Random graph modeling of key predistribution schemes in wireless sensor networks

Osman Yağan

CyLab
Carnegie Mellon University
oyagan@andrew.cmu.edu

Joint work with
Prof. Armand M. Makowski

Wireless sensor networks (WSNs)

- **Distributed** collection of **small** sensor nodes:
  - Have **limited** capability for computations and wireless communications.
  - Gather (security-sensitive) data and control (security-critical) operations.
- **Applications**
  - Military (Battlefield surveillance)
  - Health (Patient monitoring)
  - Home ("Smart" systems, home automation)
  - Environment (Monitoring fires in forests)

WSNs and security

- WSNs are usually deployed in **hostile** environments
  - Communications are monitored, and nodes are subject to capture and surreptitious use by an adversary.
- **Cryptographic protection** is needed to ensure secure communications.
- Scalable solutions with very low storage, management, and computational load are required.

Random key predistribution schemes!

Random key predistribution schemes

- A large set of (cryptographic) keys: **Key pool**.
- For each sensor i, a subset Σ_i is generated by some random mechanism: **Key ring** of i.
  - Inserted in the sensor's memory before deployment.

Principle: Sensors i and j can communicate **securely** if
i) They have a wireless communication link available, **and**
ii) They have at least one key in common, i.e.

$$\Sigma_i \cap \Sigma_j \neq \emptyset.$$

Goal: Evaluate key predistribution schemes

A vast number of different key predistribution schemes have been proposed so far.

- ◇ Differ only in the mechanism that generates random key rings.

Evaluating key predistribution schemes:

- How to **select the parameters** of a given scheme so that certain desired properties hold **with high probability**?
- How do various schemes compare with each other w.r.t. **connectivity, security, memory load,** and **scalability**?

Approach: Random graph modeling

Random graph modeling

Random graphs: Natural models for random key predistribution schemes for wireless sensor networks:

- sensor $\rightarrow$ node, secure link $\rightarrow$?
- **Communication graph:** Eg., the disk model.

$$i \sim j \quad \text{iff} \quad \|\boldsymbol{x_i} - \boldsymbol{x_j}\| < \rho$$

- **Key graph:** Induced by the key predistribution scheme.

$$i \sim j \quad \text{iff} \quad \Sigma_i \cap \Sigma_j \neq \emptyset$$

Intersecting random graphs

System model: Communication graph $\bigcap$ Key graph:

- $i \sim j$ if $\Sigma_i \cap \Sigma_j \neq \emptyset$ and $\|\boldsymbol{x_i} - \boldsymbol{x_j}\| < \rho$
- Many concerns regarding WSNs can be mapped into problems for this system model.

A simple case of interest – Full visibility

- Sensors are all within communication range of each other.
- System model = **Key graph**.
- Allows to focus on the randomized key predistribution.
- Key graph may have applications in other fields.

My dissertation

- The Eschenauer-Gligor (EG) scheme
  - Connectivity under full visibility [ISIT 2008-2009, CISS 2010, **IT 2012**]
  - Connectivity under an on-off channel model [**IT 2012**]
  - Triangle existence and small-world properties [Allerton 2009, GraphHoc 2009, **IT 2013**]
- The pairwise scheme of Chan, Perrig and Song
  - Connectivity under full visibility [ISIT 2012, **IT 2012**]
  - Connectivity under an on-off channel model [ICC, **IT 2013**]
  - Scalability (gradual deployment) [WiOpt 2011, **Perf Eval 2013**]
  - Security [PIMRC 2011, TISSEC 2013]

The punch line

	EG Scheme	Pairwise Scheme
Connectivity (Full Visibility)	$\lvert\Sigma\rvert = O(\log n)$	$\lvert\Sigma\rvert_{n,\mathrm{Avg}} = O(1)$ $\lvert\Sigma\rvert_{n,\mathrm{Max}} = O(\sqrt{\log n})$
Connectivity (On-Off Channel, p_n)	$\lvert\Sigma\rvert = O(\frac{\log n}{p_n})$	$\lvert\Sigma\rvert_{n,\mathrm{Avg}} = O(\frac{\log n}{p_n})$ $\lvert\Sigma\rvert_{n,\mathrm{Max}} = O(\frac{\log n}{p_n})$
Gradual Deployment	$\surd$	$\lvert\Sigma\rvert_{n,\mathrm{Avg}} = O(\log n)$ $\lvert\Sigma\rvert_{n,\mathrm{Max}} = O(\log n)$
Unassailability	$\lvert\Sigma\rvert = O(\sqrt{n \log n})$	$\lvert\Sigma\rvert_{n,\mathrm{Avg}} = O(1)$ $\lvert\Sigma\rvert_{n,\mathrm{Max}} = O(\sqrt{\log n})$
Unsplittability	$\lvert\Sigma\rvert = O(\sqrt{n \log n})$	$\lvert\Sigma\rvert_{n,\mathrm{Avg}} = O(w_n)$ $\lvert\Sigma\rvert_{n,\mathrm{Max}} = O(\sqrt{w_n \log n})$
Perfect Resiliency	$\times$	$\surd$
Node Authentication	$\times$	$\surd$

$|\Sigma|$: # of keys required $\quad w_n$: Any function satisfying $\lim_{n\to\infty} w_n = \infty$.

Today, we focus on the connectivity results for the Eschenauer-Gligor scheme

1. "A zero-one law for connectivity in random key graphs"
   - O. Yağan and A. M. Makowski, *IEEE Trans. Inf. Theory* **58**(5): 2983-2999, May 2012.

2. "Performance of the Eschenauer-Gligor key distribution scheme under an ON-OFF channel"
   - O. Yağan, *IEEE Trans. Inf. Theory* **58**(6):3821-3835, June 2012.

Eschenauer-Gligor (EG) scheme

Before network deployment, each node **randomly** selects a set of K **distinct** keys from a (very large) pool of P keys.

- n(# of nodes), P(key pool size), K(size of each key ring).
- $\Sigma_1, \ldots, \Sigma_n$ iid and uniform in $\mathcal{P}_K$.

$\mathcal{P}_K$: Collection of all subsets of $\{1, \ldots, P\}$ with size K.

EG graph $\mathbb{K}(n;\theta)$: Arises under the full visibility assumption.
$\theta \equiv (K, P)\,,\ V = \{1, \ldots, n\}\,,\ E = \{i \sim j : \Sigma_i \cap \Sigma_j \neq \emptyset\}$

$$\mathbb{P}\left[i \sim j\right] = 1 - \frac{\binom{P-K}{K}}{\binom{P}{K}} := 1 - q(\theta) \simeq \frac{K^2}{P}$$

Connectivity Results
Under Full Visibility

Connectivity of the EG graph (YM 2008)

Theorem 1 *Consider a scaling* $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ *such that*

$$\frac{K_n^2}{P_n} \sim c \frac{\log n}{n}$$

for some $c > 0$. *If there exists some* $\sigma > 0$ *such that* $\sigma n \leq P_n$, *then*

$$\lim_{n \to \infty} \mathbb{P}\left[\ \mathbb{K}(n;\theta_n) \text{ is connected } \right] = \begin{cases} 0 & \text{if } c < 1 \\ & \\ 1 & \text{if } c > 1. \end{cases}$$

Blackburn & Gerke (2008): Theorem 1 with $P_n = o(n)$.

Rybarcyzk (2009): Theorem 1 **without** any constraint on P_n.

Why sharp zero-one laws are useful?

- Connectivity is at **odds** with other network properties.
- Recall that $\mathbb{P}[i \sim j] \simeq \frac{K^2}{P}$
- To increase the chances of connectivity;
  - ◇ Increase K $\Rightarrow$ Larger key rings, **larger memory** req.
  - ◇ Decrease P $\Rightarrow$ Larger K/P ratio, **less resiliency** against node capture attacks.
- Sharp zero-one laws provide a **precise** threshold of connectivity.
  - ◇ Knowing the exact minimum requirements for ensuring connectivity, one can dimension the scheme without suffering performance losses in other properties.

Connectivity Results
Under non-full visibility

A simple communication model

- Assume that communication channels are mutually independent, and each channel is either **on** or **off**.
- With p in $(0,1)$, consider i.i.d. $\{0,1\}$-valued rvs with success probability p.
  - ◇ The channel between nodes i and j is available with probability p and unavailable with the complementary probability $1-p$.
  - ◇ Also known as **ON-OFF Fading Channel**.
- Can be modeled by an Erdős-Rényi (ER) graph $\mathbb{G}(n;p)$.

The overall system model is the *intersection* $\mathbb{K}(n;\theta) \cap \mathbb{G}(n;p)$.

Connectivity of EG scheme under on-off channel

Model: $\mathbb{K} \cap \mathbb{G}(n;\theta,p) \Rightarrow \mathbb{P}\left[i \sim j\right] = p(1-q(\theta))$.

Theorem 2 (Yağan, IT 2012) *Consider scalings* $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ *and* $p : \mathbb{N}_0 \to (0,1)$ *such that*

$$p_n(1-q(\theta_n)) \sim c\frac{\log n}{n}, \quad n = 1,2,\ldots$$

for some $c > 0$*. If* $\lim_{n\to\infty} p_n \log n = p^\star$ *exists and there exists some* $\sigma > 0$ *such that*

$$\sigma n \leq P_n$$

then we have

$$\lim_{n\to\infty} \mathbb{P}\left[\; \mathbb{K} \cap \mathbb{G}(n;\theta_n,p_n) \textit{ is connected } \right] = \begin{cases} 0 & \textit{if} \quad c < 1 \\ 1 & \textit{if} \quad c > 1. \end{cases}$$

Connectivity under the disk model?

- Nodes are distributed over a bounded region $\mathcal{D}$ of the plane
- **Disk model:** Nodes i and j located at $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$, respectively, in $\mathcal{D}$ are able to communicate if

$$\|\boldsymbol{x_i} - \boldsymbol{x_j}\| < \rho \tag{1}$$

- ρ : transmission range of sensors
- **Random geometric graph,** $\mathbb{H}(n;\rho)$: Induced under (1) when nodes are independently and uniformly distributed over $\mathcal{D}$
- If $\mathcal{D}$ is **unit torus**, then $\mathbb{P}\left[i \sim j\right] = \pi\rho^2$

The overall system model is the *intersection* $\mathbb{K}(n;\theta) \cap \mathbb{H}(n;\rho)$.

A natural conjecture

Conjecture 1 (Yağan, IT 2012) *Consider scalings* $K, P : \mathbb{N}_0 \to \mathbb{N}_0$ *and* $\rho : \mathbb{N}_0 \to (0, 1/\sqrt{\pi})$ *such that*

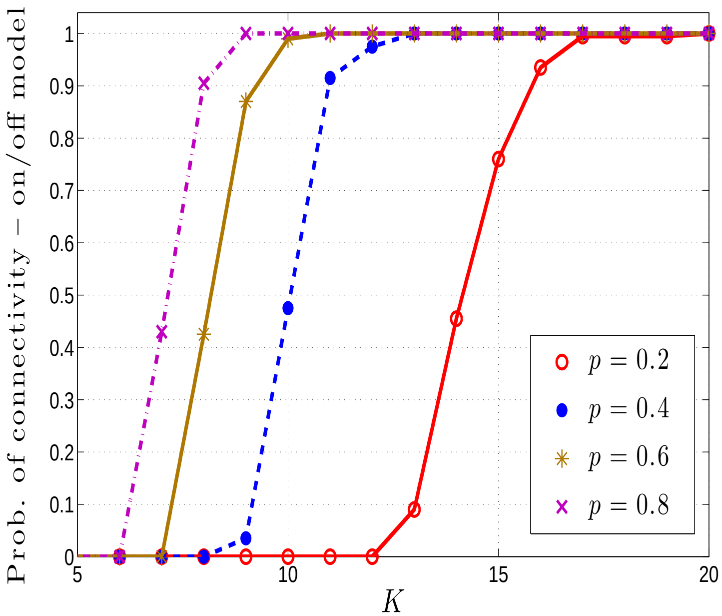
$$\pi \rho_n^2 \cdot (1 - q(\theta_n)) \sim c \frac{\log n}{n}, \quad n = 1, 2, \ldots \tag{2}$$

for some $c > 0$. *Then we have*

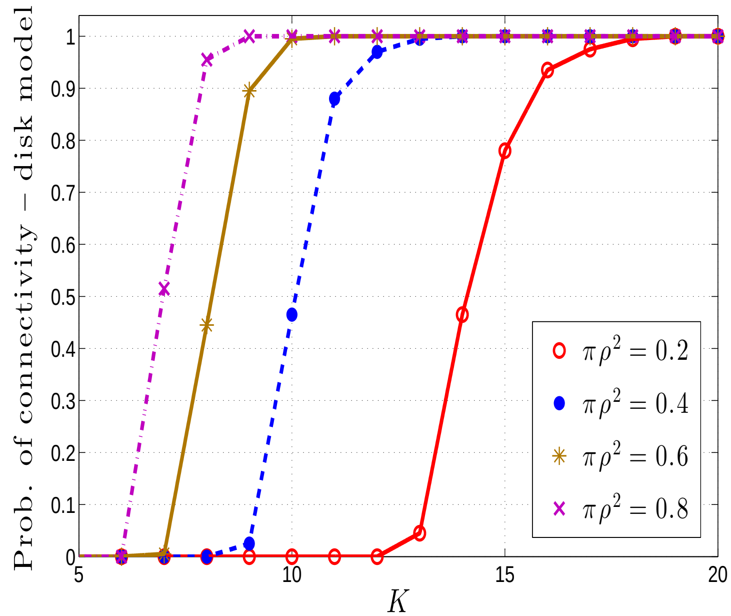
$$\lim_{n \to \infty} \mathbb{P}\left[\ \mathbb{K} \cap \mathbb{H}(n; \theta_n, \rho_n) \text{ is connected } \right] = \begin{cases} 0 & \text{if} \quad c < 1 \\ 1 & \text{if} \quad c > 1. \end{cases}$$

Conjecture: Theorem 2 holds when **On-Off** communication model is replaced by the **Disk** model.

Supporting evidence



(a)



(b)

a) Probability of connectivity under the **on-off** channel model.

b) Probability of connectivity under the **disk model** with $\pi\rho^2 = p$.

Figures are almost indistinguishable!

Partial results on the conjecture

- **Yi et al.** established the zero-law; i..e, under (2), they showed

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{K}\cap\mathbb{H}(n;\theta_n,\rho_n) \text{ is connected }\right] = 0 \qquad \text{if } c<1$$

- **Di Pietro et al.** showed that

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{K}\cap\mathbb{H}(n;\theta_n,\rho_n) \text{ is connected }\right] = 1 \qquad \text{if } c>20\pi$$

- **Krzywdzinski and Rybarczyk** showed that

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{K}\cap\mathbb{H}(n;\theta_n,\rho_n) \text{ is connected }\right] = 1 \qquad \text{if } c>8$$

No result exists for the connectivity when $1 < c \leq 8 \Rightarrow$
An important **gap** given the **trade-offs** vs. security and memory

A related conjecture by Gupta & Kumar

Model: Random geometric graph with randomly deleted links.
$\mathbb{G} \cap \mathbb{H}(n; p, \rho) \Rightarrow \mathbb{P}[i \sim j] = p(\pi\rho^2)$.

Conjecture 2 (Gupta & Kumar, 1998) *Consider scalings $p : \mathbb{N}_0 \to (0,1)$ and $\rho : \mathbb{N}_0 \to (0, 1/\sqrt{\pi})$ such that*

$$p_n \cdot \pi\rho_n^2 \sim c\frac{\log n}{n}, \quad n = 1, 2, \ldots$$

for some $c > 0$. Then, we have

$$\lim_{n\to\infty} \mathbb{P}\left[\ \mathbb{G} \cap \mathbb{H}(n; p_n, \rho_n) \textit{ is connected } \right] = \begin{cases} 0 & \textit{if} \quad c < 1 \\ 1 & \textit{if} \quad c > 1. \end{cases}$$

Not resolved! Open since 15 years!

A summary of connectivity results for intersection of random graphs

- **ER graph** $\bigcap$ **Random Geometric Graph**
  - Conjecture by Gupta & Kumar, 1998: **Open** for $1 < c \leq 8$
- **EG graph** $\bigcap$ **Random Geometric Graph**
  - Conjecture by Yağan, 2012: **Open** for $1 < c \leq 8$
- **EG Graph** $\bigcap$ **ER Graph**
  - A sharp zero-one law is **available**, Yağan, **IT** 2012.
- **Random *K*-out Graph** $\bigcap$ **ER Graph** (Not covered today)
  - A sharp zero-one law is **available**, Yağan & Makowski, **IT**.

Theorem 2 is the first* *complete* zero-one law established for the connectivity of the *intersection* of random graphs.

Thanks!

Visit `www.andrew.cmu.edu/~oyagan` **for references..**