Probability of k-hop Connection under Random Connection Model

Guoqiang Mao, *Senior Member, IEEE*, Zijie Zhang, *Student Member, IEEE*, and Brian D.O. Anderson, *Fellow, IEEE*

***Abstract*— Consider a wireless sensor network with *i.i.d.* sensors following a homogeneous Poisson distribution in a given area A in $\Re^2$. A sensor located at $\boldsymbol{x}_2 \in A$ is directly connected to a sensor located at $\boldsymbol{x}_1 \in A$ with probability $g(\boldsymbol{x}_2 - \boldsymbol{x}_1)$, independent of any other distinct pair of sensors. In this letter, we provide a recursive formula for computing $\Pr(k|\boldsymbol{x})$, the probability that a node $\boldsymbol{x} \in A$ apart from another node is connected to that node at exactly k hops, for a generic random connection function $g : \Re^2 \to [0,1]$. The recursive formula is accurate for $k = 1, 2$ and provides an approximation for $\Pr(k|\boldsymbol{x})$ for $k > 2$. The exact and approximate analytical results are validated by simulations. The knowledge of $\Pr(k|\boldsymbol{x})$ can be used in a number of areas in sensor networks.**

***Index Terms*— Random connection model, k-hop connection**

I. Introduction

Consider a wireless sensor network (WSN) with *i.i.d.* sensors following a homogeneous Poisson distribution with known density ρ in a given 2D area, denoted by A. A sensor at $\boldsymbol{x}_2 \in A$ is directly connected to a sensor at $\boldsymbol{x}_1 \in A$ with probability $g(\boldsymbol{x}_2 - \boldsymbol{x}_1)$, termed the *connection function*, where $g : \Re^2 \to [0,1]$, independent of the event that another distinct pair of sensors are directly connected. There are three related probabilities characterizing the connectivity properties of such a network, i.e. the probability that an arbitrary node is k-hops apart from another arbitrary node, $\Pr(k)$ (i.e. the length of the shortest path from the first node to the second node measured by the number of hops is k); the probability that a node $\boldsymbol{x}$ apart from another node is connected to that node in exactly k hops, $\Pr(k|\boldsymbol{x})$; and the spatial distribution of the nodes k-hops apart from another designated node, $\Pr(\boldsymbol{x}|k)$. These three probabilities are related through Bayes' formula and if one is computable, the other two will be computable using similar techniques. Therefore we call these problems collectively *the probability of k-hop connection problems*.

Solutions to the probability of k-hop connection problems can be used in a number of areas in sensor networks. The probability $\Pr(k)$ is useful in estimating the overall energy consumption, lifetime and capacity of a WSN [1], [2]. The probability $\Pr(k|\boldsymbol{x})$ can be used in the analysis of end-to-end delay and energy consumption, and reliability of packet transmission [3], [4], [5] and the probability in vehicular networks that a vehicle can access the base station within a designated number of hops [6]. The probability $\Pr(\boldsymbol{x}|k)$ is useful in estimating the distance between two nodes from their neighborhood information and obtaining variance of such estimate, which can then be used in localization [7].

The conditional probability $\Pr(k|\boldsymbol{x})$ was first investigated by Chandler [8] in which he analysed the probability that two random radio stations separated by a known distance can communicate in k or less hops where stations are uniformly distributed on a flat earth and any two stations are directly connected if their Euclidean distance is less than a given threshold r, i.e. following the *unit disk connection model* for which $g(\boldsymbol{x}) = 1$ when $\|\boldsymbol{x}\| \le r$ and $g(\boldsymbol{x}) = 0$ otherwise. Here $\|\boldsymbol{x}\|$ denotes the Euclidean norm of $\boldsymbol{x}$. Note that under an omnidirectional connection model, including the widely used unit disk model (UDM) and log-normal model [5], $\Pr(k|\boldsymbol{x}) = \Pr(k|\,\|\boldsymbol{x}\|)$. In [5] Mukherjee and Avidor presented a technique to compute $\Pr(k|\,\|\boldsymbol{x}\|)$ in a network with Poissonly distributed nodes, where the wireless links between nodes are subject to log-normal fading. In [9] Bettstetter et al. analysed $\Pr(k)$ under the UDM for $k = 1, 2$ and where a total of n nodes are *i.i.d.* in a rectangular area following uniform distribution. In [10] Miller analysed $\Pr(k)$ for $k = 1, 2$ under a similar setting as that in [9] for Gaussian node distribution. In [11], Dulman et al. investigated $\Pr(k|\,\|\boldsymbol{x}\|)$ in 1D and 2D networks with Poissonly distributed nodes under the UDM. In [1], Zorzi and Rao analyzed under the UDM the average progress per hop for Poissonly distributed nodes whose path is established using the greedy forwarding protocol (GFP). On that basis they derived $\Pr(k|\,\|\boldsymbol{x}\|)$ for networks using the GFP. In [12], [13] the authors analyzed $\Pr(k|\,\|\boldsymbol{x}\|)$ under the UDM for Poissonly distributed nodes. Further the authors pointed out the existence of the so-called *spatial dependence problem* in the analysis of the hop count statistics, i.e. in 2D networks the event that a node B is a k-hop neighbor of a node C and the event that another node D is a k-hop neighbor of the same node C are dependent for $k \ge 2$, a fact which has been incorrectly ignored in many previous studies. In this letter we advance the earlier work by giving a recursive formula for computing $\Pr(k|\boldsymbol{x})$ under the more practical generic random connection model, where a sensor located at $\boldsymbol{x}_2$ is directly connected to a sensor located at $\boldsymbol{x}_1$ with probability $g(\boldsymbol{x}_2 - \boldsymbol{x}_1)$, independent of the event that another distinct pair of sensors are directly connected. The recursive formula is accurate for $k = 1, 2$ and provides an approximation for $\Pr(k|\boldsymbol{x})$ for $k > 2$. Our result incorporates the earlier results obtained under the UDM [12], [13] and the log-normal model [5] as its two special cases. Further, the impact of boundary effects is included in the computation of $\Pr(k|\boldsymbol{x})$. Spatial dependence however causes some errors for $k > 2$, see later simulations.

G. Mao and Z. Zhang are with the University of Sydney and National ICT Australia. B.D.O. Anderson is with Australian National University and National ICT Australia.

This work is supported by the ARC discovery project: DP0877562.

II. Analysis of Probability of k-Hop Connection

Without loss of generality, consider that there is a node located at the origin. The probability that a node located at $\boldsymbol{x}$ is directly connected to the node at the origin is $g(\boldsymbol{x})$. Due to the independence of connections between nodes, the set of nodes directly connected to the node at the origin, denoted by K_1, can be shown to have an inhomogeneous Poisson distribution with density $\rho g(\boldsymbol{x})$ [14, Proposition 1.3]. Obviously $Pr(k=1|\boldsymbol{x}) = g(\boldsymbol{x})$. It follows that the set of nodes *not* directly connected to the node at the origin, denoted by $\overline{K_1}$, has an inhomogeneous Poisson distribution with density $\rho(1-g(\boldsymbol{x}))$.

Due to the inhomogeneous Poisson distribution of K_1, the probability that there is a node in K_1 within a differential area $dA_{\boldsymbol{x}}$ centered at $\boldsymbol{x}$ is $\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}}$. Without loss of generality, we assume that the node in $dA_{\boldsymbol{x}}$ is located at $\boldsymbol{x}$. The probability that there is more than one node in the differential area $dA_{\boldsymbol{x}}$ can be ignored due to the Poisson distribution of nodes. Therefore the probability that a node at $\boldsymbol{y} \notin dA_{\boldsymbol{x}}$ is *not* directly connected to *any* of the nodes within K_1 is given by

$$\begin{aligned}
& \lim_{dA_{\boldsymbol{x}}\to 0} \prod_{dA_{\boldsymbol{x}}\subset A, dA_{\boldsymbol{x}}\cap\{\boldsymbol{y}\}=\emptyset} [(1-g(\boldsymbol{y}-\boldsymbol{x}))\,\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}} \\
& + (1-\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}})] && (1)\\
= & \lim_{dA_{\boldsymbol{x}}\to 0} e^{\sum_{dA_{\boldsymbol{x}}\subset A, dA_{\boldsymbol{x}}\cap\{\boldsymbol{y}\}=\emptyset} \log(1-g(\boldsymbol{y}-\boldsymbol{x})\rho g(\boldsymbol{x})dA_{\boldsymbol{x}})} && (2)\\
= & \lim_{dA_{\boldsymbol{x}}\to 0} e^{\sum_{dA_{\boldsymbol{x}}\subset A, dA_{\boldsymbol{x}}\cap\{\boldsymbol{y}\}=\emptyset} (-g(\boldsymbol{y}-\boldsymbol{x})\rho g(\boldsymbol{x})dA_{\boldsymbol{x}})} && (3)\\
= & e^{-\int_A \rho g(\boldsymbol{y}-\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}} && (4)
\end{aligned}$$

In Eq. 1, the term $(1-g(\boldsymbol{y}-\boldsymbol{x}))\,\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}}$ represents the probability that there is a node in $dA_{\boldsymbol{x}}$ from K_1 (i.e. the $\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}}$ term) *and* the node at $\boldsymbol{y}$ is not directly connected to that node in $dA_{\boldsymbol{x}}$ at location $\boldsymbol{x}$ (i.e. the $1-g(\boldsymbol{y}-\boldsymbol{x})$ term). The term $1-\rho g(\boldsymbol{x})\,dA_{\boldsymbol{x}}$ represents the probability that there is *no* node in $dA_{\boldsymbol{x}}$ from K_1.

It follows from Eq. 4 that the probability that the node at $\boldsymbol{y}$ is directly connected to *at least one* of the nodes in K_1, denoted by $g_2(\boldsymbol{y})$, is given by

$$g_2(\boldsymbol{y}) = 1 - e^{-\int_A \rho g(\boldsymbol{y}-\boldsymbol{x})g(\boldsymbol{x})d\boldsymbol{x}} \qquad (5)$$

A node at $\boldsymbol{y}$ is connected to the node at the origin in exactly two hops if and only if it is not directly connected to the node at the origin *and* it is directly connected to at least one node in K_1. Therefore

$$\Pr(k=2|\boldsymbol{x}) = (1-g(\boldsymbol{x}))\,g_2(\boldsymbol{x})$$

For consistency in notation we also use $g_1(\boldsymbol{x})$ for $g(\boldsymbol{x})$. Due to the spatial dependence problem mentioned in [12], [13], the event that a node is directly connected to another node in k hops and the event that a third node is directly connected to the same node in k hops are *dependent* for $k \geq 2$. In this letter, we ignore such dependence and consider the above two events to be *approximately* independent. Using the independence approximation, it then follows from [14, Proposition 1.3] that the set of nodes that are directly connected to the node at the origin in exactly two hops, denoted by K_2, has an *approximate* inhomogeneous Poisson distribution with density $\rho(1-g(\boldsymbol{x}))\,g_2(\boldsymbol{x})$. The set of nodes that are *not* connected to the node at the origin within two hops, denoted by $\overline{K_1+K_2}$, has an approximate inhomogeneous Poisson distribution with density $\rho(1-g_1(\boldsymbol{x}))(1-g_2(\boldsymbol{x}))$.

Using the same steps that lead to Eq. 5 and the fact that K_2 has an *approximate* inhomogeneous Poisson distribution with density $\rho(1-g(\boldsymbol{x}))\,g_2(\boldsymbol{x})$, it can be shown that the probability that a node at $\boldsymbol{y}$ is directly connected to *at least one* of the nodes in K_2, denoted by $g_3(\boldsymbol{y})$, is given by

$$g_3(\boldsymbol{y}) = 1 - e^{-\int_A \rho g_1(\boldsymbol{y}-\boldsymbol{x})(1-g_1(\boldsymbol{x}))g_2(\boldsymbol{x})d\boldsymbol{x}} \qquad (6)$$

It then follows that

$$\Pr(k=3|\boldsymbol{x}) = (1-g_1(\boldsymbol{x}))(1-g_2(\boldsymbol{x}))\,g_3(\boldsymbol{x})$$

Using the independence approximation, it can be shown from [14, Proposition 1.3] that the set of nodes within $\overline{K_1+K_2}$ that are directly connected to at least one node in K_2, denoted by K_3, has an approximate inhomogeneous Poisson distribution with density $\rho(1-g_1(\boldsymbol{x}))(1-g_2(\boldsymbol{x}))\,g_3(\boldsymbol{x})$.

By recursion, for a positive integer $l > 1$, it can be shown that (adopting the independence assumption)

$$\Pr(k=l|\boldsymbol{x}) = g_l(\boldsymbol{x}) \prod_{i=1}^{l-1} (1-g_i(\boldsymbol{x})) \qquad (7)$$

where

$$g_l(\boldsymbol{y}) = 1 - e^{-\int_A \rho g_1(\boldsymbol{y}-\boldsymbol{x})g_{l-1}(\boldsymbol{x})\prod_{i=1}^{l-2}(1-g_i(\boldsymbol{x}))d\boldsymbol{x}}$$

Given knowledge of $\Pr(k|\boldsymbol{x})$ and $\Pr(\boldsymbol{x})$, i.e. the distribution of the displacement between two randomly selected nodes in A [15], $\Pr(\boldsymbol{x}|k)$ and $\Pr(k)$ can be readily obtained using Bayes' formula.

Note that the independence approximation is only needed for the computation of $\Pr(k=l|\boldsymbol{x})$ with $l > 2$, and is not required for the computation of $\Pr(k=1|\boldsymbol{x})$ and $\Pr(k=2|\boldsymbol{x})$. Therefore the results on $\Pr(k=l|\boldsymbol{x})$ in Eq. 7 are accurate for $l = 1, 2$ and are an approximation only for $l > 2$.

III. Simulation

In this section we use simulations to establish the accuracy of the analytical result on $\Pr(k|\boldsymbol{x})$. In the simulation, sensors are deployed in a 2000×2000 square area according to a homogeneous Poisson process with density ρ. Two most widely used connection models, i.e. the UDM and the log-normal model, are used as specific examples of the generic connection g. Under the UDM, the transmission range r is 100, i.e. $g(\boldsymbol{x}) = 1$ for $\|\boldsymbol{x}\| \leq 100$ and $g(\boldsymbol{x}) = 0$ for $\|\boldsymbol{x}\| > 100$. Simulations are conducted for a number of node densities but only the results for a node density which gives an average node degrees 40 are shown. Results using other node densities showed similar trend. Under the log-normal model, a node B is directly connected to another node C if the power received from C at B, whose propagation follows the log-normal model, is greater than a given threshold, P_T. It then follows that under the log-normal model

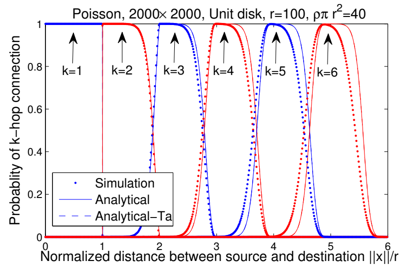

Fig. 1. Conditional probability $\Pr(k|\,\|\boldsymbol{x}\|)$ under the unit disk model for k =1 to 6. Analytical-Ta is the result from [13]. Our result is marginally more accurate than the result in [13] for $k > 2$ however the two analytical results are mostly indistinguishable.

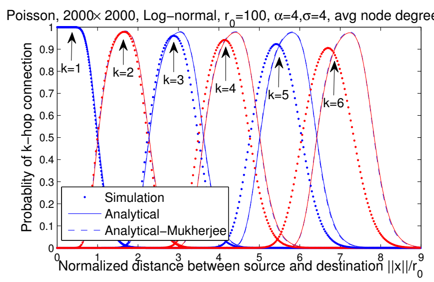

Fig. 2. Conditional probability $\Pr(k|\,\|\boldsymbol{x}\|)$ under the log-normal model for k =1 to 6. Analytical-Mukherjee is the result from [5]. Our result is marginally more accurate than the result in [5] for $k > 2$ however the two analytical results are mostly indistinguishable.

$$g(\boldsymbol{x}) = \Pr\left(P_r\left(\|\boldsymbol{x}\|\right) \geq P_T\right) = \int_{10\alpha\log_{10}\frac{\|x\|}{r_0}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2\sigma^2}} dz$$

where $r_0 = d_0 10^{\frac{P_t - PL_0(d_0) - P_T}{10\alpha}}$ and P_r is the received power in dB milliwatts, P_t is the transmitted power in dB milliwatts, $\|\boldsymbol{x}\|$ is the Euclidean distance between the two nodes, $PL_0(d_0)$ is the reference path loss in dB at a reference distance d_0, α is the path loss exponent and σ is the standard deviation of the log-normal fading. The path loss exponent depends on the environment and can vary between 2 in free space to 6 in urban areas and the value of σ can be as large as 12. Several values of α and σ have been used in the simulation, but only the result for $\alpha = 4$ and $\sigma = 4$ is shown because other results have similar accuracy. Under the log-normal shadowing model, r_0 is chosen to be 100, while ρ is chosen to give the same average degree as that in the UDM. Every point shown in the plots is the average value from 2000 simulations. As the number of instances of random networks used in the simulation is large, the confidence interval is too small to be distinguishable and hence is ignored in the following plots. Our analytical result is compared with the analytical result from [13] obtained under the UDM in Fig. 1, and the analytical result from [5] obtained under the log-normal model in Fig. 2 respectively. Note that under an omnidirectional model, such as the UDM and the log-normal model considered in the simulation, $\Pr(k|\boldsymbol{x}) = \Pr(k|\,\|\boldsymbol{x}\|)$.

As shown in Fig. 1 and Fig. 2, our result is marginally more accurate than the result in [13] obtained under the UDM and the result in [5] obtained under the log-normal model due to the inclusion of the boundary effect in our analysis. However, unsurprisingly, in most cases our analytical result is almost indistinguishable from the previous results [5], [13], reflecting the negligible impact of the boundary effect on the analysis of $\Pr(k|\boldsymbol{x})$. The main contribution of our analysis is that it is applicable for a wider class of wireless channel models whereas the results in [13] and in [5] are only valid for the particular channel model being considered.

The discrepancy between the analytical result and the simulation result starts to appear for $k > 2$. It is due to the spatial dependence problem [12], [13] and the associated independence approximation mentioned earlier. The discrepancy between the analytical and the simulation results is larger for a larger value of k. This is caused both by the independence approximation and the accumulation of errors in the recursive procedure for computing $\Pr(k|\boldsymbol{x})$. Such errors appear to be acceptable for some applications [1], [2], [7], [4], [5] but not necessarily for all. It is beyond the scope of this letter to discuss the performance implication of the error on specific applications.

References

[1] M. Zorzi and R. R. Rao, "Geographic random forwarding (geraf) for ad hoc and sensor networks: multihop performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 337–348, 2003.

[2] ——, "Geographic random forwarding (geraf) for ad hoc and sensor networks: energy and latency performance," *IEEE Transactions on Mobile Computing*, vol. 2, no. 4, pp. 349–365, 2003.

[3] 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, pp. 131–140, 2006.

[4] L. Zhao and Q. Liang, "Hop-distance estimation in wireless sensor networks with applications to resource allocation," *EURASIP Journal on Wireless Communications and Networking*, vol. 2007, p. 8, 2007.

[5] S. Mukherjee and D. Avidor, "Connectivity and transmit-energy considerations between any pair of nodes in a wireless ad hoc network subject to fading," *IEEE Transactions on Vehicular Technology*, vol. 57, pp. 1226–1242, 2008.

[6] Y. Yang, H. Hu, J. Xu, and G. Mao, "Relay technologies for wimax and lte-advanced mobile systems," *IEEE Communications Magazine*, vol. 47, no. 10, pp. 100–105, 2009.

[7] S. Dulman and P. Havinga, "Statistically enhanced localization schemes for randomly deployed wireless sensor networks," in *Proceedings of the 2004 Intelligent Sensors, Sensor Networks and Information Processing Conference*, P. Havinga, Ed., 2004, pp. 403–410.

[8] S. A. G. Chandler, "Calculation of number of relay hops required in randomly located radio network," *Electronics Letters*, vol. 25, no. 24, pp. 1669–1671, 1989.

[9] C. Bettstetter and J. Eberspacher, "Hop distances in homogeneous ad hoc networks," in *The 57th IEEE Semiannual Vehicular Technology Conference*, vol. 4, 2003, pp. 2286–2290.

[10] L. E. Miller, "Probability of a two-hop connection in a random mobile network," in *Proc. 35th Conf. on Information Sciences and Systems*, 2001.

[11] S. Dulman, M. Rossi, P. Havinga, and M. Zorzi, "On the hop count statistics for randomly deployed wireless sensor networks," *International Journal of Sensor Networks*, vol. 1, no. 1/2, pp. 89–102, 2006.

[12] X. Ta, G. Mao, and B. D. O. Anderson, "Evaluation of the probability of k-hop connection in homogeneous wireless sensor networks," pp. 1279 – 1284, 2007.

[13] ——, "On the probability of k-hop connection in wireless sensor networks," *IEEE Communications Letters*, vol. 11, no. 8, pp. 662–664, 2007.

[14] R. Meester and R. Roy, *Continuum Percolation*, ser. Cambridge Tracts in Mathematics. Cambridge University Press, 1996.

[15] L. E. Miller, "Distribution of link distances in a wireless network," *Journal of Research of the National Institute of Standards and Technology*, vol. 106, no. 2, pp. 401–412, 2001.