Unassailable Sensor Networks

Alessandro Mei*
Dept of Computer Science
Sapienza University of Rome
mei@di.uniroma1.it

Alessandro Panconesi
Dept of Computer Science
Sapienza University of Rome
ale@di.uniroma1.it

Jaikumar Radhakrishnan
School of Technology and
Computer Science
TIFR, Mumbai, India
jaikumar@tifr.res.in

ABSTRACT

We show that massive attacks against sensor networks that use random key pre-distribution schemes cannot be cheap, provided that the parameters are set in the right way. By choosing them appropriately, any adversary whose aim is to compromise a large fraction of the communication links is forced, with overwhelming probability, to capture a large fraction of the nodes. This holds regardless of the information available to the adversary to select the nodes. We consider two important security properties: We say that the network is unassailable if the adversary cannot compromise a linear fraction of the communication links by compromising a sub-linear fraction of the nodes, and that the network is unsplittable if the adversary cannot partition the network into two (or more) linear size fragments. We show how to set the relevant parameters of random key pre-distribution—pool and key ring size—in such a way that the network is not only connected, but also provably unassailable and unsplittable with high probability. Moreover, we also show how to set the parameters in such a way to form a giant component in the network, a connected subgraph including, say, 99% of the sensors. Giant components emerge by using much smaller key rings, are sparse, and, quite remarkably, are provably unassailable and unsplittable as well. All these results are supported by experiments.

Categories and Subject Descriptors

C.2.0 [Computer-Communication Networks]: General—Security and protection; C.2.1 [Computer-Communication Networks]: Network Architecture and Design—Wireless communication; E.1 [Data]: data structures—graphs and networks

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.


General Terms

Security, Design.

Keywords

Wireless sensor network, random pre-distribution of keys, security.

1. INTRODUCTION

In this paper we study the problem of massive attacks against sensor networks. A massive attack is when the adversary is able to compromise a large fraction of the communication links of the network, by means of capturing nodes and thereby gaining control of the cryptographic material they contain. Since sensor networks are typically unattended, collecting nodes is considered to be a realistic way to tamper with their security (see [1] among others). Even though it may partly fall under control of the adversary, the network as a whole can continue to operate in a useful manner if the inflicted damage is confined to a small part. This is indeed the case in many applications of sensor networks, such as monitoring of the environment and of the infrastructure. In these cases, it can be much more important to protect the global functionality of the network than few individual communication links. If the adversary has enough resources to capture a large fraction of the nodes, then there is not much that can be done to salvage the functionality of the network. Therefore, a meaningful question is to ask whether a massive attack can be effected cheaply, that is, by capturing just a small fraction of the nodes.

This is the problem we investigate in this paper. We prove in a mathematically rigorous fashion that the well-known random key pre-distribution scheme [2] makes a sensor network secure against any massive attack whatsoever.

We establish security against massive attacks regardless of the strategy used to collect nodes. For instance, the adversary could obtain information on the network by means of traffic analysis or by exploiting weaknesses of the key-discovery protocol, and use this information to select an appropriate strategy and a small set of nodes. We show that it is possible to fix the relevant parameters for the security of the network—pool and key-ring size—in such a way that, with overwhelming probability and regardless of the information available to the adversary, the network is at the same time connected and secure against massive attacks. In other words, only “expensive” massive attacks can succeed.

In the literature, security against massive attacks, and security properties in general, are often investigated only
experimentally. One of the main contributions of this paper is to show that they can be certified a-priori in a precise and rigorous way. We remark that the guarantees we provide are out of reach for experimental approaches.

Besides massive attacks, we analyse another kind of attack and certify random pre-distribution of keys against it. Informally, the question is whether the adversary can split the network into two large chunks and compromise all communications between them just by acquiring a small set of nodes. We show that this is not possible, again regardless of the information available to the adversary. Besides their intrinsic interest, we feel that our results introduce an important element of novelty at the methodological level in the large body of research on random pre-distribution of keys and, more generally, on pairwise-communication security. Furthermore, our formal framework and methodological approach is general enough to be applicable to other contexts, such as peer-to-peer networks.

We now describe our results more precisely. We begin by reviewing briefly the model, formally defined in Section 3. We are given a network of \( n \) nodes and a pool of \( K \) cryptographic keys. Each node \( u \) is given a random subset \( K_u \) of \( k \) keys, its key ring. Two nodes in the network are connected by a link if and only if they are within transmission radius and share a key. The resulting graph is called a kryptograph. The basic question is how to set the values of \( K \) and \( k \) in order to have some desirable properties, such as various forms of connectivity and, simultaneously, security. It should be noted that security and connectivity are at odds. The former likes large key rings (relative to the pool size), since not all nodes are connected and secure against the omniscient adversary is forced with the computing power to cause maximum damage, within its limited budget. We call this the omniscient adversary show that, if we are interested in protecting the network against massive attacks, we do not (asymptotically) gain any advantage by using challenge response with respect to any light-weight, non-cryptographic protocol that does not keep key distribution secret. Broadcasting lists of key indices in the clear is good as well (within constant factors).

Interestingly, our results show that it is possible to build a secure sensor network that is very strong as a whole, while every individual link that composes the same network is extremely weak—it is not hard to see that the omniscient adversary can corrupt arbitrarily chosen individual links extremely easily, it is enough to select and corrupt any node in the network that contains the same key used to secure the target link.

We prove that kryptographs enjoy another strong security property. Suppose now that the aim of the adversary is to split the network into two large sets and to compromise all links between them, i.e. the adversary wants to partition the network. We show that, with overwhelming probability, the omniscient adversary cannot do this. This result is a consequence of the following structural property of kryptographs: with high probability and simultaneously for all large (linear size) sets of vertices, the number of edges between a set and its complement is \( \Theta(n \log n) \). This fact implies good fault-tolerant properties: to partition a network into two large sets, a huge (linear fraction) of the links must go down.

Finally, we extend our results to giant components. For many applications it suffices to have a giant connected component that covers the area within which the network is deployed. We show that if

\[
\frac{k^2}{K} \sim \frac{1}{n}
\]  

then, with high probability, the network has a giant connected component, say, a connected component containing 99% of the vertices. Note that Condition 2 can be satisfied with much smaller key rings than Condition 1. This results in memory saving (smaller key ring) and computation saving (faster key discovery), an important consideration in resource starving environments typical of sensor networks. But more importantly, networks generated with Condition 2
are much sparser, not only globally but also locally at the neighborhood level, than those set up under Condition 1. This is very beneficial because it translates in a reduced amount of interference, limiting the number of packet collisions and corresponding retransmissions. Nonetheless, we prove that the giant connected components are unassailable and unsplittable against the omniscient adversary as well, just like the whole kryptograph.

The properties we have been describing can be proven in the general case, but here we present the full-visibility case only. We made this choice because in this fashion the basic argument with the underlying reasons emerge more clearly, without cluttering technicalities. The general case is dealt with in the experiments. They complement and sharpen the predictions of our theorems rather than being the sole evidence of security.

To summarize, we believe that we have introduced an interesting approach from the methodological point of view that allows the rigorous investigation of security properties, not only for sensor networks. In particular, we have also shown that random key pre-distribution ensures surprisingly strong security properties against massive attacks.

2. RELATED WORK

Key management is one of the central security issues for wireless sensor networks (see [1], among others). The idea of probabilistic key sharing for WSNs is introduced by Eschenauer and Gligor [2]. The authors also provide a simple and centralized algorithm for re-keying in a distributed WSN. Later, in [4], a few new mechanisms are described in the framework of random key pre-distribution. Among these, the q-composite random key pre-distribution scheme, a modification of the basic scheme in [2], achieves better security under small scale attacks while trading off increased vulnerability in the face of a large scale physical attack on the network sensors. Secure key discovery protocols for random pre-distribution of keys have been proposed in [2, 4], and [5].

Two schemes combine the random pre-distribution scheme with a deterministic technique to build up secure pairwise channels. The first scheme is proposed in [6]. The authors use a deterministic protocol proposed by Blom [7] 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 [6] create a set \(W\) 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 [8]. In principle, this work is similar to [6], where Blundo et al’s polynomial scheme [9] is used instead of Blom’s.

The problem of network connectivity when using random pre-distribution of keys has been addressed in [2]. Their basic idea is that the network can be considered to be a random graph in the sense of Erdős and Rényi [10]. However, notice that a kryptograph 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 Erdős and Rényi. In fact, random graphs and kryptographs have different structural properties as illustrated in [3]. There are other difficulties. The Erdős-Rényi 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 Erdős-Rényi Theorem as used in [2] ensures high probability of connectivity in the general case, when devices are not within transmission range. Later, the work in [11], that applies the well-known results by Erdős and Rényi on giants components in the random graph to the kryptograph following the same methodology in [2], shows exactly the same problems. We remark that having a precise understanding of a model is always important, but especially so when security is at stake. Moreover, note that security properties, like the ones we consider in this paper, cannot even be formulated in the Erdős-Rényi model.

In [3], for the first time a precise mathematical analysis is given of the connectivity and security properties of sensor networks that make use of the random pre-distribution of keys. In that work, the authors show that, if

\[
\frac{k^2 \log \frac{n}{K}}{K} \geq \frac{1}{\log n} \quad (3)
\]

then, with high probability, the network is simultaneously connected and secure against the random adversary, an adversary that blindly picks sensors at random to compromise the network. In this paper, we extend the result to any possible massive attack, to unsplittability, and to giant components.

Connectivity properties have been studied for non-secure wireless sensor networks as well. In [12], a geometric random model has been used to investigate minimum node degree and h-connectivity. Using a recent asymptotic result from Penrose [13], 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 [14]) 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 [14, 15, 16, 17]. 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.

3. PRELIMINARIES

In this paper we will study probability events that depend on \(n\), the number of vertices of the network. Examples of such events are “the network is connected”, “the network has a giant component” etc.

**Definition 3.1.** Let \(E_n\) denote an event that depends on \(n\). We will say that \(E_n\) holds with high probability if

\[
\lim_{n \to \infty} \Pr[E_n] = 1.
\]

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

**Definition 3.2.** Let \(K\) be the size of a finite set of keys (the pool), and let \(k \leq K\) be a fixed parameter. Let \(|K| = \ldots\)
Let \( \{1, 2, \ldots, K\} \) be the index set of the keys in the common pool of size \( K \). The graph \( G_{r,k,K}^n \) 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 \( K_u \), which is generated 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, we is an edge if (a) the two nodes are within distance \( r \); and (b) \( K_u \cap K_v \neq \emptyset \).

The resulting graph \( G_{r,k,K}^n \) is called a kryptograph with parameters \( r, k, K, n \). In the special case in which every two nodes are within transmission range \( r \geq \sqrt{2} \), the so-called full visibility case, the resulting graph is denoted as \( G_{r,K}^n \).

In the sequel, for the 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 \( G_{r,k,K}^n \) 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 this paper we study analytically only the full visibility case, even though our proofs extend to the general case. We will study the network under the assumptions

\[
\frac{k^2}{K} \sim \frac{\log n}{n} \quad (4)
\]

and

\[
\frac{k^2}{K} \sim \frac{1}{n} \quad (5)
\]

The first is a necessary and sufficient condition if we want full connectivity with high probability [3]. We show in this paper that the second condition ensures that, with high probability, the network has a giant component. We will also assume that \( k \geq 2 \).

**Definition 3.3.** The omniscient adversary is defined as follows:

- It knows the key ring of every vertex in the network;
- It selects a set of \( t \) vertices, thereby capturing all their key rings;
- A link \( uv \) is corrupted if the adversary has captured a key \( x \in K_u \cap K_v \) (one key is enough);
- The adversary must corrupt a constant fraction of the links: Its goal is to minimize \( t \).

Henceforth we will refer to the omniscient adversary simply as the adversary.

**Definition 3.4.** A network is unassailable if the adversary must collect a linear fraction of the nodes in order to corrupt a linear fraction of the links.

We will investigate another property that is useful not only for security but also for fault-tolerance. Assume that the aim of the adversary now is to find a set of vertices \( S \) such that all edges from \( S \) to \( V - S \) are corrupted. Then we say that the set \( S \) is bad.

**Definition 3.5.** A network is unsplittable if there is no bad set \( S \) of size \( 3n \leq |S| \leq \frac{2}{3}n \), where \( \beta := \frac{1}{3} \).

We will show that if the adversary captures \( t = o(n) \) many vertices then, with high probability the network is unsplittable. We remark that this result holds for all \( t \in (0, \frac{1}{2}) \). Our choice of \( \beta = \frac{1}{3} \) is for the sake of simplicity.

We now review some standard definitions and facts.

**Definition 3.6.** We use \( f \sim g \) to mean that \( f \) and \( g \) are the same up to constant factors. We say that \( f(n) = o(g(n)) \) if \( \lim_{n \to \infty} \frac{f(n)}{g(n)} = 0 \).

**Fact 3.7.** (Hoeffding bound [18, 19]). Let \( K_1, K_2, \ldots, K_t \) be subsets of \( [K] \), each of size \( pK \) (for some \( p \in [0,1] \)), chosen independently at random. Let \( A \) be a subset of \([K]\) of size \( a \). Then, for all \( t \in [0,1] \),

\[
\Pr \left[ \sum_{i=1}^{t} |K_i \cap A| \geq (p + t)\alpha \right] \leq \left( \frac{e p}{p + t} \right)^{(p+1)\alpha} ;
\]

\[
\Pr \left[ \sum_{i=1}^{t} |K_i \cap A| \leq (1 - t)p\alpha \right] \leq \exp \left( -\frac{t^2}{2p\alpha} \right).
\]

**Fact 3.8.**

\[
\left( \frac{n}{k} \right)^k \leq \left( \frac{en}{K} \right)^k
\]

### 4. UNASSAILABLE NETWORKS

A kryptograph can be thought of as a balls-and-bins experiment in which bins correspond to keys of the pool. To choose its key-ring, each vertex is given \( k \) balls. The key-ring is chosen by throwing the balls of a given vertex into \( k \) distinct bins. If a ball of vertex \( u \) lands in bin \( i \) then key \( i \) is assigned to \( u \). Thus, there is an edge \( uv \) in the graph iff there is a bin that contains a ball from \( u \) and a ball from \( v \).

The following observation is the key to limiting the power of the adversary. Let \( A \) denote the set of \( t \) bins (corresponding to the \( t \) keys) chosen by the adversary and let \( L \) denote the \( t \) largest bins. Then, denoting with \( B_i \), the number of balls in bin \( i \), we have

\[
(\#edges captured) \leq \sum_{i \in A} \binom{B_i}{2} \leq \sum_{i \in L} \binom{B_i}{2}
\]

Thus, in order to show that the adversary with \( t \) keys cannot capture many edges, it is sufficient to show that the number of pairs in the \( t \) fullest bins is small.

**Definition 4.1.** A bin is useful if it contains at least two balls. It is useless otherwise.

Our argument to show that the kryptograph is unassailable has two parts. We show that: (a) in order to capture a linear fraction of the edges, the adversary must pick \( \Theta(n \log n) \) bins; (b) no vertex can contribute more than \( O(\log n) \) useful bins. It follows that in order to capture a linear fraction of the edges, a linear fraction of the vertices must be captured. Both (a) and (b) will be shown to hold with high probability. We begin with (b).

**Lemma 4.2.** With high probability, no vertex contributes more than \( 10 \log n \) useful bins.
Proof. For a vertex $u$, let $K_u$ denote the set of bins where the balls originating from $u$ fall. Now fix some vertex $u$, and let $Y$ denote the number of useful bins in $K_u$. Let $X$ be the number of balls originating from the other vertices that land in one of the bins in $K_u$. Clearly $X \geq Y$, so that $\Pr[Y > t] \leq \Pr[X > t]$. We will bound the latter. Note, that

$$X = \sum_{v \neq u} |K_v \cap K_u|.$$  

Using Fact 3.7 (with $p = \frac{1}{n}$, $\ell = n - 1$ and $p + t = \frac{10 \log n}{(n - 1)K} \geq 10p$), we conclude that

$$\Pr[X > 10 \log n] \leq \left(\frac{e}{10}\right)^{10 \log n} \ll n^{-10}.$$  

The claim follows from this by using the union bound, by summing over all choices for $u$. \hfill \Box

Note that the above lemma says that a huge fraction of the key ring is not used. For instance, if $\eps \approx n^{1/2}$, then $\log n$ balls are at most three keys $S, V$ are at most three keys $S, V$. Thus, in order to capture $\log n$ many edges (i.e. a constant fraction), the adversary must capture at least $\log n/10$ useful bins.

Let $\alpha > 0$ be a constant. By Lemma 4.4, with high probability, no bin contains more than 5 balls, i.e. every key is used at most by $\Theta(\frac{\log n}{n}) = 10$ edges. Thus, in order to capture $\log n$ many edges (i.e. a constant fraction), the adversary must capture at least $\log n/10$ useful bins. By Lemma 4.2, with high probability, no vertex can contribute more than $10 \log n$ useful bins. It follows that, with high probability, the adversary must capture at least $\log n/100 \log n = \Theta(n)$ vertices. \hfill \Box

5. UN Splittable Networks

We now show that the adversary cannot easily partition the network into two linear size sets, corrupting all edges between them.

Given a set of vertices $S$ we denote by $(S, V - S)$ the set of edges with one endpoint in $S$ and the other in $V - S$, i.e. the set of edges that cross the cut. We will show that, with high probability, simultaneously for all $S$ with linearly many nodes, the number of edges that cross the cut is $\Theta(n \log n)$. The results of Section 4 imply that in order to corrupt all these edges the adversary must capture a linear fraction of the nodes, i.e. the network is unsplittable.

Again, we will show the result in the full visibility case under the hypothesis $K = n^\alpha$ with $\alpha \geq 2$. The result holds in the general case if $K = n \log n$ but the somewhat lengthy proof is omitted.

Lemma 5.1. With high probability, for all $S$ of size in the range $[0.1, 0.5n]$, $(S, V - S)$ contains $\Omega(n \log n)$ edges.

Proof. Fix a set $S$ of size $s \in [0.1n, 0.5n]$. To show that there are $\Omega(n \log n)$ edges leaving $S$, we proceed indirectly. Consider the set of triples

$$T = \{(v, w, g) : v \neq w \text{ and } g \in K_v \cap K_w\}.$$  

Let $T(S) = \{(v, w, g) : v \in V \text{ and } w \in V - S\}$. We show that (a) with high probability, for all $S$, $|T(S)| = \Omega(n \log n)$, and (b) with high probability for every pair $(v, w)$, there are at most 3 keys $g$, such that $(v, w, g) \in T$. Thus, the number of edges in $(S, V - S)$ is at least $|T(S)|/3 = \Omega(n \log n)$. As in the proof of Lemma 6.4, we will condition on the event $\neg \mathcal{E}(S)$. Note that $|T(S)| \geq \sum_{w \in V - S} |k(S) \cap K_w|$, and

$$\mathbb{E} \left[ \sum_{w \in V - S} |k(S) \cap K_w| \right] = \frac{|k(S)| |k(n - s)|}{K} \geq \frac{sk^2(n - s)}{10K} \geq \alpha n \log n,$$

for some constant $\alpha > 0$ (independent of $|S|$). (For the last inequality we used $\frac{k^2}{n} \sim \log n$ and $0.1n \leq s \leq 0.9n$.) We can use Lemma 3.7 (with $p = \frac{1}{n}$, $\ell = n - s$ and $a \leftarrow |k(S)|$) and a routine computation, to conclude that the probability that $|T(S)|$ is less than $(\alpha/2)n \log n$ is at most $\exp\left(-\left(c_0/2\right)n \log n\right)$. Thus,

$$\Pr[|T(S)| < (\alpha/2)n \log n] \leq \Pr[\mathcal{E}(S)] + \exp(-\left(c_0/2\right)n \log n).$$

Using the union bound we obtain (by summing over all $S$
such that $\beta n \leq |S| \leq 0.5n$,

$$\Pr[3S \beta n \leq |S| \leq 0.5n, |T(S)|] \leq \sum_S \Pr[|E(S)| + \exp(-(\alpha s/8)n \log n)] = 2^{-\Theta(n)},$$

where we use Lemma 6.2 to justify the last inequality. This completes part (a) of our argument.

For part (b), fix a pair of distinct vertices $(v, w)$. If $|K_v \cap K_w| \geq 4$, then there is a subset $J \subseteq K_v$ of size 4 such that $J \subseteq K_w$. Thus, by the union bound again

$$\Pr[|K_v \cap K_w| \geq 4] \leq \binom{k}{4} \binom{k-4}{4} \binom{K}{k}^{-1} \leq O\left(\frac{k^2}{K^4}\right).$$

Since $\frac{k^2}{K^2} \sim \frac{\log n}{n}$, this quantity is $O(n^{-3})$. There are at most $n^2$ pairs $(v, w)$, so with probability $1 - o(1)$, for all $(v, w)$ there are at most three keys $s$ such that $(v, w, s) \in T$. This establishes part (b) and completes the proof of the lemma. \qed

**Theorem 5.2.** Assume $K := n^\alpha$ with $\alpha \geq 2$. Then, with high probability, the network is unsplittable.

**Proof.** Follows immediately from Lemma 5.1 and Theorem 4.5. \qed

## 6. THE EMERGENCE OF A GIANT COMPONENT

In this section, we prove that, if $\frac{k^2}{N} \sim O$, for a appropriately large constant $C_0$, then with high probability the cryptograph has a giant connected component. (To keep the proofs simple, we do not attempt to optimize the constant.) In particular, we will show that with high probability there exists a connected component of size $0.9n$. The same argument can be used to show the existence of a connected component of size $0.9n$, for any $\alpha \in (0, 1)$.

Let $k(S)$ denote the set of keys chosen by vertices of $S$, and let $s$ denote the size of a set $S$. The expected size of $\kappa(S)$ is roughly $ks$. The next lemma shows that it is extremely unlikely that any large (linear size) set has a combined key ring that is much smaller than this. From now on, let

$$\alpha, \beta := \frac{1}{10}.$$

**Definition 6.1.** For a set $S \subseteq V$, let $E(S)$ denote the event $|k(S)| < \alpha k|S|$.

In the next lemma we assume $\frac{k^2}{N} = \frac{c}{n}$ where $c \leq k$. In this way the lemma holds for the case $\frac{k^2}{N} \sim \frac{c}{n}$, which is relevant for this section, as well as the case $\frac{k^2}{N} \sim \frac{\log n}{n}$, which is relevant for the next section.

**Lemma 6.2.** Let $\frac{k^2}{N} = \frac{c}{n}$ where $c \leq k$. Then,

$$\sum_{S \subseteq V: \beta n \leq |S| \leq \frac{c}{2}} \Pr[|E(S)|] = 2^{-\Theta(n)}.$$

**Proof.** We first bound $Pr[|E(S)|]$ for a fixed set $S$, and then sum over all choices of $S$. We carry out our calculations pretending that the vertices obtain their key rings by sampling with replacement; note that this only makes or claim stronger. Fix a set $S$, and let $s = |S|$. By the union bound, and recalling that $k \geq c$, we have

$$\Pr[|E(S)|] = \Pr[3S \beta n \leq |S| \leq 0.5n, \forall x \in T(s) \subseteq \{\alpha sk\}, \forall x \in S \subseteq \{\alpha sk\}] \leq \left(\frac{K}{\alpha sk}\right)^{-\alpha sk} \leq \left(\frac{\alpha sk}{K}\right)^{-\alpha sk} = \left(\frac{\alpha sk}{\alpha s n}\right)^{(1-\alpha)k} = \left(\frac{\alpha s}{n}\right)^{(1-\alpha)k}.$$

Now, we sum over all choices $S$ (with $\frac{n}{10} \leq |S| \leq \frac{c}{2}$),

$$\sum_{S} \Pr[|E(S)|] \leq \sum_{s=\frac{n}{10}}^{\frac{c}{2}} \binom{n}{s} \left(\frac{\alpha s}{n}\right)^{(1-\alpha)k} \leq \sum_{s=\frac{n}{10}}^{\frac{c}{2}} \left(\frac{\alpha s}{n}\right)^{(1-\alpha)k} \leq \sum_{s=\frac{n}{10}}^{\frac{c}{2}} \frac{1}{2^s} = O(1).$$

For the last inequality we used our assumption $k \geq 2$, which implies that

$$e^{\alpha k + (1-\alpha)k} \leq e^{\alpha k (1-\alpha)k} \leq (ea^{1-\alpha})^k < 1.$$ \qed

The next lemma says that with high probability all large enough subsets of nodes must be connected to a vertex outside the set. This easily implies the existence of a giant component.

**Definition 6.3.** A cut is a non-empty subset $S \subseteq V$ of size at most $\frac{c}{2}$ such that there is no edge between $S$ and $V - S$. Let $C(S)$ denote the event “$S$ is a cut”.

**Lemma 6.4.** Let $\frac{k^2}{N} = \frac{c}{n}$ where $c \leq k$. Then,

$$\Pr[3S, \beta n \leq |S|, C(S)] \leq 2^{-\Theta(n)}.$$

**Proof.** Fix a set $S$ of size $s$, such that $\beta n \leq s \leq \frac{c}{2}$. Then,

$$\Pr[C(S)] \leq \Pr[E(S)] + \Pr[C(S) \mid \neg E(S)].$$

We have already analyzed the first term in the previous
ability for larger networks (when \( n = 2,000 \) and larger, for \( n \) less than \( .01 \) when connected, meaning that full connectivity fails with probability parameters, the network is “almost always” completely connected, with the probability. A first experiment has been done to check theoretical findings. A first experiment has been done to check for the sake of brevity.

7. EXPERIMENTS

We set up a number of experiments to back up our theoretical findings. A first experiment has been done to check that, if parameters are chosen according to Condition 1, then the sensor network is actually connected with high probability (note that this first part of the experiment replicates what has been done in [3]), and, more importantly, that if parameters are chosen according to Condition 2, then a giant connected component emerges in the network with high probability.

The details of the experiment are as follows: Network size \( n \) ranges from 1,000 to 8,000 sensors, communication radius is set to \(.1\), pool size to \( n \log n \), and key ring size to \( c \log n \), where \( c \) is set to 10. Constant \( c \) depends on the communication range and must be experimentally tuned. With these parameters, the network is “almost always” completely connected, meaning that full connectivity fails with probability less than \(.01\) when \( n = 1,000 \), and with decreasing probability for larger networks (when \( n=2,000 \) and larger, for example, we have got no disconnected network in the experiments after more than 10,000 runs). Note that probability of less than \(.01\) is acceptable for our purposes, it is enough to slightly increase constant \( c \) to make it as small as desired.

To check the existence of a giant connected component with much smaller key rings (asymptotically smaller, as shown theoretically), we performed the same experiment setting the key ring size in accordance with Condition 2. To help visualize what a giant connected component looks like in a kryptograph, in Figure 1 it is shown a sensor network where gray edges represent physical links and black edges secure links. Secure links in Figure 1 form a giant connected component. The same network is shown in Figure 2 where gray edges are secure links while black edges are those in the giant connected component.

Experimentally, a key ring of size \( 11 \sqrt{\log n} \), same pool size of \( n \log n \), is enough to guarantee the emergence with high probability of a giant connected component of at least \% of the sensors for \( n \) ranging from 1,000 to 8,000. Constant 11 depends on communication range and on the size of the desired giant component. Indeed, further experiments show that a key ring of size \( 13 \sqrt{\log n} \) guarantees a giant component of 95\%, while a key ring of size \( 15 \sqrt{\log n} \) guarantees a giant component of 99\% for the same network sizes. These conclusions have been validated by thousands of runs. See Figure 3 to appreciate the large gap between the key ring size needed for full connectivity and the key ring size for the presence of a giant component. This gap results in different structural properties of the kryptograph, for example giant components are much sparser. Figure 4 shows the average number of edges of the networks for key ring size of \( 10 \log n \), \( 11 \sqrt{\log n} \), \( 13 \sqrt{\log n} \), and \( 15 \sqrt{\log n} \); that is, in the case of condition 2.

\[
\text{Pr}[\exists S, \beta n \leq |S|, \mathcal{C}(S)] \leq \sum_{S \subseteq V} \Pr[\mathcal{C}(S)] \\
\leq \sum_{S \subseteq V} \Pr[\mathcal{E}(S)] \\
+ \Pr[\mathcal{C}(S) | \sim \mathcal{E}(S)].
\]

The claim now follows from Lemma 6.2 and (7).

\[\Pr[\mathcal{C}(S) | \sim \mathcal{E}(S)] \leq 4^{-n}.\] (7)

[For the last inequality we choose \( C_0 \) large enough.] Using the union bound (summing over choices of \( S \) such that \( \beta n \leq |S| \leq \frac{s}{2} \)), we have

\[\Pr[\exists S, \beta n \leq |S|, \mathcal{C}(S)] \leq \sum_{S \subseteq V} \Pr[\mathcal{C}(S)] \\
\leq \sum_{S \subseteq V} \Pr[\mathcal{E}(S)] \\
+ \Pr[\mathcal{C}(S) | \sim \mathcal{E}(S)].\]

The claim now follows from Lemma 6.2 and (7).

\[\text{Theorem 6.5. With high probability, the kryptograph has a connected component of size at least } 0.9n.\]

**Proof.** Assume the event of the statement of Lemma 6.4. Thus, if \( S \) is a cut then it has fewer than \( 0.1n \) vertices. In particular, the graph has no connected component with size in the range \([0.1n, 0.9n)\). Now, if the largest connected component has more than \( 0.9n \) vertices, then we are done. Otherwise, it must have fewer \( 0.1n \) vertices. Start collecting components \( S_1, S_2, \ldots \) until their total size exceeds \( 0.1n \). The total size is at most \( 0.2n \), and we get a cut with more than \( 0.1n \) vertices, a contradiction.

Similarly to what has been done for the connected kryptograph, when Condition 2 holds the resulting giant connected component can be shown to be unassailable and unsplittable. The proof is slightly more technical and it is omitted for the sake of brevity.

**Figure 1:** Kryptograph with 200 sensors. Gray edges represent physical links, while black edges represent secure links. Here, there is giant connected component that includes more than 90\% of the sensor nodes.
Figure 2: Kryptograph of Figure 1, where now gray edges represent secure links, while black edges represent secure links in the giant connected component. It is possible to observe the presence of a number of very small disconnected components.

Figure 3: Plot of key ring size to get connectivity, a giant component of 90%, 95%, and 99% of the sensor nodes. Other parameters are those used in the experiments.

Figure 4: Number of secure edges in the network in case of connectivity, and in the presence of a giant connected component of 90%, 95%, and 99% of the sensor nodes.

Figure 5: Key load in a kryptograph with 8,000 sensors, communication radius .1, pool of size 103,726, and key ring of size 129. The parameters guarantee that the kryptograph is connected, unassailable, and unsplittable with high probability.

Let’s now turn to security. All the following experiments have been performed with the parameter choice tuned in the experiments above. It is important to realize there may be keys in the pool that are used (individually) to protect many links, more than a constant number. These keys are more “important”. This is a problem for security since the omniscient adversary knows these keys, and in which key rings they are stored, and can perform the attack in such a way to collect them. Since we know from our theoretical results that the adversary still cannot get a big advantage by using the optimal strategy, it must be true that important keys are few in the kryptograph. To experience this phenomenon, we have performed an experiment to draw the outline of key load for all the keys in the pool. Figure 5 shows how many links are secured using the same individual key, for every key in the pool starting from the most important to the least important. The experiment has been done with the above described choice of parameters to
get connectivity in a 8,000 node network. A large fraction of the keys are not used at all, while there are indeed keys that are used to secure as many as 14 links. However, it is also evident from the outline that there are not too many important keys in the network. Figure 6 zooms on the first 100 keys.

To see whether our first intuition is correct (the outline can vary considerably when n changes) with set up an experiment to check whether the kryptograph and the giant components are unassailable. To check experimentally this property (and the following) is not easy. The omniscient adversary is not computationally bounded, and implementing the optimal strategy takes exponential time. Since this is not feasible except when the network is composed of very few sensors, we have implemented an approximate, greedy version of the adversary. We assume that the adversary at every key choice greedily picks the key that gives the largest immediate advantage. Figure 7 shows the results: The number of “important” keys required by the (approximate) omniscient adversary to compromise 50% of the networks links grows linearly with network size, for the whole connected kryptograph and for every giant component. That is, connected kryptographs and giant connected components in kryptographs are both unassailable and unsplittable, even in networks of practical sizes.

8. REFERENCES


Figure 9: Vertical geographical cut to split a giant connected component in a kryptograph into two parts of roughly the same size. Black edges are those that cross the cut.

Figure 10: Number of important keys to collect in order to split geographically the network into two parts in a connected kryptograph and in giant connected components of different size.


