A General Probabilistic Model for Improving Key Assignment in Wireless Networks

Patrick Tague and Radha Poovendran
Network Security Lab (NSL), Department of Electrical Engineering
University of Washington, Seattle, Washington
Email: {tague,radha}@ee.washington.edu

***Abstract*— We study the problem of establishing secure communication channels in resource-constrained wireless networks using key predistribution. Pairwise communication channels between nodes are secured using link keys which are established as a function of cryptographic seeds predistributed to each node. We propose a general model for seed assignment which regulates the number of nodes sharing each seed. In addition, we provide a general model for wireless network connectivity where communication is restricted by both radio range and an independent pairwise relationship. We provide probabilistic analysis for network connectivity and resilience to node capture in terms of our seed assignment and network connectivity models. Finally, we provide a numerical example demonstrating how the proposed approach reduces key wastage while maintaining resilience to node capture of prior results.**

I. Introduction

Applications involving large-scale wireless networks deployed in hostile environments require the development of secure protocols that can operate in a decentralized manner. Due to limitations such as the wireless radio range, battery energy, and computational capability of each node, secure protocols for resource-constrained networks rely on shared symmetric keys.

A promising approach to symmetric key establishment in wireless networks is key predistribution, studied in various forms in many papers (e.g. [1]–[10]). In a key predistribution scheme, *seeds* are assigned to nodes prior to network deployment. For generality, we use the term seed to refer to any secret quantity, such as a key [4], [5], [7], [8], hashed secret [3], [6], or secret share [1], [2], [9], [10], used for key establishment. The network is then randomly deployed, suggesting that the assignment of seeds cannot rely on post-deployment node location and must be tolerant to random placement of nodes. After physical deployment, neighboring nodes must determine if they share a sufficient number of seeds to establish a *link key* for secure communication. Hence, a key predistribution scheme must specify methods for both pre-deployment *seed assignment* and post-deployment *link key establishment*. The link key establishment protocol requires each node to use the individually assigned seeds as inputs, and thus depends on the outcome of seed assignment. Hence, seed assignment plays a crucial role in providing a connected network while maximizing the resilience to captured nodes.

This work is supported in part by the following grants: ONR award, N00014-04-1-0479; ARO grant, W911NF-05-1-0491; and NSA/DOD IASP award.

The most resource efficient method of seed assignment is the assignment of a single master seed to every node in the network. This solution requires minimal storage and minimal communication overhead for key updates. However, the compromise of a single node exposes the master key and compromises the security of the entire network. A solution which prevents compromise due to node capture is the assignment of a unique pairwise key to each pair of nodes. However, this scheme requires storage for $(N-1)$ keys in each of the N nodes and a total of $\binom{N}{2}$ keys. Furthermore, addition of a single node to the network would require $\mathcal{O}(N)$ communication overhead to update every node with an additional pairwise key. Hence, methods of seed assignment for key predistribution schemes exhibit a trade-off between resilience to node capture and resource efficiency.

A. Motivation

The authors of [4] proposed random key predistribution to balance the trade-off between resilience to node capture and storage efficiency. Seed assignment in [4] consists of the assignment of a random selection of K seeds from a pool of P seeds for each node, where P and K are chosen to provide a connected network with a specified probability. The link key establishment protocol in [4] determines if a seed is shared, in which case the shared seed is used directly as the link key. We notice that each node selects a given seed randomly with a probability $\frac{K}{P}$ using this scheme. Thus, the number of nodes which share the given seed is a random variable distributed according to a binomial distribution with parameters $(N, \frac{K}{P})$. Hence, the number of nodes sharing each of the P seeds is highly variable, taking values between 0 and N with an expected value of $\mu = \frac{NK}{P}$. If the number of nodes sharing a seed is much less than μ, the probability that any two of the nodes will be within wireless communication range is very small. Hence, the probability that such a seed will be used to establish a link key is very small. On the other hand, if the number of nodes sharing a seed is much greater than μ, a large number of link keys will be established using the seed. An adversary able to recover such a seed will thus be able to compromise a large number of secure channels throughout the network. Such worst-case *tail-effects* due to assignment of seeds to a number of nodes much greater or much less

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2006**	2. REPORT TYPE	3. DATES COVERED **00-00-2006 to 00-00-2006**
4. TITLE AND SUBTITLE **A General Probabilistic Model for Improving Key Assignment in Wireless Networks**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of Washington,Department of Electrical Engineering,Seattle,WA,98195**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR’S ACRONYM(S)
		11. SPONSOR/MONITOR’S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**9**	

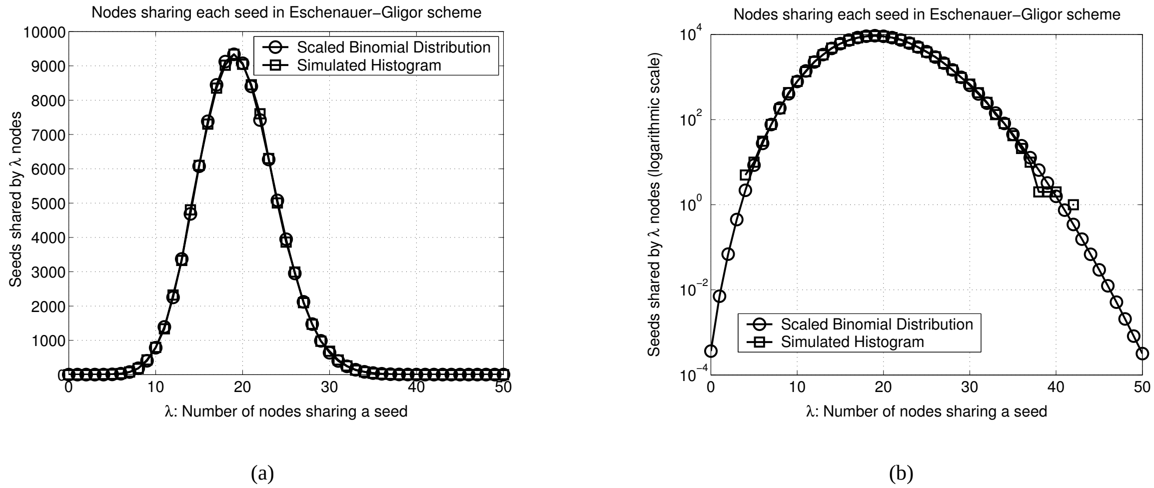

Fig. 1. The simulated histogram of the number of nodes sharing each seed satisfies a binomial distribution. The simulated and theoretical plots are given on (a) linear vertical axis and (b) logarithmic vertical axis.

than μ cannot be analyzed by an average-case probabilistic analysis, such as those provided in [5], [9], [10]. Furthermore, any key predistribution scheme which uses a similar random seed selection method suffers from the same tail-effects of the binomial distribution.

To demonstrate the binomial distribution typical of random key predistribution, we provide Fig. 1 which compares the simulated histogram of the number of nodes sharing each seed with the number of nodes given by the binomial distribution.

B. Contribution

In order to reduce key wastage due to assignment of seeds to a very small number of nodes and minimize the impact of every compromised seed on the remaining network, we propose the regulation of the number of nodes which share each seed. As will be shown, such regulation does not affect the average-case performance in terms of network connectivity or resilience to node capture. However, the occurrence of tail-effects, as discussed in Section I-A, can be reduced.

The contributions of this work are as follows. We propose a general model for seed assignment using discrete probability distributions to regulate the number of nodes which share each seed. The model preserves the average-case performance of existing schemes (e.g. [4], [10]) in terms of network connectivity and resilience to node capture. Furthermore, we propose three seed assignment algorithms for use with the seed assignment model. In addition, we propose a connectivity model for networks in which communication is restricted by radio range and an independent pairwise relationship such as the existence of a shared seed. Finally, we demonstrate the application of our seed assignment and connectivity models and analytically compare the results to previous works.

The paper is organized as follows. We propose our seed assignment model in Section II. In Section III, the appropriate network connectivity model for secure wireless networks is derived. Section IV presents the analysis of network connectivity and resilience to node capture according to the proposed models. Numerical examples and comparison to previous works are presented in Section V.

II. Proposed Seed Assignment Model

We propose a general model for seed assignment which allows for explicit control of the number of nodes which share each seed. The model is given with respect to the following definitions.

A. Model Definitions

The set $S(s)$ of nodes which are assigned the seed s is defined as the *assignment set* of seed s. The number of nodes sharing each seed is regulated by the designation of an *assignment distribution* $\mathcal{P}$ which specifies the probability $\mathcal{P}(\lambda)$ that an assignment set $S(s)$ has size λ, i.e. $\mathcal{P}(\lambda) = Pr[|S(s)| = \lambda]$. The set of values λ with non-zero probability mass is defined as the *support* Λ of the assignment distribution, i.e. $\Lambda = \{\lambda : \mathcal{P}(\lambda) > 0\}$. The average size of an assignment set under an assignment distribution $\mathcal{P}$ is denoted by

$$\mu = \sum_{\lambda \in \Lambda} \lambda \mathcal{P}(\lambda).$$

The given definitions provide the basis for the proposed seed assignment model. However, the seed assignment model further requires analytical elements to allow for the design of assignment distributions which can avoid the tail-effects discussed in Section I-A. Furthermore, given an desirable assignment distribution, the model required a seed assignment algorithm which can realize the given distribution.

B. Assignment Distribution Design

Though the design of an assignment distribution is application-dependent, we provide a brief discussion of some desirable and achievable properties of assignment distributions.

Based on the discussion in Section I that seeds should not be assigned to small or large sets of nodes, the optimal solution is to assign every seed to a fixed number of nodes, i.e. $|\Lambda| = 1$. As stated in [7], this optimal solution is not always achievable. Hence, we consider assignment distributions with $|\Lambda| \geq 1$, including the binomial distribution discussed in Section I-A.

To approximate the optimal solution with $|\Lambda| = 1$, it is highly desirable to specify the set Λ by a contiguous set of integers $\{\lambda \in \mathbb{Z} : \lambda_{min} \leq \lambda \leq \lambda_{max}\}$ such that $|\Lambda|$ is as small as possible. Furthermore, to best approximate the optimal solution, the value of the assignment distribution $\mathcal{P}$ should be larger for the values of λ nearest to μ. As will be discussed in Section II-D, trade-offs exist which further complicate the design problem.

C. Seed Assignment Algorithms

An algorithm which assigns seeds to nodes based on a desired assignment distribution $\mathcal{P}$ must take into account the assumption that each of the N nodes receives exactly K seeds. Assuming the assignment distribution $\mathcal{P}$ is specified, we propose three seed assignment algorithms based on repeated sampling of the assignment distribution $\mathcal{P}$, noting that many such algorithms can be designed. The algorithms presented in this section are the *Random Weighted Seed Selection (RWSS)*, *Random Assignment Set Selection (RASS)*, and *Random Node Partition (RNP)* algorithms. Each of the algorithms is described and presented as code, in which the function $sample(\mathcal{P})$ refers to a sample of the assignment distribution $\mathcal{P}$ and the function $select(n, X)$ refers to a random selection of n items from the set X.

RWSS Algorithm: Let Ψ represent the set of pairs (s, λ) where s is a seed and λ is a sample of the assignment distribution $\mathcal{P}$. Initially, $\Psi = \emptyset$. Since the total number of seed assignments is NK, pairs (s, λ) are generated and added to Ψ until

$$\sum_{(s,\lambda)\in\Psi} \lambda \geq N \cdot K.$$

For each of the N nodes, K pairs (s, λ) with $\lambda > 0$ are selected from Ψ. The seed s for each of the K selected pairs is assigned to the node. Each value λ for the selected pairs is decremented before replacing the pairs in Ψ. The decreasing value of λ will ensure that each seed is assigned only as many times as specified by the sample of $\mathcal{P}$. The algorithm terminates as soon as each of the N nodes has received K seeds. The RWSS algorithm is presented in Fig. 2.

RASS Algorithm: Let $\Omega = \{1, \ldots, N\}$ represent the set of N nodes. For each seed s, a sample $\lambda \in \Lambda$ of the assignment distribution $\mathcal{P}$ is generated. The assignment set $S(s)$ of λ nodes is randomly chosen from Ω, and the seed s is assigned to the nodes in $S(s)$. A node is removed from Ω as soon as it is assigned K seeds. Hence, the algorithm terminates once $|\Omega| = 0$. The RASS algorithm is presented in Fig. 3.

RNP Algorithm: The set $\{1, \ldots, N\}$ is randomly partitioned into subsets $S_1, \ldots, S_T$ such that the subset sizes $|S_i|$ are samples of the assignment distribution $\mathcal{P}$. The nodes in each subset S_i are assigned a common seed, ensuring that each node receives exactly one seed. The set-partition step is repeated a total of K times. To ensure that each $|S_i|$ reflects a sample of $\mathcal{P}$, the final partition subset S_T in each round is combined with a random selection of nodes which are then omitted from the subsequent round. The RNP Algorithm is presented in Fig. 4.

Algorithm: $RWSS(N, K, \mathcal{P})$
$\Psi \leftarrow \emptyset, j \leftarrow 1$
while $\sum_{(s,\lambda)\in\Psi} \lambda < N \cdot K$ **do**
  $\Psi \leftarrow \Psi \cup (s_j, sample(\mathcal{P}))$
  $j \leftarrow j + 1$
end while
for n from 1 to N **do**
  $\Psi_0 \leftarrow \{(s, \lambda) \in \Psi : \lambda > 0\}$
  $E \leftarrow select\,(\Psi_0, \min(K, |\Psi_0|))$
  if $|E| < K$ **then**
    $F \leftarrow select\,(\Psi \setminus E, K - |E|)$
    $E \leftarrow E \cup F$
  end if
  assign $\{s : (s, \lambda) \in E\}$ to n
  $(s, \lambda) \leftarrow (s, \lambda - 1)$ for $(s, \lambda) \in E$
end for

Fig. 2. RWSS Algorithm

Algorithm: $RASS(N, K, \mathcal{P})$
$\Omega \leftarrow \{(1, 0), \ldots, (N, 0)\}$
while $|\Omega| > 0$ **do**
  $\lambda \leftarrow sample(\mathcal{P})$
  $S \leftarrow select\,(\Omega, \min(\lambda, |\Omega|))$
  assign next seed to $\{n : (n, c) \in S\}$
  $(n, c) \leftarrow (n, c + 1)$ for $(n, c) \in S$
  $\Omega \leftarrow \Omega \setminus \{(n, c) \in S : c = K\}$
end while

Fig. 3. RASS Algorithm

D. Finite Sampling Effects

In the three algorithms presented in the previous section, a finite number of samples are taken from the assignment distribution $\mathcal{P}$. As discussed below, each of the three algorithms can result in assignment sets of size $\lambda \notin \Lambda$ near termination of the algorithm. We refer to the occurrence of such sets as *boundary effects*. In what follows, we discuss these boundary effects and how they can be avoided.

Algorithm: $RNP(N, K, \mathcal{P})$
$\Phi_1 \leftarrow \emptyset$
for i from 1 to K **do**
  $\Phi \leftarrow \{1, \ldots, N\} \setminus \Phi_1$
  while $|\Phi| > 0$ **do**
    $\lambda \leftarrow sample(\mathcal{P})$
    $S \leftarrow select\,(\Phi, \min(\lambda, |\Phi|))$
    if $|S| < \lambda$ and $i < K$ **then**
      $\Phi_1 \leftarrow select\,(\{1, \ldots, N\} \setminus S, \lambda - |S|)$
      $S \leftarrow S \cup \Phi_1$
    end if
    assign next seed to nodes in S
    $\Phi \leftarrow \Phi \setminus S$
  end while
end for

Fig. 4. RNP Algorithm

1) RWSS Algorithm: Near the end of the RWSS algorithm, the number of seeds with positive weight may be less than K. In order to avoid very small assignment sets, a random selection of seeds with $\lambda = 0$ can be combined with the remaining seeds to ensure that every node receives K seeds. This technique can result in a small number of assignment sets with size slightly greater than λ_{max} and a small number of assignment sets of size slightly less than λ_{min}.

2) RASS Algorithm: Near the end of the RASS algorithm, the number of nodes with fewer than K seeds may be smaller than the generated λ. Hence, assignment sets of size less than λ_{min} may occur before the algorithm terminates. If desired, these small sets of remaining nodes can be added to previously designated assignment sets, possibly leading to a small number of assignment sets of size greater than λ_{max}.

3) RNP Algorithm: The repeated samples of the assignment distribution $\mathcal{P}$ in each round of the RNP algorithm will often not sum to N. As discussed above, the last subset S_T can be combined with a random selection of nodes which are then omitted from the subsequent round. In the K^{th} round, however, a single assignment set of size less than λ_{min} may occur. The special case where $\Lambda = \{\lambda\}$ and $\frac{N}{\lambda}$ is an integer does not suffer from boundary effects. This special case approximates the use of combinatorial designs [7], [8].

Through extensive simulation, we notice that the occurrence of boundary effects increases as $|\Lambda|$ decreases. Hence, there exists a trade-off between minimizing the size of the support Λ and minimizing the boundary effects which occur. In order to balance this trade-off, however, the analytical properties of seed assignment in terms of the assignment distribution must be investigated. Hence, we are interested in analyzing probabilistic network connectivity and resilience to node capture. The model for network connectivity is presented in the next section, and the analysis based on this model is provided in Section IV. Furthermore, the worst-case resilience to node capture, also analyzed in Section IV, can be considered in the design of an assignment distribution.

III. Network Connectivity Model

In order to design an assignment distribution used for seed assignment, we propose a model for wireless connectivity in which communication is limited both by radio range and an independent (random or deterministic) logical restriction, such as the existence of shared predistributed seeds. We assume that N nodes are deployed uniformly at random with resulting locations $x_u \in \mathcal{A} \subset \mathbb{R}^2$ for $u = 1, \ldots, N$ and each node has an omni-directional antenna with radio range r. Based on the assumptions, we derive the probability $P_G(k)$ that the network is k-connected using graph theory and spatial statistics.

We derive the probability $P_G(k)$ using three graphs: the *physical graph*, *logical graph*, and *network graph*. The physical graph G_P models communication restricted by radio range such that a pair of nodes are adjacent in G_P if and only if they are within radio range. The logical graph G_L models logical relationships resulting from the given relation $\mathcal{R}$ such that a pair of nodes are adjacent in G_L if and only if the relation $\mathcal{R}$ is true. The network graph G is given by the edge-wise intersection of G_P and G_L, appropriately modeling the desired restrictions on the wireless network. The probability $P_G(k)$ is then given by the probability that the graph G is k-connected.

We provide the following results relating to the node degree and the connectivity of the network graph G. The final result given in Theorem 2 provides a probabilistic connectivity model which can be used to provide sufficient parameters for desired network connectivity.

Lemma 1: Given a node u with degree D in the logical graph G_L, the probability $Pr[d_u \geq k]$ that u has degree at least k in the network graph G is given by

$$Pr[d_u \geq k] = 1 - e^{-\rho \frac{D+1}{N} \pi r^2} \sum_{i=0}^{k-1} \frac{(\rho \frac{D+1}{N} \pi r^2)^i}{i!}.$$

Proof: The physical graph G_P can be modeled by a geometric random graph with vertices distributed according to a two-dimensional Poisson point process with rate $\rho = \frac{N}{|\mathcal{A}|}$. Thus, the probability distribution of the number of nodes d within distance r of the node u is given by a Poisson distribution with parameter $\rho \pi r^2$ [11]. Hence, the probability that the number of nodes d is at least a given value k in the physical graph G_P is given by

$$Pr[d \geq k] = 1 - e^{-\rho \pi r^2} \sum_{i=0}^{k-1} \frac{(\rho \pi r^2)^i}{i!}. \tag{1}$$

Given that a vertex u has degree D in G_L, the degree d_u of node u in the network graph G is at least k if and only if at least k of the D neighbors of u in G_L are within distance r of u. Since the neighbors of u in G_L are determined independently of the neighbors of u in G_P, the neighbors of u in G are uniformly distributed in the region $\mathcal{A}$. Thus, the neighbors of u in G_L form a geometric random graph G_P^u, which can be represented by a two-dimensional Poisson point process with rate $\frac{D+1}{|\mathcal{A}|} = \rho \frac{D+1}{N}$. Hence, replacing ρ by $\rho \frac{D+1}{N}$ in (1) completes the proof. ■

Theorem 2: The network graph G resulting from edge-wise intersection of a physical graph G_P and a logical graph G_L with average node degree D is k-connected with probability $P_G(k)$ given by

$$P_G(k) = \left(1 - e^{-\rho \frac{D+1}{N} \pi r^2} \sum_{i=0}^{k-1} \frac{(\rho \frac{D+1}{N} \pi r^2)^i}{i!}\right)^N$$

where $\rho = \frac{N}{|\mathcal{A}|}$ is the node density in the region $\mathcal{A}$.

Proof: The probability that the degree d_u of a node u is at least k in the network graph G is given by Lemma 1. Thus, the minimum node degree $d_{min} = \min\{d_u : u = 1, \ldots, N\}$ is at least k in G with probability given by

$$Pr[d_{min} \geq k] = Pr[d_1 \geq k, \ldots, d_N \geq k]. \tag{2}$$

Due to the properties of the Poisson point process, the probabilities given by Lemma 1 are independent and identically distributed for each node u. Hence, the minimum node degree

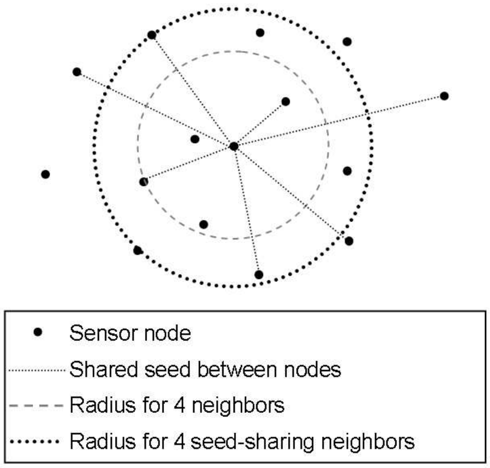

Fig. 5. The required radio range r of each node to guarantee network connectivity increases as the number of nodes D which share seeds with a given node decreases. This illustration compares the case $D = N - 1$ to $D << N$.

d_{min} is at least k with probability $Pr[d_{min} \geq k] = Pr[d_u \geq k]^N$. As r increases, a geometric random graph becomes k-connected, asymptotically, as soon as the minimum vertex degree is k with high probability [12], [13]. Hence, the probability of connectivity is given by $P_G(k) = Pr[d_{min} \geq k]$. ∎

The result of Theorem 2, with fixed values of k, N, and ρ, suggests that as the number of nodes D which share seed with a given node decreases, the radio range r of each node must increase as illustrated by Fig. 5.

IV. Analysis of Seed Assignment

In this section, we provide probabilistic analysis for seed assignment using a given assignment distribution $\mathcal{P}$. We provide a general probabilistic analysis for resilience to node capture. We compute the probability that two nodes share a given seed and the probability that two nodes share exactly i seeds for $i = 1, \ldots, K$. Finally, we compute the probability of network connectivity using Theorem 2.

A. Resilience to Node Capture

The resilience to node capture for a given key predistribution scheme can be measured by computing the fraction of links $f(x)$ which are compromised when x nodes have been randomly captured. However, the means of compromising a link depend on the link key establishment protocol and the type of seeds which are assigned. Hence, for generality, we measure resilience to node capture by deriving an approximation for the probability $p_c(m, x)$ that exactly m of the x captured nodes contain a given seed.

Lemma 3: Given uncaptured nodes u and v which share a seed s such that $\lambda = |S(s)|$ is known, if $x \ll N$ and $m \ll \lambda$, the probability $p_c(m, x, \lambda)$ that exactly m of the x captured nodes contain s can be approximated as

$$p_c(m, x, \lambda) \approx \binom{x}{m} \left(\frac{\lambda - 2}{N - 2}\right)^m \left(\frac{N - \lambda}{N - 2}\right)^{x-m}.$$

Proof: If $x \ll N$ and $m \ll \lambda$, then we can assume that each of the captured nodes contains the seed s independently, and the selection of x out of $(N-2)$ nodes can be modeled as repeated trials of selection with replacement. For each trial, the probability that the selected node contains the seed s given that $(\lambda-2)$ of the $(N-2)$ nodes contain s is $\frac{\lambda-2}{N-2}$. The assumption of independence suggests that each of the x trials can be modeled as an independent Bernoulli random variable. Hence, the probability that m of the x trials are successful is given by a binomial distribution, and the probability $p_c(m, x, \lambda)$ is as desired. ∎

Lemma 3 demonstrates the claims that a seed shared by a large set of nodes leads to a high probability of link key compromise as discussed in Section I. Hence, this lemma can be used to evaluate the worst-case resilience to node capture for a given seed assignment protocol. The following theorem can similarly be used to evaluate the average resilience to node capture.

Theorem 4: Given an assignment distribution $\mathcal{P}$ with mean μ, uncaptured nodes u and v which share a seed s, and x captured nodes, the probability $p_c(m, x)$ that exactly m of the x captured nodes contain s can be approximated as

$$p_c(m, x) \approx \binom{x}{m} \left(\frac{\mu - 2}{N - 2}\right)^m \left(\frac{N - \mu}{N - 2}\right)^{x-m}$$

where μ is the mean of a given assignment distribution $\mathcal{P}$.

Proof: This result is an approximation to the result of Lemma 3 obtained by replacing λ by the mean μ of the assignment distribution $\mathcal{P}$. ∎

B. Probability of Sharing Seeds

We next compute the probability that a given pair of nodes share i of the K assigned seeds, for $i = 1, \ldots, K$. Lemma 5 computes the probability that a given pair of nodes will share a given seed such that the number of nodes sharing the seed is known. Theorem 6 computes the probability that a given pair of nodes will share i of the K assigned seeds such that the number of nodes sharing each seed is known. This theorem can be used to evaluate the worst-case probability of sharing seeds for a given assignment distribution $\mathcal{P}$. Finally, Theorem 7 computes the average probability that a given pair of nodes will share i of the K assigned seeds for a given assignment distribution $\mathcal{P}$.

Lemma 5: A node u containing a seed s, such that $\lambda = |S(s)|$ is known, will share s with a node v with probability $p(s, \lambda) = \frac{\lambda-1}{N-1}$.

Proof: Given a node u containing s, exactly $(\lambda - 1)$ of the remaining $(N-1)$ nodes contain s. Hence, the probability that v is one of these $(\lambda - 1)$ nodes is $\frac{\lambda-1}{N-1}$. ∎

Theorem 6: A node u containing seeds $s_1, \ldots, s_K$, such that $\lambda_j = |S(s_j)|$ for $j = 1, \ldots, K$ are known, will share exactly i seeds with a node v with probability $p_s(i, \lambda_1, \ldots, \lambda_K)$

given by

$$p_s(i,\lambda_1,\ldots,\lambda_K) = \frac{1}{i!(K-i)!}\sum_{\pi}\left(\prod_{j=1}^{i}\frac{\lambda_{\pi_j}-1}{N-1} \times \prod_{j=i+1}^{K}\frac{N-\lambda_{\pi_j}}{N-1}\right)$$

where the summation is over all permutations $\pi = (\pi_1,\ldots,\pi_K)$ of $(1,\ldots,K)$.

Proof: The event that v shares s_j with u can be modeled as a Bernoulli trial with success probability $p(s_j,\lambda_j)$ given by Lemma 5. The probability that i of the K independent events occur is given by the probability that the sum of K independent Bernoulli random variables is equal to i. Since the success probabilities of the K events are not equal, the total probability is summed over all possible choices of i of the K events, represented by the first i entries of a permutation of $(1,\ldots,K)$. For a given permutation $(\pi_1,\ldots,\pi_K)$ of $(1,\ldots,K)$, the contribution to the total probability is the product of $p(s_{\pi_j},\lambda_{\pi_j})$ for $j=1,\ldots,i$ and $1-p(s_{\pi_j},\lambda_{\pi_j})$ for $j=i+1,\ldots,K$. To compensate for permutations which result in the choice of the same i events, the probability is multiplied by $\frac{1}{i!(K-i)!}$. ∎

Theorem 7: A node u will share exactly i seeds with a node v with probability $p_s(i)$ given by

$$p_s(i) = \binom{K}{i}\left(\frac{\mu-1}{N-1}\right)^i\left(\frac{N-\mu}{N-1}\right)^{K-i}$$

where μ is the mean of the assignment distribution $\mathcal{P}$.

Proof: Since the random variables $\lambda_1,\ldots,\lambda_K$ are independent, the probability $p_s(i)$ can be computed by taking the expected value of $p_s(i,\lambda_1,\ldots,\lambda_K)$, as given by Theorem 6, with respect to each of the random variables. Letting the expected value with respect to λ_j be denoted $\mathcal{E}_j[\cdot]$, the probability $p_s(i)$ is given by

$$p_s(i) = \frac{1}{i!(K-i)!}\sum_{\pi}\left(\prod_{j=1}^{i}\frac{\mathcal{E}_{\pi_j}[\lambda_{\pi_j}]-1}{N-1} \times \prod_{j=i+1}^{K}\frac{N-\mathcal{E}_{\pi_j}[\lambda_{\pi_j}]}{N-1}\right). \quad (3)$$

Identical distribution of the λ_j suggests that each $\mathcal{E}_{\pi_j}[\lambda_{\pi_j}]$ is equal to the mean μ of the assignment distribution $\mathcal{P}$. The product terms are thus independent of the index j, and the summands are independent of the permutation π, so the sum-of-products form is replaced by a single product of exponents with coefficient $\frac{K!}{i!(K-i)!} = \binom{K}{i}$. ∎

C. Network Connectivity

Theorem 2 provides the probability of network connectivity as a function of the average node degree D in the logical graph G_L. Assuming the logical graph relation $\mathcal{R}$ is true if and only if the given pair of nodes share at least one seed, we compute the average degree D. We first compute the expected degree $d(u)$ of a node u in G_L given that the sizes λ_j of the K assignment sets corresponding to the seeds assigned to node u are known. This computation is done by mapping the assignment of seeds to node u to a combinatorial occupancy problem. Each pair of nodes (u,v), for $v \neq u$, is represented by a bin, and a shared seed between nodes u and v corresponds to a ball in the bin representing (u,v). Hence, the expected degree $d(u)$ is given by the average number of non-empty bins. The average node degree D in G_L is then computed by taking the expected value of $d(u)$ with respect to the assignment distribution $\mathcal{P}$. The mapping to a combinatorial occupancy problem is illustrated in Fig. 6.

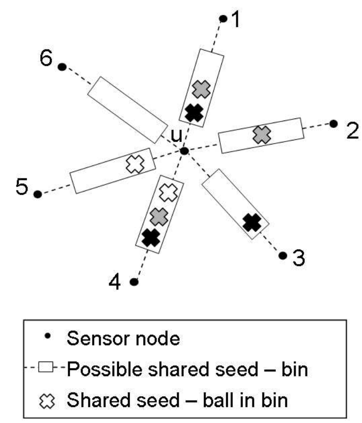

Fig. 6. Seed assignment to nodes in the network is represented by a combinatorial occupancy problem where each pair of nodes (u,v) is represented by a bin, and a shared seed between nodes u and v is indicated by a ball in the bin (u,v).

Lemma 8: Given a node u with seeds $s_1,\ldots,s_K$, such that $\lambda_j = |S(s_j)|$ for $j=1,\ldots,K$ are known, the probability $Pr[e(u)\geq E]$ that the number of nodes $e(u)$ which will not share a seed with u is at least E is given by

$$Pr[e(u)\geq E] = \sum_{m=E}^{N-1}(-1)^{m-E}\binom{m-1}{E-1}\binom{N-1}{m}\prod_{j=1}^{K}\frac{\binom{N-1-m}{\lambda_j-1}}{\binom{N-1}{\lambda_j-1}}.$$

Proof: The event that u and (λ_j-1) other nodes contain a seed s_j corresponds to placing (λ_j-1) balls in the $(N-1)$ bins $(u,v), v\neq u$. If the set of $m\geq E$ bins to remain empty is given, the number of ways to place the (λ_j-1) balls in the $(N-1-m)$ bins is $\binom{N-1-m}{\lambda_j-1}$. Thus, the total number of ways to assign K seeds in such a way that a particular set of $m\geq E$ bins remains empty is $\prod_{j=1}^{K}\binom{N-1-m}{\lambda_j-1}$. The number of ways to select the m bins to remain empty is $\binom{N-1}{m}$. By the Inclusion-Exclusion Principle [14], the number of ways $M(E)$ that K subsets of bins can be chosen such that at least

E bins remain empty is given by

$$M(E) = \sum_{m=E}^{N-1} (-1)^{m-E} \binom{m-1}{E-1} \binom{N-1}{m} \prod_{j=1}^{K} \binom{N-1-m}{\lambda_j - 1}. \quad (4)$$

Dividing $M(E)$ by the total number of ways to choose the K subsets given by $M(0)$ yields the probability that at least E bins remain empty. ■

Theorem 9: Given a node u with seeds $s_1, \ldots, s_K$, such that $\lambda_j = |S(s_j)|$ for $j = 1, \ldots, K$ are known, the expected degree $d(u)$ of u in the logical graph G_L is given by

$$d(u) = (N-1) \left(1 - \prod_{j=1}^{K} \frac{\lambda_j - 1}{N-1}\right).$$

Proof: The expected number of empty bins $\mathcal{E}[e(u)]$ can be computed using the fact that

$$\mathcal{E}[e(u)] = \sum_{E=1}^{N-1} Pr[e(u) \geq E] \quad (5)$$

since $e(u)$ is a non-negative discrete random variable [15]. Substituting the result of Lemma 8 into (5) provides an expression for $\mathcal{E}[e(u)]$. The expected degree $d(u)$ is then given by

$$d(u) = N - 1 - \mathcal{E}[e(u)] \quad (6)$$

because each non-empty bin corresponds to an edge in the graph G_L. Replacing $\mathcal{E}[e(u)]$ with the result from Lemma 8 yields

$$d(u) = N - 1 - \sum_{E=1}^{N-1} \sum_{m=E}^{N-1} (-1)^{m-E} \binom{N-1}{m} \binom{m-1}{E-1} \prod_{j=1}^{K} \frac{\binom{N-1-m}{\lambda_j - 1}}{\binom{N-1}{\lambda_j - 1}}. \quad (7)$$

Reversing the order of summation and appropriately changing the limits of summation yields

$$d(u) = N - 1 - \sum_{m=1}^{N-1} \binom{N-1}{m} \prod_{j=1}^{K} \frac{\binom{N-1-m}{\lambda_j - 1}}{\binom{N-1}{\lambda_j - 1}} \sum_{E=1}^{m} (-1)^{m-E} \binom{m-1}{E-1}. \quad (8)$$

The binomial theorem suggests that

$$\sum_{E=1}^{m} (-1)^{m-E} \binom{m-1}{E-1} = 0^{m-1}. \quad (9)$$

Since $0^0 = 1$, the only non-zero term of the summation is when $m = 1$. Hence the expected degree of node u is given by

$$d(u) = N - 1 - \binom{N-1}{1} \prod_{j=1}^{K} \frac{\binom{N-2}{\lambda_j - 1}}{\binom{N-1}{\lambda_j - 1}} \quad (10)$$

$$= (N-1) \left(1 - \prod_{j=1}^{K} \frac{N - \lambda_j}{N-1}\right). \quad (11)$$

■

Theorem 10: The expected node degree D in the logical graph $G_L(N, \mathcal{R})$ is given by

$$D = (N-1) \left(1 - \left(\frac{N-\mu}{N-1}\right)^K\right).$$

Proof: The expected node degree D is computed by taking the expected value of $d(u)$ given by Theorem 9 with respect to each of the independent random variables λ_j for $j = 1, \ldots, K$, denoted by $\mathcal{E}_j[\cdot]$. The expected node degree D is thus given by

$$D = (N-1) \left(1 - \prod_{j=1}^{K} \frac{N - \mathcal{E}_j[\lambda_j]}{N-1}\right). \quad (12)$$

Identical distribution of the λ_j suggests that $\mathcal{E}_j[\lambda_j]$ can be replaced by the mean μ of the assignment distribution $\mathcal{P}$ completing the proof. ■

Theorem 10 can then be applied directly to the result of Theorem 2 to yield the probability $P_G(k)$ that the secure network is k-connected.

V. Numerical Example and Comparison to Previous Works

In this section, we provide numerical examples demonstrating the use of the proposed seed assignment and network connectivity models. The examples are analytically compared to the existing key predistribution schemes of [4] and [10]. For both examples, we consider a network of $N = 10,000$ nodes deployed over a region $\mathcal{A}$ of area $|\mathcal{A}| = 1\ km^2$. Each node is equipped with a radio of range $r = 40\ m$ and has storage for 200 key-length quantities. Connectivity is guaranteed with probability 0.999.

A. Comparison with Random Key Predistribution [4]

Since each stored quantity in this scheme is a key, $K = 200$. For this example, we assume that no more than 24 nodes are allowed to share a given seed. The given parameters can be applied to Theorem 2 and Theorem 10 to yield a minimum average assignment set size of $\mu \geq 20.5$. Based on design parameters, we select the assignment distribution given by

$$\mathcal{P}(\lambda) = \begin{cases} \frac{\lambda-16}{20}, & \lambda \in \{17, \ldots, 20\} \\ \frac{25-\lambda}{20}, & \lambda \in \{21, \ldots, 24\} \\ 0, & \text{else} \end{cases} \quad (13)$$

which is symmetric over $\mu = 20.5$ and has support $\lambda = \{17, \ldots, 24\}$. Furthermore, we choose the RWSS protocol for illustration, noting that boundary effects may result in assignment sets of size $\lambda \notin \Lambda$ with negligible probability. The desired and simulated assignment distributions are illustrated in Fig. 7. We compare this scheme to that of [4] with the same design parameters. A random subset of $K = 200$ seeds from a set of $P = \left\lfloor \frac{NK}{\mu} \right\rfloor = 97,561$ seeds is independently selected and assigned for each node. Each assigned seed is a cryptographic key used directly as a link key. A link secured using a seed s is compromised as soon as the adversary

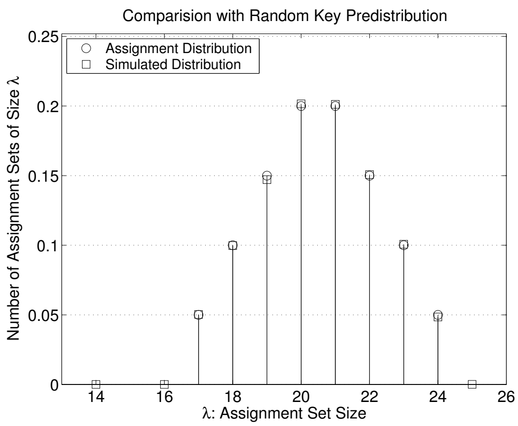

Fig. 7. The plot compares the designed assignment distribution in (13) with the simulated distribution for $N = 10,000$ and $K = 200$. The plotted values of the simulated distribution for λ outside the support $\Lambda = \{17, \ldots, 24\}$ represent the boundary effects due to finite sampling of $\mathcal{P}$.

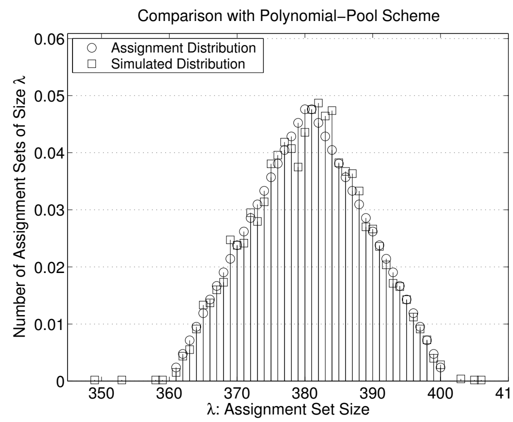

Fig. 8. The plot compares the designed assignment distribution in (16) with the simulated distribution for $N = 10,000$ and $K = 10$. The plotted values of the simulated distribution for λ outside the support $\Lambda = \{361, \ldots, 400\}$ represent the boundary effects due to finite sampling of $\mathcal{P}$.

captures a single node containing s. Hence, the probability of link compromise is approximated by Theorem 4 as

$$f(x) = 1 - p_c(0, x) \approx 1 - \left(\frac{N-\mu}{N-2}\right)^x. \quad (14)$$

The binomial distribution resulting from random key predistribution yields an average assignment set size of $\mu = \frac{NK}{P}$. Thus, the result of (14) is approximately equal to the probability of link compromise

$$f(x) = 1 - \left(1 - \frac{K}{P}\right)^x \quad (15)$$

published in [5]. However, the example given above has the distinct advantage of avoiding the tail-effects discussed in Section I-A. Based on the binomial distribution resulting from the seed assignment protocol of [4], the parameters given in the example yield a probability of $Pr[\lambda \leq 16] = 0.190$ that at most 16 nodes share a seed and a probability of $Pr[\lambda \geq 25] = 0.186$ that 25 or more nodes share a seed. Hence, the use of our protocols eliminates the 19% of seeds shared by fewer nodes than desired and the 18.6% of seeds shared by more nodes than desired, resulting in a more balanced assignment of seeds to nodes.

The worst-case probability of sharing at least one seed can be computed using Theorem 6 as $1 - p_s(0, \lambda_{min}, \ldots, \lambda_{min})$. For the scheme of [4], this probability is 0 because $\lambda_{min} = 1$. However, we compare to the simulated results shown in Fig. 1 where $\lambda_{min} = 4$, yielding a worst-case probability of 0.058. The scheme designed above yields a worst-case probability of 0.274 of sharing at least one seed, demonstrating a significant increase in the worst-case probability. The worst-case resilience is evaluated for $x = 50$ using Lemma 3 as $f(x) = 1 - p_c(0, x, \lambda_{max})$. For the scheme of [4], this probability is 1 because $\lambda_{max} = N$. However, we again compare to the simulated results shown in Fig. 1 where $\lambda_{max} = 42$, yielding a worst-case probability of $f(50) = 0.186$. The scheme designed above yields a worst-case probability of $f(50) = 0.109$, demonstrating a significant decrease in the worst-case probability.

B. Comparison with Polynomial-Pool Scheme [10]

For this scheme, each seed is given by the t coefficients of the corresponding polynomial share, so the number of seeds K and threshold t must satisfy $Kt \leq 200$. Hence, we choose $K = 10$ and $t = 20$. For this example, we assume that no more than 400 nodes are allowed to have shares of the same polynomial. The given parameters can be applied to Theorem 2 and Theorem 10 to yield a minimum average assignment set size of $\mu \geq 380.3$. Based on the design parameters, we select the assignment distribution given by

$$\mathcal{P}(\lambda) = \begin{cases} \frac{\lambda-360}{420}, & \lambda \in \{361, \ldots, 380\} \\ \frac{401-\lambda}{420}, & \lambda \in \{381, \ldots, 400\} \\ 0, & \text{else} \end{cases} \quad (16)$$

which is symmetric over $\mu = 380.5$ and has support $\lambda = \{361, \ldots, 400\}$. Furthermore, we choose the RWSS algorithm for illustration, noting that boundary effects may result in assignment sets of size $\lambda \notin \Lambda$ with negligible probability. The desired and simulated assignment distributions are illustrated in Fig. 8. We compare this scheme to that of [10] with the same design parameters. Shares of a random subset of $K = 10$ polynomials from a set of $P = \left\lfloor \frac{NK}{\mu} \right\rfloor = 262$ polynomials are assigned to each node. A link key is compromised when at least t shares of the common polynomial are recovered from captured nodes. According to Theorem 4, the average case probability that at least t of the x captured nodes contain shares of a given polynomial is

$$\begin{aligned} f(x) &= 1 - \sum_{m=0}^{t-1} p_c(m, x) \\ &\approx 1 - \sum_{m=0}^{t-1} \binom{x}{m} \left(\frac{\mu-2}{N-2}\right)^m \left(\frac{N-\mu}{N-2}\right)^{x-m}. \quad (17) \end{aligned}$$

The binomial distribution resulting from random polynomial selection yields an average assignment set size of $\mu = \frac{NK}{P}$. Thus, the result of (14) is approximately equal to the probability of link compromise

$$f(x) = 1 - \sum_{m=0}^{t-1} \binom{x}{m} \left(\frac{K}{P}\right)^m \left(1 - \frac{K}{P}\right)^{x-m} \quad (18)$$

published in [10]. However, the example given above has the distinct advantage of avoiding the tail-effects discussed in Section I-A. Based on the binomial distribution resulting from the seed assignment protocol of [4], the parameters given in the example yield a probability of $Pr[\lambda \leq 360] = 0.160$ that at most 360 nodes share a seed and a probability of $Pr[\lambda \geq 401] = 0.137$ that 401 or more nodes share a seed. Hence, the use of our protocols eliminates the 16% of seeds shared by fewer nodes than desired and the 13.7% of seeds shared by more nodes than desired, resulting in a more balanced assignment of seeds to nodes.

The worst-case probability of sharing at least one seed can be computed using Theorem 6 as $1 - p_s(0, \lambda_{min}, \ldots, \lambda_{min})$. For the scheme of [10], this probability is 0 because $\lambda_{min} = 1$. We choose to compare using the smallest value of λ such that $\mathcal{P}(\lambda) * P \geq 1$, providing the expected minimum assignment set size in lieu of simulation. This value is given by $\lambda_{min} = 341$ and yields a worst-case probability of 0.293. The scheme designed above yields a worst-case probability of 0.307 of sharing at least one seed, demonstrating a slight increase in the worst-case probability. The worst-case resilience is evaluated for $x = 500$ using Lemma 3 as $f(x) = 1 - p_c(0, x, \lambda_{max})$. For the scheme of [10], this probability is 1 because $\lambda_{max} = N$. In this case, we choose to compare using the largest value of λ such that $\mathcal{P}(\lambda) * P \geq 1$, providing the expected maximum assignment set size in lieu of simulation. This value is given by $\lambda_{max} = 416$ and yields a worst-case probability of $f(500) = 0.594$. The scheme designed above yields a worst-case probability of $f(500) = 0.523$, demonstrating a slight decrease in the worst-case probability.

The major advantage of our seed assignment model for threshold secret-sharing schemes [9], [10] arises in cases where the storage requirement is high enough to allow for a threshold t such that $t > \mu$. In such a case, the assignment distribution can be designed with $\lambda_{max} = t$, leading to a fraction of compromised links equal to 0 for all values of x.

VI. Conclusion

We proposed a general model for seed assignment which regulates the number of nodes sharing each seed using a discrete probability distribution. The proposed model enables the reduction of key wastage while providing the same resilience to node capture as existing schemes. We proposed three seed assignment algorithms based on taking samples of a probability distribution and discussed the boundary effects which result from taking only a finite number of samples. In addition, we proposed a general model for wireless network connectivity in which communication is limited by radio range and restricted by an independent pairwise relationship such as the existence of a shared seed. We analyzed the probabilistic network connectivity and resilience to node capture for seed assignment schemes using the proposed models. Finally, we provided numerical examples to illustrate the use of the seed assignment models as well as a comparison to existing key predistribution schemes.

References

[1] R. Blom, "An optimal class of symmetric key generation systems," in *Advances in Cryptology: Proceedings of EUROCRYPT '84, LNCS 209*. Berlin: Springer-Verlag, 1984, pp. 335–338.

[2] C. Blundo, A. D. Santis, A. Herzberg, S. Kutten, U. Vaccaro, and M. Yung, "Perfectly-secure key distribution for dynamic conferences," in *Advances in Cryptology: Proceedings of CRYPTO '92, LNCS 740*. Berlin: Springer-Verlag, 1993, pp. 471–486.

[3] T. Leighton and S. Micali, "Secret-key agreement without public-key cryptography," in *Advances in Cryptology: Proceedings of CRYPTO '93, LNCS 773*. Springer, 1994, pp. 456–479.

[4] L. Eschenauer and V. D. Gligor, "A key-management scheme for distributed sensor networks," in *Proceedings of the 9th ACM Conference on Computer and Communications Security*. New York: ACM Press, 2002, pp. 41–47.

[5] H. Chan, A. Perrig, and D. Song, "Random key predistribution schemes for sensor networks," in *Proceedings of IEEE Symposium on Security and Privacy*. New York: IEEE, 2003, pp. 197–213.

[6] M. Ramkumar and N. Memon, "An efficient random key pre-distribution scheme," in *Proceedings of IEEE Conference on Global Communications*. New York: IEEE, 2004, pp. 2218–2223.

[7] S. Çamtepe and B. Yener, "Combinatorial design of key distribution mechanisms for distributed sensor networks," in *Proceedings of 9th European Symposium on Research Computer Security, LNCS 3193*. Berlin: Springer-Verlag, 2004, pp. 293–308.

[8] J. Lee and D. Stinson, "Deterministic key predistribution schemes for distributed sensor networks," in *Selected Areas in Cryptography, LNCS 3357*. Springer, 2004, pp. 294–307.

[9] W. Du, J. Deng, Y. Han, and P. Varshney, "A pairwise key pre-distribution scheme for wireless sensor networks," in *Proceedings of the 10th ACM Conference on Computer and Communications Security*. New York: ACM Press, 2003, pp. 42–51.

[10] D. Liu, P. Ning, and R. Li, "Establishing pairwise keys in distributed sensor networks," *ACM Transactions on Information and System Security*, vol. 8, no. 1, pp. 41–77, February 2005.

[11] N. Cressie, *Statistics for Spatial Data*. New York: John Wiley & Sons, Inc., 1993.

[12] M. Penrose, "On k-connectivity for a geometric random graph," *Wiley Random Structures and Algorithms*, vol. 15, no. 2, pp. 145–164, 1999.

[13] C. Bettstetter, "On the minimum node degree and connectivity of a wireless multihop network," in *Proceedings of the 3rd ACM International Symposium on Mobile Ad Hoc Networking & Computing*. New York: ACM Press, 2002, pp. 80–91.

[14] B. R. Johnson, "An elementary proof of inclusion-exclusion formulas," *The American Mathematical Monthly*, vol. 87, no. 9, pp. 750–751, November 1980.

[15] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York: John Wiley & Sons, Inc., 1957, vol. 1.