IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 2, 2020, accepted February 18, 2020, date of publication February 28, 2020, date of current version March 11, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977130

A Location and Optimal Coverage Based Filtering Scheme in Wireless Sensor Networks

ZHIXIONG LIU, HUI YE, AND FANGMIN LI
School of Computer Engineering and Applied Mathematics, Changsha University, Changsha 410022, China
Hunan Province Key Laboratory of Industrial Internet Technology and Security, Changsha University, Changsha 410022, China

Corresponding author: Hui Ye (csuleo@foxmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772088, in part by the Natural Science Foundation of Hunan Province of China under Grant 2019JJ40326, in part by the Excellent Youth Project of the Educational Department of Hunan Province of China under Grant 18B410, in part by the Project of the Educational Department of Hunan Province of China under Grant 19C0171, in part by the Hunan Province Key Laboratory of Industrial Internet Technology and Security under Grant 2019TP1011, and in part by the Changsha Science and Technology Plan under Grant kc18099013.

ABSTRACT Injected false reports by attackers bring fateful consequences in wireless sensor networks, i.e., wasting the limited batteries of nodes and misleading the decision making of users. Existing security designs mainly attach some extra fields by a group of sensing nodes after the pure data, and check the correctness of attached MACs (Message Authentication Codes) in the process of report forwarding, each of which represents the agreement of sensors on the report, thus to drop the ones which failed on MAC checking. They cannot recognize the reports forged by t arbitrary compromised sensors collaboratively; the variable t is a security parameter. Furthermore, failures of reporting usually occur in sparse regions for lacking of enough detecting sensors, which is incurred by the complicated deploying region and the adopted random deploying strategy. This paper proposes a *Location and Optimal Coverage based Filtering scheme* (LOCF). It first derived the optimal coverage degree Δ by considering both the network size and covering efficiency, and then employed covering algorithm to deploy sensors accordingly. Each deployed sensor has to dispense its location to downstream sensors, through which sensors are bounded with locations. A report for the observed event must attach t endorsements along with locations of detecting sensors. In the forwarding process, intermediate sensors evaluate the correctness and rationality of both MACs and locations. Simulation results demonstrate that LOCF outperforms existing works in terms of covering effectiveness, filtering efficiency and compromise robustness.

INDEX TERMS Wireless sensor network, injected false reports, optimal coverage, location, compromise robustness.

I. INTRODUCTION

Wireless sensor network (WSN) is usually composed of a large quantity of sensor nodes which deployed in unattended environments, and plays core roles in a huge amount of scenarios, such as intrusion detection, target tracing, farm automation, etc. [1], [2]. For the fact that only being equipped with limited batteries and weak protecting abilities, sensor nodes are easily compromised by attackers, resulting in leakage of all stored secret information. The attacker can manipulate these secrets to forge reports, thus to deceive the user and also waste the limited resources of the sensors [3]. Even worse, sensors located not adjacent can forge reports together, and this behavior is considered illegal in actually deployed networks [4].

To resist the injection of faked reports, several security designs [7]–[18] have been proposed in the last years. The common idea of them is to share symmetric keys among sensors, or establish associated relationship between sink and the other nodes, and rely on forwarding sensors to authenticate the attached MACs in the reports, thus to filter out the faked ones with invalid MACs in a probabilistic manner en-route.

However, two defects reduce the efficiency of these schemes severely. First, MAC checking cannot recognize the reports faked by t compromised sensors, thus sensors located not adjacent are able to launch conspiratorial attacks. Second, sparse areas are normal in actually deployed networks due to complex environment and other factors, and events

The associate editor coordinating the review of this manuscript and approving it for publication was Rongbo Zhu.

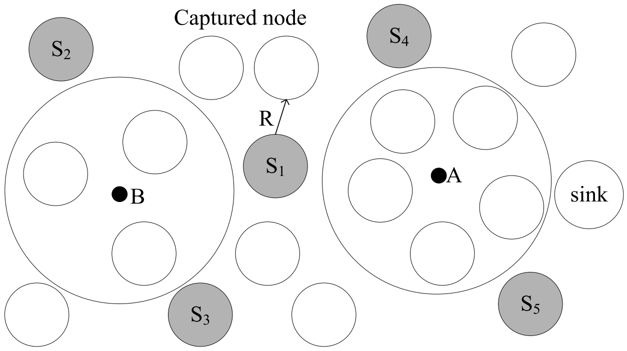


FIGURE 1. **Conspiratorial attacks and sparse covering.**

occurred in such places are easily failed of reporting for the reason of lacking t detecting sensors. As depicted in Fig. 1, assuming that $t = 5$ and sensors $S_1, \ldots, S_5$ are controlled by the attacker. Obviously, these sensors can be abused to fake reports together escaping the checks of existing schemes. And we also observe that only three sensors covering B; as a result, events occurred near B cannot be reported successfully.

To solve the above mentioned problems, we propose a LOCF scheme which guarantees the covering requirements of the monitoring region through an optimal covering algorithm, and resist the conspiratorial attacking behavior of sensors in arbitrary areas by checking the logic of locations. Theoretical analysis and simulations verified the performance shifting than existing mechanisms. The main contributions of this paper are summarized as follows:

First, through the exchange of locations among sensors after deployment, the relationship between sensors and locations are established, and sensors located nonadjacent are not able to forge reports together which sneaking through the intermediate verification successfully. As a result, after being captured, the damage scope of sensors are limited in their local areas. Experiments show that with ten sensors captured, there is only 3.2% of the forged reports escaping the security checking of LOCF, while a huge fraction (92.8%) of forged reports are unrecognizable in existing schemes.

Second, the optimal covering degree Δ is derived from covering efficiency and overhead of the network. Compared with random deployment, Δ covering strategy guarantees much better covering performance on t-threshold based filtering mechanisms. Further analysis and simulation results illustrate that, in situation of $2t$ covering degree, the probabilities in Δ covering and random deployment to guarantee t-threshold based covering are 80% and 2.6%, respectively.

The rest part of the paper is organized as follows. Related filtering schemes and covering algorithms are introduced in Section II. The derivation of optimal covering degree is illustrated in Section III. Detailed design of the proposed LOCF scheme is presented in Section IV. We evaluate the performance of LOCF in Section V. Simulations are shown in Section VI. Finally, Section VII is the conclusions.

II. RELATED WORK

Existing filtering mechanisms mainly adopt symmetric key technique [7]–[14] or public key infrastructure [15]–[17] to encrypt and verify reports. For reason that mainstream sensors only equipped with limited battery and computational abilities, symmetric key based methods may be more applicable than public key based ones [5], [6]. Fan *et al.* [7] firstly investigated false data injection problem and proposed a probabilistic en-route verifying scheme (SEF). It equips each sensor with a group which is used to generate endorsement for the reports. For the fact that the number of groups is much less than the number of nodes, each group would be shared by a group of nodes. As a result, MACs generated by detecting sensors may be verified by downstream sensors with the same group. However, the relationship between number of groups and size of the network is not investigated in the scheme.

Yu *et al.* [8] presented a grouping-based mechanism (GRSEF), which divides sensors into exactly t groups rather than n groups in SEF ($n > t$). Grouping method is achieved through multiple axes based keys distribution. It eliminated redundant groups thus improved the filtering efficiency greatly; however, the robustness to tolerate sensors failure is weakened than existing schemes. GRSEF also provides a feasible solution for filtering false data in situations of mobile sinks.

Yu *et al.* [9] presented an overhead balanced scheme (DEFS). Along multiple fixed paths, keys are distributed through the so-called *Hill Climbing* method. In this method, the number of keys distributed is inversely proportional to distance between source sensor and forwarding sensor. Consequently, overhead of sensors can be balanced to a certain extent. However, fixed paths are fragile to dynamic topology of sensor networks.

Bashir *et al.* [10] proposed a ERFS scheme to filter out RFID copies in sensor networks. The authors argue that redundant copies shorten the lifetime of the network during forwarding such useless information. Differ to related RFID copy filtering schemes which check copies in BS, ERFS performs copy detecting by cluster heads. However, the cluster based organization is unsuitable for the situation when event happens among multiple clusters.

Dobrev *et al.* [11] proposed a OARB scheme with the help of rotating and bean sensors. It aims to detect the intruder within some fixed period under the assumption that intruder is moveable and appears everywhere. Minimum number of rotating sensors and beam sensors are used to monitor attacking behavior on the path from cluster head to sink. It performs well in optimal networks, but is hard to be extended to regular networks with constrained capabilities.

Li *et al.* [12] presented a voting mechanism based filtering algorithm called PVFS. In the scheme, each intermediate node uses the probability of d_i / d_0 to store the authentication key for a node within the source cluster, and verifies the votes carried in the report during forwarding. Here d_i and d_0 mean the distance between source cluster to sink, and the distance between forwarding cluster to sink, respectively. Based on PVFS, Su *et al.* [13] utilizes context-aware architecture (CAA) to analyze sensing data, and then sends decision message to ordinary nodes. Guan *et al.* [14] judges the confidence of information through the residual signal generated by the estimators. However, the deploying cost of CAA and estimator is high, and the locating accuracy is inefficient in both of them.

Yang *et al.* [15] proposed a public key technique based scheme (CCEF). Differ than in symmetric key based schemes, it first shares a pair of public keys between each sensor and sink, and then verify reports on forwarding sensors through a witness key. Wang *et al.* [16] adopted elliptic curve cryptography technique to amend CCEF. Ayday *et al.* [17] employs the idea of random network coding to verify reports.

On sensor network covering, Wei *et al.* [18] raised a dynamic mechanism called DCA, which implements trans-regional coverage through adapting the geometric boundaries to the overlapping region. Moreover, multiple sensor nodes are scheduled synchronously utilizing the Centralized Voronoi Tessellation theory. This algorithm is able to achieve high covering efficiency.

Gil *et al.* [19] proposed a solution for a special NP-complete problem: Maximum Set Covers for DSNs (MSCD), and also presented a node dispatching method for target coverage, which takes use of genetic algorithm to seek out the optimal collection of covers, thus to prolong the monitoring life cycle of the network.

Bara'a *et al.* [20] modeled the MSCD problem, and then constructed a kind of energy-efficient sensor network to provide stable covering performance. The algorithm can further enhance the reliability of the generated set cover by a post-heuristic step.

As can see above, without checking the legitimacy of positions, existing filtering mechanisms fail to resist collusion attacks; and sensor coverage required by t-based authentication is not regulated in them which reduced the applicability of these algorithms. In this paper, we commit to find a solution to cope with these problems simultaneously.

III. OPTIMAL COVERAGE DEGREE

As discussed before, random deployment method forms some sparse areas in actual deployed networks and thus is not suitable for false data filtering. We adopt a covering algorithm to deploy sensors. For simplicity, the covering of at least t groups simultaneously is noted as $t-$k cover; For some covering algorithm with covering degree Δ, the probability of t-k cover is noted as $p_\Delta(t, \ldots, n)$, here n is total number of groups.

Definition 1: Given the covering expectation threshold θ and increment threshold ε, here $0< \theta <1$, $0< \varepsilon <1$. Assume that all sensors in the network are deployed according to some kind of covering algorithm. With Δ increasing, if $(p_\Delta(t, \ldots, n) - p_{\Delta-1}(t, \ldots, n) \geq \varepsilon)$ and $(p_{\Delta+1}(t, \ldots, n) - p_\Delta(t, \ldots, n) < \varepsilon)$ under the condition that $p_\Delta(t, \ldots, n) \geq \theta$, we say Δ is *optimal*.

Here the parameter θ is set to guarantee that the covering degree is large enough, thus to meet the requirements of t-based filtering. The selection of θ is a trade-off between energy cost and filtering ability. A larger θ brings a bigger t-k cover probability required by the filtering algorithm; while it also incurs a larger energy cost resulted by a higher covering degree. Only if the covering degree meets the requirement limited by θ, we then start to seek out the optimal covering degree. The criteria is that, given a threshold ε, (it should be small enough), with the increment of the covering degree, if the increment of $t-$k cover probability is not larger than ε, we treat the covering degree Δ as the optimal covering degree. That is because with the increment of Δ, the $t-$k cover probability only has a small increment while the energy consumption can be increased considerably. The selection of ε is similar to θ.

Theorem 1: Assuming each sensor randomly selects one from n groups. The probability for getting t different groups for totally Δ sensors is

$$p_\Delta(t, \ldots, n) = 1 - \frac{\sum_{m=1}^{t} C(n, m)C(\Delta - 1, \Delta - m)}{C(n + \Delta - 1)} \tag{1}$$

Proof: Note B_1 as the collection of Δ nodes, and B_2 as the collection of n groups. There are $C(n + \Delta - 1, \Delta)$ combinations for each element in B_1 maps one group in B_2 randomly. And there are $C(n, t)C(\Delta-1, \Delta-t)$ combinations for t sensors holding different groups. So, the probability of t sensors holding different groups is

$$p_\Delta(t) = \frac{C(n, t)C(\Delta - 1, \Delta - t)}{C(n + \Delta - 1, \Delta)} \tag{2}$$

The probabilities for each sensor to hold $1, \ldots, t$-1 groups can be calculated in the same way. Therefore, the probability for each sensor to hold at most t groups can be computed as

$$p_\Delta(1, \ldots, t) = \frac{\sum_{m=1}^{t} C(n, m) \cdot C(\Delta - 1, \Delta - m)}{C(n + \Delta - 1, \Delta)} \tag{3}$$

Thus the probability of holding at least t groups is $p = 1 - (p_\Delta(1)+p_\Delta(2)+\ldots+p_\Delta(t))$.

On the one hand, the covering degree could not be too small so as to guarantee the effectiveness of $t-$k cover; on the other hand, it could not be too large to avoid incurring huge energy cost for energy constrained sensor networks. As a result, a suitable covering degree should be selected to meet the aforementioned two requirements simultaneously. The following is to derive the optimal covering degree.

Theorem 2: Assuming sensors are deployed according to some kind of covering algorithm, and each data report is attached with t MACs. Let ε be 0.05, and θ be 0.8, we say $\Delta = 2t$ is the optimal covering degree for t-based filtering.

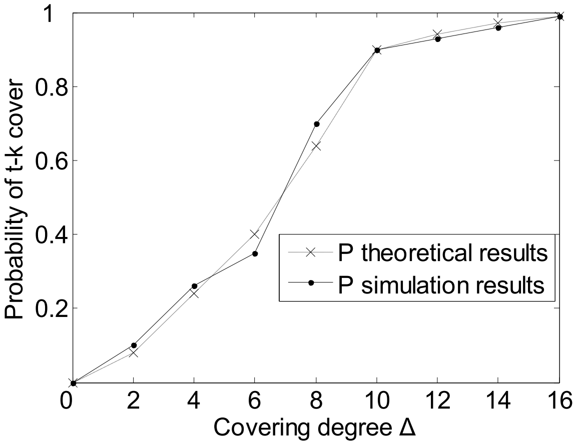


FIGURE 2. Theoretical and simulation results of covering probability.

Proof: According to theorem 1, we get,

$$P_{\Delta+1}(t) = \frac{C(n,t)\cdot C(\Delta,\Delta+1-t)}{C(n+\Delta,\Delta+1)} = \frac{\Delta\cdot(\Delta+1)}{(\Delta+1-t)\cdot(n+\Delta)}P_\Delta \quad (4)$$

As n is larger than one, when $t = 1$, we can get the following equation,

$$P_2(t) = \frac{C(n,1)\cdot C(1,1)}{C(n+1,2)} = \frac{2}{n+1} \leq 0.8 \quad (5)$$

We assume that Equation 4 is right under the condition of $t = k$, then in the case that $t = k{+}1$, we can get,

$$P_2(k+1) = \frac{C(n,k+1)C(2k+1,k+1)}{C(n+2k+1,2k+2)} = \frac{4(n-t)(2t+1)^2}{(t+1)(n+2t+1)(n+2t)}P_{2t} \quad (6)$$

Obviously,there is,

$$\frac{4(n-t)(2t+1)^2}{(t+1)(n+2t+1)(n+2t)} \leq 1 \quad (7)$$

Consequently, we get $P_2(k+1) \leq P_2(t) \leq 0.8$. Using derivative methods, in case that $\Delta \geq 2t$, there is $\text{In}(\Delta+1,\Delta) \leq 0.05 = \varepsilon$. In the similar way, we can compute $\text{In}(\Delta,\Delta-1) \geq 0.05 = \varepsilon$. From the above we know, for the corresponding covering algorithm, only when its covering degree Δ equals to $2t$, it is the optimal one for t-based filtering.

For simplicity, we investigate the changes of p according to Δ by treating both of the number of groups n and the system threshold t as constants. Fig. 2 illustrates the mathematical and experimental curves of p when $t = 5$, $n = 10$, $\theta = 0.8$, and $\varepsilon = 0.05$. The experimental values are averaged on 500 randomized experiments. From the curve we know, when $\Delta < 2t$, with the increasing of covering degree, p increases quickly. For example, when the covering degree increases from seven to eight, the increment of p is 0.02 (larger than ε). While when $\Delta > 2t$, p can only get a small increment according to the increase of covering degree, e.g, the increment of p is $0.04 < \varepsilon$ when the covering degree increases from 11 to 12. Therefore, both theoretical analysis and experimental results show that $\Delta = 2t$ is the optimal covering degree.

IV. THE LOCF SCHEME

LOCF includes four stages: network initialization, Event reporting, intermediate verification and sink authentication. For faked reports escaping from verification successfully, sink executes the final judgement. As the sink guarding phase is almost the same as in existing en-route verification strategies, we are going to omit this here.

A. NETWORK MODEL AND ATTACK MODEL

The target network considered in this paper includes a huge amount of ordinary sensors, which are all not equipped with anti-tamper devices. The sensing and transmission radii of ordinary sensors are denoted as r_s and r_c, respectively. These battery-powered sensors only have weak capabilities on computing, transmission and self-protection. As a comparison, the sink node has ample energy and powerful data processing ability. As the ultimate guard of the security design, sink is able to detect out all faked reports through the full knowledge of all secrets.

We assume that the sensors are intensively deployed, so that each area is usually covered by a group of sensors simultaneously. But sparse regions are also common in some complicated terrain. After detecting an event, the sensor with the strongest signal (called SS) on this event among all detecting sensors is selected as the represent to generate a data report. We assume that all other detecting sensors can communicate with the SS using transmission radius r_c. The SS summarizes all the detecting results and synthesizes a legitimate report, which is then forwarded to sink through traversing several hops.

We assume that sensor network has a short safe bootstrapping phase after network deployment, during which sensor nodes are safe (attackers cannot compromise them) to distribute keys, locations and neighbor information. After the bootstrapping phase, the adversary can compromise multiple sensor nodes, obtain their security information and take full control of them. We further assume that the attacker cannot compromise the sink. In this paper, we focus on collaborative false report injection attacks [1], [20], i.e., a group of compromised nodes from different geographical areas inject forged data reports into the network; however, if the adversary has compromised enough nodes within a circular area with radius r_s, it is able to launch various kinds of attacks to ruin the network completely, which is out of the scope of this paper.

B. NETWORK INITIALIZATION

Before being deployed, each sensor takes k keys from a group including m keys randomly ($k < m$). There are totally n groups with equal size divided from N keys. All sensors are then deployed according to the covering algorithm DCA presented in [18]. According to Theorem 2 proved in section III, we set the covering degree to $2t$ here.

After deployment, the localization algorithm SPAWN presented in [21] is adopted to obtain sensors' positions. We use L_i:(X_i, Y_i) to denote the position of S_i, here X_i and Y_i are x coordinate and y coordinate of the sensor, respectively. Then, c pieces of package (S_i, L_i, U_i) are disseminated through multicast algorithm [22], in which seeds are distributed uniformly to each sensor with a probability of $c/\, N_a$, where U_i marks the group index in S_i, and N_a denotes the size of the network.

The parameter c affects the filtering efficiency and security performance. In the extreme situation that c is zero, the defending ability of LOCF is equal to SEF. For caching positions of other sensors, some more storage overhead is needed. For example, assume a key occupies 64bits and a position occupies 10bits, then we need 1KB to cache 100keys and 100positions. This is affordable in mainstream sensors equipping with 128KB ROM [19].

C. EVENT REPORTING

WSN is usually deployed to monitor some kind of physical signal and generate reports in some fixed period. When there is the need to report an event, several detecting sensors collaborate with each other in the following manner. First, the representative sensor SS calculates the position of the event L_e through three anchor sensors. Second, SS sends its detecting value to $t-1$ one hop neighbors, which are selected randomly. Third, each designated neighbor produces a MAC for the signal if the received value matching with the value of its own; other, event reporting fails. The produced MAC is then sent to SS as the endorsement of the event. Finally, SS synthesizes the information of t detecting sensors and produces a report as: $\{E, L_e; T_e; MAC_{i1}, MAC_{i2}, \ldots, MAC_{it}; L_{j1}, L_{j2}, \ldots, L_{jt}\}$, here $MAC_i = K_i(e)$, which is produced by key K_i, i_t and j_t denote index of the key and ID of the sensor, respectively.

D. INTERMEDIATE VERIFICATION

In the route of forwarding report R to sink, the intermediate sensor first executes regular verification, i.e., whether the report carried enough needed information: t different groups, t sensor IDs, t MACs and t positions. If any of the attached information is inadequate, the forwarding is abandoned. Then the sensor verifies whether all these positions included are rational according to the equation dis $|L_e, L_{jt}| \le r_s$, here r_s is the sensing radius, and dis $|L_e, L_{jt}|$ represents the distance between the stimulus and detecting sensor. If dis $|L_e, L_{jt}| > r_s$, the position L_{jt} is treated as irrational and the forwarding is also abandoned. Next, the sensor has to check the validity of positions and MACs in the report. It traversals all attached positions and cached ones, and executes consistency checking. Moreover, it also traversals all attached key indexes and cached ones, and carries out the verification as follows: produces a new MAC using its cached key, and see if the result matches the corresponding one attached. When the verification checks, the forwarding is abandoned and the report is dropped. There are also situations that the sensor cannot execute correctness checking of positions and MACs, i.e., the stored information is other than the attached information. In this case, the sensor forwards the report to next hop. If all verification passed, the report is also forwarding to next sensor. The pseudo-codes for intermediate verification is given in Algorithm 1.

Algorithm 1 Intermediate Verification in LOCF

```
//* on obtaining the data report $R$*//
Verify whether $t$ pieces of $\{i_k, M_{ik}\}$ blocks attached in $R$; abandon forwarding otherwise.
Verify whether the $t$ indexes $\{i_k, 1 \le k \le t\}$ come from different groups; abandon forwarding otherwise.
Verify whether $t$ pieces of $\{j_k, L_{jk}\}$ blocks attached in $R$; abandon forwarding otherwise.
For each attached $L_{jk}$, check equation dis $|L_e, L_{jk}| \le r_s$; abandon forwarding otherwise.
If cached a position $L$ for some detecting sensor, it check the consistency of the two positions; abandon forwarding otherwise.
If cached a key $K$ for some attached MAC, it check the consistency of the two MACs through re-producing and comparison; abandon forwarding otherwise.
Send $R$ to the next hop.
```

E. SINK AUTHENTICATION

The sink is the only node in this deployed network with strong capabilities on computation, transmission, storage and self-protection. Thus it is reasonable for sink to execute as the final guard for the filtering mechanism. Upon receiving the final report, the sink node can authenticate all attached information in the report. That is because it possesses all keys and locations of all nodes. At first, sink execute the ordinary check of the number of MACs and locations, and then also verify the correctness of each MAC through re-computing and comparison. It also has to check the correctness and legitimacy of all locations carried in the report. Only when all authentications passed, the report is accepted by the sink. Finally, the sink needs to respond to the message carried in the report according to some predefined rules, which is out of the scope of this paper. The pseudo-codes for sink authentication is shown in Algorithm 2.

V. PERFORMANCE EVALUATION

A. COMPROMISE ROBUSTNESS

LOCF first validates the correctness of all MACs in the report as existing mechanisms, and then judges the rationality of positions of all detecting sensors. Hence, arbitrary sensors captured cannot be abused to forge irrational reports. Considering the conspiratorial attack in Fig. 3, here we assume t is five and the attacker has captured five sensors $S_1, \ldots, S_5$ with different key groups. If a forged report R is like this: $\{e, L_e, T_e; M_{i1}, M_{i2}, \ldots, M_{i5}; L_{j1}, L_{j2}, \ldots, L_{j5}\}$, then R will

Algorithm 2 Sink Authentication

```
//* on obtaining the data report $R^*$//
1. Verify whether $t$ $\{i_k, M_{ik}; j_k, L_{jk}\}$ blocks attached in $R$; if not, reject the data report.
2. Verify whether the $t$ indexes $\{i_k, 1 \le k \le t\}$ come from different groups; if not, reject the data report.
3. Check each attached location according to equation dis $|L_e, L_{jk}| \le r_s$; reject the data report otherwise.
4. Check the correctness of each attached location according to the prior knowledge, reject the data report otherwise.
5. Authenticate the correctness of each attached MACs through re-producing and comparison; if outmatches occurs, reject the report.
6. Accept and respond to $R$.
```

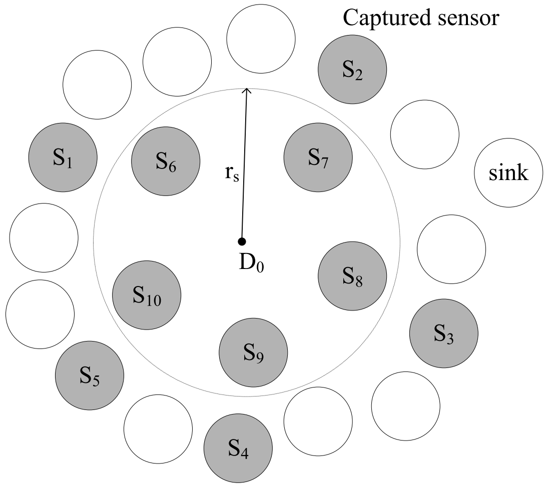


FIGURE 3. Resisting conspiratorial attack.

be judged as irrational by forwarding sensors quickly. For the interval between S_1 and S_3 is too bigger ($\geq 2r_s$), there is no way to construct dis $|L_e, L_1| \le r_s$ and dis $|L_e, L_3| \le r_s$ simultaneously.

To fake a report which can sneak the verification of LOCF successfully, the attacker should obtain t different groups within some πr_s^2 region. Only in this situation, can it forge t rational positions and t correct MACs.

Theorem 3: In LOCF, after compromising N_c sensors randomly ($N_c \ge t$) in the monitoring region with size D, the probability for the attacker getting at least t different groups within some πr_s^2 region is

$$p_L = \sum_{i=t}^{N_c} A(i,n)\frac{C(i,N_c)\cdot(\pi r_s^2/D)^i\cdot(1-\pi r_s^2/D)^{N_c-i}}{n^i} \quad (8)$$

Proof: Note the πr_s^2 region as D_0 in Fig. 3. Each sensor lies in D_0 with probability $p_0 = \pi r_s^2/D$. Then the probability of t sensors lying in D_0 can be computed as $p_u = C(t, N_c)p_0^t(1-p_0)^{Nc-t}$. And each of these sensors has probability $p_r = A(i,n)/n^t$ to hold a different group. Thus t different groups lying in D_0 with probability $p_r p_u$.

The probability that the attacker holding exactly $t+1, \ldots, n$ groups in D_0 can be calculated in the same way as holding t groups. As a result, the probability p_L can be calculated through accumulation.

The SEF scheme is a basic framework for most of previous filtering mechanisms, so it is reasonable to be taken as a reference on performance evaluation of LOCF. Fig. 4 compares the analytic and experimental curves p_S and p_L as the parameters t, D_0/D and n are set to 5, 1/4 and 20. Here p_S is the probability to ruin SEF according to Eq. 1. The experimental curves are averaged over 500 runs. From Fig. 4 we know, SEF can be ruined quite possibly only with little captured sensors. But a huge amount of captured sensors are necessary to ruin LOCF with a big probability. For example, the probabilities for SEF and LOCF to be ruined by 10 captured sensors are 0.928 and 0.032, respectively. Thus, both analysis and experiments show LOCF exhibits better robustness than SEF on sensor compromise.

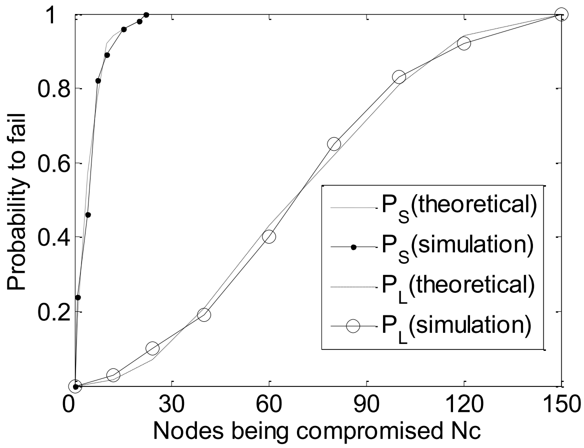


FIGURE 4. Analytic and experimental curves of p_S and p_L.

B. FILTERING ABILITY

In situation that totally N_c ($N_c < t$) sensor nodes within a πr_s^2 region are captured by the attacker, a forged *appearing*rational report should carry exactly t-N_c faked MACs and t-N_c faked positions. Thus an intermediate sensor node has probability to verify one of the faked MACs as $p_a = (k/m)$ $(t$-$N_c)/n = k(t$-$N_c)/N$.

And the probability that S_i possesses one of the faked locations is $p_b = 1-(1-c/N_c)^{t-Nc}$. Let P_1 be $P_a+P_b-P_aP_b$, then the portion of detected reports in h hops is computed as $p_h = 1-(1-p_1)^h$. As a result, the transmitting hops for a faked report is

$$H = \sum_{i=1}^{\infty} p_1 \cdot i(1-p_1)^{i-1} \quad (9)$$

C. ENERGY COST

The energy cost of LOCF mainly includes four aspects: (1)the communication cost on sensors deployment according to covering algorithm and position dissemination during initialization; (2)the communication cost among all detecting sensors on collaborated reports production; (3)the computation cost of verification; (4)the communication cost on transmitting reports.

As the authors in [13] pointed out, compared with the energy cost on transmitting reports, the cost of MAC calculation, transmitting short packages among detecting sensors on reports production and position dissemination are much smaller, thus they are omitted on the following analysis. Note the size of a pure report as I_e, the size of sensor ID as I_s, the size of position as I_p, and the size of MAC as I_m, respectively. Then, the size of reports in LOCF and SEF is $I_{e0} = I_e + I_p + t(I_s + I_p + I_m)$ and $I_{e1} = I_e + t(I_s + I_m)$, respectively. Obviously, compared with SEF and the other derivative en-route filtering mechanisms, the redundant information incurred in LOCF increases certain transmitting cost, but it is rational for bringing extra ability on defending conspiratorial attacks. Further, the proposed LOCF also reduces cost through the quick filtering of faked reports, which is going to be verified in the simulation results part.

VI. SIMULATION RESULTS

To further test the effectiveness of the proposed LOCF scheme, we construct an experimental platform using Python and C++ language. Due to limited space, we only evaluate the performance on $t - k$ cover effectiveness, compromise robustness, filtering capacity and energy cost, under the situation that $c = 25, 50$, $\theta = 0.8$, and $\varepsilon = 0.05$. The simulated environment is illustrated as below. In the given 50×50 m^2 region, 500 sensors are deployed: (1) uniformly; (2) according to DCA covering algorithm. A sink and a source node locate in two ends of the region. The data transfer time for each report is set at 10ms. For both SEF and LOCF, we use 20 groups each of which containing 15 keys. The averaged keys cached by each sensor is set at seven, and other settings of parameters are illustrated in Table 1. All results are mean value for 15 tests.

TABLE 1. Variables setting.

Variables	Setting
Delivery interval	2s
Quantity of packets	150
Transmitting radius	3m
Sensing radius	12m
Transmitting cost	6 mj
Reception cost	1.2 mj
Quantity of MACs attached in report (t)	5
Delivery interval	2s

We choose four representative sensors, S_1, S_2, S_3 and S_4 whose coordinates are (0, 50), (25, 50), (25, 25) and (35, 35) respectively, to compare the t−k cover effectiveness of two deployment strategies: random deployment and DCA covering algorithm. From Table 2 we can see that using random deployment, there are only small amounts of t−k covers of the selected four sensors, while with DCA covering algorithm, the number of t−k cover is much larger that random deployment do. For example, for the two boundary sensors, there are only three t−k covers in random deployment, compared with 43 t−k covers in DCA.

TABLE 2. Comparison of t−k cover.

Deploying model	Times of t-k cover
random deployment (S_1)	1
random deployment (S_2)	2
random deployment (S_3)	5
random deployment (S_4)	3
DCA (S_1)	21
DCA (S_2)	22
DCA (S_3)	25
DCA (S_4)	24

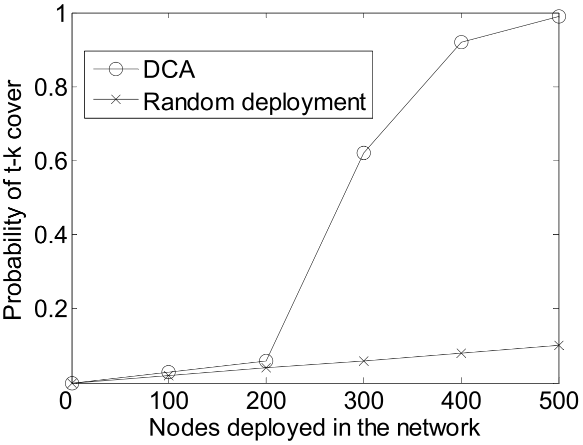


FIGURE 5. t−k cover changes with deployed sensors.

Fig. 5 shows the change of t−k cover according to the number of deployed sensors in the network. From Fig. 5 we get the following observations: (1) when deployed small amounts of sensors, both the performance on t−k cover of random deployment and DCA is weak. For example, with 200 deployed sensors, the t−k cover probabilities of them are only 6% and 4%, respectively. This is because small deploying density usually leads to bad covering effectiveness; (2) According to the increase of deployed sensors, t−k cover effectiveness of random deployment increases slowly, e.g., there is only an increment of 3% of the covering probability in random deployment from 200 deployed sensors to 400. While under the same situation, there is a big increment of 86% of t−k cover effectiveness in DCA. The reason is that some sparse regions (with less than t sensors with different groups covered) in the network formed through random deployment but not in DCA.

Fig. 6 plots how f changes according to N_c, here f is a filtering metric which means the additive reports abandoned during forwarding. From Fig. 6 we know: (1) LOCF performs better than SEF in compromise robustness. The number of captured sensors that SEF, and LOCF can resist is 12 and 130, respectively. (2) With N_c increasing, the filtering efficiency of SEF declines rapidly, while in LOCF it declines

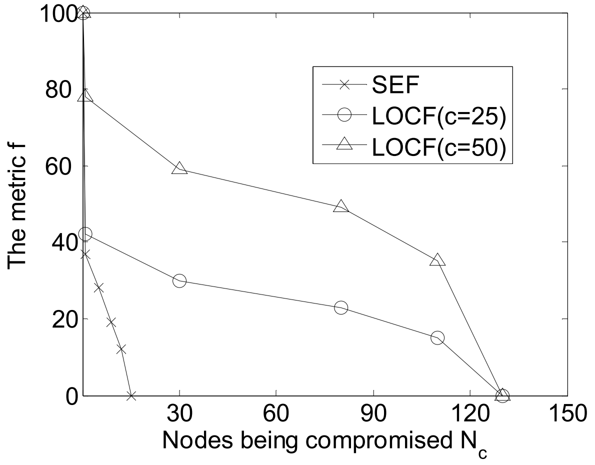


FIGURE 6. Filtering efficiency changes with N_c.

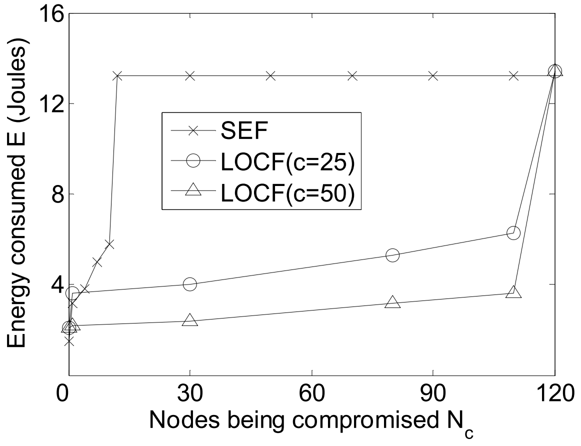


FIGURE 8. Energy cost as a function of captured sensors.

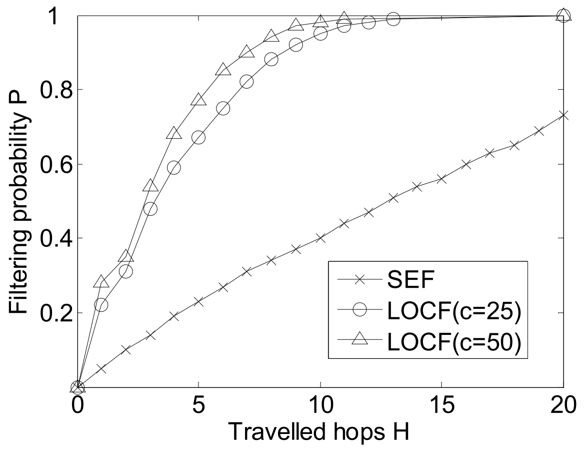


FIGURE 7. Proportion of abandoned reports changes with hops.

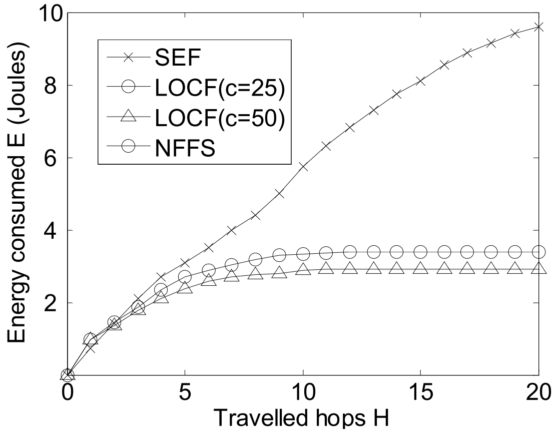


FIGURE 9. Energy cost E changes with H.

gracefully. The reason is that LOCF can limit the conspired destructiveness of compromised nodes through the bounding of sensors and locations. (3) The filtering efficiency in LOCF is much better than SEF with a fixed N_c. Take $N_c = 10$ as an example, f gets the value of 21 in SEF, 38 in LOCF ($c = 25$), and 75 in LOCF ($c = 50$), respectively. (4) The parameter c has a positive linear relationship with filtering efficiency.

Fig. 7 shows the proportion of abandoned reports changes with the forwarding hops H. From Fig. 7 we observe that as being forwarded further, the proportions of dropped reports are getting larger gradually both in SEF and LOCF. And within some fixed hops, LOCF exhibits better filtering performance than SEF. Take $H = 5$ for instance, the percentage of dropped faked reports is 24% in SEF, 66% in LOCF ($c = 25$), and 78% in LOCF ($c = 50$), respectively. This is because that LOCF can drop more faked reports than SEF through additional location rationality checking.

Fig. 8 illustrates how energy cost E changes with the number of compromised sensors N_c. From Fig. 8 we note that, LOCF has better energy cost performance than SEF in most situations. For example, with 30 compromised sensors, the energy cost is 13.4Joules in SEF, 3.9Joules in LOCF ($c = 25$), and 2.3Joules in LOCF ($c = 50$), respectively. The exception is in situation of four compromised sensors. The reason for this is that the filtering efficiency of LOCF is equal to SEF under $c = 25$ and N_c <4, but the size of packet in LOCF is larger. Combined with these two factors, LOCF performs a little weaker than SEF in this situation.

Fig. 9 plots the curves of energy cost along with transmitting hops. From Fig. 9 we get that, when data reports traveling more than two hops in the network, LOCF outperforms both SEF and NFFS in energy cost for the earlier filtering of faked reports. This is because SEF and NFFS only verify the attached MACs thus is weaker in filtering efficiency than LOCF which authenticate additionally the legitimacy of locations in the report.

VII. CONCLUSION

Faked reports injection is one of the critical threat in wireless sensor networks, which depletes the constrained energy of nodes and deceives users on decision making easily. Existing MAC verification based security designs cannot defend conspiratorial attacks by a group of captured sensors, and also fails to regulate sensors coverage required by t-based authentication. The novel solution in this paper is first to deploy sensors through deriving the optimal covering degree rather than the random deployment method adopted in almost all existing mechanisms, and then verify both the validity of attached MACs and positions of detecting sensors, and the rationality of positions. Analytic and experimental results demonstrate that LOCF is able to defend conspiratorial attacks efficiently and guarantee $t-$k cover with a high probability. As for next step, we are going to study the case of multiple sinks and seek a way to extend the results to protect data security in internet of things.

REFERENCES

[1] G. Mao, B. Fidan, and B. D. O. Anderson, "Wireless sensor network localization techniques," *Comput. Netw.*, vol. 51, no. 10, pp. 2529–2553, Jul. 2007.

[2] S. Jaggi, M. Langberg, S. Katti, T. Ho, D. Katabi, M. Medard, and M. Effros, "Resilient network coding in the presence of byzantine adversaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2596–2603, Jun. 2008.

[3] S.-L. Peng, S.-S. Li, X.-K. Liao, Y.-X. Peng, and N. Xiao, "Estimation of a population size in large-scale wireless sensor networks," *J. Comput. Sci. Technol.*, vol. 24, no. 5, pp. 987–997, Sep. 2009.

[4] H. Yang, F. Ye, Y. Yuan, S. Lu, and W. Arbaugh, "Toward resilient security in wireless sensor networks," presented at the MobiHoc, May 2005.

[5] K. Naresh, K. P. Pradeep, and K. S. Sathish, "An active en-route filtering scheme for information reporting in wireless sensor networks," *Int. J. Comput. Sci. Inf. Technol.*, vol. 2, no. 4, pp. 1812–1819, Feb. 2011.

[6] Z. Su, "Key management schemes and protocols for wireless sensor networks," *J. Softw.*, vol. 18, no. 5, pp. 1218–1231, May 2007.

[7] F. Ye, H. Luo, and L. Zhang, "Statistical en-route filtering of injected false data in sensor networks," presented at the INFOCOM, Mar. 2004.

[8] L. Yu and J. Z. Li, "Grouping-based resilient statistical en-route filtering for sensor networks," in *Proc. 28th Annu. Joint Conf. IEEE Comput. Commun. Soc.*, Apr. 2009, pp. 1782–1790.

[9] F. Yang, X. H. Zhou, and Q. Y. Zhang, "Multi-dimensional resilient statistical en-route filtering in wireless sensor networks," in *Advances in Grid and Pervasive Computing* (Lecture Notes in Computer Science). Berlin, Germany: Springer-Verlag, 2010, pp. 130–139.

[10] A. K. Bashir, S.-J. Lim, C. S. Hussain, and M.-S. Park, "Energy efficient in-network RFID data filtering scheme in wireless sensor networks," *Sensors*, vol. 11, no. 7, pp. 7004–7021, Jul. 2011.

[11] S. Dobrev, L. Narayanan, and J. Opatrny, "Optimal sensor networks for area monitoring using rotating and beam sensors," *Theory Comput. Syst.*, vol. 54, no. 4, pp. 622–639, Jun. 2014.

[12] F. Li and J. Wang, "A probabilistic voting-based filtering scheme in wireless sensor networks," presented at the IWCMC, Jul. 2006.

[13] M. N. Su and T. H. Cho, "Context-aware architecture for probabilistic voting-based filtering scheme in sensor networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 99, pp. 2751–2763, Oct. 2017.

[14] Y. Guan and X. Ge, "Distributed attack detection and secure estimation of networked cyber-physical systems against false data injection attacks and jamming attacks," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 4, no. 1, pp. 48–59, Mar. 2018.

[15] H. Yang and S. Lu, "Commutative cipher based en-route filtering in wireless sensor networks," presented at VTC, Sep. 2004.

[16] H. Wang and Q. Li, "PDF: A public-key based false data filtering scheme in sensor networks," presented at WASA, Aug. 2007.

[17] E. Ayday, F. Delgosha, and F. Fekri, "Location-aware security services for wireless sensor networks using network coding," presented at the 26th INFOCOM, May 2007.

[18] H. X. Wei and Q. Mao, "A dynamic covering algorithm of wireless sensor network based on CVT," presented at the 20th RTCSA, Aug. 2014.

[19] J.-M. Gil and Y.-H. Han, "A target coverage scheduling scheme based on genetic algorithms in directional sensor networks," *Sensors*, vol. 11, no. 2, pp. 1888–1906, Feb. 2011.

[20] B. A. Attea and S. M. Hameed, "A genetic algorithm for minimum set covering problem in reliable and efficient wireless sensor networks," *Iraqi J. Sci.*, vol. 55, no. 1, pp. 224–240, Jan. 2014.

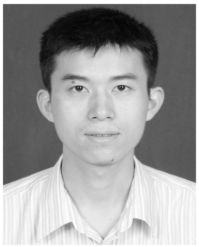
[21] W. Henk, L. Jaime, and Z. W. Moe, "Cooperative localization in wireless networks," *Proc. IEEE*, vol. 97, no. 2, pp. 427–450, Mar. 2009.

[22] P. Bose, B. Morin, I. Stojmenovic, and J. Urrutia, "Routing with guaranteed delivery in ad hoc wireless networks," *Wireless Netw.*, vol. 7, no. 6, pp. 609–616, Nov. 2001.

ZHIXIONG LIU received the B.S. degree in computer science and technology from Henan Normal University, Henan, China, in 2004, and the M.S. and Ph.D. degrees in computer science and technology from Central South University, Changsha, China, in 2012.

From 2012 to 2017, he was a Lecturer with the School of Computer Engineering and Applied Mathematics, Changsha University, where he has been an Assistant Professor with the School of Computer Engineering and Applied Mathematics, since 2017. His research interests include wireless sensor networks and the Internet of Things.

Dr. Liu is a member of China Computer Federation (CCF).



HUI YE received the B.S. degree in computer science and technology from Changsha University, Hunan, China, in 2000, the M.S. degree from Hunan University, Changsha, China, in 2005, and the Ph.D. degree in computer science and technology from Central South University, Changsha, in 2010.

From 2011 to 2016, he was a Lecturer with the School of Computer Engineering and Applied Mathematics, Changsha University, where he has been an Assistant Professor with the School of Computer Engineering and Applied Mathematics, since 2017. His research interests include wireless sensor networks, algorithm optimization, and the Internet of Things.

FANGMIN LI received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.S. degree from the National University of Defense Technology, Changsha, China, in 1997, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2001, all in computer science.

He is currently a Professor with the School of Computer Engineering and Applied Mathematics, Changsha University. He has authored over 150 academic articles in wireless networks, and also holds ten Chinese patents. His current research interests include wireless communications and networks security, computer systems and architectures, and embedded systems.

Dr. Li is a Senior Member of China Computer Federation (CCF) and a Committee Member of the Technical Committee on Sensor Network, CCF.