Filtering Injected False Data in Wireless Sensor Networks by Bandwidth Efficient Co-operative Authentication Scheme

Priyanka C, Sunaina P, SupriyaV, Vatsala N Gowda
City Engineering College,Bangalore
priyankagp.c@gmail.com,gracefulsunaina@gmail.com,supri.naidu@gmail.com
n.gowdavatsala@gmail.com

Abstract

Injecting false data attack is a well known serious threat to wireless sensor network, for which an adversary reports bogus information to sink causing error decision at upper level and energy waste in en-route nodes. In this paper, we propose a novel bandwidth efficient cooperative authentication scheme for filtering injected false data. Based on the random graph characteristics of sensor node deployment and the cooperative bit-compressed authentication technique, the proposed scheme can save energy by early detecting and filtering the majority of injected false data with minor extra overheads at the en-route nodes. In addition, only a very small fraction of injected false data needs to be checked by the sink, which thus largely reduces the burden of the sink. Both theoretical and simulation results are given to demonstrate the effectiveness of the proposed scheme in terms of high filtering probability and energy saving.

1. Introduction

Due to fast booming of microelectro mechanical systems, wireless sensor networking has been subject to extensive research efforts in recent years. It has been well recognized as a ubiquitous and general approach for some emerging applications, such as environmental and habitat monitoring, surveillance and tracking for military [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. A wireless sensor network is usually composed of a large number of sensor nodes which are interconnected through wireless links to perform distributed sensing tasks. Each sensor node is low-cost but equipped with necessary sensing, data processing, and communicating components. There-fore, when a sensor node generates a report after being triggered by a special event, e.g., a surrounding temperature change, it will send the report to a data collection unit (also known as sink) through an established routing path [17].

Wireless sensor networks are usually deployed at unattended or hostile environments. Therefore, they are very vulnerable to various security attacks, such as selective forwarding, wormholes, and Sybil attacks [12], [18]. In addition, wireless sensor networks may also suffer from injecting false data attack [10]. For an injecting false data attack, an adversary first compromises several sensor nodes, accesses all keying materials stored in the compromised nodes, and then controls these compromised nodes to inject bogus information and send the false data to the sink to cause upper-level error decision, as well as energy wasted in en-route nodes.

For instance, an adversary could fabricate a wildfire event or report wrong wildfire location information to the sink, and then expensive resources will be wasted by sending rescue workers to a non-existing or wrong wildfire location.

Therefore, it is crucial to filter the false data as accurately as possible in wireless sensor networks. At the same time, if all false data are flooding into the sink simultaneously, then not only huge energy will be wasted in the en-route nodes, but also heavy verification burdens will undoubtedly fall on the sink. As a result, the whole network could be paralyzed quickly. Therefore, filtering false data should also be executed as early as possible to mitigate the energy waste. To tackle this challenging issue, some false data filtering mechanisms have been developed [7], [8], [9], [10], [11], [12], [13]. Since most of these filtering mechanisms use the symmetric key technique, once a node is compromised, it is hard to identify the node. In other words, the compromised node can abuse its keys to generate false reports, and the reliability of the filtering mechanisms will be degraded.

In this paper, we propose a novel cooperative authentication scheme for filtering injected false data. Compared with the previously reported mechanisms, the scheme achieves not only high filtering

probability but also high reliability. The main contributions of this paper are threefold.

- First, we study the random graph characteristics of wireless sensor node deployment, and estimate the probability of k-neighbors, which provides the necessary condition for this authentication scheme.
- Second, we propose the scheme to filter the injected false data with cooperative bit-compressed authentication technique. With the proposed mechanism, injected false data can be early detected and filtered by the en-route sensor nodes. In addition, the accompanied authentication information is bandwidth-efficient.
- Third, we develop a custom Java simulator to demonstrate the effectiveness of the proposed scheme in terms of en-routing filtering probability and false negative rate on true reports.

The remainder of this paper is organized as follows: In Section 2, we introduce the system model and design goal. In Section 3, we review some preliminaries including TinyECC-based non-interactive keypair establishment and message authentication code in Z_2^n. Then, we present the this scheme in Section 4, followed by the security analysis and performance evaluation in Section 5 and Section 6, respectively. We review some related works in Section 7. In the end, we draw our conclusions in Section 8.

2. Model and Design Goal

In this section, we formulate the network model, the security model, and identify the design goal.

2.1 Network Model

We consider a typical wireless sensor network which consists of a sink and a large number of sensor nodes $N=\{N_0,N_1,\ldots\}$ randomly deployed at a certain interest region(CIR) with the area S. The sink is a trustable and powerful data collection device, which has sufficient computation and storage capabilities and is responsible for initializing the sensor nodes and collecting the data sensed by these nodes. Each sensor node Ni is stationary in a location. For differentiation purpose, we assume each sensor node has a unique nonzero identifier. The communication is bidirectional, i.e., two sensor nodes within their wireless transmission range (R) may communicate with each other. Therefore, if a sensor node is close to the sink, it can directly contact the sink. However, if a sensor node is far from the transmission range of the sink, it should resort to other nodes to establish a route and then communicate with the sink. Formally, such a

Wireless sensor network, as shown in Fig. 1, can be represented as an undirected graph G=(V,E) where $V=\{V_1,V_2\ldots\}$is the set of all sensors $N=\{N_0,N_1\ldots\}$ plus the sink, and $E=\{(v_i,v_j) \mid v_i,v_j \in V$ is the set of edges. Let d(vi, vj) denote as the distance between v_i and v_j, then each e_{ij}, which indicates whether there exists a communication edge between two nodes vi and v_j or not, is defined as,

$$e_{ij} = \begin{cases} 1, & d(v_i, v_j) \le R; \\ 0, & d(v_i, v_j) > R. \end{cases} \quad (1)$$

Let v1 denote the sink. All sensor nodes $V/\{v_1\}=\{v_2,v_3,\ldots\}$ can run the Dijkstra shortest path algorithm find their shortest paths to the sink v_1, only if the graph G is fully connected.

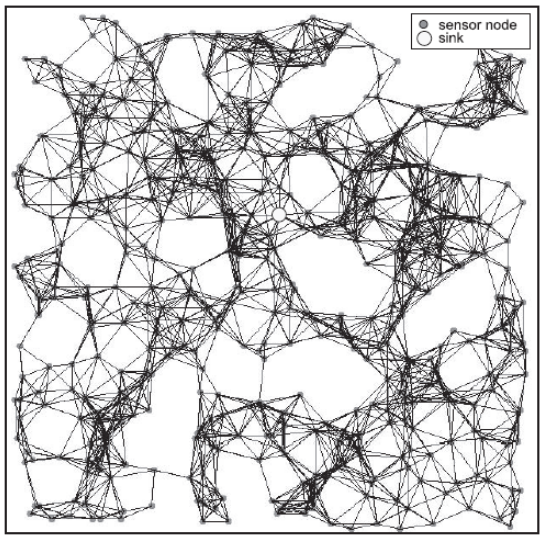


Fig.1.Wireless sensor network under consideration.

Probability of fully connected G. Assume that the positions of these vertexes V are uniformly distributed in the area S with network density λ, where $\lambda = \frac{|V|}{S}$ and |V| denotes the cardinality of V. Based on the random graph theory, the probability that there are n nodes in an arbitrary region A with the area A is,

$$P(N=n|\mathcal{A}) = \binom{|\mathcal{V}|}{n}\left(\frac{\lambda\cdot\mathcal{A}}{|\mathcal{V}|}\right)^{n}\cdot\left(1-\frac{\lambda\cdot\mathcal{A}}{|\mathcal{V}|}\right)^{|\mathcal{V}|-n} = \binom{|\mathcal{V}|}{n}\left(\frac{\mathcal{A}}{\mathcal{S}}\right)^{n}\cdot\left(1-\frac{\mathcal{A}}{\mathcal{S}}\right)^{|\mathcal{V}|-n}. \quad (2)$$

To calculate the full connection probability P_{con}, we first compute P_{iso}, the isolation probability of any node in G, where a node is called isolated if there exists no link between it and any other nodes. In other words, in some circle coverage with the area πR^2, except one node lies at the center, no other node exists. If the border effects are neglected, we have,

$$P_{iso} = P(N=0|\pi R^2) = \binom{|\mathcal{V}|-1}{0}\cdot\left(\frac{\pi R^2}{\mathcal{S}}\right)^{0}\cdot\left(1-\frac{\pi R^2}{\mathcal{S}}\right)^{|\mathcal{V}|-1} = \left(1-\frac{\pi R^2}{\mathcal{S}}\right)^{|\mathcal{V}|-1}. \quad (3)$$

Based on the isolation probability P_{iso}, we can compute the full connection probability P_{con} [20] as

$$P_{con} \geq (1-P_{iso})^{|\mathcal{V}|} = \left(1-\left(1-\frac{\pi R^2}{\mathcal{S}}\right)^{|\mathcal{V}|-1}\right)^{|\mathcal{V}|}. \quad (4)$$

Fig. 2 shows the full connection probability P_{iso} versus different transmission range R and |V|. It can be seen that the expected fully connected G can be achieved by choosing proper R and |V|.

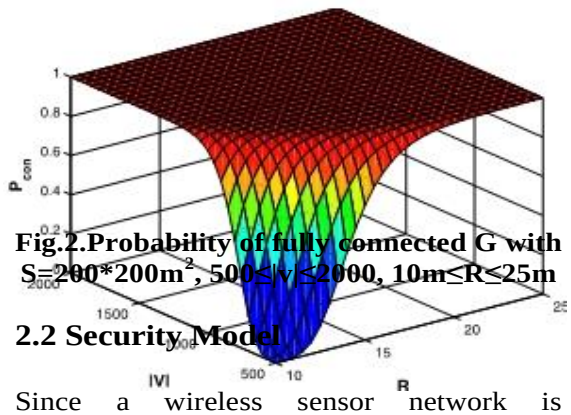


Fig.2.Probability of fully connected G with $S=200*200m^2$, $500\leq|v|\leq2000$, $10m\leq R\leq25m$

2.2 Security Model

Since a wireless sensor network is unattended, a malicious adversary may readily launch some security attacks to degrade the network functionalities. In addition, due to the low-cost constraints, sensor nodes are not equipped with expensive tamper-proof device and could be easily compromised in such an unprotected wireless sensor network. Therefore, in our security model, we assume an adversary A can compromise a fraction of sensor nodes and obtain their stored keying materials. Then, after being controlled and reprogrammed by the adversary A, these compromised sensor nodes can collude to launch some injected false data attacks.

Since our work focuses on filtering injected false data attack, other attacks launched by the compromised sensor nodes in wireless sensor network, such as building bogus routing information, selectively dropping true data packet, and creating routing loops to waste the energy of network [18], are not addressed in this paper.

2.3 Design Goal

The design goal is to develop an efficient cooperative bandwidth efficient authentication scheme for filtering the injected false data. Specifically, the following two desirable objectives will be achieved.

2.3.1 Early Detecting the Injected False Data by the En-Route Sensor Nodes

The sink is a powerful data collection device. Nevertheless, if all authentication tasks are fulfilled at the sink, it is undoubted that the sink becomes a bottleneck. At the same time, if too much injected false data floods into the sink, the sink will surely suffer from the Denial of Service (DoS) attack. Therefore, it is critical to share the authentication tasks with the en-route sensor nodes such that the injected false data can be detected and discarded early. The earlier the injected false data is detected, the more energy can be saved in the whole network.

2.3.2 Achieving Bandwidth-Efficient Authentication

Since sensors are low cost and energy constraint, it is desirable to design bandwidth efficient authentication scheme.

3. Preliminaries

3.1 TinyECC-Based Non-interactive Keypair Establishment

TinyECC is a configurable library for Elliptic Curve Cryptography (ECC), which allows flexible integration of ECC-based public key cryptography in sensor network applications. A substantially experimental evaluation using representative sensor platforms, such as MICAz [21] and Imote2 [22], is performed, and the results show that the ready-to-use TinyECC is suitable for wireless sensor networks to provide convenient authentications and pair key establishments [19]. Let p be a large prime and E(Fp) represent an elliptic curve defined over Fp. Let G € E (Fp) be a base point of prime order q. Then, each sensor node Ni € N can preload a TinyECC based public-private key pair (Yi, xi), where the private key xi is randomly chosen from Zq and the public key Yi=xiG.

Noninteractive keypair establishment. For any two sensor nodes vi, vj€G= (V, E), no matter what eij€ (0,1) is, sensor nodes vi with the key pair (Yi, xi) and vj with the key pair (Yj, xj) can establish a secure Elliptic Curve Diffie-Hellman (ECDH) keypair without direct contacting [23], where

$$k_{ij} = x_i Y_j = x_i x_j G = x_j x_i G = x_j Y_i = k_{ji}. \quad (5)$$

Because of the hardness of Elliptic Curve Discrete Logarithm (ECDL) problem, only vi and vj can secretly share a key. At the same time, the established keys are independent. In other words, if a sensor node vi is compromised, then the key kij shared between vi and vj will be disclosed. However, the key kjj[1] shared between vj and another sensor node vj[1] is not affected. For unattended wireless sensor networks, the property of key independence is useful, since it can limit the scope of key disclosure to the adversary A.

3.2 Message Authentication Code in Z_2^n

Message authentication code (MAC) provides assurance to the recipient of the message which came from the expected sender and has not been altered in transit [24]. Let h(.) be a secure cryptographic hash function [25]. A MAC in Z_2^n can be considered as a keyed hash, and defined as

$$MAC(m, k, n) = h(m \| k) \bmod 2^n, \quad (6)$$

where m, k, n are a message, a key, and an adjustable parameter, respectively. When n=1, MAC(m, k, 1) provides one-bit authentication, which can filter a false message with the probability 1/2; while n =α, MAC(m, k, α) can filter a false message with a higher probability $1-1/2^{\alpha}$.

4. Proposed Scheme

In this section, we will propose this scheme for filtering injected false data in wireless sensor networks. Before proceeding this scheme, the design rationale is introduced.

4.1 Design Rationale

To filter the false data injected by compromised sensor nodes, this scheme adopts cooperative neighbor*router (CNR)-based filtering mechanism. As shown in Fig. 3, in the CNR-based mechanism,

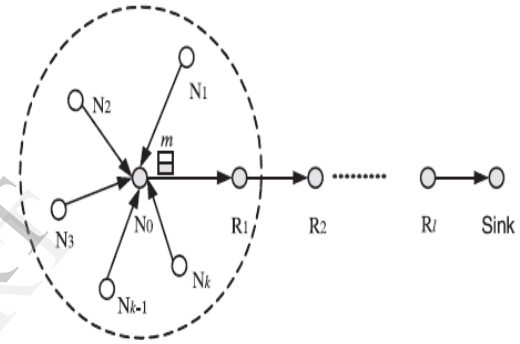


Fig. 3. Cooperative CNR-based authentication mechanism.

when a source node N_0 is ready to send a report m to the sink via an established routing pathR_{N0}: [R1→R2→… Rl→Sink], it first resorts to its k neighboring nodes N_{N0}: {N1, N2, Nk} to cooperatively authenticate the report m, and then sends the report m and the authentication information MAC from N_0 U N_{N0} to the sink via routing R_{N0}, where,

$$\mathbf{MAC} = \begin{pmatrix} mac_{01} & \cdots & mac_{0l} & mac_{0s} \\ mac_{11} & \cdots & mac_{1l} & mac_{1s} \\ mac_{21} & \cdots & mac_{2l} & mac_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ mac_{k1} & \cdots & mac_{kl} & mac_{ks} \end{pmatrix}, \quad (7)$$

each macij, 0 ≤ i ≤ k, 1 ≤ j ≤ l, represents Ni's MAC on m for Rj's authentication, and each macis represents Ni's MAC on m for the sink's authentication. As indicated in network model, the sink initializes all sensor nodes, and then each sensor node shares its private key with the sink. At the same time, according to the TinyECC-based non-interactive keypair establishment [19], the

full bipartite key graph between N_0 U N_N and R_N can be established, as shown in Fig. 4. Because of the existence of full bipartite key graph, the MAC design is reasonable. Therefore, when a compromised sensor node sends a false data to the sink, the false data can be filtered if there is at least one uncompromised neighboring node participating in the reporting. To achieve the bandwidth-efficient authentication, each mac_{ij} is set as one bit and each mac_{is} is bits by using the above MAC in Z_2^n technique. Then, the scale of MAC is only (l+α) *(k + 1) bits.

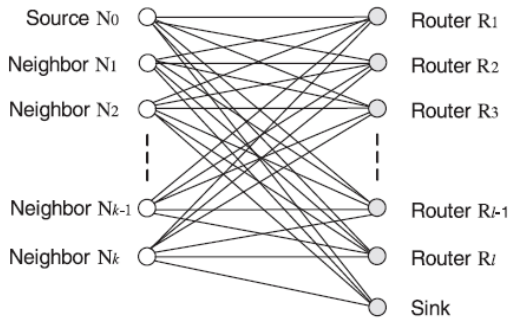


Fig. 4. Bipartite graph representing the relationships between the (source + neighbors) and (sink + routers).

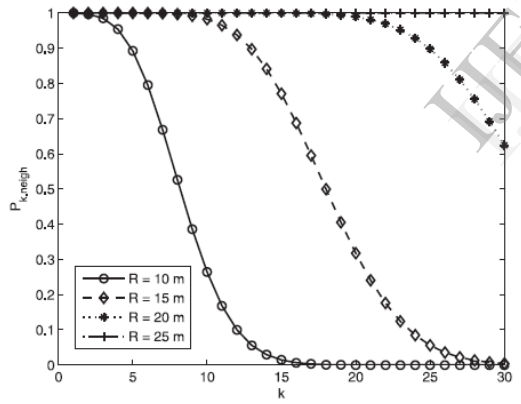


Fig. 5. Probability of k neighbors $P_{k.neigh}$ with S = 200*200 m^2, |V | = 1,000, R = 10, 15, 20, 25 m, and 1 ≤ k ≤ 7.

Probability of k neighbors. In the cooperative CNR-based authentication, if the number of the neighbors of the source node is less than a preset threshold k, the MAC authentication does not work. Let $P_{k.neigh}$ denote the probability that there are at least k neighbors in the transmission range of a source node, then

$$
\begin{aligned}
P_{\text{k-neigh}} &= P(N \geq k|\pi R^2) \\
&= 1 - P(N < k|\pi R^2) \\
&= 1 - \sum_{j=0}^{k-1} P(N = j|\pi R^2) \\
&= 1 - \sum_{j=0}^{k-1} \binom{|\mathcal{V}|-1}{j} \cdot \left(\frac{\pi R^2}{\mathcal{S}}\right)^j \cdot \left(1 - \frac{\pi R^2}{\mathcal{S}}\right)^{|\mathcal{V}|-j-1}.
\end{aligned}
\tag{8}
$$

Fig. 5 shows the probability $P_{k.neigh}$ in a parameterized wireless sensor network with different k, (1 ≤ k ≤ 30). It can be seen the expected high probability can be achieved when choosing a proper k, i.e., k ≤ 6. As a result, the CNR based MAC authentication mechanism is feasible.

4.2 Description of Authentication

This authentication scheme consists of two phases: sensor nodes initialization and deployment, and sensed results reporting protocol.

4.2.1 Sensor Nodes Initialization and Deployment

Given the security parameter, the sink first chooses an elliptic curve (E(Fp), G, q) defined over Fp, where p is a large prime and G € E(Fp) is a base point of prime order q with |q| = k . Then, the sink selects a secure cryptographic hash function h(), where h : (0, 1)*→ Z_q* . Finally, the sink sets the public parameters as params = { E(Fp), G, q, h() }.

To initialize sensor nodes N = {N_0, N_1, N_2, ..} the sink invokes the Algorithm 1. Then, the sink deploys these initialized sensor nodes at a CIR in various ways, such as by air or by land. Given the rich literature in wireless sensor node deployment [26], [27], we do not address the deployment in detail. Without loss of generality, we assume that all sensor nodes are uniformly distributed in CIR after deployment. When these sensor nodes are not occupied by the reporting task, they cooperatively establish or adjust their routing to the sink either a shortest path or a path adapted to some resource constrains with some existing routing protocol. Note that, the established routing path can accelerate the reporting. Once an event occurs, a report can be immediately relayed along the established routing path.

Algorithm 1. Sensor Nodes Initialization Algorithm
1: **Procedure** SENSORNODESINITIALIZATION
Input: *params* and un-initialized $\mathcal{N} = \{N_0, N_1, N_2, \ldots\}$
Output: initialized $\mathcal{N} = \{N_0, N_1, N_2, \ldots\}$
2: **for** each sensor node $N_i \in \mathcal{N}$ **do**
3: preload N_i with TinyECC, *params* and energy

4: choose a random number $x_i \in \mathbb{Z}_q^*$ as the private key, compute the public key $Y_i = x_i G$, and install (Y_i, x_i) in N_i
5: **end for**
6: **return** initialized $\mathcal{N} = \{N_0, N_1, N_2, \ldots, N_n\}$
7: **end procedure**

4.2.2 Sensed Results Reporting Protocol

When a sensor node generates a report m after being triggered by a special event, e.g., a temperature change or in response to a query from the sink, it will send the report to the sink via an established routing. Assume that, the sensor (source) node N_0 has sensed some data m and is ready to report m to the sink via the routing path R_{N0}: [R1 → R2→ …Rl → Sink], as shown in Fig. 3, the following protocol steps will be executed:

Step 1. The source node N_0 gains the current timestamp
T, chooses k neighboring nodes N_{N0}: {N1, N2. . . Nk}, and sends the event (m, T) and routing R_{N0} to N_{N0}.

Step 2. With (m, T, R_{N0}) as input, each sensor node Ni Є (N_{N0} U {N_0}) invokes the Algorithm 2 to generate a row authentication vector

$$Row_i = (mac_{i1}, mac_{i2}, \ldots, mac_{il}, mac_{is}), \quad (9)$$

and reports Row_i to the source node N_0.

Algorithm 2. CNR Based MAC Generation
1: **procedure** CNRBASEDMACGENERATION
Input: *params*, $N_i \in (N_{N_0} \cup N_0)$, m, T, R_{N_0}
Output: Row_i
2: N_i uses the non-interactive keypair establishment to compute shared keys with each node in $R_{N_0} : [R_1 \rightarrow R_2 \rightarrow \cdots \rightarrow R_l \rightarrow Sink]$ as $k_{i1}, k_{i2}, \ldots, k_{il}, k_{is}$, where k_{is} is N_i's private key distributed by the *sink*
3: **if** N_i believes the report m is true **then** ▷ a neighboring node is assumed having the same ability to detect a true event as the source node and correctly judge the report m.
4: **for** $j = 1$ to l **do**
5: $mac_{ij} = MAC(m \| T, k_{ij}, 1)$
6: **end for**
7: $mac_{is} = MAC(m \| T, k_{is}, \alpha)$
8: **else**
9: **for** $j = 1$ to l **do**
10: mac_{ij} is set as a random bit
11: **end for**
12: mac_{is} is set as a random bit string of length α
13: **end if**
14: **return** $Row_i = (mac_{i1}, mac_{i2}, \ldots, mac_{il}, mac_{is})$
15: **end procedure**

Step 3. After the source node N_0 aggregates all row vectors (Row_0, Row_1, . . . ,Row_k), it formats the authentication information MAC as

$$\mathbf{MAC} = \begin{pmatrix} Row_0 \\ Row_1 \\ Row_2 \\ \vdots \\ Row_k \end{pmatrix} = \begin{pmatrix} mac_{01} & \cdots & mac_{0l} & mac_{0s} \\ mac_{11} & \cdots & mac_{1l} & mac_{1s} \\ mac_{21} & \cdots & mac_{2l} & mac_{2s} \\ \vdots & \vdots & \vdots & \vdots \\ mac_{k1} & \cdots & mac_{kl} & mac_{ks} \end{pmatrix}, \quad (10)$$

and reports (m, T, MAC) as well as N_{N0} to the sink along the routing R_{N0}.

4.2.3 En-Routing Filtering

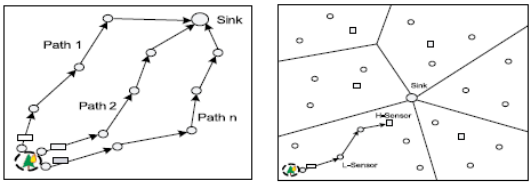
When each sensor node Ri, ($1 \le i \le l$), along the routing R_{N0} receives (m, T, MAC) from its upstream node, it checks the integrity of the message m and the timestamp T . If the timestamp T is out of date, the message (m, T, MAC) will be discarded. Otherwise, Ri invokes the Algorithm 3. If the returned value is "accept," Ri will forward the message (m, T, MAC) to its downstream node, Otherwise, (m, T, MAC) will be discarded.

Algorithm 3. CNR Based MAC Verification
1: **procedure** CNRBASEDMACVERIFICATION
Input: *params*, $R_j \in \{R_1, \ldots, R_l\}$, m, T, N_{N_0}
Output: *accept* or *reject*
2: R_j uses the noninteractive keypair establishment to compute shared keys with each node in $\{N_0, N_1, \ldots, N_k\}$ as $k_{0j}, k_{1j}, \ldots, k_{kj}$
3: set returnvalue = "*accept*"
4: **for** $i = 0$ to k **do**
5: $\overline{mac}_{ij} = MAC(m \| T, k_{ij}, 1)$
6: **if** $\overline{mac}_{ij} \oplus mac_{ij} \neq 0$ **then**
7: set returnvalue = "*reject*"
8: **break**
9: **end if**
10: **end for**
11: **return** returnvalue
12: **end procedure**

4.2.4 Sink Verification

If the sink receives the report (m,T,MAC), it checks the integrity of the message m and the timestamp T . If the timestamp is out of date, the report (m, T, MAC) will be immediately discarded. Otherwise, the sink looks up all private keys k_{is} of Ni, $0 \le i \le k$, and invokes the Algorithm 4. If the returned value of Algorithm 4 is "accept," the sink accepts the report m; otherwise, the sink rejects the report.

Algorithm 4. Sink Verification
1: **procedure** SINKVERIFICATION
Input: *params*, $k_{0s}, k_{1s}, \ldots, k_{ks}$, m, T
Output: *accept* or *reject*



(a) (b)

Fig. 6. Reliability and Scalability of the scheme. (a) Reliability with multi reports. (b) Scalability with heterogonous deployment.

```
2:    set returnvalue = "accept"
3:    for i = 0 to k do
4:       mac_is = MAC(m||T, k_is, α)
5:       if mac_is ⊕ mac_is ≠ 0 then
6:          set returnvalue = "reject"
7:          break
8:       end if
9:    end for
10:   return returnvalue
11: end procedure
```

Reliability and scalability. For the scheme, once a compromised sensor node participates in the report confirmation, the report will be polluted and cannot reach the sink. To improve the reliability, multi reports solution is naturally introduced in the scheme. As shown in Fig. 6a, once a true wildfire event occurs, multisource nodes close to the event independently choose k different neighbors, produce the multi reports and send them to the sink via different paths. Only if one report reaches the sink, the true event will successfully reported. As a result, the reliability of the scheme can be improved. In the scheme, the additional (l + α) * (k + 1) authentication bits are in linear with the length of the path l. If l is too long, the authentication bits become large. To resolve the scalability issue, we can devise a large-scale sensor network into a heterogenous sensor network [28], where each partition consists of a powerful High-end sensor (H-sensor) and a number of Low-end sensors (L-sensors), as shown in Fig. 6b. Each H-sensor serves as a cluster header. When a L-sensor senses some event, it can report to the nearby H-sensor, but not to the remote sink. Therefore, the heterogenous deployment can provide a good solution to the scalability issue of scheme.

5. Security Analysis

In this section, we analyze the security of the authentication scheme with respect to our main design goal, i.e., the effectiveness of filtering the injected false data.

5.1 Theoretical Analysis

Since the timestamp T is embedded in the report, the replay attack, a special injecting false data attack, can be filtered obviously. Therefore, how the scheme is resistant to the generic injecting false data attack will be studied here. Because the adversary A can compromise some sensor nodes in the network, without loss of generality, we assume the compromised probability for each sensor node is, and study the filtering probability.

Let a compromised sensor node N_0 be ready to report an injected false data m with a valid timestamp T to the sink. According to the protocol, N_0 should select k neighboring sensor nodes to generate the authentication information MAC together, and then send (m, T, MAC) to the sink via the routing R_{N0}: [R1 → R2 →… Rl → Sink]. In the selected k neighboring sensor nodes N_{N0} : $\{N_1,N_2,...N_k\}$,as we know, with the probability $\binom{k}{i}\rho^i(1-\rho)^{k-i}$ there are I compromised nodes. At the same time, in the routing R_{N0}, with the probability $\binom{l}{j}\rho^j(1-\rho)^{l-j}$ there are j compromised nodes among l routing nodes. Because all keys are key independence, then in order to pass the false data (m*, T*, MAC) to the authentication, the sensor node N_0 must correctly guess all authentication bits between k-i uncompromised neighboring nodes and l-j uncompromised routing nodes plus the sink. Therefore, the guess probability is,

$$Pr = P_{\text{k-neigh}} \cdot \binom{k}{i}\rho^i(1-\rho)^{k-i} \cdot \binom{l}{j}\rho^j(1-\rho)^{l-j} \cdot \frac{1}{2^{(k-i)(l+\alpha-j)}}. \tag{11}$$

Then, the false positive authentication probability is,

$$\text{FPA} = \sum_{i=0}^{k}\sum_{j=0}^{l} P_{\text{k-neigh}} \cdot \binom{k}{i}\binom{l}{j}\rho^{i+j}(1-\rho)^{k+l-i-j} \cdot \frac{1}{2^{(k-i)(l+\alpha-j)}}. \tag{12}$$

Furthermore, we can obtain the filtering probability under this circumstance as

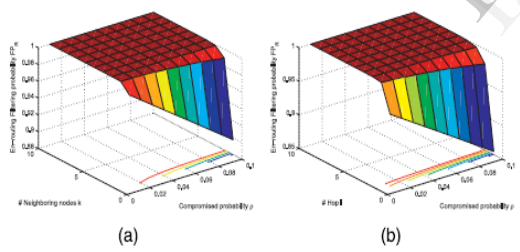
$$\text{FP} = 1 - \text{FPA} = 1 - P_{\text{k-neigh}} \cdot \sum_{i=0}^{k}\sum_{j=0}^{l}\binom{k}{i}\binom{l}{j}\rho^{i+j} (1-\rho)^{k+l-i-j} \cdot \frac{1}{2^{(k-i)(l+\alpha-j)}}. \tag{13}$$

When $\alpha = 0$, FP is rewritten as

$$\mathrm{FP}_R = 1 - P_{\text{k-neigh}} \cdot \sum_{i=0}^{k} \sum_{j=0}^{l} \binom{k}{i} \binom{l}{j} \rho^{i+j} (1-\rho)^{k+l-i-j} \cdot \frac{1}{2^{(k-i)(l-j)}}. \quad (14)$$

which represents the en-routing filtering probability of the scheme and measures how much injected false data can be filtered as early as possible before their reaching the sink, in such a way the energy waster can be reduced, and the sink can avert the DoS attack due to large number of injected false data.

Fig. 7 plots how the en-routing filtering probability FP_R varies with the number of neighboring node k, the number of en-routing nodes l, and the compromised probability p. From the figure, when k and l are properly set, FP_R approaches to 1 in theory. However, in reality, when an experienced and astute adversary A launches an attack, it may first choose those compromised nodes as its neighbors participating in the injecting false data attack to increase the success probability, then the FP_R would be reduced. Therefore, it is of interest to use simulation to evaluate the en-routing filtering probability FP_R of the scheme.



(a) (b)

Fig. 7. The en-routing filtering probability FP_R as the functions of the number of neighboring nodes k and the compromised probability p, and the number of en-routing node l and the compromised probability p (a) FP_R versus k and, (b) FP_R versus l and p

5.2 Simulation-Based En-Routing Filtering Evaluation

In the simulation, the en-routing filtering probability can be tested as,

$$\mathrm{FP}_R = \frac{\text{number of false data filtered by en-route nodes}}{\text{total number of false data}}. \quad (15)$$

In what follows, we provide the simulation results for FP_R.

5.2.1 Simulation Settings

We study FP_R of the scheme using a simulator built in Java. In the simulations, 1,000 sensor nodes with a transmission range R are randomly deployed in a CIR of region 200*200 m^2 interest region. We consider each sensor node could be compromised with the probability. In Table 1, we list the simulation parameters. Then, we test the networks when the number of en-routing nodes in the interest areas is varied from 5 to 15 in increment of 1. For each case, 10,000 networks are randomly generated, and the average of en-routing filtering probabilities over all of these randomly sampled networks are reported.

5.2.2 Simulation Results

Fig. 8 shows the en-routing filtering probability FPR in terms of different number of en-routing nodes. As the number of routing nodes increases, FPR increases. At the same time, by choosing more neighboring nodes involved in the protocol, i.e., the parameter k increases, FP_R will further increase, even the compromised probability is 5 percent. Further observing the FPR with different transmission range R, we can see a relatively low FP_R for R = 20 m compared with that for R = 15m. The reason is that, under the same settings, when the transmission range increases, the number of compromised neighboring nodes will also increase, so the experienced and astute A has more chances to choose more compromised nodes participating in the attack to increase the success attack probability.

TABLE 1
Parameter Settings

Parameter	Value
Simulation area	200m × 200m
Number of sensor nodes	1000
Transmission range R	15m, 20m
Compromised probability ρ	2%, 5%
# neighboring nodes k	4, 6
# routing nodes l	$5, \cdots, 15$

Based on these observations, we have the following theorem.

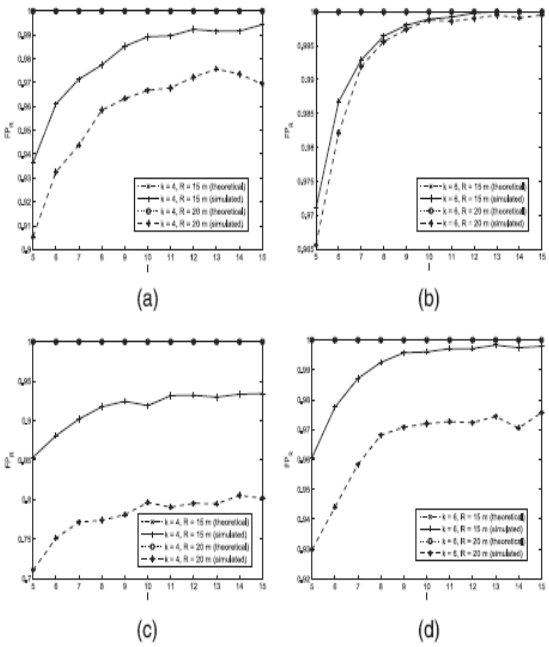


Fig. 8. En-routing filtering probability FP_R versus the different number of routing nodes l, where $5 \leq l \leq 15$. (a) p = 2%, k = 4, (b) p = 2%, k = 6, (c) p = 5%, k = 4 (d) p = 5%, k = 6.

Theorem 1. The scheme can effectively resist the injecting false data attack launched by the experienced and astute A, only if the number of compromised nodes in the transmission range R is less than the security parameter k.

Proof. From (13)-(14), we have the following relationship between FP and FP_R, i.e.,

$$\mathrm{FP} = 1 - (1 - \mathrm{FP}_R) \cdot \frac{1}{2^{\alpha}}, \qquad (16)$$

where 1-FP_R is the success probability of injecting false attack escaping from the en-routing filtering, which consists of two parts:1)$FPA|_{Nc=k}$, the false positive probability when the number of participating neighboring compromised nodes Nc = k in the attack, and 2) $FPA|_{N\,c<k,}$ the false positive probability when Nc < k. Therefore, we have

$$\mathrm{FP} = 1 - (\mathrm{FPA}_{|N_c=k} + \mathrm{FPA}_{|N_c<k}) \cdot \frac{1}{2^{\alpha}}. \qquad (17)$$

When the parameter α is well chosen, the item $FPA_{|Nc<k}$ $.1/2^{\alpha} \rightarrow 0$. However, $1/2^{\alpha}$ does not affect $FPA_{|Nc<k,}$ since all participating t nodes are compromised. Thus, we have $FP=1-FPA_{|nc<k}$. because the condition Nc=k is determined by the number of compromised nodes in transmission range r, if this condition does not hold, FP=1-$FPA_{|Nc<k=}$1. Therefore only if number of compromised nodes in transmission range R is less than parameter k, this scheme can effectively the injecting false data attack launched by experienced and astute A.

Fig. 9 also shows the filtering ratio at each en-routing node Ri in R_{N0}, where $1 \leq i \leq 10$. The results confirm our design goal, i.e., the injected false data can be early detected and filtered by the en-routing sensor nodes. Thus, the energy wasted in relaying injected false data can be reduced.

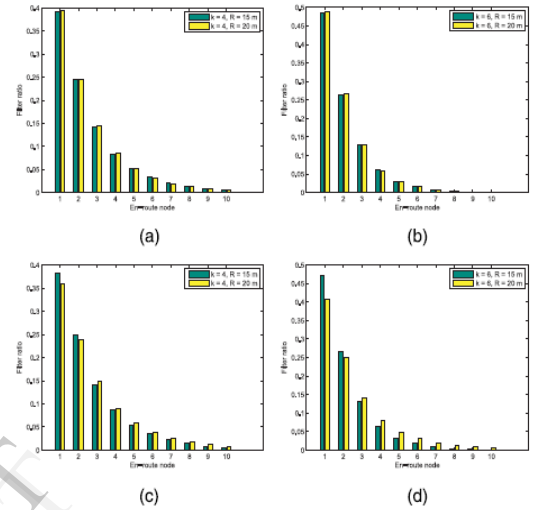


Fig.9. The filtering ratio at each routing node Ri in R_N, where $1 \leq i \leq 10$. (a) p = 2%,k = 4, (b) p = 2%, k = 6, (c) p = 5%, k = 4 (d) p = 5%, k = 6.

Reliability of the scheme. In addition to the high (en-routing) filtering probability, the scheme also has high reliability, i.e., even though some sensor nodes are compromised, the true event reports still can reach the sink with high probability. Let FNR be the false negative rate on the true reports and tested as,

$$\mathrm{FNR} = \frac{\text{number of true data that cannot reach the } \textit{sink}}{\text{total number of true data}}. \qquad (18)$$

If FNR is small, the scheme is demonstrated high reliability. Note that, selectively dropping true report attack [18] can increase the FNR. However, its adverse impact can affect any routing algorithm. Thus, for fairness, we only consider FNR that caused by 1) the number of uncompromised neighboring sensor nodes being less than k, or 2) some compromised sensor nodes polluting the true report. Fig. 10 shows

the false negative rate FNR versus different number of reports. It can be seen, when the number of independent reports increases, the FNR decreases. Especially, when the number is five, the FNR is less than 10 percent. In reality, when a true wildfire event takes place, usually several independent entities report the event. Thus, the multi reports technology in scheme fits to the realistic scenarios. As a result, the scheme can achieve high reliability.

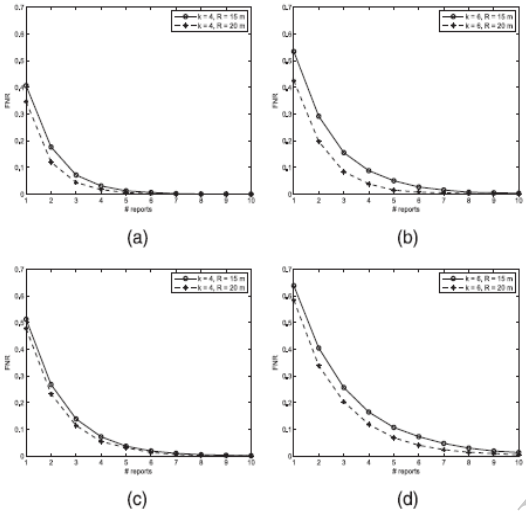


Fig. 10. The false negative rate FNR in terms of different number of independent reports, where the number is from 1 to 10. (a) p = 2%, k=4, (b) p = 2%, k = 6, (c) p = 5%, k = 4, and (d) p = 5%, k = 6.

5.3 Discussion on Gang Injecting False Data Attack

In this section, we introduce a new stronger injecting false data attack, called gang injecting false data attack, in wireless sensor networks. This kind of attack is usually launched by a gang of compromised sensor nodes controlled and moved by an adversary A. As shown in Fig. 11, when a compromised source node is ready to send a false data, several compromised nodes will first move and aggregate at the source node, and then collude to inject the false data. Because of the mobility, the gang injecting false data attack is more challenging and hard to resist.

To tackle this kind of attack, a possible solution with the scheme is to require each participating sensor node to provide its position information. If the current position is not consistent with the previous ones, the gang attack can be detected. Nevertheless, how to prevent/mitigate the gang injecting false data attack from mobile compromised sensor nodes is still worthy of the further investigation.

6. Performance Evaluation

Energy saving is always crucial for the lifetime of wireless sensor networks. In this section, the performance of the proposed scheme is evaluated in terms of energy efficiency.

6.1 Energy Consumption in Non-interactive Keypair Establishments

The additional computation costs of the proposed scheme are mainly due to the expensive ECDH operations during the noninteractive keypair establishments. Fortunately, since the noninteractive keypair establishments are averagely distributed in each sensor node and

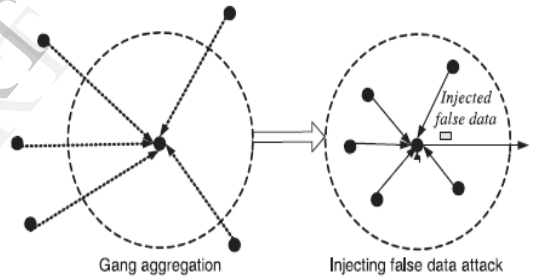


Fig. 11. Gang injecting false data attack.

only executed once during the routing establishment, the ECDH operation is not a heavy burden. When designing TinyECC-based sensor node, we can choose a 160-bit elliptic curve for achieving the same security level as 1,024-bit RSA [25]. Assume that, each sensor node is equipped with a low-power high performance sensor platform, i.e., MICAz [21]. Then, according to [19], this type of sensor platform only requires 50.82 mJ to establish a noninteractive shared key.

7. Related Work

Recently, some research works on bandwidth-efficient filtering of injected false data in wireless sensor networks have been appeared in the literature in [9], [10], [11], [12], [13]. In [9], Ye et al. propose a statistical en-routing filtering mechanism called SEF. SEF requires that each

sensing report be validated by multiple keyed message authenticated (MACs), each generated by a node that detects the same event. As the report being forwarded, each node along the way verifies the correctness of the MACs at earliest point. If the injected false data escapes the en-routing filtering and is delivered to the sink, the sink will further verify the correctness of each MAC carried in each report and reject false ones. In SEF, to verify the MACs, each node gets a random subset of the keys of size k from the global key pool of size N and uses them to producing the MACs. To save the bandwidth, SEF adopts the bloom filter to reduce the MAC size. By simulation, SEF can prevent the injecting false data attack with 80-90 percent probability within 10 hops. However, since n should not be large enough as described above, the filtering probability at each en-routing node p is relatively low. Besides, SEF does not consider the possibility of en-routing nodes' compromise, which is also crucial to the false data filtering. In [10], Zhu et al. present an interleaved hop-by-hop authentication (IHA) scheme for filtering of injected false data. In IHA, each node is associated with two other nodes along the path, one is the lower association node, and the other is the upper association node. An en-routing node will forward received report if it is successfully verified by its lower association node. To reduce the size of the report, the scheme compresses t + 1 individual MACs by XORing them to one. By analyses, only if less than t nodes are compromised, the sink can detect the injected false data. However, the security of the scheme is mainly contingent upon the creation of associations in the association discovery phase. Once the creation fails, the security cannot be guaranteed. In addition, as pointed in [7], Zhu et al.'s scheme, similar as SEF, also adopts the symmetric keys from a key pool, which allows the compromised nodes to abuse these keys to generate false reports. Location-Based Resilient Secrecy (LBRS) is proposed by Yang et al. [11], which adopts location key binding mechanism to reduce the damage caused by node compromise, and further mitigate the false data generation in wireless sensor networks. In [12], Ren et al. propose more efficient location-aware end-to-end data security design (LEDS) to provide end-to-end security guarantee including efficient en-routing false data filtering capability and high-level assurance on data availability. Because LEDS is a symmetric key based solution, to achieve en-routing filtering, it requires location-aware key management, where each node should share at least one authentication key with one node in its upstream/downstream report auth cell. In [13], Zhang et al. provide a public key based solution to the same problem. Especially, they propose the notion of location-based keys by binding private keys of individual nodes to both their IDs and geographic locations and a suite of location-based compromise-tolerant security mechanisms. To achieve en-routing filtering, additional 20 bytes authentication overheads are required. Bit-compressed authentication technology can achieve bandwidth-efficient, which has been adopted in some research works [29], [30]. In [29], Canetti et al. use one-bit authentication to achieve multicast security. The basic idea in multicast is very similar to the BECAN scheme, where a source knows a set of keys R = {K1, . . . , Kl}, each recipient u knows a subset Ru € R. When the source sends a message M, it authenticates M with each of the keys, using a MAC. Each recipient u verifies all the MACs which were created using the keys in its subset Ru. If any of these MACs is incorrect, the message M will be rejected. To achieve the bandwidth efficiency, each MAC is compressed as single bit. The security of the scheme is based on the assumption that the source is not compromised. However, once the source is compromised, the scheme obviously does not work. Therefore, it cannot be applied to filter false data injected by compromised nodes in wireless sensor networks. In [30], Benenson et al. also use 1-bit MACs to decide whether a query is legitimate in wireless sensor networks. However, similar as that in [29], once the source is compromised, the 1-bit MACs also do not work. Different from the above works, the proposed scheme adopts CNR based filtering mechanism together with multireports technology. Because of noninter-active key establishment, does not require a complicated security association [10], [12]. In addition, it considers the scenario that each node could be compromised with probability, i.e., some enrouting nodes could be compromised. To avoid putting all eggs in one basket, scheme distributes the en-routing authentication to all sensor nodes along the routing path. To save the bandwidth, it also adopts the bit-compressed authentication technique. Therefore, it is compromise-tolerant and suitable for filtering false data in wireless sensor networks.

8. Conclusion

In this paper, we have proposed a novel scheme for filtering the injected false data. By theoretical analysis and simulation evaluation, the scheme has been demonstrated to achieve not only high en-routing filtering probability but also high reliability with multi-reports. Due to the simplicity and effectiveness, the scheme could be applied to other fast and distributed authentication scenarios, e.g., the efficient authentication in wireless mesh network [31]. In our future work, we will investigate how to prevent/mitigate the gang injecting false data attack from mobile compromised sensor nodes [32].

References

[1] R. Szewczky, A. Mainwaring, J. Anderson, and D. Culler, "An Analysis of a Large Scale Habit Monitoring Application," *Proc. Second ACM Int'l Conf. Embedded Networked Sensor Systems (Sensys '04), 2004.*

[2] L. Eschenauer and V.D. Gligor, "A KeyManagement Scheme for Distributed Sensor Networks," *Proc. Ninth ACM Conf. Computer and Comm. Security (CCS '02), 2002.*

[3] R. Lu, X. Lin, C. Zhang, H. Zhu, P. Ho, and X. Shen, "AICN: An Efficient Algorithm to Identify Compromised Nodes in Wireless Sensor Network," *Proc. IEEE Int'l Conf. Comm. (ICC '08), May 2008.*

[4] X. Lin, R. Lu, and X. Shen, "MDPA: Multidimensional Privacy-Preserving Aggregation Scheme for Wireless Sensor Networks," *Wireless Comm. and Mobile Computing, vol. 10, pp. 843-856, 2010.*

[5] X. Lin, "CAT: Building Couples to Early Detect Node Compromise Attack in Wireless Sensor Networks," *Proc. IEEE GLOBECOM '09, Nov.-Dec. 2009.*

[6] K. Ren, W. Lou, and Y. Zhang, "Multi-User Broadcast Authentication in Wireless Sensor Networks," *Proc. IEEE Sensor Ad Hoc Comm. Networks (SECON '07), June 2007.*

[7] L. Zhou and C. Ravishankar, "A Fault Localized Scheme for False Report Filtering in Sensor Networks," *Proc. Int'l Conf. Pervasive Services, (ICPS '05), pp. 59-68, July 2005.*

[8] Z. Zhu, Q. Tan, and P. Zhu, "An Effective Secure Routing for False Data Injection Attack in Wireless Sensor Network," *Proc. 10th Asia-Pacific Network Operations and Management Symp. (APNOMS '07), pp. 457-465, 2007.*

[9] F. Ye, H. Luo, S. Lu, and L. Zhang, "Statistical En-Route Detection and Filtering of Injected False Data in Sensor Networks," *Proc. IEEE INFOCOM '04, Mar. 2004.*

[10] S. Zhu, S. Setia, S. Jajodia, and P. Ning, "An Interleaved Hop-by-Hop Authentication Scheme for Filtering of Injected False Data in Sensor Networks," *Proc. IEEE Symp. Security and Privacy, 2004.*

[11] H. Yang, F. Ye, Y. Yuan, S. Lu, and W. Arbaugh, "Toward Resilient Security in Wireless Sensor Networks," *Proc. Sixth ACM Int'l Symp. Mobile Ad Hoc Networking and Computing (MobiHoc '05), pp. 34-45, 2005.*

[12] K. Ren, W. Lou, and Y. Zhang, "LEDS: Providing Location-Aware End-to-End Data Security in Wireless Sensor Networks," *Proc. IEEE INFOCOM '06, Apr. 2006.*

[13] Y. Zhang, W. Liu, W. Lou, and Y. Fang, "Location-Based Compromise-Tolerant Security Mechanisms for Wireless Sensor Networks," *IEEE J. Selected Areas in Comm., vol. 24, no. 2, pp. 247-260, Feb. 2006.*

[14] C.-M. Yu, C.-S. Lu, and S.-Y. Kuo, "A Dos-Resilient En-Route Filtering Scheme for Sensor Networks," *Proc. Tenth ACM Int'l Symp. Mobile Ad Hoc Networking and Computing (MobiHoc '09), pp. 343-344, 2009.*

[15] J. Chen, Q. Yu, Y. Zhang, H.-H. Chen, and Y. Sun, "Feedback Based Clock Synchronization in Wireless Sensor Networks: A Control Theoretic Approach," *IEEE Trans. Vehicular Technology, vol. 59, no. 6, pp. 2963-2973, June 2010.*

[16] S. He, J. Chen, Y. Sun, D.K.Y. Yau, and N.K. Yip, "On Optimal Information Capture by Energy-Constrained Mobile Sensors," *IEEE Trans. Vehicular Technology, vol. 59, no. 5, pp. 2472-2484, June 2010.*

[17] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Ad Hoc Networks, vol. 3, no. 3, pp. 325-349, May 2005.*

[18] V.C. Giruka, M. Singhal, J. Royalty, and S. Varanasi, "Security in Wireless Sensor Networks," *Wireless Comm. and Mobile Computing, vol. 8, no. 1, pp. 1-24, Jan. 2008.*

[19] A. Liu and P. Ning, "TinyECC: A Configurable Library for Elliptic Curve Cryptography in Wireless Sensor Networks," *Proc. Seventh Int'l Conf. Information Processing in Sensor Networks (IPSN '08), pp. 245-256, Apr. 2008.*

[20] J. Dong, Q. Chen, and Z. Niu, "Random Graph Theory Based Connectivity Analysis in Wireless Sensor Networks with Rayleigh Fading Channels," *Proc. Asia-Pacific Conf. Comm. (APCC '07), pp. 123-126, Oct. 2007.*

[21] MICAz: Wireless Measurement System, http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICAz_Datasheet.pdf, 2010.

[22] Imote2: High-Performance Wireless Sensor NetworkNode,http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/Imote2_Dat asheet.pdf, 2010.