2011 Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing

# Analysis of Multi-Path Random Key Pre-Distribution for Wireless Sensor Networks

Chun-Wei Tsai*, Wei-Shuo Li†, Min Chen‡, Wen-Shyong Hsieh†§, and Chu-Sing Yang¶
*Applied Geoinformatics, Chia Nan University of Pharmacy & Science, Taiwan, R.O.C.
†Computer Science and Information Engineering, Shu-Te University, Taiwan, R.O.C.
‡Computer Science and Engineering, Seoul National University, Seoul, Korea
§Computer Science and Engineering, National Sun Yat-sen University, Taiwan, R.O.C.
¶Electrical Engineering, National Cheng Kung University, Taiwan, R.O.C.

***Abstract*—Most wireless sensor networks require that every large enough node contain certain properties. By using the Szemerédi's regularity lemma, one can approximate a complex network by a much simpler object in such a way that the approximation is "regular" for most pairs of partitions of this network. After obtaining a more traceable network, we establish bounds for the probability of the property that a random key pre-distribution subgraph satisfies that each node has a path of length $\ell$ to its $\ell^{\text{th}}$-hop neighbors. The end result is a sharp threshold $p \geq Cn^{-(\ell-1)/\ell}$ that satisfies this property and that can be considered as an application of the sparse Szemerédi's regularity lemma.**

***Keywords*-Sensor network, security, random key pre-distribution.**

## I. Introduction

The so-called wireless sensor networks (WSN) has been a buzzword and a popular research topic over the past ten years or so because it provides new solutions to problems such as temperature and humidity monitoring, analysis of the motion of tornadoes, fire detection, military defense, and health monitoring [1], [2], [3], [4]. Especially, compared to the traditional sensors, the sensors of WSN are small, cheap, intelligent, and energy-efficient. That is why we use them in some hostile environments. To better understand whether traditional technologies can be applied to WSN or not, Akyildiz et al. [3] paid special interest in the differences between the wireless sensor networks and the wireless ad hoc networks. Among them, a striking difference is in that the topologies of a sensor network are changing very frequently compared with the wireless ad hoc networks. In the same study, Akyildiz et al. also pointed out that the sensors of WSN are limited in power, computational capacity, memory, and storage. For these reasons, most efficient algorithms, applications, and policies for wireless ad hoc networks are not suitable for wireless sensor networks. That is why new solutions are needed for WSN.

Recently, several studies [2], [3], [5], [6] have tried to discuss and review the research on WSN from different perspectives. In [6], Yick et al. divided WSN into five types: terrestrial WSN, underground WSN, underwater WSN, multimedia WSN, and mobile WSN, each of which can be adopted to the particular approach in question. More recently, a promising research direction, i.e., *security* [7], [8], [9] that may affect the performance and reliability of WSN has been addressed. Du and Xiao [7] did a good survey on the security of the sensor networks. They introduced several attacks on the sensor networks based on the layers. This classification helps differentiate the security research issues on WSN. For instance, two well known attacks on the physical layer introduced by this study are jamming and tampering. The jamming attack uses a small number of randomly distributed jamming nodes to disrupt the network in such a way that all of the nodes in the sensor network cannot be served. The tampering attack means that the adversary (i.e., the attacker) can capture the sensor node and extract some important information by tampering with the node they captured. By using this classification, the system manager can easily find the solution when the system undergoes the security problem. However, many studies [7], [8] pointed out that because the cost of tamper-resistant sensor nodes is very high, most of today's WSN nodes are, by design, not tamper-resistant. For this reason, when the adversary compromises a node, the whole sensor networks will be compromised [7], [8]. The random key pre-distribution scheme described herein avoids such a problem, by reducing the number of nodes compromised. Nevertheless, this may somehow affect the connectivity. The question is, how do we balance such a tradeoff, i.e., the number of nodes compromised and the connectivity?

In this paper, we analyze the question of the threshold probability function for the property "every pair of vertices lies in a secure path" with respect to a security parameter $p$ [9]. The goal is to obtain a threshold function $p_0(n)$ such that if $p \gg p_0$, then the probability of a security subgraph having the above property is close to 1 whereas if $p$ is slightly smaller than $p_0$, then the probability is close to 0. We show that in almost all security graphs, the number of secure paths of length $\ell$ is close to its expectation provided that $\varepsilon$ is sufficiently small, $p \geq Cn^{-(\ell-1)/\ell}$, and $n$ is sufficiently large.

## II. Related Works

Random key pre-distribution (*RKP*) based schemes [9], [10], [11], [12], [13] have been developed to guarantee the existence of specific properties, such as disjoint secure paths and disjoint secure cliques, to achieve a secure cooperation between

978-0-7695-4372-7/11 $26.00 © 2011 IEEE
DOI 10.1109/IMIS.2011.100

nodes [13], [11]. In the *RKP* wireless sensor networks, such subgraph count problems are difficult to attack because there are several levels of dependencies among the problems. In [7], Du et al. introduced several key pre-distribution schemes. Of them are fully pairwise key scheme, $\lambda$-secure key scheme, $p$-probabilistic key pre-distribution scheme, $q$-composite key pre-distribution scheme, random pairwise key scheme, and multi-space key scheme. Key pre-distribution schemes can also be classified according to the security model, be it deterministic or probabilistic, and the properties of the group keys. The primary threat of a security model is the ability of a compromised node to obtain a group key. The deterministic key establishment schemes can guarantee that a communication group $C \in \mathcal{C}$ is able to establish a common key (or pairwise key). In contrast, the probabilistic key establishment schemes of a communication group have a certain probability of being secure. This kind of group key schemes is dependent on the combinatorial methods. In our previous works [14], [15], we proposed a scheme to enhance the connectivity by using path-key. We also presented a rather precise probabilistic model.

Following are several reasons why the random graph theory is not applicable to subgraph count of RKP networks.

1) *Subgraph count is positively correlated.* If some subgraphs occur, the probability of remaining subgraphs that occur will increase. It is easy to show that the condition amounts to knowing information about the occurrence of some edges, and therefore we are allowed to use incomplete edges to find a subgraph that is non-disjointed.
2) *A physical link is not independent of the existence of other links.* The sensor nodes are deployed on a geometry space and therefore counting of substructures is far from independent due to the triangle inequality. When dealing with a subgraph count in a geometry space, one may assume that $\mathcal{X} = v_1, \ldots, v_n$ are randomly scattered in a $d$-dimension cube of volume $n/\lambda$, where $\lambda$ is a Poisson parameter. Let $|C_i|$ denote the size of the component of $G(\mathcal{X}, \lambda)$ containing vertex $i$. Then we have

$$P\left(n^{-1}\sum_{i=1}^{n}\mathbf{1}[|C_i| = k] \sim p_k(\lambda)\right) \to 1.$$

Note that $p_k(\lambda)$ was defined in [16] as

$$p_{k+1}(\lambda) = (k+1)\lambda^k \int_{(R^d)^k} h(x_1, \ldots, x_k) \exp(-\lambda B(0, x_1, \ldots, x_k))\, dx_1 \ldots dx_k \quad (1)$$

where $h(x_1, \ldots, x_k)$ is the indicator function if $G(0, x_1, \ldots, x_k)$ is connected. Moreover, $B(0, x_1, \ldots, x_k)$ is the volume of the union of 1-balls centered at $0, x_1, \ldots, x_k$. Thus, the number of vertices of $G(\mathcal{X}, \lambda)$ lying in a subgraph of order $k$ divided by $n$ converges to $p_k(\lambda)$.

3) *The secure links are not independent.* The study of random intersection graph $G(n, K, k)$ is motivated by its application to RKP. Generally speaking, the classical RKP to accomplish this is due to Eschenauer and Gligor [17]. Each sensor node is loaded with $k$ distinct encryption keys, randomly taken from a pool of $K$ possible keys, before deployment. Two sensors can form a secure link if they are within wireless communication range and share at least one encryption key.
Most of the subsequent literature model $G(n, K, k)$ by the Erdös-Rényi random graph $G(n, p)$ [10], [13] where

$$p = 1 - ((K-k)!)^2/((K-2k)!K!).$$

Modeling $G(n, K, k)$ in this way is *unsatisfactory* since the behavior of $G(n, p)$ and $G(n, K, k)$ is essentially different. As far as the subgraph count problem on RKP is concerned, the graph model must exploit the fact that many more subgraphs will be in $G(n, K, k)$ than in $G(n, p)$, especially when $k$ is small.

Intuitively, it is complicated to model the subgraph count problem in RKP by combining random geometry graph $G(\mathcal{X}, \lambda)$ with random intersection graph $G(n, K, k)$. This is due to the fact that in both models, the edges involve different levels of dependency. Therefore, the goal in this paper is to study pseudorandom graphs as models for networks with an additional tolerance $\varepsilon$ to control the dependency error.

## III. Preliminaries and Notations

We use $vG$ and $v(G)$ to denote the number of vertices of a graph $G$ and $e_G$ and $e(G)$ to denote the number of edges. We identify a path with $\ell + 1$ vertices by $P^\ell$. For any vertex set $W \subseteq V$, let $G[W]$ denote the graph *induced* by $W$. We denote the number of edges in $G[W]$ by $e_G(W) = e(W)$.

For any pair of possibly disjoint subsets $U, W \subseteq V(G)$, let $E_G(U, W) = E(U, W)$ denote the set of edges in $G$ with one endpoint in $U$ and the other endpoint in $W$. We also define $e_G(U, W) = e(U, W) := |E_G(U, W)|$.

The expression $|\Gamma(v)$ refers to the neighborhood of a vertex $v$ in $G$. Since we often have to consider the neighborhood into a subset $W \subseteq V$, we abbreviate $\Gamma(v) \cap W$ by $\Gamma_W(v)$. Accordingly, we define $d_W(v) := |\Gamma_W(v)|$, which is the degree of vertex $v$ into the set $W$. The neighborhood spanned by a set of vertices $Q$ is denoted by $\Gamma(Q)$; that is,

$$\Gamma(Q) := \bigcup_{v \in Q} (\Gamma(v).$$

For neighborhoods inside specific partitions, we will use the abbreviations $\Gamma_i(v) := \Gamma_i(v) \cap V_i$ and $d_i(v) = |\Gamma_i(v)|$. For $F \subseteq E$, we let $\Gamma_i^F(v) = \{u \in V_i | \{u, v\} \in F\}$ and define $d_i^F(v)$ accordingly. Moreover, let $\Gamma^i$ denote the neighborhoods in $V_i$.

The Szemerédi's regularity lemma is a powerful tool for tackling many problems in combinatorics and graph theory [18], [19], [20], [21]. The notion of *$\varepsilon$-regular* graphs was introduced by Szemerédi in [22]. It describes a property that is typically observed in random bipartite graphs. Therefore,

$\varepsilon$-regular graphs are frequently referred to as pseudorandom graphs. Those graphs have gained importance since it was shown in [22], which states that *any* sufficiently large graph allows for a nearly complete partitioning into constantly many vertex classes of equal size such that most of them are pairwise $\varepsilon$-*regular*. In general, the regularity lemma imposes an additional structure on very large graphs, which can be used to identify certain small substructures like graphs of fixed size.

To simplify our discussion that follows, the following notations are used. Let $e(U,W)$ be the number of edges between $U$ and $W$ where $U,W \in V$ are two disjoint subsets. Furthermore,

$$d_p(U,W) = \frac{e_H(U,W)}{p|U||W|}. \tag{2}$$

We say that the pair $(U,W)$ is $(\varepsilon,p)$-*regular*, if for all $U' \subset U$ and $W' \subset W$ with $|U'| \geq \varepsilon|U|$ and $|W'| \geq \varepsilon|W|$, we have

$$|d_p(U',W') - d_p(U,W)| \leq \varepsilon.$$

Below, we shall sometimes use the expression $\varepsilon$-*regular with respect to density* $p$ to mean that $(U,W)$ is an $(\varepsilon,p)$-regular pair. If $B = (U,W;E)$ is a bipartite graph and $(U,W)$ is an $(\varepsilon,B,p)$-regular pair, we say that $B$ is an $(\varepsilon,p)$-*regular bipartite graph*.

Intuitively, a bipartite graph $(V_i \cup V_2, E)$ is $(epsilon)$-*regular* if its edges are distributed in a random-like way. The parameter $\varepsilon$ reflects the uniformity of this distribution. The smaller the $\varepsilon$, the more uniform the edge distribution in $G$.

**Definition 1.** For a path $P^\ell$, let $\mathcal{G}(\ell,n,m)$ be the family of graphs on vertex set $V = \bigcup_{u \in V(P^\ell)} V_u$, where the sets $V_u$ are pairwise disjoint sets of vertices of size $n$, and edge set $E = \bigcup_{\{u,v\} \in E(P^\ell)} E_{uv}$ where $E_{uv} \subseteq V_u \times V_v$ and $|Exy| = m$. Let $\mathcal{G}(H,n,m,e) \subseteq \mathcal{G}(\ell,n,m)$ denote the set of graphs in $G(\ell,n,m)$ satisfying that each $(V_u \cup V_v, E_{xy})$ is an $(\varepsilon)$-*regular* graph.

**Definition 2.** Let $\mathcal{B}(\ell,n,m,\delta) \subseteq \mathcal{G}(\ell,n,m)$ denote the subfamily consisting of all members that contain fewer than

$$(1-\delta)n^{\ell+1}\left(\frac{m}{n^2}\right)^\ell$$

copies of $P^\ell$.

Observe that apart from the factor $(1-\delta)$, if the edges are randomly distributed between the vertex sets of any graph in $\mathcal{G}(H,n,m)$, we expect

$$n^{\ell+1}\left(\frac{m}{n^2}\right)^\ell$$

copies of $P^\ell$ in any member of this family.

## IV. Main Results

Given a arbitrary graph $G$ (WSN) with $n$ vertices (sensor nodes) and $m$ edges (links), the behavior of the ***RKP*** sensor network can be taken as a *random subgraph* $S_{G_p} = S(G,p)$ of $G$ by keeping each edge of $G$ with probability $p$ independently.

Our main result of this paper is the proof of Theorem 2, a counting version of the KLR-Conjecture, in the case when $H$ is the path $P^\ell$ of size $\ell$ and $m \geq Cn^{(\ell+1)/\ell}$.

**Theorem 1.** (Bad graphs) For $3 \leq \ell \leq \log n$ and every $\delta,\beta > 0$, there exists $\varepsilon_0$ such that for all $\varepsilon \leq \varepsilon_0$, there exists a constants $C$ such that

$$|\mathcal{B}(\ell,n,m,\delta) \cap \mathcal{G}(\ell,n,m)| \leq \beta^m \binom{n^2}{m}^\ell,$$

for $m \geq Cn^{(\ell+1)/\ell}$.

**Theorem 2.** (Bad extension) For $3 \leq \ell \leq \log n$ and every $\delta,\beta > 0$, there exists $\varepsilon_0$ such that for all $\varepsilon \leq \varepsilon_0$, there exists a constant $C$ such that

$$|\mathcal{G}\left(\ell,n,dn^2,\varepsilon : \mathcal{B}(\ell,n,m,\varepsilon')\right)| \leq \beta^m \binom{n^2}{m}^\ell,$$

for $m \geq Cn^{(\ell+1)/\ell}$.

Theorem 1 asserts that the number of bad graphs is very small while Theorem 2 states that there are at most $\beta^m\binom{n^2}{m}^\ell$ graphs belonging to $\mathcal{G}(\ell,n,dn^2,\varepsilon)$ that could be extended by bad $(\varepsilon',p)$-*regular* graphs.

The following corollary asserts that for a given graph $G \in \mathcal{G}(n,m,\varepsilon)$, at least $(1-\varepsilon)n$ vertices can establish a secure path to its $\ell$-hop neighbors.

**Corollary 3.** For $3 \leq \ell \leq \log n$ and every $\delta,\beta > 0$, there exists $\varepsilon_0$ such that for all $\varepsilon \leq \varepsilon_0$, $p \gg 1/\log n$, there exists a constant $C$ such that every ***RKP*** subgraph satisfying $p \geq Cn^{-(\ell-1)/\ell}$ has all but at most $\varepsilon n$ nodes that can establish a secure path to its $\ell$-hop neighbors with probability $1-o(1)$.

### A. Counting Lemmas

The objects to be counted will be referred to as "bad," because later we will identify them with substructures of graphs for which the occurrence of a complete subgraph cannot be guaranteed. These structures must be shown to occur very rarely in order to prove Corollary 3. Since almost all vertices in an $(\varepsilon,d)$-*regular* graph have $\Theta(m/n)$ neighbors, one can deduce that the neighborhood of $\Theta(n^2/m)$ vertices typically has size $\Theta(m/n \cdot n^2/m) = \Theta(n)$. Furthermore, the neighborhoods of the vertices have disjoint parts of size $\Theta(m/n)$; that is, every vertex contributes an equal part to the combined neighborhood of all vertices. In the sequel, we give a formal definition of this structure [18].

**Definition 3.** (Covering vertices) Let $P \subseteq V_i$. $I_j(P;\nu)$ be defined as an (ordered) subset of $\tilde{P} \subseteq P$ with maximum cardinality such that the following property is satisfied. Let $\tilde{P} = \{v_1,\ldots,v_k\}$. Then, there exist pairwise disjoint sets $W_1 \subseteq \Gamma_j(v_1),\ldots,W_k \subseteq \Gamma_j(v_k)$ with $|W1| = \cdots = |W_k| = q_\nu := (1-\nu)n$. For $P \subseteq V_i$ the sets $(W_i)_{i=1}^k$ are called *covering neighborhoods*.

**Definition 4.** ($\nu$-*cover*) Consider a graph $G \in \mathcal{G}(\ell,n,m,\varepsilon)$ and $\nu > 0$. A set $P \subseteq V_i$ with $|P| = p_\nu := \nu/(3d)$ is called a $\nu$-*cover* of $V_j$ if $|I_j(P;\nu)| \geq (1-\varepsilon')p_\nu$ such that $|\Gamma_j(P)| \geq (1-\nu)n$.

**Definition 5.** ($\nu$-*multicovers*) Let $\nu,\varepsilon' > 0$. We call a set $Q \subseteq Vi$ a $\nu$-*multicover* of $V_j$ if there exist pairwise disjoint

subsets $P_1, \ldots, P_r \subseteq Q$ with $r =:= |Q|/p_\nu$ and $|P_1| = \cdots = |P_r| = p_\nu$ such that $P_i$ is a $\nu$-cover of $V_j$, $i = 1, \ldots, r$.

We define two kinds of bad structures contained in each $G \in \mathcal{B}(\ell, n.m, \varepsilon)$.

**Definition 6.** (Bad graphs) Let $\alpha > 0$,

(a) for $1 \le i \le \ell$, there exist sets $W_i \subset V_i$ of size $|W_i| \le \alpha n$ such that for $1 \le i \le k$ and $1 \le j \le \alpha n/x$, there exist less than $(1-\alpha)n/x$ pairwise disjoint subsets $X_1, \ldots, X_{\ell+1}$ that form an $(\varepsilon'', p)$-*regular* for each pair.

(b) for $\ell \le i \le \ell+1$, there exists a set $X \subset V_\ell$ of size at least $\delta n$ such that for all $q \ge Cp^{-(\ell-1)/\ell}$, at least $\beta^q \binom{n}{q}$ are not $\nu$-*multicover*

Now we construct all possibilities of bad events for $\ell$-tuple $V_1, \ldots, V_\ell$. Let $Q_1$ be a subset in $V_1$. A set $\varepsilon'' n \le Q' \subseteq Q_1$ is called bad if $|\Gamma(Q', Q_2)| \le (1-\delta)x$. The following lemma states that if we choose a bad set $Q$ with respect to $v$ within $V_1$, then the number of such sets that are allowed to be selected is very small.

**Lemma 1.** (Count (a)) For $\beta, \nu, \alpha > 0$, there exists $\varepsilon_0 = \varepsilon(\beta, \nu) > 0$ such that for $\varepsilon \le \varepsilon_0$, all but at most $\beta^y \binom{x^2}{y}$ partitions of $V_i$ into subsets $Q_i^j$ with size $(1+\varepsilon')x$, $1 \le j \le \alpha n/x$ contain at most

$$\beta^x \binom{n}{(1+\varepsilon')x}^{\ell+1}$$

bad $\ell$-*tuples* $(Q_i^j, \ldots, Q_\ell^j)$.

**Lemma 2.** (Count (b)) For $\beta, \nu, \varepsilon' > 0$, there exists a constant $\varepsilon_0 = \varepsilon_0(\beta, \nu, \varepsilon')$ such that for $\varepsilon \le \varepsilon_0$, any pair $(V_i, V_j)$ satisfies the number of sets with size $s$ that are not $\nu$-*multicover* is at most

$$\beta^s \binom{n}{s},$$

provided that $m^{(\ell+1)/\ell} \le n^2/4$ and $p_\nu \le s \le q$.

Observe that if all families of disjoint sets $Q_i$ that satisfy a certain undesired property satisfy $\sum_i |Q_i| \le \delta n$, then we can delete at most $\delta n$ vertices such that none of them have this bad property.

## V. The Proofs

Theorem 1 is an easy consequence from Lemma 1 and Lemma 2, we omit its proof here. Theorem 2 is more complicated to prove. Our proof strategy is as follows. First, we define a family of *bad* graphs to be excluded from the family $\mathcal{G}(\ell, n, m, \varepsilon)$. Then, we shall show that there are merely a tiny fraction of all graphs in $\mathcal{B}(\ell, n, m, \delta) \cap \mathcal{G}(\ell, n, m, \varepsilon)$ that can be extended to an $(\varepsilon, d)$-*regular* graph.

*Proof:* (of Theorem 2) Let $x \ge \frac{n^2}{m} \log n$ and $\varepsilon' \frac{m}{n^2} \le y \le (1-\varepsilon')\frac{m}{n^2}x^2$. In particular, Lemma 1 shows that all but a $\beta^x$ fraction of all tuples are *good*. We prove Theorem 2 by constructing all graphs that satisfy the above properties. First, we select $m$ edges between the $(\varepsilon', p)$-*regular pairs* $(V_i, V_j)$, $1 \le i < j \le \ell$. There are at most

$$\binom{n^2}{m}^{\ell-1}$$

ways to do that. Second, we choose pairwise disjoint sets $Q_i^j$, $j \le \ell$. There are less than

$$\binom{n}{x}^{\ell\alpha n/x + (\alpha n/x)^2} \le n^{\ell\alpha n + (\alpha n)^2/x}$$

ways to do so.

Next we distribute the $y$ edges between the pair $(Q_i^j, Q_i^j)$. There are at most

$$\begin{aligned}
&\prod_{1\le i\le i'\le \ell-1} \prod_{j=1}^{\alpha n/x} \binom{|E(Q_i^j, Q_{i'}^j)|}{y} \\
&\le \prod_{1<i<i'<\ell-1} \binom{\sum_{j=1}^{\alpha n/x} |E(Q_i^j, Q_{i'}^j)|}{\frac{\alpha n}{x} y} \\
&\le \prod_{1<i<i'\le \ell-1} \binom{m}{\frac{\alpha n}{x} y} \qquad (3)
\end{aligned}$$

ways to choose the graph $G_k^j$.

Then, we want to choose extensions of $G_k^j$ to satisfy $(\varepsilon, d)$-*regular*. As these extensions have to belong to $\mathcal{G}(\ell - 1, x, dx^2, \varepsilon)$, the numbers of ways to select the neighborhoods of vertices in $V_j$ is thus at most

$$\begin{aligned}
&\prod_{v\in B} \left(\beta^x \binom{n}{(1+\varepsilon')x}\right) \prod_{v\notin B} \binom{n-(1+\varepsilon')x}{d_j(v)-(1+\varepsilon')x} \\
&\le \beta^{\varepsilon' nx} \prod_v \left(4^{d_j(v)} \binom{n}{d_j(v)}\right) \\
&\le \beta^{\varepsilon' nx} 4^m \prod_v \binom{n}{d_j(v)} \\
&\le \beta^{\varepsilon' nx} 4^m \binom{(d-\varepsilon)n^2}{m} \\
&\le \beta^{\varepsilon' nx} 4^m (d/2)^m \binom{n^2}{m} \\
&\le \hat{\beta}^m \binom{n^2}{m} \qquad (4)
\end{aligned}$$

Similarly, there are at most $\beta'^m \binom{n^2}{m}$ possibilities to choose the pair $(V_\ell, V_{\ell+1})$. Then, we show that there are at most

$$\hat{\beta}^m \binom{n^2}{m} + \beta'^m \binom{n^2}{m} = \tilde{\beta}^m \binom{n^2}{m}$$

ways to extend it to $V_\ell + 1$.

Put all of these together, we obtain the upper bound

$$\begin{aligned}
&\binom{n^2}{m}^{\ell-1} n^{\ell\alpha n + (\alpha n)^2/x} \prod_{1<i<i'\le k} \binom{m}{\frac{\alpha n}{x} y} \tilde{\beta}^m \binom{n^2}{m} \\
&\le 2^{(\ell\alpha n^2/x)\log n + \ell^2 m} \tilde{\beta}^m \binom{n^2}{m} \\
&\le 2^{\ell\alpha m + \ell^2 m} \tilde{\beta}^m \binom{n^2}{m} \qquad (5)
\end{aligned}$$

Let $\beta'' = 2^{\ell\alpha m+\ell^2 m}\tilde{\beta}^m$. We complete the proof. ∎

Now we prove Corollary 3.

*Proof:* First of all, let $d$ and $\varepsilon'$ be given. We set

$$\beta = \left(\frac{1}{4}\right)^{1/2\ell}\left(\frac{d}{e}\right)^{2}\left(\frac{1}{e}\right)^{2\ell/d}.$$

By applying $\beta$ and $\varepsilon'$ to Theorem 2, we obtain $\varepsilon_0$ and $C$. Then let $\varepsilon = \min\{d/2, \varepsilon'/4, \varepsilon_0\}$. Suppose $V = \{V_1, V_2, \ldots, V_\ell\}$ is an $\ell$-partite partition graph from $\mathcal{B}(P^\ell, n, T, \varepsilon', \delta)$. We shall show that an *RKP* subgraph of $\mathcal{B}$ is unlikely to appear in $S^{(r)}_{G_p}$. Since the bipartite subgraph $S_G(V_i, V_j)$ contains at least $(d-\varepsilon)pn^2$ edges each. By the definition of $\mathcal{F}$, there is a set $W \subset V_1$ with $|W| \le \delta n$ such that $|\Gamma_1(\Gamma_2(\ldots\Gamma_{\ell-1})| \le (1-\delta)n$. Then, Theorem 2 infers that there are at most

$$\beta^m\binom{n^2}{m}$$

possibilities for choosing all such subgraphs. Now the number of all possible graphs belonging to $(\varepsilon, d)$-*regular* are at most

$$\sum_{(d-\varepsilon)n^2\ge m\ge(p-\varepsilon')n^2}\binom{N}{n}^{\ell+1}\binom{(d-\varepsilon)n^2}{m}^{\ell}.$$

Hence, the probability of the random subgraph $S_G \in \mathcal{F}$ is bounded from above by

$$\begin{aligned}
&\sum_{\substack{(d-\varepsilon)\\ \ge m \ge \\ (p-\varepsilon')n^2}}\binom{N}{n}^{\ell+1}\binom{(d-\varepsilon)n^2}{m}^{\ell}\beta^m\binom{n^2}{m}^{\ell}p^{2m\ell}\\
&\le \sum_{\substack{(d-\varepsilon)\\ \ge m \ge \\ (p-\varepsilon')n^2}}\left(\frac{eN}{n}\right)^{(\ell+1)n}\left(\frac{e(d-\varepsilon)n^2p}{m}\right)^{2m\ell}\left(\beta^{1/\ell}\frac{e}{d-\varepsilon}\right)^{m\ell}\\
&\le n^{2\ell}\left(\frac{eN}{n}\right)^{(\ell+1)n}\left(e^{1/d}\left(\frac{e}{d-\varepsilon}\right)^{1/2}\beta^{1/2\ell}\right)^{2dpn^2\ell}\\
&\le exp(2\ell\log n + \ell n\log N - 4\ell n^{-1-1/\ell})\\
&\to o(1), \qquad (6)
\end{aligned}$$

The result now follows as noted by Markov inequality. ∎

Now we prove Lemma 1.

*Proof:* (of Lemma 1) Since we are supposed to build a bad pair of $(Q_1, Q_2)$ that satisfies condition (a), there are at least $q'/2$ indices $i$ such that $|\Gamma(v_i, Q)| \le (1-\delta)x/(q'/2)$. Let $\delta \le 2^{-3\varepsilon''n}$. Since there only $\delta x$ choices for $v_i$, the number of bad sets $Q'$ is at most

$$\binom{q'}{\lceil q'/2\rceil}\delta^{q'/2}\frac{x^{q'}}{q'!} \le 2^{q'}\delta^{\varepsilon''n/2}e^{q'}\binom{n}{\varepsilon''n} \le \delta^{\varepsilon''n}\binom{n}{\varepsilon''n} \quad (7)$$

where $\varepsilon''n \le q' \le n/4$.

Since that subsets $Q'_1, \ldots, Q'_r$ of size $x$ are chosen sequentially, any choice of a bad subset $Q'_i$ may depend on the previously chosen subsets $(Q'_1, \ldots, Q'_{i-1})$. Moreover, at least $\delta r = \delta\alpha n/x$ sets $Q'_i$ must be bad and do not depend on the order of choice. Hence, the probability of a subset $Q \ge \varepsilon' x$ being bad is at most

$$2^r\left(\frac{\delta^{\varepsilon''n}\binom{n}{\varepsilon''n}}{\binom{n-\varepsilon''n-x}{x}}\right)^{\alpha n/x} \le 2^r(\delta e)^{\varepsilon''n\alpha r} \le \delta^{\alpha x/2} \quad (8)$$

In other words, there are at most $\delta^{\alpha x/2}\binom{n}{(1+\varepsilon')x}$ bad sets $Q$. Then, the proof follows. ∎

Finally comes the proof of Lemma 2.

*Proof:* (of Lemma 2) Let $Q$ denote the family of sets of size $s$ in $V_i$ that is a *$\nu$-multicover* of $V_j$, $1 \le i < j \le \ell+1$. First, Lemma 1 states that the probability of a randomly chosen set $X \subseteq Q$ with $|X| = p_\nu$ is at most $\beta^{p_\nu}\binom{n}{p_\nu}$. Moreover, at least $\delta r$ sets $X_i$ must be bad and do not depend on the order of choice. Hence, we first count the number of sets $Q$ that contain $\delta r$ bad subsets. We start by partitioning the set $Q$ into $r := \lfloor s/p_\nu \rfloor$ disjoint sets of size $s$ and an additional set of size $s - rp_\nu$. Hence, we have at most

$$\begin{aligned}
&\binom{r}{\delta r}(\beta^{p_\nu})^{\delta r}\left(\prod_{i=0}^{r-1}\binom{n-ip_\nu}{p_\nu}\right)\binom{n-rp_\nu}{s-rp_\nu}\\
&\le \left(\frac{e}{\delta}\right)^{\delta r}\beta^{s/2}\frac{s!}{p_\nu!^r(s-rp_\nu)!}\binom{n}{s} \qquad (9)
\end{aligned}$$

ways to do that. Since there are $\frac{s!}{p_\nu!^r(s-rp_\nu)!}\binom{n}{s}$ ways to partition, the lemma follows. ∎

## VI. Simulation

As depicted in Figs. 1, 2, and 3, the threshold phenomenon of the property we desire "every node has a secure path to $\ell$-hop neighbors" is given in this section, with a minor modification. Let AEVP denote the property "at least 1-$\varepsilon$ node has a secure path to the $\ell$-hop neighbors." One can see that the probability that satisfies AEVP is rapidly increased to 1. In what follows, we use $n$ to denote the number of nodes and $d$ the average degree. Although the main results in Section IV show that the probability is $1 - o(1)$. Here one can observe that the probabilities in the simulation have exponential decay behaviors. The reason is similar to the difference between binomial model and uniform model. In other words, we fix the number of edges and choose each bad graph uniformly at random in our proofs. However, for simulations, we take each pair of vertices to be an edge with probability $p$ independently. Consequently, the probabilities in the simulations are equivalent to those in the main results.

## VII. Conclusion and Future Work

In this paper, we showed that pseudorandom subgraphs with size smaller than the density of the given $\varepsilon$ graphs still inherit good property with high probability. Our main contribution is determined a desired RKP probability $p$ such that every vertex can find a secure path to its $\ell$-hop neighbors, provided that $p \ge Cn^{-(\ell-1)/\ell}$ and some appropriate constants are selected. The most challenging part is when the sensor network is rather sparse. In this case, the number of all possible RKP subgraphs is too large to attack the exponential bounds. Thus approaches based on large deviation inequality generally seem to fail to

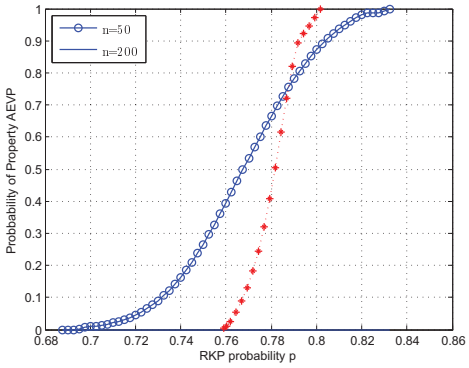


Fig. 1. Different number of nodes with $d = n \log^2 n$, $\ell = 2$, $\varepsilon = 0.1$.

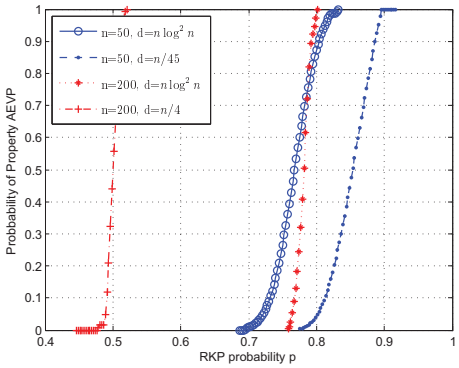


Fig. 2. Different number of nodes compared to different degree with $\ell = 2$, $\varepsilon = 0.1$.

provide the results we described here (see Rödl's work for more details). To see why this may be expected, the expected value of path $P^\ell$ has the order of $O(n)$ in a sparse case, rather than $O(n^2)$ in the dense case, while the number of sets we need to control is $\exp(\Omega(n))$, and most common large deviation inequalities need such density $d$ in Theorem 1 to be large compared with $p$. In the future, it seems valuable to take some geometric properties into account. The WSN application is based on many previous researches. In recent years, lots of key distribution schemes for heterogeneous WSN

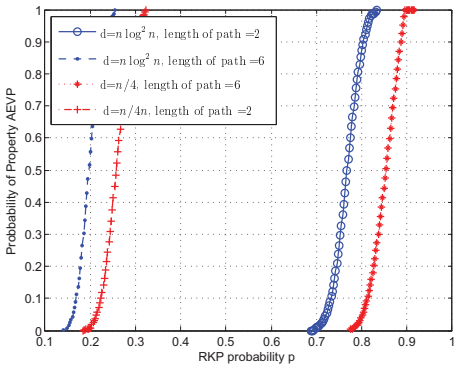


Fig. 3. All parameter are the same except the length of path.

are presented, thus modifying the regularity lemma to fit the heterogeneous/multiple environment could be more suitable for WSN.

## ACKNOWLEDGMENT

This work was supported in part by National Science Council, Taiwan, R.O.C., under Contracts NSC100-2219-E-006-001 and NSC100-2218-E-041-001-MY2.

## REFERENCES

[1] K. C. Rahman, "A Survey on Sensor Network," *Journal of Computer and Information Technology*, vol. 1, pp. 76–87, 2010.
[2] S. Tilak, N. B. Abu-Ghazaleh, and W. Heinzelman, "A taxonomy of wireless micro-sensor network models," *ACM SIGMOBILE, Mobile Computing and Communication Review*, vol. 6, no. 2, pp. 28–36, 2002.
[3] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
[4] M. Tubaishat and S. Madria, "Sensor networks: an overview," *IEEE Potentials*, vol. 22, no. 2, pp. 20–23, 2003.
[5] K. Akkaya and M. F. Younis, "A survey on routing protocols for wireless sensor networks," *Ad Hoc Networks*, vol. 3, no. 3, pp. 325–349, 2005.
[6] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer Networks*, vol. 52, no. 12, pp. 2292–2330, 2008.
[7] X. Du and Y. Xiao, "A survey on sensor network security," in *Wireless Sensor Networks and Applications*, Y. Li, M. T. Thai, and W. Wu, Eds. Springer US, 2008, pp. 403–421.
[8] Y. Wang, G. Attebury, and B. Ramamurthy, "A survey of security issues in wireless sensor networks," *IEEE Communications Surveys and Tutorials*, vol. 8, no. 1-4, pp. 2–23, 2006.
[9] X. Chen, K. Makki, K. Yen, and N. Pissinou, "Sensor network security: a survey," *IEEE Communications Surveys & Tutorials*, vol. 11, no. 2, pp. 52–73, 2009.
[10] H. Chan, A. Perring, and D. Song, "Random key predistribution schemes for sensor network," *IEEE Symposium on Research in Security and Privacy*, pp. 197–213, 2003.
[11] R. Pietro, L. Mancini, and A. Mei, "Random key assignment secure wireless sensor networks," *ACM workshop on Security of Ad Hoc and Sensor Networks*.
[12] W. Du, J. Deng, Y. Han, P. Varshney, and J. Katz, "A pairwise key predistribution scheme for wireless sensor networks," *ACM Transactions on Information and System Security (TISSEC)*, 2005.
[13] J. Huang and Y. Kim, "Revisiting random key pre-distribution for sensor networks," *ACM Workshop on Security of Ad Hoc and Sensor Networks*, 2004.
[14] W.-S. Li, T.-S. Su, and W.-S. Hsieh, "Multi-neighbor random key pre-distribution: A probabilistic analysis," *IEEE Communication Letters.*, vol. 13, pp. 306–308, May 2009.
[15] W.-S. Li and W.-S. Hsieh, "Performance analysis of connectivity on multi-hop random key pre-distribution," *New Trends in Information and Service Science, NISS*, pp. 1325–1330, 2009.
[16] M. D. Penrose, "Random geometric graphs," *OxfordStudies in Probability. Oxford University Press,*, vol. 5, May.
[17] L. Eschenauer and V. Gligor, "A key-management scheme for distributed sensor networks," *The Computer and Communications Security, ACM Press*, pp. 41–47, 2002.
[18] T. Schickinger, "Complete subgraphs of random graphs, ph.d. thesis," *Department of Computer Science, Techical University Munich, Germany*, 2002.
[19] S. Gerke, Y. Kohayakawa, V. Rödl, and A. Steger, "Small subsets inherit sparse $\varepsilon$-regularity," *Journal of Combinatorial Theory, Series B*, vol. 97, pp. 34–56, 2007.
[20] Y. Kohayakawa and V. Rödl, "Regular pairs in sparse random graphs I," *Recent Advances in Algorithmic Combinatorics (B. Reed and C. Linhares-Sales, eds.), CMS Books Math./Ouvrages Math. SMC, Springer, New York*, vol. 11, pp. 289–351, 2003.
[21] Y. Ishigami, "The algorithmic aspects of the regularity lemma," *Preprint*.
[22] E. Szemerédi, "Regular partitions of graphs, j.-c. bermond, j.-c. fournier, m. las vergnas, d. sotteau (eds.), problèmes en combinatoire et thèorie des graphes," *Colloque Inter. CNRS*, pp. 399–401, 1978.