Dealing with Random and Selective Attacks in Wireless Sensor Systems

JINFENG NI, LI ZHOU, and CHINYA V. RAVISHANKAR
University of California, Riverside

We present a framework for analyzing the effects of random and selective compromises (using order statistics) in sensor networks. We discuss the problem of ensuring data integrity at the source and during transit in sensor networks, and present an analysis of the reliability of reports from mobile collectors. No analysis has appeared in the literature of source integrity for mobile nodes, or of selective attacks in sensor networks. We address transit data integrity by presenting mGKE, a key establishment scheme for general group-based sensor deployments, and present a detailed analytical and experimental comparison of mGKE with current schemes. mGKE outperforms current methods in terms of resilience, connectivity, and memory and communication overhead.

Categories and Subject Descriptors: C.2.0 [**Computer Communication Networks**]: General—*Security and protection*; C.2.1 [**Computer Communication Networks**]: Network Architecture and Design—*Wireless Communications*

General Terms: Security, Design, Performance, Reliability

Additional Key Words and Phrases: Attack models, fault tolerance, mobile sensors

ACM Reference Format:
Ni, J., Zhou, L., and Ravishankar, C. V. 2010. Dealing with random and selective attacks in wireless sensor systems. ACM Trans. Sensor Netw. 6, 2, Article 15 (February 2010), 40 pages.
DOI = 10.1145/1689239.1689245 http://doi.acm.org/10.1145/1689239.1689245

1. INTRODUCTION

Compromises in sensor networks are a serious problem, but no general framework exists for modeling compromises. This article presents both a key-distribution scheme for securing sensor deployments, as well an analysis framework for modeling sensor compromises.

This work was supported in part by contract number N00014-07-C-0311 with the Office of Naval Research and by a grant from Tata Consultancy Services, Inc.
This article is based on the paper "Supporting Secure Communication and Data Collection in Sensor Mobile Networks," which was presented at IEEE INFOCOM. © IEEE 2006.
Author's address: C. V. Ravishankar; email: ravi@cs.ucr.edu.

DOI 10.1145/1689239.1689245 http://doi.acm.org/10.1145/1689239.1689245

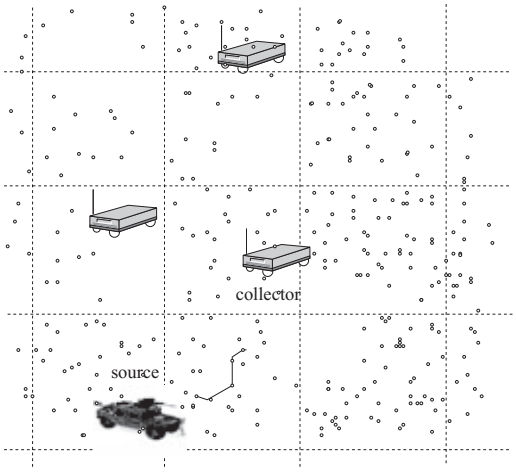


Fig. 1. Mobile Robomotes.

Sensors may be static or mobile, and are often deployed in groups to improve reliability [Liu and Ning 2003b; Du et al. 2004; Huang et al. 2004]. Within each group, data reports can be mutually validated by sensors and otherwise aggregated, and sent to base stations for further analysis. When data is collected by mobile collectors [Ye et al. 2002; Zhang et al. 2003; Kansal et al. 2004; Tirta et al. 2002; Zhang et al. 2005], validation occurs at the base station, using reports from several mobile sensors.

One must ensure the integrity of data reports as well as that of data in transit [Ye et al. 2004; Zhu et al. 2004; Zhou and Ravishankar 2005; Yang et al. 2005]. Cryptographic keys are essential to guarantee data integrity for each hop. One must secure messages between static sensors, between static sensors and mobile collectors, and between mobile collectors. Sensor nodes are resource-limited, so public key cryptosystems can prove expensive. Preloading shared keys is difficult, since a sensor's neighbors may be unknown in advance, given mobility, and in ad-hoc, on-demand deployments.

We distinguish *source data integrity*, which ensures that source reports are trustworthy, from *transit integrity*, which ensures that data remains trustworthy in transit. Source integrity typically requires aggregation of reports arriving from several sources, so transit integrity is a precondition. Source integrity is well studied for static sensors [Ye et al. 2004; Zhou and Ravishankar 2005], but not for mobile sensors. Transit integrity in wireless environments requires cryptographic means.

Figure 1 shows a battlefield where static sensors collaborate to collect data, to be sent to the base station via mobile Robomotes [Sibley et al. 2002]. Soldiers may also carry backpacks with mobile collectors having more memory, computing, battery power, and transmission range than static sensors. In an ocean or river water monitoring scenario, static sensors may be anchored, but data collectors may float, and move with the water.

1.1 Our Contributions

A major contribution of our work is a novel framework, based on order statistics, for analyzing the effects of selective attacks on sensor networks. Selective attack is far more serious than random attack [Huang et al. 2004], but no analysis of selective attack has appeared in the literature, since it poses major technical challenges. We apply our framework to analyze the resilience of PIKE [Chan and Perrig 2005] and mGKE (see Section 5).

We also describe mGKE, a *G*roup-based *K*ey *E*stablishment scheme, an extension of the GKE scheme [Zhou et al. 2005a, 2005b] to ensure transit integrity in the presence of mobility. Previous random key predistribution schemes [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2004; Du et al. 2003; Liu and Ning 2003a, 2003b] require sensors to be densely and uniformly distributed, but mGKE can establish unique pairwise keys in connected networks regardless of sensor density or distribution. Communication in sensor networks is mostly between neighboring nodes. mGKE establishes pairwise keys between neighbors using only local communication, unlike PIKE [Chan and Perrig 2005], which require network-wide communication. mGKE security also degrades gracefully with the number of compromised sensors, improving resilience against node compromise.

Related work appears in Section 2, and our assumptions and metrics in Section 3. We analyze mobile collectors compromises in Section 4. We present the mGKE scheme in Section 5, analyze its security in Section 6, and evaluate its performance in Section 8. Section 9 concludes the article.

2. RELATED WORK

Mobile collectors are typically also privileged, so their compromise has nonlocal consequences [Zhang et al. 2005]. Schemes are proposed in Zhang et al. [2005] to limit mobile collector privileges using privilege-dependent pairwise keys between mobile collectors and sensors. Successful key establishment serves as proof of privilege. However, these schemes may break down with as few as 200 node compromises, due to sensor memory limitations. This low number precludes their use in large sensor networks, which may consist of thousands of nodes.

It has been argued that mobility facilitates security and authentication in MANETs [Kong et al. 2001; Balfanz et al. 2002; Capkun et al. 2003]. Kong et al. [2001] and Capkun et al. [2003] use public-key cryptosystems, which are expensive for resource-constrained sensor networks, the domain we address. Secure associations between nodes and certificate service are the chief concerns in [Kong et al. 2001; Balfanz et al. 2002; Capkun et al. 2003]. We deal with the impact of compromised mobile nodes on data integrity.

2.1 Key Predistribution

Random key predistribution (RKP) schemes [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2004] preload each sensor with a random subset of keys from a global key pool. Now, any pair of sensors will share a key with a certain probability. Two sensors can choose any element in the intersection of their

subsets as their pairwise key. If these subsets are disjoint, they may establish a *path key* using intermediary sensors. These schemes are based on results from random graph theory [Erdös and Renyi 1959], which guarantee that a random graph is connected with high probability if the number of edges in it exceeds a threshold. To improve resilience to node capture, Du et al. [2003] and Liu and Ning [2003a, 2003b] proposed *structured random key predistribution* (SRKP) schemes, which have a nice property: When fewer than a threshold number of sensors are compromised, keys between uncompromised sensors are affected with negligible probability.

RKP and SRKP suffer from two major problems. First, they require deployment densities high enough to ensure connectivity, seriously hindering their use in sparse sensor deployments, as when sensors fail over time, or when the deployments are themselves sparse. Second, their approach to key (or key space) sharing degrades resilience to node capture. Compromising a sensor also compromises all keys (or key spaces) in it, weakening the security of all other sensors using keys from this set (or space).

PIKE [Chan and Perrig 2005] arranges sensors in a logical grid; sensor $s_{i,j}$ is at grid node (i, j). Sensors on the same row or column share preloaded pairwise keys. If $s_{i,j}, s_{k,l}$ share no preloaded pairwise key, they can establish a path key via either $s_{i,l}$ or $s_{k,j}$. PIKE is more secure than SRKP since pairwise keys are unique. However (Section 8.2.1), over 98% of neighboring sensor pairs in PIKE do not share preloaded keys, and must establish path keys, which requires $O(\sqrt{n})$ communication overhead [Chan and Perrig 2005]. If the bivariate polynomial's degree is set equal to the group size γ, the grid-based predistribution model of [Liu and Ning 2003a] is γ-collusion-resistant, achieving the same security level as PIKE. However, PIKE can halve its memory requirement using cryptographic hash functions. This does not seem possible for [Liu and Ning 2003a].

While most work considers flat network architecture, [Chen and Drissi 2005; Qingguang et al. 2006] address the problem of key management in hierarchical sensor networks.

2.2 Group-Based Key Predistribution

Several group-based key predistribution schemes [Du et al. 2004; Liu and Ning 2003a; Huang et al. 2004] assume groups are assigned to subregions statically, so that group adjacency are known before sensor deployment. In practice, however, it could prove difficult, or even impossible to deploy groups to predetermined subregions.

More flexible group-based deployment models appeared in Zhou et al. [2005b] and in Liu et al. [2005], which allow a group to be deployed *any* subregion, simplifying sensor deployment. [Liu et al. 2005] uses the proposed framework in combination with schemes such as random key predistribution (RKP) [Eschenaer and Gligor 2002], the random subset and the polynomial-based scheme [Liu and Ning 2003a], improving security significantly.

The basic ideas underlying our current work first appeared in Zhou et al. [2005b] and [Zhou et al. 2006], where we proposed mGKE, a pairwise key

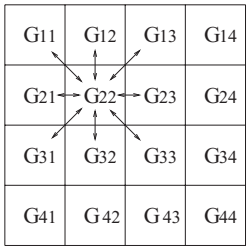


Fig. 2. Groups.

predistribution framework for general group-based sensor deployment. In mGKE, sensors within a group share unique pairwise preloaded keys, and t pairs of agents across each pair of groups are preloaded with unique pairwise keys, to facilitate intergroup key establishment.

This idea was subsequently extended in Liu et al. [2008] to a hash key-based scheme in which sensors within a group are preloaded with unique pairwise keys. Given groups of size γ, each pair of groups contain γ pairs of "bridges" preloaded with unique pairwise keys for inter-group key establishment. Clearly, this hash key-based scheme is a special case of the scheme in Zhou et al. [2005b], when the number t of agents equals γ.

Among the contributions of our paper are an in-depth analysis of the ideas of [Zhou et al. 2005b], and a novel framework to analyze the security impact of selective attacks, which are much more serious than random attacks. Earlier work, such as Zhou et al. [2005b], Liu et al. [2005, 2008], has only analyzed security for random attacks.

3. NETWORK ASSUMPTIONS AND THREAT MODEL

We assume that attackers may listen to, intercept, record or manipulate all traffic in the network. Compromising a node compromises all keys stored at the node. Attackers may inject false reports using compromised nodes, as well as compromise communications between sensor pairs by compromising the keys shared pairwise between them. All path keys are established at deployment time. This does not weaken our model, since attackers may record key establishment messages, and recover path keys when intermediaries are compromised.

As in [Liu and Ning 2003b; Du et al. 2004; Huang et al. 2004], we deploy sensors in group-based fashion. Current group-based schemes [Du et al. 2004; Huang et al. 2004] assume that group adjacency are known before sensor deployment, and tie sensor groups to subregions statically. In Figure 2, the sensor group that is tied to cell (i, j) is labeled G_{ij}, since it may be deployed only into cell (i, j). In contrast (see Figure 3), mGKE allows sensor group $G_{i,j}$ to be deployed into any cell (k, l) in the grid, making our sensor deployment more flexible.

We assume that sensors have resource limitations typical of current sensors, such as MICA2 motes [Crossbow]. We present separate schemes for the case when mobile collectors are resource-rich devices, and when they are resource-limited as in [Sibley et al. 2002; Bergbreiter and Pister 2003]. Our analysis for

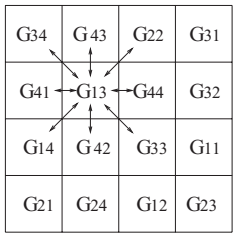


Fig. 3. mGKE.

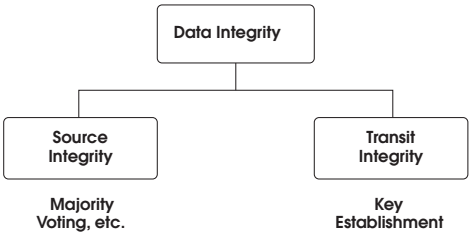


Fig. 4. Data integrity.

mobile collectors assumes the Random Waypoint model [Bettstetter et al. 2004], which is commonly used for wireless mobile networks, but our analysis may be extended to other mobility models. We show in Section 4 that this model can improve data consistency if compromised mobile collectors mount false report attacks [Ye et al. 2004].

3.1 Data Integrity Challenges

Data integrity may be compromised at two levels. First, the source generating the data may have been compromised, raising the issue of *source integrity*. Second, the data may be compromised by adversaries en route, raising the issue of *transit integrity*. Source integrity is addressed by ensuring that a report from a single sensor is never trusted, and that no report is trusted unless it is validated using reports from multiple sensors.

We will not be concerned with the specifics of how source reports are validated; we are interested mainly in ensuring that the base station receives reports from enough sensors to ensure validation. We ensure *transit integrity* by establishing secure pairwise keys between communicating nodes, so that reports can be signed or encrypted to prevent tampering. Key establishment is a major focus of our work.

3.2 Selective v/s Random Attack Model

Previous schemes [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2003; Liu and Ning 2003a, 2003b; Du et al. 2004; Chan and Perrig 2005] have analyzed the *random* attack model, in which attackers compromise sensors at random. However, this model is simplistic; clever attackers can adopt a *selective* attack model [Du et al. 2004; Huang et al. 2004], choosing targets deliberately

to maximize the benefits of attack. An adversary targeting a certain region will target sensors in that region. Similarly, an adversary may target sensors that hold the largest numbers of uncompromised keys, to maximize the number of key compromises at the next attack step.

As [Huang et al. 2004] show for RKP and SRKP, selective attacks are deadlier than random attacks. To compromise 50% of the communication links among uncompromised sensors in a 10,000-node network under RKP, one must compromise 230 sensors under random attack but only 160 under selective attack. Under SKRP, the attacker must compromise 200 sensors under random attack, but only 125 sensors under selective attack.

3.3 Security Metrics

We characterize the security of schemes by their *resilience* and *connectivity*.

3.3.1 *Resilience.* Let $\mathcal{U}$ be the set of uncompromised sensors. Let $L(\mathcal{U})$ be the set of links between them, and $\widehat{L}(\mathcal{U}) \subseteq L(\mathcal{U})$ be the subset of compromised links. Resilience is the ratio $|\widehat{L}(\mathcal{U})|/|L(\mathcal{U})|$, and measures how compromises affect other parts of the network.

This definition of resilience is similar to those used in previous random key predistribution schemes [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2003; Liu and Ning 2003a; Du et al. 2004; Huang et al. 2004]. However, in our definition, a link is secured either by a preloaded pairwise key or by a path key. In contrast, the previous schemes consider only the links secured by preloaded keys [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2004] or keys derived from preloaded key space [Du et al. 2003; Liu and Ning 2003a; Huang et al. 2004]. As Du et al. [2003] point out, a path key is compromised if an attacker deciphers messages during key establishment or compromises any intermediary. One must hence consider the security of path keys to properly evaluate the effects of sensor compromise.

3.3.2 *Key Connectivity.* Key connectivity is the probability that a sensor network is securely connected. In Section 6.3, we show that mGKE can allow a sensor network to be securely connected with 100% probability, as long as the network is physically connected.

3.4 Resource Metrics

We measure resource requirements in terms of communication and memory overhead.

3.4.1 *Communication Overhead.* We measure communication overhead as the average number of hops that messages must travel to establish a key. Since security degrades gracefully with the number of compromises only in mGKE and PIKE (see Figure 10), we compared these methods in terms of communication overhead. For both methods, we measure only the overhead for sending the encrypted path key messages, and ignore routing overhead. As indicated in Chan and Perrig [2005], routing overhead is highly dependent on the underlying routing protocol, which is beyond the scope of our paper. Also, mGKE

Table I. Experiment and Analysis Setup

Parameter	Setting (Default)	Parameter	Setting (Default)
# static sensors (n_s)	10000–50000 (10000)	# groups (g)	100–500 (100)
# mobile collectors (n_m)	100 (100)	# keys/sensor for other groups μ	30 (30)
Deployment density (δ)	20–100 (50)	# agents between two groups (t)	30 (30)
Communication range (r)	40 m (40 m)	# compromised sensors	0–2000 (0–2000)
Group size (γ)	100 (100)	# groups MS is associated with (g')	30 (30)
Time interval T	100s–300s (200s)	Pause time at waypoint (ω)	5s (5s)
Mobile collector speed (v)	10m/s–30m/s (20m/s)	Monitored area (A)	$\frac{n_s \pi r^2}{\delta}$ ($10^3 \times 10^3$m)
Number of trials	100 (100)	Fraction of compromised (ρ)	0.2 (0.2)

introduces lower routing overhead than PIKE for any given routing protocol (see Section 5.4). Therefore, neglecting the communication overhead of routing for both mGKE and PIKE does not favor mGKE in any way over PIKE. Rather, it helps us focus on the efficiency of the two key establishment techniques.

3.4.2 *Memory Overhead.* As is standard practice, we quantify memory overhead by the number of keys preloaded into each sensor. We do not count temporary storage used in establishing new keys, or the memory to store them.

3.5 Experimental Validation of Analysis

Although our primary focus was on analytical methods, we verified our analysis through experiments. We used a simulator implemented in C++ under Linux and the LEDA library[1] to simulate the network topology and to compute shortest paths between nodes. Our data analysis was performed using the GNU Scientific Library (GSL).[2] Each set of simulations consisted of 100 trials, and averages computed over these 100 trials. Table I summarizes the setup for our simulation and analysis.

4. AN ANALYSIS OF SOURCE INTEGRITY FOR MOBILE COLLECTORS

Let n_m mobile collectors move about in a region R of area A_R. Under the Random Waypoint mobility model, waypoints are distributed uniformly in the region. At step i, a mobile node moves at constant velocity v from its current waypoint P_{i-1} to a new random waypoint P_i, where it pauses for a constant time w to communicate with neighboring sensors. It does not communicate with sensors while in transit.

A collector sends to the base station the data it collects from sensors in each visited subregion. Consider a subregion r of area A_r. During an interval T of

[1]LEDA Library. http://www.mpi-inf.mpg.de/LEDA/.
[2]GNU Scientific Library. http://www.gnu.org/software/gsl/.

interest, several mobile collectors may visit subregion r, so the base station has several reports of data for r. The base station uses majority voting when reports are in conflict. Given κ compromised mobile collectors, an attacker can fabricate at most κ false reports for subregion r.

Definition 1 (*Resilience*). Let $Y_r(T)$ uncompromised mobile collectors visit subregion r in the interval T. The system is κ-resilient for subregion r if $Y_r(T) > \kappa$.

Intuitively, since any mobile collector visits any subregion at each step with the same probability, the expected number of uncompromised mobile collectors visiting a subregion during interval T increases with number of steps taken, which increases with T. We will analyze the probability of κ-resilience as a function of n_m and κ. To the best of our knowledge, this is the first work to present such an analysis for any mobility model.

4.1 Data Consistency under Random Waypoint Mobility Model

Let $\mathbf{C}(\boldsymbol{\kappa})$ be the event that κ of the n_m mobile collectors have been compromised. Let $Y_r(T)$ be the number of uncompromised mobile collectors visiting subregion r during the time interval T. Let $\mathbf{K}(\mathbf{r})$ be the event that the system is resilient for region r. Now,

$$\Pr[\mathbf{K}(\mathbf{r})|\ \mathbf{C}(\boldsymbol{\kappa})] = \Pr[Y_r(T) > \kappa].$$

Let the j^{th} collector $M^{\langle j\rangle}$ take $\tau^{\langle j\rangle}$ steps in the interval T. Let $l_i^{\langle j\rangle}$ be the length of its i^{th} step, which takes time $t_i^{\langle j\rangle} = (l_i^{\langle j\rangle}/v + w)$. The distance between two random points in a unit square has the density [Ghosh 1951]

$$f(l) = \begin{cases} 2l(l^2 - 4l + \pi) & \text{for } 0 \le l \le 1 \\ 2l(4\sqrt{l^2-1} - (l^2 + 2 - \pi) - 4\tan^{-1}(\sqrt{l^2-1})) & \text{for } 1 < l \le \sqrt{2}, \end{cases} \quad (1)$$

from which we can calculate the expectation $E[l_i^{\langle j\rangle}] = s(\frac{2+\sqrt{2}+\sinh^{-1}1}{15})$. Hence,

$$E\left[t_i^{\langle j\rangle}\right] = \frac{s}{v}\left(\frac{2+\sqrt{2}+\sinh^{-1}1}{15}\right) + w. \quad (2)$$

The number of steps $\tau^{\langle j\rangle}$ that mobile collector $M^{\langle j\rangle}$ takes in time T is

$$\tau^{\langle j\rangle} = \sup\{k : S_k \le T\},\ \text{where } S_k = \sum_{i=1}^{k} t_i^{\langle j\rangle}. \quad (3)$$

Let u be a random variable such that $T = S_k + u$, so that $E[S_k] = T - E[u]$. Since $E[S_k] = E[k]E[t_i^{\langle j\rangle}]$, $E[k] = \frac{T-E[u]}{E[t_i^{\langle j\rangle}]}$. Clearly, $0 < E[u] < E[t_i^{\langle j\rangle}]$, so that $\frac{T}{E[t_i^{\langle j\rangle}]} - 1 < E[k] < \frac{T}{E[t_i^{\langle j\rangle}]}$, or using $E[k] = E[\tau^{\langle j\rangle}]$,

$$\frac{T}{E\left[t_i^{\langle j\rangle}\right]} - 1 < E\left[\tau^{\langle j\rangle}\right] < \frac{T}{E\left[t_i^{\langle j\rangle}\right]}, \quad (4)$$

where $E[t_i^{\langle j\rangle}]$ is as in Equation (2). Let $\mathbf{V^{\langle j\rangle}}$ be the event that $M^{\langle j\rangle}$ visits subregion r at least once, and let $\boldsymbol{\tau}^{\langle \mathbf{j}\rangle}$ be the event that $M^{\langle j\rangle}$ takes $\tau^{\langle j\rangle}$ steps in

interval T. Since regions R and r have areas A_R and A_r, the probability of $M^{\langle j \rangle}$ visiting r at any given step is $p = A_r/A_R$. Now, $\Pr[\mathbf{V}^{\langle \mathbf{j} \rangle} | \boldsymbol{\tau}^{\langle \mathbf{j} \rangle}] = 1 - q^{\tau^{\langle j \rangle}}$, where $q = 1 - p$. Since $l_i^{\langle j \rangle}$ is bounded by the length $s\sqrt{2}$ of the square region's diagonal, $\tau^{\langle j \rangle} \geq \tau_{\min} = T/(\sqrt{2}s/v + w)$. Using a step size of zero gives $\tau^{\langle j \rangle} \leq \tau_{\max} = T/w$. Hence,

$$\Pr[\mathbf{V}^{\langle \mathbf{j} \rangle}] = \sum_{\tau^{\langle j \rangle} = \tau_{\min}}^{\tau_{\max}} \Pr[\mathbf{V}^{\langle \mathbf{j} \rangle} | \boldsymbol{\tau}^{\langle \mathbf{j} \rangle}] \Pr[\boldsymbol{\tau}^{\langle \mathbf{j} \rangle}] = \sum_{\tau^{\langle j \rangle} = \tau_{\min}}^{\tau_{\max}} (1 - q^{\tau^{\langle j \rangle}}) \Pr[\boldsymbol{\tau}^{\langle \mathbf{j} \rangle}] = 1 - E[q^{\tau^{\langle j \rangle}}]. \tag{5}$$

We now proceed to bound $E[q^{\tau^{\langle j \rangle}}]$. From the Edmundson-Madansky [Madansky 1959] inequality, we know that for any convex function $f : \mathbb{R} \to \mathbb{R}$ and a random variable ξ with support in $[a, b]$ and mean value μ,

$$E[f(\xi)] \leq \frac{bf(a) - af(b)}{b - a} + \mu \frac{f(b) - f(a)}{b - a} \tag{6}$$

We apply Inequality (6), with $a = \tau_{\min}$, $b = \tau_{\max}$, $f(x) = q^x$, $\xi = \tau^{\langle j \rangle}$, and $\mu = E[\tau^{\langle j \rangle}]$, to get

$$\begin{aligned} E[q^{\tau^{\langle j \rangle}}] &\leq \frac{\tau_{\max} q^{\tau_{\min}} - \tau_{\min} q^{\tau_{\max}}}{\tau_{\max} - \tau_{\min}} + E[\tau^{\langle j \rangle}] \frac{q^{\tau_{\max}} - q^{\tau_{\min}}}{\tau_{\max} - \tau_{\min}} \\ &= \frac{w}{T} \left(1 + \frac{vw}{s\sqrt{2}}\right) \left((\tau_{\max} q^{\tau_{\min}} - \tau_{\min} q^{\tau_{\max}}) + E[\tau^{\langle j \rangle}](q^{\tau_{\max}} - q^{\tau_{\min}})\right) \end{aligned} \tag{7}$$

We can preserve the inequality in Equation (7) by taking a conservative approach, and using the upper bound for $E[\tau^{\langle j \rangle}]$ from Equation (4). Now, Equations (5) and (7) yield

$$\begin{aligned} \Pr[\mathbf{V}^{\langle \mathbf{j} \rangle}] &= 1 - E[q^{\tau^{\langle j \rangle}}] \\ &\geq 1 - \frac{w}{T} \left(1 + \frac{vw}{s\sqrt{2}}\right) \left((\tau_{\max} q^{\tau_{\min}} - \tau_{\min} q^{\tau_{\max}}) + \frac{T(q^{\tau_{\max}} - q^{\tau_{\min}})}{E[t_i^{\langle j \rangle}]}\right) \end{aligned} \tag{8}$$

Using equality in Equation (8) gives us a conservative estimate of the probability that mobile collector $M^{\langle j \rangle}$ visits subregion r. Now, there are n_m collectors in all, of which $n_m - \kappa$ are uncompromised. Each mobile collector will visit subregion r at least once with probability $\beta = \Pr[\mathbf{V}^{\langle \mathbf{j} \rangle}]$, so that $Y_r(T)$ is Binomially distributed with success probability β. That is,

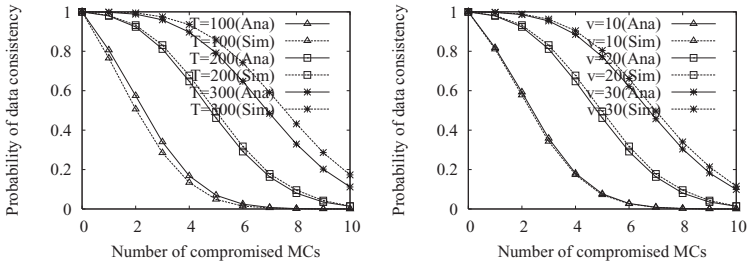
$$\Pr[Y_r(T) = y] = \binom{n_m - \kappa}{y} \beta^y (1 - \beta)^{n_m - \kappa - y}.$$

Therefore, we have

$$\Pr[Y_r(T) > \kappa] = 1 - \sum_{y=0}^{\kappa} \Pr[Y_r(T) = y] = 1 - \sum_{y=0}^{\kappa} \binom{n_m - \kappa}{y} \beta^y (1 - \beta)^{n_m - \kappa - y}$$

4.2 Validation of Analysis by Experiments

Consider a $1,000m \times 1,000m$ region with 100 equal-sized subregions, with 100 collectors moving at speeds of $v = 10m/s$, $20m/s$ or $30m/s$, pausing at



(a) Consistency vs. T (v=20). (b) Consistency vs. speed (T=200).

Fig. 5. Security with compromised mobile collectors (MCs).

Table II. Our Notation

Notation	Description	Notation	Description
s_i	the i-th static sensor	γ	the group size
m_j	the j-th mobile collector	g	the number of groups
G_u	the u-th static sensor group	δ	avg. # sensors in any sensor's range
n_s	the number of static sensors	μ	# preloaded keys per sensor for other groups
n_m	the number of mobile collectors	t	the number of agents for other groups

waypoints for $w = 5s$. Let base stations collect data every $T = 100s$, $200s$ and $300s$. Figure 5(a) compares our analytical and experimental results for the probability of κ-resilience for any subregion for various T, when $v = 20m/s$. Figure 5(b) compares our analytical and experimental results for κ-resilience for various speeds v when $T = 200$. Our analytical results closely match those of our experiments, confirming the accuracy of our analysis.

5. TRANSIT INTEGRITY: THE MGKE SCHEME

We present mGKE, a key-establishment scheme for group-based sensor deployments [Liu and Ning 2003b; Du et al. 2004; Huang et al. 2004], in which a sensor belongs to a single group, but each group may be deployed to any subregion in the sensor field, as in Figure 3. Each node is preloaded with a set of keys, each shared pairwise with one other node. We use the notation in Table II. We refer to the pairwise key between a pair of sensors as an S-S key, and the pairwise key between a mobile node and a sensor as an M-S key.

Let there be n_s sensors and n_m mobile collectors. We will denote the ith static sensor by s_i and the jth mobile collector by m_j. We arrange the static sensors into g groups G_i, $1 \le i \le g$, each of which has $\gamma = n_s/g$ sensors. Group G_u will comprise sensors s_i such that $(u-1)\gamma < i \le u\gamma$. Let $\langle G_u, s_i \rangle$ denote sensor s_i from group G_u. We will replace $\langle G_u, s_i \rangle$ by s_i, when no confusion can arise.

5.1 Outline of mGKE

Two nodes are *associated* if they share a *preloaded* pairwise key. Sensors in the same group are preconfigured to be associated with each other. Each sensor is also associated with sensors in other groups, in a pattern designed to ensure several sensor associations across each pair of groups. Any sensor s_i can now estab-

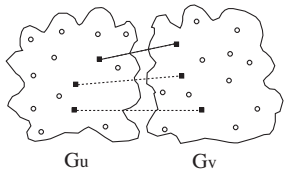


Fig. 6. 3-association.

lish a unique pairwise key with any other s_j. If s_i and s_j are from the same group, they start off associated. If they are from different groups, there will exist *multiple* associations between their groups, so they can establish a pairwise key using any pair of these associated sensors as intermediaries. Unlike PIKE [Chan and Perrig 2005], this process involves only localized communication.

We present two approaches for key establishment between a mobile collector and a static sensor. The first is usable when the mobile collector has $O(n_s)$ memory, but the second method is more general. In our second approach, the base station preloads each mobile collector m_i with keys ensuring several associations with each of a set of selected groups. m_i can now establish a unique pairwise key with any static sensor $\langle G_u, s_j \rangle$ using its associations in G_u (or in any nearby group, since all groups are associated). A mobile collector pair can use this method to establish a path key.

5.2 Key Establishment between Communicating Sensor Pairs

5.2.1 *S-S Key Predistribution.* We preload each pair of sensors from the same group with a unique pairwise key. Memory requirements are modest if the group size is chosen appropriately; if the group size is 100 as in [Du et al. 2004; Huang et al. 2004], each sensor must store 99 keys. For 64-bits keys, each sensor requires 792 bytes. This is doable for a Mica2 Mote sensor that has 4KB SRAM [Crossbow]. This memory requirement can be halved, as in Chan and Perrig [2005], so that 396 bytes suffice to assign unique keys to all sensor pairs in the same group. We now consider key establishment across groups.

Definition 2. $\langle G_u, s_i \rangle$ is an agent for G_v in G_u, if $\langle G_u, s_i \rangle$ is associated with some $\langle G_v, s_j \rangle$ in G_v. G_u and G_v are *t-associated* if they have t agents for each other.

Though group adjacencies are unknown prior to deployment, we require each group to be associated with every other group making it easy to establish key across groups. Sensors from groups G_u and G_v can establish path keys using any pair of associated agents as intermediaries. If there are g groups, and each sensor has enough memory to hold μ inter-group pairwise keys, each group can have up to $t = \lceil \frac{\mu\gamma}{g-1} \rceil$ agents in each of the other groups. Algorithm 1 shows how to define group associations. We use functions $\mathcal{F}_i$ $(1 \leq i \leq t)$ which uniformly map group pairs from $[1, g] \times [1, g]$ to $[1, n_s]$. $\mathcal{F}_i(G_u, G_v)$ selects the ith agent for G_v in G_u, as follows

$$\mathcal{F}_i(G_u, G_v) = \big(t(v-1) + i\big) \pmod{\gamma} + (u-1)\gamma.$$

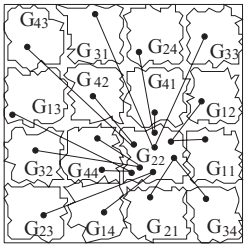


Fig. 7. Keys ($t = 1$).

G_u comprises sensors s_i with $(u-1)\gamma < i \leq u\gamma$. Hence $\mathcal{F}_1(G_u, G_v), \ldots, \mathcal{F}_t(G_u, G_v)$ select t sensors, with indices between $\big(t(v-1)+1\big) \mod \gamma + (u-1)\gamma$ and $tv \mod \gamma + (u-1)\gamma$ as agents for G_v.

Algorithm 1. Inter-group S-S key predistribution

```
t = ⌈μγ/(g−1)⌉
for each pair of groups G_u, G_v do
  for i = 1 to t do
    s_x = F_i(G_u, G_v)
    s_y = F_i(G_v, G_u)
    assign a unique pairwise key to s_x and s_y
  end for
end for
```

Figure 7 shows the inter-group S-S key predistribution for sensors in group G_{22}. For simplicity, we only show the scenario when each group pair has one agent pair. Accordingly, each sensor is required to be preloaded with $\mu = 2$ keys shared with sensors in distinct groups. Algorithm 1 has several attractive features. Each sensor works as agent for the same number of groups. This balances loads and creates no high-value targets, since no sensor holds more keys than any other. The existence of multiple agents improves resilience for establishing path keys. Finally, agents can be discovered easily using the functions $\mathcal{F}_1, \ldots, \mathcal{F}_t$, rather than by lookups.

A nice feature of mGKE is that its security properties remain robust even if t is reduced (see Section 6). We can hence tune t to meet a sensor's memory constraints. mGKE will work very well even when sensors have limited memory.

5.2.2 *S-S Key Establishment.* A unique pairwise key is preloaded for every intra-group sensor pair. For a pair of communicating sensors from different groups, we adopt the Highest Random Weight technique [Thaler and Ravishankar 1998] to choose agents for path key generation, using a hash function $\mathcal{H}$ to realize distributed agreement. Sensors $\langle G_u, s_i \rangle$ and $\langle G_v, s_j \rangle$ generate a path key as follows (see Figure 8).

One principal, say $\langle G_u, s_i \rangle$, computes $\mathcal{H}(s_i, s_j, p)$ for $1 \leq p \leq t$, selects the p yielding the highest $\mathcal{H}$ value. It now uses $\mathcal{F}_p$ to pick an associated sensor pair

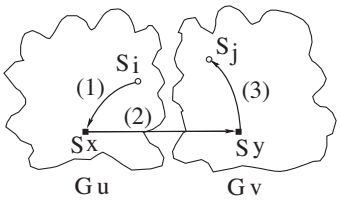


Fig. 8. Intergroup keys.

$\langle G_u, s_x \rangle$ and $\langle G_v, s_y \rangle$. Now, s_i randomly generates a key K_{ij} and sends it to agent s_x, encrypted with the key K_{ix} it shares with s_x.

$$s_i \rightarrow s_x : (K_{ij}, G_v)_{K_{ix}}.$$

s_x decrypts this message, encrypts it with the key K_{xy} shared with s_y, and sends it to s_y.

$$s_x \rightarrow s_y : (K_{ij})_{K_{xy}}.$$

s_y decrypts this packet, re-encrypts it with the key K_{jy} it shares with s_j, and sends it to s_j.

$$s_y \rightarrow s_j : (K_{ij})_{K_{jy}}.$$

s_j first applies $\mathcal{H}$ to select the same associated pair $\langle G_u, s_x \rangle$ and $\langle G_v, s_y \rangle$ that s_i selected, and recovers K_{ij} using K_{jy}, its preloaded association key with s_y.

5.3 Key Establishment between Mobile Collectors and Sensors

We call $\langle G_u, s_i \rangle$ an agent for mobile collector m_i in G_u if m_i and $\langle G_u, s_i \rangle$ are associated.

5.3.1 *Mobile Collectors with $O(n_s)$ Memory.* Static sensors are memory-limited but mobile nodes are not. Each sensor s_j is preloaded with a secret key K_{s_j} shared pairwise with the base station. Sensor s_j communicates securely with mobile collector m_i using key $K_{ij} = \mathcal{R}(K_{s_j}, m_i)$, where $\mathcal{R}$ is a pseudo-random function (PRF) [Goldreich et al. 1986]. Each mobile collector m_i is preloaded with the set of keys $\{K_{ij}\}$ for all sensors s_j. Sensor s_j can compute a unique pairwise key shared with every mobile collector m_i on-demand. However, mobile collectors have enough memory to store the keys they need. While $\mathcal{R}$ may be easy to compute, the overhead can be high if the number of mobile collectors is high.

5.3.2 *Mobile Collectors with Limited Memory.* We create associations between each mobile collector m_i and sensors from some selected g' groups, in a pattern that ensures that m_i is t-associated with each of the g' groups. The g' groups can be selected using g' functions analogous to the $\mathcal{F}_i$ functions defined in Section 5.2.1, to ensure that each group is likely to be chosen by the same number of mobile collectors. This balances loads and reduces high-value targets, since no group holds more keys than any other. Also, agents for mobile collectors can be chosen using functions $\mathcal{F}'_i$ analogous to the functions $\mathcal{F}_i$ in

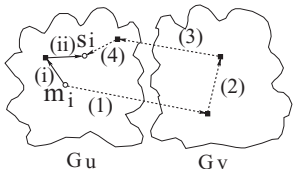


Fig. 9. M-S Keys.

Section 5.2.1, in whose definition we can treat m_i as a group. The function $\mathcal{F}_i'(G_u, m_i)$ is used to select the ith agent for m_i in G_u.

5.3.3 *M-S Key Establishment.* A mobile collector m_i and a sensor $\langle G_u, s_i \rangle$ generate a path key as follows. If m_i has agents in G_u, we use HRW as in Section 5.2.2 to choose an agent for path key generation. Otherwise, m_i finds an agent in an adjacent group (say G_v), and uses that agent and the agent pair between G_u and G_v as intermediaries to establish path keys. To further reduce the communication overheads at sensors, we may allow m_i to move to the agent.

5.4 Features of mGKE

Resilience to Impersonation. Since all messages above are secured with the preloaded pairwise keys shared between sender and receiver, no attacker can impersonate the intermediaries without the preloaded keys.

Failure Resilience. We can guarantee resilience as in [Thaler and Ravishankar 1998]. Each association between groups or mobile sensor has $t = \lceil \frac{\mu\gamma}{g-1} \rceil$ agent pairs. If a pair of intermediaries selected for a path key using $\mathcal{H}$ fails, we simply select the pair corresponding to the index q that yields the second biggest $\mathcal{H}$ value, and use $\mathcal{F}_q$ to determine the new agent pair for path key generation. We can continue until we find an agent pair that is alive.

Routing Protocol. Routing is an issue orthogonal to our work. PIKE uses the geographic routing protocol GPSR [Karp and Kung 2000] with a globally addressable infrastructure GHT [Ratnasamy et al. 2002] to find routes to the intermediate nodes. mGKE can also use GPSR and GHT to find routes from nodes to agents or between agents.

However, the overhead of routing in mGKE is much smaller than that in PIKE. Routing to trusted intermediary nodes in PIKE involves network-wide route discovery, since these intermediaries may not always be in the vicinity. In contrast, the static/mobile node and the agent are either within the same group or within nearby groups in mGKE, so discovering a route to the agent only involves route discovery within the group or nearby groups. Route discovery between agents is also local since they are in adjacent groups. mGKE can accomplish key establishment even without a globally addressable infrastructure.

6. ANALYSIS OF RESILIENCE UNDER RANDOM ATTACK

We compare mGKE with SRKP [Du et al. 2003], the scheme in Du et al. [2004], and PIKE [Chan and Perrig 2005] in terms of resilience against node capture and connectivity.

6.1 S-S Keys Shared between Static Sensors

Let s_i and s_j be *any* two static nodes, $L_{\mathbf{ij}}$ be the communication link between them, and K_{ij} be the key used to secure this link. Let $\mathbf{\Lambda(K_{ij})}$ be the event that K_{ij} is a preloaded key, and let $\mathbf{\Pi(K_{ij})}$ be the event that K_{ij} is a path key. Let $\widetilde{\mathbf{L_{ij}}}$ be the event that link L_{ij} is compromised, and $\widetilde{\kappa}$ be the event that κ static sensors have been compromised. If $\mathbf{U_{ij}}$ is the event that sensors s_i and s_j are both uncompromised, we defined resilience (Section 3.3) as the probability of event L_{ij} given events $\mathbf{U_{ij}}$ and $\widetilde{\kappa}$. This probability is

$$\begin{aligned}\Pr[\widetilde{\mathbf{L_{ij}}}\,|\,(\mathbf{U_{ij}} \wedge \widetilde{\kappa})] = \Pr[\widetilde{\mathbf{L_{ij}}}\,|\,(\mathbf{U_{ij}} \wedge \widetilde{\kappa}) \wedge \mathbf{\Lambda(K_{ij})}] \times \Pr[\mathbf{\Lambda(K_{ij})}|\,\mathbf{U_{ij}}] \\ + \Pr[\widetilde{\mathbf{L_{ij}}}\,|\,(\mathbf{U_{ij}} \wedge \widetilde{\kappa}) \wedge \mathbf{\Pi(K_{ij})}] \times \Pr[\mathbf{\Pi(K_{ij})}|\,\mathbf{U_{ij}}]. \quad (9)\end{aligned}$$

Pairwise keys in [Eschenaer and Gligor 2002; Chan et al. 2003; Du et al. 2003; Liu and Ning 2003a; Du et al. 2004; Huang et al. 2004] are randomly selected from a global pool, so the compromise of one sensor may compromise a number of pairwise keys for other sensors. This is impossible in mGKE, since our preloaded pairwise keys are unique. The only way to compromise a link secured by a preloaded key in mGKE is to compromise one of its endpoints. That is, *mGKE achieves perfect resilience against node capture* for preloaded keys. We hence exclude the event $\mathbf{\Lambda(K_{ij})}$, reducing Equation (9) to

$$\Pr[\widetilde{\mathbf{L_{ij}}}\,|\,(\mathbf{U_{ij}} \wedge \widetilde{\kappa})] = \Pr[\widetilde{\mathbf{L_{ij}}}\,|\,(\mathbf{U_{ij}} \wedge \widetilde{\kappa}) \wedge \mathbf{\Pi(K_{ij})}] \times \Pr[\mathbf{\Pi(K_{ij})}|\,\mathbf{U_{ij}}]. \quad (10)$$

Henceforth, K_{ij} will implicitly be a path key. Let $\mathbf{\Pi_2(K_{ij})}$ be the event that the path key K_{ij} is generated using two agents, and $\mathbf{\Pi_1(K_{ij})}$ be the event that the path key K_{ij} is generated using a single agent, as in the case when s_i or s_j is itself the agent for the other's group. Under random attack, the attacker compromises a random sensor at each step, so the event of s_i or s_j being compromised is independent of the event that K_{ij} is a path key. In other words, $\Pr[\mathbf{\Pi(K_{ij})}|\mathbf{U_{ij}}] = \Pr[\mathbf{\Pi(K_{ij})}]$, $\Pr[\mathbf{\Pi_1(K_{ij})}|\,\mathbf{U_{ij}}] = \Pr[\mathbf{\Pi_1(K_{ij})}]$, and $\Pr[\mathbf{\Pi_2(K_{ij})}|\,\mathbf{U_{ij}}] = \Pr[\mathbf{\Pi_2(K_{ij})}]$. Consequently, we can rewrite Equation 10 as

$$\begin{aligned}\Pr[\widetilde{\mathbf{L_{ij}}}|(\mathbf{U_{ij}} \wedge \widetilde{\kappa})] = \Pr[\mathbf{\Pi(K_{ij})}] \times \big(\Pr[\widetilde{\mathbf{L_{ij}}}|(\mathbf{U_{ij}} \wedge \widetilde{\kappa}) \wedge \mathbf{\Pi_1(K_{ij})}] \times \Pr[\mathbf{\Pi_1(K_{ij})}] \\ + \Pr[\widetilde{\mathbf{L_{ij}}}|(\mathbf{U_{ij}} \wedge \widetilde{\kappa}) \wedge \mathbf{\Pi_2(K_{ij})}] \times \Pr[\mathbf{\Pi_2(K_{ij})}]\big). \quad (11)\end{aligned}$$

Let there be g groups each of size γ, and let each sensor hold μ preloaded keys for sensors in other groups. As shown in Section 5.2.1, each group has $t = \frac{\mu\gamma}{g-1}$ agents in every other group. If α is the probability that a given sensor is an agent for the other's group, then

$$\alpha = \frac{\binom{\gamma-1}{t-1}}{\binom{\gamma}{t}} = \frac{t}{\gamma}.$$

$\Pr[\mathbf{\Pi(K_{ij})}]$ is the ratio of the number of path keys to the total number of keys among all pairs of communicating sensors s_i and s_j. We note first that when K_{ij} is a path key, $L_{\mathbf{ij}}$'s endpoints s_i and s_j must belong to different groups. We take $s_i \in G_1$ and $s_j \in G_2$ to be *any* pair of sensors from different groups G_1, G_2. Now, K_{ij} is a path key unless s_i and s_j are both agents sharing a preloaded key, which event occurs with probability $\frac{\alpha^2}{t}$. Thus, $\Pr[\mathbf{\Pi(K_{ij})}] = 1 - \frac{\alpha^2}{t}$. Event $\mathbf{\Pi_1(K_{ij})}$ occurs if either

E_1. exactly one of s_i or s_j is an agent for the other's group, or
E_2. both s_i and s_j are agents for the other's group, but they share no preloaded key.

The probability of E_1 is $2\alpha(1-\alpha)$. For E_2, the probability that s_i and s_j are each agents for the other's group is α^2. The t agents G_1 and G_2 have for each other can be matched pairwise in $t!$ ways. Further, there are $(t-1)(t-1)!$ matchings in which any given agent $s_i \in G_1$ is not paired with agent $s_j \in G_2$. Therefore, the probability of E_2 is

$$\alpha^2 \frac{(t-1)(t-1)!}{t!} = \alpha^2 \frac{t-1}{t}, \text{ so that}$$

$$\Pr[\mathbf{\Pi_1(K_{ij})}] = \frac{2\alpha(1-\alpha) + \alpha^2 \frac{t-1}{t}}{1 - \frac{\alpha^2}{t}}.$$

Event $\mathbf{\Pi_2(K_{ij})}$ occurs when neither s_i nor s_j is an agent for the other's group, so

$$\Pr[\mathbf{\Pi_2(K_{ij})}] = \frac{(1-\alpha)^2}{1 - \frac{\alpha^2}{t}}.$$

Let there be n_s static sensors, of which x are compromised. The probability that the agent used to transmit the path key K_{ij} is not compromised, when s_i and s_j are uncompromised is $\binom{n_s-3}{\kappa}/\binom{n_s-2}{\kappa}$. Thus $\Pr[\widetilde{\mathbf{L_{ij}}} \,|\, (\mathbf{U_{ij}} \wedge \widetilde{\boldsymbol{\kappa}}) \wedge \mathbf{\Pi_1(K_{ij})}]$ can be computed as

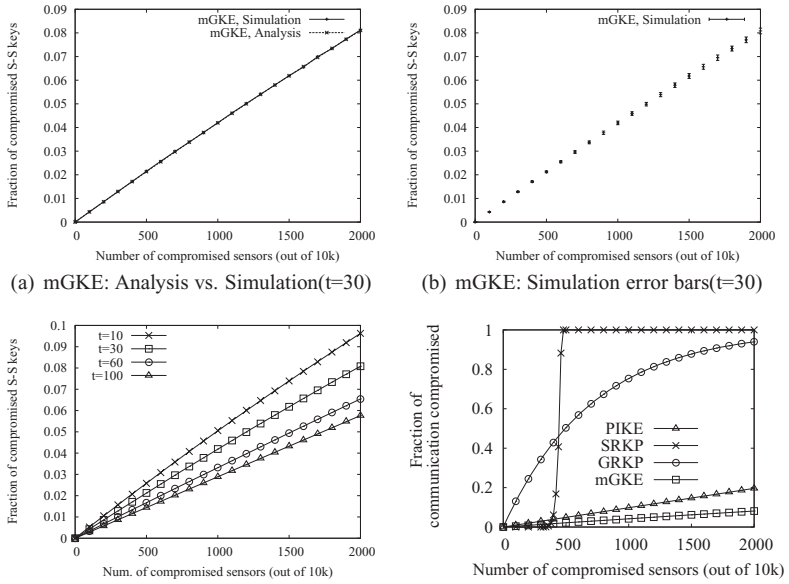
$$\Pr[\widetilde{\mathbf{L_{ij}}} \,|\, (\mathbf{U_{ij}} \wedge \widetilde{\boldsymbol{\kappa}}) \wedge \mathbf{\Pi_1(K_{ij})}] = 1 - \frac{\binom{n_s-3}{\kappa}}{\binom{n_s-2}{\kappa}} = \frac{\kappa}{n_s - 2}.$$

Similarly,

$$\Pr[\widetilde{\mathbf{L_{ij}}} \,|\, (\mathbf{U_{ij}} \wedge \widetilde{\boldsymbol{\kappa}}) \wedge \mathbf{\Pi_2(K_{ij})}] \;=\; 1 - \frac{\binom{n_s-4}{\kappa}}{\binom{n_s-2}{\kappa}} = 1 - \frac{(n_s - \kappa - 2)^{\underline{2}}}{(n_s-2)^{\underline{2}}},$$

where $a^{\underline{k}}$ is the falling factorial function $a(a-1)\cdots(a-k+1)$. Equation (11) now becomes

$$\Pr[\widetilde{\mathbf{L_{ij}}} \,|\, (\mathbf{U_{ij}} \wedge \widetilde{\boldsymbol{\kappa}})] = \left\{(1-\alpha)^2 \left(1 - \frac{(n_s - 2 - \kappa)^{\underline{2}}}{(n_s-2)^{\underline{2}}}\right) + \left(2\alpha(1-\alpha) + \alpha^2 \frac{t-1}{t}\right)\left(\frac{\kappa}{n_s - 2}\right)\right\} \times \frac{\Pr[\mathbf{\Pi(K_{ij})}]}{1 - \frac{\alpha^2}{t}}.$$



(a) mGKE: Analysis vs. Simulation(t=30) (b) mGKE: Simulation error bars(t=30)

(c) Link compromise vs. t (d) mGKE vs. SRKP, GRKP, PIKE

Fig. 10. Links compromised between uncompromised sensors ($n_s = 10^4$, $\delta = 50$).

We estimate Pr[$\mathbf{\Pi(K_{ij})}$] both analytically and experimentally in Section 8.2.1.

Figure 10(a) shows that our analysis for the number of link compromises closely matches simulation results. Figure 10(b) plots the error bars over 100 trials. There is minimal variance, so we can compare average values from simulations.

Figure 10(c) shows that mGKE's resilience drops modestly as t decreases. Lowering t by a factor of 10 (from 100 to 10) increases the fraction of compromised S-S keys only by a factor of around 2. mGKE is clearly a good choice when sensor memory is very limited, since we can afford to reduce t without unduly compromising security.

Figure 10(d) compares the resilience of mGKE with that of SRKP [Du et al. 2003], GRKP [Du et al. 2004], and PIKE [Chan and Perrig 2005]. SRKP resilience is computed using the analysis in Du et al. [2003], preloading each sensor with 200 keys drawn from 4 key spaces chosen from 50 key spaces. GRKP resilience is computed as in Du et al. [2004], with a key space size of 100,000 and connectivity of 99.99%. (mGKE's connectivity is 100%. See Section 6.3.) The analysis in Du et al. [2004] only considers links secured by preloaded keys, so that the fraction of compromised links in GRKP will be even higher if we include path keys. Under this definition, the resilience graphs of PIKE and mGKE would both be lines of zero, representing perfect resilience. We handicap mGKE and PIKE by including both preloaded keys and path keys.

In Figure 10(d), SRKP's resilience drops dramatically when around 350 of 10,000 sensors are compromised. PIKE and mGKE show graceful degradation of resilience, so that the system is not compromised by compromising a a few

sensors. Figure 10(d) shows that the resilience of mGKE is about twice as high as that of PIKE, since a significantly larger fraction of links are secured by pairwise keys in mGKE (see Section 8.2). In Section 8.2, we show that mGKE also has significantly lower communication overhead than PIKE.

6.2 M-S Keys Shared between Mobile and Static Nodes

Let $s_i \in G_u$ be an uncompromised sensor and m_j be an uncompromised mobile collector. The link $L_{\mathbf{ij}}$ between nodes s_i and m_j is compromised only if K_{ij} is a path key established using a compromised sensor. Let $\mathbf{\Pi_a}(\mathbf{K_{ij}})$ be the event that m_j is associated with G_u but s_i is not associated with m_j. In this case, K_{ij} is established through a sensor s_k, $k \neq i$ in G_u. Let $\mathbf{\Pi_{\bar{a}}}(\mathbf{K_{ij}})$ be the event that m_j is not associated with G_u, so that K_{ij} must be established using an intermediary $s_k \in G_v$ with which m_j is associated, and G_v is a nearby group. The probability $\Pr[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}}]$ of $\mathrm{L_{ij}}$ with κ sensors compromised is

$$\Pr\left[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}} \wedge \mathbf{\Pi_a}(\mathbf{K_{ij}})\right] \times \Pr\left[\mathbf{\Pi_a}(\mathbf{K_{ij}})\right] + \Pr\left[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}} \wedge \mathbf{\Pi_{\bar{a}}}(\mathbf{K_{ij}})\right] \times \Pr\left[\mathbf{\Pi_{\bar{a}}}(\mathbf{K_{ij}})\right].$$

Since m_j is associated with g' out of g groups, we get

$$\Pr[\mathbf{\Pi_a}(\mathbf{K_{ij}})] = \left(\frac{g'}{g}\right) \text{ and } \Pr[\mathbf{\Pi_{\bar{a}}}(\mathbf{K_{ij}})] = 1 - \frac{g'}{g}.$$

Proceeding as in the analysis for $L_{\mathbf{ij}}$ between static sensors, we get

$$\Pr[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}} \wedge \mathbf{\Pi_a}(\mathbf{K_{ij}})] = (1-\alpha)\left(1 - \frac{\binom{n_s-2}{\kappa}}{\binom{n_s-1}{\kappa}}\right) = (1-\alpha)\frac{\kappa}{n_s-1}, \text{ and}$$

$$\Pr[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}} \wedge \mathbf{\Pi_{\bar{a}}}(\mathbf{K_{ij}})] = \frac{\alpha^2}{t}\left(1 - \frac{\binom{n_s-2}{\kappa}}{\binom{n_s-1}{\kappa}}\right) + (1-\alpha)^2\left(1 - \frac{\binom{n_s-4}{\kappa}}{\binom{n_s-1}{\kappa}}\right) + \left[2\alpha(1-\alpha) + \alpha^2\left(1-\frac{1}{t}\right)\right]\left(1 - \frac{\binom{n_s-3}{\kappa}}{\binom{n_s-1}{\kappa}}\right).$$

Combining these expressions, simplifying, and using the falling factorial notation,

$$\Pr[\widetilde{\mathbf{L}_{\mathbf{ij}}}|\,\widetilde{\boldsymbol{\kappa}}] = \frac{g'(1-\alpha)\kappa}{g(n_s-1)} + \left(1-\frac{g'}{g}\right)\left\{\frac{\alpha^2\kappa}{t(n_s-1)} + (1-\alpha)^2\left(1 - \frac{(n_s-\kappa-1)^{\underline{3}}}{(n_s-1)^{\underline{3}}}\right) + \left[2\alpha(1-\alpha) + \alpha^2\left(1-\frac{1}{t}\right)\right]\left(1 - \frac{(n_s-\kappa-1)^{\underline{2}}}{(n_s-1)^{\underline{2}}}\right)\right\}.$$

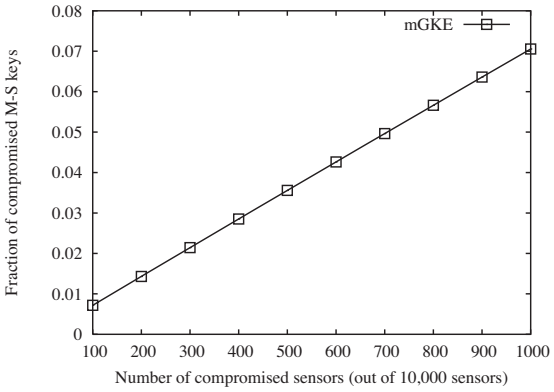


Fig. 11. Links compromised.

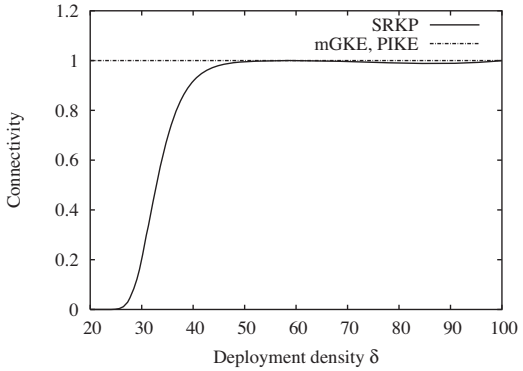


Fig. 12. Connectivity.

Figure 11 shows that in mGKE, the resilience of links between static and mobile nodes degrades linearly with the number of compromised static sensors, which is the best theoretically possible. We used $n_s = 10^4$, $\delta = 50$, $\frac{g'}{g} = 0.3$, $\alpha = \frac{t}{\gamma} = 0.3$. It is not meaningful to compare our scheme with SRKP and PIKE. SRKP's resilience degrades dramatically even for the static case. PIKE needs a globally addressable infrastructure to find intermediaries, and cannot be directly adopted to support mobile sensor networks.

6.3 Resilience and Connectivity

RKP and SRKP require high density deployments to ensure path key establishment with high probability [Hwang and Kim 2004; Chan and Perrig 2005]. In contrast, any two sensors can establish a path key in mGKE, regardless of the sensor distribution, given a physically connected network. Figure 12 compares SRKP, PIKE and mGKE connectivity for a 10,000-sensor network. For SRKP, each sensor has 4 key spaces chosen from a pool of 50 key spaces, and preloaded with 200 keys, a typical configuration from Du et al. [2003]. SRKP connectivity decreases dramatically for sensor densities less than 50, and is almost surely

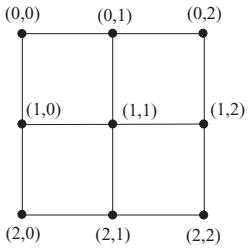


Fig. 13. PIKE.

disconnected when the density is around 25. In contrast, PIKE and mGKE retain full connectivity regardless of sensor density. Remarkably, only 55 keys are required for the mGKE scheme to achieve full connectivity among static sensors when groups are 10-associated (See Section 8.1).

7. ANALYZING SELECTIVE ATTACKS USING ORDER STATISTICS

Sensor deployments are far more vulnerable to selective attack than to random attack. Yet no analysis of selective attack has been attempted in the literature, since it is technically challenging. We will now present a general framework for analyzing selective attack, and apply it to PIKE and mGKE.

An attacker can accelerate compromise of communication links by targeting, at each step, the sensor s_t whose compromise reveals the largest number of unknown pairwise keys. The attacker gains all preloaded keys at s_t, *and all path keys mediated by* s_t, since he may record earlier messages. Let $[s_j, s_k]$ represent the path key between s_j and s_k, and let $M(s_i) = \{[s_{i_{11}}, s_{i_{12}}], [s_{i_{21}}, s_{i_{22}}], \cdots\}$ be the set of path keys mediated by s_i.

7.1 The Yield Metric

Let $\mathcal{S} = \{s_1, s_2, \ldots, s_{n_s}\}$ be the set of sensors, and let $\mathcal{C} \subseteq \mathcal{S}$ and $\mathcal{U} \subseteq \mathcal{S}$ be the set of compromised and uncompromised sensors, respectively. Initially, $\mathcal{C}$ is empty. The *yield* $Y_{\mathcal{C}}(s_i)$ of sensor s_i represents how much *new* key information the compromise of s_i would reveal about the *rest* of the network, given that the sensors in $\mathcal{C}$ have been compromised. Since all keys involving nodes in $\mathcal{C}$ are already known, we define the yield as

$$Y_{\mathcal{C}}(s_i) = M(s_i) \setminus \{[s_j, s_k] \mid s_j \in \mathcal{C} \text{ or } s_k \in \mathcal{C}\}. \tag{12}$$

$Y_{\mathcal{C}}(s_i)$ would be defined differently for each key establishment scheme. Each selective attack step targets the sensor s_t with the largest yield. That is, $Y_{\mathcal{C}}(s_t) = \max\{Y_{\mathcal{C}}(s_i)\}, s_i \in \mathcal{U}$. Now, we show how to define this metric for PIKE and mGKE.

7.2 Selective Attack Illustrated

Figure 13 shows the 3×3 logical grid for a 9-sensor network using PIKE. For simplicity, we use a sensor's logical position as its ID. Sensors on the same logical row or column share preloaded keys. Other neighboring sensors can establish

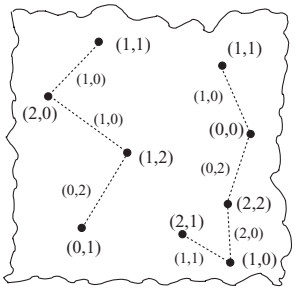


Fig. 14. PIKE Path keys.

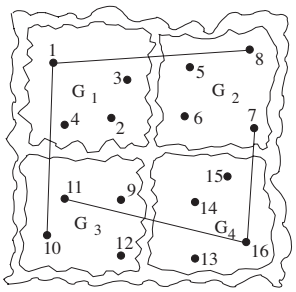


Fig. 15. mGKE associations ($t = 1$).

a path key via a intermediary on the same logical row or column. When this intermediate sensor is compromised, the path key that is established via this intermediate sensor will also be compromised. However, a preloaded key is compromised only if one of the two sensors sharing the key is compromised.

Figure 14 shows how path keys work in PIKE. The dashed line between two sensors denotes the path key they share, with the sensor ID near the dashed line denoting the trusted intermediary that facilitates the path key establishment. For example, the neighboring sensor pair (1,1) and (2,0) establish a path key via the intermediary (1,0). Figure 14 illustrates that initially, the yield of (1,0) is 3, and the yields of (0,2), (1,2), and (0,1) are all 1. All other sensors have yields of 0. Thus a clever attacker will select the next target as sensor (1,0), as it mediates the largest number of path keys, and has the largest yield.

In mGKE, sensors from the same group share pairwise keys. Agents mediate path keys between sensor pairs not sharing preloaded keys. Figure 15 shows the group associations in mGKE. Group G_1 and G_2 are associated via sensors 1 and 8. Figure 16 shows path keys in the 16-sensor network. Sensor pairs (3,5), (3, 6) and (2, 6) establish path keys via agent pairs (1, 8). Since sensor 1 is also the agent for group G_3, it also mediates path keys between the neighboring pairs (4, 11), and (2, 9). Thus the initial yield of sensor 1 is 5. Similarly, the initial yield of sensor 8 and 16 is 3, that of sensor 10 and 11 is 2, that of sensor 7 is 1, and others is 0. The attacker will attack sensor 1 next.

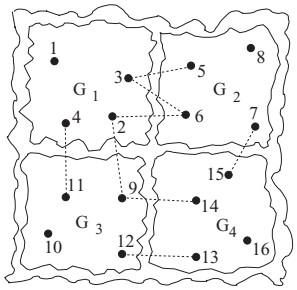


Fig. 16. mGKE path keys.

7.3 Framework for Analysis of Resilience under Selective Attack

Since PIKE does not allow mobility, we consider a static deployment of n_s sensors. For convenience, we will write $\kappa = |\mathcal{C}|$. We bound the number of compromises by assuming $\kappa \leq \rho n_s$ for some ρ. Let $s_t \in \mathcal{U}$, having yield $Y_{\mathcal{C}}(s_t) = \max_i\{Y_{\mathcal{C}}(s_i)\}$, be the node targeted next (see Section 7.2). We will call $Y_{\mathcal{C}}(s_t)$ the *marginal yield* of the next compromise given $\mathcal{C}$, and write $\widehat{Y}_{\mathcal{C}} = Y_{\mathcal{C}}(s_t)$.

Let s_i mediate a path key between a node pair (s_u, s_v). and let s_i have yield $Y_{\emptyset}(s_i)$ when $\mathcal{C} = \emptyset$. Consider a later time, when $\mathcal{C} \neq \emptyset$. If $s_i \notin \mathcal{C}$, but $s_u \in \mathcal{C}$, it will have a lower yield $Y_{\mathcal{C}}(s_i)$, since s_i's compromise will yield fewer fresh keys (see Equation (12)). Hence, $Y_{\mathcal{C}}(s_i) \leq Y_{\emptyset}(s_i)$.

Let the n_s nodes have initial yields $Y_{\emptyset}(s_1), \ldots, Y_{\emptyset}(s_{n_s})$. Let these yields now be sorted into the list $Y_{\emptyset}^{\langle 1\rangle} \leq Y_{\emptyset}^{\langle 2\rangle} \leq \cdots \leq Y_{\emptyset}^{\langle n_s\rangle}$. The value $Y_{\emptyset}^{\langle k\rangle}$ is called the *k-th order statistics* [Arnold et al. 1993] of the n_s random variables $Y_{\emptyset}(s_i)$. Similarly, after the κ nodes in $\mathcal{C}$ are compromised, let the yields of the $n_s - \kappa$ uncompromised nodes be sorted into $Y_{\mathcal{C}}^{\langle 1\rangle} \leq Y_{\mathcal{C}}^{\langle 2\rangle} \leq \cdots \leq Y_{\mathcal{C}}^{\langle n_s-\kappa\rangle}$.

Under selective attack, the attacker always picks the node with maximum yield as its target. Hence, the first node compromised has yield $Y_{\emptyset}^{\langle n_s\rangle}$. After the set $\mathcal{C}$ of nodes is compromised, the next node targeted in set $\mathcal{U}$ has yield $Y_{\mathcal{C}}^{\langle n_s-|\mathcal{C}|\rangle}$. In other words, the marginal yield given $\mathcal{C}$ is always

$$\widehat{Y}_{\mathcal{C}} = Y_{\mathcal{C}}^{\langle n_s-\kappa\rangle}, \quad \text{where } \kappa = |\mathcal{C}|.$$

7.4 Estimating $Y_{\mathcal{C}}$

Since $\mathcal{C} = \emptyset$ immediately after deployment, the initial yield $Y_{\emptyset}(s_i)$ of s_i equals the number of path keys it mediates for the *entire network*. Subsequent node compromises make $\mathcal{C} \neq \emptyset$, lowering s_i's yield to $Y_C(s_i)$. Denote this decrease by $D_{\mathcal{C}}(s_i) = Y_{\emptyset}(s_i) - Y_C(s_i)$ for a node $s_i \in \mathcal{U}$, with $Y_C(s_i) = 0$ for $s_i \in \mathcal{C}$.

A link is compromised as soon as one of its endpoints is compromised. Let $s_i \in \mathcal{U}$ mediate a path key for an uncompromised link (s_u, s_v). Clearly, s_i's yield drops by 1 when the first of s_u or s_v is compromised, but drops no further when the other endpoint is compromised. The existence of each such compromised link indicates that the the yield of its (uncompromised) intermediary has dropped by 1, that is, that $D_{\mathcal{C}}(s_i)$ has increased by 1.

The initial yield $Y_{\emptyset}(s_i)$ is the number of path keys s_i mediates between the other $n_s - 1$ sensors. In contrast, $D_{\mathcal{C}}(s_i)$ is the number of path keys s_i mediates for links already having at least one end point in $\mathcal{C}$. When ρ is modest ($\rho \leq 0.2$, say), $|\mathcal{C}|$ is also small relative to n_s. If s_i's initial yield $Y_{\emptyset}(s_i)$ is one of the ρn_s largest yields $Y_{\emptyset}^{\langle n_s \rangle}, Y_{\emptyset}^{\langle n_s - 1 \rangle}, \ldots, Y_{\emptyset}^{\langle n_s - \rho n_s \rangle}$, $D_{\mathcal{C}}(s_i)$ can be expected to be small relative to $Y_{\emptyset}(s_i)$.

In uniform deployments, nodes with the highest initial yields are also likely distributed uniformly. Selective attack targets nodes with the highest yields first, so sensor compromises are likely distributed uniformly, so the yield distribution flattens out quickly. Hence, node compromises are likely to have similar effects on the yields of all nodes in $\mathcal{U}$, and the variance in $D_{\mathcal{C}}(s_i)$ is likely small compared to the initial yield values $Y_{\emptyset}(s_i)$. Hence, $D_{\mathcal{C}}(s_i)$ is likely to affect $Y_{\mathcal{C}}^{\langle n_s - \kappa \rangle}$ less than $Y_{\emptyset}(s_i)$ does.

Let the nodes $s_{\emptyset}^{\langle 1 \rangle}, s_{\emptyset}^{\langle 2 \rangle}, \ldots, s_{\emptyset}^{\langle n_s \rangle}$ have initial yields $Y_{\emptyset}^{\langle 1 \rangle}, Y_{\emptyset}^{\langle 2 \rangle}, \ldots, Y_{\emptyset}^{\langle n_s \rangle}$, and the nodes $s_{\mathcal{C}}^{\langle 1 \rangle}, s_{\mathcal{C}}^{\langle 2 \rangle}, \ldots, s_{\mathcal{C}}^{\langle n_s - \kappa \rangle}$ have yields $Y_{\mathcal{C}}^{\langle 1 \rangle}, Y_{\mathcal{C}}^{\langle 2 \rangle}, \ldots, Y_{\mathcal{C}}^{\langle n_s - \kappa \rangle}$. We use the approximation $Y_{\mathcal{C}}^{\langle n_s - \kappa \rangle} = Y_{\mathcal{C}}(s_{\mathcal{C}}^{\langle n_s - \kappa \rangle}) \approx Y_{\emptyset}(s_{\emptyset}^{n_s - \kappa \rangle}) - D_{\mathcal{C}}(s_{\emptyset}^{\langle n_s - \kappa \rangle})$, so that

$$E[\widehat{Y}_{\mathcal{C}}] \approx E\left[Y_{\emptyset}^{\langle n_s - \kappa \rangle}\right] - E\left[D_{\mathcal{C}}\left(s_{\emptyset}^{\langle n - \kappa \rangle}\right)\right]. \tag{13}$$

The expectation of k-th order statistics $Y_{\emptyset}^{\langle k \rangle}$ of $Y_{\emptyset}(s_i)$, $0 \leq i < n_s$, where all $Y_{\emptyset}(s_i)$ are identically and independently distributed between $[0, n_s^2]$ can be obtained as follows [Arnold et al. 1993]. Let $F(y)$ denote the cumulative distribution function (CDF) of $Y_{\emptyset}(s_i)$, so that $F(y) = \Pr(Y_{\emptyset}(s_i) \leq y)$. Now, if $Y_{\emptyset}^{\langle k \rangle}$ is no more than y, at least k of the $Y_{\emptyset}(s_i)$ are no more than y. Therefore, the cumulative distribution function of $Y_{\emptyset}^{\langle k \rangle}$ is

$$F_{\emptyset}^{\langle k \rangle}(y) = \sum_{j=k}^{n} F(y)^j (1 - F(y))^{n-j}.$$

Now, the expectation of $Y_{\emptyset}^{\langle k \rangle}$ is computed as

$$\begin{aligned} E[Y_{\emptyset}^{\langle k \rangle}] &= \sum_{y=0}^{n_s^2} y\left[F_{\emptyset}^{\langle k \rangle}(y) - F_{\emptyset}^{\langle k \rangle}(y-1)\right] = \sum_{y=0}^{n_s^2} \left[1 - F_{\emptyset}^{\langle k \rangle}(y)\right] \\ &= \sum_{y=0}^{n_s^2} \left[1 - \sum_{j=k}^{n_s} F(y)^j (1 - F(y))^{n_s - j}\right]. \end{aligned}$$

Using Equation 13, we obtain

$$E[\widehat{Y}_{\mathcal{C}}] \approx \sum_{y=0}^{n_s^2} \left(1 - \sum_{j=n_s-\kappa}^{n_s} F(y)^j (1 - F(y))^{n_s - j}\right) - \bar{D}_{\mathcal{C}}. \tag{14}$$

We now compute the CDF of the initial yields of nodes and $\bar{D}_{\mathcal{C}}$ for PIKE and mGKE.

7.5 Analysis of Resilience to Selective Attack for PIKE

Let node s_i be at row u and column v in a PIKE logical grid. If some node s_u from row u and some node s_v from column v are neighbors, we know that

s_i will act as the path-key intermediary for link (s_u, s_v) with probability 1/2. There are $(\sqrt{n_s}-1)^2$ such (s_u, s_v) pairs, but s_i will mediate path keys only for neighboring pairs. Since s_u and s_v are distributed uniformly in the region, they are neighbors with probability $\pi r^2/A$, where r is the communication radius and A the region size. Clearly, the yield $Y_\emptyset(s_i)$ follows a Binomial distribution with success probability $\frac{1}{2}(\frac{\pi r^2}{A})$ over $(\sqrt{n_s}-1)^2$ trials. In practice, $A \approx 1000m \times 1000m$, and $r \approx 40m$, so $\pi r^2 \ll A$, so we can approximate the Binomial as a Gaussian with mean $\mu_P = (\sqrt{n_s}-1)^2 \frac{\pi r^2}{2A}$ and variance $\sigma_P = (\sqrt{n_s}-1)^2 \frac{\pi r^2}{2A}(1-\frac{\pi r^2}{2A})$. Now, $Y_\emptyset(s_i)$ has density

$$y_\emptyset(x) = \frac{1}{\sigma_P\sqrt{2\pi}} e^{-\frac{(x-\mu_P)^2}{2\sigma_P^2}}. \tag{15}$$

To estimate $E[D_{\mathcal{C}}(s_\emptyset^{\langle n_s-\kappa\rangle})]$, we note that there are $\kappa(n_s-\kappa)$ pairs (s_u, s_v) with exactly one of s_u and s_v in $\mathcal{C}$, and $\frac{1}{2}\kappa(\kappa-1)$ pairs with both s_u and s_v in $\mathcal{C}$. Sensor $s_\emptyset^{\langle n_s-\kappa\rangle}$ mediates $Y_\emptyset(s_\emptyset^{\langle n_s-\kappa\rangle})$ path keys, so it mediates the path key for any given link with probability

$$\omega(s_\emptyset^{\langle n_s-\kappa\rangle}) = E\left[\frac{Y_\emptyset\left(s_\emptyset^{\langle n_s-\kappa\rangle}\right)}{\sum_{s_i\in\mathcal{S}} Y_\emptyset(s_i)}\right] \approx \frac{E\left[Y_\emptyset(s_\emptyset^{\langle n_s-\kappa\rangle})\right]}{E[\sum_{s_i\in\mathcal{S}} Y_\emptyset(s_i)]} = \frac{E[Y_\emptyset^{\langle n_s-\kappa\rangle}]}{n_s\mu_P}. \tag{16}$$

We could use equality above if $Y_\emptyset(s_\emptyset^{\langle n_s-\kappa\rangle})$ and $\sum_{s_i\in\mathcal{S}} Y_\emptyset(s_i)$ were stochastically independent [Dubey 1963]. Our approximation is nonetheless justified; the sum $\sum_{s_i\in\mathcal{S}} Y_\emptyset(s_i)$ is determined almost entirely by terms other than $Y_\emptyset(s_\emptyset^{\langle n_s-\kappa\rangle})$, since $|\mathcal{S}| \gg 1$.

Next, we note that the probability that any pair (s_u, s_v) is a link is $\pi r^2/A$, the probability that a link uses a path key is $\mathbf{\Pi(K_{ij})}$ (see Section 8.2.1), and the probability that this path-key intermediary is $s_\emptyset^{\langle n_s-\kappa\rangle}$ is $\omega(s_\emptyset^{\langle n_s-\kappa\rangle})$. Hence, $D_{\mathcal{C}}(s_i)$ is a Binomial process with success probability $\frac{\pi r^2}{A}\mathbf{\Pi(K_{ij})}\omega(s_\emptyset^{\langle n_s-\kappa\rangle})$ over $\kappa(n_s-\kappa)+\frac{1}{2}\kappa(\kappa-1)$ trials, or

$$D_{\mathcal{C}}\left(s_\emptyset^{\langle n_s-\kappa\rangle}\right) \sim B\left(\kappa(n_s-\kappa)+\frac{\kappa(\kappa-1)}{2},\ \frac{\pi r^2}{A}\cdot\mathbf{\Pi(K_{ij})}\cdot\omega\left(s_\emptyset^{\langle n_s-\kappa\rangle}\right)\right).$$

It follows that

$$E\left[D_{\mathcal{C}}\left(s_\emptyset^{\langle n_s-\kappa\rangle}\right)\right] = \left(\kappa(n_s-\kappa)+\frac{\kappa(\kappa-1)}{2}\right)\frac{\pi r^2}{A}\cdot\mathbf{\Pi(K_{ij})}\cdot\omega\left(s_\emptyset^{\langle n_s-\kappa\rangle}\right) \tag{17}$$

The marginal yield $E[\widehat{Y}_{\mathcal{C}}]$ with respect to κ is now readily computed from Equations (14), (15), (16), and (17) (see Figure 18(c)).

7.6 Analysis of Resilience to Selective Attack for mGKE

We now estimate the initial yields $Y_\emptyset(s_i)$ for s_i in group G_0, under mGKE. Each group has t pairs of agents for every other group, and each node s_i is associated with μ groups. Node s_i functions as an intermediary with probability $1/t$ for the path key between s_u and s_v, under three conditions. First, s_u must be from

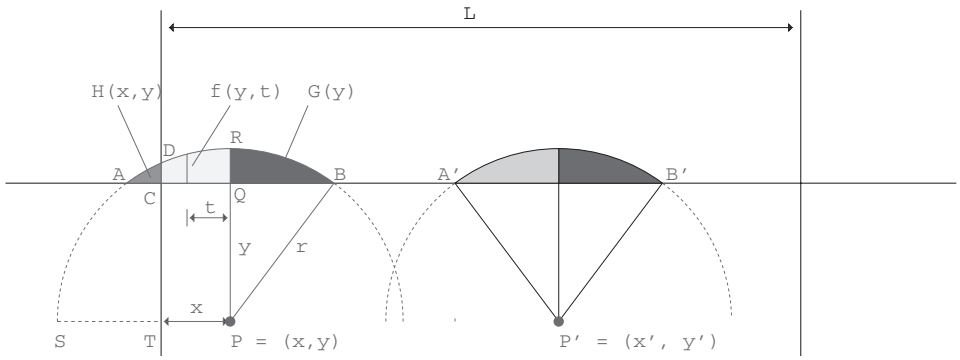


Fig. 17. Computing $p^{\oplus}$ and $p^{\otimes}$.

the same group G_0 as s_i. Second, s_v must belong to a group which s_i is associated with and is adjacent to group G_0. Third, s_u and s_v must be neighbors. Ignoring boundary effects in the monitored area, each group has four horizontally/vertically adjacent groups, and four diagonally adjacent groups.

We call two groups are $\oplus$-adjacent if they are vertically or horizontally adjacent, and $\otimes$-adjacent if they are diagonally adjacent. Let $0 \leq N^{\oplus} \leq 4$ be the number of groups $\oplus$-adjacent to group G_0 and associated with s_i. Let $0 \leq N^{\otimes} \leq 4$ be the number of groups $\otimes$-adjacent to group G_0 and associated with s_i. Let $p^{\oplus}$ be the probability that an $s_u \in G_0$ neighbors an s_v from a $\oplus$-adjacent group. Let $p^{\otimes}$ be the probability that an $s_u \in G_0$ neighbors an s_v from a $\otimes$-adjacent group. We now estimate $p^{\oplus}$ and $p^{\otimes}$.

Figure 17 shows a section of grid, with cells of size $L \times L$. Let sensor s_i have range r, and be deployed at position $P = (x, y)$, with x measured from the left edge of the cell and y measured from its top edge. A sensor from the vertically adjacent grid cell will be a neighbor of s_i if and only if it lies in the truncated circular segment CDRBQC. A sensor from the diagonally adjacent grid cell will be a neighbor of s_i if and only if it lies in the truncated segment ACDA. Let us denote the area of any region W by $||W||$.

The probabilities $p^{\oplus}$ and $p^{\otimes}$ are the expectations, over all x and y, of $\frac{||CDRBQC||}{L^2}$ and $\frac{||ACDA||}{L^2}$. Clearly, $||CDRBQC|| = ||CDRQC|| + ||BQRB|| = ||CDRQC|| + G(y)$. Also, $f(y,t) = \sqrt{r^2 - t^2} - y$, and $BQ = \sqrt{r^2 - y^2}$. The following three cases arise.

The circular segment is truncated by the cell's left edge when $x < \sqrt{r^2 - y^2}$, and by the right edge when $x > L - \sqrt{r^2 - y^2}$. When $\sqrt{r^2 - y^2} \leq x \leq L - \sqrt{r^2 - y^2}$, the segment is untruncated, as for a sensor at position P' in Figure 17. We start by noting that

$$\int f(y,t)dt = \int \left(\sqrt{r^2 - t^2} - y\right) dt = \frac{1}{2}\left(t\sqrt{r^2 - t^2} + r^2 \sin^{-1}\left(\frac{t}{r}\right) - 2ty\right).$$

If we define $F(x, y) = ||CQRDC||$, we can write

$$G(y) = \int_0^{\sqrt{r^2-y^2}} f(y,t)\,dt = \frac{1}{2}\left(r^2\cos^{-1}\left(\frac{y}{r}\right) - y\sqrt{r^2-y^2}\right)$$

$$F(x, y) = \int_0^{x} f(y,t)\,dt = \frac{1}{2}\left(x\sqrt{r^2-x^2} + r^2\sin^{-1}\left(\frac{x}{r}\right) - 2xy\right).$$

For a sensor at position (x, y), let the overlap area of the circular segment with the vertically adjacent cell be $\Omega^{\oplus}(x, y)$. We have the following cases.

$$\Omega^{\oplus}(x, y) = \begin{cases} 0 & \text{if } y \geq r, \\ 2G(y) & \text{if } y < r, \text{ and } \sqrt{r^2-y^2} \leq x \leq L - \sqrt{r^2-y^2}, \\ F(x, y) + G(y) & \text{if } y < r, \text{ and } x < \sqrt{r^2-y^2} \text{ or } x \leq L - \sqrt{r^2-y^2}. \end{cases}$$

The total overlap are with the vertically disposed cell for sensors in the region $0 \leq y \leq r$ is therefore

$$\begin{aligned} \Omega^{\oplus} &= \int_0^r \left(2\int_0^{\sqrt{r^2-y^2}} \big(F(x, y) + G(y)\big)\,dx + \int_{\sqrt{r^2-y^2}}^{L-\sqrt{r^2-y^2}} 2G(y)\,dx\right) dy \\ &= \int_0^r \int_0^{\sqrt{r^2-y^2}} 2F(x, y)\,dx\,dy + \int_0^r \left(L - \sqrt{r^2-y^2}\right) 2G(y)dy. \end{aligned} \tag{18}$$

We note that

$$L\int_0^r 2G(y)\,dy = L\left(\frac{(r^2-y^2)^{\frac{3}{2}}}{3} - r^2\sqrt{r^2-y^2} + r^2 y\cos^{-1}\left(\frac{y}{r}\right)\right)\Big|_0^r = \frac{2Lr^3}{3} \tag{19}$$

and that

$$\begin{aligned} \int_0^{\sqrt{r^2-y^2}} 2F(x, y)dx = \Big(r^2\sqrt{r^2-x^2} + r^2 x\sin^{-1}\left(\frac{x}{r}\right) - \frac{(r^2-x^2)^{\frac{3}{2}}}{3} \\ + y(r^2-x^2)\Big)\Big|_0^{\sqrt{r^2-y^2}} = \frac{2}{3}(y^3-r^3) + r^2\sqrt{r^2-y^2}\cos^{-1}\left(\frac{y}{r}\right), \end{aligned}$$

so that

$$\int_0^r \int_0^{\sqrt{r^2-y^2}} 2F(x, y)dx\,dy = -\frac{r^4}{2} + r^2\int_0^r \sqrt{r^2-y^2}\cos^{-1}\left(\frac{y}{r}\right)dy \tag{20}$$

Further,

$$\int_0^r \sqrt{r^2-y^2}\,2G(y)\,dy = r^2\int_0^r \sqrt{r^2-y^2}\cos^{-1}\left(\frac{y}{r}\right)dy - \int_0^r y(r^2-y^2)\,dy \tag{21}$$

From Equations (18), (19), (20), and (21), we get

$$\Omega^{\oplus} = -\frac{r^4}{2} + \frac{2Lr^3}{3} + \int_0^r y(r^2 - y^2)\,dy = \frac{2Lr^3}{3} - \frac{r^4}{4}, \text{ so that}$$
$$p^{\oplus} = \frac{\Omega^{\oplus}}{L^4} = \frac{2}{3}\left(\frac{r}{L}\right)^3 - \frac{1}{4}\left(\frac{r}{L}\right)^4 \tag{22}$$

We compute the overlap area for the diagonally adjacent group as

$$H(x, y) = ||PRSTP|| - ||TCAST|| - ||TPQCT|| - ||CQRDQ||$$

$$\Omega^{\otimes} = \frac{\pi}{4}r^2 - \int_0^y \sqrt{r^2 - z^2}dz - \int_0^x \sqrt{r^2 - z^2}dz + xy$$
$$= \frac{\pi}{4}r^2 - \frac{1}{2}\left(y\sqrt{r^2 - y^2} + r^2 \sin^{-1}\left(\frac{y}{R}\right) + x\sqrt{r^2 - x^2} + r^2 \sin^{-1}\left(\frac{x}{R}\right) - 2xy\right),$$

where $0 \le x, y \le r$ and $x^2 + y^2 \le r^2$. Proceeding as before,

$$p^{\otimes} = \frac{2}{L^4}\int_0^x dx \int_0^{\sqrt{r^2-x^2}} \Omega^{\otimes}\,dy$$
$$= \frac{2}{L^4}\int_0^x dx \int_0^{\sqrt{r^2-x^2}} \Big(\frac{\pi}{4}r^2 - \frac{1}{2}\Big(y\sqrt{r^2 - y^2} + r^2 \sin^{-1}\left(\frac{y}{r}\right)$$
$$- x\sqrt{r^2 - x^2} - r^2 \sin^{-1}\left(\frac{x}{r}\right) + xy\Big)\Big)dy$$
$$= \frac{r^4}{8L^4}. \tag{23}$$

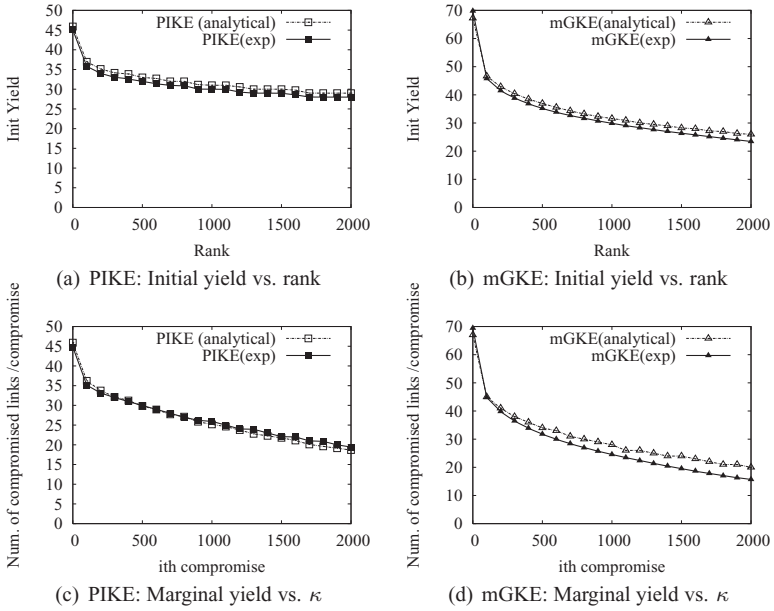
7.6.1 *Estimating the Yield.* Let $X^{\oplus}$ be the number of path keys mediated by s_i between some $s_u \in G_0$ and some s_v in a $\oplus$-adjacent group. Let $X^{\otimes}$ be the number of path keys mediated by s_i between some $s_u \in G_0$ and some s_v in a $\otimes$-adjacent group. Therefore,

$$Y_{\phi}(s_i) = \sum_{j=0}^{N^{\oplus}} X_j^{\oplus} + \sum_{j=0}^{N^{\otimes}} X_j^{\otimes}, \tag{24}$$

where $X_j^{\oplus}$ and $X_j^{\otimes}$ are random variables distributed as $X^{\oplus}$, and $X^{\otimes}$, respectively. Clearly, $X^{\oplus}$ and $X^{\otimes}$ are Binomially distributed. If groups G_1 and G_0 share t agent pairs, there will be $\gamma^2 - t$ path keys between them using one of the t agents in group G_0. Hence if $p^{\oplus}$ and $p^{\otimes}$ are as in Equations (22) and (23),

$$X^{\oplus} \sim B(\gamma^2 - t, p^{\oplus}/t), \text{ and } X^{\otimes} \sim B(\gamma^2 - t, p^{\otimes}/t).$$

Since the success probabilities are low, we can approximate these Binomial distributions as Poissons with parameters $\lambda^{\oplus} = (\gamma^2 - t)\frac{p^{\oplus}}{t}$, and $\lambda^{\otimes} = ((\gamma^2 - t))\frac{p^{\otimes}}{t}$, respectively. We get the CDF for $Y_{\phi}(s_i)$ as follows from Equation (24). We recall that that $N^{\oplus}$ and $N^{\otimes}$ are the number of groups $\oplus$-adjacent or $\otimes$-adjacent to the group containing s_i, and associated with s_i. Since the sum of Poissons is also



(a) PIKE: Initial yield vs. rank
(b) mGKE: Initial yield vs. rank
(c) PIKE: Marginal yield vs. κ
(d) mGKE: Marginal yield vs. κ

Fig. 18. Initial yields and marginal yields.

Poisson, we have

$$F(w) = \Pr[Y_{\emptyset}(s_i) \leq w] = \sum_{j=0}^{4} \sum_{k=0}^{4} \Pr[S_{j,k} \leq w] \cdot \Pr[N^{\oplus} = j] \cdot \Pr[N^{\otimes} = k], \quad (25)$$

where random variables $S_{j,k}$ are Poisson distributed with parameter $\lambda^{j,k} = j\lambda^{\oplus} + k\lambda^{\otimes}$. We compute $\Pr[N^{\oplus} = j]$ as follows. The analysis for $\Pr[N^{\otimes} = j]$ is identical.

Let s_i occur in group G_0. The other $g - 1$ groups are randomly placed into the cells in the region. Exactly four of these will be $\oplus$-adjacent to G_0; let us color these groups black. mGKE ensures that s_i is associated with exactly μ of the $g - 1$ peer groups. Hence, the number of black groups associated with s_i is simply the number of black items in a sample of size μ drawn without replacement from a population of $g - 1$ items, of which four are black. This is a hypergeometric distribution, so that

$$\Pr[N^{\oplus} = j] = \Pr[N^{\otimes} = j] = \binom{4}{j}\binom{g-5}{\mu-j} \Big/ \binom{g-1}{\mu}$$

We estimate $E[D_C(s_{\emptyset}^{\langle n_s - \kappa \rangle})]$ as for PIKE in Section 7.5, using Equation (17). The marginal yield $E[\widehat{Y}_C]$ with respect to κ is now readily computed from Equations (14), (25), (16), and (17).

7.7 Analytical and Experimental Results

Figures 18(a) and 18(b) show that the analytical and experimental values for initial yields in PIKE and mGKE match very well, showing the accuracy of our

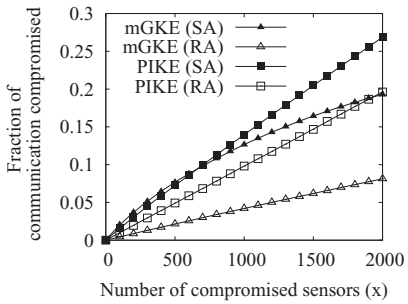


Fig. 19. PIKE & mGKE resilience.

analysis based on order statistics. Figures 18(c) and 18(d) show that the analytical and experimental values for the marginal yields under selective attack also match very well. Our analysis framework for selective attack clearly captures its true characteristics. As expected, the marginal yield decreases with κ for both PIKE and mGKE.

Figure 19 compares the resilience of PIKE and mGKE under random attack (RA) and selective attack (SA). As expected, SA is more effective than RA, since the attacker targets nodes to maximize new key information at each selective attack step. These results are consistent with the results for the resilience of RKP and SRKP under the two attack models in Huang et al. [2004]. Pairwise keys being unique (Section 6), PIKE's resilience decreases roughly linearly and mGKE's sublinearly, with the number of compromises. In contrast, RKP and SRKP resilience degrades dramatically after a threshold under both random and selective attack [Huang et al. 2004].

Figure 19 shows that under selective attack, PIKE has slightly better resilience than mGKE for small κ (less than 7%). However, as κ increases, mGKE outperforms PIKE. Resilience under selective attack depends on the fraction of path keys and the skew in the distribution of yields of uncompromised nodes. Greater skew lets the attacker discover more keys from the rest of network, by compromising the nodes with highest yields.

For uniform deployments, each node in PIKE is intermediary for approximately the same number of neighboring node pairs. In mGKE, agent nodes for neighboring groups tend have higher yields than other nodes, so mGKE starts off with higher skew for yields than PIKE (See Figure 18(a) and Figure 18(b)). PIKE's resilience is slightly better than mGKE's at the outset. However, as high-yield nodes are quickly compromised in selective attack, the effect of path-keys begins to dominate. Since mGKE has a much lower fraction of path keys, (see Section 8.2), it outperforms PIKE as more nodes are compromised.

8. MEMORY AND COMMUNICATION OVERHEAD

We now compare mGKE's and PIKE's memory and communication costs. Communication in sensor networks is mostly between neighboring nodes; mGKE's protocols recognize this fact and localize communications, but PIKE does not.

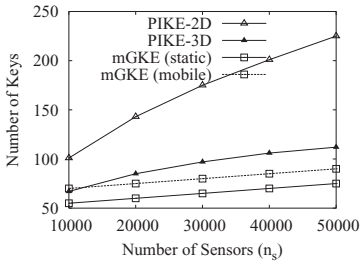


Fig. 20. Memory requirements.

This comparison is hence fair; that mGKE is a group-based scheme while PIKE is not material from this point of view.

We used 100 mobile collectors, and between 10,000 and 50,000 static sensors, with a default of 10,000. The communication range for each sensor was $40m$. The deployment density δ, the average number of static sensors in a sensor's transmission range, varied from 20 to 100, to represent low- to high-density deployments. The deployment area A is determined by the number of static sensors n_s, sensor density δ, and the communication range, with $A = \frac{n_s \pi r^2}{\delta}$. The group size γ was set to be 100, as in [Du et al. 2004; Huang et al. 2004], and the number of groups varied from 100 to 500 accordingly. Sensors in each group were uniformly distributed within a region of area A/g.

8.1 Memory Overhead

mGKE has low memory requirements. Given n_s static sensors, with group size γ, mGKE requires each sensor to be preloaded with $\gamma - 1$ pairwise keys shared with sensors from the same group and $t(g-1)/\gamma$ pairwise keys shared with sensors in different groups. Further, we use the method in Chan and Perrig [2005] to reduce the memory requirement by a factor of two. Therefore, the memory needed per sensor to establish S-S key is $\lceil \frac{1}{2}(\gamma - 1) \rceil + \lceil \frac{(n-\gamma)t}{2\gamma^2} \rceil$ keys. To establish M-S keys with n_m mobile collectors, each of which is t-associated with g' groups, each sensor must also be preloaded with an additional $\lceil \frac{g' t n_m}{2n_s} \rceil$ keys.

In contrast, PIKE-2D and PIKE-3D have memory overheads $\lceil \sqrt{n_s} \rceil + 1$ and $3\lceil \sqrt[3]{n_s} \rceil + 1$, respectively [Chan and Perrig 2005]. As noted in Section 5.4, PIKE requires global addressability, and cannot directly support mobility. Figure 20 shows the memory requirements of PIKE-2D, PIKE-3D and mGKE (t = 30). For mGKE, the solid line shows the memory overhead for supporting static sensors only, while the dashed line shows the memory needed to support mobile sensor networks with $g'/g = 0.3$.

8.2 Communication Overhead for S-S keys

mGKE requires messages only for path key establishment. If H is the average number of hops when a path key K_{ij} is established, the average communication overhead is $H \times \Pr[\mathbf{\Pi(K_{ij})}]$.

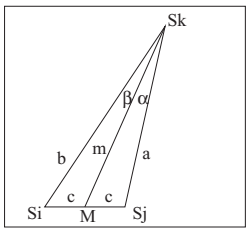


Fig. 21. PIKE.

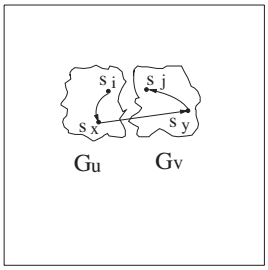


Fig. 22. mGKE.

Path key establishment between sensors s_i and s_j in PIKE (Figure 21) requires a message via intermediary s_k, requiring $h(s_i, s_k) + h(s_k, s_j)$ hops. To traverse physical distance $\bar{\lambda}$, a message needs at least $\bar{\lambda}/r$ hops, where r is the transmission radius, so $\bar{\lambda}/r$ is a lower bound for the average hop distance. In Figure 21, this distance is $a + b$. From trigonometry, $m = \frac{1}{2}\sqrt{a^2 + b^2 + 2ab\cos(\alpha + \beta)} \leq \frac{1}{2}(a + b)$, so we can use $2m$ as a lower bound for $h(s_i, s_k) + h(s_k, s_j)$. Two choices are available in PIKE for s_k, of which the one yielding the smaller distance is chosen. That is, we must choose the smaller of the distances from M to two intermediaries randomly placed in the square. This corresponds to the smaller of the distances from M to two randomly chosen points, or more precisely, to the first-order statistic for two randomly chosen values from the distribution of distances in the square. For samples drawn from distribution $F(x)$ with density $\frac{d}{dx}F(x) = f(x)$, the first-order statistic has distribution [Arnold et al. 1993]

$$f_{X_1}(x) = 2\left(f(x) - F(x)f(x)\right), \text{ with expectation } E[f_{X_1}] = \int_0^\infty x f_{X_1}(x)\, dx. \tag{26}$$

Using Equations (1) and (26), we obtain the expectation $E[m] = E[f_{X_1}] = 0.379475$, but omit the details for lack of space. A lower bound for the number of hops for PIKE's path key establishment in a square of area A is $\frac{2m\sqrt{A}}{r} = \frac{0.75895\sqrt{A}}{r}$.

Establishing a path key between s_i and s_j in mGKE (see Figure 22) requires messages from s_i to s_x, from s_x to s_y, and from s_y to s_j. If $h(s_p, s_q)$ denotes the hop distance between s_p and s_q, the number of hops required for path key establishment is $H(s_i, s_j) = h(s_i, s_x) + h(s_x, s_y) + h(s_y, s_j)$. If H_{mGKE} is the expected

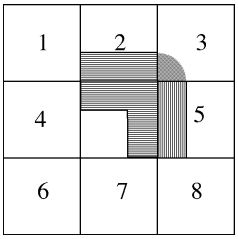


Fig. 23. Adjacencies.

number of hops for path key establishment in mGKE, linearity of expectation leads to

$$H_{mGKE} = 2 * \bar{h}_{mGKE} + \bar{h'}_{mGKE},$$

where $\bar{h}_{mGKE}$ is the expected hop distance between any two nodes in a group, and $\bar{h'}_{mGKE}$ is the expected hop distance between any two sensors from adjacent groups.

In mGKE, each group of sensors is in an $a \times a$ square, where $a = \sqrt{A/g}$. Since s_i and s_x are in the same square, the expected physical distance between them [Ghosh 1951] is $\bar{\lambda}_{mGKE} = 0.52a$. To get $\bar{\lambda'}_{mGKE}$, the expected distance between s_x and s_y, we note that two adjacent squares may touch (see Figure 23) along an edge or at a corner. Let $\bar{\lambda'}_{\oplus}$ be the expected distance between two random points picked randomly from neighboring squares that are vertically (or horizontally) disposed. Let $\bar{\lambda'}_{\otimes}$ be the expected distance between two random points picked from neighbors that are diagonally disposed. We now have $\bar{\lambda'}_{mGKE} = \Pr[\oplus]\bar{\lambda'}_{\oplus} + \Pr[\otimes]\bar{\lambda'}_{\otimes}$, where $\Pr[\oplus]$ (or $\Pr[\otimes]$) is the probability that neighboring squares are horizontally or diagonally disposed, respectively.

It is known [Ghosh 1951] that the expected distance between two random points in an $a \times 2a$ rectangle is $0.804a$. Since these two points are from the same square half with probability 0.5 and from different square halves with probability 0.5, we get $0.804a = 0.5\bar{\lambda'}_{\oplus} + 0.5\bar{\lambda}_{mGKE}$. That is, $\bar{\lambda'}_{\oplus} = 1.088a$.

To get $\bar{\lambda'}_{\otimes}$, consider two random points in a $2a \times 2a$ square, which consists of four $a \times a$ squares. Clearly, the expected distance between two random points in a $2a \times 2a$ square is $0.52 \times 2a = 1.04a$. Since these two points are from the same $a \times a$ square with probability 0.25, from two horizontally or vertically adjacent $a \times a$ squares with probability 0.5, and from two diagonally disposed $a \times a$ squares with probability 0.25, we can use linearity of expectation to write $1.04a = 0.25\bar{\lambda}_{mGKE} + 0.5\bar{\lambda'}_{\oplus} + 0.25\bar{\lambda'}_{\otimes}$. Since we know $\bar{\lambda}_{mGKE} = 0.52a$, $\bar{\lambda'}_{\oplus} = 1.088a$, we get $\bar{\lambda'}_{\otimes} = 1.464a$.

We estimate $\Pr[\oplus]$ and $\Pr[\otimes]$ as follows. s_i needs to communicate with a node s_j in a vertically adjacent square (see Figure 23) only if s_i is within distance r from the top edge of its cell. Similarly, s_i needs to communicate with a node in a diagonally disposed square only if s_i is within distance r from the corner, that is, inside the quarter circle shown. Hence, we have $\Pr[\oplus] = \frac{ar}{ar+0.25\pi r^2}$, and $\Pr[\otimes] = \frac{0.25\pi r^2}{ar+0.25\pi r^2}$.

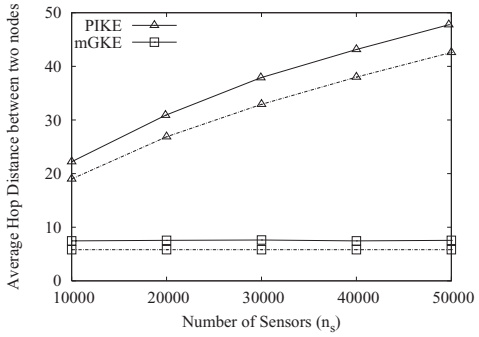


Fig. 24. $\bar{h}_{mGKE}$ and $\bar{h}_{PIKE}$.

Now, $\bar{\lambda'}_{mGKE} = \frac{ar}{ar+0.25\pi r^2}\bar{\lambda'}_{\oplus} + \frac{0.25\pi r^2}{ar+0.25\pi r^2}\bar{\lambda'}_{\otimes}$, yielding $\bar{\lambda'}_{mGKE} = \frac{4.35a^2+1.46\pi ra}{4a+\pi r}$. Consequently,

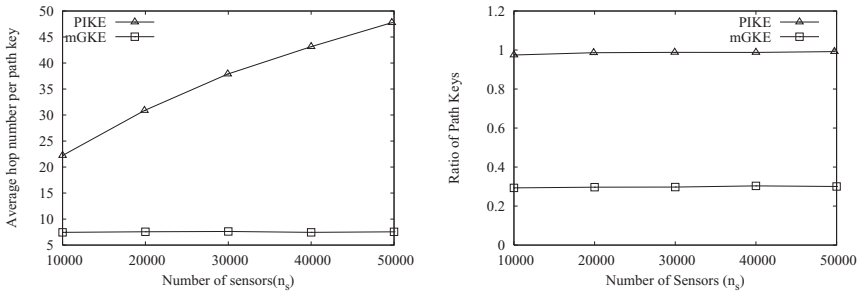
$$H_{mGKE} = 2 * \bar{h}_{mGKE} + \bar{h'}_{mGKE} \geq \frac{1.04\sqrt{A/g}}{r} + \frac{4.35A/g + 1.46\pi r\sqrt{A/g}}{(4\sqrt{A/g} + \pi r)r},$$

In Figure 24, the solid line shows the experimental results and the dashed line shows theoretical lower bound $\bar{h}$ for PIKE and mGKE, using a density $\delta = 50$. For both schemes, the experimental results match the lower bound quite closely. Therefore, we may use this lower bound to approximate $\bar{h}$.

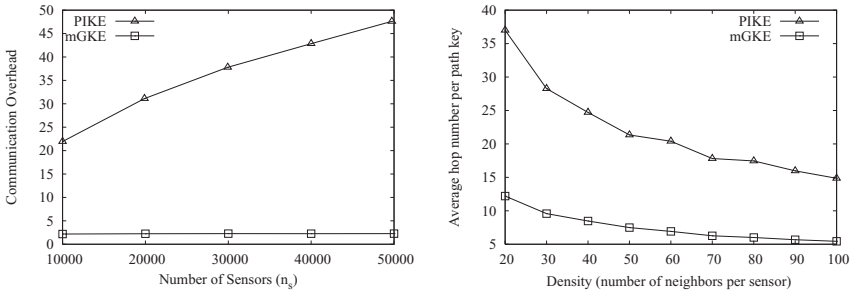
Figure 25(a) shows simulation results for the average number of hops H to establish path keys in PIKE and mGKE, for a density of 50. For fixed group size, H_{mGKE} remains constant as the network grows; network size has no impact on the communication overhead because the communication for establishing mGKE path keys is localized to two adjacent groups. In contrast, establishing a path key in PIKE requires network-wide communication, and thus H_{PIKE} increases as the network size increases.

8.2.1 *Path Key Fraction for S-S Keys.* Let (s_i, s_j) be a pair of neighbors picked globally in the system at random, and let $\mathbf{\Pi_G(K_{ij})}$ be the event that the key $K_{\mathbf{ij}}$ is a path key. The path key fraction is clearly the probability $\Pr[\mathbf{\Pi_G(K_{ij})}]$. Consider a PIKE grid of n nodes. For a given s_i, there are at most $2(\sqrt{n}-1)$ choices for s_j from on the same row or column, out of a total of $(n-1)$ choices system-wide, so that the probability of a shared key is at most $\frac{2}{\sqrt{n}+1}$. Hence, *neighbors in PIKE hardly ever share preloaded keys, so that* $\Pr[\mathbf{\Pi_G(K_{ij})}] \approx 1$. Our simulation results (Figure 25(b)) confirm this calculation.

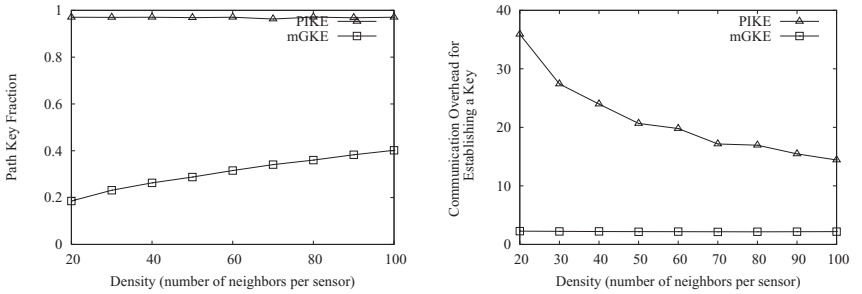
We estimate the global path key fraction in mGKE as follows. The expected number of neighboring sensor pairs is $\frac{n_s^2}{2}\frac{\pi r^2}{A}$. For any pair of groups which are vertically or horizontally adjacent, the probability that a pair of sensors from each group are neighbors is $p^{\oplus}$ (See Section 7.6), and such a pair will have a path key with probability $1 - \frac{\alpha^2}{t}$ (see Section 6.1). Hence, each pair of groups



(a) No. of hops vs. NW size ($\delta = 50$) (b) Path keys vs. NW size ($\delta = 50$)



(c) Communic. vs. NW size ($\delta = 50$) (d) No. of hops vs. density (n =10,000)



(e) Path keys vs. density (n =10,000) (f) Communic. vs. density (n =10,000)

Fig. 25. Average number of hops, path key fraction and communication overhead.

will have on average $\gamma^2 p^{\oplus}(1 - \frac{\alpha^2}{t})$ path keys. Since groups are deployed in a $\sqrt{g} \times \sqrt{g}$ grid, the total number of group pairs vertically or horizontally adjacent is $2\sqrt{g}(\sqrt{g} - 1)$. These adjacent group pairs will have on average $\gamma^2 p^{\oplus}(1 - \frac{\alpha^2}{t})2\sqrt{g}(\sqrt{g} - 1)$ path keys. Similarly, we can estimate the average number of path keys between diagonally adjacent groups as $\gamma^2 p^{\otimes}(1 - \frac{\alpha^2}{t})2(\sqrt{g} - 1)^2$. Hence,

$$\Pr[\mathbf{\Pi_G(K_{ij})}] = \frac{4A\gamma^2(1 - \frac{\alpha^2}{t})(\sqrt{g} - 1)\left(p^{\oplus}\sqrt{g} + p^{\otimes}(\sqrt{g} - 1)\right)}{n_s^2 \pi r^2}.$$

This equation yields a path-key fraction of 28% for the default mGKE configuration, which is in perfect agreement with the experimental ratio shown in Figure 25(b).

8.2.2 *Communication Overhead.* Figure 25(c) shows $H \times \Pr[\mathbf{\Pi_G(K_{ij})}]$, an estimate of the path key establishment overhead. mGKE's overhead is 10 times lower than PIKE's for 10,000 nodes, and over 20 times lower for 50,000 nodes. Clearly, mGKE is especially suitable for very large sensor networks.

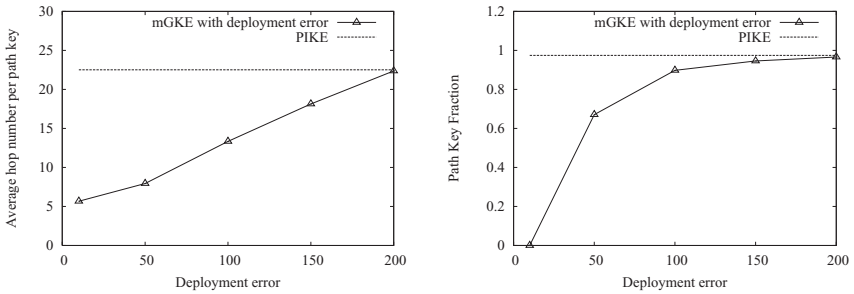
Two major differences between PIKE and mGKE result in a big difference in their communication overheads. First, sensors use local intermediaries when establishing path keys in mGKE, so only local communication is needed to transmit key establishment messages. In contrast, intermediaries in PIKE could be *anywhere* in the entire target region, so that network-wide communication is required. Second, a larger fraction of keys are path keys in PIKE than in mGKE. When sensors are deployed in groups, sensors from the same group are more likely to be neighbors. In mGKE, all sensors from the same group share preloaded pairwise keys, but in PIKE, only sensors on the same grid column or row do. Consequently, the fraction of path keys in PIKE is significantly higher than that of mGKE.

PIKE distributes sensors uniformly in the region, while mGKE uses a flexible group deployment method, and exploits the locality inherent in group-based deployments to reduce the number of path keys. Since the PIKE work [Chan and Perrig 2005] does not discuss group-based deployment, and its logical grid is clearly intended to be system-wide, there is no sound basis for discussing its performance in group-based deployments.

8.2.3 *Low Density Deployments.* Figure 25(d) shows the average number of hops for establishing path keys for a network of 10,000 nodes deployed at low densities. As expected, the average number of hops decreases in both PIKE and mGKE with density, but mGKE incurs much lower communication overhead than PIKE. Figure 25(e) shows the ratio of path keys in PIKE and mGKE, and Figure 25(f) plots the communication overhead in PIKE and mGKE, varying the network density from 20 to 100. Clearly, mGKE has much lower communication overheads than PIKE even when the network density is low.

8.2.4 *Effects of Deployment Error.* In practice, sensor deployment is subject to errors. We will now examine how deployment errors affect mGKE's communication costs, assuming, as in Liu et al. [2008], that sensors in a group follow a 2-D Gaussian distribution with standard deviation σ, with mean location being the center of grid cell.

Figures 26(a) and 26(a) present the average number of hops per path key establishment, and the path key fraction as σ increases from 10m to 200m. We assume 100m$\times$100m grid cells and a 1000m$\times$1000m region. As we increase σ, the sensors in a group fall into a wider area, reducing the value of group-based deployment. At a σ of 200m, sensors in a group are likely to scatter over the entire 1000m$\times$1000m region, nullifying the benefits of group-based deployments. At this extreme value for σ, mGKE has communication cost similar to that of PIKE. For more moderate values of σ, mGKE outperforms PIKE.



(a) No. of hops vs. deployment error (b) Path keys vs. deployment error

Fig. 26. Communication cost under deployment error(n =10,000, $\delta = 50$).

8.3 Communication Overhead for M-S Keys

Establishing a key between $\langle G_u, s_i \rangle$ and m_j requires two intragroup messages if G_u is associated with m_j, or two intergroup messages and two intragroup messages otherwise. Let $\frac{g'}{g} = 0.3$. Then, m_j is associated with G_u with probability $\Pr[\mathbf{\Pi_a}(\mathbf{K_{ij}})] = \frac{g'}{g} = 0.3$. Otherwise m_j is associated with at least one of G_u's eight adjacent groups with probability $P_2 = 1-\Pr[\mathbf{\Pi_a}(\mathbf{K_{ij}})]-(1-\Pr[\mathbf{\Pi_a}(\mathbf{K_{ij}})])^9 = 0.66$. We consider only these two cases, since they occur with probability close to 1.

In the first case, the two intragroup messages require on average $\bar{h}_1 = 2 \times 0.52a/r = 1.04a/r$ hops, for $a \times a$ cells. For the second case, the two intergroup messages require $\bar{h}_2 = 2 \times \frac{1}{2}(\bar{\lambda'}_{\oplus} + \bar{\lambda'}_{\otimes})/r = 2.57a/r$. Now, the average number of hops to establish an M-S key is $\Pr[\mathbf{\Pi_a}(\mathbf{K_{ij}})] \times \bar{h}_1 + P_2 \times \bar{h}_2$.

To demonstrate that our scheme supports mobile sensor networks, we evaluate the fraction of total available energy in the sensor networks consumed to establish keys between mobile collectors and static sensors. Let the region be a $1,000m \times 1,000m$ square, divided into 100 subregion of size $100m \times 100m$. Let the network have $n_m = 100$ mobile collectors, and $n_s = 10,000$ static sensors, divided into $g = 100$ groups with size $\gamma = 100$. Mobile collectors move at constant speed $v = 10m/s$, and pause $w = 5s$ at waypoints. Let $\frac{g'}{g} = 0.3$, and the transmission radius is $r = 40m$.

From our analysis, we know the average number of hops per M-S key is about 7. When a mobile collector moves to a subregion, it will establish keys with all the 100 static sensors in the subregion. As analyzed in Section 4, the average time for each data collection is $E[t_i] = 57s$. On average, the 100 mobile collectors will establish $100 \times 100 = 10^4$ keys in all, requiring 7×10^4 transmissions, every $57s$.

If the energy to transmit a packet per hop is approximately $0.48mJ$ [Madden et al. 2003], the energy consumption for key establishment will be $7 \times 10^4 \times 0.48 = 3.36 \times 10^4 mJ$. If each sensor has two AA batteries, each with average capacity 2,850 mAh [Malan et al. 2004], the total energy capacity of 10,000 sensors would be $10,000 \times 2,850 \times 2 = 5.7 \times 10^7$ mAh. This capacity will keep the sensor network alive for about $57s \times (3 \times 5.7 \times 10^7 \times 3600)/(3.36 \times 10^4) = 12087$ days when the energy is only used for M-S key establishment.

9. CONCLUSIONS

In this work, we address the problem of ensuring data integrity at the source and during transit in sensor networks, and the related challenge of analyzing the resilience of sensor networks to selective and random attacks. For source integrity, we present an analysis of the impact of mobile collector compromises, and the circumstances under which trustworthiness can be guaranteed. Our analysis forms a sound basis for quantifying the tradeoffs to be made to ensure security of deployments with mobile collectors.

We also present mGKE, a new group-based key predistribution scheme for large sensor networks, with a number of advantages over current methods. First, it accommodates very flexible deployment models as well as mobility. Second, it enables any pair of sensors to establish a unique pairwise key in any physically connected network, regardless of sensor density or distribution. Third, mGKE is nearly perfectly resilient against node capture attacks, due to the uniqueness of pairwise keys. Unlike competing group-based methods, system security in mGKE does not degrade dramatically when the number of compromised sensors reaches a certain threshold. mGKE is remarkably resilient to compromises. Finally, mGKE uses only local communication to establish path keys, and has very low overhead.

A major contribution of our article is a novel framework for analyzing the impact of selective sensor compromises in sensor networks, using the theory of order statistics. Selective attack, in particular, poses difficult technical challenges, and our work is the first such analysis to appear in the literature. We have applied this framework to perform detailed analysis of the effects of selective attack on PIKE and mGKE, and compared the results of analysis with those of experiment. Our analytical and experimental results match extremely well, confirming the correctness of our analysis.

REFERENCES

ARNOLD, B., BALAKRISHNAN, N., AND NAGARAJA, H. 1993. *A First Course in Order Statistics*. Wiley-Interscience Publication.

BALFANZ, D., SMETTERS, D., STEWART, P., AND WONG, H. 2002. Talking to strangers: Authentication in ad hoc wireless networks. In *Proceedings of the 9th Annual Network and Distributed System Security Symposium (NDSS)*.

BERGBREITER, S. AND PISTER, K. S. J. 2003. Cotsbots: An off-the-shelf platform for distributed robotics. In *Proceedings of the International Conference on Intelligent Robots and Systems*.

BETTSTETTER, C., HARTENSTEIN, H., AND COSTA, X. P. 2004. Stochastic properties of the random waypoint mobility model. *Wirel. Netw. 10*, 5, 555–567.

CAPKUN, S., HUBAUX, J.-P., AND BUTTY, L. 2003. Mobility helps security in ad hoc networks. In *Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing*. ACM Press, 46–56.

CHAN, H. AND PERRIG, A. 2005. PIKE: Peer intermediaries for key establishment in sensor networks. In *Proceedings of the IEEE Conference on Computer Communications (InFOCOM'05)*.

CHAN, H., PERRIG, A., AND SONG, D. 2003. Random key predistribution schemes for sensor networks. In *Proceedings of the IEEE Symposium on Security and Privacy*.

CHEN, X. AND DRISSI, J. 2005. An efficient key management scheme in hierarchical sensor networks. In *Proceedings of the IEEE Mobile Adhoc and Sensor Systems Conference*.

CROSSBOW. MICA2: Wireless Measurement System.
`http://www.xbow.com/Product/pdf/files/wireless_pdf/6020-0042-0_MICA2.pdf`.

DU, W., DENG, J., HAN, Y., CHEN, S., AND VARSHNEY, P. 2004. A key management scheme for wireless sensor networks using deployment knowledge. In *Proceedings of the IEEE Conference on Computer Communications (InFOCOM'04)*.

DU, W., DENG, J., HAN, Y., AND VARSHNEY, P. 2003. A pairwise key predistribution scheme for wireless sensor networks. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS'03)*.

DUBEY, S. D. 1963. A theorem on a ratio of random variables. *Oper. Res. 13*, 3, 476–477.

ERDÖS AND RENYI. 1959. On random graphs I. *Publicationes Mathematicae 6*, 290–291.

ESCHENAER, L. AND GLIGOR, V. D. 2002. A key-management scheme for distributed sensor networks. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS'02)*.

GHOSH, B. 1951. Random distances within a rectangle, and between two rectangles. *Bull. Calcutta Math. Soc. 43*, 17–24.

GOLDREICH, O., GOLDWASSER, S., AND MICALI, S. 1986. How to construct random functions. *J. ACM 33*, 4, 792–807.

HUANG, D., MEHTA, M., MEDHI, D., AND HARN, L. 2004. Location-aware key manageent scheme for wireless sensor networks. In *Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN'04)*.

HWANG, J. AND KIM, Y. 2004. Revisiting random key pre-distribution schemes for wireless sensor networks. In *Proceedings of the ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN'04)*.

KANSAL, A., SOMASUNDARA, A. A., JEA, D. D., SRIVASTAVA, M. B., AND ESTRIN, D. 2004. Intelligent fluid infrastructure for embedded networks. In *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services (MobiSys'04)*. ACM Press, 111–124.

KARP, B. AND KUNG, H. T. 2000. GPSR: greedy perimeter stateless routing for wireless networks. In *Proceedings of the Annual ACM Conference on Mobile Computing and Networking (MobiCom'00)*. 243–254.

KONG, J., ZERFOS, P., LUO, H., LU, S., AND ZHANG, L. 2001. Providing robust and ubiquitous security support for mobile ad hoc networks. In *Proceedings of the 9th International Conference on Network Protocols (ICNP'01)*. IEEE Computer Society, 251.

LIU, D. AND NING, P. 2003a. Establishing pairwise keys in distributed sensor networks. In *Proceedings of the ACM Conference on Computer and Communications Security (CCS'03)*.

LIU, D. AND NING, P. 2003b. Location-based pairwise key establishments of static sensor networks. In *Proceedings of the ACM Workshop in Security in Ad Hoc and Sensor Networks (SASN'03)*.

LIU, D., NING, P., AND DU, W. 2005. Group-based key pre-distribution in wireless sensor networks. In *Proceedings of the 4th ACM Workshop on Wireless Security (WiSe'05)*. 11–20.

LIU, D., NING, P., AND DU, W. 2008. Group-based key predistribution for wireless sensor networks. *ACM Trans. Sens. Netw. 4*, 2.

MADANSKY, A. 1959. Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Stat. 30*, 743–746.

MADDEN, S., FRANKLIN, M. J., HELLERSTEIN, J. M., AND HONG, W. 2003. The design of an acquisitional query processor for sensor networks. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD'03)*. 491–502.

MALAN, D., WELSH, M., AND SMITH, M. 2004. A public-key infrastructure for key distribution in tinyos based on elliptic curve cryptography. In *Proceedings of the 1st IEEE International Conference on Sensor and Ad Hoc Communications and Networks*.

QINGGUANG, Z., YANLING, C., AND JUAN, L. 2006. A lightweight key management protocol for hierarchical sensor networks. In *Proceedings of the International Conference on Parallel and Distributed Computing, Applications and Technologies*.

RATNASAMY, S., KARP, B., YIN, L., ESTRIN, D., GOVINDAN, R., AND SHENKER, S. 2002. GHT: A geographic hash table for data-centric storage. In *Proceedings of the ACM Workshop on Wireless Sensor Networks and Applications (WSNA'02)*.

SIBLEY, G. T., RAHIMI, M. H., AND SUKHATME, G. S. 2002. Robomote: A tiny mobile robot platform for large-scale sensor networks. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02)*.

THALER, D. AND RAVISHANKAR, C. V. 1998. Using name-based mappings to increase hit rates. *IEEE/ACM Trans. Netw. 6*, 1, 1–14.

TIRTA, Y., LI, Z., LU, Y., AND BAGCHI, S. 2002. Efficient collection of sensor data in remote fields using mobile collectors. In *Proceedings of the 13th International Conference on Computer Communications and Networks (ICCCN'04)*.

YANG, H., YE, F., YUAN, Y., LU, S., AND ARBAUGH, W. 2005. Toward resilient security in wireless sensor networks. In *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05)*. ACM Press, 34–45.

YE, F., LUO, H., CHENG, J., LU, S., AND ZHANG, L. 2002. A two-tier data dissemination model for large-scale wireless sensor networks. In *Proceedings of the 8th Annual International Conference on Mobile Computing and Networking (MobiCom'02)*. ACM Press, 148–159.

YE, F., LUO, H., LU, S., AND ZHANG, L. 2004. Statistical en-route detection and filtering of injected false data in sensor networks. In *Proceedings of the IEEE Conference on Computer Communications (InFOCOM'04)*.

ZHANG, W., CAO, G., AND PORTA, T. L. 2003. Data dissemination with ring-based index for wireless sensor networks. In *Proceedings of the IEEE International Conference on Network Protocols (ICNP'03)*.

ZHANG, W., SONG, H., ZHU, S., AND CAO, G. 2005. Least privilege and privilege deprivation: Towards tolerating mobile sink compromises in wireless sensor networks. In *Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05)*. ACM Press, 378–389.

ZHOU, L., NI, J., AND RAVISHANKAR, C. 2005a. (short paper) gke: Efficient group-based key establishment for large sensor networks. In *Proceedings of the 1st IEEE/CreateNet International Conference on Security and Privacy for Emerging Areas in Communication Networks (SecureComm'05)*.

ZHOU, L., NI, J., AND RAVISHANKAR, C. V. 2005b. Efficient key establishment for group-based wireless sensor deployments. In *Proceedings of the 4th ACM Workshop on Wireless Security (WiSe'05)*. 1–10.

ZHOU, L., NI, J., AND RAVISHANKAR, C. V. 2006. Supporting secure communication and data collection in mobile sensor networks. In *Proceedings of the IEEE Conference on Computer Communications (InFOCOM'06)*.

ZHOU, L. AND RAVISHANKAR, C. 2005. A fault localized scheme for false report filtering in sensor networks. In *Proceedings of the IEEE International Conference on Pervasive Services*.

ZHU, S., SETIA, S., JAJODIA, S., AND NING, P. 2004. An interleaved hop-by-hop authentication scheme for filtering false data injection in sensor networks. In *Proceedings of the IEEE Symposium on Security and Privacy*.

Received August 2008; revised April 2009; accepted July 2009