Journal of Information & Computational Science 10:12 (2013) 3781–3789 August 10, 2013
Available at http://www.joics.com

Spatial Correlation Based Secure Data Aggregation Scheme in Wireless Sensor Networks ⋆

Guorui Li [a,*], Ying Wang [b]

[a] *School of Computer and Communication Engineering, Northeastern University at Qinhuangdao, Qinhuangdao 066000, China*

[b] *Department of Information Engineering, Qinhuangdao Institute of Technology, Qinhuangdao 066004, China*

Abstract

In wireless sensor networks, the communication cost is often several orders of magnitude higher than that of computation. Data aggregation is an essential technique to reduce the communication cost and prolong network lifetime. As wireless sensor networks are usually deployed in unattended or even hostile circumstances to collect sensitive information, sensor nodes are prone to node compromise attacks. Therefore, security should be considered for data aggregation schemes where high data reliability and high data accuracy are both required. We propose a spatial correlation based secure data aggregation scheme which combines abnormal behaviors detection with data aggregation in this paper. It can detect and exclude the exceptional data values within the cluster in order to reduce the influence of malicious attacks and abnormal transmission errors. We show through experiments that our proposed scheme can provide high detection accuracy ratio and low false alarm ratio.

Keywords: Wireless Sensor Networks; Data Aggregation; Security; Spatial Correlation

1 Introduction

Wireless sensor networks consist of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants [1]. These sensor nodes can coordinate among themselves to form a communication network in an ad hoc multi-hop way. The typical application fields of wireless sensor networks include industrial process control, security and surveillance, environmental sensing, and structural health monitoring, etc [2].

⋆Project supported by the Fundamental Research Funds for the Central Universities of China (Grant No. N100323001 and N120423005), the Research Fund for the Doctoral Program of Higher Education of China (Grant No. 20120042120009), the Natural Science Foundation of Hebei Province, China (Grant No. F2012501014), the Science and Technology Project of Liaoning Province, China (Grant No. 2010302005), the Scientific Research Foundation of the Higher Education Institutions of Hebei Province, China (Grant No. Z2010215) and the Science and Technology Research and Development Project of Qinhuangdao (Grant No. 2012021A029).

*Corresponding author.
Email address: lgr@mail.neuq.edu.cn (Guorui Li).

DOI: 10.12733/jics20101994

In wireless sensor networks, the communication cost is often several orders of magnitude higher than that of computation. Generally, the data transmission can account for up to 70% of the power consumed in typical sensor nodes [3]. In order to save the limited battery power of sensor node, data aggregation technique is often adopted as an effective way. The inherent redundancy in the raw data collected from the sensor nodes can often be eliminated by in-network data aggregation. Moreover, such operation is also useful for extracting application specific information from the raw data. Therefore, it is critical for the network to support data aggregation scheme in order to conserve energy for a longer network lifetime [4].

As wireless sensor networks are usually deployed in unattended or even hostile circumstances to collect sensitive information, sensor nodes are prone to node compromise attacks and security issues such as data confidentiality and integrity are extremely important. Therefore, wireless sensor network schemes, e.g., data aggregation scheme, must be designed with security in mind. There have been a few researches in the recent past on secure data aggregation in wireless sensor networks. A comprehensive survey on secure data aggregation schemes of wireless sensor networks was presented in [5] . We will briefly review some representative schemes in Section 2. However, there is a strong conflict between security and data aggregation scheme. We should design data aggregation scheme and security scheme together so that data aggregation can be achieved without sacrificing security.

We proposed a spatial correlation based secure data aggregation scheme in this paper which combines abnormal behaviors detection with data aggregation. The proposed scheme can detect and exclude the exceptional data value within the cluster in order to reduce the influence of malicious attacks and abnormal transmission errors. It includes spatial correlation based clustering algorithm which partitions the sensor network into a number of data correlated clusters and secure data aggregation algorithm which using Mahalanobis distance and OGK estimation to detect the abnormal behaviors. We show through experiments that our proposed scheme can provide quite high detection accuracy ratio and low false alarm ratio. Therefore, the final aggregated results will be more reliable and dependable without the influence of abnormal behaviors.

This paper is organized as follows. In the next section, we review some related works. In Section 3, we present our spatial correlation based secure data aggregation scheme. In Section 4, we describe our experiment settings and evaluation results. Finally, we draw our conclusion of this paper in Section 5.

2 Related Work

In general, the proposed secure data aggregation schemes can be classified into two categories based on the requirement of decrypting sensor data at data aggregators: secure data aggregation scheme using plain sensor data and secure data aggregation scheme using encrypted sensor data [5].

2.1 Secure Data Aggregation Scheme Using Plain Sensor Data

In this kind of secure data aggregation scheme, data aggregator should decrypt every message it received, aggregate the messages according to the corresponding aggregation function, and encrypt the aggregated result before forwarding to the next node. It can provide hop-by-hop

data confidentiality but result in latency for the reason of encryption/decryption operations.

Secure Information Aggregation (SIA) scheme is proposed in [6] by constructing efficient random sampling mechanisms and interactive proofs. The authors argue that it is possible for the user to verify the aggregated value is a good approximation of the true value even when the aggregator and a fraction of the sensors are compromised. In Secure Data Aggregation and Verification (SecureDAV) scheme, data aggregator can combine the partial signatures form member nodes to form a full signature of the aggregated data and send it to the sink [7]. However, the communication overhead on data validation is very high and only the average aggregation function is supported. In Secure Aggregation Tree (SAT) scheme, a weighted voting method is employed to decide whether the data aggregator is misbehaved [8]. If it is a misbehaving node, then the secure aggregation tree is rebuilt to exclude it from the aggregation tree. In Secure hop-by-hop Data Aggregation Protocol (SDAP), the topology tree is dynamically partitioned into multiple logical sub-trees of similar sizes using a probabilistic approach [9]. Therefore, fewer nodes are located under a high level sensor node in the logical sub-tree resulting in reduced potential security threat by high level node compromising. In Secure and rELiable Data Aggregation (SELDA) scheme, sensor nodes observe behaviors of their neighbors to develop trust levels and exchange their trust levels with neighboring nodes to from a web of trust which allows them to determine secure and reliable paths to data aggregators [10]. In Data Aggregation and Authentication (DAA) scheme, the monitoring nodes of data aggregator also perform data aggregation and compute the corresponding message authentication codes to provide authentication and verification function [11]. And the sensors within two consecutive data aggregators also verify the data integrity on the encrypted data to support confidential data transmission.

2.2 Secure Data Aggregation Scheme Using Encrypted Sensor Data

In this kind of secure data aggregation scheme, data aggregator does not require to decrypt the sensed value when performing data aggregation. Some of these schemes use symmetric cryptography, others employ asymmetric cryptography. They can provide end-to-end data confidentiality and result in less latency. However, those schemes are applicable to only a subset of common aggregation functions, such as sum and average.

In Concealed Data Aggregation (CDA) scheme, data aggregators perform data aggregation functions that are applied to the encrypted data using privacy homomorphism technique which is an encryption transformation that allows direct computation on encrypted data [12]. In Concealed Data Aggregation using Privacy homomorphism (CDAP) scheme, a set of resource-rich sensor nodes, called aggregator nodes, are employed to aggregate the encrypted data using asymmetric key based privacy homomorphic cryptography [13]. The n-layers data aggregation (n-LDA) scheme proposed in [14] ensures that an attacker cannot get access to any aggregated data from the network when a certain number of nodes are compromised. When more nodes are compromised, only the aggregated values received by the captured nodes can be acquired by the attacker.

3 Secure Data Aggregation Scheme

The following notations in Table 1 are used throughout the paper.

Table 1: Notations

Notation	Meaning
d_i	Sampled data value of node i
n	The dimension of sampled data value
d_{ij}	The distance between d_i and d_j
$N(i)$	The neighbor sensors of node i
T	The election time
δ	The threshold of dissimilarity among sampled data values within each cluster
$\hat{\mu}$	Mean estimator
$\hat{\Sigma}$	Covariance estimator
$d(x_i)$	Mahalanobis distance
$X_n(\alpha)$	Upper (100α)th percentile of chi-square distribution with n degrees of freedom

3.1 Spatial Correlated Weigh

In sensor networks, the wireless sensor node's sampled data value at a certain sample time can be described by an n dimensional vector. If the sampled values of node i and j are denoted as $d_i = (x_1, x_2, \cdots, x_n)$ and $d_j = (y_1, y_2, \cdots, y_n)$ respectively, the distance between those two measurements can be computed as

$$d_{ij} = \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2 + \cdots + |x_n - y_n|^2} \tag{1}$$

Then, the expectation of d_{ij} is

$$E(d_{ij}) = \frac{1}{|N(i)|}\Sigma_{j\in N(i)} d_{ij} \tag{2}$$

The deviation of d_{ij} is

$$D(d_{ij}) = \frac{1}{|N(i)|}\Sigma_{j\in N(i)}(d_{ij} - E(d_{ij}))^2 \tag{3}$$

According to [15], the spatial correlated weight w_i of node i can be computed as

$$w_i = \frac{(\Sigma_{j\in N(i)}|d_{ij} - E(d_{ij})|)^2}{|N(i)|^2 D(d_{ij})} \tag{4}$$

According to the Cauchy-Schwarz inequality,

$$(\Sigma_{j\in N(i)}|d_{ij} - E(d_{ij})|)^2 \leqslant |N(i)|\Sigma_{j\in N(i)}(d_{ij} - E(d_{ij}))^2 \tag{5}$$

We can assert that $0 \leqslant w_i \leqslant 1$. Generally speaking, the spatial correlated weight w_i indicates the average spatial measurement deviation among node i and its h hop neighbors. The larger the value of w_i, the higher the spatial correlation of node i with its h hop neighbors. Therefore, it should be selected as an aggregator preferentially.

3.2 Relative Energy Level

If the residual energy at sensor node i is denoted as e_i, the relative energy level re_i can be computed by comparing e_i with its h hop neighbors' average residual energy. That is, the relative energy level of node i can be computed as

$$re_i = \frac{e_i + \Sigma_{i \in N(i)} e_i}{e_i \times (|N(i)| + 1)} \tag{6}$$

The smaller the value of re_i, the more energy of node i remains than its h hop neighbors. Therefore, it should be selected as an aggregator primarily.

3.3 Spatial Correlation Based Secure Data Aggregation Scheme

The spatial correlation based secure data aggregation scheme includes spatial correlation based clustering algorithm and secure data aggregation algorithm. The spatial correlation based clustering algorithm is shown in Table 2.

Table 2: Spatial correlation based clustering algorithm

```
Broadcast sensor's sampled data value and residual energy
Compute w_i and re_i
Wait (re_i / w_i)T time
{
  If receive clustering message and has not joined any cluster
  {
    Compute d_ij
    If (d_ij ⩽ δ/2)
      Join cluster and forward clustering message
  }
}
If (re_i / w_i)T time is up
{
  Create a cluster and mark itself as cluster head
  Broadcast clustering message
}
```

In spatial correlation based clustering algorithm, each sensor node i broadcasts its sampled data value d_i and residual energy e_i to its h hop neighbors. It then computes spatial correlated weight w_i and relative energy level re_i according to Eq. (4) and Eq. (6), respectively. Sensor node i waits $(\frac{re_i}{w_i})T$ time to create a new cluster. If it receives a clustering message from its neighbor within this time period and has not joined any cluster, it will compute the distance d_{ij} according to Eq. (1). If $(d_{ij} \leqslant \frac{\delta}{2})$, the Triangle Inequality ensures the metric distance between any two sensors of the same cluster is at most δ. Therefore, we can ensure the sensor nodes of the same cluster have the similar observation. If the sensor has not joined any cluster after $(\frac{re_i}{w_i})T$ time, it will create a new cluster and mark itself as cluster head and broadcast clustering message to its h hop neighbors.

According to the spatial correlation based clustering algorithm, the resulting clusters are comprised of sensors that are spatially close to each other and have similar observations. Then, the cluster head executes the secure data aggregation algorithm which is shown in Table 3 to exclude the abnormal sensed data caused either by malicious attacks or by transmission errors.

Table 3: Secure data aggregation algorithm

Collect sensed value from cluster members
Calculate mean estimator $\hat{\mu}$ and covariance estimator $\hat{\Sigma}$ using OGK estimation
Calculate Mahalanobis distance $d(x_i) = ((d_i - \hat{\mu})^T \hat{\Sigma} (d_i - \hat{\mu}))^{\frac{1}{2}}$
If $(d(x_i) > X_n(\alpha))$
{
Identify sensor node i's sensed value as abnormal
Exclude sensor node i's sensed value from collected sensed value set
}
Aggregate collected sensed value using required aggregation function(s)

The cluster head collects sensed data value from cluster members and calculates mean estimator $\hat{\mu}$ and covariance estimator $\hat{\Sigma}$ according to OGK (Orthogonalized Gnanadesikan Kettenring) estimation [16]. It then calculates Mahalanobis distance $d(x_i)$ and compares $d(x_i)$ with $X_n(\alpha)$, which is the upper (100α)th percentile of chi-square distribution with n degrees of freedom.

Note that, d_i is distribute as $N_n(\mu, \Sigma)$, following a multivariate normal distribution with mean μ and variance-covariance matrix Σ. The Mahalanobis squared distance $(d(x_i))^2$ is distributed as chi-square distribution X_n^2 with n degree of freedom. Therefore, the probability that d_i satisfies $(d(x_i))^2 > X_n^2(\alpha)$ is α, where α is upper (100α)th percentile. Sensed data value d_i will be identified as abnormal and excluded from collected sensed value set if $d(x_i) > X_n(\alpha)$ or $(d(x_i))^2 > X_n^2(\alpha)$. After filtering abnormal sensed data value, cluster head will continue to aggregate collected sensed value using required aggregation function(s).

The reason why we decide to choose Mahalanobis distance measurement is because it includes the inter-attribute dependencies so we can compare the attribute combination and get more precise results. We choose OGK estimation is because it ensures a high breakdown point with some missing data and can compute quickly with a lower computational cost.

4 Experiments

We used the real sensed data collected from 54 Mica2Dot sensors deployed in the Intel Berkeley Research Lab between 28 February and 5 April 2004 to demonstrate the performance of our spatial correlation based secure data aggregation scheme. The collected data include humidity, temperature, light and voltage values along with timestamp information collected once every 31 seconds. We randomly add some noise following the normal distribution to the tested data in order to simulate the abnormal behaviors. The experiments were carried on using R with robustbase and rrcov packages.

Fig. 1 shows an instance of the spatial correlation based clustering algorithm in the Intel Berkeley Research Lab. We partitioned the sensor network into seven clusters and marked each

cluster head with a red circle. The corresponding spatial correlated weight of the instance shown in Fig. 1 is presented in Table 4, where the space distance threshold is set to 10 meters. We can see that all spatial correlated weights of the elected cluster heads are above 70%.

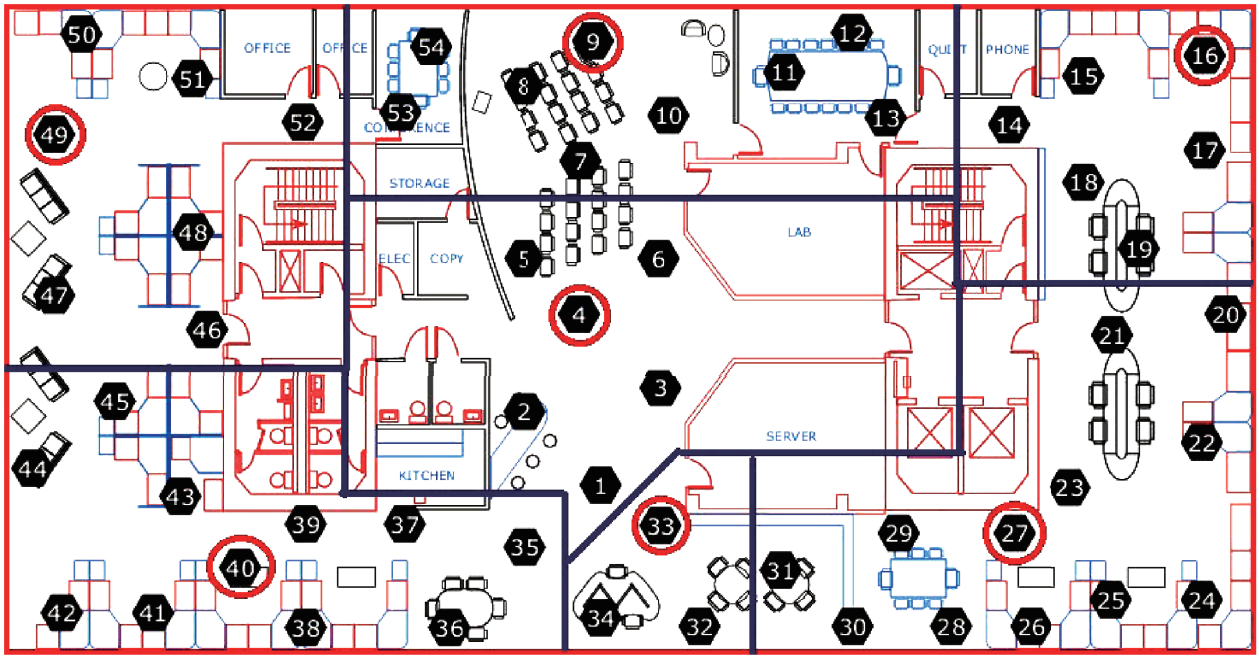

Fig. 1: An instance of the spatial correlation based clustering algorithm

Table 4: An instance of spatial correlated weight

Cluster ID	Cluster head	w_i
1	27	0.9426
2	49	0.8794
3	40	0.8577
4	9	0.7991
5	4	0.7543
6	16	0.7501
7	33	0.7046

The performance of our proposed secure data aggregation scheme is evaluated by the following two indicators, detection accuracy ratio and false alarm ratio. We calculated the detection accuracy ratio and false alarm ratio as the the percentage of abnormal behaviors which can be detected successfully and that of the normal behaviors which were regarded as the abnormal behaviors falsely.

Fig. 2 shows the detection accuracy ratios of the spatial correlation based secure data aggregation scheme with abnormal deviation 15% and 20%, respectively. We can see that the detection accuracy ratio of the proposed scheme remains very high. And the detection accuracy ratio increases with the increase of the abnormal deviation.

Fig. 3 shows the false alarm ratio of the spatial correlation based secure data aggregation scheme with abnormal deviation 15% and 20%, respectively. We can see that the false alarm

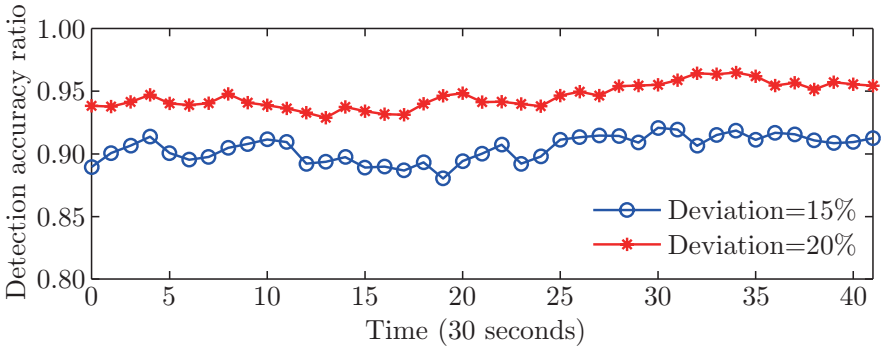

Fig. 2: The detection accuracy ratio of the spatial correlation based secure data aggregation scheme

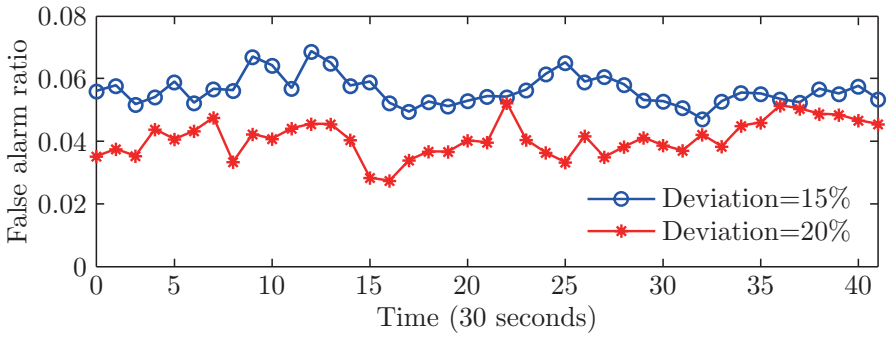

Fig. 3: The false alarm ratio of the spatial correlation based secure data aggregation scheme

ratio remains very low. And the false alarm ratio decreases with the increase of the abnormal deviation. According to Fig. 2 and Fig. 3, we can conclude that the larger the abnormal deviation is, the better the detection performance that our scheme can achieve.

5 Conclusion

In this paper, we proposed the spatial correlation based secure data aggregation scheme for wireless sensor networks with the goal of detecting and excluding exceptional data values within the network. The whole sensor network is partitioned into several data correlated clusters according to the measurement of spatial correlated weight and relative energy level. Mahalanobis distance and OGK estimation are employed to identify exceptional behaviors. We show through experiments that our proposed scheme can achieve a low false alarm rate and a high detection accuracy rate. Furthermore, the detection performance of the scheme rises with the increase of the abnormal deviation.

Acknowledgement

The authors would like to thank the anonymous reviews for their constructive comments and suggestions on improving the presentation of this work.

References

[1] I. Akyldiz, W. Su, Y. Sankarasubramanian, E. Cayirci, A survey of sensor networks, IEEE Communications Magazine, 40 (2002), 102-114

[2] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Computer Networks, 52 (2008), 2292-2330

[3] H. Cam, S. Ozdemir, P. Nair, D. Muthuavinashiappan, H. Sanli, Energy-efficient secure pattern based data aggregation for wireless sensor networks, Computer Communications, 29 (2006), 446-455

[4] S. Lee, S. Kim, D. Ko, S. Kim, S. An, Prediction based mobile data aggregation in wireless sensor network, in: Proc. Advances in Grid and Pervasive Computing'09, 2009, 328-339

[5] S. Ozdemir, Y. Xiao, Secure data aggregation in wireless sensor networks: A comprehensive overview, Computer Networks, 53 (2009), 2022-2037

[6] B. Przydatek, D. Song, A. Perrig, SIA: Secure information aggregation in sensor networks, in: Proc. SenSys'03, 2003, 255-265

[7] A. Mahimkar, T. Rappaport, SecureDAV: A secure data aggregation and verification protocol for wireless sensor networks, in: Proc. Globecom'04, 2004, 2175-2179

[8] K. Wu, D. Dreef, B. Sun, Y. Xiao, Secure data aggregation without persistent cryptographic operations in wireless sensor networks, Ad Hoc Networks, 5 (2007), 100-111

[9] Y. Yang, X. Wang, S. Zhu, G. Cao, SDAP: A secure hop-by-hop data aggregation protocol for sensor networks, ACM Transactions on Information and System Security, 11 (2008), 18-43

[10] S. Ozdemir, Secure and reliable data aggregation for wireless sensor networks, LNCS, 4836 (2007), 102-109

[11] Y. Xiao, Security in Distributed Grid Mobile and Pervasive Computing, CRC Press, New York, 2007

[12] D. Westhoff, J. Girao, M. Acharya, Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption key distribution and routing adaptation, IEEE Transactions on Mobile Computing, 5 (2006), 1417-1431

[13] S. Ozdemir, Concealed data aggregation in heterogeneous sensor networks using privacy homomorphism, in: Proc. Pervasive Services'07, 2007, 165-168

[14] I. Rodhea, C. Rohner, N-LDA: n-layers data aggregation in sensor networks, in: Proc. ICDCS'08, 2008, 400-405

[15] Y. Ma, Y. Guo, X. Tian, M. Ghanem, Distributed clustering-based aggregation algorithm for spatial correlated sensor networks, IEEE Sesnsors Journal, 11 (2011), 641-648

[16] G. Li, J. He, Y. Fu, Group-based intrusion detection system in wireless sensor networks, Computer Communications, 31 (2008), 4324-4332