Secure k-Connectivity Properties of Wireless Sensor Networks

Yee Wei Law[1] Li-Hsing Yen[2]
Roberto Di Pietro[3]
Marimuthu Palaniswami[1]

Depart. of Electrical and Electronic Engineering[1],
The University of Melbourne,
Australia
{ywlaw,swami}@ee.unimelb.edu.au
Department of Computer Science[2],
National University of Kaohsiung,
Taiwan ROC
lhyen@nuk.edu.tw
Department of Mathematics[3],
Università di Roma Tre,
Italy
dipietro@di.uniroma1.it

Abstract

A k-connected wireless sensor network (WSN) allows messages to be routed via one (or more) of at least k node-disjoint paths, so that even if some nodes along one of the paths fail, or are compromised, the other paths can still be used. This is a much desired feature in fault tolerance and security. k-connectivity in this context is largely a well-studied subject. When we apply the random key pre-distribution scheme to secure a WSN however, and only consider the paths consisting entirely of secure (encrypted and/or authenticated) links, we are concerned with the secure k-connectivity of the WSN. This notion of secure k-connectivity is relatively new and no results are yet available. The random key pre-distribution scheme has two important parameters: the key ring size and the key pool size. While it has been determined before the relation between these parameters and 1-connectivity, our work in k-connectivity is new. Using a recently introduced random graph model called kryptograph, we derive mathematical formulae to estimate the asymptotic probability of a WSN being securely k-connected, and the expected secure k-connectivity, as a function of the key ring size and the key pool size. Finally, our theoretical findings are supported by simulation results.

1. Introduction

Many mission-critical and military applications are envisaged for wireless sensor networks (WSNs). For such applications, securing the communications between sensor nodes is critical. Denote an n-node network by the undirected graph $G(V_n, E)$, where the vertex set V_n represents the nodes, and the edge set E represents *secure* communication links. In the extreme cases, either the same key is stored in every $v_i \in V_n$, or for every $(v_i, v_j) \in E$, a key is stored in v_i and v_j $(1 \leq i, j \leq n)$. In the former case, the whole network is vulnerable to a single-key compromise; and in the latter case, the required amount of storage per node does not scale. One distributed approach is to establish pairwise keys *probabilistically* between every two sensors and provide some assurance that the WSN is connected [8]. However, when end-to-end communications are at stake, even if a secure path (i.e., a path consisting entirely of secure links) can be established between two nodes, message confidentiality could be lost if any of the secure links along the path is compromised. Hence, to improve security, it is useful to be able to establish multiple secure node-disjoint paths between any two nodes.

In this paper, we investigate the secure k-connectivity properties of WSNs. Our motivation starts with a mechanism called *multipath routing*, i.e., routing messages through multiple paths. Multipath routing is important for load-balancing the traffic between source and destination nodes, and for increasing the reliability of data delivery [9]. However, how extensively multipath routing can be used depends on the connectivity of the network. There are two important parameters in the key pre-distribution scheme: the key ring size K, and the key pool size P (to be elaborated in the next section). Usually, K is fixed by hardware constraints. Intuitively, the lower P is, the higher the connectivity of the network becomes; but in terms of security, the higher P is, the more resilient the network becomes to node capture. For network designers, a goal is to strike a balance between network connectivity and resilience to node capture.

The central metric in our study is the minimum number of nodes to be compromised/removed before any pair of nodes are disconnected. In graph theory, this is called *vertex connectivity*, denoted κ, and if $\kappa = k$, we call the network k-connected. In a *securely* k-connected network, there exist at least k node-disjoint *secure* paths between each pair of nodes. In this paper, we are interested in these two metrics:

- The survivor function $\Pr\{\kappa \geq k\}$, i.e., the probability that a graph is k-connected.

- The expected connectivity, i.e., the mean of the connectivity computed over all the possible graph instances.

Our contribution is an analytical framework for evaluating how the key ring size K and the key pool size P influence the above two metrics. In particular, we provide close formulae for the two metrics that depend on K and P. To the best of our knowledge, we are the first to address secure k-connectivity of WSNs using the random key predistribution scheme, which is widely adopted in the literature. Finally, simulations results confirm our analytical findings.

The rest of this paper is organized as follows. Section 2 introduces the reference model. Section 3 discusses our analytic results. Simulation results are given in Section 4. Section 5 discusses related work. Finally, some concluding remarks are reported in Section 6.

2. The reference model

We begin by describing Eschenauer et al.'s random key pre-distribution scheme [8]. In this scheme, each sensor is pre-assigned a *key ring* of K secret keys randomly drawn from a common pool of P keys. The sensor nodes are then randomly deployed in the field. Two sensors share a secure communication link if they lie within each other's communication range and if they share at least a common pre-assigned key. A fundamental challenge is to choose K and P such that the network is connected with secure links alone.

The underlying theory of this scheme is best studied from two perspectives – from a random graphical perspective, and from a combinatorial perspective. A random graph $G(n, p_s)$ is a graph of n vertices (sensor nodes) for which the probability that an edge (a secure link) exists between any two vertices is independently determined by a coin flip of probability p_s [2]. If p_s is zero, then the graph is disconnected, and if p_s is one, the graph is fully connected, so there must exist a certain value of p_s such that the graph is almost surely connected. For a graph that is connected with probability P_c, Erdös et al. [7] show that

$$p_s = \frac{\ln n - \ln(-\ln P_c)}{n} \tag{1}$$

From a combinatorial viewpoint, the probability that any two nodes share at least one key is given by

$$p_s = 1 - \frac{\binom{P-K}{K}}{\binom{P}{K}} \tag{2}$$

Combining Equation 1 and Equation 2, and fixing n and the intended P_c, we can determine P from K, which is usually fixed by hardware contraints.

However, the random graph model used by Eschenauer et al. does not capture the real nature of a WSN, in particular, because the model does not take into account the distance metric between each pair of nodes. A model that has been proved to be more appropriate is the *kryptograph* model, introduced by Di Pietro et al. [6]. In the following, we give the definition of a *random geometric graph* (also called *unit disk graph*), and then the definition of a kryptograph.

Definition 1 (Penrose [12]) A random geometric graph $G(V_n; r)$ is an undirected graph with vertex set V_n uniformly distributed in a d-dimensional unit cube, and with undirected edges connecting all the pairs $v_i, v_j \in V_n$ that satisfy $L_p(v_i, v_j) \leq r$, where L_p is a Lebesgue metric.

Note: Lebesgue metric can be thought of as a "universal" metric [3]. Both the Euclidean distance and the toroidal distance are Lebesgue metrics.

Definition 2 A kryptograph $G(V_n; r, p_s)$ is a subgraph of a random geometric graph $G(V_n; r)$, with edge set $E = \{e | e \in G(V_n; r)$; Both ends of e share at least a key, subjected to a probability of $p_s\}$.

It is important to note that the notion of a k-connected kryptograph, and the notion of a securely k-connected random geometric graph are equivalent. We will use these definitions to derive the k-connectivity properties of kryptographs in the next section.

3. k-Connectivity of kryptographs

In the following, we will study these two metrics: (1) the survivor function $\Pr\{\text{connectivity} \geq k\}$; and (2) the expected connectivity.

3.1. Survivor function $\Pr\{\text{connectivity} \geq k\}$

The problem is formulated as such: given the key ring size K, what should the key pool size P be to achieve k-connectivity? We attack the problem by applying Theorem 1, which is in turn based on Theorem 2. Although both of these theorems are originally formulated for random geometric graphs, we will prove in Proposition 1 that they are also applicable to kryptographs. Table 1 summarizes the notation used in this paper.

Theorem 1 **(Bettstetter [1])** For $n \gg 1$,

$$\Pr\{G(V_n; r) \text{ is } k\text{-connected}\} \approx \Pr\{d_{\text{min}} \geq k\}$$

Table 1. Notation.

n	Total number of nodes
N	$n-1$
K	Key ring size
P	key pool size
κ	Vertex connectivity
r	Communication range
$\rho(V_n;\kappa \geq k)$	The k-connectivity threshold, i.e. the minimum value of r at which $G(V_n;r)$ becomes k-connected
$\rho(V_n;\delta \geq k)$	The k-nearest neighbor distance, i.e. the minimum value of r at which the minimum degree of $G(V_n;r)$ becomes k
d_i	Degree of node i. (d = degree of *a* node.)
d_{min}	Minimum degree of a graph
p	Probability that two nodes are within range
q	$1-p$
p_s	Probability that two nodes share at least one key
q_s	$1-p_s$
A	Network deployment area
n'	Average number of neighours of a node

Theorem 2 (Penrose [14]) Given $k \geq 1$, and the Lebesgue metric L_p is defined for $p > 1$,

$$\lim_{n\to\infty} \Pr\{\rho(V_n;\kappa \geq k) = \rho(V_n;\delta \geq k)\} = 1$$

Note: L_2 ($p = 2$) is the Euclidean norm or 2-norm.

The following theorem is our first core result. It specifies the necessary conditions on the key ring size and the key pool size for achieving k-connectivity in kryptographs.

Theorem 3 For $n \gg 1$, $k \geq 2$

$$\begin{aligned}
&\Pr\{G(V_n;r,p_s) \text{ is } k\text{-connected}\} \\
&\approx \Bigg\{1-(1-pp_s)^N - \sum_{i=1}^{k-1}\binom{N}{i}p^iq^{N-i}(1-q_s^i) \\
&\quad - \sum_{i=k}^{N}\left[\binom{N}{i}p^iq^{N-i}\sum_{j=1}^{k-1}\binom{i}{j}p_s^jq_s^{i-j}\right]\Bigg\}^n
\end{aligned}$$

Proof. Given all nodes are uniformly distributed and ignoring border effects, the probability that a node lies in the range of another node is $p = \pi r^2/A$. The probability that these two nodes share at least a key is p_s, as given by Equation 2. In general, we have

$$\Pr\{d = k\} = \sum_{i=k}^{N}\binom{N}{i}p^iq^{N-i}\binom{i}{k}p_s^kq_s^{i-k} \qquad (3)$$

The probability that a node is not isolated (i.e. has non-zero degree) is therefore

$$\begin{aligned}
&\Pr\{d \geq 1\} \\
=&1-\Pr\{d=0\} \\
=&1-\left[q^N + \binom{N}{1}pq^{N-1}q_s + ... + p^Nq_s^N\right] \\
=&1-(q+pq_s)^N = 1-(1-pp_s)^N \qquad (4)
\end{aligned}$$

Using Equations 3 and 4, for $k \geq 2$, with detailed derivation in Appendix A, we have

$$\begin{aligned}
&\Pr\{d \geq k\} \\
&\quad = 1-(1-pp_s)^N - \sum_{i=1}^{k-1}\binom{N}{i}p^iq^{N-i}(1-q_s^i) \\
&\quad - \sum_{i=k}^{N}\left[\binom{N}{i}p^iq^{N-i}\sum_{j=1}^{k-1}\binom{i}{j}p_s^jq_s^{i-j}\right] \qquad (5)
\end{aligned}$$

Applying Theorem 1 and the approximation $\Pr\{d_{min} \geq k\} \approx \Pr\{d \geq k\}^n$, we finally have Theorem 3. $\blacksquare$

In the following, we show that Theorem 1 and Theorem 2, although originally formulated for random geometric graphs, are also valid for kryptographs. We start by observing that in the proof of Theorem 2, instead of V_n, Penrose [14] considers a homogeneous *Poisson* point process, $\mathscr{P}_n$, of rate n (i.e., n points per unit space) on the unit cube C. Denote $\mathscr{E}(k,n,r)$ as the expected number of points with degree k in $G(\mathscr{P}_n;r)$; and $v_r(x)$ as the Lebesgue volume of the sphere with radius r centered at coordinates x. Then, out of n points, the average number of points with degree k is

$$\mathscr{E}(k,n,r) = n\int_C e^{-nv_r(x)}\frac{[nv_r(x)]^k}{k!}dx \qquad (6)$$

Theorem 2 is based on the hypothesis that, given $\alpha \in \mathbb{R}$, it is possible to find a sequence $(r_n)_{n\geq 1}$ satisfying Equation 7.

$$\lim_{n\to\infty}\mathscr{E}(k,n,r_n) = e^{-\alpha} \qquad (7)$$

Note that Equation 7 is valid for $G(\mathscr{P}_n;r)$ and not $G(V_n;r)$, but Penrose, using the "de-Poissonization" method [13, Section 6], shows that provided the sequence $(r_n)_{n\geq 1}$ satisfies Equation 7, for $G(V_n;r)$,

$$\lim_{n\to\infty}\Pr\{\rho(V_n;\delta \geq k) \leq r_n\} = e^{e^{-\alpha}} \qquad (8)$$

Equation 8 is used by Penrose to prove Theorem 2.

Here is our strategy. Using Proposition 1, we extend the applicability of Equation 7 to kryptographs. An implication of this proof is that Equation 8 is applicable to kryptographs as well. This in turn implies that Theorem 2, and hence Theorem 1 are applicable to kryptographs.

Proposition 1 If given $\alpha \in \mathbb{R}$, it is possible to find a sequence $(r_n)_{n\geq 1}$ satisfying Equation 7 for $G(V_n; r)$, then for $G(V_n; r, p_s)$, given $\beta \in \mathbb{R}$, it is also possible to find a sequence satisfying

$$\lim_{n\to\infty} \mathscr{E}(k, n, r_n) = e^{-\beta}$$

Proof. Following Penrose, we replace V_n by a Poisson process $\mathscr{P}_n$ of rate n on the unit cube C. Furthermore, for each value of k, we denote the corresponding α by α_k. By the definition of $\mathscr{E}(k, n, r)$,

$$\begin{aligned}\mathscr{E}(k,n,r) &= n\int_C \left(\sum_{i=k}^{N} e^{-nv_r(x)}\frac{[nv_r(x)]^i}{i!}\binom{i}{k}p_s^k q_s^{i-k}\right)dx\\ &= p_s^k q_s^{-k}\sum_{i=k}^{N}\binom{i}{k}q_s^i n\int_C\left(e^{-nv_r(x)}\frac{[nv_r(x)]^i}{i!}\right)dx\end{aligned}$$

Taking limit on both sides,

$$\lim_{n\to\infty}\mathscr{E}(k,n,r) = p_s^k q_s^{-k}\sum_{i=k}^{N}\binom{i}{k}q_s^i e^{-\alpha_i}$$

Since the RHS of the above equation is a positive linear combination of $e^{-\alpha_i}$, there must exist $\beta \in \mathbb{R}$ such that the RHS equals $e^{-\beta}$. ■

3.2. Expected connectivity

The expected connectivity can be expressed as:

$$E[\kappa] = \sum_{k=1}^{N} k\Pr\{\kappa = k\}$$

The problem reduces to determining $\Pr\{\kappa = k\}$, i.e. the probability of getting a graph with connectivity that is *exactly* k. Theorem 2 and Proposition 1 imply that provided $n \to \infty$, as r reaches the $(k+1)$-nearest neighbor distance $\rho(V_n; \delta \geq k+1)$, $G(V_n; r, p_s)$ becomes $(k+1)$-connected. During the time before r reaches $\rho(V_n; \delta \geq k+1)$ and after r leaves $\rho(V_n; \delta \geq k)$, $G(V_n; r, p_s)$ remains k-connected. During this time, there are probably n, $n-1$, ..., 1 nodes with degree k, while the rest of the nodes have a degree of at least $k+1$. Let $x_k = \Pr\{d = k\}$ and $y_{k+1} = \Pr\{d \geq k+1\}$, then

$$\begin{aligned}&\Pr\{\kappa = k\}\\ &\quad= x_k^n + \binom{n}{n-1}x_k^{n-1}y_{k+1} + \ldots + \binom{n}{1}x_k y_{k+1}^{n-1}\\ &\quad= (x_k + y_{k+1})^n - y_{k+1}^n\end{aligned}$$

Therefore,

$$E[\kappa] = \sum_{k=1}^{N} k[(\Pr\{d=k\} + \Pr\{d \geq k\})^n - \Pr\{d \geq k\}^n] \tag{9}$$

4. Simulation results

Our simulations are performed using Mathematica, with the following parameters: $n = 100$, $n' = 20$, $K = 4$ (following Di Pietro et al. [5]), $A = 1$, $r = \sqrt{(n'+1)/(n\pi)}$. Figure 1 to 6 compare the simulated and theoretical survivor functions $\Pr\{\text{connectivity} \geq k\}$ for $P \in \{4, 10, 15, 20, 25, 30\}$. Note that when $P = 4 = K$, every pair of nodes share all keys – this degenerate case is equivalent to using a network-wide key. For each case, 50 network topologies are randomly generated and the connectivities of these topologies are calculated using toroidal distances. On a Pentium D 3 GHz, a computation time of 30-40 minutes is needed for each topology.

Note that the root mean square errors (RMSEs) of the predictions using Theorem 1, and the RMSEs of our predictions using Theorem 3 are mostly in the same order of magnitude. This reinforces the validity of Theorem 3 for kryptographs.

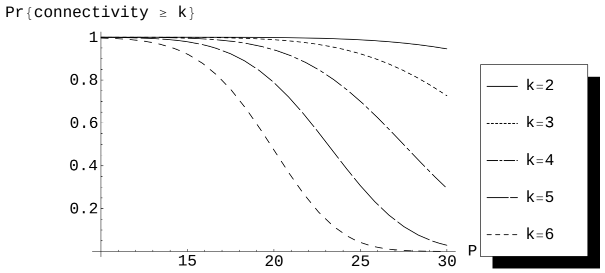

Figure 7. $\Pr\{\text{connectivity} \geq k\}$ **vs** P

Figure 7 plots $\Pr\{\text{connectivity} \geq k\}$ as a function of P. As an example of the usefulness of this plot, say a network is required to have a connectivity of 4, then P should *at most* be 27 where $\Pr\{\text{connectivity} \geq 4\}$ is slightly larger than 0.5. At $P = 27$, the expected connectivity is 3.53 according to Equation 9.

Table 2. Expected connectivities $E[\kappa]$ from simulation and from Equation 9.

	$P =$					
	4	10	15	20	25	30
$E[\kappa]$ (sim.)	11.60	10.48	7.48	5.24	3.66	3.06
$E[\kappa]$ ((9))	11.40	10.07	7.36	5.34	3.96	2.99

Table 2 compares the average connectivities obtained from simulations and the expected connectivities calculated from Equation 9. In rounded figures, these estimations are nearly perfect.

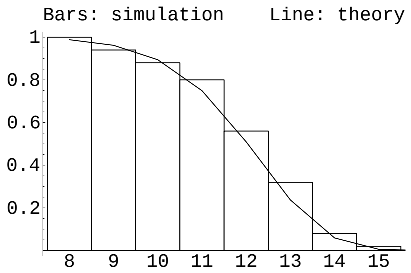

Figure 1. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 4$**. RMSE=0.0411.**

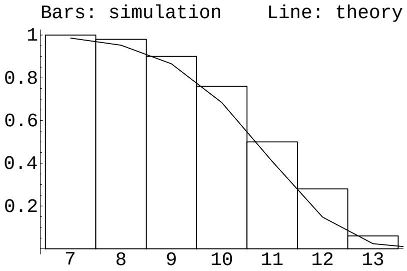

Figure 2. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 10$**. RMSE=0.0707.**

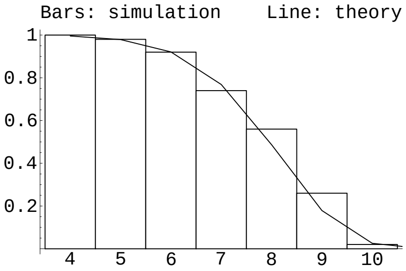

Figure 3. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 15$**. RMSE=0.0427.**

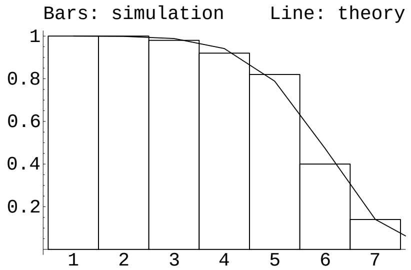

Figure 4. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 20$**. RMSE=0.0340.**

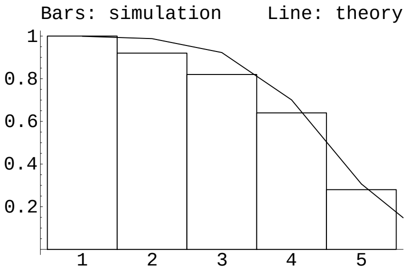

Figure 5. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 25$**. RMSE=0.0626.**

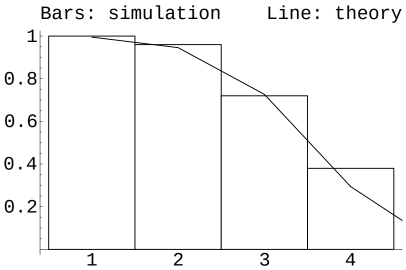

Figure 6. $\Pr\{\text{connectivity} \geq k\}$ **vs** k**, for** $P = 30$**. RMSE=0.0436.**

5. Related work

The study of connectivity, and k-connectivity in particular are founded on recent advances in random graph theory. One of the traditional random graph models is the graph $G(n, p_s)$ (see Section 2). Eschenauer et al. use this traditional model to model WSNs that implement their random key pre-distribution scheme [8]. Chan et al.'s extension to Eschenauer et al.'s scheme [4] is also based on the same traditional model. However, this model does not take into account the distance between vertices, and hence is inadequate.

The first work that uses random geometric graphs to model ad hoc wireless networks is probably due to Huson et al. [10]. Di Pietro et al. [5] avoid the pitfalls of the traditional random graph model and introduce the kryptograph model as defined in Section 2. They prove that as long as the key ring size $K \geq 2$ and key pool size $P = n \ln n$, the probability of a kryptograph being connected is $1 - o(1)$. In one of their examples, $n = 500$, $r = 0.2$, $K = 4$ and $A = 1$, $P = n/(2 \ln n) \approx 40$ guarantees the network to be connected. As a comparison, we derive Figure 8 from Equation 4. The figure clearly confirms Di Pietro et al.'s results [5]. Di Pietro et al. [6] also establish some conditions on which the WSN is "redoubtable". A redoubtable network forces an attacker that captures nodes at random with the aim of compromising a constant fraction of the links to capture at least a constant fraction of the nodes.

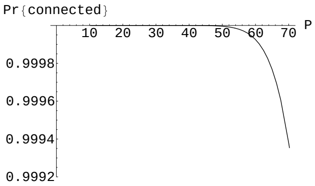

Figure 8. Probability of $G(V_{500}; 0.2, p_s)$ **being connected.** p_s **is related to** P **by Equation 2.**

For traditional random graphs, Łuczak [11] proves that given $k \geq 3$, if $\exists \epsilon$ such that $\epsilon < E[d]$, then the corresponding k-core (subgraphs of $G(n, p_s)$ all of whose vertices have a degree of at least k) is almost surely either empty or k-connected. Theorem 2 by Penrose [14] can be thought of as an analogue of Łuczak's theorem, for random geometric graphs. Bettstetter applies Penrose's result to derive Theorem 1 [1]. In terms of k-connectivity, this paper is the first work done on kryptographs.

Sun et al. propose a metric called average pairwise connectivity (APC), defined as the average connectivity among all pairs of nodes in the network [15]. The problem with this metric is that an APC of a topology might not be the same as the APC of another topology, so it makes more sense to calculate the *average* APC. The real problem is that Sun et al. estimate the APC using the expectation of the upperbound of the connectivity between any node pair, when there is still no theoretical basis for such an estimation.

6. Conclusion

We model WSNs that implement the random key predistribution scheme as “kryptographs”. Based on this definition, we successfully quantify the k-connectivity of kryptographs, by extending some relevant theorems in random geometric graphs to kryptographs. In particular, we derive analytical formulae that describe the asymptotic probability of k-connectivity, as well as the expected k-connectivity of kryptographs. Our theoretical findings are supported by simulation results. A practical application of our results is determining the key pool size that provides, on average, a certain k-connectivity guarantee, for a given key ring size that usually represents the hardware constraint. Finally, we are currently further refining our model.

References

[1] C. Bettstetter. On the minimum node degree and connectivity of a wireless multihop network. In *MobiHoc ’02: Proceedings of the 3rd ACM international symposium on Mobile ad hoc networking & computing*, pages 80–91. ACM Press, 2002.

[2] B. Bollobás. *Random Graphs*. Academic Press Inc., 1985.

[3] M. Capinski and P. E. Kopp. *Measure, Integral and Probability*. Springer-Verlag, 2nd edition, 2004.

[4] H. Chan, A. Perrig, and D. Song. Random key predistribution schemes for sensor networks. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*. IEEE Computer Society, 2003.

[5] R. Di Pietro, L. Mancini, A. Mei, A. Panconesi, and J. Radhakrishnan. Connectivity properties of secure wireless sensor networks. In *2nd ACM workshop on Security of ad hoc and sensor networks*, pages 53–58. ACM Press, 2004.

[6] R. Di Pietro, L. V. Mancini, A. Mei, A. Panconesi, and J. Radhakrishnan. How to design connected sensor networks that are provably secure. In *Proceedings of the 2nd IEEE International Conference on Security and Privacy for Emerging Areas in Communication Networks (SecureComm 2006)*. IEEE Press, 2006.

[7] P. Erdös and A. Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.

[8] L. Eschenauer and V. Gligor. A key-management scheme for distributed sensor networks. In *Proc. 9th ACM conference on Computer and communications security*, pages 41–47. ACM Press, 2002.

[9] D. Ganesan, R. Govindan, S. Shenker, and D. Estrin. Highly resilient, energy efficient multipath routing in wireless sensor networks. *Mobile Computing and Communications Review (MC2R)*, 1(2), 2002.

[10] M. Huson and A. Sen. Broadcast scheduling algorithms for radio networks. In *Military Communications Conference (MILCOM ’95)*, volume 2, pages 647–651. IEEE, 1995.

[11] T. Łuczak. Size and connectivity of the k-core of a random graph. *Discrete Math.*, pages 61–68, 1991.

[12] M. Penrose. *Random Geometric Graphs*. Oxford University Press, 2003.

[13] M. D. Penrose. The longest edge of the random minimal spanning tree. *The Annals of Applied Probability*, 7(2):340–361, May 1997.

[14] M. D. Penrose. On k-connectivity for a geometric random graph. *Random Struct. Algorithms*, 15(2):145–164, 1999.

[15] F. Sun and M. Shayman. On the average pairwise connectivity of wireless multihop networks. In *Global Telecommunications Conference (GLOBECOM ’05)*, volume 3, pages 1762–1766. IEEE, 2005.

Appendix A. Derivation of Equation 5

$$
\begin{aligned}
&\Pr\{d \geq k\} \\
&= \Pr\{d \geq 1\} - \sum_{i=1}^{k-1} \Pr\{d = i\} \\
&= 1 - (1 - pp_s)^N - \binom{N}{1} p q^{N-1} \Big[p_s \Big] \\
&\quad - \binom{N}{2} p^2 q^{N-2} \left[\binom{2}{1} p_s q_s + p_s^2 \right] - \ldots \\
&\quad - \binom{N}{k-1} p^{k-1} q^{N-k+1} \left[\binom{k-1}{1} p_s q_s^{k-2} + \ldots + p_s^{k-1} \right] \\
&\quad - \binom{N}{k} p^k q^{N-k} \left[\binom{k}{1} p_s q_s^{k-1} + \ldots + \binom{k}{k-1} p_s^{k-1} q_s \right] - \ldots \\
&\quad - \binom{N}{N} p^N \left[\binom{N}{1} p_s q_s^{N-1} + \ldots + \binom{N}{k-1} p_s^{k-1} q_s^{N-k+1} \right] \\
&= 1 - (1 - pp_s)^N - \binom{N}{1} p q^{N-1} [1 - q_s] \\
&\quad - \binom{N}{2} p^2 q^{N-2} [1 - q_s^2] - \ldots \\
&\quad - \binom{N}{k-1} p^{k-1} q^{N-k+1} [1 - q_s^{k-1}] \\
&\quad - \binom{N}{k} p^k q^{N-k} \left[\sum_{i=1}^{k-1} \binom{k}{i} p_s^i q_s^{k-i} \right] - \ldots \\
&\quad - \binom{N}{N} p^N \left[\sum_{i=1}^{k-1} \binom{N}{i} p_s^i q_s^{N-i} \right]
\end{aligned}
$$

Simplifying the above gives us Equation 5.