On Optimality of Key Pre-distribution Schemes for Distributed Sensor Networks

Subhas Kumar Ghosh*

October 12, 2006

Abstract

We derive the optimality results for key pre distribution scheme for distributed sensor networks, and relations between interesting parameters. Namely, given a key-pool of size n we derive the optimal value that is jointly achievable for parameters like, Size optimality: using less memory per node - while supporting large network, Connectivity optimality: possibility of establishing secure communication between any two nodes over short path, and Resiliency optimality: large fraction of network remains working under compromise or node capture. We characterize this relation in graph theoretic framework. Our result shows that the desired graph (a combination of network topology graph on which key-share graph is embedded) must have small clique and independent set and must have high expansion properties, in other words Expander graphs are best suited for forming secure networks.

Keywords: key pre distribution scheme, distributed sensor networks, expander graph, combinatorial design, random key distribution

1 Introduction

Security is an important issue in deploying distributed sensor networks (DSN). One of the approaches to establish secure communication is constructing protocol based on pre-distributed keys. Basic model we consider here is key pre-distribution mechanism (**KPS**) as defined by Eschenauer and Gligor [13] (see Section-2). Several extensions and variations can be found in the works of Chan, Perrig, and Song [10], Liu and Ning [17], Du, Deng, Han and Varshney[11], which considers random key pre-distribution mechanism (*randomized*-**KPS**). These works extended the basic random key assignment schemes using conference key distribution schemes of Blom [4] and Blundo et el [5]. A random subset assignment key predistribution scheme, and a hypercube-based key predistribution scheme was studied in [19]. A closest pairwise keys predistribution scheme and a location-based pairwise keys scheme which takes advantage of sensors' expected locations was described in [18].

Under the basic model of key pre-distribution mechanism for a secure sensor network, Çamtepe and Yener [6] considered a deterministic pre-key distribution mechanism (*deterministic*-**KPS**). Çamtepe and Yener's method uses block design techniques in combinatorial design theory. Very similar approaches based on combinatorial design theory are proposed in [16, 15] along with probabilistic approaches yielding hybrid designs (*hybrid*-**KPS**) to support arbitrary network sizes. Recently Chakrabarti, Maitra and Roy [9] considered combinatorial design followed by a probabilistic merging applied to key pre-distribution in sensor nodes. They used a transversal design to construct a configuration and then used random merging of blocks to form sensor nodes providing flexibility in adjusting the number of common keys between any two nodes.

All previous works derive bounds on properties of the network under respective schemes. However, it is not clear what properties are desired, and what are possible bounds on achievable parameters, as well as their

*Honeywell Technology Solutions Laboratory, 151/1, Doraisanipalya, Bannerghatta Road, Bangalore, India, 560076, Email:subhas.kumar@honeywell.com

relations. This is precisely what we analyze in this work. In this work we consider both *randomized*-**KPS** and *deterministic*-**KPS** and look at their optimality for three most desired set of parameters, namely, (*i*) using less memory per node - while supporting large network, (*ii*) possibility of establishing secure communication between any two nodes in the network over a short path (possibly $\mathcal{O}(\log n)$, where n is the order of the network), and (*iii*) resiliency under compromise or node capture. In this work we derive optimality results for key pre-distribution scheme, and trade-off relations between parameters.

1.1 Contribution

In this paper, we focus on key pre-distribution problem for DSN. We first define the efficiency and security properties of a key pre-distribution procedure. Then we show how these parameters map to properties of shared key graph combined with network graph. Our main contribution comes from deriving optimality results for key pre-distribution scheme on this combined graph expressed as a function of properties of the resulting graph. We are able to show that if one requires a key pre-distribution mechanism which is both efficient and secure then combined network graph with shared key graph need to have good expansion properties.

1.2 Remainder of this paper

In section-2 we define some basic terminologies. We describe random key pre-distribution scheme and show how a set system or combinatorial design can be used as a deterministic key pre-distribution scheme. In section-3 we derive optimality results for key pre-distribution scheme. Following which we compare the derived parameters in this work with other known results in section-4.

2 Key Pre-distribution Mechanism

2.1 Randomized KPS

First we describe the random key pre-distribution schemes. Random key pre-distribution scheme works by selecting a *pool of keys* X from some pre-specified *key-space* of size $|X| = n$. Each node is then assigned a random subset of keys generated by sampling X by a fixed number of times l. The idea is that after deployment, any two nodes can initiate communication with the common key if they have one. Thus a *randomized*-**KPS** is defined by a set X and a family of subsets of X, $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$, where $\forall i, |A_i| = l$ is the size of *key-ring*.

2.2 Deterministic KPS

Now we describe deterministic key pre-distribution schemes. A *set system* or *combinatorial design* $(X, \mathcal{A})$ consists of a finite set X of elements called *points*, $X = \{x_1, x_2, \ldots, x_n\}$, and a family of subsets of X, i.e. $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$, with $A_j \subseteq X, \forall j : 1 \leq j \leq m$ called *blocks*. If all blocks are of same size, say l, then $(X, \mathcal{A})$ is said to be *uniform design* (of rank l). A set system $(X, \mathcal{A})$ can be used to design a *deterministic*-**KPS** for DSNs as follows. Let us denote the sensor nodes by $\mathcal{U} = \{u_1, u_2, \ldots, u_m\}$. We then identify each block $A_i \in \mathcal{A}$ with one sensor node $u_i \in \mathcal{U}$, and we identify the ground set X as the set of n keys. Then, for $1 \leq j \leq m$, sensor node u_i receives the set of keys in block A_i.

2.3 Vital Parameters

Let $(X, \mathcal{A})$ be a **KPS** (deterministic or randomized). As noted in previous section, several proposals exists for key pre-distribution in DSN, however, with a possible loss in specific cases, one could unify them. Almost all **KPS** has three important design objectives, namely: (1) A pre-distribution of shares, (2) Shared key discovery, and establishing path keys when two nodes do not share a key, (3) Resiliency of the network under node capture. We have already described the step pre-distribution of shares. In shared key discovery

phase nodes interact in their neighborhood to assess possible common shares. Hence, possibility of finding a neighbor having common key depends on distribution scheme, as well as network topology. Objective of path key establishment is to have a secure communication link between a pair of nodes which do not share a common key, but through their neighbors with which they have common key they can establish a secure path. For resiliency observe that when a sensor node is compromised all its l keys will become unusable. We need to consider how many secure links will still exist in the network. Assume u_a, u_b and u_c are three nodes sharing a common key. If u_c is compromised then u_a and u_b can no longer communicate, as their only common key is assumed to be compromised now, however they may still be able to establish a path key (with extra communication cost). In sequel we formally define these parameters and conditions under which we say optimality is achieved.

2.4 Optimality Considerations

For any of the above key pre-distribution schemes, parameters we will be interested about are from two directions, *efficiency*, and *security*. Efficiency in fact captures several aspects of the design. First, for every network of size $m > 1$ we must be able to provide a construction using as small share per-node as possible. We capture this in the following definition:

Definition 2.1. (Share Size Optimal KPS): *A key pre-distribution scheme $(X, \mathcal{A})$ with $|X| = n$, and $|\mathcal{A}| = m$, where the size of each set $A_i \in \mathcal{A}$ is l will be calld a (m, l, n)-***KPS***. For a fixed $|X| = n$, a (m, l, n)-***KPS*** is share size optimal if it achieves minimum l and maximum m.*

Few comments are in place about Definition-2.1. Note that X is identified with the key-pool. Hence, idea of defining optimality with respect to fixed $|X| = n$ and achieving smallest l and maximum m implies supporting maximum size network (m) using smallest share size per node (l). A more constructive definition is indeed possible, where one may require that, given m, l, n and i one can generate the share for ith node A_i efficiently (possibly in time polynomial in $(\log m, l, n)$, however we will not consider the optimality issue from computational perspective in this work.

Second aspect of an efficient design is the ability to form a secure network. As described above this has two objectives, having shared key in the neighborhood, and having path key between any pairs. Two sensor nodes u_i and $u_j \in \mathcal{U}$, share a key if and only if $A_i \cap A_j \neq \emptyset$, and they are neighbors on the network graph. We would like to have at least one node in the neighborhood for a node with which it shares a key.

Condition to ensure the existence of path key between any pair of nodes is slightly complicated. To be able to establish a path key between u_i and $u_j \in \mathcal{U}$ we must have following: Let $P_{i,j}$ be the set of all possible paths between u_i and u_j, then to be able to establish a path key there must be at least one path (denoted by natural sequence of vertices on network graph) $p = u_i u_k \dots u_j \in P_{i,j}$ such that there is a common key between every consecutive pair of vertices in p. Formally:

Definition 2.2. (Connectivity Optimal KPS): *Let $(X, \mathcal{A})$ be a* **KPS** *(deterministic or randomized). Then we shall call $(X, \mathcal{A})$ a (ϵ, k)-connectivity optimal* **KPS** *if for any pair of nodes u and v, the probability that they can establish a path-key over shortest path of length k is greater than ϵ, where the probability is taken over all pairwise vertices of the network graph.*

Final consideration is the security. Again looking at the key sharing graph and the network graph, we can estimate that any chosen pair of nodes will not have a secure link when a fraction of nodes are compromised in the network.

Definition 2.3. (Resiliency Optimal KPS): *Let $(X, \mathcal{A})$ be a* **KPS** *(deterministic or randomized). Then for $\gamma, \delta \in \mathbb{R}$ and $0 \leq \gamma, \delta \leq 1$ we shall call $(X, \mathcal{A})$ a (γ, δ)-resiliency optimal* **KPS** *if with probability greater than γ remaining $(1 - \delta)$ fraction of the network can establish secure connection even when a δ fraction of nodes are compromised.*

Following Table (Table-1) lists the parameters.

Table 1: Design Parameters

Parameter	Meaning	Objective
n	Key-pool size	Fixed
m	Network size	Maximize
l	Share size	Minimize
ϵ	Key sharing probability	Maximize
k	Key path length	Minimize
γ	Resiliency	Maximize
δ	Fraction compromised	Maximize

3 Optimality of KPS

We have described the set of desired properties of key-pre-distribution scheme for a DSN. Next we will derive the condition that achieves the optimal values for the set of parameters as described in definition-2.1, definition-2.2, and in definition-2.3. In following we shall denote the key sharing graph by $\mathcal{G}$ with a associated set-system $(X, \mathcal{A})$. We shall also by $\mathcal{H}$ denote the network graph of the sensor nodes. Graph $\mathcal{G}$ has vertex set $\mathcal{A}$, and two vertex A_i and A_j are connected by an edge iff $A_i \cap A_j \neq \emptyset$.

Modeling $\mathcal{H}$ is somewhat delicate. Since sensor nodes are randomly deployed in a geographic region, it might appear a random graph might be suitable as a network model. In their work Eschenauer and Gligor [13] considered random graph model and used classical result of Erdös and Rényi [12] on connectivity. Erdös and Rényi model allows one to relate local connectivity(i.e., the probability that two nodes are connected) and the global connectivity (i.e., the probability that the whole network is connected). However, model used in [13], and many other subsequent works on *randomized*-**KPS** has been questioned recently ([14, 21]), as most of these works assumed that the underlying physical network is dense enough to enable their key pre-distribution to be effective. This issue can be stated as follows: assuming sensors are deployed in a two-dimensional plane, by placing each node uniformly and independently at a random location, let us also assume that nodes can transmit at various power levels, then a combination of these two factor determines the relation between global vs. local connectivity. As an example, if we assume that the given network is connected - what can we say about the degree of each node?

To characterize $\mathcal{H}$ we will use a result of Xue and P. R. Kumar [22]. Let $\mathcal{S}$ be a unit square in $\mathbb{R}^2$, and suppose n nodes identified with set $V : |V| = m$ are placed uniformly and independently in $\mathcal{S}$. Then $\mathcal{H}$ is the network graph formed when each node is connected to its neighborhood. More precisely, there exists an edge between $u, v \in V$ if when $u \in \mathcal{N}(v)$, where $\forall u \in V, \mathcal{N}(u) \stackrel{\Delta}{=} \{v : (u, v) \in E\}$. We also assume this implies $v \in \mathcal{N}(u)$, in other words we consider undirected network graphs. If $\forall u \in V, |\mathcal{N}(u)| = \phi_m$, then main result by Xue and P. R. Kumar can be stated as:

Theorem 3.1. [22] :*For $\mathcal{H}(m, \phi_m)$ to be asymptotically connected, $\Theta(\log m)$ neighbors are necessary and sufficient. Specifically, there are two constants $0 < c_1 < c_2$ such that:*

$$\lim_{m\to\infty} \mathbf{Pr}\left[\mathcal{H}(m, c_1 \log m) \text{ is disconnected }\right] = 1 \text{ and} \quad (1)$$

$$\lim_{m\to\infty} \mathbf{Pr}\left[\mathcal{H}(m, c_2 \log m) \text{ is connected }\right] = 1. \quad (2)$$

It is possible to choose $c_1 = 0.074$ and $c_2 = 5.1774$, however the critical value of the constant is unknown.

With Theorem-3.1,when we write $\mathcal{H}$, we will essentially mean $\mathcal{H}(n, \phi_m)$, where ϕ_m is of order $\Theta(\log m)$. Given $\mathcal{G} = (V, E)$ and $\mathcal{H} = (V, T)$ (note, both of them are defined on the same set of vertices), to "embed" $\mathcal{G}$ over $\mathcal{H}$ we consider the product $\mathcal{G} \cdot \mathcal{H}$ - defined as the graph whose vertex set is the Cartesian product $V \times V$ in which vertices (u, v) and (u', v') are adjacent if u, u' are either equal or adjacent in $\mathcal{G}$ and v, v' are either equal or adjacent in $\mathcal{H}$. Observe that when (u, u') has non-empty intersection on $\mathcal{G}$ and (v, v') are adjacent on $\mathcal{H}$, (u, v) and (u', v') has a secure link on $\mathcal{G} \cdot \mathcal{H}$.

3.1 Outline of Proof

Now observe that in order to establish the optimality of the desired parameters, we need to express them in terms of the properties of the graph $\mathcal{G} \cdot \mathcal{H}$. Our proof is based on techniques used in [1] and also following methods and tools used by Alon in [3]. This method (dimension argument) can be described as follows. We wish to bound the size of some finite combinatorially defined set of objects A. To do this, we first map the elements of A to a linearly independent set of vectors in some vector space V. Then we apply the dimension argument: $|A| \leq \dim V$. On this direction we will take following steps:

1. First we will require a common mechanism to represent $\mathcal{G}$ and $\mathcal{H}$. We will do this by considering a subspace of the space of polynomials in m variables over $\mathbb{R}$. A representation of a graph of order m over such subspace is an assignment of a polynomial f_v to each vertex v along with a point of evaluation of the polynomial c_v in $\mathbb{R}^m$ to each vertex such that $f_v(c_v) \neq 0$ but if $(u, v) \in E(G)$ then $f_v(c_u) = 0$. To combine $\mathcal{G}$ and $\mathcal{H}$ we consider the Tensor product of corresponding vector spaces which represents $\mathcal{G}$ and $\mathcal{H}$, and obtain a combined graph as tensor product $\mathcal{G} \cdot \mathcal{H}$.

2. Once we have a representation of the graph $\mathcal{G} \cdot \mathcal{H}$ we derive the desired parameters expressed in terms of the vector space representing $\mathcal{G} \cdot \mathcal{H}$. A simple proposition suggests that in order to ensure any two node have a secure key-path, they must be in same connected component of the product graph $\mathcal{G} \cdot \mathcal{H}$.

3. We consider $\mathcal{G} \cdot \mathcal{H}$ as a d-regular undirected graph (we justify this choice). Now we observe that a graph is connected iff its second largest eigenvalue of the adjacency matrix λ_2 is greater than 0 (first eigenvalue λ_1 is same as the degree d), and $\lambda_1 > \lambda_2$.

4. This characterization allows us to choose l optimally so that the graph $\mathcal{G} \cdot \mathcal{H}$ is connected, and based on which we derive other parameters.

3.2 Combining Graphs

In following to present our results, we first provide a geometric view of the combined key sharing graph $\mathcal{G}$ and network graph $\mathcal{H}$. Let $G = (V, E)$ be a graph of order m and let $\mathcal{F}$ be a subspace of the space of polynomials in m variables over $\mathbb{R}$. A representation of G over $\mathcal{F}$ is an assignment of a polynomial $f_v \in \mathcal{F}$ to each vertex $v \in V$ along with an assignment of a point $c_v \in \mathbb{R}^m$ to each $v \in V$ such that following two conditions hold:

1. For each $v \in V$, $f_v(c_v) \neq 0$.

2. If $(u, v) \in E(G)$ then $f_v(c_u) = 0$.

We need few notations first. Let $[n] = \{1, 2, \ldots, n\}$. With each set $A_i \subseteq [n]$ we associate its characteristic vector $v_i = (v_{i1}, v_{i2}, \ldots, v_{in}) \in \mathbb{R}^n$ where $v_{ij} = 1$ if $j \in A_i$ and $v_{ij} = 0$ otherwise. For $x, y \in \mathbb{R}^n$, let $x \cdot y = \sum_{i=1}^{n} x_i y_i$ denote their standard inner product.

Lemma 3.1. *Key-sharing graph $\mathcal{G}(m, l, n)$ has a representation over $\mathcal{P}$, where $\mathcal{P}$ is a subspace of the space of polynomials in n-variables of degree at most $l - 1$ over $\mathbb{R}$.*

Proof. Let $(X, \mathcal{A})$ be the **KPS**. $\mathcal{G}$ has vertex set $V(\mathcal{G})$ where each $v \in V(\mathcal{G})$ can be identified with a subset $A_i \subseteq X$, and $|A_i| = l$. Let $A_1, A_2, \ldots, A_m$ be the vertex set of $\mathcal{G}$. By construction, in the vertex set of $\mathcal{G}$, between vertex A_i and A_j there is an edge iff $|A_i \cap A_j| \neq \emptyset$. Using notations defined above clearly we have $v_i \cdot v_j = |A_i \cap A_j|$, where v is the characteristic vector of set A. For $i = 1, \ldots, m$, for each $A_i \in \mathcal{A}$ let us define the following polynomials $P_i(x_1, \ldots, x_n)$ on n variables:

$$\forall i : 1 \leq i \leq m, P_i(x_1, \ldots, x_n) \triangleq \prod_{k=1}^{l-1} \left(\sum_{j \in A_i} x_i - k \right)$$

For each set A_i let us assign polynomial P_i and a point $c_i = v_i$, the characteristic vector of the set A_i to vertex i. Clearly,

$$\begin{aligned} \forall i : 1 \leq i \leq m, P_i(c_i) &\neq 0, \text{ and} \\ \forall i,j : 1 \leq i \neq j \leq m, \text{ and } A_i \cap A_j \neq \emptyset, P_i(c_j) &= 0. \end{aligned}$$

Let p_i be the multilinear polynomial obtained from the standard representation of P_i as a sum of monomials by using, repeatedly, the relations $x_i^2 = x_i$. Since the vectors $c_i = v_i$ have $\{0,1\}$ coordinates, $p_i(c_j) = P_i(c_j)$ for all A_i and A_j, and graph $\mathcal{G}$ has a representation over $\mathcal{P}$, where $\mathcal{P}$ is a subspace of the space of polynomials in n-variables. Also observe that the degree of multilinear polynomial $\deg(p_i) \leq l-1$, completing the proof. □

For a graph $G = (V, E)$ define neighborhood graph $NB(G)$ as: $\forall i \in V(G)$ let $\mathcal{N}_i = \{j : (i,j) \in E(G)\}$, then $NB(G) = (V_{NB}, E_{NB}) : V_{NB} = \{\mathcal{N}_i\}_{\forall i \in V(G)}$ and $(\mathcal{N}_i, \mathcal{N}_j) \in E_{NB} \iff \mathcal{N}_i \cap \mathcal{N}_j \neq \emptyset$.

Lemma 3.2. *Let $\mathcal{H}(m, \phi_m)$ be a network-graph, then $NB(\mathcal{H}(m, \phi_m))$ has a representation over $\mathcal{Q}$, where $\mathcal{Q}$ is a subspace of the space of polynomials in m-variables of degree at most $\phi_m - 1$ over $\mathbb{R}$.*

Proof. Recall, ϕ_m is the cardinality of neighborhood for every vertex $i \in V(\mathcal{H})$, and observe that for the graph $\mathcal{H}$ to be disconnected, following must hold:

1. $\forall i \in V(\mathcal{H}), |\mathcal{N}_i| = \phi_m$, and
2. There exists at least two connected component $H_1 = (V_1, E_1)$ and $H_2 = (V_2, E_2)$, such that $V(\mathcal{H}) = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$, and $E(\mathcal{H}) = E_1 \cup E_2$.

For each vertex $i \in V(\mathcal{H})$ define neighborhood of i as the set A_i. For the neighborhood of each vertex $1 \leq i \leq m, \mathcal{N}_i$ let us define the following polynomials $Q_i(y_1, \ldots, y_m)$ on m variables:

$$\forall i : 1 \leq i \leq m, Q_i(y_1, \ldots, y_m) \triangleq \prod_{t=1}^{\phi_m - 1} \left(\sum_{j \in \mathcal{N}_i} y_j - t \right)$$

For each set $\mathcal{N}_i$ let us assign polynomial Q_i and a point $d_i = v_i$, the characteristic vector of the set $\mathcal{N}_i$. Lemma follows using similar argument as above considering space of multilinear polynomials q_i of degree $\deg q_i \leq \phi_m - 1$. □

Lemma-3.2 is for the neighborhood graph of $\mathcal{H}$. While we need to show that $\mathcal{H}$ has a similar property. Following lemma asserts that.

Lemma 3.3. *$\mathcal{H}(m, \phi_m)$ has a representation over $\mathcal{Q}$ iff $NB(\mathcal{H}(m, \phi_m))$ has a representation over $\mathcal{Q}$ and $\mathcal{H}(m, \phi_m)$ is connected, moreover $\mathcal{Q}$ is a subspace of the space of polynomials in m-variables of degree at most ϕ_m over $\mathbb{R}$.*

Proof. By Lemma-3.2, $NB(\mathcal{H}(m, \phi_m))$ has a representation over $\mathcal{Q}$ as

$$\left\{ \prod_{t=1}^{\phi_m - 1} \left(\sum_{j \in \mathcal{N}_i} y_j - t \right), d_i : i \in V(\mathcal{H}) \right\}$$

Then we can represent $\mathcal{H}$ over $\mathcal{Q}$ by considering

$$\left\{ Q_i' = y_i \cdot \prod_{t=1}^{\phi_m - 1} \left(\sum_{j \in \mathcal{N}_i} y_j - t \right), (0, 0, \ldots, 1, \ldots, 0)^T + d_i : i \in V(\mathcal{H}) \right\}$$

for ith vertex in $V(\mathcal{H})$, only ith entry of the vector $v_i' = (0, 0, \ldots, 1, \ldots, 0)^T$ is 1 and other entries are 0. Observe that $Q_i'(v_j' + d_j) \neq 0 \iff i = j$. Also, polynomials are defined over the space $\mathcal{Q}$ and have degree at most ϕ_m. □

Lemma-3.1 and Lemma-3.3 provides a characterization of the graphs $\mathcal{G}$ and $\mathcal{H}$ as a vector space of polynomials $\mathcal{P}$ and $\mathcal{Q}$ over the same field, and we can combine them now considering the space spanned by the polynomials $p(x_1, x_2, \ldots, x_n) \cdot q(y_1, y_2, \ldots, y_m)$ where $p \in \mathcal{P}$ and $q \in \mathcal{Q}$ using *tensor product* $\mathcal{P} \otimes \mathcal{Q}$ of the two vector spaces. We have the following lemma.

Lemma 3.4. *[3]: Let $\mathcal{G} = (V, E)$ and $\mathcal{H} = (V, T)$ be two graphs. Assume $\mathcal{G}$ has a representation $\{p_v(x_1, x_2, \ldots, x_l), c_v : v \in V\}$ over $\mathcal{P}$ and $\mathcal{H}$ has a representation $\{q_u(y_1, y_2, \ldots, y_r), d_u : u \in V\}$ over $\mathcal{Q}$, where $\mathcal{P}$ and $\mathcal{Q}$ are spaces of polynomials over the same field $\mathbb{R}$.*
Then $\{p_v \cdot q_u, c_v d_u : (v, u) \in V \times V\}$ is a representation of the graph product of $\mathcal{G}$ and $\mathcal{H}$ as $\mathcal{G} \cdot \mathcal{H}$ over $\mathcal{P} \otimes \mathcal{Q}$, where $c_v d_u$ denotes the concatenation of c_v and d_u.

Proof. Observe, for every $((u, v), (u', v')) \in V \times V$,

$$p_v \cdot q_u(c_{v'} d_{u'}) = p_v(c_{v'}) \cdot q_u(d_{u'}).$$

Product is non-zero when $c_v d_u = c_{v'} d_{u'}$ and it is zero when $((u, v), (u', v'))$ is an edge. Hence this is indeed a representation of $\mathcal{G} \cdot \mathcal{H}$ over $\mathcal{P} \otimes \mathcal{Q}$. □

3.3 Expressing Parameters by Properties of Product Graph

While, we have a characterization of the graph $\mathcal{H}$ to be connected by Theorem-3.1, we need to establish such criteria for the graph $\mathcal{G} \cdot \mathcal{H}$ in terms of ϕ_m and l. Observe that, by Theorem-3.1, choosing ϕ_m we can ensure connectivity of the network graph $\mathcal{H}$ and also there exists a choice of l (though smallest l which will ensure connected graph might not be trivial to derive) which ensures the connectivity in key-sharing graph $\mathcal{G}$, we need to establish a condition on the product graph $\mathcal{G} \cdot \mathcal{H}$ such that any two node can establish a secure key-path. Following proposition provides a necessary condition for that.

Proposition 3.1. *Let $\mathcal{H}$ be the network graph and let $\mathcal{G}$ be the key-sharing graph. On $\mathcal{H}$ any two vertices u and u' can establish a secure key path $p = uu_1 \ldots u_k u'$ such that there is a common key between every consecutive pair of vertices in p iff product graph $\mathcal{G} \cdot \mathcal{H}$ has a connected component $T = V_T, E_T$ and vertices (v, u) and (v', u') such that both $(v, u), (v', u') \in V_T$.*

Proof. Proof of this proposition is straightforward. First note that a connected component of a graph induces a partition on its vertex set. Now, between two vertices (v, u) and (v', u') on the product graph there is an edge, iff (v, v') is an edge in $\mathcal{G}$ and (u, u') is an edge in $\mathcal{H}$. Extending it for path is simple and we omit it. □

We can use the simple proposition for choosing ϕ_m (we will do that as per Theorem-3.1), with a bound on l so that $\mathcal{G} \cdot \mathcal{H}$ is connected. However, to make things simpler, we would like the graph $\mathcal{G} \cdot \mathcal{H}$ to be d-regular, for some $d > 0$. Recall, a graph G is d-regular when every vertex in G has exactly d neighbors. Justification for this choice can be found in the following proposition:

Proposition 3.2. *For every integer $m, d > 0$, if there exists a graph on m vertices with average degree $\delta(G) = d$, then there exists a d-regular graph G' which is at least as much optimal (connectivity, and resiliency) as much G is.*

Proof. Let G' be a d-regular graph which is (ϵ', k')-connectivity optimal, and (γ', δ')-resiliency optimal. We need to show that for any G with average degree $\delta(G) = d$, such that G is (ϵ, k)-connectivity optimal, and (γ, δ)-resiliency optimal. Then when $k = k'$ we have $\epsilon \leq \epsilon'$. Similarly, when $\delta = \delta'$ we have $\gamma \leq \gamma'$.
Connectivity:

$$\begin{aligned}
\epsilon &= \mathbf{Pr}_{u,v \in G}[\exists \text{ a path of length } k \text{ between } u, v] \\
&= \mathbf{Pr}_{u,v \in G}\left[(u, v) \in E\left(G^k\right)\right] \\
&= \mathbf{Pr}_{u,w \in G}\left[(u, w) \in E\left(G^{\lceil k/2 \rceil}\right)\right] \mathbf{Pr}_{u,w \in G}\left[(w, v) \in E\left(G^{\lfloor k/2 \rfloor}\right)\right] \\
&= \left(\mathbf{Pr}_{u',v' \in G}[(u', v') \in E(G)]\right)^k \\
&\leq \left(\mathbf{Pr}_{u',v' \in G'}[(u, v) \in E(G')]\right)^k \leq \left(\mathbf{Pr}_{u,v \in G'}\left[(u, v) \in E\left(G'^k\right)\right]\right) \leq \epsilon'
\end{aligned}$$

Where, the third line follows by considering the probability that there exists a $t < k$ and a vertex w such that $(u,w) \in E\left(G^t\right)$ and $(w,v) \in E\left(G^{k-t}\right)$, and repeating it ($G^r$ is the rth power of G, where every r length path in G is an edge in G^r). Inequality follows from the fact that $\mathbf{Pr}_{u,v\in G}\left[(u,v) \in E\left(G\right)\right] \leq \mathbf{Pr}_{u,v\in G'}\left[(u,v) \in E\left(G'\right)\right]$, by noting a regular graph induces uniform probability distribution and G has average degree equal to d. A similar result can be obtained for resiliency. □

Now our objective is to choose a bound on l so that graph $\mathcal{G} \cdot \mathcal{H}$ is connected. We have $\phi_m = \Theta\left(\log m\right)$ and we will assume that the combined graph is regular, and we are sure by Proposition-3.2 that this will not be any loss of generality.

Lemma 3.5. *If $\phi_m = \Theta\left(\log m\right)$ and $l \geq \mathcal{O}\left(1\right)$ then graph $\mathcal{G} \cdot \mathcal{H}$ is a (m, d, λ)-connected graph with number of vertice $m = \left(en/l\right)^l$, degree $d = 2l \log en/l$, and expansion ratio $1 - \lambda$.*

Proof. We first present a result for a graph $G = (V, E)$ of order m which is d-regular, and then extend it to our graph $\mathcal{G} \cdot \mathcal{H}$. Let A_G be the $m \times m$ symmetric adjacency matrix of a graph G. A_G is real and symmetric, and has m real eigenvalues which we denote by $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$. An associated orthonormal system of eigenvectors are $v_1, \ldots, v_m$ with $Av_i = \lambda_i v_i$. Note that $\lambda_1 = d$, and it is obtained for all one vector $\mathbf{1}$. Note that eigenvalues are closely related to connectedness of a graph. A graph is connected iff $\lambda_1 > \lambda_2$.

Let $S \subset V\left(G\right)$ be a subset of the vertex set. A cut of G is $S \subset V\left(G\right)$ such that $\left|E\left(S, \bar{S}\right)\right| = e\left(S, \bar{S}\right) = 0$, where for any $S \subset V\left(G\right)$ by $\bar{S}$ we denote the set $V \setminus S$, and $E\left(X, Y\right)$ denotes the set of cross edges between $X, Y \subset V\left(G\right)$. We note that a graph is connected iff it has no cuts.

Consider vector $\mathbf{v} = \left|\bar{S}\right| v_S - \left|S\right| v_{\bar{S}}$, where $v_X \in \{0,1\}^m$ denotes the characteristic vector of set X. Clearly $\mathbf{v} \perp \mathbf{1}$, and

$$\left\|\mathbf{v}\right\|^2 = \left|\bar{S}\right|^2 \left|S\right| + \left|S\right|^2 \left|\bar{S}\right| = \left|\bar{S}\right| \left|S\right| \left(\left|S\right| + \left|\bar{S}\right|\right) = m \left|S\right| \left|\bar{S}\right|$$

Also,

$$\mathbf{v}A\mathbf{v}^T = 2\left(\left|E\left(S\right)\right| \left|\bar{S}\right|^2 + \left|E\left(\bar{S}\right)\right| \left|S\right|^2 - \left|S\right| \left|\bar{S}\right| \left|E\left(S, \bar{S}\right)\right|\right)$$

As, G is d-regular so substituting $2\left|E\left(S\right)\right| = d\left|S\right| - \left|E\left(S, \bar{S}\right)\right|$ and $2\left|E\left(\bar{S}\right)\right| = d\left|\bar{S}\right| - \left|E\left(S, \bar{S}\right)\right|$, we obtain:

$$\mathbf{v}A\mathbf{v}^T = md\left|S\right| \left|\bar{S}\right| - m^2 \left|E\left(S, \bar{S}\right)\right|$$

This allows us to compute λ_2 by computing the *Rayleigh quotient*:

$$\lambda_2 \geq \frac{\mathbf{v}A\mathbf{v}^T}{\left\|\mathbf{v}\right\|^2} = d - \frac{m\left|E\left(S, \bar{S}\right)\right|}{\left|S\right| \left|\bar{S}\right|} \geq d - 2\left|E\left(S, \bar{S}\right)\right| / \left|S\right|, \text{ with } \left|S\right| \leq m/2$$

Now observe that we have $\lambda_1 = d > \lambda_2 \geq d - 2\left|E\left(S, \bar{S}\right)\right| / \left|S\right|$, in order to ensure that the graph is connected. Hence for every $S \subseteq V\left(G\right), \left|S\right| \leq m/2$, we must ensure that the term $2\left|E\left(S, \bar{S}\right)\right| / \left|S\right|$ is away from 0, but less than d. Define

$$h\left(G\right) \triangleq \min_{S:\left|S\right| \leq m/2} \left|E\left(S, \bar{S}\right)\right| / \left|S\right|.$$

as the expansion ratio of graph G. Define a graph G with $\lambda_2/d \leq \lambda : \lambda < 1$ as (m, d, λ)-graph G. We have following claim:

Claim 3.1. *If $\mathcal{G}$ is an (m, d_g, λ_g)-graph and $\mathcal{H}$ is an (m, d_h, λ_h)-graph, then $\mathcal{G} \cdot \mathcal{H}$ is an $(m^2, d_g d_h, \max\left(\lambda_g, \lambda_h\right))$-graph.*

Proof. Considering the normalized adjacency matrix of graph $\mathcal{G}$ and $\mathcal{H}$ and observing that normalized adjacency matrix of graph $\mathcal{G} \cdot \mathcal{H}$ is the tensor product of these two matrices. Thus eigenvalues are pairwise products of eigenvalues of $\mathcal{G}$ and $\mathcal{H}$. Largest eigenvalue is thus $\mathbf{1} \cdot \mathbf{1}$, and second largest eigenvalue must be either $\mathbf{1} \cdot \lambda_h$ or $\lambda_g \cdot \mathbf{1}$. □

To obtain a bound on l we must relate $h(\mathcal{G}\cdot\mathcal{H})$ to l. By the last paragraph, it will sufficient for us to consider $\mathcal{G}$ as an (m, d_g, λ_g)-graph with $\lambda_g \geq \lambda_h$. Let d be the degree of $\mathcal{G}\cdot\mathcal{H}$, and choose $d_g = d/d_h = d/\log m$, note that this choice is feasable by Lemma-3.4. Thus we have $h(\mathcal{G}\cdot\mathcal{H}) \geq d(1-\lambda_g)/2\log m$. On the other hand if $l < n/2$ size subsets of $[n]$ are chosen, then degree of any vertex on $\mathcal{G}$ can be at most

$$\sum_{i=1}^{l-1}\binom{n}{i} \leq l\left(\frac{en}{l}\right)^l$$

We can possibly support large network (i.e. have a large m) if we allow large degree, but our objective will be to have a low degree graph with good expantion ratio. This implies d must be of order $\mathcal{O}(\log m)$. This gives us the (m, d, λ)-connected graph with number of vertice $m = (en/l)^l$, degree $d = 2l\log(en/l)$, and expansion ratio $1-\lambda$, when $l = \mathcal{O}(1)$.

□

3.4 Optimality Results

Now we establish bounds on the desired parameters. Intuitively, on the product graph $\mathcal{G}\cdot\mathcal{H}$ vertices can establish secure link directly (resp. by a key path) if they have an edge (resp. they are in same connected component). Notice that all vertices that belongs to a clique on product graph $\mathcal{G}\cdot\mathcal{H}$ has pair-wise secure link. Thus cliques on this graph ensure more number of secure connectivity. On the other hand when a vertex is compromised, all other vertices belonging to the same clique are also compromised, thus larger independent set ensures more resiliency. This seems contradictory - but not so, when graph has small cliques and many of them. Concretely, if there is no small set S for which $\mathcal{G}\cdot\mathcal{H}\setminus S$ has one very large component and many small ones then both requirements are met. In other words we show that graph having neither large clique, nor large independent set are best for product graph $\mathcal{G}\cdot\mathcal{H}$ if they also have good expansion ratio. We note the following fact (known as Expander Mixing Lemma) implying large expansion ratio implies the graph is nearly random:

Lemma 3.6. *Let $G = (V, E)$ be a d-regular graph. Then for all $S, T \subseteq V(G)$:*

$$\left|\,|E(S,T)| - \frac{d\,|S|\,|T|}{m}\right| \leq \lambda_2\sqrt{|S|\,|T|}. \tag{3}$$

Now we need to answer the following question concerning definition-2.2: for a randomly chosen pair of vertices u and v, what is the probability that they have a path of length k.

Lemma 3.7. *Let $\mathcal{G}\cdot\mathcal{H}$ be a $\left((en/l)^l, 2l\log(en/l), \lambda\right)$ graph, then corresponding* **KPS** *is $(\mathcal{O}(\log n), \mathcal{O}(1))$ - connectivity optimal.*

Proof. The distance $d(u, v)$ between vertices u and v in a graph $G = (V, E)$ is the length of the shortest path between them. The diameter of G can be defined as $\max_{u,v}\{d(u, v)\}$. Also $B(u, r) = \{v : d(u, v) \leq r\}$, is the ball or radius r around u. We claim that an (m, d, λ)-graph G has diameter $\mathcal{O}(\log m)$. This follows from fact that $|B(u, r)| > m/2$ for every vertex u and some $r \in \mathcal{O}(\log m)$. This, in turn follows from G's expansion properties. Namely, we show that $|B(u, r+1)| > (1+\epsilon)|B(u, r)|$ for some fixed $\epsilon > 0$ as long as $|B(u, r)| \leq m/2$. We have by Lemma-3.6 that $|E(S, S)|/|S| \leq d\cdot(|S|/m + \lambda)$ for every set S. Therefore, $\left|E\left(S, \bar{S}\right)\right|/|S| \geq d\cdot((1-\lambda) - |S|/m)$. But S has at least $\left|E\left(S, \bar{S}\right)\right|/d$ neighbors outside of itself, so the claim follows with $\epsilon = 1/2 - \lambda$.

Now to compute connectivity parameter, we note following:

$$\mathbf{Pr}_{u,v}[u \text{ and } v \text{ has path of length } k \leq \mathcal{O}(\log m)] \geq \mathcal{O}(1)$$

□

Similarly, for resiliency:

Lemma 3.8. *Let $\mathcal{G} \cdot \mathcal{H}$ be a $\left((en/l)^l, 2l\log(en/l), \lambda\right)$ graph, then corresponding* **KPS** *is $((1-\delta/[(1-\lambda)\, l \log n]), \delta)$-resiliency optimal for all $\delta > 0$.*

Proof. It is sufficient for us to to estimate the probability that δm is a cut on the product graph $\mathcal{G} \cdot \mathcal{H}$, i.e.:

$$\begin{aligned} \gamma &= \mathbf{Pr}_{u,v}\left[u, v \text{ has path when } \delta \text{ fraction of nodes are compromised }\right] \\ &\geq 1 - \mathbf{Pr}\left[\delta m \text{ is a minimum cut }\right] \\ &= 1 - \frac{\delta m}{md(1-\lambda)} \end{aligned}$$

□

Finally, we combine Lemma-3.5, 3.7 and 3.8 to obtain the following theorem:

Theorem 3.2. *There exists a* **KPS** *with*
n size key-pool, $m = \mathcal{O}\left(n^l\right)$ size network, $l \geq \mathcal{O}(1)$ size key ring,
$k = \mathcal{O}(\log n)$ with $\epsilon = \mathcal{O}(1)$,
$\gamma = (1 - \delta/[(1-\lambda)\, l \log n])$ with δ fraction compromised nodes.

4 Concluding Remarks

In this work we have introduced a realistic model to analyze the efficiency, connectivity and security properties of any key pre-distribution mechanism for DSN. Under this model local vs. global secure connectivity properties has been analyzed using novel linear algebraic methods, and optimality trade-off has been expressed in terms of the expansion properties of the underlying graph. This is, to the best of our knowledge first such asymptotic analysis considering joint optimality and trade-off between storage, connectivity, and resiliency at the same time. Our results are existential, and suggests that the product graph $\mathcal{G} \cdot \mathcal{H}$ shall have good expansion properties to achieve optimal parameters. While a graph can be connected with a bottleneck set S which is small and $G \setminus S$ has one large connected component and many small components. This will not be good for constructing a secure sensor networks. As compromising S will be sufficient to ensure a non-functioning network. It is desired that the product graph $\mathcal{G} \cdot \mathcal{H}$ shall be a expander graph(c.f. [2], A graph $G = (V, E)$ is ϵ-edge-expanding if for every partition of the vertex set V into X and $\bar{X} = V \setminus X$, where X contains at most a half of the vertices, the number of cross edges $e\left(X, \bar{X}\right)$ are greater than $\epsilon\,|X|$) . Challenge for DSN is following. One need to design a **KPS** such that when $\mathcal{H}$ is a randomly generated network, embedding $\mathcal{G}$ on $\mathcal{H}$ shall ensure a ϵ-edge-expanding graph $\mathcal{G} \cdot \mathcal{H}$ for some $\epsilon > 0$, and ϵ is bounded away from zero as m, the order of graph grows. Finally, we note that a complementary thought was explored in [20], where authors have shown that some topologies arising naturally in the context of (secure) wireless networking are low-degree, expander graphs. Another interesting recent result strengthens our thought is [8]. In [8], authors have presented a deterministic key distribution scheme based on Expander Graphs. Their paper shows how to map the parameters (e.g., degree, expansion, and diameter) of a Ramanujan Expander Graph to the desired properties of a key distribution scheme for a physical network topology. In other words their work is complementary to this work in exhibiting an explicit example of designing *deterministic*-**KPS** using Expander graphs.

We conclude this section with comparison of parameters derived in this work with some of the existing works. It must be noted that there are significant tradeoffs between the parameters we have discussed. Also it must be noted that each of these works improve on one or more parameters while loosing on the other. What we derive in this work is the optimal value that is jointly achievable for parameters. We refer reader to an excellent survey on key distribution mechanisms by Çamtepe and Yener [7]. Following table (Table-2) lists design parameters derived in this work along with (a) exhaustive pair-wise key distribution scheme, (b) basic probabilistic scheme of [13], (c) random pair-wise scheme of [10], and (d) symmetric BIBD based deterministic scheme of [6].

Table 2: Comparison of Design Parameters

Solution	Ref.	n	m	l	k	ϵ	γ	δ
Exhaustive Pair-wise	Folklore	-	n	$2(n-1)$	d	$\mathcal{O}(1)$	$\mathcal{O}(1)$	any
Probabilistic	[13]	n	-	$2l$	d	$\frac{((n-l)!)^2}{((n-2l)!n!)}$	l/n	-
Random Pair-wise	[10]	n	-	$2np$	d	p	$\mathcal{O}(1)$	any
symmetric BIBD	[6]	n^2+n+1	n^2	$n+1$	d	$\mathcal{O}(1)$	$1/n$	-
-	this work	n	n^l	$\mathcal{O}(1)$	$\log n$	$\mathcal{O}(1)$	$1-\frac{\delta}{\log n}$	δ

Acknowledgments.

Author would like to thank Debapriyay Mukhopadhyay, Ranjeet K. Patro and Satyajit Banerjee for several discussions on these problems. Author would also like to thank the anonymous reviewers of ESAS 2006, for their valuable comments.

References

[1] N. Alon, L. Babai, and H. Suzuki. Multilinear polynomials and Frankl-Ray-Chaudhuri-Wilson type intersection theorems. *J. Comb. Theory Ser. A*, 58(2):165–180, 1991.

[2] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986.

[3] Noga Alon. The shannon capacity of a union. *Combinatorica*, 18(3):301–310, 1998.

[4] R Blom. An optimal class of symmetric key generation systems. In *Proc. of the EUROCRYPT 84 workshop on Advances in cryptology: theory and application of cryptographic techniques*, pages 335–338, New York, NY, USA, 1985. Springer-Verlag New York, Inc.

[5] Carlo Blundo, Alfredo De Santis, Amir Herzberg, Shay Kutten, Ugo Vaccaro, and Moti Yung. Perfectly-secure key distribution for dynamic conferences. In *CRYPTO '92: Proceedings of the 12th Annual International Cryptology Conference on Advances in Cryptology*, pages 471–486, London, UK, 1993. Springer-Verlag.

[6] Seyit Ahmet Çamtepe and Bülent Yener. Combinatorial design of key distribution mechanisms for wireless sensor networks. In *ESORICS*, pages 293–308, 2004.

[7] Seyit Ahmet Çamtepe and Bülent Yener. Key distribution mechanisms for wireless sensor networks: a survey. RPI Technical Report TR-05-07, RPI, 2005.

[8] Seyit Ahmet Çamtepe, Bülent Yener, and Moti Yung. Expander graph based key distribution mechanisms in wireless sensor networks. In *to appear in IEEE International Conference on Communications (ICC) 2006*, 2006.

[9] Dibyendu Chakrabarti, Subhamoy Maitra, and Bimal K. Roy. A key pre-distribution scheme for wireless sensor networks: Merging blocks in combinatorial design. In *ISC*, pages 89–103, 2005.

[10] Haowen Chan, Adrian Perrig, and Dawn Song. Random key predistribution schemes for sensor networks. In *SP '03: Proceedings of the 2003 IEEE Symposium on Security and Privacy*, page 197, Washington, DC, USA, 2003. IEEE Computer Society.

[11] 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.

[12] P. Erdös and A. Rényi. On the evolution of random graphs. *Institute of Mathematics Hungarian Academy of Sciences*, 5:17–61, 1960.

[13] Laurent Eschenauer and Virgil D. Gligor. A key-management scheme for distributed sensor networks. In *CCS '02: Proceedings of the 9th ACM conference on Computer and communications security*, pages 41–47, New York, NY, USA, 2002. ACM Press.

[14] Joengmin Hwang and Yongdae Kim. Revisiting random key pre-distribution schemes for wireless sensor networks. In *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 43–52, New York, NY, USA, 2004. ACM Press.

[15] Jooyoung Lee and Douglas R. Stinson. Deterministic key predistribution schemes for distributed sensor networks. In *Selected Areas in Cryptography*, pages 294–307, 2004.

[16] Jooyoung Lee and Douglas R. Stinson. A combinatorial approach to key predistribution for distributed sensor networks. In *IEEE Wireless Communications and Networking Conference*, volume 2, pages 1200–1205, 2005.

[17] 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.

[18] Donggang Liu and Peng Ning. Location-based pairwise key establishments for static sensor networks. In *SASN '03: Proceedings of the 1st ACM workshop on Security of ad hoc and sensor networks*, pages 72–82, New York, NY, USA, 2003. ACM Press.

[19] Donggang Liu, Peng Ning, and Rongfang Li. Establishing pairwise keys in distributed sensor networks. *ACM Trans. Inf. Syst. Secur.*, 8(1):41–77, 2005.

[20] Alessandro Panconesi and Jaikumar Radhakrishnan. Expansion properties of (secure) wireless networks. In *SPAA '04: Proceedings of the sixteenth annual ACM symposium on Parallelism in algorithms and architectures*, pages 281–285, New York, NY, USA, 2004. ACM Press.

[21] Roberto Di Pietro, Luigi V. Mancini, Alessandro Mei, Alessandro Panconesi, and Jaikumar Radhakrishnan. Connectivity properties of secure wireless sensor networks. In *SASN '04: Proceedings of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pages 53–58, New York, NY, USA, 2004. ACM Press.

[22] Feng Xue and P. R. Kumar. The number of neighbors needed for connectivity of wireless networks. *Wirel. Netw.*, 10(2):169–181, 2004.