An Efficient Concealed Data Aggregation Scheme for Wireless Sensor Networks

Gwoboa Horng*, Chien-Lung Wang and Tzung-Her Chen

***Abstract*—Soon after wireless sensor networks (WSNs) have attracted much attention both in industry and academia, maintaining the security of WSNs, especially end-to-end confidentiality, becomes a challenging problem. A sensor device has the limited computation capability, battery power, less memory size, and unreliable communication protocols. In order to save the overall energy resources and maintain the security of WSN, we need to reduce the amount of encrypted data transmitted. One approach is to consolidate the encrypted data along the routing path. This is called concealed data aggregation (CDA). In this paper, a novel end-to-end CDA scheme based on the concept of secret sharing is proposed to achieve simultaneously the goals of saving power and securely sending the concealed data.**

***Index Terms*—Concealed data aggregation (CDA), end-to-end encryption, secret sharing, security, wireless sensor network.**

I. Introduction

With wireless sensor networks (WSNs) being in a growing area both in industry and academia, the demand of keeping the sensed data secret from malicious outsiders will "grow from a ripple to a wave". Over the past decades, there have been more and more investigations engaged in WSNs. In other words, WSNs are quickly gaining popularity due to the fact that they are potentially low cost solutions to a variety of real-world challenges [1]. Wireless sensor networks are widely used in a variety of applications, including environment monitors (such as seismaesthesia, barometric pressure, temperature and humidity) as well as other ecological distribution monitors, especially, used in hostile environments (such as military sensing, tracking).

A wireless sensor network usually consists of a huge number of tiny autonomous devices called sensor nodes. A typical sensor node is equipped with Mhz processors rather than GHz processors, limited memory size, short-range radio communication, and powered by battery/solar energy. For representative example, the MICA2 [6] which is designed by Berkely, and its size is a several cubic inch. A MICA2 is composed of an 8 MHz processor, 128 kb of instruction memory, 4 kb of RAM for data, 512 kb of flash memory, 19.2 kbps bandwidth, and the communication range is 10-20 meter. In practice, a MICA2 with full energy can run about 2 weeks in work model, and almost 1 year in sleep model. In a word, sensor nodes have severe resource constraints due to their lack of powerful computing capability, data storage and energy. All of these represent major obstacles to the implementation of traditional computer security methods in WSNs.

One of the most notable characteristics of WSNs is that the sensor nodes collect the monitored data from the outside environment and then deliver them to a central point, hereafter simply called a sink node which is assumed to locate in a secure place. Since the number of sensor nodes may be up to ten thousands, the sensing data is huge. As pointed out in [13], for a sensor node, transmitting one bit consumes the same amount of energy as executing 50 to 150 instructions. Since the tiny device is limited in power, without an efficient scheme to process data, the energy will be exhausted quickly. Therefore, reducing the energy consumption is one of the most important issues in WSNs.

For reducing the energy consumption and increasing the WSN's overall lifetime, some studies focus on reducing the energy consumption by aggregating the sensed data [3]. Other studies go one step further, taking security into consideration, to aggregate concealed data [4, 8, 14]. These schemes allow for the end-to-end encryption between sensor nodes and a sink node and enable aggregators to apply aggregation function over ciphertexts directly.

The main advantages of the concealed data aggregation (CDA) lie in reducing the package size by aggregating the sensed data and eliminating the need of decrypting sensitive data and encrypting again after aggregation, the so-called hop-to-hop encryption. In a word, for avoiding the battery power being exhausted quickly, apply CDA to aggregate the encrypted sensing data can, on one hand, reduce large communication cost between sensor nodes and a sink node and, on the other hand, protect the sensitive data from revealing out of the aggregator nodes.

To enable concealed data aggregation, the nodes in a WSN are divided into three classes, namely, the sensor nodes S_1, S_2,…

Manuscript received September 27, 2007. This research was supported by the National Science Council of the Republic of China under contract NSC- 95-2221- E- 005- 080 and NSC- 96- 2628- E- 005- 076- MY3.

Gwoboa Horng is with the Department of Computer Science, National Chung-Hsing University, 250 Kuo-Kuang Road, Taichung 402, Taiwan, ROC. (corresponding author to provide phone: 886-042284-0497#924; e-mail: gbhorng@cs.nchu.edu.tw).

Chien-Lung Wang is with the Department of Computer Science, National Chung-Hsing University, 250 Kuo-Kuang Road, Taichung 402, Taiwan, ROC. (e-mail: phd9004@cs.nchu.edu.tw).

Tzung-Her Chen is with the Department of Computer Science and Information Engineering, National Chiayi University Chia-Yi City, Taiwan 60004, ROC. thchen@mail.ncyu.edu.tw. (e-mail: thchen@mail.ncyu.edu.tw).

, S_l, the aggregation nodes A_1, A_2,… , A_m, and the sink node R. The sensor node S_i encrypts its sensed data m_i' resulting in where $m_i' = E_{key}(S_i)$ before transmitting data to an aggregation node A_j. Then, the aggregation node will consolidate the encrypted data it received from the sensor nodes with a suitable function f. Let $y_j' = f(m_1', m_2', ..., m_l')$. The aggregation node A_j delivers the y_j' to the sink node R. Finally, R will compute $y = D_{key}(y')$. The process is shown in Fig.1.

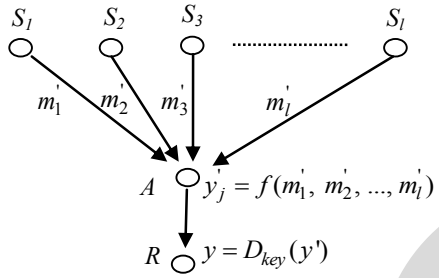

Fig.1. Concealed data aggregation

Although data aggregation can reduce the communication cost significantly, unfortunately it makes security more difficult to achieve. For instance, data aggregation does interfere with data encryption. Straightforwardly, the sensed data cannot be encrypted using a unique key shared between each sensor node and the aggregator node because the aggregator node should decrypt the data before aggregation. It's absolutely not a feasible way to risk sharing an identical key among sensor nodes and aggregator node. Otherwise, an attacker who has compromised a sensor node to obtain the key will have full control to the entire network.

The rest of the paper is organized as follows. In section 2, some related studies are addressed. The proposed scheme is described in section 3. Its security analysis and some discussions are addressed in section 4. Finally, section 5 concludes the whole paper.

II. RELATED WORK

In order to enable end-to-end encryption in WSNs, the homomorphic characteristics of a privacy homomorphism (PH) [7] is usually adopted to guarantee the feasibility and security for concealed data aggregation schemes. It's natural to adopt public-key-based PH algorithm, for example RSA, into CDA schemes since sensor nodes only need to store the non-sensitive public key. Unfortunately, public-key-based PH schemes require expensive computations and long keys, implying large messages, which would quickly deplete the battery of tiny sensor devices and thus do not suit the WSN scenario.

In WSNs authentication is divided into authentication between sensor nodes [16] and sensing data authentication [9, 12]. In this paper, we focus on sensing data aggregation which relates to reduce the power consumption, and the data authentication is not our focus. We will assume that data authentication is achieved by [12]. There are many papers on secure aggregation of data [2, 4, 5, 8, 9, 10, 11, 14, 15]. The followings are two most recent papers on CDA.

A. Girao et al.

In 2006, Girao et al. [15] presented a CDA scheme based on the symmetric additive PH scheme proposed by Domingo-Ferrer. Except efficiency, there are some drawbacks in [8] and [15]. But the scheme does not consider the problem of the non-response IDs. The scheme is described as follows.

Initial phase:

(1) One of the public parameters is a large integer g. It is pointed out that g should have many small divisors and at the same time there should be many integers less than g that can be inverted modulo g.

(2) The secret key $k=(r, g')$. The value $r \in Z_g$ is chosen such that, r^{-1} mod g exists and $\log_{g'} g$ is an integer with small g'.

At sensor node, s_i:

(1) Compute $m_i' = E_{key}(S_i)$ and transmit m_i' to the A.

At aggregation node, A_i:

(1) Aggregate all of m_i' into $y' = (\sum_{i=1}^{n} m_i')$ then delivers to R.

At sink node, R:

(1) Compute $y = D_{key}(y') = \sum_{y=1}^{d} m_{i.j} \bmod g'$.

Firstly, the aggregation is done using a key that is applied on each node in the network. Secondly, Girao et al.'s scheme inherits the disadvantage of size grow from Domingo-Ferrer's PH scheme. From the aspects of security and energy consumption, Girao et al.'s scheme is not a good candidate.

B. Castelluccia et al.

In [4] Castelluccia et al. proposed a symmetric CDA scheme based on key stream. However, one of disadvantage in this scheme is large consumption while the numbers of these problematic nodes are large. The scheme is shown as follows:

Initial phase:

(1) Represent message m as integer $m \in [0, M\text{-}1]$ where M is large integer.

(2) Let k be a randomly generated keystream, where $k \in [0, M\text{-}1]$.

At sensor node, s_i:

(1) Compute $m'=E(m, k, M)= m+ k \ (mod\ M)$ and transmit m' to the A.

At aggregation node, A_i:

(1) Aggregate all of m' into $y'=f(\sum_{i=1}^{n} m')$ then delivers to R.

At sink node, R:

(1) Compute $D(m', k, M)=m'-k \ (mod\ M)$ to get m.

Addition of ciphertexts:

(1) Let $m_1'=E(m_1, k_1, M)$ and $m_2'=(m_2, k_2, M)$.

(2) For $k=k_1+k_2$, $D(m_1'+m_2', k, M)= m_1+ m_2$.

Note that the scheme applied a unique key on each sensor node and preserves the small ciphertext size. In such a way, it is suitable for the application in WSNs. However, the sink node needs to know the ID list of the nodes that contribute to the

received aggregated data, i.e. so-called ID-problem [12].

III. The Proposed Scheme

In our work, we assume a fixed base station that can establish secrets with the ad hoc wireless nodes before deployment, so we do not address key management issues further.

A. Assumption

(1) The sink node is powerful and can broadcast messages to all nodes directly. Sensor devices are low power and can only communicate with nearby nodes, such like communicate with aggregators or nearby sensor nodes.
(2) The sensor nodes are deployed on the target field with uniform distribution and collect information to transmit to sink node. Then the aggregators are located on the center of sensor nodes, than these aggregators will collect and route packages by self-organization to sink node. In a word, an aggregator performs data collection and package forwarding.
(3) The network is spread out enough so there are likely to be many hops between a typical node and the sink node. The network is dense enough so that there are usually several nodes within one-hop distance of any particular node. And the routing paths are known in aggregators.
(4) Another important assumption is a secure sink node. An attacker can compromise any nodes in a WSN, except the sink node. And the shared secret will embed into sensor nodes and aggregators previously before deploying. In a word, the network environment is static and in which nodes are not mobile.

B. Notation

Item: *Description*
R: the sink node
A_i: the i-th aggregation node
S_i: the i-th sensor node
n: the number of nodes delivering sensed data
m_i: the monitored information from i-th sensor node
ID_i: the identification of i-th sensor node
k_i: the secret key of i-th sensor node, which is shared with R
d_i: the information encrypted by i-th sensor node

C. The Concealed Data Aggregation Based on Secret Sharing

The proposed secure CDA scheme, consisting of four phases: initialization, data concealment, concealed data aggregation and concealed data disclosure, is shown as follows.

Initialization phase:

Initially, the sink node R should do the following operations.

(1) Define a random polynomial over Z_p , $f(x) = ax+b$, where p is a prime and $p > \sum_{i=1}^{n} m_i$, $p > \sum_{i=1}^{n} ID_i$. The parameters a and b are kept secret.
(2) Compute the secret key $k_i = f(ID_i)$ mod p for each sensor node S_i, where ID_i is S_i's identify, $1 \le i \le n$.
(3) Share (t_1, t_2) with S_i and send (ID_i, k_i) to S_i in a secure way.

Data concealment phase:

For each S_i with (ID_i, k_i) and (t_1, t_2), assume it captures a sensed data m_i. It should do the following operations.

(1) Define a polynomial over Z_p , $g_i(x) = \alpha_i x + m_i$, where $\alpha_i = (k_i - m_i)/t_2$. That implies that $k_i = g_i(t_2)$.
(2) Compute the encrypted sensed data $d_i = g_i(t_1)$ mod p.
(3) Send (ID_i, d_i) to A.

Concealed data aggregation phase:

Assume the aggregator A receives m pairs of (ID_i, d_i) from the sensor nodes, it should do the following operations.

(1) Compute $u = \sum_{i=1}^{m} ID_i$, $v = \sum_{i=1}^{m} d_i$.
(2) Send (m, u, v) to R.

Concealed data disclosure phase:

For the sink node R, upon receiving (m, u, v), it should do the following operations.

(1) Compute $r = au + mb$ mod p. Note that r is also equal to $\sum_{i=1}^{m} k_i$ mod p.
(2) Compute the disclosed aggregated data $T = \sum_{i=1}^{m} m_i = (t_2 * v - t_1 * r)/(t_2 - t_1)$ mod p.

Finally, the sink node obtains the aggregated data $T = \sum_{i=1}^{m} m_i$, i.e., the summary of the sensed data. Then, the average of the sensed data will be T/m. Note some sensor nodes do not send the message out for some reasons and, of course, the aggregator node does not aggregate these messages.

D. Prove the Correctness

The correctness in the concealed data disclosure phase will be demonstrated as follows.

$$
\begin{aligned}
T &= (t_2 v - t_1(au + mb))/(t_2 - t_1) \\
&= (t_2 \textstyle\sum_{i=1}^{m} d_i - t_1(a \sum_{i=1}^{m} ID_i + mb))/(t_2 - t_1) \\
&= (t_2 \textstyle\sum_{i=1}^{m} (\alpha_i t_1 + m_i) - t_1(\sum_{i=1}^{m} k_i))/(t_2 - t_1) \\
&= (t_2 \textstyle\sum_{i=1}^{m} (t_1(k_i - m_i)/t_2 + m_i) - t_1(\sum_{i=1}^{m} k_i))/(t_2 - t_1) \\
&= (t_1 \textstyle\sum_{i=1}^{m} k_i - t_1 \sum_{i=1}^{m} m_i + t_2 \sum_{i=1}^{m} m_i - t_1 \sum_{i=1}^{m} k_i)/(t_2 - t_1) \\
&= (t_2 - t_1) \textstyle\sum_{i=1}^{m} m_i / (t_2 - t_1) \\
&= \textstyle\sum_{i=1}^{m} m_i \bmod p
\end{aligned}
$$

Since $p > \sum_{i=1}^{n} m_i \ge \sum_{i=1}^{m} m_i$, so T is $\sum_{i=1}^{n} m_i$.

IV. Security Analysis and Discussions

Due to the shared-medium nature of the wireless links, an adversary can easily intercept legitimate traffic, tamper the original traffic, or inject superfluous traffic, even compromise sensor nodes in a wireless sensor network to collapse the network. To deal with the malicious attacks in WSNs, the proposed scheme focuses on protecting the encryption of sensed data, and the goal is securely delivers the concealed aggregation data from nodes to sink node. In this paper, we address two attack models, active attacks and passive attacks to show the proposed scheme is secure.

A. Security Analysis

First, the active attacks are considered. A malicious attacker can compromise nodes and gets the secure data, like ID_i, k_i, and d_i.

However, he can not reconstruct the function $f(x)$ and $g(x)$.Because the parameters a and b is secure and keeps by sink node. Next, we consider passive attacks. It is impossible for an attack intends to reveal the sensed data m_i form ID_i and d_i. Since then is an unknown in $g_i(x)$, namely α_i . Similarly, the key k_i is secure too. An attacker cannot compute k_i from d_i and ID_i since $\alpha_i = k_i - m_i$, without knowing α_i , k_i can be any value.

B. Disscussion

We adopt Shamire 2-out-of-2 threshold scheme to share the sensed data m_i between sensor node ID_i and the sink. The secret polynomial is $g_i(x) = \alpha_i x + m_i$. The two shares are (t_1, k_i) and (t_2, d_i). Since k_i is known to the sink, only d_i is required to send to the sink. We also use the following property of linear function: $\sum_i g_i(x) = \sum_i (\alpha_i x + m_i) = \sum_i \alpha_i x + \sum_i m_i$. Therefore, we can use the aggregation of the IDs to compute $\sum_i k_i$.

Therefore, we can think that the sink and the sensor nodes share the sent $\sum_{i=1}^{m} m_i$ using Shamir 2-out-of-2 threshold scheme where the secret polynomial is $G(x) = \sum_{i=1}^{m} \alpha_i x + \sum_{i=1}^{m} m_i$ and the two shares are $(t_1, \sum_{i=1}^{m} k_i)$ and $(t_2, \sum_{i=1}^{m} d_i)$.

C. Energy Consumption

For easy description of energy consumption, we follow [4] and reconstruct a multi-level WSN model with degree 3. There are 2187 sensor nodes, 1092 aggregators, and only one sink node in this scenario and same with [4]. We refer Castelluccia et al's scheme as CMT and use the average operator to compare the performance of CMT, hop-by-hop protocol (HBH), and no-aggregation (No-Agg). We follow CMT's assumptions where each ciphertext is $log(M)=log(t)+log(n)$ bits long and the package header is fixed as 56 bits. Note that $log(t)$ is the size of the plaintext, and $log(n)$ is the size of the ID of sensor nodes.

We compare the performance of using average operator in CMT, our proposed scheme, HBH, and No-agg. The results are shown in Table 1. In CMT, the number of bits sent by the leaves is larger with the aggregation methods (CMT with A(0%): 56+ $log(t)$+ $log(n)$= 75 bits) than when no aggregation is used (No-agg: 56+$log(t)$=63 bits) where the A(0%) means that A is average operator and all sensor nodes send their sensing data. In our scheme, the number of bits sent by sensor nodes in level 7 of A(0%) is $log(M)$= 56+ $log(u)$+ $log(v)$+ $log(m)$ = 56+ 22+ 13+ 2= 93 bits, and $log(M)$= 56+ $log(u)$+ $log(v)$+ $log(m)$ = 56+ 22+ 13+ 4+ 64= 95 bits in level 6. Where $log(u) = \sum_{i=1}^{m} ID_i$ and needs 22 bits to recode all nodes' ID. And we need $log(v) = \sum_{i=1}^{m} d_i$ = 13 bits for aggregating all sensing data, assuming each sensed data is 7 bits long.

In WSNs, if assume all sensing data will be sent back to sink node is unrealistic. In table 1, it is extremely to show our scheme needs more bits than CMT in all sensor nodes send their sensing data back. However, our scheme is more efficient than CMT in practical scenario, especially when many sensor nodes are breakdown.

D. Main Contributions

We proposed a novel end-to-end CDA scheme based on the concept of secret sharing scheme. It enjoys the following properties:

(1) It provides end-to-end encryption on the sensed data between the sensor node and the sink node;
(2) It can limit the damage from compromised sensor nodes since sensor nodes have distinct keys;
(3) It preserves small size of ciphertext during transmission;
(4) It is scalable to large sensor networks due to its lightweight computation and easy key management.

V. CONCLUSION

Wireless sensor networks are widely used in a variety of applications. Maintaining the security and increasing the lifetime of WSNs are essential to the success of their applications. In order to save the overall energy resources and maintain the security of WSN, we need to reduce the amount of encrypted data transmitted. One approach is to consolidate the encrypted data along the routing path. This is called concealed data aggregation (CDA). In this paper, a novel end-to-end CDA scheme based on the concept of secret sharing is proposed to achieve simultaneously the goals of saving power and securely sending the concealed data. Our scheme is more bandwidth-efficient than the CMT scheme.

REFERENCE

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci," A Survey on Sensor Networks," *IEEE Communications Magazine*, vol.40, no.8, pp.102-114, 2002.
[2] M. Anand, E. Cronin, M. Sherr, M. Blaze, Z. Ives, and I. Lee," Sensor Network Security: More Interesting Than You Think," *Proc. USENIX Workshop on Hot Topics in Security*, 2006.
[3] A. Boulis, S. Ganeriwal, and M.B. Srivastava," Aggregation in Sensor Networks: An Energy-Accuracy Trade-Off," *Elsevier Journal of Ad Hoc Networks*, vol.1, pp.317-331, 2003.
[4] C. Castelluccia, E. Mykletun, and G. Tsudik," Efficient Aggregation of Encrypted Data in Wireless Sensor Networks," *Proc. Mobile and Ubiquitous Systems: Networking and Services*, pp.109-117, 2005.
[5] H. Chan, A. Perrig, D. Song," Secure Hierarchical in-Network Aggregation in Sensor Networks," *Proc. ACM conference on Computer and Communications Security*, pp.278-287, 2006.
[6] Crossbow Technology Inc., Motes: Smart Dust Sensors, Wireless Sensor Networks," Webpage. [Online]. Available: http://www.xbow.com.
[7] J. Domingo-Ferrer," A Provably Secure Additive and Multiplication Privacy Homomorphism," *Proc. Information Security Conference, LNCS* vol.2433, pp.471-483, 2002.
[8] J. Girao, D. Westhoff and M. Schneider," CDA: Concealed Data Aggregation for Wireless Sensor Networks," *Proc. IEEE International Conference on Communications*, pp.3044-3049, 2005.
[9] L. Hu and D. Evans," Secure Aggregation for Wireless Networks," *Proc. Symposium on Applications and the Internet Workshops*, pp.384-391, 2003.
[10] P. Jadia, A. Mathuria," Efficient Secure Aggregation in Sensor Networks," *Proc. High Performance Computing, LNCS,* vol.3296, pp.40-49, 2004.
[11] A. Mahimkar, T.S. Rappaport," SecureDAV: A Secure Data Aggregation and Verification Protocol for Sensor Networks," *Proc. IEEE Global Telecommunications Conference*, vol.4 pp.2175-2179, 2004.
[12] S. Peter, K. Piotrowski, and P. Langendoerfer," On Concealed Data Aggregation for Wireless Sensor Networks," *Proc. IEEE Consumer Communications and Networking Conference*, pp.192-196, 2007.

[13] K. Piotrowski, P. Langendoerfer and S. Peter," How Public Key Cryptography Influences Wireless Sensor Node Lifetime," *Proc. ACM Workshop on Security of Ad Hoc and Sensor Networks*, pp.169-176, 2006.

[14] B. Przydatek, D. Song and A. Perrig," SIA: Secure Data Aggregation in Sensor Networks," *Proc. First ACM Workshop Sensor Systems*, Nov. 2003.

[15] 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*, vol.5, pp.1417-1431, 2006.

[16] S. Zhu, S. Setia, and S. Jajodia," LEAP: Efficient Security Mechanism for Large-Scale Distributed Sensor Networks," *Proc. ACM conference on Computer and Communications Security*, pp.62-72, 2003.

Table 1. Number of bits sent per node for each level in four schemes.

Levels	Nun Nodes	CMT					Our scheme					HBH-A	No-Agg
		A(0%)	A(10%)	A(30%)	A(50%)	A(70%)	A(0%)	A(10%)	A(30%)	A(50%)	A(70%)		
1	3	75	950	2699.4	4449	6198.6	103	102	102	102	101	73	68859
2	9	75	366	949.8	1533	2116.2	101	101	100	100	99	72	22932
3	27	75	172	366.6	561	755.4	99	99	99	98	98	70	7623
4	81	75	107	172.2	237	301.8	98	98	97	97	96	68	2520
5	243	75	85	107.4	129	150.6	96	96	96	95	95	67	819
6	729	75	78	85.8	93	100.2	95	95	94	94	93	65	252
7	2187	75	75	75	75	75	93	93	93	92	92	63	63