How to Design Connected Sensor Networks that Are Provably Secure

ROBERTO DI PIETRO, LUIGI V. MANCINI, ALESSANDRO MEI,
ALESSANDRO PANCONESI
University of Rome "La Sapienza", Italy

JAIKUMAR RADHAKRISHNAN
Tata Institute of Fundamental Research, Bombay, India

***Abstract*— We give, for the first time, a precise mathematical analysis of the connectivity and security properties of sensor networks that make use of the random pre-distribution of keys. We also show how to set the parameters– pool and key-ring size– in such a way that the network is not only connected with high probability via secure links, but also provably secure, in the following sense: we formally show that any attacker that captures sensors at random with the aim of compromising a constant fraction of the secure links, must capture at least a constant fraction of the nodes of the network. In the context of wireless sensor networks where random pre-distribution of keys is employed, we are the first to provide a mathematically precise proof, with a clear indication of parameter choice, that two crucial properties – connectivity via secure links and resilience against malicious attacks– can be obtained simultaneously. Our theoretical results are complemented by extensive simulations that reinforce our main conclusions.**

***Index Terms*— Wireless security, sensor networks.**

I. Introduction

A Wireless Sensor Network (WSN) is a collection of sensors whose size can range from a few hundred sensors to a few hundred thousand or possibly more. The sensors do not rely on any pre-deployed network architecture, thus they communicate via an ad-hoc wireless network. The power supply of each individual sensor is provided by a battery, whose consumption for both communication and computation activities must be optimized. Distributed in irregular patterns across remote and often hostile environments, sensors should autonomously aggregate into collaborative, peer-to-peer networks. Sensor networks must be robust and survivable in order to overcome individual sensor failure and intermittent connectivity (due, for instance, to a noisy channel or a shadow zone).

It is widely believed that WSNs can be useful in many diverse settings [1]. In many applications establishing secure pair-wise communications is very important and, in some cases, critical. In particular, it is a pre-requisite for the implementation of secure routing, and can be useful for secure group communications as well. However, due to the scarceness of resources, public key cryptography may not be a viable solution. In this case confidentiality has to be enforced by using symmetric key algorithms [17]. Key management is thus a central issue in secure wireless sensor networks. One of the most promising approaches is the so-called *random pre-distribution of keys* introduced in [11]. This model is the object of study of this paper.

We begin by describing the model introduced in [11]. A *Secure Wireless Sensor Network* (SWSN) is composed of N sensors. Each sensor is pre-assigned a *key ring* of k secret keys randomly drawn from a common pool of K random keys. The sensors are then randomly deployed in a given geographical area. Two sensors share a secure communication link if they lie within communication range and they share a common pre-assigned key. A first fundamental problem in secure wireless sensor networks is to choose proper k and K such that the network is connected by using secure links alone. This problem is addressed by Eschenauer and Gligor in [11]. Their basic idea is that a SWSN can be considered to be a random graph in the sense of Erdös and Rényi [10]. According to this well-known model, a *random graph* of N vertices and parameter p is defined as follows: for every pair of vertices u and v the edge uv is inserted with probability p, by flipping a coin. Crucially, for every potential edge a new coin flip, independent of the previous ones, is performed. Therefore edges exist independently of each other.

Notice that a SWSN is generated by a completely different random process and it is not clear that this process can be approximated and if so, to what extent, by a random graph in the sense of Erdös and Rényi. To appreciate the problem, consider for instance the following situation. There are three sensors x,y and z, all within transmission range, whose key rings are of size 2, and suppose that the pool size is $K = 10^4$. Assume that we know already that edges xy and yz exist. What is the probability that edge xz also exists? If we assume independence, following Erdös and Rényi, then this probability is $\sim \frac{1}{5000}$, but in reality $\Pr[xz \text{ exists} \mid \text{both } xy \text{ and } yz \text{ exist}] \sim \frac{1}{2}$. This choice of parameters is made for the sake of clarity: the problem is present in every practical situation.

In fact, random graphs and SWSNs have different

structural properties as illustrated in Figure 1, that reports the clustering coefficients and the number of triangles for these two kinds of graphs. The clustering coefficient of a vertex u is the number of links between neighbours of u divided by the number of all such potential links, i.e. $\binom{\deg(u)}{2}$. The clustering coefficient of a graph is the average clustering coefficient, taken over all vertices. This quantity was introduced in the seminal paper by Watts and Strogatz in their study of small-world networks, and is a useful structural parameter of a network [20] (see also [15] and references therein). All this clearly shows a discrepancy between the classical Erdös-Renyi model proposed in [11] and the real networks generated by random pre-distribution of keys.

There are other difficulties. The Erdös-Renyi model assumes full-visibility– any two devices can be connected by a direct link regardless of their geographical position. There is no guarantee that the Erdös-Renyi Theorem as used in [11] ensures high probability of connectivity in the general case, when devices are not within transmission range.

In this paper, we present a precise mathematical analysis of SWSNs. We remark that having a precise understanding of a model is always important, but especially so when security is at stake. Note a very important point in this respect. If we keep K fixed and let k grow, the probability of connectivity increases but the security of the network decreases. Intuitively, if key-ring size is large then capturing just a few sensors is likely to be enough to reconstruct the entire pool and compromise the whole network. Viceversa, if k is very small when compared to K, every key will be used by very few links, which is good for security, but the network is likely to be disconnected. Thus, security and connectivity are in conflict. It is a very important problem whether there exists a choice of key-ring size and pool size that ensures both. In this paper we show that the answer is yes.

Concerning connectivity, we show that if $K \geq N$ and

$$\frac{k^2}{K} \sim \frac{\log N}{N} \tag{1}$$

then the network is connected with high probability. (In fact, the condition $K \geq N$ can be replaced by a weaker but more technical condition.) The ratio $\frac{k^2}{K}$ has a meaning. When k is small when compared to K, which is not only usually the case in the applications but also highly desirable for security reasons, the ratio is roughly equal to the probability that a link exists between a given pair of nodes since

$$\Pr[\text{link exists}] = 1 - \left(1 - \frac{k}{K}\right)^k \sim \frac{k^2}{K}.$$

When $K \geq N$, condition 1 is in some sense optimal. As we prove in this paper, if the ratio $\frac{k^2}{K}$ is smaller the network is likely to be disconnected. (If the ratio is larger the probability of connectivity can only improve.)

Our results concerning connectivity holds both in the full-visibility case (i.e. every two nodes are within transmission range of each other) and in the general case.

Let us now turn to security. A first observation is that that the kind of stochastic dependency exhibited by the model that was described earlier is bad for security. Assume that an attacker is able to collect a subset of the sensors and extract their keys. Clearly, all links incident on captured nodes will be compromised but can the damage extend to other links? The example above shows that once a node u is captured not only are the links incident to u compromised, but it is likely that many links between neighbours of u are also compromised. In general, the stochastic dependencies of the model can give raise to unexpected correlations that an adversary can exploit and must therefore be carefully analysed. If we make use of the Erdös-Renyi model this crucial aspect is completely overlooked.

It is not hard to give examples where the adversary can compromise the entire network just by capturing a sublinear fraction of the vertices. The relevant question is whether there exists a choice of the relevant parameters– key-ring size and pool size– such that, in essence, the damage is limited to the edges incident on the captured nodes. In this paper we show that such a choice of the parameters exists. Crucially, the same choice also ensures connectivity. The following definition embodies the notion of security just described. We say that a network is *redoubtable* if the following holds: any attacker that captures sensors at random with the aim of compromising a constant fraction of the links, must capture at least a constant fraction of the nodes. The random attacker is the one commonly considered in the literature.

We formally prove that, if $K \geq N \log N$ and k is chosen to satisfy (1), the network is not only connected with high probability, but it is also redoubtable. For instance, we can choose $k \sim \log N$ and $K \sim N \log N$. To the best of our knowledge, this is the first asymptotic bound on the resilience of wireless sensor networks using random pre-distribution of keys. Note that results of this kind cannot even be formulated in the Erdös-Renyi model.

To summarize, we show by a precise mathematical analysis how to design sensor networks that are at the same time connected (with high probability) and provably secure. Thus, our results put on a firm theoretical foundation the large body of experimental work that followed the original paper of Eschenauer and Gligor, and lay the foundations for the rigorous investigation of security properties.

Lastly, we complement the above mentioned theoretical results with extensive simulations. The experiments support the conclusion that our design guidelines guarantee both connectivity and resilience for a wide interval of practical network sizes and communication ranges.

II. Related Work

The idea of probabilistic key sharing for WSNs is introduced by Eschenauer and Gligor [11]. The authors also provide a simple and centralized algorithm for re-keying in a distributed WSN. Later, in [6], three mechanisms are described in the framework of random key pre-distribution. First of all, the *q-composite* random key pre-distribution scheme, a modification of the basic scheme in [11], achieves better security under small scale attack while trading off increased vulnerability in the face of a large scale physical attack on the network sensors. Secondly, the multi-path key reinforcement protocol substantially increases the security of the channel by leveraging the security of other links. Lastly, the random-pairwise keys scheme assigns private pairwise keys to randomly selected pairs of sensors so as to guarantee that the rest of the network remains fully secure even when some of the sensors have been compromised. Moreover, this latter scheme supports node to node authentication.

Two schemes build up a secure pairwise channel which combine a deterministic technique with a pre-distribution random scheme. The first scheme is proposed in [9]. The authors use a deterministic protocol proposed by Blom [3] that allows any pair of nodes in a network to find a pairwise secret key. As a salient feature, Blom's scheme guarantees a so called λ*-secure property*: as long as no more than λ nodes are compromised, the network is perfectly secure. A λ-secure data structure built this way is called a key space. The authors in [9] create a set $\mathscr{W}$ composed of ω key spaces, and randomly assign up to τ spaces per sensor. Two nodes can find a common secret key if they have picked a common key space. The second scheme is proposed in [14]. In principle, this work is similar to [9], where Blundo et al's polynomial scheme [4] is used instead of Blom's.

Pairwise secure channel establishment is a key requirement in order to perform in-networking processing in a framework of confidentiality [21]. However, note that to support pairwise key establishment, neighboring sensors have to discover the keys they possibly share. To this aim, the solution in [7],[8] address the issue of key discovery in a very efficient way, trading off communications with local computations, while not weakening the overall security of the established links. Note that these solutions can be applied to the model we propose in this paper.

Connectivity properties have been studied for non-secure wireless sensor networks as well. In [2], a geometric random model has been used to investigate minimum node degree and h-connectivity. Using a recent asymptotic result from Penrose [16], Bettstetter experimentally shows how to compute a communication range r such that, for a given number of nodes and a given integer h, the network is guaranteed to be h-connected. Equivalently, it is possible to compute how many sensors are needed to cover a given geographical area with an h-connected network.

In 1945, E Marczewski (see [13]) considered graphs where sets where associated with vertices and two vertices were connected if their associated sets had an element in common. Recently, graphs obtained by choosing the sets randomly have been investigated [13], [12], [18], [19]. In these works, the sets associated with the vertices are usually large. For a certain choice of parameters this model of random graphs is shown to be similar to the $G(n,p)$ model of Erdős-Rényi. However, for the range of parameters of interest to us these results are not applicable.

III. Preliminaries

We say that $f(n) = o(1)$ if $f(n)$ goes to zero as n goes to infinity. If an event (depending on n) happens with probability $1-o(1)$, we say that it occurs with *high probability* or *almost surely*.

Fact 3.1: (Union Bound) Let $E_1,\ldots,E_m$ be m events. Then,

$$\Pr\left[\bigcup_{i=1}^{m} E_i\right] \leq \sum_{i=1}^{m} \Pr[E_i].$$

We recall some basic facts and definitions from graph theory (see for instance [5]). As customary $V(G)$ and $E(G)$ denote the vertex and the edge set of a graph G, respectively. Given a graph $G=(V,E)$ a *cut* is a proper subset $S \subseteq V$ such that there is no edge connecting a vertex in S with a vertex in $V-S$.

Fact 3.2: A graph G is connected if and only if it has no cuts.

The terms point, node and vertex will be used interchangeably.

IV. Connectivity of Secure Wireless Sensor Networks

The following definition captures exactly the kind of networks that are generated with randon pre-distribution of keys.

Definition 4.1: Let K be the size of a finite set of keys (the *pool*), and let $k \leq K$ be a fixed parameter. Let $[K] = \{1,2,\ldots,K\}$ be the index set of the keys in the common pool of size K. The graph $G^N_{r,k,K}$ is defined as the geometric random graph obtained by the following procedure:

- First, each node u is assigned a subset of keys, its *key ring*, whose indexes are in $K_u \subseteq [K]$ by sampling $[K]$ without replacement k times.
- Second, the N nodes are distributed uniformly at random on the given square geographical area, that, without loss of generality, we assume to be of side one (called the *unit square*).
- Third, uv is an edge if (a) the two nodes are within distance r; and (b) $K_u \cap K_v \neq \emptyset$.

The resulting graph $G^N_{r,k,K}$ is called a *kryptograph* with parameters r, k, K and N. In the special case in which

every two nodes are within transmission range, the so-called *full visibility* case, the resulting graph is denoted as $G^N_{k,K}$.
In the sequel, for sake of simplicity we shall identify $[K]$ with the set of keys and K_u with the key ring of a vertex u.

Note that all links of $G^N_{r,k,K}$ are secure by definition (edge uv exists only if vertices u and v share at least one key). Therefore if the kryptograph is connected it is so via secure links alone.

In the proof of connectivity we will assume that the keyrings are generated by sampling with replacement. This simplifies the analysis of connectivity without loss of generality. In fact, sampling without replacement can only be better, as it can be seen by the following coupling argument. Suppose each node picks a set of size k in the following way. It first picks a set by sampling with replacement k times. Now, if did not pick k distinct elements it picks whatever more is needed by sampling without replacement. So, in the end it has a set of size exactly k, and the distribution of this key ring is uniform. Thus, we can always assume that key rings sampled without replacement were generated by this process, but the key rings we consider in the proofs (i.e. the first k samples) are actually subsets of the actual sets the nodes hold. So, if there is connectivity using sampling with replacement there must be connectivity using sampling without replacement.

We now proceed to establish almost sure connectivity in the full visibility case under the following three assumptions on the parameters k and K. The first is,

$$\frac{k^2}{K} = c\frac{\log N}{N} \tag{2}$$

where $c > 8$ is a constant. The term k^2/K is (very nearly) the probability that a secure link exists between two given endpoints. It can be shown that if $k^2/K = o(\log N/N)$ then the graph is disconnected with positive probability. Thus the condition $k^2/K = \Omega(\log N/N)$ is necessary to establish that the graph is connected almost surely. If we establish connectivity under condition (2), the result will follow immediately for higher values k^2/K (by an easy coupling argument). The second assumption is

$$K \geq N. \tag{3}$$

In what follows, the weaker, but uglier condition

$$\binom{N}{s} \leq \binom{K}{ks/4}$$

where s is the size of a vertex set, would do. We keep condition (3) because it is cleaner and always satisfied in practice. The last condition is

$$k \geq 5 \tag{4}$$

which, again, is easily satisfied in practice. To express our results in a parameterized fashion we shall define

$$k := 2\alpha \tag{5}$$

where $\alpha \geq \frac{5}{2}$. The probability of connectivity will depend on the constants c and α. As a rule of thumb the larger c and α the higher the probability of connectivity. In practical application, these conditions on the parameters are easily seen to hold.

Definition 4.2: Let S be a set of vertices, and let $k(x)$ be the set of keys chosen by vertex x. We define

$$k(S) := \bigcup_{x \in S} k(x).$$

When the size of a set of vertices S is "small" the expected size of $k(S)$ is (roughly) sk. The next lemma says that it is unlikely that we deviate far below the expectation.

Lemma 4.3: Assume conditions (2)–(4). Let $\mathscr{S}$ be the collection of non-empty sets of vertices of size at most $\min\{K/k, N/2\}$. Then

$$\Pr[\exists S \in \mathscr{S},\ |k(S)| \leq |S|k/4] \leq N\left(\frac{ec\log N}{N}\right)^{\alpha}.$$

Proof: By the union bound

$$\Pr[\exists S \in \mathscr{S},\ |k(S)| \leq |S|k/4] \leq \\ \leq \sum_{s \leq N/2} \Pr[\exists S,\ |S| = s,\ |k(S)| \leq sk/4].$$

To estimate $\Pr[\exists S,\ |S| = s,\ |k(S)| \leq sk/4]$, we first choose a set $S \subseteq V$ of size s. There are

$$\binom{N}{s}$$

many ways to do this. We then fix a set $T \subseteq [K]$ of keys of size $sk/4$. This can be done in

$$\binom{K}{ks/4}$$

different ways. Finally, we need to compute the probability that $k(S)$ is included in the set T. This probability is equal to

$$\left(\frac{ks}{4K}\right)^{ks}.$$

Therefore, recalling that $N \leq K$ and the basic inequality

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$$

we have

$$\begin{aligned}\Pr[\exists S,\ |S|=s,\ \ |k(S)|\le sk/4]\le\\ &\le\binom{N}{s}\binom{K}{ks/4}\left(\frac{ks}{4K}\right)^{ks}\\ &\le\binom{K}{s}\binom{K}{ks/4}\left(\frac{ks}{4K}\right)^{ks}\\ &\le\binom{K}{ks/4}^2\left(\frac{ks}{4K}\right)^{ks}\\ &\le\left(\frac{4eK}{ks}\right)^{ks/2}\left(\frac{ks}{4K}\right)^{ks}\\ &=e^{ks/2}\left(\frac{ks}{4K}\right)^{ks/2}\\ &=\left(\frac{eks}{4K}\right)^{ks/2}\end{aligned}$$

Let

$$p(s):=\left(\frac{eks}{4K}\right)^{ks/2}.$$

To compute the maximum value of this quantity as s varies, write it as

$$(z^z)^t$$

with

$$z:=\frac{eks}{4K}$$

and $t=2K/e$. Note that $z\le 1$ and that t does not depend on s. The function z^z is monotone decreasing in the range $z\le 1$. So, the maximum is achieved for the value of s, where z is as small as possible, that is $s=1$. The value of $p(s)$ at $s=1$ is

$$p(1)=\left(\frac{ek}{4K}\right)^{k/2}\le\left(\frac{ec\log N}{N}\right)^{\alpha}\qquad(6)$$

■

By fixing the parameters c and α we can bound $\Pr[\exists S\in\mathscr{S},\ |k(S)|\le|S|k/4]$ by any inverse polynomial (for large enough N), i.e. N^{-t} for any fixed t. In the following corollary we commit to a particular choice of the parameters for the rest of the section.

Corollary 4.4: Assume conditions (2)–(4). Let $\mathscr{S}$ be the collection of non-empty sets of vertices of size at most $\min\{K/k,N/2\}$. Then, for N large enough,

$$\Pr[\exists S\in\mathscr{S},\ |k(S)|\le|S|k/4]\le\frac{1}{N}.$$

Proof: If $\alpha\ge\frac{5}{2}$ (i.e. $k\ge 5$) the quantity

$$\left(\frac{ec\log N}{N}\right)^{\alpha}$$

is (much) smaller than $1/N^2$, for N large enough. ■

Lemma 4.5: Let S be a proper set of vertices and let $x\in V-S$. If $k(x)\cap k(S)\neq\emptyset$ then S is not a cut.

Proof: Let $a\in k(x)\cap k(S)$. By definition, $k(S)=\bigcup_{y\in S}k(y)$. Therefore there exists $z\in S$ such that $a\in k(z)$. But then z and x have a key in common, and therefore $xz\in E$. ■

The next theorem uses the previous two lemmas to establish that the graph is connected almost surely. The basic strategy is to prove that almost surely no set of vertices is a cut . Lemma 4.3 says that, for all "small" sets S simultaneously, the set of keys $k(S)$ is "large" almost surely. Technically, this is the difficult claim to establish, since if S is "big" the odds that $k(S)$ is "large" are easily seen to be overwhelming (this is proven in the next theorem). Therefore, for all sets S simultaneously, $k(S)$ is "large" with high probability. If we consider now the sets of vertices of type $V-S$ we see in the following proof that almost surely there is a vertex $x\in V-S$ such that $k(x)$ intersects both $k(S)$ and its $k(V-S)$. But this, by Lemma 4.5, implies that S is not a cut.

Theorem 4.6: Assume conditions (2)–(4). Then, almost surely, the graph is connected.

Proof: We will show that, almost surely, there is no cut in the graph. Given any non-trivial $S\subseteq V$, either S or $V-S$ has size at most $N/2$. Therefore without loss of generality we can assume that $|S|\le N/2$. Let $\mathscr{E}$ be the event: for all sets of vertices of size at most K/k, the size of $k(S)$ is at least $|S|k/4$. By Corollary 4.4 this event happens almost surely. We want to estimate the probability that $k(V-S)$ does not intersect $k(S)$. Let s be fixed. Recalling the basic inequality $1-x\le e^{-x}$, by the union bound we have,

$$\begin{aligned}\Pr[\exists S,|S|=s,S\text{ is a cut}\mid\mathscr{E}]=\\ &=\Pr[\exists S,|S|=s,k(S)\cap k(V-S)=\emptyset\mid\mathscr{E}]\le\\ &\le\binom{N}{s}\left(1-\frac{sk}{4K}\right)^{k(N-s)}\\ &\le N^s\exp(-sk^2(N-s)/4K)\\ &\le N^s\exp(-sk^2N/8K)\\ &=N^{-(c/8-1)s}\end{aligned}$$

where the last inequality follows from assumption (2). By summing over all sizes s and recalling that $c>16$, we have that the probability of having a cut in the graph is at most

$$\sum_{s=1}^{N/2}N^{-(c/8-1)s}\le\sum_{s=1}^{\infty}N^{-s}\sim\frac{1}{N}.$$

If $K/k\ge N/2$, then we have already covered all cases. Otherwise, we have to consider the sets of vertices of size s in the range $K/k\le s\le N/2$. There are at most 2^N such sets. Assuming $\mathscr{E}$, each such set has a key ring of size at least $K/4$ (by restricting attention to a subset of size K/k). So, there exists an $\varepsilon>0$ such that the probability that one such set of vertices is a cut is at most, if $k\ge 5$,

$$2^N\left(1-\frac{1}{4}\right)^{kN/2}\le 2^{-\varepsilon N}.$$

To sum up:

$$\begin{aligned}\Pr[\exists S,S\text{ is a cut}]&\le\Pr[\exists S,S\text{ is a cut}\mid\mathscr{E}]+\Pr[\mathscr{E}^c]\\ &\le\frac{1}{N}+2^{-\varepsilon N}+\frac{1}{N}\ll\frac{3}{N}\end{aligned}$$

for N large enough. The claim follows. ■

The proof of connectivity shows that the probability that the graph is not connected goes to zero as N grows. How quickly depends on the parameters c and α. In the experimental section we will show that for the kind of parameters that reflect practical usage the probability of connectivity is overwhelming. Here we try to get a feeling directly from the formulae. By analysing the proofs one can see that the probability that the graph is disconnected is at most

$$p := \left(\frac{ek}{4K}\right)^{\alpha} + 2N^{-(c/8-1)} + 2^{N}\left(\frac{3}{4}\right)^{kN/2}$$

Assume $N = 2^8 = 256$, $K = 2^{14} = 16384$, and that each sensor is given $k = 2^7 = 128$ keys, which implies $\alpha = 64$. With this choice we get $c = 32$. Assuming furthermore for the sake of simplicity that logarithms are to the base 2, we have that

$$p \approx 2^{-23}.$$

Essentially the same results can be proven for the general, and more practical case, when visibility is not full. Calculations are omitted for the sake of brevity.

V. Secure Wireless Sensor Networks Are Provably Secure

While connectivity is a fundamental property of SWSNs (we cannot really call it a network if it is disconnected), also fundamental is it to understand how resilient is a SWSN against external attacks.

We want to model the following attack: an external entity tampers with the sensors and collects all the keys in the keyrings. The sensors to tamper with are chosen randomly among the ones yet to be compromised. The attacker's goal is to collect enough keys to decrypt as many network communications as possible. This is the kind of an attacker that is considered in the large majority of works on security for wireless sensor networks in the literature. An equivalent attack can be carried on by a subset of the sensors in the network that are actually malicious and cooperate to subvert the communication confidentiality by using all their knowledge and keys.

Definition 5.1: A *collusion* in a secure wireless sensor network is a subset of the network sensors. Given a collusion, a key of the pool is compromised if it belongs to the keyring of some sensor w in the collusion. We also say that, given two sensors u and v, secure link uv is compromised if and only if u and v share some key (that is, the secure link exists) and all the shared keys compromised.

When random key pre-distribution is used, the attacker, by compromising a sensor w, not only does compromise all the communication links from sensor w, also compromises a number of other links in the network, those using the same compromised keys. This is a weakness. So, it is possible that the attacker takes control over a constant fraction of the network by compromising a sub-linear number of the sensors. When this is *not* possible, we say that the network is redoubtable, that is, essentially secure against massive attacks.

Definition 5.2: A Secure Wireless Sensor Network is *redoubtable* if the probability that a collusion of $o(N)$ nodes uniformly chosen at random in the network can compromise a constant fraction of the network links is zero.

Theorem 5.3: If a secure wireless sensor network is built in such a way that

$$\frac{k}{K} \sim \frac{1}{N}, \tag{7}$$

then the network is redoubtable.

Proof: Let a collusion of $c = o(N)$ nodes be uniformly chosen at random in the network. Assume that the collusion can compromise a constant fraction of the network links with probability ε. Since every sensor is assigned k of keys, the collusion as a whole has a collection of $o(kN) = o(K)$ compromised keys at most. Consider a secure link uv in the network. Clearly, $K_u \cap K_v$ contains at least one key. Therefore, link uv is compromised with probability $o(1)$. As a consequence, at most a fraction of $o(1)$ edges are compromised, on average. But this is impossible if $\varepsilon > 0$, so, it must be $\varepsilon = 0$. ■

This result shows that secure wireless sensor networks can be designed in such a way to be provably secure. One of the key results of our work is to show that there exists a way to choose the parameters k and K such that the network is *both* redoubtable and connected with high probability. This is claimed in the following corollary.

Corollary 5.4: If a secure wireless sensor network is built in such a way that $K \geq N \log N$ and such that

$$\frac{k^2}{K} \sim \frac{\log N}{N}, \tag{8}$$

then the network is redoubtable and connected with high probability.

Proof: Combine Theorem 4.6 and Theorem 5.3. ■

For instance, we can choose $k \sim \log N$ and $K \sim N \log N$. Actually, this is the choiche that minimizes the keyring size, that can be important especially in devices with small memory like sensors. If we take a larger keyring, say $k \sim \sqrt{N}$, and set the poolsize accordingly, that is $K \sim N^2/\log N$, we get a network that is provably secure and connected with high probability, but not essentially more secure against massive attacks than the previous one, in spite of the much larger keyring.

VI. Simulation Results

As discussed in the introduction, the Erdös and Rényi random graph has been used in the literature to model wireless sensor networks with random pre-distribution of keys. There are a number of problems with this approach. To show that the structure of a random graph and of a kryptograph are different, we have set up an experiment that measure the clustering coefficient of the two networks. In this experiment, we have set the key

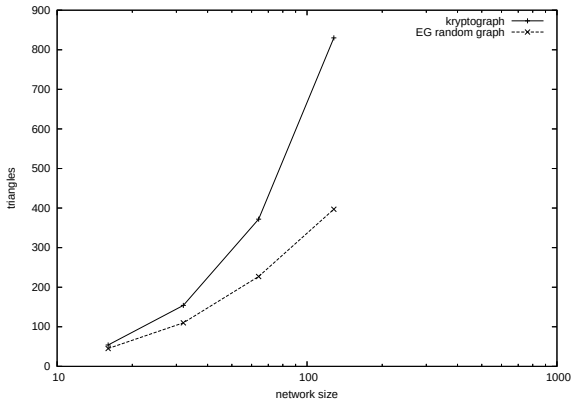

Fig. 1. Number of triangles in the Erdös and Rényi random graph and in the kryptograph as a function of the network size.

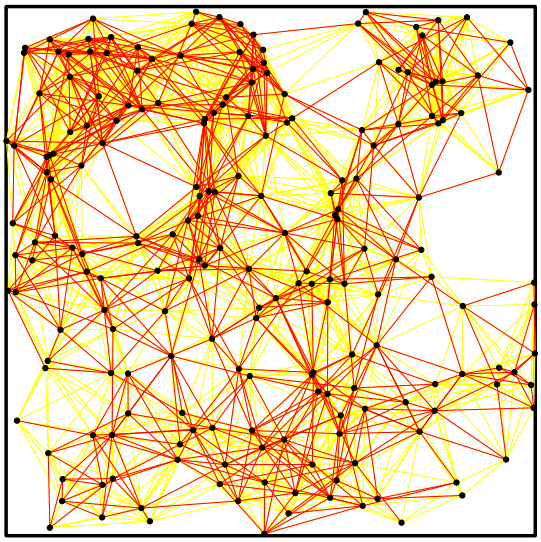

Fig. 2. Randomly generated secure sensor network of size 200. Communication range is 0.2, key ring size is 4, and pool size is 50. Lighter lines mean physical visibility, darker lines secure visibility. This graph is connected by using secure links alone.

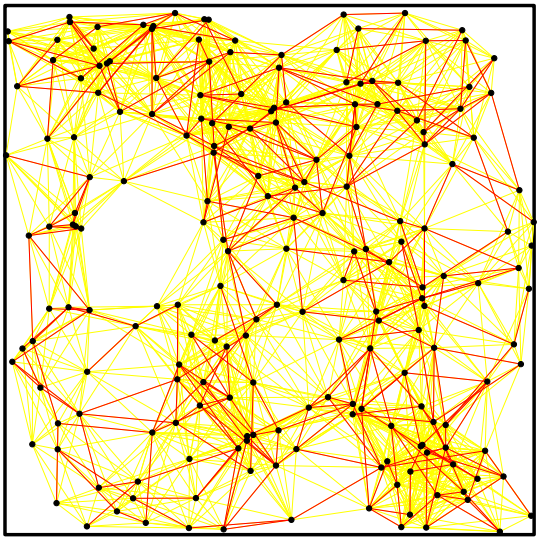

Fig. 3. Randomly generated secure sensor network of size 200. Communication range is 0.2, key ring size is 4, and pool size is 100. Lighter lines mean physical visibility, darker lines secure visibility. The network has a few isolated sensors.

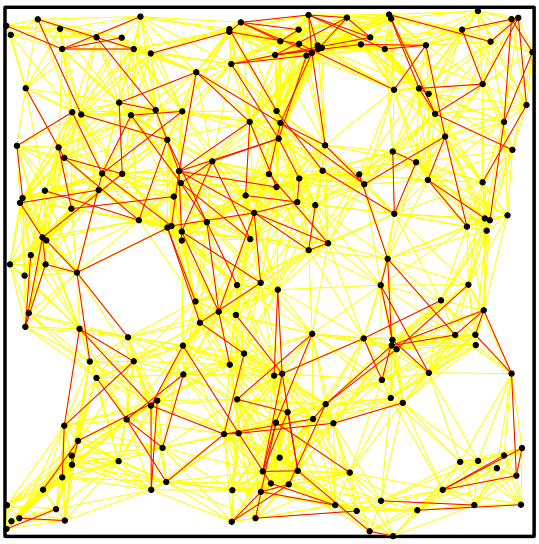

Fig. 4. Randomly generated secure sensor network of size 200. Communication range is 0.2, key ring size is 4, and pool size is 150. Lighter lines mean physical visibility, darker lines secure visibility. The network has a slightly larger number of isolated sensors and even some very small disconnected components.

ring size to $\log N$ and the pool size to $(1/2)N\log N$. This is enough to get connectivity in the full visibility case. Figure 1 shows the results: The kryptograph has a much higher number of triangles (and consequently a much higher clustering coefficient). While this difference is small when $N = 16$, it gets larger and larger as the network size grows. When $N = 128$, the kryptograph shows twice the number of triangles of the random graph, and the gap keeps growing in larger networks.

To help visualize the structure of secure wireless sensor networks, Figures 2, 3, and 4 show three similar networks where the pool size is increased from 50 to 100 and then to 150. Note that, as soon as the pool size is too big to guarantee connectivity, isolated sensors start to appear in the graph. This is perfectly analogous to what is predicted by well-known graph-theoretic results on other random models and very important from a practical point of view. Indeed, even in the remote probability that our design methodology generate a disconnected graph, it is almost surely connected except a very small number of isolated points.

In our experiments, we set the communication range to .2, while the network size N ranges from 1,000 to 10,000 sensors. To choose the parameters, we set $c = 32$ in Equation 1. Constant c depends on network density. Experimentally, $c = 32$ guarantees that high probability of connectivity holds from very low density networks. So, if we fix the keyring size to $\log N$, then the pool size is equal to $(1/32)N\log N$. Similarly, if we fix the keyring size to $2\log N$, then the pool size is equal to

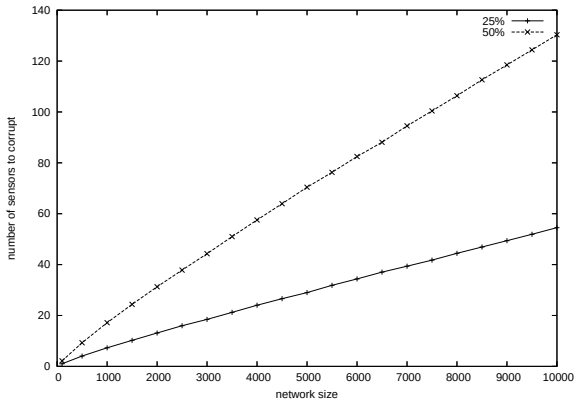

Fig. 5. Number of sensors that the attacker has to collect to compromise 50% and 25% of the network links. Pool size K is set to $N/\log N$.

$(1/8)N\log N$. Finally, a key ring size of $8\log N$ implies a pool size of $2N\log N$. We performed a large set of experiments for all of the above options. These networks are virtually "always" connected, meaning that we got no disconnected network among the 10,000 generated per each parameter choice and network size.

Lastly, we performed experiments to validate our theoretical results on network resilience. Figure 5 shows the number of sensors that a coalition must have to compromise 50% of the secure links in a network where the pool size is set to $N/\log N$ and the key ring to a constant. This graph can be almost exactly interpolated by function $hN/\log N$, for some constant h. This experimental evidence supports the asymptotic result in Theorem 5.3.

VII. Conclusion

In this paper, we have shown that wireless sensor networks using random pre-distribution of keys can be designed in such a way to be both provably connected with high probability and provably secure against massive attacks. To the best of our knowledge, this is the first fundamental result that rigorously combines two central properties of this kind of networks.

References

[1] I. F. Akyildiz, Y. Sankarasubramaniam, W. Su, and E. Cayirci. Wireless sensor networks: a survey. *Computer Networks*, 38:393–422, 2002.

[2] C. Bettstetter. On the minimum node degree and connectivity of a wireless multihop network. In *Proceedings of the* 3^{rd} *ACM international symposium on Mobile ad hoc networking and computing*, pages 80–91, 2002.

[3] R. Blom. An optimal class of symmetric key generation systems. In *Advances in Cryptology: Proceedings of EUROCRYPT '84, volume 338 of LNCS*, 1985.

[4] C. Blundo, A. De 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, volume 740 of LNCS*, 1993.

[5] B. Bollobas. *Modern Graph Theory*. Springer, 1998.

[6] Haowen Chan, Adrian Perrig, and Dawn Song. Random key predistribution schemes for sensor networks. In *Proceedings of the IEEE Symposium on Security and Privacy*, pages 197–213, Oakland, California, USA, 11-14 May 2003.

[7] Roberto Di Pietro, Luigi V. Mancini, and Alessandro Mei. Efficient and resilient key discovery based on pseudo-random key pre-deployment. In *Proceedings of the* 18^{th} *IEEE International Parallel and distributed Processing Symposium (IPDPS'04), 26-30 April 2004, Santa Fe, New Mexico, USA*, pages 217–224, 2004.

[8] Roberto Di Pietro, Luigi V. Mancini, and Alessandro Mei. Energy efficient node-to-node authentication and communication confidentiality in wireless sensor networks. *ACM/Kluwer Wireless Networks, to appear*, 2005.

[9] Wenliang Du, Jing Deng, Yunghsiang S. Han, and Pramod K. Varshney. A pairwise key pre-distribution scheme for wireless sensor networks. In *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, pages 42–51, New York, NY, USA, 2003. ACM Press.

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

[11] Laurent Eschenauer and Virgil D. Gligor. A key-management scheme for distributed sensor networks. In *Proceedings of the* 9^{th} *ACM conference on Computer and communications security*, pages 41–47. ACM Press, 2002.

[12] K. B. Singer-Cohen J. A. Fill, E. R. Schenerman. Random intersection graphs when $m = \omega(n)$: An equivalence theorem relating the evolution of the $g(n,m,p)$ and $g(n,p)$ models. *Random Structures and Algorithms*, 16:156–176, 2000.

[13] M. Karoński, E. R. Sheinerman, and K. B. Singer-Cohen. On random intersection graphs: the subgraph problem. *Combinatorics, Probability and Computing*, 8:131–159, 1999.

[14] Donggang Liu and Peng Ning. Establishing pairwise keys in distributed sensor networks. In *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, pages 52–61, New York, NY, USA, 2003. ACM Press.

[15] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

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

[17] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and J.D. Tygar. SPINS: Security Protocols for Sensor Networks. In *Proceedings of the* 7^{th} *Annual International Conference on Mobile Computing and Networking*, pages 189–199. ACM Press, 2001.

[18] K. B. Singer-Cohen. *Random intersection graphs*. PhD thesis, Department of Mathematical Sciences, The Johns Hopkins University, 1995.

[19] D. Stark. The vertex degree distribution of random intersection. *Random Structures and Algorithms*, 24:249–258, 2004.

[20] D. J. Watts and S. H. Strogatz. Collective dynamics of "small world" networks. *Nature*, 393:440–442, 1998.

[21] Sencun Zhu, Sanjeev Setia, and Sushil Jajodia. Leap: efficient security mechanisms for large-scale distributed sensor networks. In *CCS '03: Proceedings of the 10th ACM conference on Computer and communications security*, pages 62–72, New York, NY, USA, 2003. ACM Press.