidean distance. In [11], an expression for the expected number n of relay nodes, hence expected value of hop distance, between two randomly located nodes is derived using an analytical expression of connectivity in n or less number of hops. The expression for connectivity in one or two hops is derived and connectivity in more than two hops is calculated using an iterative formula. However, the generalization from two hops to more than two hops, hence the validity of the general connectivity expression, is based on intuition alone, without the support of a comparison with any experimental results. In [9], the experimental investigation of the impact of directional transmission on hop distance and the impact of network diameter on maximum hop distance are presented.

The problem of estimation of the maximum Euclidean distance that can be covered in a given number of hops is not addressed extensively in the literature. In fact, the maximum Euclidean distance corresponding to a certain hop distance is a practical metric for modeling spatially random sensor networks. Specifically, the capacity of multi-hop communication in terms of estimating the area of multi-hop packet reception is defined by the maximum distance that can be covered in multi-hop paths. Hence, this metric is related with coverage area estimations and data dissemination capability of sensor networks. Furthermore, the maximum Euclidean distance is directly related with estimation of hop distance since hop distance is equal to the least number of hops over all multi-hop paths between any two locations.

Cheng and Robertazzi [1] derive the probability density function of a single hop maximal distance in a one-dimensional network. In [1], the expected value of the single hop maximal distance is used for the purpose of determining the expected number of broadcast cycles formed before broadcast percolation ceases. Furthermore, it is claimed that this expected distance value is true for any single hop. Although the inter-dependency of single hop distances is mentioned, no results are derived for the probability density function of the maximal Euclidean distance for a given hop distance.

Similar to the problem of distance estimation for a given hop distance, the reverse problem of determination of the hop distance between two sensor nodes at a certain Euclidean distance is of profound importance for several sensor network functions. For instance, estimation of the minimum number of retransmissions in a multi-hop path, hence the hop distance, is closely related with the estimation of required total energy consumption. However, when sensor nodes are randomly located in the network, hop distance is not deterministic, but a random variable, which is dependent on the communication range of sensor nodes and the node density of the network. Therefore, the hop distance between two sensors that are a certain Euclidean distance apart from each other is randomly distributed with a conditional probability mass function.

In this paper, the distribution of the maximum Euclidean distance for a given hop distance is analyzed for one-dimensional sensor networks. This analysis provides a mathematical model of the information dissemination capability in applications such as vehicular networks [3]. For instance, the locations of vehicles on a road at a specific time instance forms spatially random distribution of sensors embedded to the vehicles that are parts of the same network. Furthermore, distributing sensors to a road by a moving vehicle to monitor road activity creates a spatially random network of sensors. Vehicular networks is a direct application of the analysis, applicability of the one-dimensional distribution results is not limited to vehicular networks. In fact, one-dimensional distance distribution is a powerful information that can be applied to any one-dimensional network with random node locations.

Extension of the one-dimensional network analysis to two-dimensional spatially random networks poses a challenging problem. The definition of maximum Euclidean distance as well its modeling and analysis in two-dimensional networks are more complicated than one-dimension networks. This is due to the geometric complexity of the spatial distribution of node locations and the definition of hop distance in two dimensions. Although a single line uniquely defines the path between two locations in one-dimensional networks, in two-dimensional networks, there can be multiple paths between any two sensor nodes. In this paper, in addition to the one-dimensional analysis, a model for analyzing the maximum Euclidean distance for a given hop distance in two-dimensional networks is outlined. The analysis of two-dimensional networks is a future study using this model.

In this paper, the theoretical expressions for the *expected value* and the *standard deviation* of Euclidean distance are presented. Since these expressions are computationally costly, efficient approximation methods are proposed. Furthermore, the similarity between the distribution of the multi-hop distance and the Gaussian distribution is evaluated using the mean square difference and the *kurtosis* [2] of the multi-hop distance distribution. Finally, the probability mass function of hop distance for a given Euclidean distance is determined.

The remainder of the paper is organized as follows: In Section 2, the distribution of the multi-hop distance for a given hop distance is studied. Theoretical expressions for multi-hop distance distribution are presented in Section 2.1. Furthermore, approximation equations for multi-hop distance statistics along with their derivations are provided in Section 2.4. The Gaussianity of multi-hop-distance distribution is studied in Section 3. The kurtosis of a single-hop-distance is approximated in Section 3.2, and the method to obtain an approximation to the kurtosis of the multi-hop-distance is derived in Section 3.3. In Section 4, the probability mass function of hop distance corresponding to a given Euclidean distance is derived. Performance analysis of the proposed approximation methods are provided in Section 5 with comparisons to experimental and theoretical results. Section 6 provides a proposed model for analyzing maximum Euclidean distances in two-dimensional sensor networks as a future work. Finally, Section 7 concludes the paper.

2. Distribution of multi-hop-distance for a given hop distance

In this section, the distribution of the maximum distance covered in a given number of hops is studied. Sensor nodes are uniformly distributed and have a fixed communication range R. Furthermore, sensors are able to receive every packet within their communication range. It is assumed that there is no mobility and no node failure.

Before proceeding to the derivations and related discussions, two definitions that are frequently used throughout the text are provided as follows:

Single-hop-distance (**r**): the maximum possible distance covered in a single hop. This variable is used with the corresponding hop number, e.g. $\mathbf{r_i}$, which designates the single-hop-distance of hop i and $\mathbf{r_i} \leqslant R$ for all $i = 1, 2, 3, \ldots$

Multi-hop-distance (**d**): the maximum possible distance covered in multiple hops. This variable is used with the given hop distance N as $\mathbf{d_N}$. Multi-hop-distance $\mathbf{d_N}$ is the sum of individual single-hop-distances $\mathbf{r_i}$ for $i = 1, 2, \ldots, N$.

The distribution $f_{\mathbf{d_N}}(d_N)$ of the multi-hop-distance $\mathbf{d_N}$ is investigated by the expected value $E[\mathbf{d_N}]$ and standard deviation $\sigma_{\mathbf{d_N}}$ of $\mathbf{d_N}$. Determination of $E[\mathbf{d_N}]$ and $\sigma_{\mathbf{d_N}}$ is achieved by studying the expected value $E[\mathbf{r_i}]$ and standard deviation $\sigma_{\mathbf{r_i}}$ of single-hop-distance. Therefore, the study of $f_{\mathbf{d_N}}(d_N)$ involves the probability distribution of a single-hop-distance $f_{\mathbf{r_i}}(r_i)$. Furthermore, the similarity of $f_{\mathbf{d_N}}(d_N)$ and a Gaussian distribution with the same mean and standard deviation is inspected.

In the following two sections, theoretical expressions of $f_{\mathbf{r_i}}(r_i)$ and $f_{\mathbf{d_N}}(d_N)$ are provided.

2.1. Theoretical expressions for $f_{\mathbf{r_i}}(r_i)$ and $f_{\mathbf{d_N}}(d_N)$

2.1.1. Probability density function $f_{\mathbf{r_i}}(r_i)$ of single-hop-distance

The probability density function $f_{\mathbf{r_i}}(r_i)$ of the single-hop-distance $\mathbf{r_i}$ is provided in the study of Cheng and Robertazzi [1] for one-dimensional networks. The following two definitions are taken from [1] and used throughout this paper.

Definition. $\mathbf{P_i}$: furthest point covered in hop i. The distance between P_i and the furthest point P_{i-1} of hop $i-1$ is equal to the single-hop-distance $\mathbf{r_i}$ of hop i.

Definition. $\mathbf{r_{e_{i-1}}}$: the length of the *vacant* segment from the point P_i to the location R away from the point P_{i-1} at the same direction, where $\mathbf{r_{e_i}} \leqslant R$ for all $i = 1, 2, 3, \ldots$. By the definition of $\mathbf{r_i}$ and $\mathbf{r_{e_i}}$, the following is true for all hops i:

$$\mathbf{r_{e_i}} + \mathbf{r_i} = R. \tag{1}$$

The random variables $\mathbf{r_i}$ and $\mathbf{r_{e_i}}$ for $i = 1, 2, 3, \ldots, N$ are shown in Fig. 1.

In [1], $f_{\mathbf{r_i}}(r_i)$ is given by

$$f_{\mathbf{r_i}}(r_i) = \frac{\lambda e^{-\lambda(R-r_i)}}{1 - e^{-\lambda(R-r_{e_{i-1}})}}, \tag{2}$$

where λ is node density.

Since $\mathbf{r_i} + \mathbf{r_{e_i}} = R$, $f_{\mathbf{r_{e_i}}}(r_{e_i}) = f_{\mathbf{r_i}}(R - r_{e_i})$ is true with a change of variables. Hence, using Eq. (2), the pdf $f_{\mathbf{r_{e_i}}}(r_{e_i})$ is obtained as

$$f_{\mathbf{r_{e_i}}}(r_{e_i}) = \frac{\lambda e^{-\lambda r_{e_i}}}{1 - e^{-\lambda(R-r_{e_{i-1}})}}. \tag{3}$$

Calculating $E[\mathbf{r_i}]$ using $f_{\mathbf{r_{e_i}}}(r_{e_i})$ is possible due to the relation in Eq. (1), and is easier than the calculation with $f_{\mathbf{r_i}}(r_i)$. Hence, $E[\mathbf{r_i}]$ is determined with the pdf $f_{\mathbf{r_{e_i}}}(r_{e_i})$.

The dependency between consecutive single hops is obvious in Eqs. (2) and (3) and is mentioned in [1]. Due to this dependency, the pdf of multi-hop-distance is a very complex expression and is not discussed in [1]. In the following section, $f_{\mathbf{d_N}}(d_N)$ is provided and its first-order statistics are calculated.

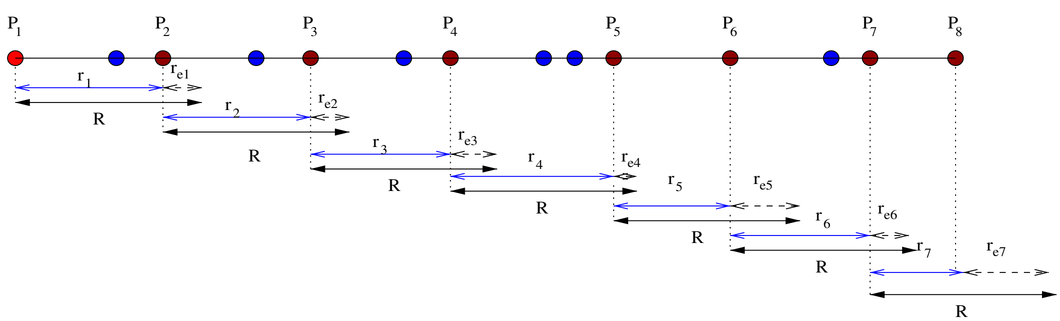


Fig. 1. Illustration of single-hop-distances r_i and furthest points P_i.

2.1.2. Probability density function $f_{\mathbf{d_N}}(d_N)$ of multi-hop-distance

The probability distribution $f_{\mathbf{d_N}}(d_N)$ of the multi-hop-distance $\mathbf{d_N}$ can be written as follows:

$$f_{\mathbf{d_N}}(d_N) = f_{\mathbf{d_{e_N}}}(NR - d_N)$$

$$f_{\mathbf{d_{e_N}}}(d_{e_N}) = \int_0^R \int_0^{R-r_{e_1}} \cdots \int_0^{R-r_{e_{N-1}}} \frac{\lambda e^{-\lambda r_{e_N}}}{1-e^{-\lambda R - r_{e_{N-1}}}} \cdots \frac{\lambda e^{-\lambda r_{e_2}}}{1-e^{-\lambda R - r_{e_1}}} \frac{\lambda e^{-\lambda r_{e_1}}}{1-e^{-\lambda R}} dr_{e_N} \cdots dr_{e_2} dr_{e_1}, \quad (4)$$

where $\mathbf{d_{e_N}}$ is the sum of the N vacant lengths $\mathbf{r_{e_i}}$ for $i = 1, 2, \ldots, N$ given by

$$\mathbf{d_{e_N}} = \sum_{i=1}^{N} \mathbf{r_{e_i}}. \quad (5)$$

Using Eq. (4) for individual single-hop-distances, the following expression is obtained for multi-hop-distance $\mathbf{d_N}$ and $\mathbf{d_{e_N}}$:

$$\mathbf{d_N} = NR - \sum_{i=1}^{N} \mathbf{r_{e_i}}. \quad (6)$$

Since NR is a deterministic quantity, $f_{\mathbf{d_N}}(d_N) = f_{\mathbf{d_{e_N}}}(NR - d_N)$ holds.

$f_{\mathbf{d_N}}(d_N)$ in Eq. (4) is a nested integral which cannot be reduced to a closed-form expression to the best of our knowledge. Since a distribution can be evaluated with statistical values, such as expected value and standard deviation, $E[\mathbf{d_N}]$ and $\sigma_{\mathbf{d_N}}$ are studied.

The expressions for the expected value of the multi-hop-distance and that of the variance can be written as follows:

$$E[\mathbf{d_N}] = \int_0^R \int_0^{R-r_{e_1}} \cdots \int_0^{R-r_{e_{N-1}}} \frac{d_N \lambda e^{-\lambda r_{e_N}}}{1-e^{-\lambda R - r_{e_{N-1}}}} \cdots \frac{\lambda e^{-\lambda r_{e_2}}}{1-e^{-\lambda R - r_{e_1}}} \frac{\lambda e^{-\lambda r_{e_1}}}{1-e^{-\lambda R}} dr_{e_N} \ldots dr_{e_2} dr_{e_1}, \quad (7)$$

$$\sigma^2_{\mathbf{d_N}} = \int_0^R \int_0^{R-r_{e_1}} \cdots \int_0^{R-r_{e_{N-1}}} \frac{(d_N - \overline{d_N})^2 \lambda e^{-\lambda r_{e_N}}}{1-e^{-\lambda R - r_{e_{N-1}}}} \cdots \frac{\lambda e^{-\lambda r_{e_2}}}{1-e^{-\lambda R - r_{e_1}}} \frac{\lambda e^{-\lambda r_{e_1}}}{1-e^{-\lambda R}} dr_{e_N} \cdots dr_{e_2} dr_{e_1}, \quad (8)$$

where $\mathbf{d_N}$ is given in Eq. (6) and $\overline{d_N} = E[\mathbf{d_N}]$. Eqs. (7) and (8) are nested integrals like the pdf expression in Eq. (4), which makes them computationally costly. Therefore, some approximation methods are derived as will be presented in Section 2.4.

Before proceeding to the approximation of $E[\mathbf{d_N}]$ and $\sigma_{\mathbf{d_N}}$, expected distance $E[\mathbf{r_i}]$ and standard deviation $\sigma_{\mathbf{r_i}}$ of single-hop-distance $\mathbf{r_i}$ are calculated in the following section. These values are used in $E[\mathbf{d_N}]$ and $\sigma_{\mathbf{d_N}}$ approximations.

2.2. Expected single-hop-distance

As mentioned in Section 2.1.1, it is easier to calculate $E[\mathbf{r_{e_i}}]$ with $f_{\mathbf{r_{e_i}}}(r_{e_i})$ than $E[\mathbf{r_i}]$ with $f_{\mathbf{r_i}}(r_i)$. The relation in Eq. (1) implies $E[\mathbf{r_i}] + E[\mathbf{r_{e_i}}] = R$. Hence, $E[\mathbf{r_{e_i}}]$ can be used calculate to $E[\mathbf{r_i}]$. In [1], $E[\mathbf{r_{e_i}}]$ is calculated to be the following:

$$E[\mathbf{r_{e_i}}] = \frac{1 - e^{-\lambda(R - r_{e_{i-1}})}(1 + \lambda(R - r_{e_{i-1}}))}{\lambda(1 - e^{-\lambda(R - r_{e_{i-1}})})}. \quad (9)$$

In Eq. (9), $\mathbf{r_{e_{i-1}}}$ is replaced with its expected value $E[\mathbf{r_{e_{i-1}}}]$ which is then substituted by $E[\mathbf{r_{e_{i-1}}}] = R - E[\mathbf{r_{i-1}}]$.

The equivalence of the expected single-hop-distances is assumed in [1]. Hence, $\mathbf{r_i}$ can be replaced by its expectation $E[\mathbf{r_i}]$ to give

$$E[\mathbf{r_i}] = E[\mathbf{r_{i-1}}] = \bar{r}. \quad (10)$$

With the assumption in Eq. (10), $E[\mathbf{r_{i-1}}]$ can be further replaced by $\bar{r}$. Similarly, $E[\mathbf{r_{e_i}}]$ on the left-hand side of Eq. (9) is equal to $E[\mathbf{r_{e_i}}] = \overline{r_e} = R - E[\mathbf{r_i}] = R - \bar{r}$. As a result, the following expression for finding the approximation of the expected single-hop-distance is obtained:

$$\ln\left(1 - \frac{\lambda\bar{r}}{\lambda R - \lambda\bar{r} - 1}\right) = \lambda\bar{r}. \quad (11)$$

Expression (11) is an implicit expression. Therefore, the values of $\bar{r}$ are obtained numerically. $\bar{r}$ is used to derive closed form equations of the statistics of the single-hop-distance, which in turn prove to be handy in calculating the statistics of the multi-hop-distance.

2.3. Variance of single-hop-distance

The standard deviation $\sigma_{\mathbf{r_i}}$ of single-hop-distance $\mathbf{r_i}$ can be calculated using the variance of $\mathbf{r_{e_i}}$ more easily than a calculation using $f_{\mathbf{r_i}}(r_i)$. After calculating $\sigma_{\mathbf{r_{e_i}}}$, the following proposition is used to obtain $\sigma_{\mathbf{r_i}}$.

Proposition. *The variance of the single-hop-distance* $\mathbf{r_i}$ *is equal to the variance of total of the singlehop vacant region length*, $\mathbf{r_{e_i}}$. *Hence*,

$$\sigma^2_{\mathbf{r_i}} = \sigma^2_{\mathbf{r_{e_i}}}. \quad (12)$$

Proof. The variance of $\mathbf{r_i}$ can be written as

$$\begin{aligned}E[(\mathbf{r_i}-\bar{r})^2] &= E[(R-\mathbf{r_{e_i}}-\bar{r})^2]\\ &= E[(R-\mathbf{r_{e_i}})^2 - 2\bar{r}(R-\mathbf{r_{e_i}})+\bar{r}^2]\\ &= E[R^2-2R\mathbf{r_{e_i}}+\mathbf{r_{e_i}^2}-2\bar{r}R+2\bar{r}\mathbf{r_{e_i}}+\bar{r}^2]\\ &= R^2-2R\overline{r_e}+E[\mathbf{r_{e_i}^2}]-2\bar{r}R+2\bar{r}\overline{r_e}+\bar{r}^2\\ &= R^2-2R\overline{r_e}+E[\mathbf{r_{e_i}^2}]-2(R-\overline{r_e})R\\ &\quad +2(R-\overline{r_e})\overline{r_e}+R^2-2\overline{r_e}R+\overline{r_e}^2\\ &= E[\mathbf{r_{e_i}^2}]-\overline{r_e}^2,\end{aligned}$$

$$\sigma^2_{\mathbf{r_i}} = \sigma^2_{\mathbf{r_{e_i}}}.$$

Hence, the variance of $\mathbf{r_i}$ can be found by determining the variance of $\mathbf{r_{e_i}}$. □

Then, the variance of single-hop-distance becomes

$$\sigma^2_{\mathbf{r_i}} = \sigma^2_{\mathbf{r_{e_i}}} = E[\mathbf{r_{e_i}^2}]-\overline{r_e}^2 = E[\mathbf{r_{e_i}^2}]-(R-\bar{r})^2. \tag{13}$$

The second moment of $\mathbf{r_{e_i}}$ in Eq. (13) can be found similar to the way outlined in [1] for finding $E[\mathbf{r_{e_i}}]$ and is obtained as follows:

$$\begin{aligned}E[\mathbf{r_{e_i}^2}] = &-\frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}(R-r_{e_{i-1}})^2\\ &-\frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}\frac{2}{\lambda}(R-r_{e_{i-1}})\\ &+\frac{1}{1-e^{-\lambda(R-r_{e_{i-1}})}}\left\{\frac{2}{\lambda^2}-\frac{2}{\lambda^2}e^{-\lambda(R-r_{e_{i-1}})}\right\}.\end{aligned} \tag{14}$$

In Eq. (14), $\mathbf{r_{e_{i-1}}}$ is replaced by its expected value $E[\mathbf{r_{e_{i-1}}}]$ which is further substituted by $R-\mathbf{r_{e_{i-1}}}=\bar{r}$ to give:

$$E[\mathbf{r_e^2}] = \frac{-\bar{r}^2e^{-\lambda\bar{r}}-\frac{2}{\lambda}\bar{r}e^{-\lambda\bar{r}}+\frac{2}{\lambda^2}-\frac{2}{\lambda^2}e^{-\lambda\bar{r}}}{1-e^{-\lambda\bar{r}}}. \tag{15}$$

Substituting Eq. (15) into Eq. (13), the variance of the single-hop-distance can be found. However, taking $E[\mathbf{r_{e_i}^2}]=E[\mathbf{r_e^2}]$ in Eq. (15) is not a result of the approximation in Eq. (10). In fact, it is assumed that *single-hop-distances are identically distributed but not independent*. In Section 3, higher order statistical values such as $E[\mathbf{r_i^4}]$ of the single-hop-distance are taken to be identical for every single hop, which is a result of this assumption.

2.4. Expected value and standard deviation of multi-hop-distance

In this section, $E[\mathbf{r_i}]=\bar{r}$ is used for the derivation of expected multi-hop-distance $E[\mathbf{d_N}]$ and standard deviation $\sigma_{\mathbf{d_N}}$.

Since multi-hop-distance $\mathbf{d_N}$ is the sum of single-hop-distances as given by Eq. (6), $E[\mathbf{d_N}]$ can be approximated using the assumption that single-hop-distance variables $\mathbf{r_i}$ are identically distributed with mean given by Eq. (10). Hence, the expected value of the multi-hop-distance for N hops is approximated by

$$E[\mathbf{d_N}] = E\left[\sum_{i=1}^{N}\mathbf{r_i}\right] = N\bar{r}. \tag{16}$$

The standard deviation $\sigma_{\mathbf{d_N}}$ of the multi-hop-distance for N hops can be calculated with the following equation:

$$\begin{aligned}E\left[(\mathbf{d_N}-\overline{d_N})^2\right] &= E\left[\left(\left(\sum_{i=1}^{N}\mathbf{r_i}\right)-\left(\sum_{i=1}^{N}\overline{r_i}\right)\right)^2\right]\\ &= E\left[\left(\sum_{i=1}^{N}\mathbf{r_i}\right)^2-2\left(\sum_{i=1}^{N}\mathbf{r_i}\right)\left(\sum_{i=1}^{N}\overline{r_i}\right)\right]\\ &\quad +E\left[\left(\sum_{i=1}^{N}\overline{r_i}\right)^2\right]\\ &= E\left[\left(\sum_{i=1}^{N}\mathbf{r_i}\right)^2\right]-N^2\bar{r}^2.\end{aligned} \tag{17}$$

Eq. (17) requires $E[d_N^2]=E[(\sum_{i=1}^{N}\mathbf{r_i})^2]$. This term can be calculated by Eqs. (29) and (30) presented in Section 3.3.

3. Gaussianity of the multi-hop-distance distribution

Experimental distribution $f_{\mathbf{d_N}}(d_N)$ of multi-hop-distance $\mathbf{d_N}$ is illustrated in Fig. 2 for $N=1,2,3,\ldots,10$. Each experimental distribution $f_{\mathbf{d_N}}(d_N)$ is compared with a corresponding Gaussian distribution with mean $E[\mathbf{d_N}]$ and standard deviation $\sigma_{\mathbf{d_N}}$.

In Fig. 2, the multi-hop-distance distribution curves become more similar to a Gaussian curve with increasing hop distance. This similarity suggests that *the multi-hop distance* $\mathbf{d_N}$ distribution can be approximated by the well-*known Gaussian distribution*, which is only described by its mean and standard deviation. Furthermore, Eqs. (16) and (17) provide effective ways of approximating the mean and the variance of $\mathbf{d_N}$.

Due to the statistical dependence between consecutive single-hop-distances, Central Limit Theorem does not apply to claim that multi-hop-distance

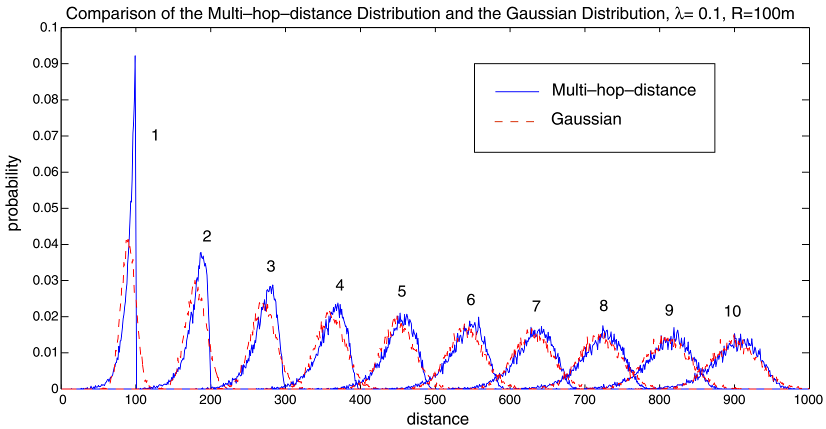


Fig. 2. Comparison of the multi-hop-distance distribution with Gaussian distribution.

has a Gaussian distribution. Hence, it should be noted that Gaussian pdf is an approximation to the multi-hop-distance density function.

Although Fig. 2 presents an illustration of increasing Gaussian character in multi-hop-distance distribution, a quantitative method to test the Gaussianity of multi-hop-distance is required. Section 3.1 introduces such a measure called *kurtosis*, and Sections 3.2 and 3.3 explain how the kurtosis values of single-hop-distance and multi-hop-distance distributions are determined, respectively.

3.1. Kurtosis as a measure of Gaussianity

As a quantitative measure of similarity to Gaussian pdf, a statistical quantity called "kurtosis" can be used [2]. Kurtosis is a measure of the peakedness of a probability distribution. Gaussian pdf has a kurtosis value of 3. In general, the kurtosis of a random variable $\mathbf{x}$ is defined as the ratio of the fourth-order central moment and the square of the variance:

$$\mathrm{kurt}(\mathbf{x}) = \frac{E[(\mathbf{x}-\bar{x})^4]}{E[(\mathbf{x}-\bar{x})^2]^2}. \quad (18)$$

Another definition of kurtosis is "*kurtosis excess*", which is 3 less than the general kurtosis expression (Eq. (18)). For a normal distribution, the kurtosis excess of a normal distribution is zero. Hence, *in order to obtain a more Gaussian-like distribution, the absolute value of kurtosis-excess should be minimized.* In the remainder of the paper, the term *kurtosis* is used to refer to *kurtosis excess*.

In Fig. 3, the experimental kurtosis of the multi-hop-distance is shown as a function of number of hops N in case of a node density of $\lambda = 0.05$ nodes/m and a communication range of $R = 100$ m. The kurtosis value and hence the non-Gaussianity has a decreasing character with increasing number of hops. After approximately $N = 50$ hops, the kurtosis value settles within a range of [0 0.2].

3.2. Gaussianity of the single-hop-distance distribution

In this part of our study, the Gaussianity of the pdf of the single-hop-distance $\mathbf{r}$ is inspected. To achieve this, Eq. (18) is expanded and the variable $\mathbf{x}$ is replaced by $\mathbf{r}$ to obtain the kurtosis of the single-hop-distance as follows:

$$\mathrm{kurt}(\mathbf{r}) = \frac{E[\mathbf{r}^4] - 4E[\mathbf{r}^3]E[\mathbf{r}] + 6E[\mathbf{r}^2]E[\mathbf{r}]^2 - 3E[\mathbf{r}]^4}{(E[\mathbf{r}^2] - E[\mathbf{r}]^2)^2}. \quad (19)$$

In this section, any single-hop-distance is referred as $\mathbf{r}$ and any single-hop vacant region as $\mathbf{r_e}$, without using a subscript i to denote the hop count. In Eq. (19), all terms are moments $E[\mathbf{r}^j_{\mathbf{e_i}}]$ of the single-hop-distance $\mathbf{r}$ without a reference to hop number due to the assumption that single-hop-distances are identically distributed. Hence, $E[\mathbf{r}^j_{\mathbf{e_i}}] = E[\mathbf{r}^j_{\mathbf{e}}]$.

In Eq. (19) the kurtosis expression involves the fourth and the third-order moments of the single-hop-distance $\mathbf{r}$. Calculating the third and the fourth moments of the vacant length $\mathbf{r_e}$ is much easier than calculating the corresponding moments of the distance $\mathbf{r}$. Therefore, the moments of the single-hop-distance r are calculated with the moments of the vacant

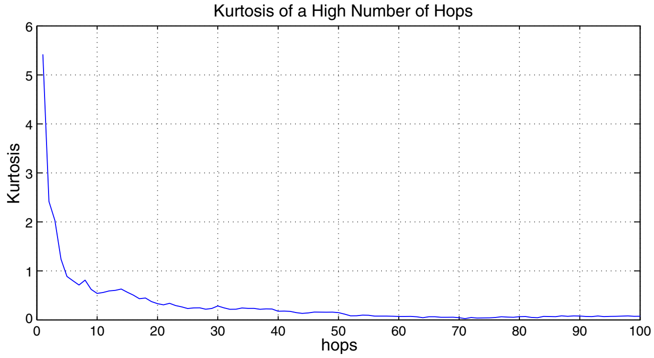


Fig. 3. Kurtosis of Experimental d_N. $R = 00$ m, $\lambda = 0.05$ nodes/m.

region $\mathbf{r_e}$. Since $\mathbf{r_e} + \mathbf{r} = R$ holds, the relationship between the moments of $\mathbf{r_e}$ and $\mathbf{r}$ is found as follows:

$$E[\mathbf{r}^3] = R^3 - 3R^2(R - E[\mathbf{r}]) + 3RE[\mathbf{r}_\mathbf{e}^2] - E[\mathbf{r}_\mathbf{e}^3], \tag{20}$$

$$E[\mathbf{r}^4] = R^4 - 4R^3(R - E[\mathbf{r}]) + 6R^2E[\mathbf{r}_\mathbf{e}^2] - 4RE[\mathbf{r}_\mathbf{e}^3] + E[\mathbf{r}_\mathbf{e}^4]. \tag{21}$$

The expressions for the third moment $E[\mathbf{r}_{\mathbf{e_i}}^3]$ and fourth moment $E[\mathbf{r}_{\mathbf{e_i}}^4]$ of $\mathbf{r}$ are calculated as

$$\begin{aligned} E[\mathbf{r}_{\mathbf{e_i}}^3] = {} & \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}(-(R-r_{e_{i-1}})^3) \\ & - \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}\frac{3}{\lambda}(R-r_{e_{i-1}})^2 \\ & - \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}\left(\frac{6}{\lambda^2}(R-r_{e_{i-1}}) + \frac{6}{\lambda^3}\right) \\ & + \frac{6}{\lambda^3(1-e^{-\lambda(R-r_{e_{i-1}})})}, \end{aligned} \tag{22}$$

$$\begin{aligned} E[\mathbf{r}_{\mathbf{e_i}}^4] = {} & \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}(-(R-r_{e_{i-1}})^4) \\ & - \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}\frac{4}{\lambda}(R-r_{e_{i-1}})^3 \\ & - \frac{e^{-\lambda(R-r_{e_{i-1}})}}{1-e^{-\lambda(R-r_{e_{i-1}})}}\left(\frac{12}{\lambda^2}(R-r_{e_{i-1}})^2 + \frac{24}{\lambda^4}\right) \\ & + \frac{1}{1-e^{-\lambda(R-r_{e_{i-1}})}}\left(\frac{24}{\lambda^4} - \frac{24}{\lambda^3}(R-r_{e_{i-1}})\right). \end{aligned} \tag{23}$$

After deriving $E[\mathbf{r}_{\mathbf{e_i}}^j]$, where $j = 1, 2, 3, \ldots$, the variables $\mathbf{r}_{\mathbf{e_i}}$ and $\mathbf{r}_{\mathbf{e_{i-1}}}$ are both replaced by $\mathbf{r_e}$. Similarly, $\mathbf{r_i}$ and $\mathbf{r_{i-1}}$ are both changed to be $\mathbf{r}$. In the same way, the third and the fourth moments of $\mathbf{r_e}$ can be found as follows:

$$E[\mathbf{r}_\mathbf{e}^3] = \frac{e^{-\lambda\bar{r}}}{1-e^{-\lambda\bar{r}}}\left(-\bar{r}^3 - \frac{3}{\lambda}\bar{r}^2 - \frac{6}{\lambda^2}\bar{r} - \frac{6}{\lambda^3}\right) + \frac{6}{\lambda^3(1-e^{-\lambda\bar{r}})}, \tag{24}$$

$$E[\mathbf{r}_\mathbf{e}^4] = \frac{e^{-\lambda\bar{r}}}{1-e^{-\lambda\bar{r}}}\left(-\bar{r}^4 - \frac{4}{\lambda}\bar{r}^3 - \frac{12}{\lambda^2}\bar{r}^2 - \frac{24}{\lambda^4}\right) + \frac{1}{1-e^{-\lambda\bar{r}}}\left(\frac{24}{\lambda^4} - \frac{24}{\lambda^3}\bar{r}\right). \tag{25}$$

The expected single-hop-distance $\mathbf{r}$ in Eq. (10) is substituted in Eqs. (24) and (25) to get the third and the fourth moments of $\mathbf{r_e}$, respectively. Then, the results are used in Eqs. (20) and (21) to get the third and fourth moments of $\mathbf{r}$, respectively. Using the third moment (Eq. (20)), the fourth moment (Eq. (21)), the second moment (Eqs. (13) and (15)) and the expected value of the single-hop-distance (Eq. (10)), Eq. (19) is evaluated to obtain the kurtosis of the single-hop-distance.

3.3. Gaussianity of the distribution of multi-hop-distance d_N

In this section, the Gaussian character of the multi-hop-distance destination is investigated. Using the approximation for the expected single-hop-distance, Eq. (10), and the definition of kurtosis in Eq. (18), the expression to compute the kurtosis of the multi-hop-distance kurt($\mathbf{d_N}$) can be obtained as follows:

$$\begin{aligned}\mathrm{kurt}(\mathbf{d_N}) &= \frac{E[(\mathbf{d_N}-\bar{d}_N)^4]}{E[(\mathbf{d_N}-\bar{d})^2]^2} \\ &= \frac{E\left[\left(\sum_1^N \mathbf{r_i} - \sum_1^N \overline{r_i}\right)^4\right]}{E\left[\left(\sum_1^N \mathbf{r_i} - \sum_1^N \overline{r_i}\right)^2\right]^2} \\ &= \frac{E\left[\left(\sum_1^N \mathbf{r_i}\right)^4 - 4\left(\sum_1^N \mathbf{r_i}\right)^3\left(\sum_1^N \overline{r_i}\right)\right]}{E\left[\left(\sum_1^N \mathbf{r_i}\right)^2 + \left(\sum_1^N \overline{r_i}\right)^2 - 2\left(\sum_1^N \mathbf{r_i}\right)\left(\sum_1^N \overline{r_i}\right)\right]} \\ &\quad + \frac{E\left[6\left(\sum_1^N \mathbf{r_i}\right)^2\left(\sum_1^N \overline{r_i}\right)^2\right]}{E\left[\left(\sum_1^N \mathbf{r_i}\right)^2 + \left(\sum_1^N \overline{r_i}\right)^2 - 2\left(\sum_1^N \mathbf{r_i}\right)\left(\sum_1^N \overline{r_i}\right)\right]} \\ &\quad - \frac{E\left[4\left(\sum_1^N \mathbf{r_i}\right)\left(\sum_1^N \overline{r_i}\right)^3 - \left(\sum_1^N \overline{r_i}\right)^4\right]}{E\left[\left(\sum_1^N \mathbf{r_i}\right)^2 + \left(\sum_1^N \overline{r_i}\right)^2 - 2\left(\sum_1^N \mathbf{r_i}\right)\left(\sum_1^N \overline{r_i}\right)\right]}. \end{aligned} \tag{26}$$

Eq. (26) can be simplified by distributing the expectation operations since expectation is a linear operator. The terms like $4(\sum_1^N \mathbf{r_i})^3(\sum_1^N \overline{r_i})$, where $E[\mathbf{r_i}] = \overline{r_i}$ can also be recognized. In such multiplications, $(\sum_1^N \overline{r_i})$ is the sum of the N expected single-hop-distances. Since the sum of expected values is a constant, $(\sum_1^N \overline{r_i})$ becomes a scalar multiplier of $E[(\sum_1^N \mathbf{r_i})^3]$. In general, the expectation of any term of the form $(\sum_1^N \mathbf{r_i})^p(\sum_1^N \overline{r_i})^s$, where p and s are some integers and $p,s > 0$, involving $(\sum_1^N \overline{r_i})$ as a multiplier, can be evaluated as follows:

$$E\left[\left(\sum_1^N \mathbf{r_i}\right)^p\left(\sum_1^N \overline{r_i}\right)^s\right] = \left(\sum_1^N \overline{r_i}\right)^s E\left[\left(\sum_1^N \mathbf{r_i}\right)^p\right]. \tag{27}$$

Using Eqs. (27), (26) can be modified to obtain:

$$\begin{aligned}\mathrm{kurt}(d) &= \frac{E\left[\left(\sum_1^N \mathbf{r_i}\right)^4\right] - 4\left(\sum_1^N \overline{r_i}\right)E\left[\left(\sum_1^N \mathbf{r_i}\right)^3\right]}{E\left[\left(\sum_1^N \mathbf{r_i}\right)^2\right] - 2E\left[\left(\sum_1^N \mathbf{r_i}\right)\right]\left(\sum_1^N \overline{r_i}\right) + \left(\sum_1^N \overline{r_i}\right)^2} \\ &\quad + \frac{6\left(\sum_1^N \overline{r_i}\right)^2 E\left[\left(\sum_1^N \mathbf{r_i}\right)^2\right] - 4\left(\sum_1^N \overline{r_i}\right)^3 E\left[\left(\sum_1^N \mathbf{r_i}\right)\right]}{E\left[\left(\sum_1^N r_i\right)^2\right] - 2E\left[\left(\sum_1^N \mathbf{r_i}\right)\right]\left(\sum_1^N \overline{r_i}\right) + \left(\sum_1^N \overline{r_i}\right)^2} \\ &\quad + \frac{\left(\sum_1^N \overline{r_i}\right)^4}{E\left[\left(\sum_1^N \mathbf{r_i}\right)^2\right] - 2E\left[\left(\sum_1^N \mathbf{r_i}\right)\right]\left(\sum_1^N \overline{r_i}\right) + \left(\sum_1^N \overline{r_i}\right)^2}. \end{aligned} \tag{28}$$

In Eq. (28), the terms in the form of $E[(\sum_1^N \mathbf{r_i})^j]$ for $j = 2,3,4$ are calculated as follows: let $f_j(N)$ be defined as $f_j(N) \equiv E[(\sum_1^N \mathbf{r_i})^j]$, where N is the hop distance and j is the power term. Then, $f_j(N)$ is computed recursively as

$$\begin{aligned} f_2(N) &= f_2(N-1) + 2(N-1)\bar{r}^2 + E[\mathbf{r}^2], \\ f_3(N) &= f_3(N-1) + 3f_2(N-1)\bar{r} \\ &\quad + 3(N-1)\bar{r}E[\mathbf{r}^2] + E[\mathbf{r}^3], \\ f_4(N) &= f_4(N-1) + 4f_3(N-1)\bar{r} \\ &\quad + 6f_2(N-1)E[\mathbf{r}^2] \\ &\quad + 4(N-1)\bar{r}E[\mathbf{r}^3] + E[\mathbf{r}^4]. \end{aligned} \tag{29}$$

The functions $f_j(N)$ can be computed directly for $N = 2$ and $j = 2,3,4$:

$$\begin{aligned} f_2(2) &= 2E[\mathbf{r}^2] + 2\bar{r}^2, \\ f_3(2) &= 2E[\mathbf{r}^3] + 6E[\mathbf{r}^2]\bar{r}, \\ f_4(2) &= 2E[\mathbf{r}^4] + 6E[\mathbf{r}^2]^2 + 8E[\mathbf{r}^3]\bar{r}. \end{aligned} \tag{30}$$

The values of the terms like $E[(\sum_1^N \overline{r_i})^j]$ for $j = 2,3,4$ are very easy to compute since there is no random quantity inside the expectation operator. According to Assumption (10), $E[\mathbf{r_i}] = \overline{r_i}$ is simply $\bar{\mathbf{r}}$, for all i. Hence, $E[(\sum_1^N \overline{r_i})^j] = E[(N\bar{r})^j]$ holds. $N\bar{r}$ and the powers of $N\bar{r}$ are deterministic quantities. The expectation of $(\sum_1^N \overline{r_i})^j$ for $j = 1,2,3,\ldots$ is then:

$$E\left[\left(\sum_1^N \overline{r_i}\right)^j\right] = E[(N\bar{r})^j] = N^j\bar{r}^j. \tag{31}$$

Eqs. (31) and (29) are then used in Eq. (28) to get the kurtosis of a multi-hop-distance consisting of N single-hop-distances.

4. Estimation of the probability mass function of hop distance for a given euclidean distance

In this section, the pmf $P_{\mathbf{N}|D}(N|D)$ of hop distance $\mathbf{N}$ corresponding to a given Euclidean distance D is estimated in a one-dimensional sensor network. Using Bayes' Rule, $P_{N|D}(N|D)$ can be stated as follows:

$$P_{N|D}(N|D) = \frac{P_{D|N}(D|N)P(N)}{\sum_{i=1}^{\infty} P_{D|N}(D|i)P(i)}. \tag{32}$$

In Eq. (32), $P_{D|i}(D|i)$ is the probability of reaching distance D in i hops and $P(i)$ is the probability of the hop distance i. Note that hop distance $i = 0$ is not included in the sum at the denominator of Eq. (32) since this is equivalent to loss of connectivity. $P(i)$ is the probability of the hop distance i and is equal to the ratio of the expected number of nodes that have a hop distance i to the total number of nodes in the network. Hence, using the definition of maximal distance d_i in Section 2, $P(i)$ can be found by

$$P(i)=\frac{\text{Expected number of nodes found on }[d_{i-1},d_{i-1}+R]}{\text{Total number of nodes}}. \tag{33}$$

Since sensor nodes are uniformly distributed with density λ and sensor nodes have equal communication range R, the expected number of nodes that can be found on any distance interval of length R is a function of R and λ, and is independent of the location of the interval. Hence, the numerator of Eq. (33) is the same for all $i = 1,2,3,\ldots$ and $P(i)$ is uniformly distributed with respect to hop distance i. Therefore, Eq. (32) can be simplified to

$$P_{\mathbf{N}|D}(N|D) = \frac{P_{D|N}(D|N)}{\sum_{i=1}^{\infty} P_{D|N}(D|i)}. \tag{34}$$

The term $P_{D|i}(D|i)$ in Eq. (34) is the probability of reaching distance D for the given hop distance i. The denominator in Eq. (34) includes all positive integer values of hop distance i. Note that $P_{D|N}(D|N)$ is different than the pdf of the maximum distance covered in N hops which is approximated as following a Gaussian distribution as shown in Section 3. However, this sum includes redundant terms and can be reduced using Lemma 1 as will be presented shortly.

4.1. Preliminary discussion

Before proceeding with the derivation of $P_{\mathbf{N}|D}(N|D)$, some definitions given in previous sections need to be re-emphasized and some additional ones provided. In Fig. 4a, the source node S broadcasts a packet in a one-dimensional sensor network. Multiple nodes can receive the packet at the same number of hops. Such nodes have the same hop distance to S. For a particular hop distance i, let the set of nodes that are i hops away from S be denoted as S_i. Among the sensors in S_i, one sensor node, P_i has the maximum distance to S. The distance of P_i to S is $\mathbf{d_i}$, and the single-hop-distance (the distance between nodes P_{i-1} and P_i) of hop i is $\mathbf{r_i}$, as defined in Section 2. Hence, $\mathbf{d_i}$ is defined as the maximum distance covered in i hops. For instance, $\mathbf{d_4} = d_4$ in Fig. 4a is the maximum hop distance to source node S among all nodes in set S_4.

In Fig. 4b, set S_4 consists of nodes X, Y, and P_4. The multi-hop-distance d_4 is the distance of P_4 to node S. On the other hand, the distance values D_X and D_Y of nodes X and Y to S are less than d_4.

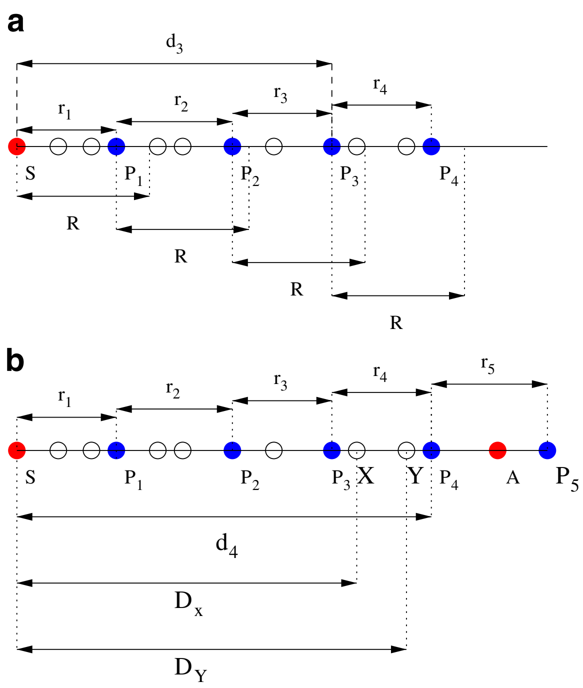


Fig. 4. Definitions of P_i, $\mathbf{d_i}$, $\mathbf{r_i}$, and redundant hop.

Therefore, these nodes do not contribute to the multi-hop-distance $\mathbf{d_4}$ covered in *4* hops. In fact, a multi-hop-distance, as defined in Section 2, is possible by hopping through nodes P_i, $i = 1,2,3,\ldots$, which provide the furthest single-hop-distance in each hop. Hence, the single hops from P_3 to X and from P_3 to Y are "redundant" in forming the multi-hop-distance $\mathbf{d_4}, \mathbf{d_5}, \ldots$, etc. Such single hops are called "redundant hops".

To find the hop distance between a node A and source S in Fig. 4b, the single-hop-distance values $r_1, r_2, r_3, \ldots$ are taken into account. Hence, $P_1, P_2, P_3, \ldots$ are the nodes of the multi-hop path from S to A. The last hop from P_4 to A in Fig. 4b, does not provide the maximum distance in hop 5, since node P_5 is more distant to P_4 compared to the distance between nodes A and P_4. In general, the last hop distance does not have to be maximized since the selected node, such as node A in Fig. 4b, does not have to be one of the nodes P_i, $i = 1,2,3,\ldots$

4.2. Mathematical foundations for the derivation of $P_{\mathbf{N}|D}(N|D)$

In this section, some mathematical axioms and lemmas that are required for deriving $P_{\mathbf{N}|D}(N|D)$ are provided. All lemmas apply to one-dimensional sensor networks with random sensor locations. The following definitions are used throughout the section.

- D: the given Euclidean distance for which $P_{\mathbf{N}|D}(N|D)$ is to be determined.
- R: sensor communication range.
- S: source node.
- S_i: set of all nodes that are at a hop distance i to S.
- P_i: the furthest node to S in S_i.
- $\mathbf{d_i}$: (multi-hop-distance) maximum distance covered in i hops (also equal to the distance of P_i to S).
- $\mathbf{r_i}$: (single-hop-distance) maximum distance covered in the ith hop.
- $\mathbf{m_i}$: minimum distance covered in i hops.

Lemma 1. $\mathbf{N}$ *is upper bounded by an integer* $N_{\max}$ *and lower bounded by an integer* $N_{\min}$ *such that* $N_{min} = \lceil \frac{D}{R} \rceil \leqslant \mathbf{N} \leqslant N_{max} = 2\lfloor \frac{D}{R} \rfloor + 1$.

Proof of Lemma 1. For a single-hop-distance $\mathbf{r_i}$, $\xi < \mathbf{r_i} < 1$ with $0 < \xi \ll 1$ is true by definition. The multi-hop-distance is minimized without having any redundant hops if $R < \mathbf{r_i} + \mathbf{r_{i+1}} = R + \xi$ holds for all consecutive single-hop-distance pairs. The minimum multi-hop-distance is then equal to the sum $\mathbf{r_1} + \mathbf{r_2} + \mathbf{r_3} + \mathbf{r_4} + \cdots = \xi + R + \xi + R + \cdots$.

Hence, the multi-hop-distance for a given hop distance M is lower bounded by $\lfloor \frac{M}{2} \rfloor R$. On the other hand, if all single-hop-distance values are chosen to be $\mathbf{r_i} = R$, then maximum distance is covered and is equal to MR. Therefore, the multi-hop-distance for M hops is bounded as

$$\left\lfloor \frac{M}{2} \right\rfloor R \leqslant \sum_{i=1}^{M} \mathbf{r_i} \leqslant MR. \tag{35}$$

For a given multi-hop-distance D, the following expression holds for at least one integer N:

$$\left\lfloor \frac{N}{2} \right\rfloor R \leqslant D \leqslant NR, \tag{36}$$

$$\left\lfloor \frac{N}{2} \right\rfloor \leqslant \frac{D}{R} \leqslant N. \tag{37}$$

Since N is an integer and $\frac{D}{R} \leqslant N$, the minimum number of hops $N_{\min}$ that covers distance D is

$$N_{\min} = \left\lceil \frac{D}{R} \right\rceil \leqslant N. \tag{38}$$

Similarly, since N is an integer and $\lfloor \frac{N}{2} \rfloor \leqslant \frac{D}{R}$, the maximum number of hops $N_{\max}$ that covers distance D satisfies

$$\left\lfloor \frac{N_{\max}}{2} \right\rfloor = \left\lfloor \frac{D}{R} \right\rfloor. \tag{39}$$

Hence,

$$\left\lfloor \frac{D}{R} \right\rfloor \leqslant \frac{N_{\max}}{2} < \left\lfloor \frac{D}{R} \right\rfloor + 1, \tag{40}$$

$$2\left\lfloor \frac{D}{R} \right\rfloor \leqslant N_{\max} < 2\left\lfloor \frac{D}{R} \right\rfloor + 2. \tag{41}$$

Eq. (41) is satisfied by $N_{\max} = 2\lfloor \frac{D}{R} \rfloor$ and $N_{\max} = 2\lfloor \frac{D}{R} \rfloor + 1$. The latter choice gives a larger N. Hence, $\mathbf{N}$ is lower bounded by an integer $N_{\min}$ and upper bounded by an integer $N_{\max}$ as

$$\left\lceil \frac{D}{R} \right\rceil \leqslant \mathbf{N} \leqslant 2\left\lfloor \frac{D}{R} \right\rfloor + 1, \tag{42}$$

which concludes the proof of Lemma 1. □

According to Lemma 1 and Eq. (42), $P_{\mathbf{N}|D}(N|D)$ given by Eq. (34) reduces to

$$P_{\mathbf{N}|D}(N|D) = \frac{P_{D|N}(D|N)}{\sum_{i=\lceil \frac{D}{R} \rceil}^{2\lfloor \frac{D}{R} \rfloor + 1} P_{D|i}(D|i)}, \tag{43}$$

where $\lceil \frac{D}{R} \rceil \leqslant N \leqslant 2\lfloor \frac{D}{R} \rfloor + 1$ and R is the sensor communication range.

Computation of $P_{\mathbf{N}|D}(N|D)$ with Eq. (43) requires the computation of the term $P_{D|i}(D|i)$, where $i \in \{1, 2, 3, \ldots\}$. $P_{D|i}(D|i)$ stands for the probability of reaching D at a given hop distance i, as defined before. It must be noted that $P_{D|i}(D|i)$ is not equal to $f_{d_i}(d_i)$ which is the probability of reaching a maximum distance of d_i at a given hop distance i. However, derivation of $P_{D|i}(D|i)$ can be made using $f_{d_i}(d_i)$ using Lemma 3 as will be provided shortly. Since the proof of Lemma 3 requires Lemma 2, Lemma 2 is first provided as follows.

Lemma 2. *If a sensor node A in* S_N *has a distance D to source node S, then* $D - R \leqslant \mathbf{d_{N-1}} < D$.

Lemma 2 states that if a node A has a Euclidean distance D and a hop distance N from the source node S, then P_{N-1} must be located within one communication range of A towards S. However, this is a necessary but not sufficient condition for node A to have a hop distance of N to S and Lemma 3 provides the sufficient conditions. The following is the proof of Lemma 2.

Proof of Lemma 2. Assume $\mathbf{d_{N-1}} < D - R$. Then, $\mathbf{d_{N-1}} + R < D$. Since $\mathbf{d_{N-1}} < \mathbf{d_N} \leqslant \mathbf{d_{N-1}} + R$, $\mathbf{d_N} < D$ must hold. However, this implies that the hop distance of node A to S is larger than N. Hence, $\mathbf{d_{N-1}} \geqslant D - R$ by contradiction.

Secondly, assume $\mathbf{d_{N-1}} > D$. Since $m_N > \mathbf{d_{N-1}}$, $m_N > D$ holds with this assumption. However, this implies that A is not in S_N. Hence, for A to be in S_N, $\mathbf{d_{N-1}} < D$ must hold. □

Using Lemma 2, Lemma 3 can be presented and proved as follows:

Lemma 3. *A node A at a distance D to the source node S has a hop distance N from S, if* $D - R \leqslant \mathbf{d_{N-1}} < D$ *and if* $\mathbf{d_j} < D - R$ *for all* $j < N - 1$.

Lemma 3 simply states that for a node A, which is a distance D away from the source node S, to be in the set S_N, there are two necessary and sufficient conditions. The first condition is that node P_{N-1} must be located within one communication range of A towards S. In other words, $D - R \leqslant \mathbf{d_{N-1}} < D$. The second condition is that all the previous maximal points P_j before $N - 1$ with $j < N - 1$ must be located at more than one communication range distance to A towards S. Hence, $0 < \mathbf{d_j} < D - R$.

Proof of Lemma 3. According to Lemma 2, $D - R \leqslant \mathbf{d_{N-1}} < D$ if node A has a hop distance N to source node S. This is shown in Fig. 5a. However, the reverse statement is not necessarily true. By definition, $\mathbf{d_j} < \mathbf{d_{N-1}}$ for all $j < N - 1$. However, if $D - R \leqslant \mathbf{d_j} < d_{N-1}$ holds true for any $j < N - 1$, then node A is in the communication range of P_j, since $\mathbf{d_j} < D \leqslant \mathbf{d_j} + R$, as shown in Fig. 5b. This requires node A to have a hop distance $j + 1$ to S, but not N. Furthermore, P_{N-1} is now in the communication range of P_j, which forces P_{N-1} to be in S_{j+1}. However, this is a contradiction since if $j < N - 2$, then point P_{N-1} becomes a previous furthest point P_{j+1}. Moreover, if $j = N - 2$, then P_{N-1} is either node A, which is further away, or some other node even further than node A within the communication range of P_j. Hence, $\mathbf{d_j} < D - R$ for all $j < N - 1$ must be true in addition to the requirement that $D - R \leqslant \mathbf{d_{N-1}} < D$. □

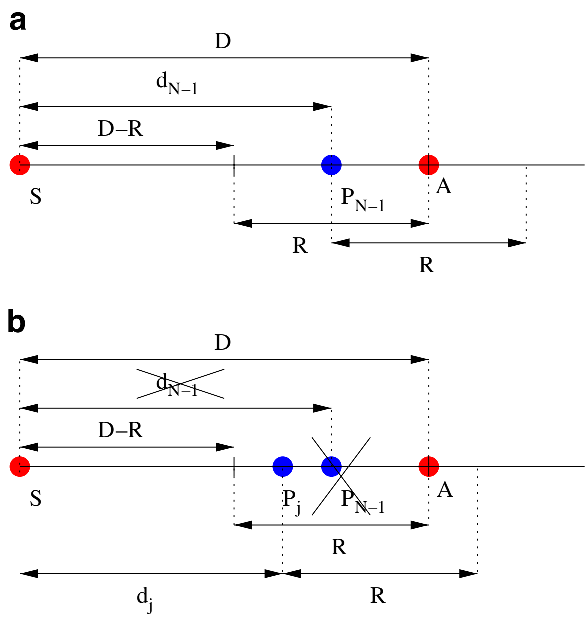


Fig. 5. Proof of Lemma 3.

4.3. Derivation of $P_{\mathbf{N}|D}(N|D)$

According to Lemma 3, the probability of reaching distance D at a given hop distance i is obtained as follows:

$$P_{D|i}(D|i) = \text{Prob}(D - R \leqslant \mathbf{d_{i-1}} < D) \cdot \prod_{j=1}^{i-2} \text{Prob}(0 < \mathbf{d_j} < D - R). \quad (44)$$

Computation of Eq. (44) requires the pdf of d_i, $p_{\mathbf{d_i}}(d_i)$, for $i \in \{1, 2, 3, \ldots\}$. In previous sections, $p_{\mathbf{d_i}}(d_i)$ is shown to be well estimated with a Gaussian distribution. The computation of $p_{\mathbf{d_i}}(d_i)$ requires only the mean multi-hop-distance, $E[\mathbf{d_i}]$, and the standard deviation of the multi-hop-distance, $\sigma_{\mathbf{d_i}}$, as shown in the following equation:

$$p_{\mathbf{d_i}}(d_i) = \frac{1}{\sqrt{2\pi}\sigma_{\mathbf{d_i}}} \mathrm{e}^{\frac{-(d_i - E[\mathbf{d_i}])^2}{2\sigma_{\mathbf{d_i}}^2}}, \quad (45)$$

where $E[\mathbf{d_i}]$ and $\sigma_{\mathbf{d_i}}$ are found by Eqs. (16) and (17), respectively.

Using the Gaussian pdf in Eqs. (45) and (44) can be written as

$$\text{Prob}(D \text{ in } i \text{ hops}) = \left(\int_{D-R}^{D} p_{\mathbf{d_{i-1}}}(x) \mathrm{d}x \right) \cdot \prod_{j=1}^{i-2} \left(\int_{0}^{D-R} p_{\mathbf{d_j}}(x) \mathrm{d}x \right). \quad (46)$$

Finally, Eq. (46) is substituted in Eq. (43) for individual values of hop distance i to compute $P_{\mathbf{N}|D}(N|D)$ as follows:

$$P_{\mathbf{N}|D}(N|D) = \frac{\left(\int_{D-R}^{D} p_{\mathbf{d_{N-1}}}(x)\mathrm{d}x\right) \cdot \prod_{j=1}^{N-2}\left(\int_{0}^{D-R} p_{\mathbf{d_j}}(x)\mathrm{d}x\right)}{\sum_{i=\lceil\frac{D}{R}\rceil}^{2\lceil\frac{D}{R}\rceil}\left[\left(\int_{D-R}^{D} p_{\mathbf{d_{i-1}}}(x)\mathrm{d}x\right) \cdot \prod_{j=1}^{i-2}\left(\int_{0}^{D-R} p_{\mathbf{d_j}}(x)\mathrm{d}x\right)\right]}. \quad (47)$$

The integral terms in the computation of $P_{\mathbf{N}|D}(N|D)$ in Eq. (47) can be substituted by equivalent "error functions" since the integrants are Gaussian pdfs. The following two definitions are used to simplify Eq. (47):

$$Y_1(i) = \mathrm{erf}\left(\frac{D - E[\mathbf{d_i}]}{\sigma_{\mathbf{d_i}}\sqrt{2}}\right) - \mathrm{erf}\left(\frac{D - R - E[\mathbf{d_i}]}{\sigma_{\mathbf{d_i}}\sqrt{2}}\right), \quad (48)$$

$$Y_2(i) = \mathrm{erf}\left(\frac{D - R - E[\mathbf{d_i}]}{\sigma_{\mathbf{d_i}}\sqrt{2}}\right) + \mathrm{erf}\left(\frac{E[\mathbf{d_i}]}{\sigma_{\mathbf{d_i}}\sqrt{2}}\right). \quad (49)$$

Eq. (50) below is the final form of the conditional probability mass function $P_{\mathbf{N}|D}(N|D)$ using Eqs. (48) and (49):

$$P_{\mathbf{N}|D}(N|D) = \frac{Y_1(N-1)\prod_{j=1}^{N-2} Y_2(j)}{\sum_{i=\lceil\frac{D}{R}\rceil}^{2\lfloor\frac{D}{R}\rfloor+1}\left(Y_1(i-1)\prod_{j=1}^{i-2} Y_2(j)\right)}. \quad (50)$$

The computation of $P_{\mathbf{N}|D}(N|D)$ in Eq. (50) is simple with the use of error functions. The parameters of Eq. (50) are the given sensor communication range R, multi-hop distance D, and the values $E[\mathbf{d_i}]$ and $\sigma_{\mathbf{d_i}}$, $i \in 1, 2, 3, \ldots$, which are computed by Eqs. (16) and (17).

5. Performance evaluation

In this section, the numerical results of theoretical and approximated expressions for single-hop and multi-hop Euclidean distance distributions $f_{\mathbf{r}}(r)$ and $f_{\mathbf{d_N}}(d_N)$ are introduced. Experimental results of distance distributions are provided and compared with the theoretical and approximation results to evaluate their validity. Furthermore, the derived approximation of the hop distance distribution $P_{\mathbf{N}|D}(N|D)$ is evaluated by comparing it with the experimentally obtained hop distance pmf.

The expressions that are evaluated are as follows:

- Expected single-hop and expected multi-hop-distances (Eq. (7)).
- Standard deviation of single-hop and multi-hop-distances (Eq. (8)).
- Kurtosis of single-hop-distance (Eq. (19)).
- Kurtosis for multi-hop-distance (Eq. (28)).
- Probability mass function of hop distance for a given Euclidean distance (Eq. (50)).

First, which particular mathematical expressions are used in performance evaluation is explained. Then, the obtained graphs for distance distributions are provided and discussed. Theoretical, experimental, and approximation results are compared to demonstrate the validity of our approximations. Afterwards, how the kurtosis graphs reflect the Gaussian behavior of multi-hop-distance distributions is discussed. Finally, the approximation for hop distance pmf is compared with experimental results.

5.1. Implementations

5.1.1. Theoretical expressions

The theoretical distribution expressions of the expected single-hop-distance $E[\mathbf{r}]$ and expected multi-hop-distance $E[\mathbf{d_N}]$ are implemented according to Eq. (7). To find the expected single-hop-distance of a particular hop, say k, Eq. (7) is evaluated for $N = k$ and $N = k - 1$ separately. The difference of these two results gives the theoretical expectation of the single-hop-distance in hop k. Similarly, the theoretical expressions of the standard deviation of single-hop-distance $\sigma_{\mathbf{r}}$ and standard deviation of multi-hop-distance $\sigma_{\mathbf{d_N}}$ are implemented according to Eq. (8). Taking $N = 1$, the standard deviation of the first single-hop-distance is obtained, while for $N > 1$ the standard deviation of multi-hop-distance is found.

The maximum number of hops that is implemented for multi-hop-distance case is limited to 5 due to the computational cost of the multi-hop expressions which are in the form of nested integrals. As will be demonstrated in the following section, our approximation methods applied to these expressions are quite accurate.

5.1.2. Approximation expressions

The approximation expressions are derived to avoid the computational cost of the theoretical ones. Eq. (10), which is the assumption in [1], is used to obtain the expected single-hop-distance (Eq. (11)) and expected multi-hop-distance (Eq. (16)). Using the approximated single-hop-distance value of Eq. (10), the standard deviation of single-hop and multi-hop-distances (Eqs. (13) and (17)) are calculated. Furthermore, kurtosis of single-hop-distance and multi-hop-distance are calculated according to (Eq. (19)) and Eq. (28), respectively. The pmf

$P_{\mathbf{D}|N}D|N$ of hop distance for a given Euclidean distance is approximated using Eq. (50).

5.1.3. Experimental study

In order to evaluate the validity of the theoretical and approximation results, 10,000 independent experiments are performed using one-dimensional spatially random sensor networks with uniform node density and fixed sensor communication range.

5.2. Results

5.2.1. Expectation and standard deviation of single-hop-distance

In Fig. 6, the theoretical value of the expected single-hop-distance $E[\mathbf{r}]$ is found to be approaching to the approximation. The hop distances are found to have a decaying oscillatory character around $E[\mathbf{r}]$. This shows that the approximation in Eq. (10) is not accurate for the first few hops and improves for larger hop numbers to match the approximated value. Due to the computational cost of the numerical implementation of the theoretical expressions, the expected distance values for up to five hops is shown in this figure.

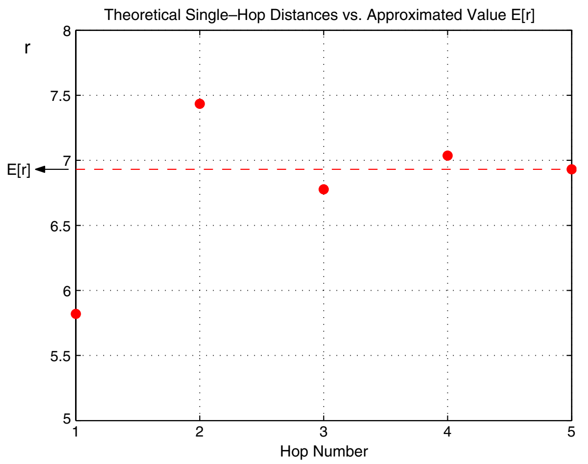


Fig. 6. Comparison of theoretical values of r_i with $E[r]$. $R = 10$, $\lambda = 0.1$, and $i = 1, 2, 3, 4, 5$.

In Fig. 7a, the analytical, experimental, and approximated values of the expectation of a single-hop-distance are presented. As it can be observed, the approximated values are very accurate since the curve of the theoretical values is closely matched. The approximated and the theoretical results resemble the experimental results almost perfectly, which shows their validity. Fig. 7b illustrates the comparison of the analytical, experimental, and approximated values of the standard deviation of a single-hop-distance. As it is the case in Fig. 7a, the approximated and the theoretical results match the experimental results almost perfectly, which demonstrates the success of the approximations and the validity the theoretical expressions for calculating standard deviations.

5.2.2. Expectation and standard deviation of multi-hop-distance

In Fig. 8, the change in the standard deviation of multi-hop-distance in N hops for changing node density λ is shown for $N = 1, 2, 3, \ldots, 10$. The increase in deviation with decreasing node density is apparent. This shows that the distribution of

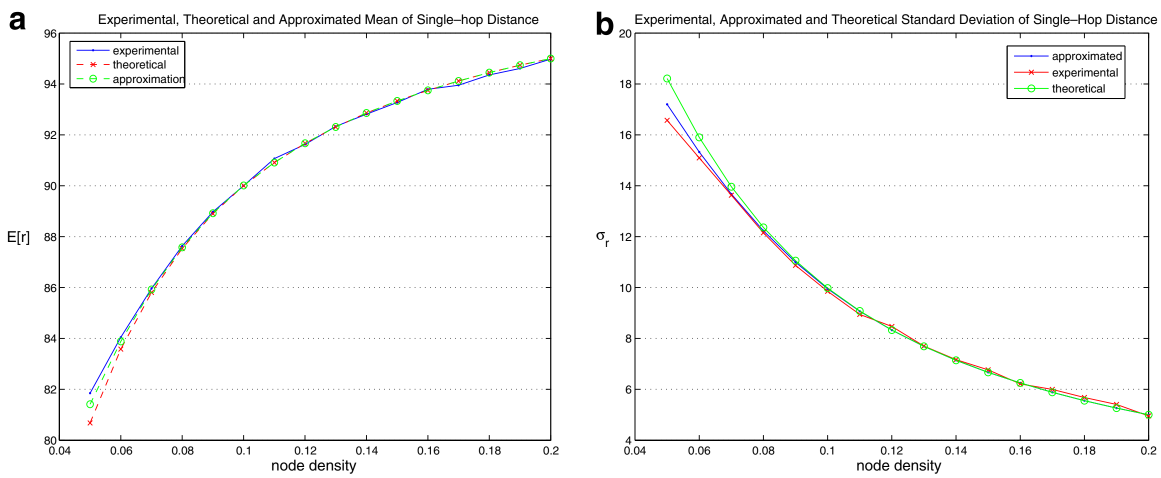


Fig. 7. Comparison of theoretical, approximated and experimental $E[r]$ and $\sigma_{\mathbf{r}}$ for changing λ. $R = 100$: (a) $E[r]$ and (b) σ_r.

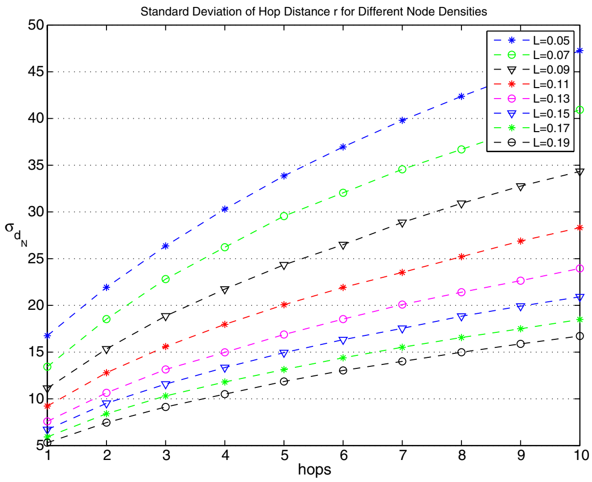


Fig. 8. Experimental values of the standard deviation of the multi-hop-distance with N hops, $\sigma_{\mathbf{d_N}}$, for $N = 1, 2, 3, \ldots, 10$, in case of different node densities. $R = 100$.

multi-hop-distance becomes less concentrated around its mean as the node density decreases. This is an expected result since less node density causes more randomness in sensor locations. Another observation in Fig. 8 is that the standard deviation curves are not linearly proportional to the number of hops N. This is an indication that the *single-hop-distances are not independent*, since the standard deviation of the sum of identically distributed independent random variables is $\sigma_Y = M\sigma_X$, where X is a single random variable and Y is the sum of M identically distributed single random variables. This is also an expected result, since it is known that the distribution of the single-hop-distance in a hop is dependent on the single-hop-distance in the previous hop.

5.2.3. Effect of node density on multi-hop-distance

Fig. 9a shows the change of the mean of a multi-hop-distance for two selected node densities. The results of theoretical, approximated, and experimental implementations are compared. Since the computations of the integral expressions are costly, results of up to $N = 5$ hops are illustrated. As it can be observed, the three results overlap which shows the validity of our approximation for the expected multi-hop-distance value. The expected value of multi-hop-distance is found to be linearly proportional to the mean of single-hop-distance as suggested in Eq. (16).

Fig. 9b, illustrates the expected distance of N hops, $N = 1, 2, 3, \ldots, 10$, obtained by experimentation and approximation. In order to present a larger number of hops, theoretical results are not shown in Fig. 9b. The standard deviations $\sigma_{\mathbf{d_N}}$ of the total distance $\mathbf{d_N}$ are shown by using error bars around the expected distance values. The figure shows the results for selected high ($\lambda = 0.2$ nodes/m) and low ($\lambda = 0.05$ nodes/m) node densities. The experimental and approximated results are quite similar and they overlap. The linearity of experimental expected distance curves suggests that $E[\mathbf{d_N}]$ is linearly proportional to the number of hops as claimed in Eq. (16). The standard deviation values found by Eq. (17) and the experimental values are also quite

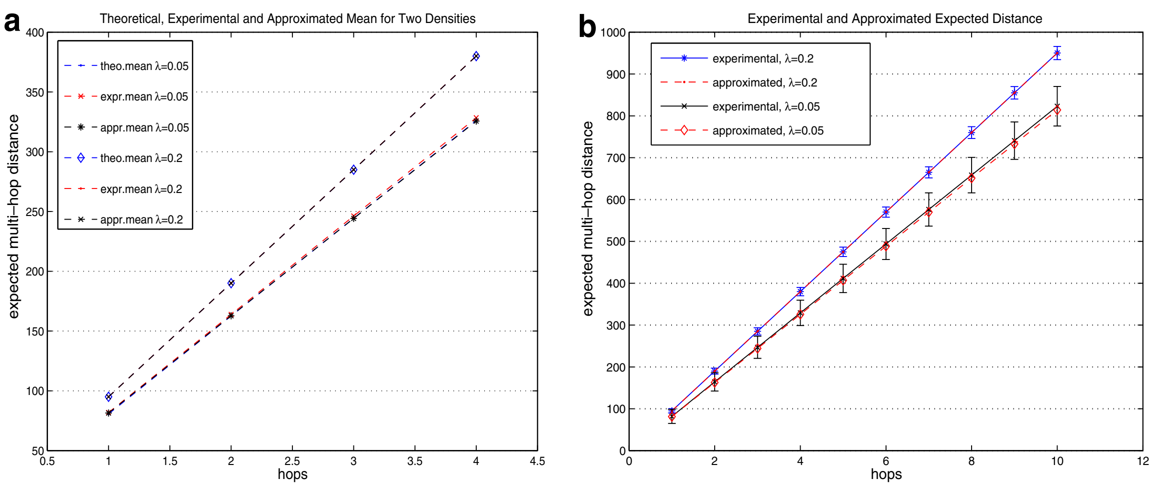


Fig. 9. Comparison of theoretical, approximated and experimental $E[\mathbf{d_N}]$ and $\sigma_{\mathbf{d_N}}$. $R = 100$, $\lambda = 0.05$ nodes/m and 0.2 nodes/m: (a) experimental, approximated and theoretical. $N = 1, 2, 3, 4$ and (b) experimental and approximated. $N = 1, 2, 3, \ldots, 10$.

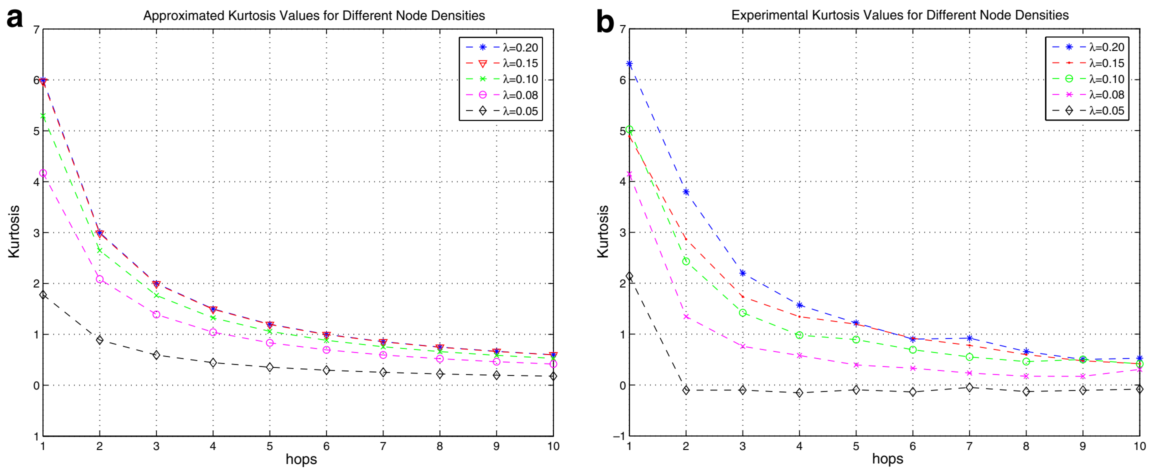


Fig. 10. Effect of node density on Gaussianity, obtained by experiments and approximations. $R = 100$: (a) approximate kurtosis of $\mathbf{d_N}$ and (b) experimental kurtosis of $\mathbf{d_N}$.

similar and they appear to be represented by a single deviation error bar. The increase in deviation with decreasing node density and increasing number of hops can be observed in Fig. 9b, which was also shown in more detail in Fig. 8.

5.2.4. *Gaussianity of multi-hop-distance*

In Fig. 10, it can be observed that the behaviors of the kurtosis curves with changing node density λ and changing number of hops N are similar in experiments and approximation results. Generally, it is observed that Gaussianity of the multi-hop-distance is higher for large number of hops and also for low node density. This is an expected result since Gaussian distribution is the most "random" distribution among those distributions having the same mean and standard deviation and for low densities the distribution of nodes present a more random character. For large densities, the number of nodes found per unit area is close to the overall node density, whereas for smaller node densities this is more unlikely. Fig. 10b illustrates how the kurtosis, hence the Gaussianity of the multi-hop-distance, is affected by the change in node density. This figure is obtained using Eq. (28), hence it shows approximated values. Experimental results for the effect of node density on the Gaussianity of the multi-hop-distance are shown in Fig. 10a.

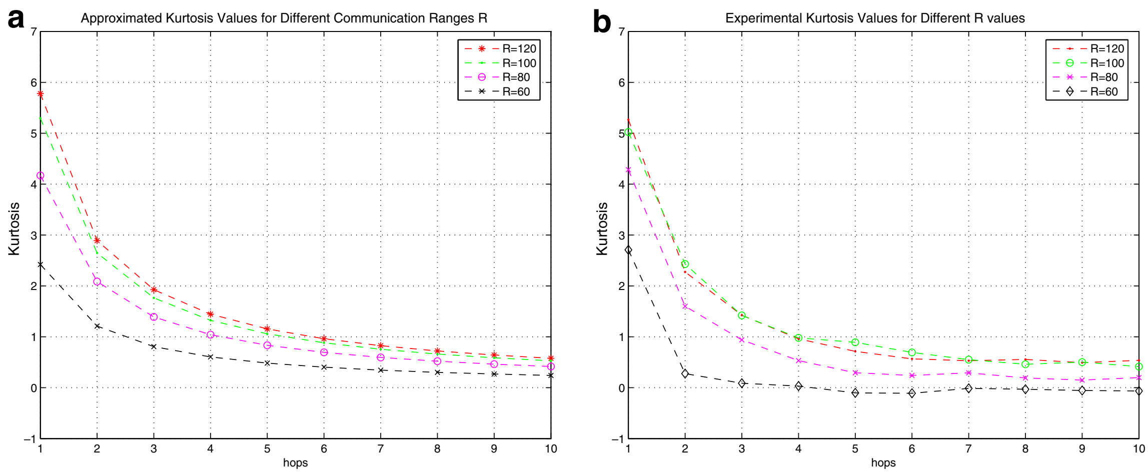


Fig. 11. Effect of communication range on Gaussianity, obtained by experiments and approximations. $\lambda = 0.1$: (a) approximated kurtosis of $\mathbf{d_N}$ and (b) experimental kurtosis of $\mathbf{d_N}$.

Increasing the communication range R is found to cause a decrease in the Gaussianity of multi-hop-distance $\mathbf{d_N}$, as shown in Fig. 11a and b. Fig. 11a illustrates how the non-Gaussianity of multi-hop-distance is affected by the change in sensor communication range R. This figure is obtained by Eq. (28), hence it shows approximated values. The experimental results are illustrated in Fig. 11b. In both figures, the multi-hop-distance distribution becomes more Gaussian for larger number of hops. The consistency between the experimental results and the approximations shows the accuracy of our kurtosis approximation expression given by Eq. (28).

An alternative way of investigating the similarity between two probability distributions is to find the mean square error between them. For a particular distribution p, a decrease in mean square error between p and a Gaussian distribution means an increasing Gaussian character of p. The mean square error (MSE) curves between the multi-hop-distance distribution and Gaussian distribution are

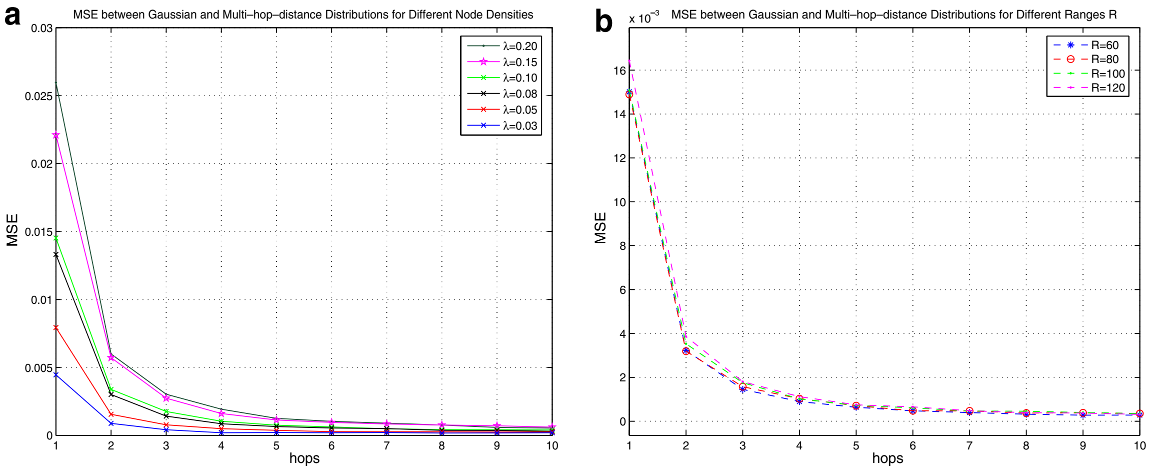


Fig. 12. Mean square error between experimental gaussian distribution and multi-hop-distance distribution: (a) changing λ. $R = 100$ m and (b) changing R. $\lambda = 0.1$ nodes/m.

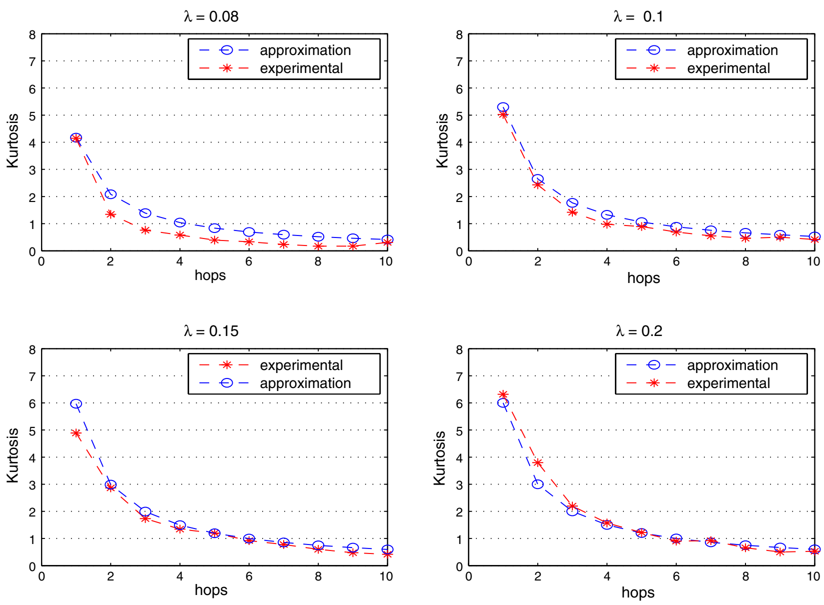


Fig. 13. Comparison of experimental and approximated kurtosis of multi-hop-distance $\mathbf{d_N}$ for changing node density.

shown in Fig. 12a and b for changing node density and changing communication range, respectively. In these figures, the decaying nature of the non-Gaussianity for increasing number of hops is obvious. The MSE results are consistent with the results in Figs. 10b and 11a. In Fig. 12b, changing the R values seems to have little effect on MSE, yet in fact, Gaussianity decreases for increasing R.

The experimental and the approximated values of the kurtosis of the multi-hop-distance $\mathbf{d_N}$ for $N = 1, 2, 3, \ldots, 10$ are compared for four different node densities in Fig. 13. Approximation results are closer to experimental results if node density λ is larger, although a higher Gaussianity of multi-hop-distance is observed for low node density. In Fig. 14, the effect of changing the communication range R on the kurtosis of $\mathbf{d_N}$ is shown. Approximation results are closer to experimental results if communication range R is larger, although a higher Gaussianity of multi-hop-distance is observed for smaller communication ranges. In general, the graphs illustrate that the approximation results are found to be similar to experimental results.

5.2.5. The pmf $P_{\mathbf{N}|D}(N|D)$

The pmf $P_{\mathbf{N}|D}(N|D)$ of hop distance $\mathbf{N}$ for a given Euclidean distance D is calculated by Eq. (50) in Section 4.3. In this section, the comparison of experimental and approximation results obtained for $P_{\mathbf{N}|D}(N|D)$ are provided. Furthermore, we also elaborate on the effect of node density and the choice of the distance D on $P_{\mathbf{N}|D}(N|D)$.

The comparison between the approximated and experimental pmfs $P_{\mathbf{N}|D}(N|D)$ is shown in Fig. 15. In this figure, the horizontal axis represents distance D. In each plot, there are five values of distance D. The third distance value is equal to the maximum distance of the fifth hop $D = E[d_5]$ and the samples of D are chosen with increments of 20 m. The choice of the fifth hop is completely random. The pmf of hop distance N is represented with a corresponding bar plot over each distance value in Fig. 15. As it can be observed, the change in D affects the pmf of N. Furthermore, when D is changed additional N values with a non-zero probability can emerge while some others can disappear. For instance, in Fig. 15a, hop number $N = 4$ exists for $D = 367$ m while it disappears for $D = 427$ m. Moreover, hop number $N = 6$ is negligibly small for $D = 367$ m while it becomes the dominant component of the pmf when $D = 447$ m. In fact, Fig. 15a–c illustrate the transition from a pmf with hop number $N = 5$ being the dominantly large component to a pmf with hop number $N = 6$ being the largest component for node densities of $\lambda = 0.05$ nodes/m, $\lambda = 0.1$ nodes/m, and $\lambda = 0.2$ nodes/m, respectively. Furthermore, the probability of $N = 5$ in Fig. 15a first increases and then decreases with increasing distance D and the rising values of the probability of $N = 6$ can be observed. If the pmf is shown for a larger range of distance D, it is observed that this

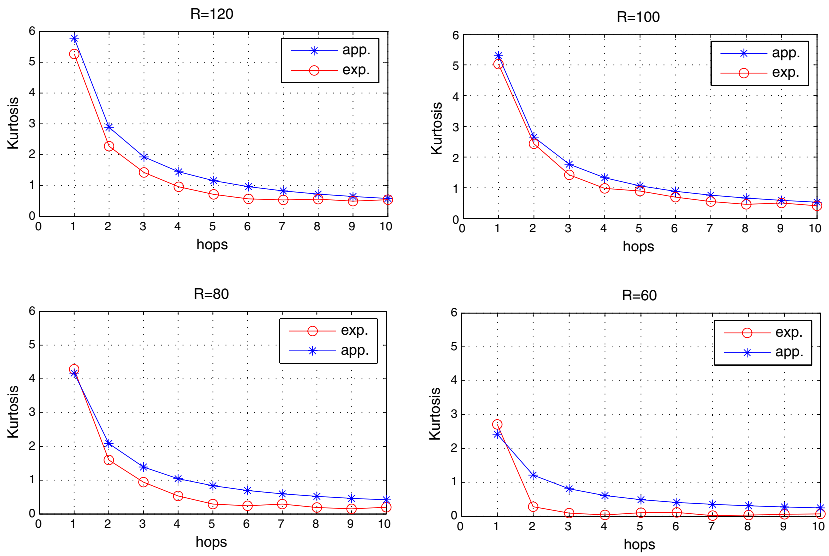


Fig. 14. Comparison of experimental and approximated kurtosis of the multi-hop-distance $\mathbf{d_N}$ for changing communication range.

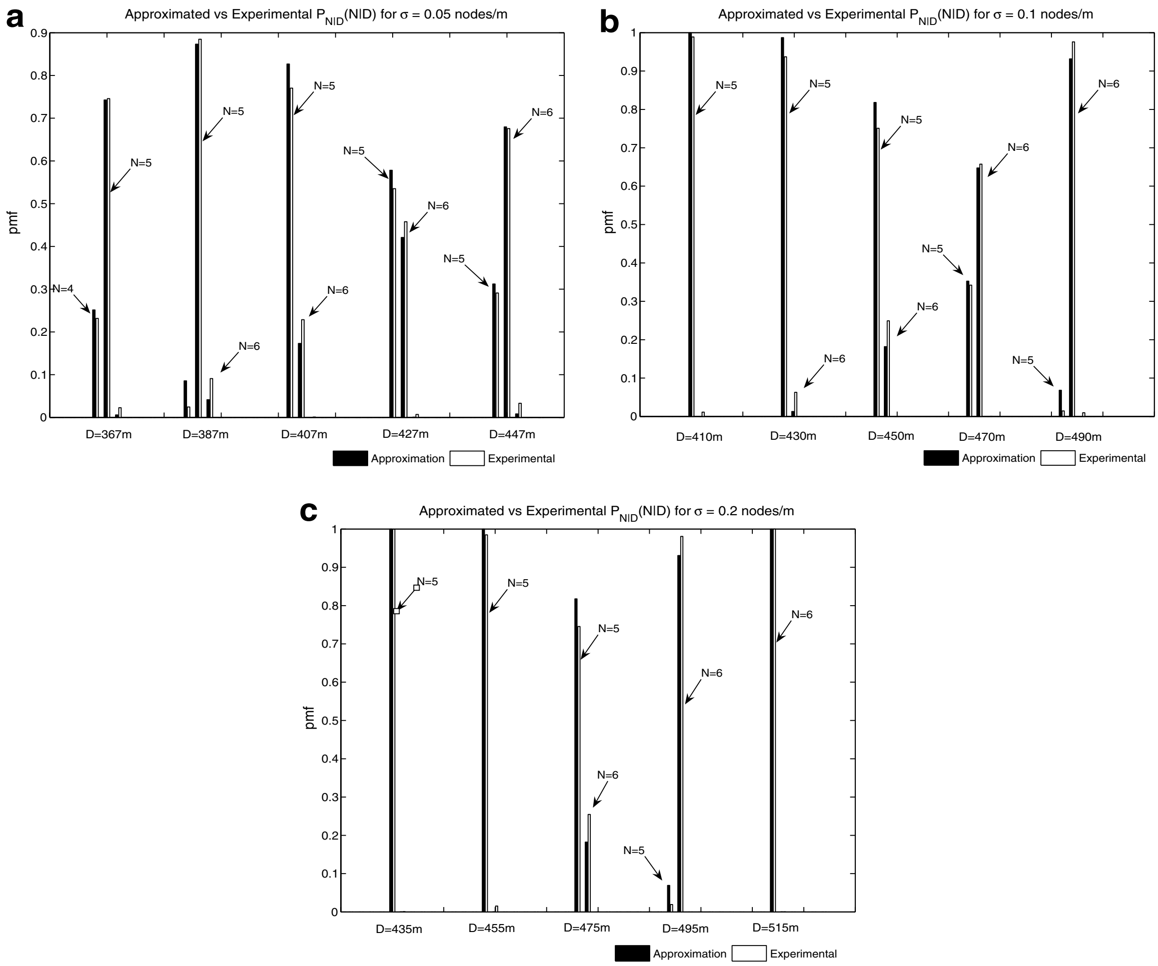


Fig. 15. Comparison between approximated and experimental $P_{\mathbf{N}|D}(P|D)$: (a) $\lambda = 0.05$ nodes/m; (b) $\lambda = 0.1$ nodes/m and (c) $\lambda = 0.2$ nodes/m.

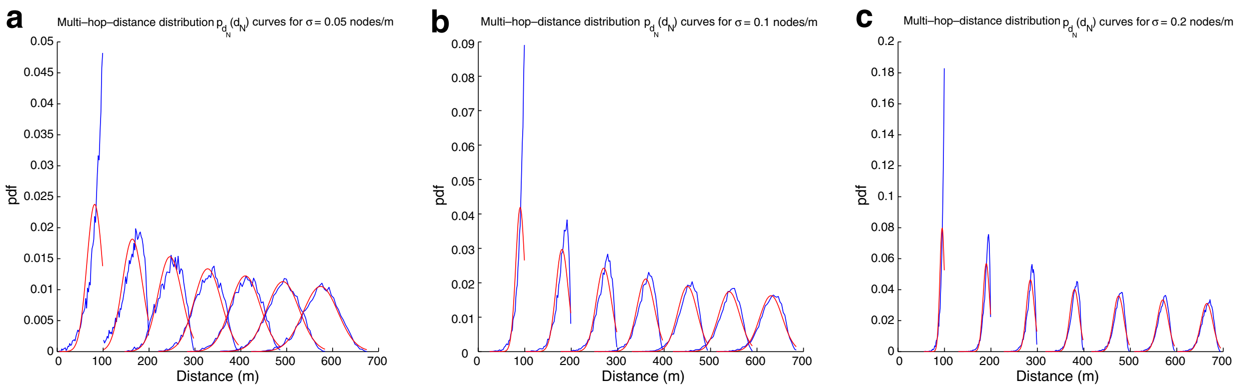


Fig. 16. Effect of node density on multi-hop-distance pdf curves: (a) $\lambda = 0.05$ nodes/m; (b) $\lambda = 0.1$ nodes/m and (c) $\lambda = 0.2$. nodes/m.

behavior of the probability of $N = 5$ is observed for all values of N.

Fig. 15 shows that the largest difference between the experimental and the approximation results is observed at the third distance values in each plot. In fact, these distance values are found at locations where the largest difference between the Gaussian pdf approximation and the multi-hop-distance pdf occurs. Such locations can be observed around the peak of the pdf curves in Fig. 16. These locations designate the expected furthest points of each hop, hence the expected starting point of the next hop. Therefore, it is not surprising to see that these distances also correspond to the starting point of the above mentioned transition in the pmf of hop distance in Fig. 15.

The effect of node density on multi-hop-distance distributions is illustrated in Fig. 16. As observed in this figure, individual multi-hop-distance distribution curves corresponding to different hop numbers become separated from each other with increasing node density. The separation of distance pdf curves suggest that as the node density is increased, the hop distance N corresponding to a given Euclidean distance D becomes more deterministic. In other words, increasing the node density is expected to cause the pmf $P_{\mathbf{N}|D}(N|D)$ to get closer to unity at a single value of N. This effect of node density on pmf can also be observed in Fig. 15. When node density is increased, the pmf of N becomes more deterministic and has one dominantly large component for individual values of distance D as observed in Fig. 15c with λ=0.2 nodes/m except $D = 475$ m where hop distances of $N = 5$ and $N = 6$ are observed simultaneously.

The effect of node density on $P_{\mathbf{N}|D}(N|D)$ is specifically shown in Fig. 17 for three randomly selected values of distance D. Two observations can be made in this figure. Firstly, when the node density is increased, a dominantly large component of the pmf emerges, which makes $P_{\mathbf{N}|D}(N|D)$ more deterministic. Secondly, as the node density increases, the probability given to small hop distance values increases while the probability of larger hop distance values decreases. Here, it must be noted that the range of possible hop distance values N for a given distance D is calculated by Eq. (42). For instance, in Fig. 17 when $D = 443$ and $\lambda = 0.05$ nodes/m, the probability of $N = 6$ is larger than the probability of $N = 5$. However, when $\lambda = 0.1$ nodes/m, probability of $N = 5$ is much larger than that of $N = 6$ and is almost 1 when $\lambda = 0.2$ nodes/m. This can be explained as follows: The multi-hop-distance, hence the covered distance corresponding an individual hop distance N, increases with increasing node density. Therefore, smaller values of N have increasing probability of covering a given distance D. This in turn causes low values of N to have increasing weights in $P_{\mathbf{N}|D}(N|D)$.

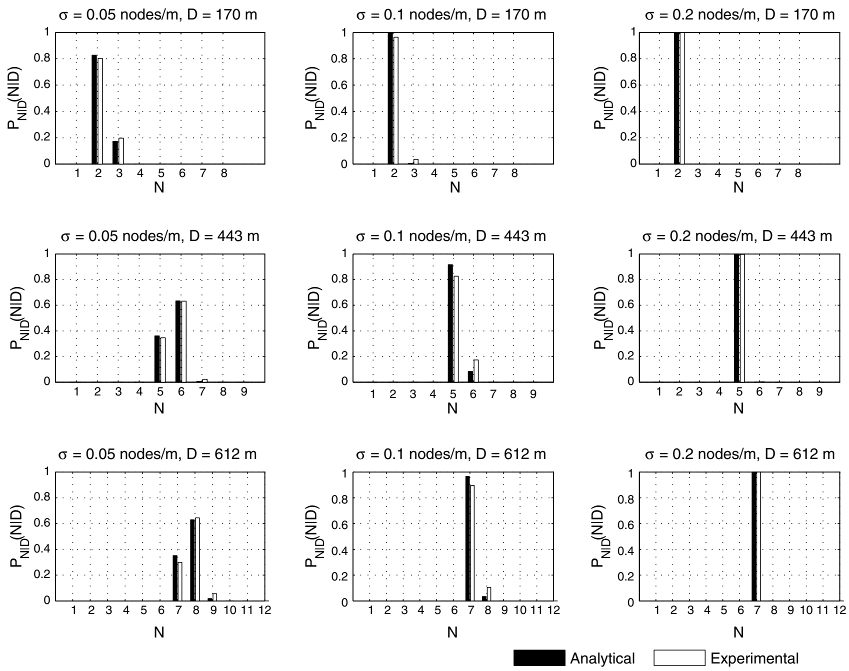


Fig. 17. Experimental and approximated $P_{\mathbf{N}|D}(N|D)$, D: randomly selected.

6. A model for the analysis of two-dimensional maximum multi-hop Euclidean distance

The analysis of the distance distribution in two-dimensional sensor networks poses a more challenging problem compared to the one-dimensional networks. The definition, modeling, and calculation of distance distributions is complicated due to the geometric complexity of these networks. Multiple paths can exist between two given locations and these paths are randomly located due to random node locations. In contrast, the path between two nodes is restricted to a line in a one-dimensional network. Fig. 18 is provided for the purpose of observing the hop distances of sensors to a specific sensor. This sensor may be considered as a source sensor broadcasting packets to the network. The packets propagate radially outward hop by hop and at each hop, a number of sensors receive the broadcast packet. This outwards propagation of hop distance to a specific sensor node further illustrates the increasing Euclidean distances to this sensor for increasing hop distances.

The definition of the maximum Euclidean distance in a two-dimensional network is however not trivial. This definition requires a definition of the direction of propagation. When the propagation direction is fixed, the results of one-dimensional network analysis can be applied to two-dimensional networks. A thin rectangular corridor along the fixed direction of propagation can be used. This thin corridor can be regarded as a line with a small thickness. However, sensors of multi-hop paths are not found on one-dimensional paths in general. Hence, searching the multi-hop paths inside a corridor can provide accurate results only for sufficiently dense networks.

Another model for analyzing the maximum Euclidean distance in two-dimensional networks can be proposed as follows. Fig. 18 shows that longer multi-hop paths can be created by searching the sensor nodes in directions outwards from the source sensor P_1. Hence, the furthest sensors to the source sensor can be located in each hop of a multi-hop path with a given hop distance to maximize the Euclidean distance to the source sensor. This model is shown in Fig. 19. This second model considers a series of single hops to establish a multi-hop path. At each single hop, next sensor of the path is searched within an angular communication area. The angular range of such an area is α and the radial range is R which is the sensor communication range. At each hop, the furthest sensor to the source sensor is located within an angular communication area which is directed outwards from the direction from the source sensor. Furthermore,

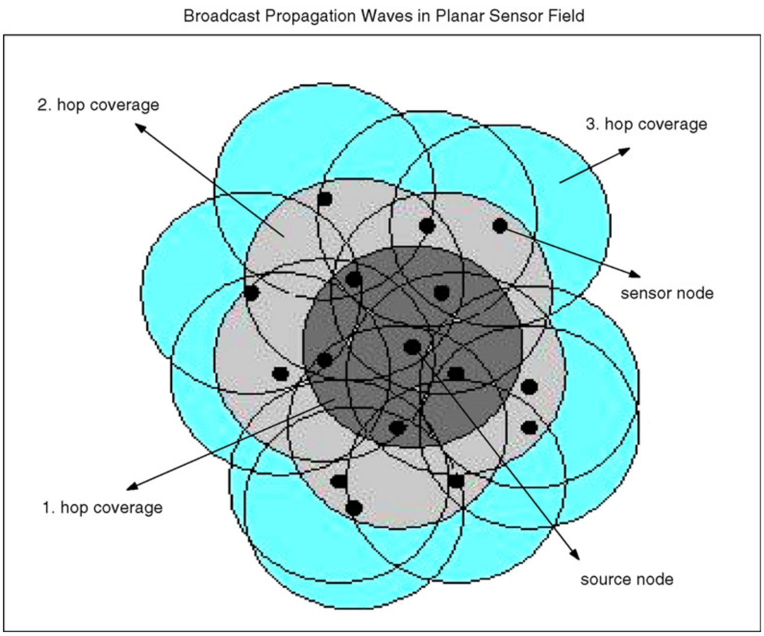


Fig. 18. Coverage areas of broadcast propagation for three hops in a two-dimensional network.

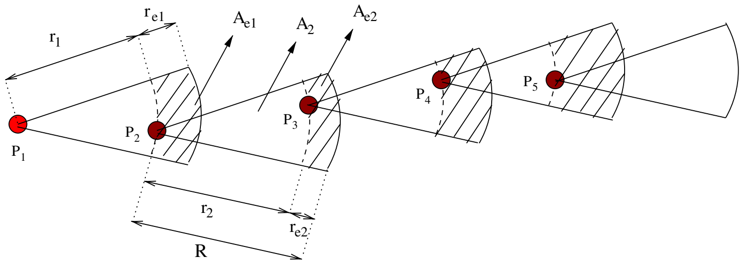


Fig. 19. Directional propagation model.

an area, designated as A_i in Fig. 19, is known to be vacant as the furthest node P_i is located at hop i.

Since the direction of propagation changes at each hop but remains to be outwards from the source sensor, this model is named as the "Directional Propagation Model". The vacant area A_i in this model are analogous to vacant line segment length r_{e_i} in the one-dimensional analysis. Furthermore, the length of a single hop r_i corresponds to the radial distance of the located sensor P_i to the previous sensor P_{i-1}. However, multi-hop-distance is not the scalar sum of the single-hop-distance values. The model should consider the change of the direction angle at each hop of the multi-hop path to determine the multi-hop-distance. Directional Propagation Model is a topic of future study to approximate the distribution of maximum Euclidean distances corresponding to given hop distances for two-dimensional sensor networks.

7. Conclusion

In this paper, the probability distribution of the maximum Euclidean distance for a given hop distance is inspected in case of a one-dimensional sensor network with uniform distribution of sensors. The maximum distance distribution provided is related with the information propagation capability in sensor networks. Hence, it can be applied to applications such as coverage area estimation, and distance estimations in emergency scenarios where information should be quickly moved to as far locations as possible. Furthermore, the reverse problem of estimation of the hop distance between any two location is an equally important topic. Hop distance designates the least number of hops over all multi-hop paths between two locations in a sensor network and provides valuable information to network engineers. Applications such as estimation of multi-hop transmission delays between two given sensor nodes and multi-hop power consumption in packet transmissions are directly related with hop distance. In this paper, the relation between hop distance and maximum Euclidean distance is approximated with Gaussian pdf and analytical methods to calculate the parameters of the Gaussian pdf are provided. Furthermore, the maximum Euclidean distance distribution is used to estimate the distribution of the hop distance corresponding to a given Euclidean distance.

The analysis can be summarized as follows. Approximations of the mean and standard deviation of the distance in a single hop is used to derive approximation expressions for the distance in multiple hops. Additionally, the similarity between the multi-hop-distance distribution and the Gaussian distribution is inspected and Gaussianity of multi-hop distance is shown to be higher for increasing hop distance values. A statistical measure called "kurtosis" is used for measuring the Gaussianity of multi-hop-distance and an effective way of approximating the kurtosis formula for multi-hop-distance is derived. The effect of node density and sensor communication range on the expected multi-hop distance and Gaussianity of multi-hop-distance is investigated. Furthermore, an approximation expression for the pmf of the hop distance corresponding to a given distance is derived using the multi-hop-distance distribution modeled with the Gaussian pdf. Comments on the effect of node density and the choice of the given Euclidean distance on the pmf of hop distance are provided.

An extensive number of experiments are conducted for the purpose of evaluating the validity of our approximation expressions. The approximations and the numerical results of the theoretical expressions are found to be highly consistent with the experimental results.

References

[1] Y. Cheng, T. Robertazzi, Critical connectivity phenomena in multihop radio models, IEEE Transactions on Communications 37 (1989) 770–777.

[2] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, 2001.

[3] G. Korkmaz, E. Ekici, F. Ozguner, U. Ozguner, Urban multi-hop broadcast protocols for inter-vehicle communication systems, in: Proceedings of First ACM Workshop on Vehicular Ad Hoc Networks (VANET 2004), October 2004, pp. 76–85.

[4] R. Nagpal, H. Shrobe, J. Bachrach, Organizing a global coordinate system from local information on an ad hoc sensor network, in: Proceedings of the Second International Workshop on Information Processing in Sensor Networks (IPSN '03), April 2003.

[5] D. Niculescu, B. Nath, Ad hoc positioning system (aps), in: IEEE GlobeCom, November 2001.

[6] D. Niculescu, B. Nath, Dv based positioning in ad hoc networks, Telecommunication Systems 22 (1–4) (2003) 267–280.

[7] C. Savarese, J. Rabay, K. Langendoen, Robust positioning algorithms for distributed ad-hoc wireless sensor networks, in: Proceedings of the General Track: 2002 USENIX Annual Technical Conference, 2002, pp. 317–327.

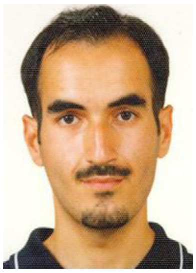
[8] S. Vural, E. Ekici, Wave addressing for dense sensor networks, in: Proceedings of Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004), August 2004, pp. 56–66.

[9] R. Vilzmann, C. Bettstetter, D. Medina, C. Hartmann, Hop distances and flooding in wireless multihop networks with randomized beamforming, in: Proceedings of Eighth ACM International Symposium on Modeling, Analysis and Simulation of Wireless and Mobile Systems, 2005, pp. 20–27.

[10] C. Bettstetter, J. Eberspächer, Hop distances in homogeneous ad hoc networks, in: Proceedings of 57th IEEE Semiannual Vehicular Technology Conference, April 2003, pp. 2286–2290.

[11] S.A.G. Chandler, Calculation of number of relay hops required in randomly located radio network, Electronics Letters 25 (1989) 770–777.

[12] L.E. Miller, Distribution of link distances in a wireless network, Journal of Research of the National Institute of Standards and Technology 106 (2001) 401–412.



Serdar Vural received his BS degree at The Electrical Engineering Department, at the Bogazici University, Istanbul, Turkey, and his MS degree at the Electrical and Computer Engineering Department, at The Ohio State University. He is currently a PhD student at the same department. His research interest is in sensor networks.

Eylem Ekici has received his BS and MS degrees in Computer Engineering from Bogazici University, Istanbul, Turkey, in 1997 and 1998, respectively. He received his Ph.D. degree in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta, GA, in 2002. Currently, he is an assistant professor in the Department of Electrical and Computer Engineering of the Ohio State University, Columbus, OH. His current research interests include wireless sensor networks, next generation wireless systems, space-based networks, and vehicular communication systems, with a focus on routing and medium access control protocols, resource management, and analysis of network architectures and protocols. He also conducts research on interfacing of dissimilar networks.