

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Theoretical Issues in Modeling of Large-Scale Wireless Sensor Networks

Di Ma, Meng Joo Er, Senior Member, IEEE
*Intelligent Systems Center, Nanyang Technological University, Research Techno Plaza,
 Singapore*

1. Introduction

In large-scale WSNs (wireless sensor networks), packets of a source node are often transmitted via multi-hop relays to reach their sink nodes. The hop-count, h , of a sink node is related to a particular hop-count originator (namely the source node) and it is defined as the least number of multi-hop relays required to send one packet from the source to the sink in this chapter. A source node can use a simple controlled flooding to set up the hop-counts relative to itself for all other nodes [1]. Let d denote the Euclidean distance between a source and its sink (the source-to-sink distance is denoted as SS-distance in the sequel), various statistical models that characterize the relationships between h and d are some of the fundamental problems in modeling large-scale WSNs. These results can be applied to address many other WSN research issues such as range-free localization, communication protocol design and evaluation, throughput optimization, transmission power control and etc.

Given a WSN whose nodes are distributed randomly according to a two-dimensional homogeneous Poisson point process of density λ , we investigate two statistical relationships between hop-count h and SS-distance d in this chapter. First of all, we propose a method termed CSP (Convolution of Successive Progress) to compute the K -hop connection probability for a two-dimensional network. The K -hop connection probability is defined as the conditional probability that a sink has a hop-count $h = K$ with respect to a source given that the SS-distance is d . Mathematically, the K -hop connection probability is defined as the conditional probability¹ $P(h = K | d)$. The CSP method is also extended into three-dimensional networks.

Secondly, based on the results of K -hop connection probability, we also present a method to compute the PDF (probability density function) of the SS-distance d conditioned on hop-count $h = K$, namely the PDF of SS-distance d of all nodes with a hop-count $h = K$. Mathematically, this conditional PDF² is denoted as $f(d | h = K)$. Simulation studies show that the proposed methods are able to achieve significant error reduction in computing these hop-distance statistics (i.e. the conditional probability/PDF) compared with existing methods.

¹ Unless otherwise specified, the conditional probability is referred to $P(h = K | d)$ in this paper.

² Unless otherwise specified, the conditional PDF is referred to $f(d | h = K)$ in this paper.

The rest of the chapter is organized as follows. Section 2 reviews some related works. Section 3 described the system models used in the analysis. The CSP algorithms for both two-dimensional and three-dimensional networks are presented in Section 4. A method of computing the conditional PDF $f(d|h=K)$ is then discussed in Section 5. Section 6 presents simulation results and discussions. Finally, Section 7 concludes the chapter.

2. Related works

Evaluation of various statistical models that characterize the relationships between hop-count h and SS-distance d are some of the fundamental problems in modeling large-scale WSNs. In [2], the solution of the conditional PDF $f(d|h=K)$ for one-dimensional networks is presented. The authors of [3] and [4] derived the conditional probability $P(h=K|d)$ based on the results of [2] for one-dimensional networks. However, only models are provided to the solution of $P(h=K|d)$ in two-dimensional networks. In [5], the exact analytical solution of the conditional probability $P(h=K|d)$ is derived for $K=1$ and $K=2$ in two-dimensional networks. However, probabilities $P(h=K|d)$ for $K>2$ were studied by analytical bounds and extensive simulations. The author of [6] proposed an iterative formulation of $P(h=K|d)$. Though not stated explicitly in his paper, its derivation is based on an “independent assumption”. This independent assumption is then highlighted in [7]. The analytical solutions proposed in [6] and [7] are the analytical approximations of $P(h=K|d)$ and only converges to the true statistics when the node density λ in the network tends to infinity. Due to the complexity of two-dimensional multi-hop percolation process, the authors of [8] proposed a simulation-based attenuated Gaussian approximation to model the conditional PDF $f(d|H=K)$. All previously mentioned results are based on the deployment model that nodes are distributed according to a one-dimensional or two-dimensional Poisson process. The author of [9] provided an analytical approximation for the probability of 2-hop connection between two randomly selected nodes in network whose nodes are distributed according to a Gaussian distribution.

Another interesting relationship between hop-count and distance is known as progress per hop. Works related to this are reported in [10] - [13]. The authors of [10] derived a solution for the expected per hop progress in two-dimensional networks. The expected per hop progress problem was also studied in [11] and [12] with different applications in protocol evaluation and localization. The authors of [13] presented an analytic approach to capture statistical bounds on hop-count for a given SS-distance in a greedy routing approach.

The statistical relationships between hop-count and SS-distance can be applied to address many WSN research problems. For example, many range-free WSN localization algorithms employ the principle of hop-count-to-distance transformation to localize nodes. The DV-hop WSN localization algorithm of [14] employs a heuristics approach known as correction factor to estimate the distance between two nodes based on the hop-count between them. Similar methods are also reported in [15] and [16]. The network throughput problem and optimal transmission radii were also studied in [10] based on the analytical expression of expected per hop progress.

3. System model

In this section, we discuss radio channel model, node deployment model and network topology model used in this chapter.

A. Channel model

As in many WSN literatures, we adopt an unit disk channel model (in some literatures, it is also known as lossless model) in this chapter. A source can communicate directly with all nodes within a disk centered at itself with a communication range r , but cannot communicate with nodes beyond the disk. The link between a pair of nodes is also assumed to be symmetric, namely, if node A can receive packets from node B directly, then node B also can receive packets from node A directly. The disk model is often used in the theoretical analysis of WSNs. There are other realistic channel models available (such as log-normal shadowing model [17]). These realistic channel models take the random effect of noise, attenuation, shadowing and etc. into consideration. The communication between a pair of nodes is thus characterized by PRR (packet reception rate) or LP (link probability) rather than the absolute SS-distance. In this chapter, since our study is a general analysis of the statistical relationship between hop-count and SS-distance, we adopt the unit disk channel model so that the radio channel can be regarded as lossless, invariant and homogeneous.

B. Deployment model

We assume all nodes with the same communication range r are randomly deployed in a plane according to a two-dimensional homogeneous Poisson point process of density λ . Furthermore, we require that all nodes deployed can form a fully-connected network. We make use of the WSN connectivity results presented in [18] to determine suitable values of λ so that all nodes deployed can form a connected network almost surely (probability greater than 0.99).

C. Topology model

Based on the channel and deployment model, we can model the WSN as an undirect graph G . A graph $G(V,E)$ consists a set of vertices (nodes) and edges (links), the set of vertices and edges are denoted by V and E respectively. The shortest network path between vertices i ($i \in V$) and vertices j ($j \in V$) is defined as the path that travels through the minimum number of edges from i to j . Therefore, the number of edges traveled is actually the hop-count between i and j . There are many algorithms proposed in the literature to set up the shortest network path in a WSN. These algorithms are out of the scope of this work. In this chapter, we assume that the network paths and the hop-counts between all node pairs have already been set up in advance.

4. Convolution of Successive Progress

In this section, we propose a method termed CSP (Convolution of Successive Progress) to compute the K -hop connection probability in two-dimensional WSNs, i.e. $P(h = K | d)$. We then extend the CSP method to compute the K -hop connection probability in three-dimensional networks.

A. Problem statement

A formal definition of the K -hop connection probability is given in this subsection. Let h and d denote the hop-count and distance between a pair of source and sink. The K -hop connection probability is an important hop-distance statistic and it is defined as the probability that the source can reach the sink in K number of multi-hop relays given that the

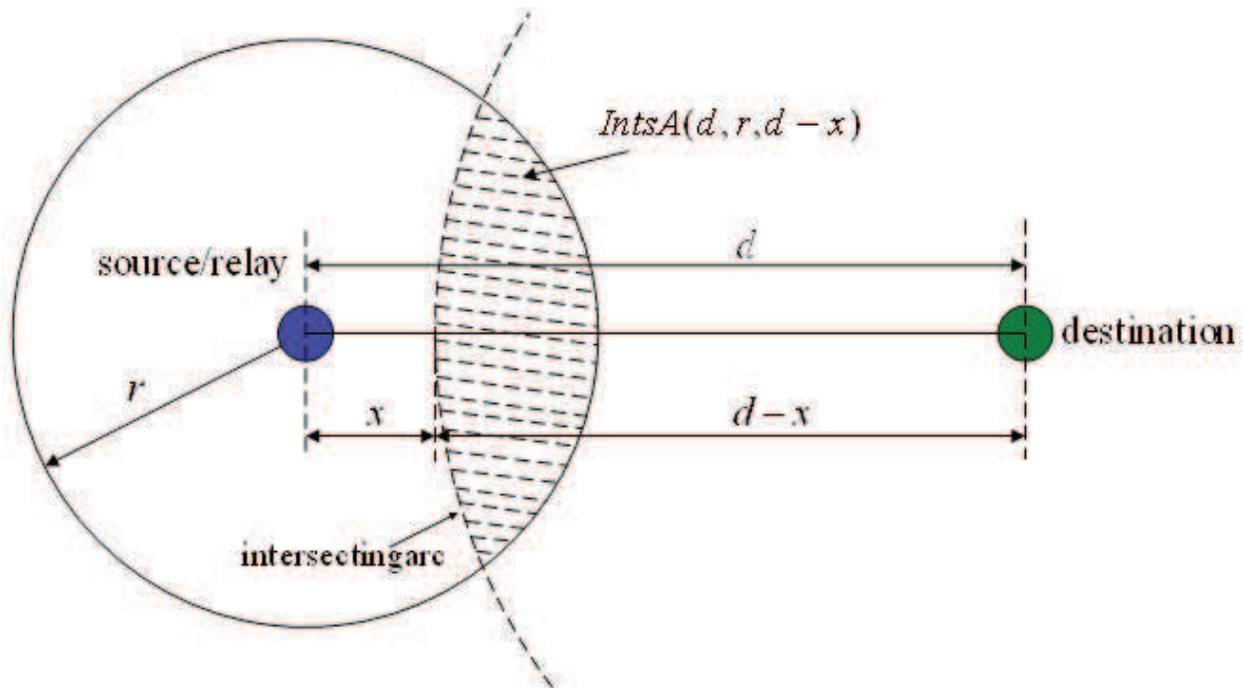


Fig. 1. Maximum progress from the source to the sink. Note that the source node can be a relay node as well. The dotted arc is the intersecting arc.

distance between the source and the sink is d , i.e. $P(h = K | d)$. In order to solve for $P(h = K | d)$, we construct a multi-hop network path from the source to the sink and computes the distribution of the length of the network path. First of all, we discuss how to compute the PDF of the length for a single hop.

B. Probability density function of maximum progress towards sink

We first introduce a random variable X (small letter x denotes an instance of X) known as maximum progress [10] per communication hop. It is a measure of the maximum progress in Euclidean distance from the source to the sink in one hop relay. Let $A(d_{12}, r_1, r_2)$ represent the intersectional area of two disks given the radii of the disks are r_1 and r_2 respectively and the distance between the centers of the two disks is d_{12} . The analytical form of $A(d_{12}, r_1, r_2)$ is given as follows:

$$A(d_{12}, r_1, r_2) = r_1^2 \cos^{-1} \left(\frac{d_{12}^2 + r_1^2 - r_2^2}{2d_{12}r_1} \right) + r_2^2 \cos^{-1} \left(\frac{d_{12}^2 + r_2^2 - r_1^2}{2d_{12}r_2} \right) - \frac{1}{2} \sqrt{(-d_{12} + r_1 + r_2)(d_{12} + r_1 - r_2)(d_{12} - r_1 + r_2)(d_{12} + r_1 + r_2)} \quad (1)$$

As depicted in Fig. 1, the distance between the source and the sink is d . A circle with a radius $d - x$ centered at the sink intersects with the coverage disk of the source and creates an intersectional region (shaded area). Therefore, nodes falling on the intersecting arc have the same distance $d - x$ to the sink. Furthermore, a node falling on the intersecting arc is selected as the relay node if there are no nodes in the shaded intersection region (the source has no neighboring nodes which has a shorter distance to the sink than the relay node falling on the intersecting arc). The CDF (Cumulative Distribution Function) of X , $F(x)$, can

be computed as the probability that there are no nodes in the shaded intersectional region. The CDF and PDF of X are computed as follows:

$$F(x) = P(X \leq x) = \exp(-\lambda \cdot A(d, r, d-x)) \quad (2)$$

$$f(x) = \frac{dF(x)}{dx} = -\lambda \exp(-\lambda \cdot A(d, r, d-x)) \frac{d}{dx} A(d, r, d-x) = \psi(\lambda, d, r, x) \quad (3)$$

$$\text{where, } \psi(\lambda, d, r, x) = -\lambda \exp(-\lambda \cdot A(d, r, d-x))(C1 + C2 + C3 + C4) \quad (4)$$

$$C1 = \frac{r(2d-2x)}{d \left(4 - \frac{(d^2 + r^2 - (d-x)^2)^2}{d^2 r^2} \right)^{\frac{1}{2}}}$$

$$C2 = -2(d-x) \cos^{-1} \left(\frac{d^2 + (d-x)^2 - r^2}{2d(d-x)} \right)$$

$$C3 = \frac{(d-x)^2 \left(\frac{2}{d} - \frac{d^2 + (d-x)^2 - r^2}{d(d-x)^2} \right)}{\left(4 - \frac{(d^2 + (d-x)^2 - r^2)^2}{d^2 (d-x)^2} \right)^{\frac{1}{2}}}$$

$$C4 = \frac{(-d+x)(-r^2 - 2dx + x^2)}{\sqrt{(-x+r)(2d-x-r)(x+r)(2d-x+r)}}$$

It can be seen from Eq. (3) that the PDF of X , $f(x)$, is parameterized on the node density λ , the source-to-sink distance d and the communication range r . Therefore, we use the function $\psi(\lambda, d, r, x)$ given by Eq. (4) to simplify the representation of $f(x)$. Let $r = 10$, the PDF of X for different SS-distance d given that $\lambda = 0.03537$ and $\lambda = 0.08842$ are plotted in Fig. 2 and Fig. 3 respectively. Given a fixed communication range, it can be seen that the shape of the function ψ is largely controlled by its first parameter node density λ and weakly depends on its second parameter SS-distance d . In the later discussion, we shall make use of this property to simplify some computations.

The PDF of maximum progress from the source to the sink is given by $\psi(\lambda, d, r, x)$. The function ψ can also be used for the PDF of maximum progress from a relay node to the sink. Suppose the distance between a relay node and the sink is d_{rs} , the PDF of maximum progress from the relay node to the sink is then given by $\psi(\lambda, d_{rs}, r, x)$. In the later discussion, we use X_i ($i = 0, 1, \dots$) to represent the random variable of maximum progress from the hop i relay node ($i = 0$ for the source) to the sink.

C. $K = 1$ and $K = 2$

For $K = 1$, $P(h = 1 | d) = 1$ if $d \leq r$ and $P(h = 1 | d) = 0$ otherwise. For $K = 2$, $P(h = 2 | d) = P(d - X_0 < r)$ if $r < d \leq 2r$ and $P(H = 2 | d) = 0$ otherwise. Since X_0 is the maximum progress from the source to the sink, $d - X_0$ is the remaining distance to be covered/progressed. Therefore,

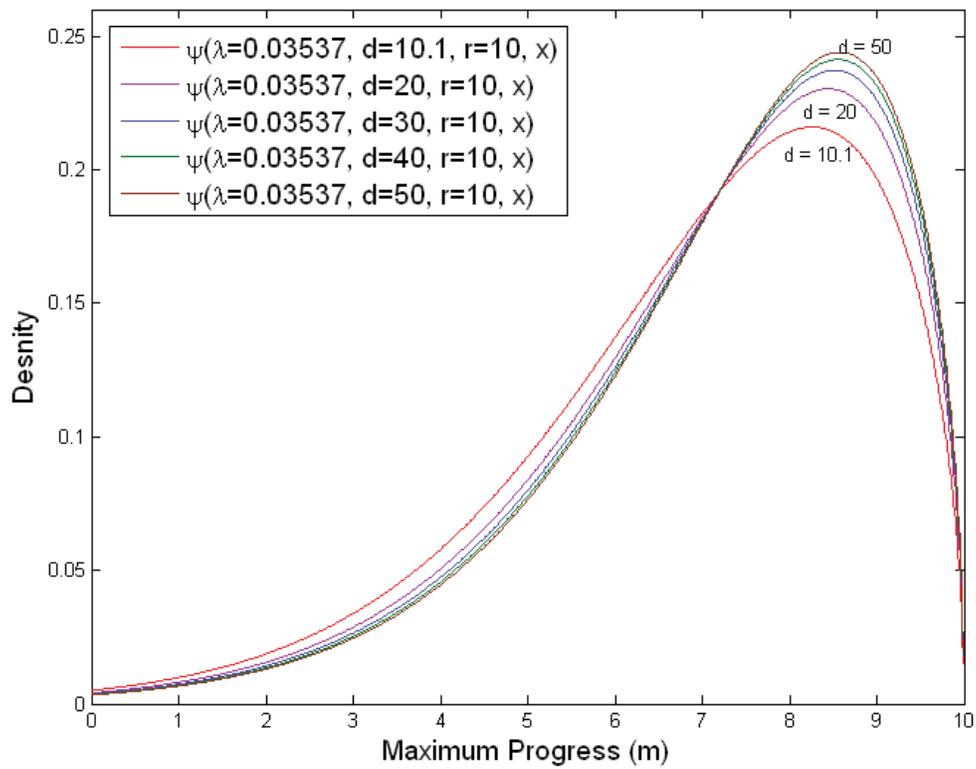


Fig. 2. Probability density function of maximum progress given different SS-distance, $\lambda = 0.03537$ and $r = 10$.

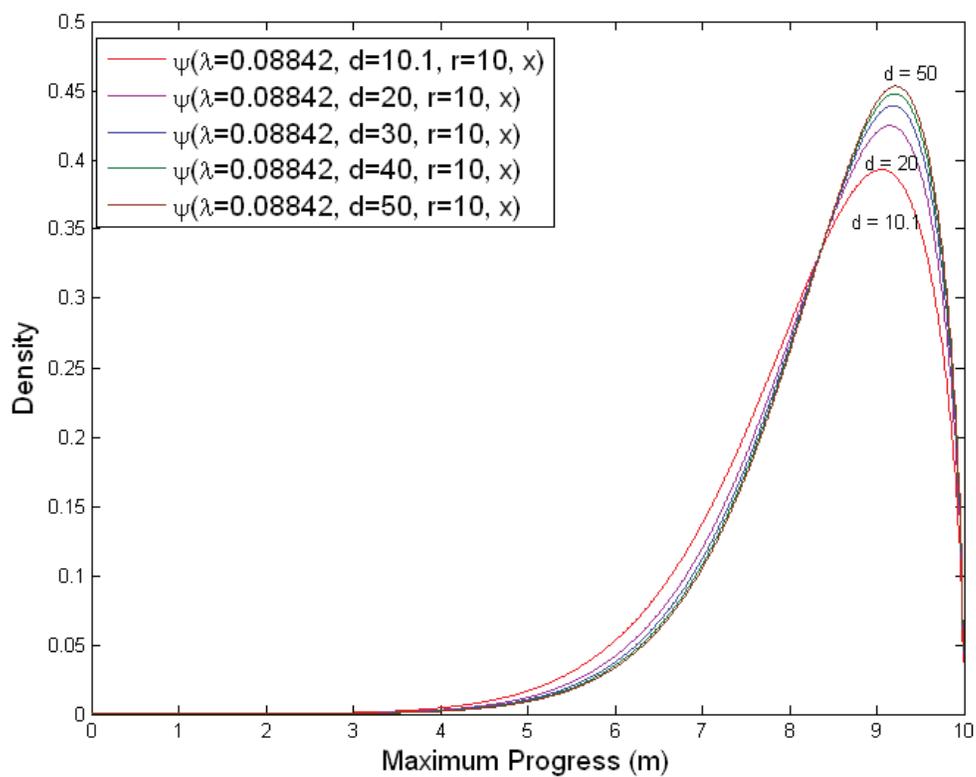


Fig. 3. Probability density function of maximum progress given different SS-distance, $\lambda = 0.08842$ and $r = 10$.

$P(d - X_0 < r)$ is the probability that a hop-1 relay node at the maximum progress from the source to the sink is within the coverage of the sink node. We can show that this approach of formulating $P(h = 2 | d)$ yields exactly the same form as the K -hop connection probability given in [7]. In particular, for $r < d \leq 2r$, $P(H = 2 | d)$ is given as follow:

$$\begin{aligned} P(h = 2 | d) &= P(d - X_0 < r) = 1 - P(X_0 \leq d - r) \\ &= 1 - \exp(-\lambda \cdot A(d, r, d - (d - r))) = 1 - \exp(-\lambda \cdot A(d, r, r)) \end{aligned} \quad (5)$$

D. $K > 2$

For $K > 2$, the conditional probability $P(h = K | d)$ is computed recursively based on the method of convolution of successive progress. We shall discuss how to compute $P(h = 3 | d)$ first.

Starting from the source, after one hop of maximum progress X_0 from the source to the sink, the remaining distance to be progressed is $d - X_0$ and a hop-1 node at the maximum progress becomes the relaying node. From this hop-1 relay node, the maximum progress towards the sink is X_1 . Therefore, $P(h = 3 | d)$ can be formulated as the product of the probability that a hop-2 relay node at the maximum progress is within the coverage disk of the sink node, and the probability that the sink node's hop-count is neither $h = 1$ nor $h = 2$. Mathematically, it is given as follows:

$$P(h = 3 | d) = P(d - X_0 - X_1 < r) \left(1 - \sum_{i=1}^2 P(h = i | d)\right) \quad (6)$$

Similar to the meaning of $P(d - X_0 < r)$, $P(d - X_0 - X_1 < r)$ is the probability that the hop-2 relay node at the maximum progress is within the coverage disk of the sink node. The PDF of X_0 and X_1 are given by $\psi(\lambda, d, r, x_0)$ and $\psi(\lambda, d - X_0, r, x_1)$ respectively, noting that $d - X_0$ is the RS-distance (distance between the hop-1 relay node and the sink). It can be seen that the PDF of X_1 depends on the random variable X_0 . Recall the property that the shape of function ψ is largely controlled by its first parameter λ rather than the second parameter SS-distance/RS-distance, therefore, we assume that X_1 is independent of X_0 . Define a new random variable Z_1 as the sum of X_0 and X_1 , the PDF of Z_1 can be calculated as follows:

$$f(z_1) = f(x_0) \otimes f(x_1) = \int_0^r f(x_0) f(z_1 - x_0) dx_0 = \int_0^r \psi(\lambda, d, r, x_0) \psi(\lambda, d - X_0, r, z_1 - x_0) dx_0 \quad (7)$$

where \otimes is the convolution operator. Now, $P(d - X_0 - X_1 < r) = P(d - Z_1 < r) = 1 - P(Z_1 \leq d - r)$. The conditional probability $P(h = 3 | d)$ is then computed as follows:

$$\begin{aligned} P(h = 3 | d) &= (1 - P(Z_1 \leq d - r)) \left(1 - \sum_{i=1}^2 P(h = i | d)\right) \\ &= \left(1 - \int_0^{d-r} f(z_1) dz_1\right) \left(1 - \sum_{i=1}^2 P(h = i | d)\right) \\ &= \left(1 - \int_0^{d-r} \int_0^r \psi(\lambda, d, r, x_0) \psi(\lambda, d - X_0, r, z_1 - x_0) dx_0 dz_1\right) \left(1 - \sum_{i=1}^2 P(h = i | d)\right) \end{aligned} \quad (8)$$

Similarly, for $P(h = 4 | d)$, it can be computed as $P(d - X_0 - X_1 - X_2 < r)(1 - \sum_{i=1}^3 P(h = i | d))$. In general, for $K > 2$, the conditional probability is given as follow:

$$\begin{aligned}
 P(h = K | d) &= P(d - X_0 - X_1 \dots - X_{K-2} < r)(1 - \sum_{i=1}^{K-1} P(h = i | d)) \\
 &= P(d - Z_{K-2} < r)(1 - \sum_{i=1}^{K-1} P(h = i | d)) \\
 &= (1 - \int_0^{d-r} f(z_{K-2}) dz_{K-2})(1 - \sum_{i=1}^{K-1} P(h = i | d))
 \end{aligned} \tag{9}$$

where, $Z_{K-2} = X_0 + X_1 + X_2 + \dots + X_{K-2}$

PDF of $Z_{K-2} : f(z_{K-2}) = f(x_0) \otimes f(x_1) \otimes f(x_2) \dots \otimes f(x_{K-2})$

PDF of $X_0 : f(x_0) = \psi(\lambda, d, r, x_0)$

PDF of $X_1 : f(x_1) = \psi(\lambda, d - X_0, r, x_1)$

PDF of $X_2 : f(x_2) = \psi(\lambda, d - X_0 - X_1, r, x_2)$

...

PDF of $X_{K-2} : f(x_{K-2}) = \psi(\lambda, d - X_0 - X_1 \dots - X_{K-3}, r, x_{K-2})$

It can be seen that the random variable Z_{K-2} is the progress of the multi-hop network path from the source to the hop $K - 1$ relay node and its PDF can be computed by convolving the PDFs of successive maximum progresses. Therefore, the name CSP which reflects the algorithm's computation methodology is phrased.

E. Extension to three-dimensional networks

The CSP algorithm can be extended to three-dimensional network based on the same intuition of its two-dimensional counterpart. In this case, we assume that nodes with communication range r_{3D} are deployed in a space according to a three-dimensional homogeneous Poisson point process of density λ_{3D} .

1) *Probability Density Function of Maximum Progress*: We have derived the analytical form of the PDF of maximum progress in two-dimensional networks. In three dimensional networks, the radio coverage of a node is a sphere rather than a disk. Therefore, the intersectional area as shown in Fig. 1 becomes an intersection volume between two spheres. Similarly, let X (small letter x denotes an instance of X) represent the maximum progress from a source to a sink in three-dimensional networks. Let $V(d_{12}, r_1, r_2)$ represent the intersectional volume of two spheres given the radii of the spheres are r_1 and r_2 respectively and the distance between the centers of the two spheres is d_{12} . The analytical form of $V(d_{12}, r_1, r_2)$ is given as follows:

$$V(d_{12}, r_1, r_2) = \frac{\pi(r_1 + r_2 - d)^2(d^2 + 2d(r_1 + r_2))}{12d} \tag{10}$$

Visualize the Fig. 1 in three-dimensional space, the distance between the source and the sink is still d . A sphere with a radius $d - x$ centered at the sink intersects with the radio coverage sphere of the source and creates an intersectional volume (visualize the shaded part as a

volume). Therefore, nodes falling on the intersecting surface (visualize the dotted intersecting arc as a surface) have the same distance $d-x$ to the sink. Furthermore, a node falling on the intersecting surface is selected as the relay node if there are no nodes in the shaded intersectional volume. The CDF of X , $F(x)$, can be computed as the probability that there are no nodes in the shaded intersectional volume. The CDF and PDF of X are computed as follows:

$$F(x) = P(X \leq x) = \exp(-\lambda_{3D} \cdot V(d, r_{3D}, d-x)) \quad (11)$$

$$f(x) = \frac{dF(x)}{dx} = -\lambda_{3D} \exp(-\lambda_{3D} \cdot V(d, r_{3D}, d-x)) \frac{d}{dx} V(d, r_{3D}, d-x) = \psi_{3D}(\lambda_{3D}, d, r_{3D}, x) \quad (12)$$

$$\psi_{3D}(\lambda_{3D}, d, r_{3D}, x) = -\lambda_{3D} \exp(-\lambda_{3D} \cdot V(d, r_{3D}, d-x))(D1 + D2) \quad (13)$$

$$D1 = -\frac{\pi(r_{3D} - x)(d^2 + 2dr_{3D} - 3r_{3D}^2 + 2d(d-x) + 6r_{3D}(d-x) - 3(d-x)^2)}{6d}$$

$$D2 = \frac{\pi(r_{3D} - x)^2(4d - 6r_{3D} - 6x)}{12d}$$

Similarly, we use the function $\psi_{3D}(\lambda_{3D}, d, r_{3D}, x)$ given by Eq. (13) to simplify the representation of the PDF of maximum progress X in three dimensional space. In the following discussion, we use X_i ($i = 0, 1, \dots$) to represent the random variable of maximum progress from the hop i relay node ($i = 0$ for the source) to the sink in three-dimensional networks.

2) $K = 1, 2$: Similarly, $P(h = 1 | d) = 1$ if $0 < d \leq r_{3D}$ and $P(h = 1 | d) = 0$ otherwise. For $P(h = 2 | d)$, it is given as follows:

$$P(h = 2 | d) = \begin{cases} 1 - \exp(-\lambda_{3D} \cdot V(d, r_{3D}, r_{3D})), & r_{3D} < d \leq 2r_{3D}; \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

$$(15)$$

where $V(d, r_{3D}, r_{3D}) = \frac{\pi(2r_{3D}-d)^2(d^2+4dr_{3D})}{12d}$.

3) $K > 2$: For $K > 2$, the CSP method for computing $P(h = K | d)$ in three-dimensional network is exactly the same as Eq. (9) except that the PDF of maximum progress is changed from function to function ψ_{3D} given in Eq. (13).

V. Probability density function of SS-distance conditioned on hop-count

We have proposed a method of computing the K -hop connection probability $P(h = K | d)$. In this section, we derive the solution of the PDF of SS-distance d given that hop-count $h = K$ based on the result of K -hop connection probability. Mathematically, this conditional PDF is denoted as $f(d | h = K)$.

We have assumed that nodes are deployed according to a Poisson process in the previous sections, therefore, the sensor node deployment region and the SS-distances of all nodes are unbounded. To simplify the analysis in this case, we define a disk shape deployment region of radius R ($R \gg r$). Without loss of generality, we assume that the source node is located at the center of the deployment disk. All other nodes falling inside the deployment disk can be

approximated as uniformly distributed in the deployment disk³. Since all nodes are distributed in a disk of radius R , the SS-distance d of all nodes satisfy $0 < d \leq R$.

A. Standard approach: computing $f(d|h=K)$ using Bayes Theorem

Since we have computed the conditional probability $P(h=K|d)$, we can derive the PDF $f(d|h=K)$ using Bayes Theorem as follows:

$$f(d|h=K) = \frac{P(h=K|d)f(d)}{P(h=K)} = \frac{P(h=K|d)f(d)}{\int_0^R P(h=K|d)f(d)dd} \quad (16)$$

Given that all nodes are distributed uniformly in a disk of radius R and the source node is located at the center of the disk, the PDF of the SS-distance d of all nodes is given as $f(d) = \frac{2d}{R^2}$. The PDF $f(d|h=K)$ can be computed by substituting $f(d)$ and $P(h=K|d)$ into Eq. (16). However, computation of $f(d|h=K)$ using Bayes Theorem requires that the value of $P(h=K|d)$ is known priorly. Furthermore, the conditional probability $P(h=K|d)$ is recursively formulated and it depends on the results of $P(h=K'|d)$ ($K' < K$). Therefore, computing $f(d|h=K)$ using Bayes Theorem is heavy from a computational point of view. We propose an alternative approximation method to compute $f(d|h=K)$ without evaluating $P(h=K|d)$ in the first place.

B. Alternative approach: random truncation

Since the SS-distances of all hop-1 nodes are less than r , $f(d|h=1)$ can be derived as a truncated distribution of $f(d)$ as follows:

$$f(d|h=1) = f(d|d \leq r) = \frac{1}{F(r)} \cdot \frac{2d}{R^2} = \frac{1}{\int_0^r \frac{2d}{R^2} dd} \cdot \frac{2d}{R^2} = \frac{2d}{r^2} \quad (17)$$

where $F(r) = P(d \leq r) = \int_0^r \frac{2d}{R^2} dd$. Similarly, the PDF $f(d|h=2)$ can also be derived as a truncated distribution of $f(d)$ as $f(d|h=2) = f(d|r < d \leq X_0 + r)$. Since $P(h=2|d) = P(d < X_0 + r)$ if $r < d \leq 2r$. The upper bound on d , $X_0 + r$, is simply the maximum SS-distance of all hop-2 nodes.

Unlike the case of $f(d|h=1)$, the upper truncation $X_0 + r$ itself is random in this case. Define a new random variable Y as $Y = X_0 + r$ and $r \leq Y \leq 2r$. Since the PDF of X_0 is given as $\psi(\lambda, d, r, x_0)$, the PDF of Y is the shifted version of the PDF of X_0 and is given by $f(y) = \psi(\lambda, d, r, y - r)$. Therefore, the PDF $f(d|h=2)$ is computed as follows:

$$\begin{aligned} f(d|h=2) &= f(d|r < d \leq X_0 + r) = f(d|r < d \leq Y) = \frac{1}{F(X_0 + r) - F(r)} \int_r^{2r} f(d) \text{rect}(d, r, y) f(y) dy \\ &= \frac{1}{F(X_0 + r) - F(r)} \int_r^{2r} \frac{2d}{R^2} \cdot \text{rect}(d, r, y) \psi(\lambda, d, r, y - r) dy \end{aligned}$$

³ The number of nodes falling inside a region of fixed area can be approximated as a Poisson random variable given that all nodes are distributed uniformly in a bounded deployment region.

$$\begin{aligned}
&\approx \frac{1}{F(\mathbb{E}[X_0]+r)-F(r)} \int_r^{2r} \frac{2d}{R^2} \cdot \text{rect}(d,r,y) \psi(\lambda,d,r,y-r) dy \\
&= \frac{1}{\int_r^{\mathbb{E}[X_0]+r} \frac{2d}{R^2} dd} \int_r^{2r} \frac{2d}{R^2} \cdot \text{rect}(d,r,y) \psi(\lambda,d,r,y-r) dy
\end{aligned} \tag{18}$$

where $\text{rect}(d, r, y) = 1$ if $r < d \leq y$ and $\text{rect}(d, r, y) = 0$ otherwise. In Eq. (18), we approximate $F(X_0 + r)$ as $F(\mathbb{E}[X_0] + r)$ in order to simplify the computation.

We can use the same approach to derive the PDF $f(d|h=3)$. The lower truncation of $f(d|h=3)$ is simply the upper truncation of $f(d|h=2)$. The upper truncation of $f(d|h=3)$ is the maximum SS-distance of all hop-3 nodes and is given by X_0+X_1+r . Define a random variable Y_1 as $Y_1 = X_0+X_1+r$ and $r \leq Y_1 \leq 3r$. As we have computed the PDF of $Z_1 = X_0 + X_1$ in Eq. (7), the PDF of Y_1 is the shifted version of the PDF of Z_1 :

$$f(y_1) = \int_0^r \psi(\lambda, d, r, x_0) \psi(\lambda, d - X_0, r, y_1 - r - x_0) dx_0 \tag{19}$$

Therefore, we compute $f(d|h=3)$ as follows:

$$\begin{aligned}
f(d|h=3) &= f(d | X_0 + r < d \leq X_0 + X_1 + r) = f(d | Y < d \leq Y_1) \\
&= \frac{1}{F(X_0 + X_1 + r) - F(X_0 + r)} \int_r^{3r} \int_r^{2r} f(d) \text{rect}(d, y, y_1) f(y) f(y_1) dy dy_1 \\
&\approx \frac{1}{F(\mathbb{E}[X_0 + X_1] + r) - F(\mathbb{E}[X_0] + r)} \int_r^{3r} \int_r^{2r} \frac{2d}{R^2} \cdot \text{rect}(d, y, y_1) f(y) f(y_1) dy dy_1 \\
&= \frac{1}{\int_{\mathbb{E}[X_0]+r}^{\mathbb{E}[X_0+X_1]+r} \frac{2d}{R^2} dd} \int_r^{3r} \int_r^{2r} \frac{2d}{R^2} \cdot \text{rect}(d, y, y_1) f(y) f(y_1) dy dy_1
\end{aligned} \tag{20}$$

where $f(y)$ and $f(y_1)$ are the PDF of Y and Y_1 respectively. In general, the PDF $f(d|h=K)$ for $K > 3$ can be computed as follows:

$$\begin{aligned}
f(d|h=K) &= f(d | \sum_{i=0}^{K-3} X_i + r < d \leq \sum_{i=0}^{K-2} X_i + r) = f(d | Y_{K-3} < d \leq Y_{K-2}) \\
&= \frac{1}{F(\sum_{i=0}^{K-2} X_i + r) - F(\sum_{i=0}^{K-3} X_i + r)} \int_r^{K_r(K-1)r} \int_r^{K_r(K-2)r} f(d) \text{rect}(d, y_{K-3}, y_{K-2}) f(y_{K-3}) f(y_{K-2}) dy_{K-3} dy_{K-2} \\
&\approx \frac{1}{F(\mathbb{E}[\sum_{i=0}^{K-2} X_i] + r) - F(\mathbb{E}[\sum_{i=0}^{K-3} X_i] + r)} \int_r^{K_r(K-1)r} \int_r^{K_r(K-2)r} \frac{2d}{R^2} \cdot \text{rect}(d, y_{K-3}, y_{K-2}) f(y_{K-3}) f(y_{K-2}) dy_{K-3} dy_{K-2} \\
&= \frac{1}{\int_{\mathbb{E}[\sum_{i=0}^{K-3} X_i]+r}^{\mathbb{E}[\sum_{i=0}^{K-2} X_i]+r} \frac{2d}{R^2} dd} \int_r^{K_r(K-1)r} \int_r^{K_r(K-2)r} \frac{2d}{R^2} \cdot \text{rect}(d, y_{K-3}, y_{K-2}) f(y_{K-3}) f(y_{K-2}) dy_{K-3} dy_{K-2}
\end{aligned} \tag{21}$$

where, $Y_{K-3} = \sum_{i=0}^{K-3} X_i + r$, PDF of $Y_{K-3} : f(y_{K-3}) = f(\sum_{i=0}^{K-3} x_i + r)$

$Y_{K-2} = \sum_{i=0}^{K-2} X_i + r$, PDF of $Y_{K-2} : f(y_{K-2}) = f(\sum_{i=0}^{K-2} x_i + r)$

$f(\sum_{i=0}^{K-3} X_i) = f(x) \otimes f(x_1) \otimes f(x_2) \dots \otimes f(x_{K-3})$

$f(\sum_{i=0}^{K-2} X_i) = f(x) \otimes f(x_1) \otimes f(x_2) \dots \otimes f(x_{K-2})$

6. Simulation studies

In this section, we study the accuracy of the proposed method in computing $P(h = K | d)$ and $f(d | h = K)$ through simulation studies. Firstly, we compare the accuracy of the two different methods in computing the conditional probability $P(h = K | d)$, namely the method proposed in [7] (this method is termed as Independent Assumption Method (IAM) in the sequel) and the CSP algorithm proposed in this chapter. Secondly, we study the accuracy of the proposed method in computing the conditional PDF $f(d | h = K)$.

A. Simulation settings

For two-dimensional networks, the communication range r is set to 10. We vary the density λ of the network and carry out 1000 independent simulations. The simulated statistics of $P(h = K | d)$ and $f(d | h = K)$ are then derived from the simulation data. For the three-dimensional network, all settings are the same except the communication range $r_{3D} = 15$.

B. K-hop connection probability $P(h = K | d)$

1) *Mean Absolute Error*: We adopt the accuracy metric known as MAE (Mean Absolute Error) introduced in [7] as the performance measure. For comparison purpose, we partition the domain of d into M bins. MAE_K is then defined as follows:

$$MAE_K = \frac{1}{M} \sum_{i=1}^M |P_a(h = K | d_i) - P_s(h = K | d_i)| \quad (22)$$

where $P_a(h = K | d_i)$ and $P_s(h = K | d_i)$ are the analytical and the corresponding simulation results of the conditional probability respectively. The index K on MAE indicates the particular value of K -hop considered. Fig. 4 shows the simulation results and the analytical results (computed using IAM) of the conditional probability $P(h = K | d)$ ($K = 2$ to 5) for networks with density $\lambda = 0.03537$. We observe that the analytical result of $P(h = 2 | d)$ closely matches the simulation result of $P(h = 2 | d)$. This is expected since the analytical solution of $P(h = 2 | d)$ given in Eq. (5) is exact. However, the analytical results deviate from the simulation results as K increases. It can be observed that the discrepancy becomes significantly large for $P(h = 5 | d)$. On the other hand, Fig. 5 shows the simulation results and the analytical results (computed using the CSP method) of the conditional probability $P(h = K | d)$ ($K = 2$ to 5) for networks with density $\lambda = 0.03537$. It is observed that the analytical results closely match the simulation results for all values of K . In fact, even without the MAE metric, it is not difficult to conclude from the observation that computation of $P(h = K | d)$ based on the CSP method is more accurate than the method of IAM.

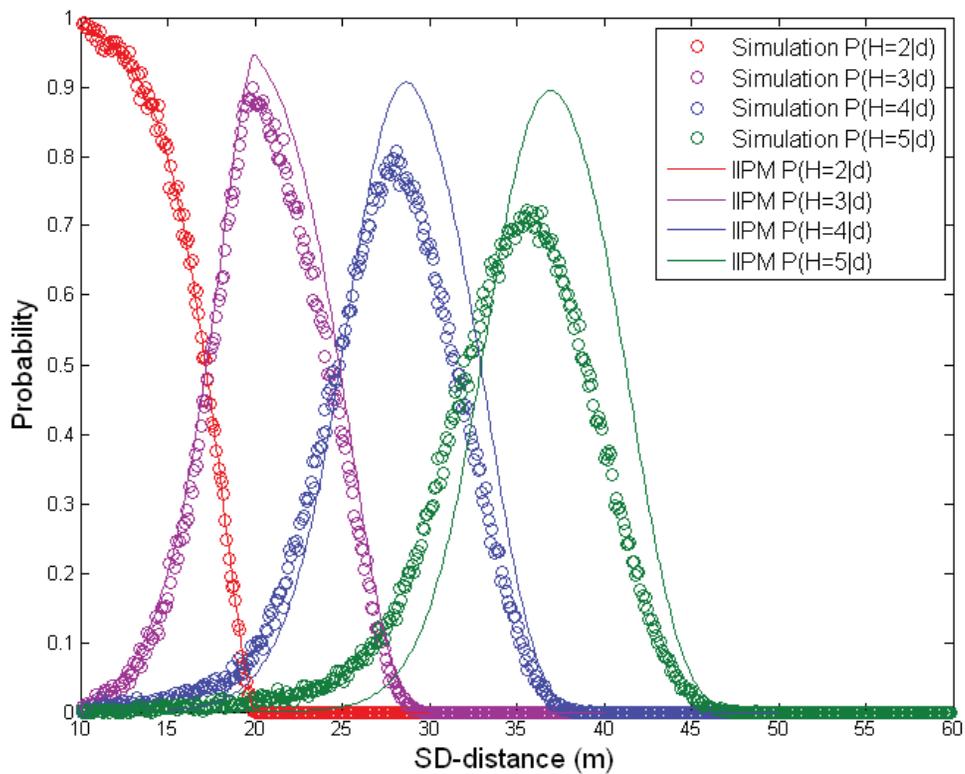


Fig. 4. Simulation results and analytical results (IAM method) of $P(h = K | d)$ for $K = 2$ to 5 . Network density $\lambda = 0.03537$.

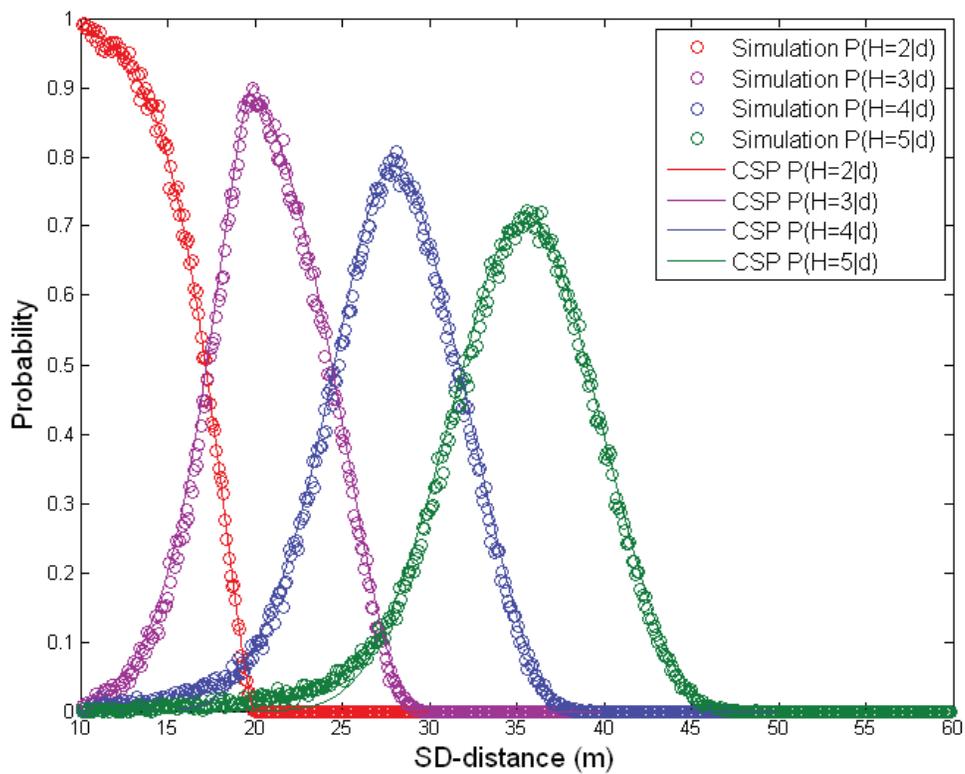


Fig. 5. Simulation results and analytical results (CSP method) of $P(h = K | d)$ for $K = 2$ to 5 . Network density $\lambda = 0.03537$.

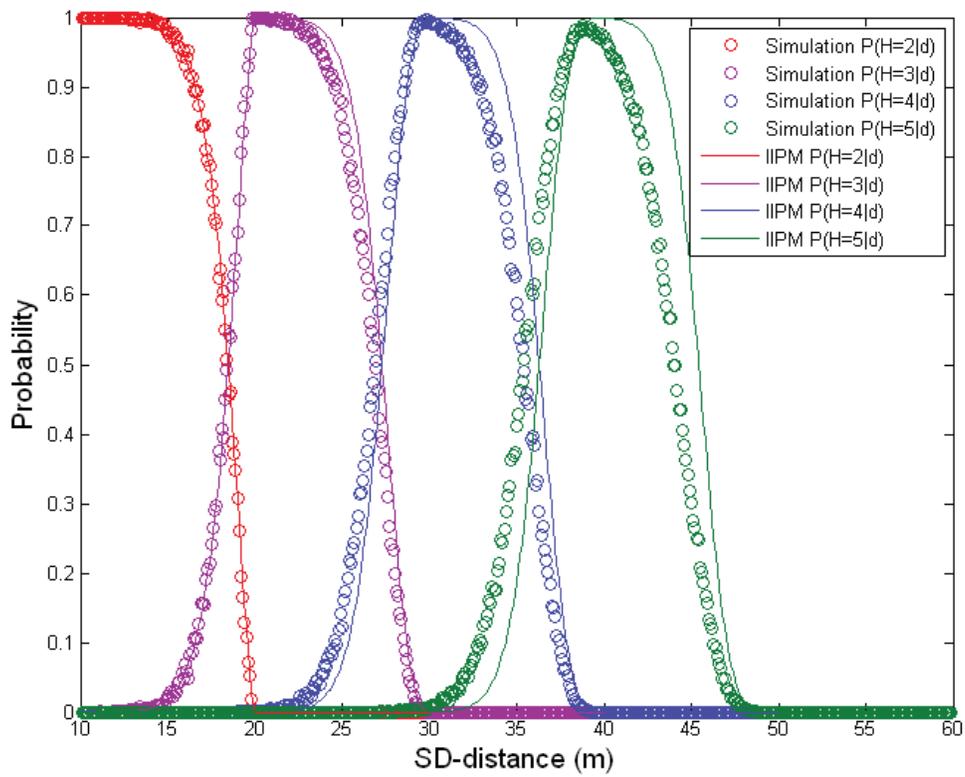


Fig. 6. Simulation results and analytical results (IAM method) of $P(h = K | d)$ for $K = 2$ to 5 . Network density $\lambda = 0.08842$.

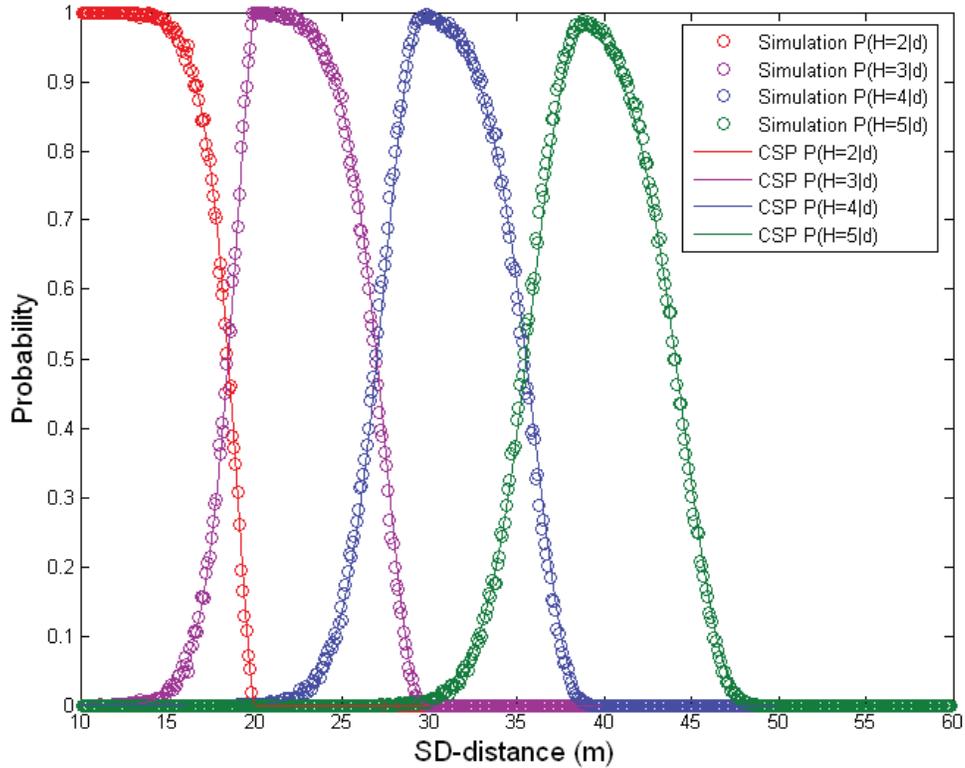


Fig. 7. Simulation results and analytical results (CSP method) of $P(h = K | d)$ for $K = 2$ to 5 . Network density $\lambda = 0.08842$.

We also plot the analytical results of $P(h = K | d)$ for networks with density $\lambda = 0.08842$ based on the method of IAM and the CSP in Fig. 6 and 7 respectively. Simulation results of $P(h = K | d)$ are also plotted in each figure. In Fig. 6, we observe that the discrepancy between the analytical results of the method of IAM and the simulation results reduces as compared to the previous case where density $\lambda = 0.03537$. The reason is that the independent assumption of the IAM method only holds when λ tends to infinity. As a consequence, computation of $P(h = K | d)$ based on the IAM method becomes more accurate as λ increases. Although the discrepancy is reduced as compared to Fig. 4, the error observed in Fig. 6 is still non-negligible. On the other hand, it can be observed in Fig. 7 that the analytical results of the CSP method concur with the simulation results.

To compare the performance of the IAM and the CSP methods quantitatively, we plot the MAE of both methods against the node density λ in Fig. 8, 9 and 10. Since the analytical solution of $P(h = 2 | d)$ is exact, error in computing $P(h = 2 | d)$ is due to computer simulation (such as finite steps taken in the numerical integration) rather than methodology. Therefore, we plot $MAE_{K=2}$ in all figures as a benchmark indicator. In Fig. 8, we plot $MAE_{K=3}$ of the IAM and the CSP methods. It is observed that the $MAE_{K=3}$ of the IAM method is much larger than the $MAE_{K=3}$ of the CSP method. Although the $MAE_{K=3}$ of the CSP method is larger than the benchmark indicator $MAE_{K=2}$, it is observed that the $MAE_{K=3}$ approaches the benchmark $MAE_{K=2}$ as the number of nodes increases. Fig. 9 and 10 shows the $MAE_{K=4}$ and $MAE_{K=5}$ of both methods. It is observed that the MAE_K of the CSP method are smaller than the MAE_K of the IAM method for all values of K . For instance, the $MAE_{K=5}$ of the CSP method is about 10 times smaller than the $MAE_{K=5}$ of the IAM method.

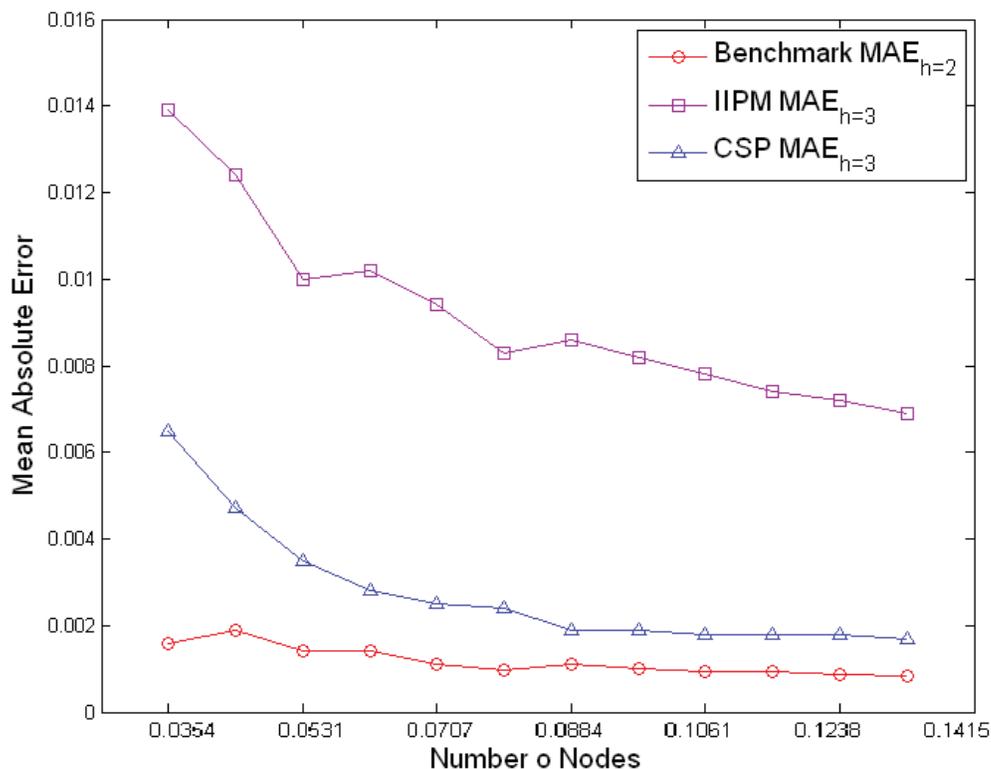


Fig. 8. The $MAE_{K=3}$ of the IAM method and the CSP method against the node density λ . The $MAE_{K=2}$ is plotted as a benchmark indicator.

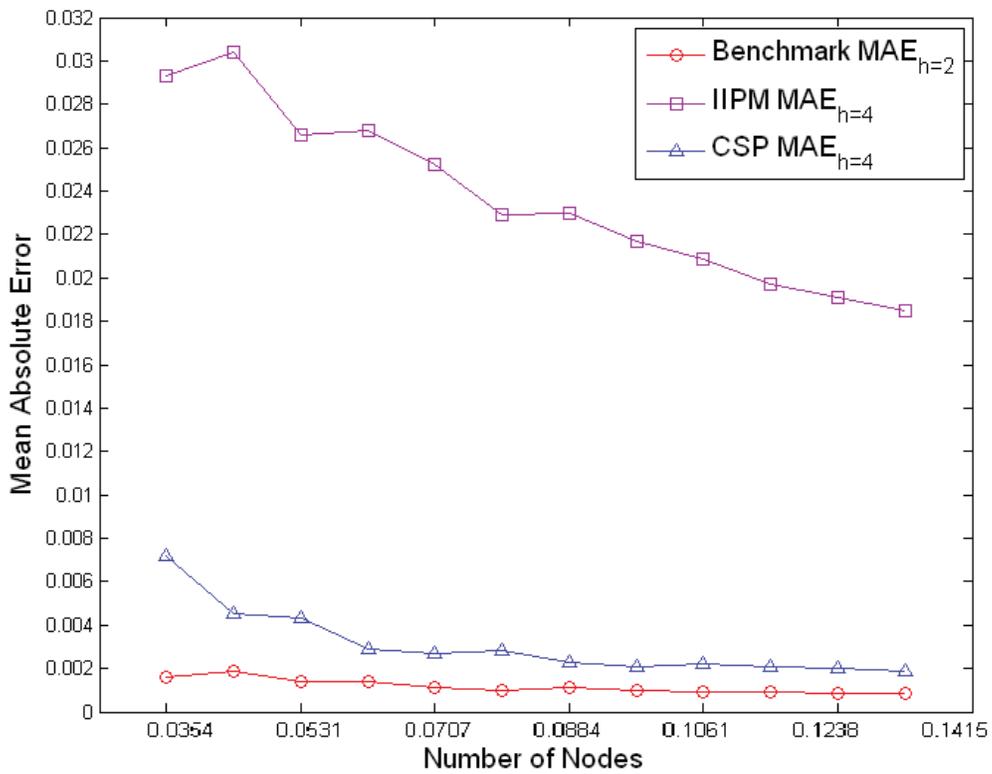


Fig. 9. The $MAE_{K=4}$ of the IAM method and the CSP method against the node density λ . The $MAE_{K=2}$ is plotted as a benchmark indicator.

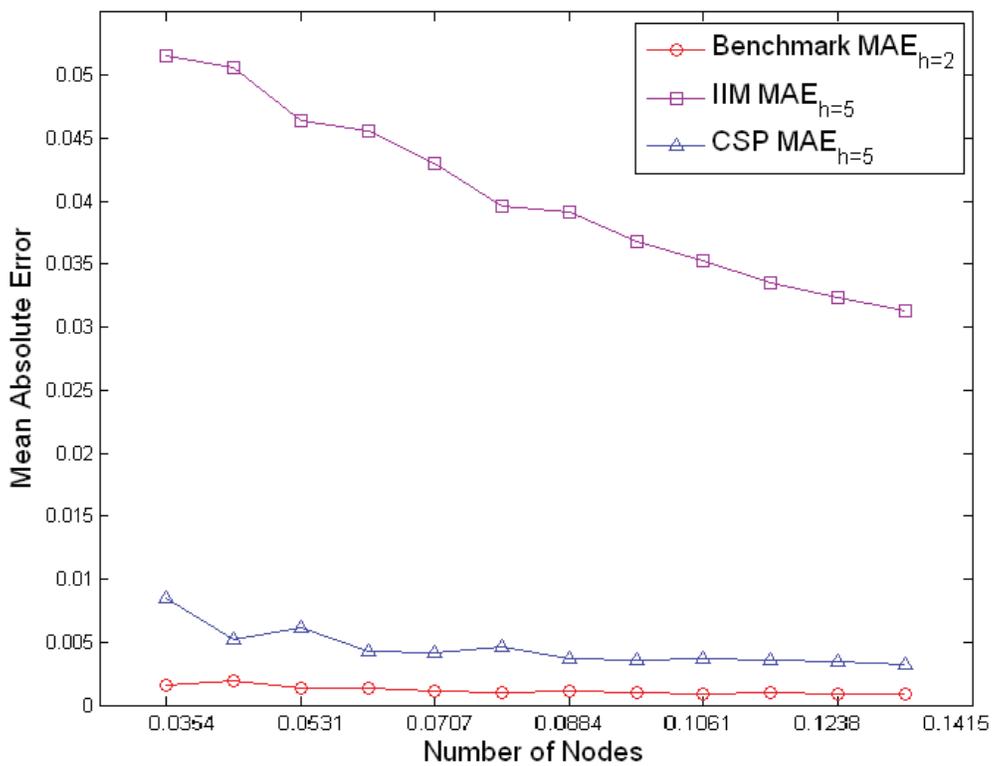


Fig. 10. The $MAE_{K=5}$ of the IAM method and the CSP method against the node density λ . The $MAE_{K=2}$ is plotted as a benchmark indicator.

2) *Snowball Effect*: Since both IAM and CSP methods are formulated recursively, errors in computing $P(h = K | d)$ will propagate to/accumulate in the computation of $P(h = K' | d)$ ($K' > K$). We compare this snowball effect of both methods. In Fig. 11, we plot the MAE of both methods against hop-count K for three network densities, namely $\lambda = 0.03537$, 0.08842 and 0.1326 . As shown in the figure, the snowball effect of the IAM method is prominent since the MAE of the IAM method increases rapidly as K increases. For the CSP method, although the formulation of the conditional probability is also recursive, the snowball effect is negligible.

3) *K-hop Outage Probability*: The analytical result of the K -hop connection probability $P(h = K | d)$ can be easily extended to the solution of a problem known as K -hop outage probability. The K -hop outage probability is defined as the probability that a sink can be connected to a source in less than or equal to K number of multi-hop relays, given that the SS-distance is d . The K -hop outage probability can be evaluated as the sum of individual K -hop connection probability, i.e. $\sum_{i=1}^K P(h = i | d)$. Fig. 12 shows the K -hop outage probability plots based on the simulation results and the analytical results for networks of density $\lambda = 0.05305$. It can be seen that the analytical result concur with the simulation results.

4) *Three-dimensional Network*: We have extended the CSP method into three-dimensional networks. To validate the method, we plot the analytical results and simulation results of $P(h = K | d)$ for three-dimensional network of density $\lambda = 0.001658$ and $\lambda = 0.002210$ in Fig. 13 and 14 respectively. It can be observed from the figures that the analytical results concur with the simulation results.

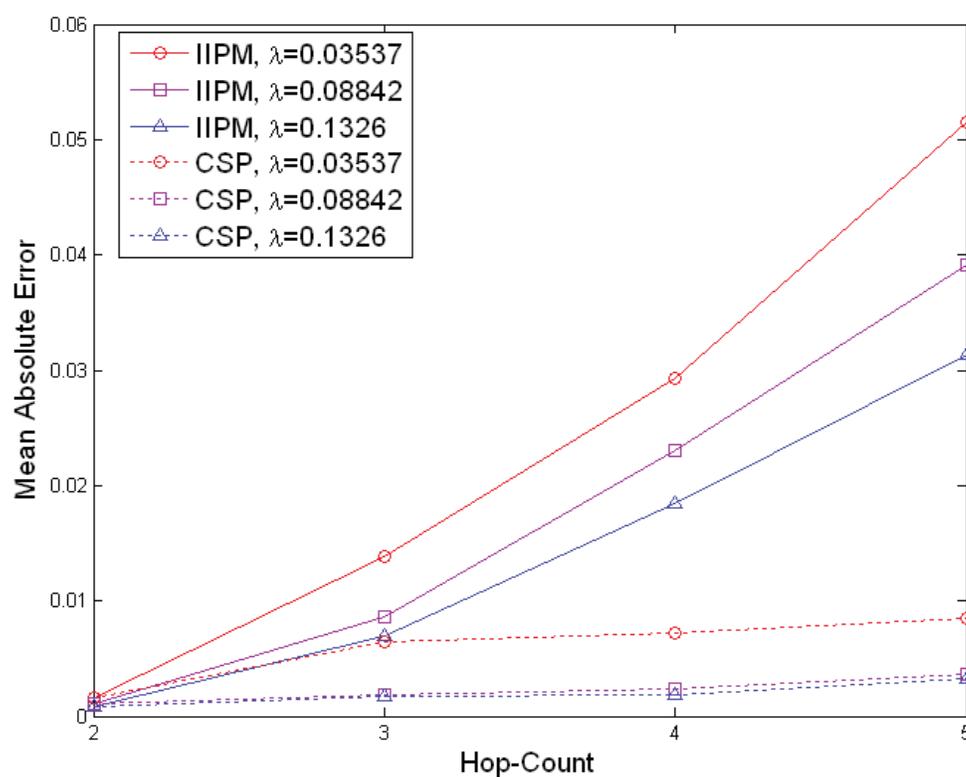


Fig. 11. Illustration of the snowball effect. The MAE of the IAM method increases rapidly as hop-count increases. On the other hand, the MAE of the CSP method does not increase much as hop-count K increases.

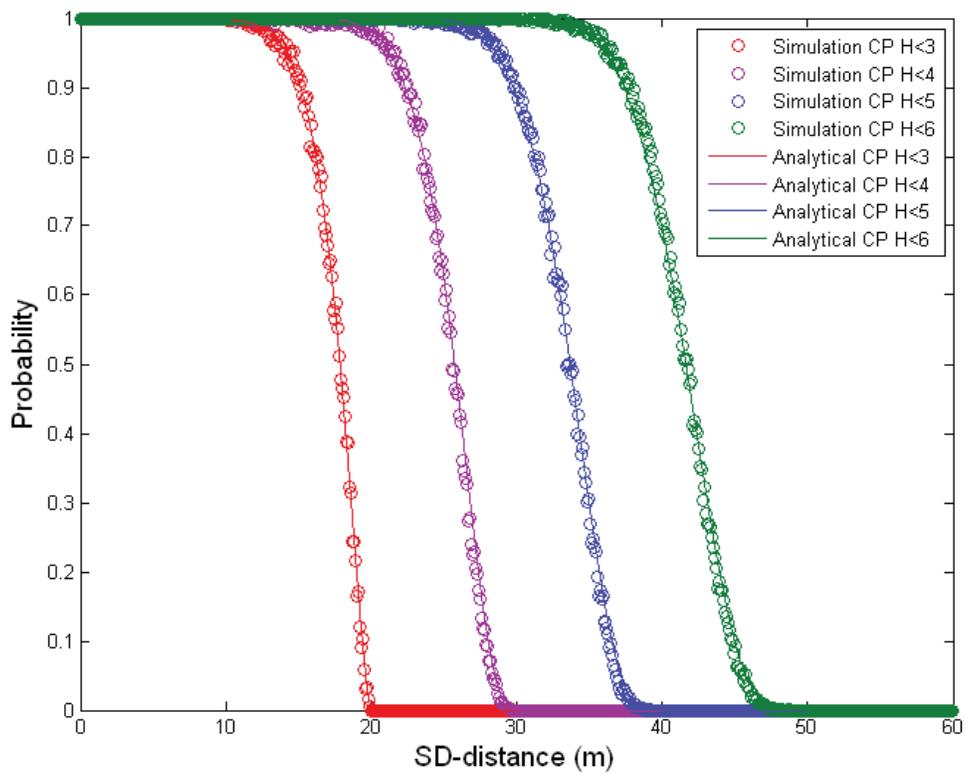


Fig. 12. Illustration of the K -hop outage Probability for $K = 2$ to 5 , $\lambda = 0.05305$

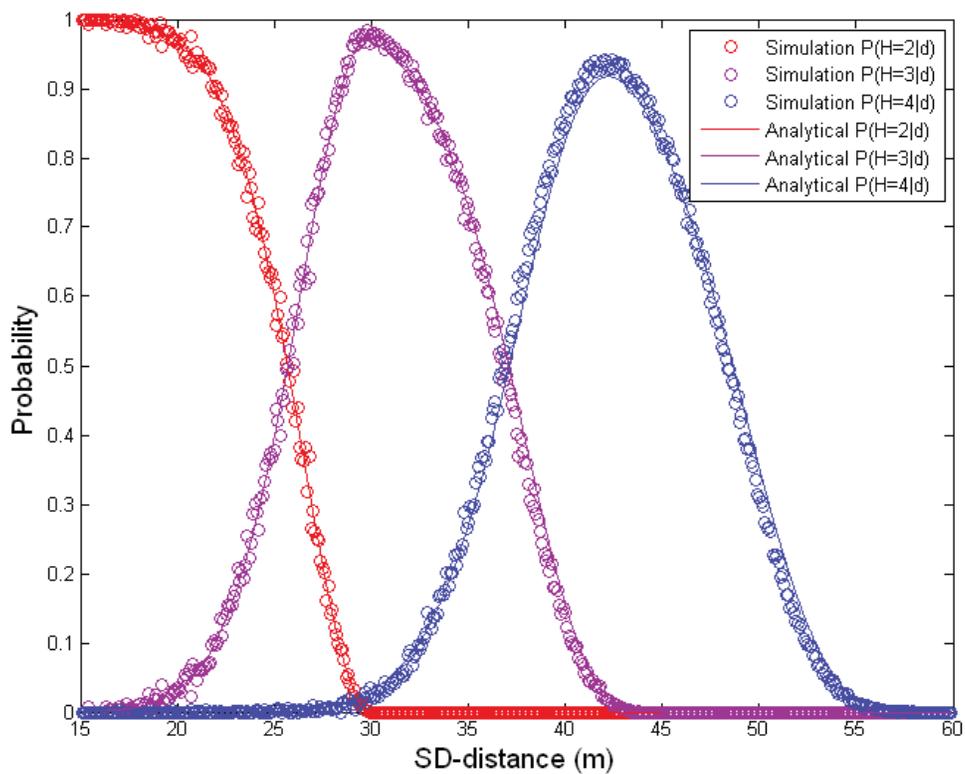


Fig. 13. Comparison of the conditional probability $P(h = K | d)$ in three-dimensional network based on the simulation results and the analytical results, $\lambda = 0.001658$.

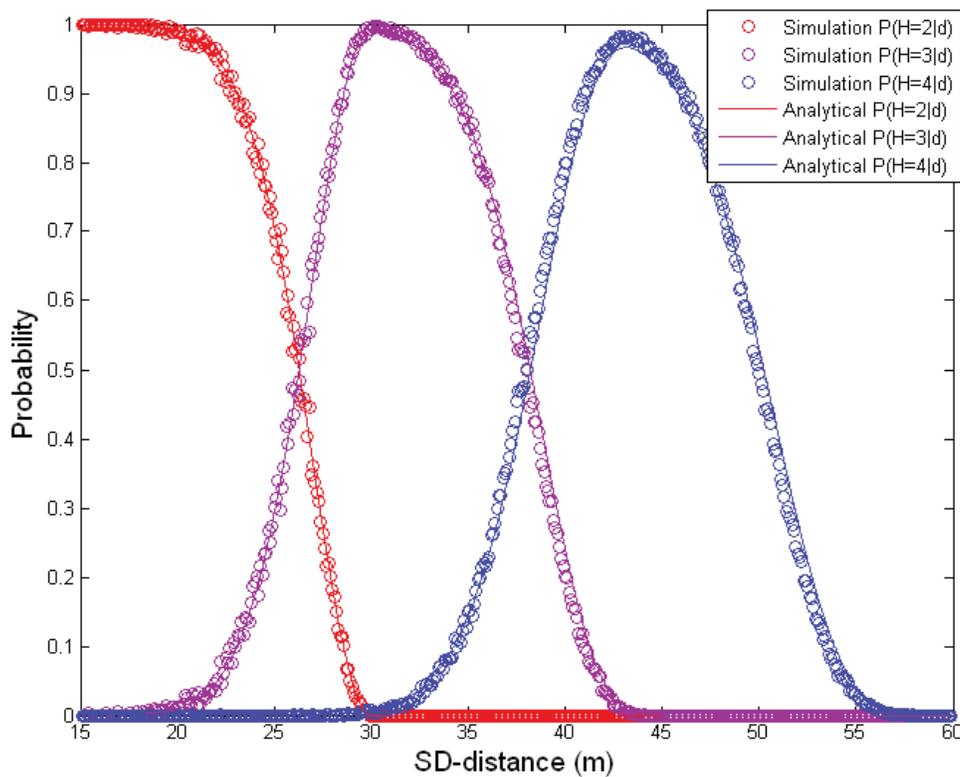


Fig. 14. Comparison of the conditional probability $P(h = K | d)$ in three-dimensional network based on the simulation results and the analytical results, $\lambda = 0.002210$.

C. Conditional probability density function $f(d | H = K)$

In this sub-section, we compare the simulation results and the analytical results of the conditional PDF $f(d | h = K)$. The simulation results and the analytical results of $f(d | h = K)$ ($K = 2$ to 5) are plotted in Fig. 16, 17, 18 and 19 respectively. The density of the network is set to $\lambda = 0.03537$. It is observed that the analytical results match the simulation results very well. Some literatures ([3] and [8]) study the Guassianity of the conditional PDF $f(d | h = K)$ by using Skewness and Kurtosis as a measure, and then use Gaussian PDF to fit the simulation data to approximate the conditional PDF $f(d | h = K)$. These heuristic approaches of approximating $f(d | h = K)$ require statistical fitting. The method proposed in this chapter, on the other hand, is a general formulation of the conditional PDF $f(d | h = K)$. We also plot the simulation results and the analytical results of the PDF $f(d | h = K)$ ($K = 2$ to 5) for networks with density $\lambda = 0.08842$ in Fig. 20, 21, 22 and 23 respectively. We observe that the analytical results also match the simulation results very well. Quantitatively, we plot the MAE_K error against the density λ of network in Fig. 15. In this case, MAE_K is defined as follows:

$$MAE_K = \frac{1}{M} \sum_{i=1}^M |f_a(d_i | h = K) - f_s(d_i | h = K)| \quad (23)$$

where $f_a(d | h = K)$ and $f_s(d | h = K)$ are the analytical and simulation result respectively. As shown in Fig. 15, the MAE is of the order of 10^{-3} . The error is indeed very small.

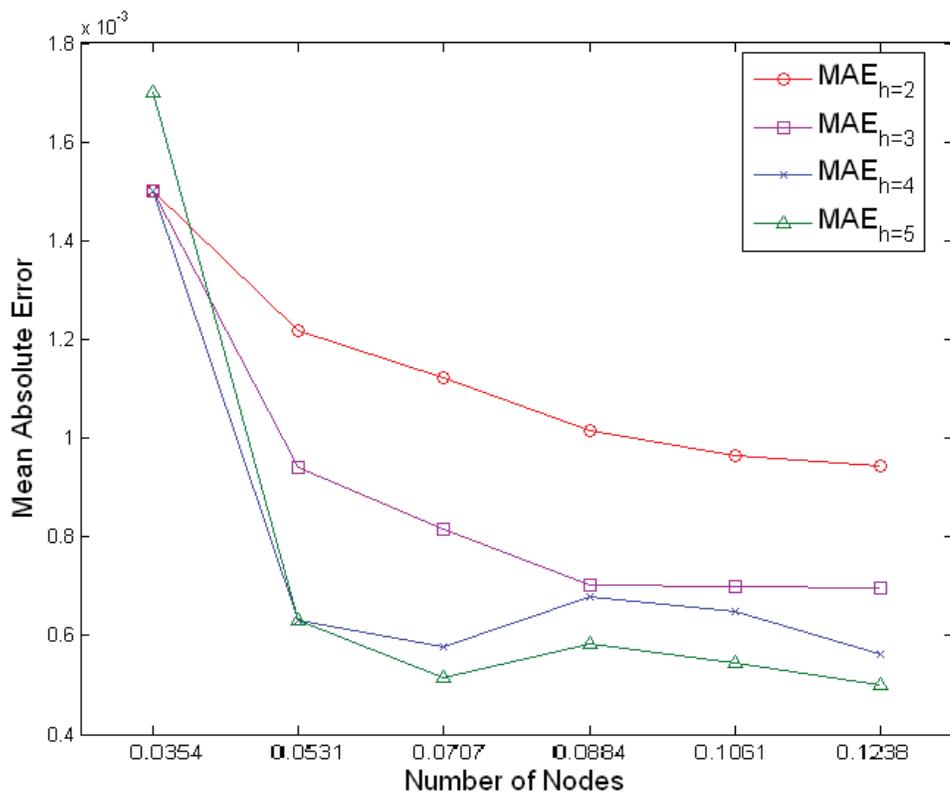


Fig. 15. Mean Absolute Error of the analytical result $f(d|h = K)$ against the number of node deployed.

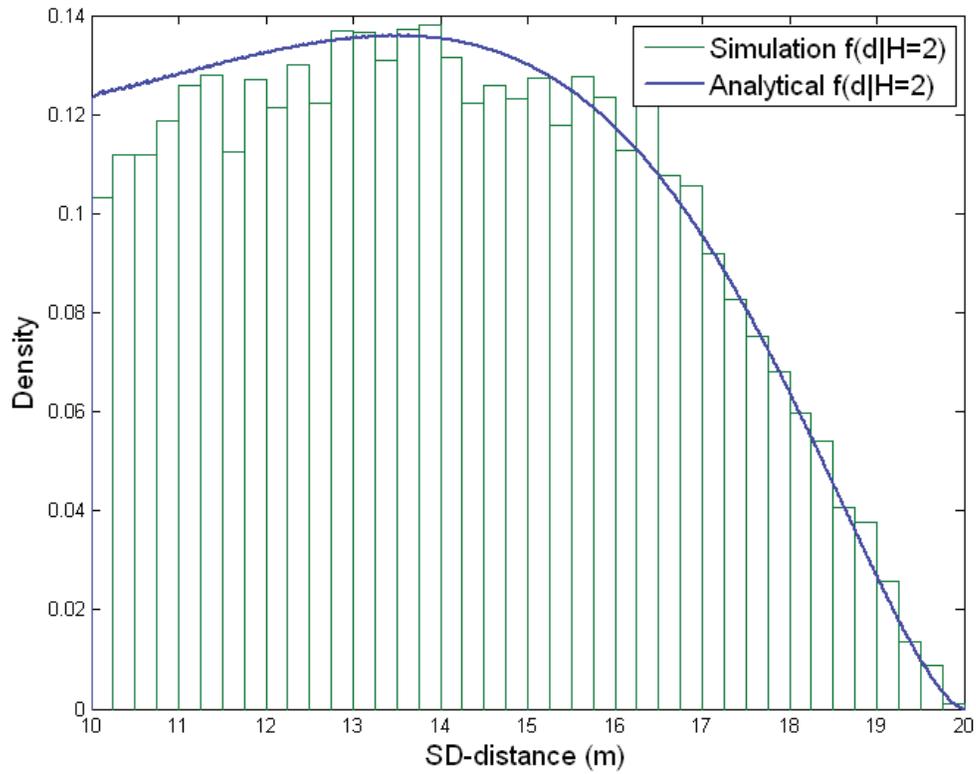


Fig. 16. The conditional pdf $f(d|h = 2)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.03537$.

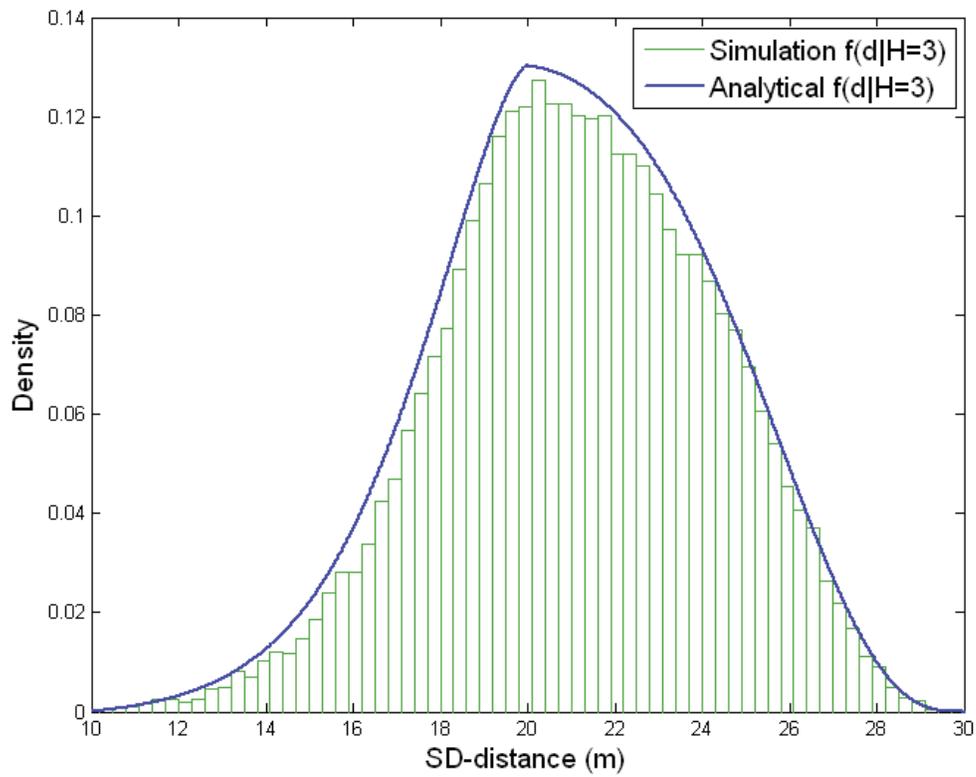


Fig. 17. The conditional pdf $f(d|h=3)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.03537$.

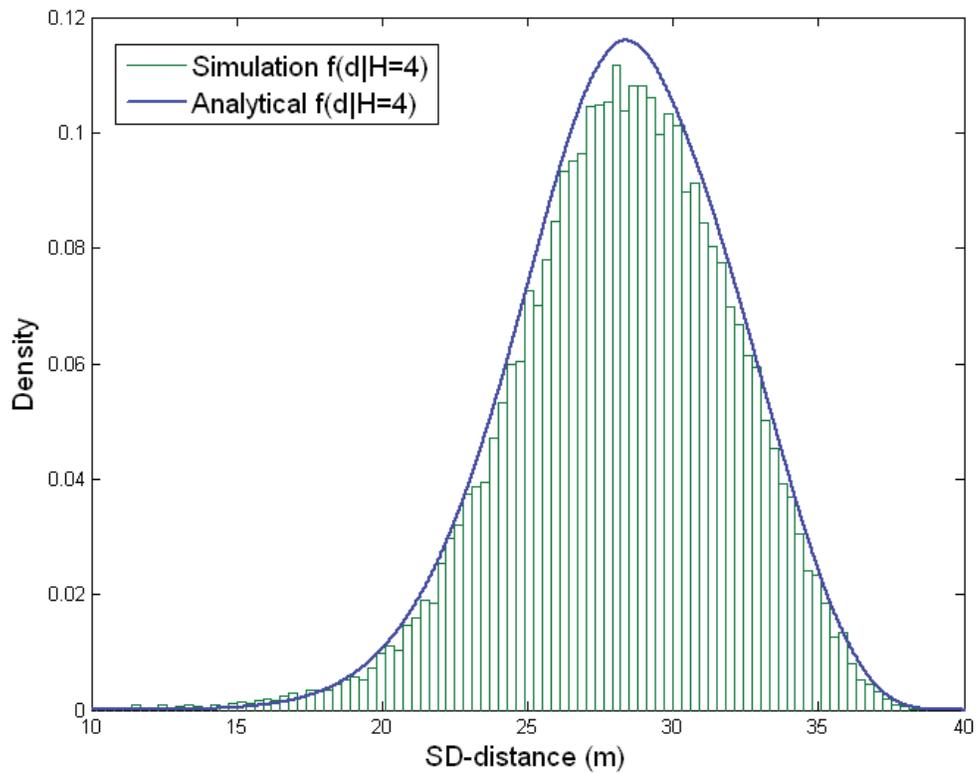


Fig. 18. The conditional pdf $f(d|h=4)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.03537$.

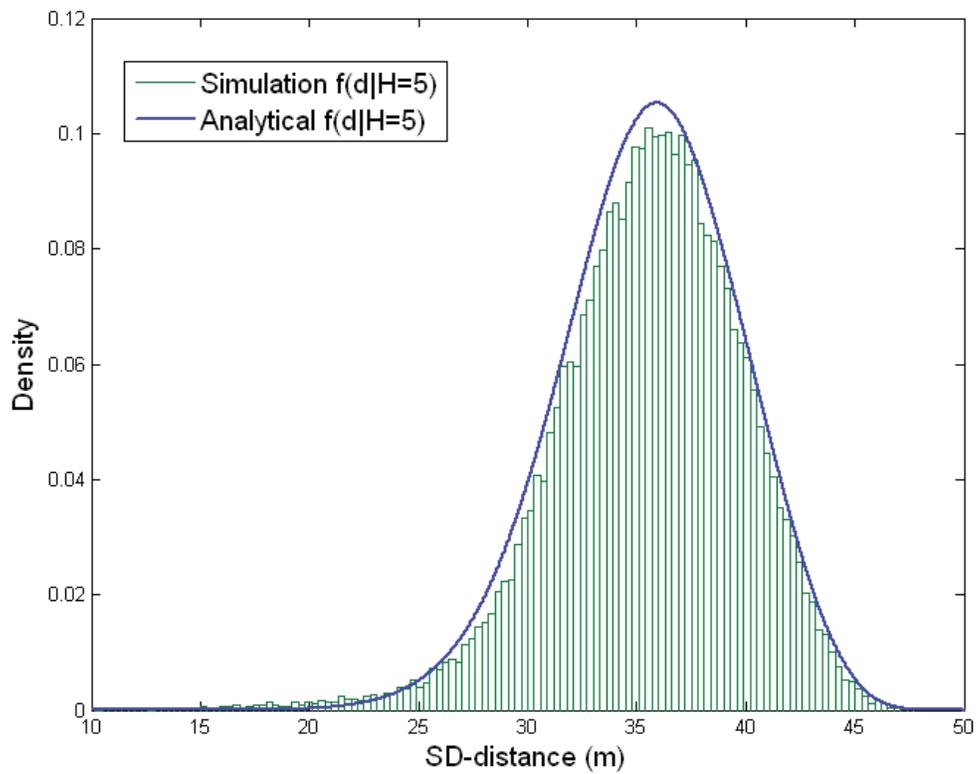


Fig. 19. The conditional pdf $f(d | h = 5)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.03537$.

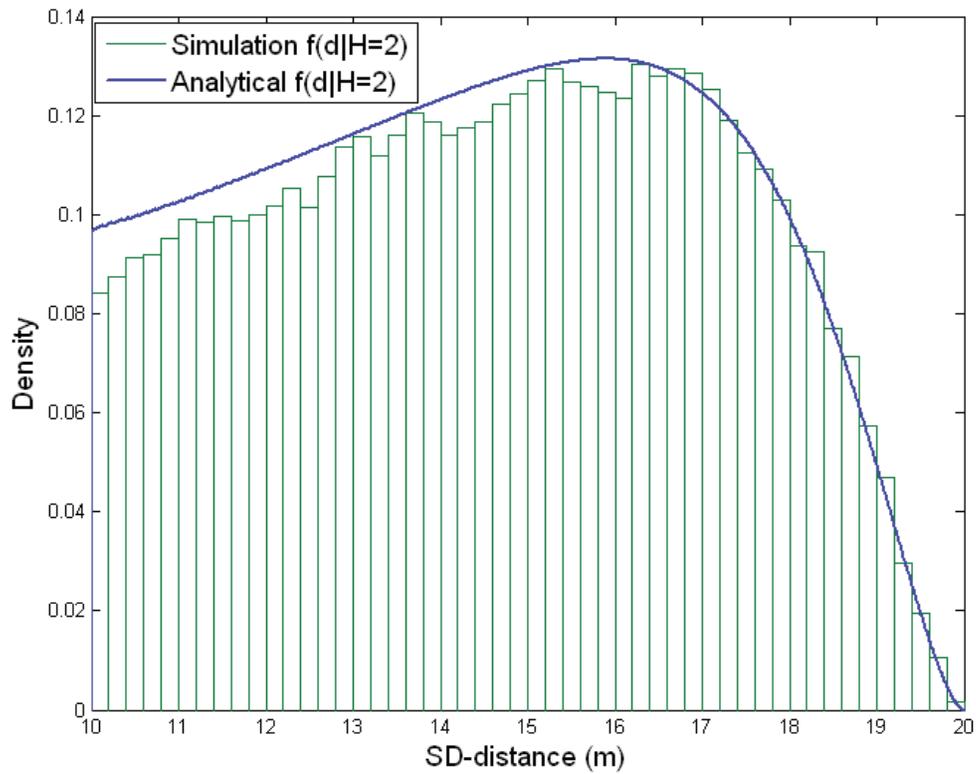


Fig. 20. The conditional pdf $f(d | h = 2)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.08842$.

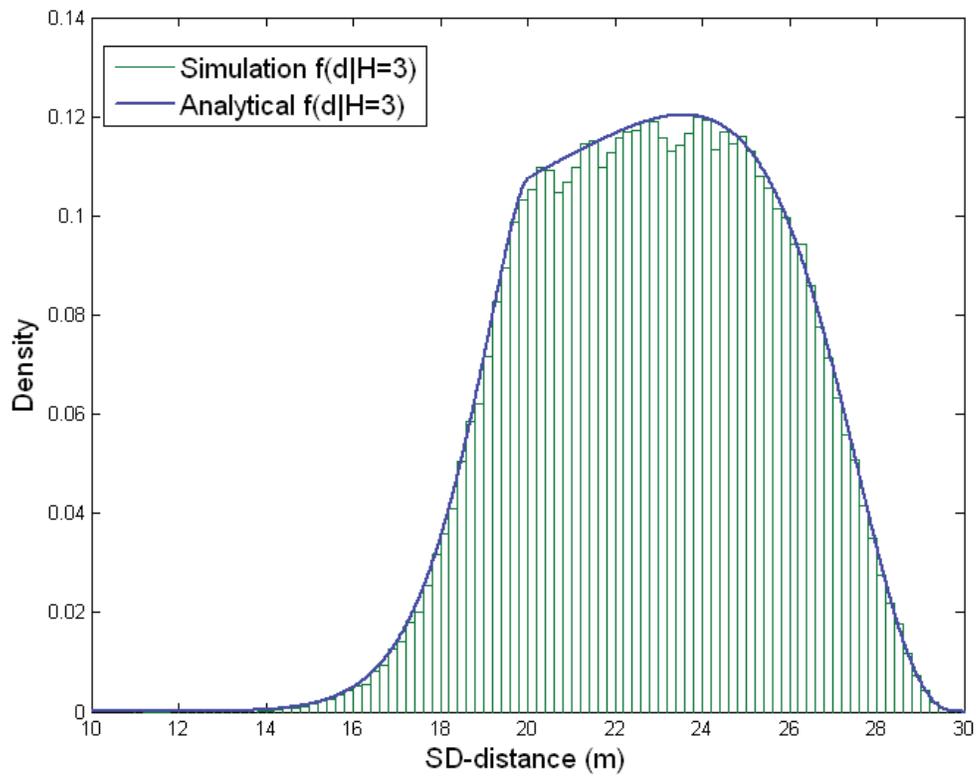


Fig. 21. The conditional pdf $f(d|h=3)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.08842$.

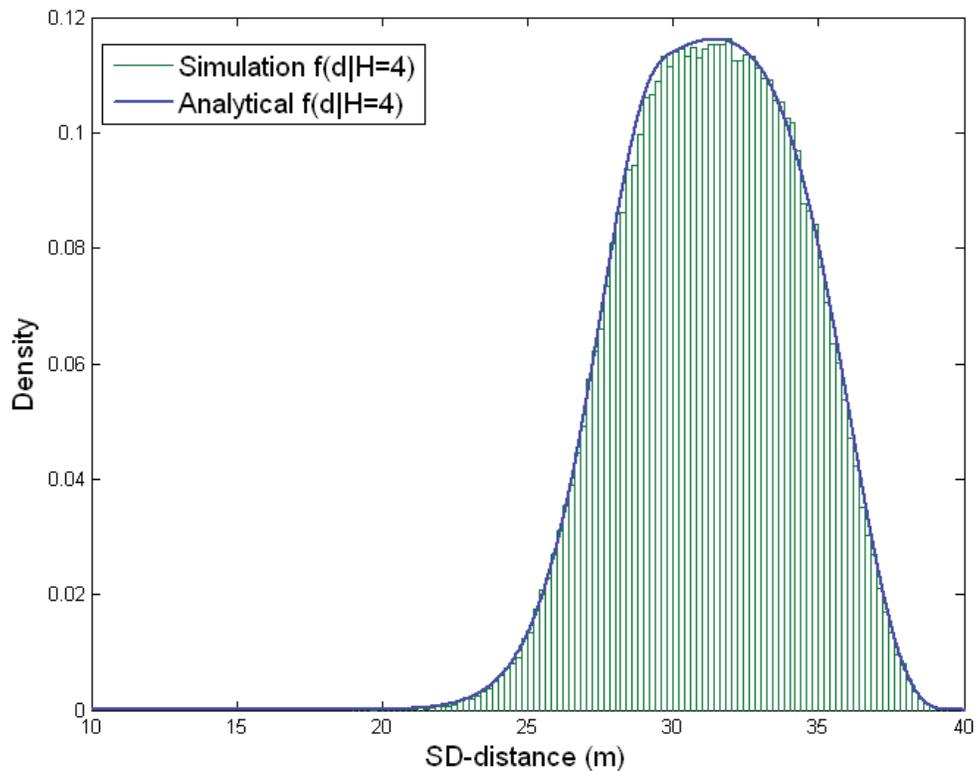


Fig. 22. The conditional pdf $f(d|h=4)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.08842$.

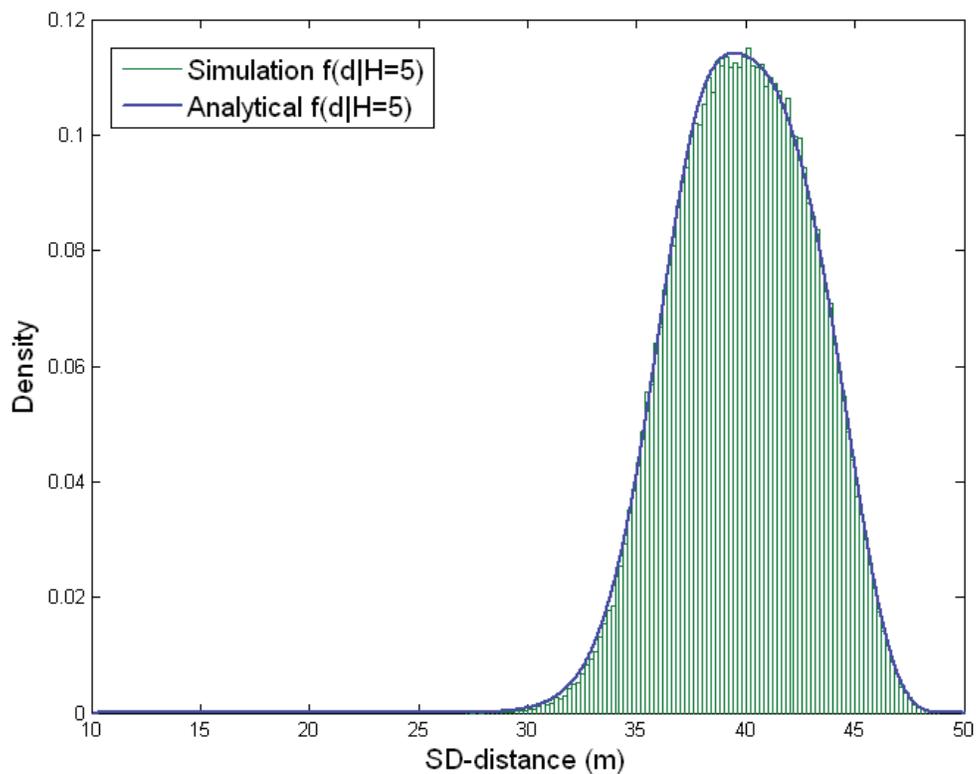


Fig. 23. The conditional pdf $f(d|h=5)$ based on the simulation data (histogram plot) and the analytical result, $\lambda = 0.08842$.

7. Conclusions

Evaluation of various statistical relationships between hop-count and source-to-sink distance are some of the fundamental research problems in large-scale WSNs. In this chapter, we investigate two statistical relationships between hop-count and SS-distance, namely the conditional probability $P(h = K|d)$ and the conditional PDF $f(d|h = K)$. We propose a method termed CSP to compute $P(h = K|d)$ in both two and three-dimensional networks. This method is also extended to compute the conditional PDF $f(d|h = K)$.

Simulation results show that significant error reduction in computing the conditional probability and the conditional PDF can be achieved compared with the existing methods.

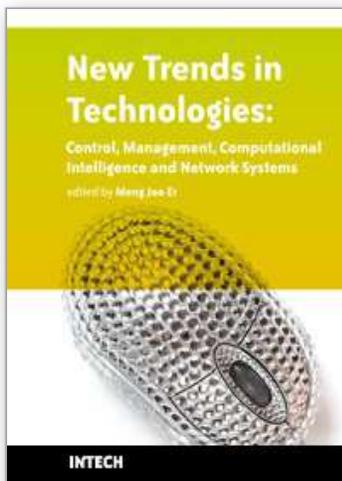
8. References

- [1] B. Wang, C. Fu and H. B. Lim, "Layered diffusion based coverage control in wireless sensor networks," in *IEEE the 32rd Conference on Local Computer Networks (LCN)*, 2007, pp. 504–511.
- [2] Y. Cheng and T. G. Robertazzi, "Critical connectivity phenomena in multihop radio models," *IEEE Transactions on Communications*, vol. 37, no. 7, 1989.
- [3] S. Vural and E. Ekici, "Probability distribution of multi-hop-distance in one-dimensional sensor networks," *Computer Networks (Elsevier)*, vol. 51, no. 3, pp. 3727–3749, 2007.
- [4] S. Dulman, M. Rossi, P. Havinga and M. Zorzi, "On the hop count statistics for randomly deployed wireless sensor networks," *Int. J. Sensor Networks*, vol. 1, 2006.
- [5] C. Bettstetter and J. Eberspacher, "Hop distances in homogeneous ad hoc networks," *Vehicular Technology Conference*, vol. 4, pp. 2286 – 2290, 2003.
- [6] S. Chandler, "Calculation of number of relay hops required in randomly located radio network," *IEEE Electronic Letter*, vol. 25, no. 24, 1989.
- [7] X. Ta, G. Mao and B. D. Anderson, "On the probability of k-hop connection in wireless sensor networks," *IEEE Communication Letters*, vol. 11, no. 9, 2007.
- [8] L. Zhao and Q. Liang, "Hop-distance estimation in wireless sensor networks with applications to resources allocation," *EURASIP Journal on Wireless Communication and Networking*, 2007.
- [9] L. E. Miller, "Probability of a two-hop connection in a random mobile network," *Proc. 35th Conf. on Information Sciences and Systems*, 2001.
- [10] L. Kleinrock and J. Silvester, "Optimum transmission radii for packet radio networks or why six is a magic number," *IEEE National Telecommunications Conference*, 1978.
- [11] J. Kuo and W. Liao, "Hop count distribution of multihop paths in wireless networks with arbitrary node density modeling and its applications," *IEEE Transactions on Vehicular Technology*, vol. 56, no. 4, 2007.
- [12] Y. Wang, X. Wang, D. Wang and D. P. Agrawal, "Range-free localization using expected hop progress in wireless sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, 2008.
- [13] S. De, A. Caruso, T. Chaira and S. Chessa, "Bounds on hop distance in greedy routing approach in wireless ad hoc networks," *International Journal of Wireless and Mobile Computing*, vol. 1, no. 2, 2006.
- [14] D. Niculescu and B. Nath, "Ad hoc positioning system (aps)," *Global Telecommunications Conference*, vol. 5, pp. 2926 – 2931, 2001.
- [15] S. Y. Wong, J. G. Lim, S. Rao and W. K. Seah, "Density-aware hop-count localization (dhl) in wireless sensor networks with variable density," *Wireless Communications and Networking Conference, IEEE*, vol. 3, pp. 1848 – 1853, March 2005.
- [16] S. Yang, J. Yi and H. Cha, "Hcrl: A hop-count-ratio based localization in wireless sensor networks," *IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pp. 31–40, June 2007.
- [17] T. S. Rappaport, *Wireless communications: principles and practice*, Prentice Hall PTR, 1996.

- [18] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," *ACM international symposium on Mobile ad hoc networking and computing*, 2002.

IntechOpen

IntechOpen



New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems

Edited by Meng Joo Er

ISBN 978-953-307-213-5

Hard cover, 438 pages

Publisher Sciyo

Published online 02, November, 2010

Published in print edition November, 2010

The grandest accomplishments of engineering took place in the twentieth century. The widespread development and distribution of electricity and clean water, automobiles and airplanes, radio and television, spacecraft and lasers, antibiotics and medical imaging, computers and the Internet are just some of the highlights from a century in which engineering revolutionized and improved virtually every aspect of human life. In this book, the authors provide a glimpse of the new trends of technologies pertaining to control, management, computational intelligence and network systems.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Di Ma and Meng Joo Er (2010). Theoretical Issues in Modeling of Large-Scale Wireless Sensor Networks, *New Trends in Technologies: Control, Management, Computational Intelligence and Network Systems*, Meng Joo Er (Ed.), ISBN: 978-953-307-213-5, InTech, Available from: <http://www.intechopen.com/books/new-trends-in-technologies--control--management--computational-intelligence-and-network-systems/theoretical-issues-in-modeling-of-large-scale-wireless-sensor-networks>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2010 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](#), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen