ISSN 2277-3061

SIAVA: Secret Information Aggregation Design for Various Applications in Wireless Sensor Networks

T.Kavitha [1], S.Chinnadurai [2], P.Kumaran [3], M.Ezhilvendan [4]

PG-Student [1], Assistant Professor [2, 3, 4]

Srinivasan Engineering College, Perambalur [1, 2],

Veltech Multitech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai[3]

Jawahar Engineering College, Chennai [4]

kavithadevaraju50@gmail.com [1], duchinna198227@gmail.com [2],

kumaran.0991@gmail.com [3], vendannetwork@gmail.com [4]

Abstract—

In general information aggregation design that reduces a large amount of transmission is the most practical technique. In previous studies, homomorphic encryptions have been applied to conceal communication during aggregation such that enciphered data can be aggregated algebraically without decryption. Since aggregators collect data without decryption, adversaries are not able to forge aggregated results by compromising them. However, these schemes are not satisfy multi-application environments. Second, these schemes become insecure in case some sensor nodes are compromised. Third, these schemes do not provide secure counting; thus, they may suffer unauthorized aggregation attacks. Therefore, we propose a new concealed data aggregation scheme extended from Boneh et al.'s homomorphic public encryption system. The proposed scheme has three contributions. First, it is designed for a multi-application environment. The base station extracts application-specific data from aggregated cipher texts. Next, it mitigates the impact of compromising attacks in single application environments. Finally, it degrades the damage from unauthorized aggregations. To prove the proposed scheme's robustness and efficiency, we also conducted the comprehensive analyses and comparisons in the end.

Index Terms—Concealed data aggregation; elliptic curve cryptography; homomorphic encryption; wireless sensor networks.

Council for Innovative Research

Peer Review Research Publishing System

Journal: International Journal of Computers & Technology

Vol 12, No 2

editor@cirworld.com

www.cirworld.com, member.cirworld.com

ISSN 2277-3061

1 INTRODUCTION

WIRELESS sensor networks (WSNs) consist of thousands of sensor nodes (SN) that gather data from deployed environments. Currently, there are plenty of rich applications proposed for WSNs, such as environment monitoring, accident reporting, and military investigation [1]. Depending on the purpose of each application, SN are customized to read different kinds of data (e.g., temperature, light, or smoke). Typically, SN are restricted by the resources due to limited computational power and low battery supply; thus, energy saving technologies must be considered when we design the protocols. For better energy utilization, cluster-based WSNs [2] have been proposed. In cluster-based WSNs, SN resident in nearby area would form a cluster and select one among them to be their cluster head (CH). Although data aggregation could significantly reduce transmission, it is vulnerable to some attacks. For instance, compromising a CH will allow adversaries to forge aggregated results [4] as similar as compromising all its cluster members. The decrypted aggregated result will be incorrect. The only solution is to aggregate the cipher texts of different applications separately. As a result, the transmission cost grows as the number of the applications increases. By CDAMA, the cipher texts from different applications can be encapsulated into “only” one cipher text. Conversely, the base station can extract application-specific plaintexts via the corresponding secret keys. CDAMA mitigates the impact of compromising SN through the construction of multiple groups. An adversary can forge data only in the compromised groups, not the whole system.

2 SYSTEM MODEL

Here, we state two models for further uses, aggregation model and attack model. The aggregation model defines how aggregation works; the attack model defines what kinds of attacks a secure data aggregation scheme should protect from.

2.1 Aggregation Model

In WSNs, SN collect information from deployed environments and forward the information back to base station (BS) via multihop transmission based on a tree or a cluster topology. The accumulated transmission carries large energy cost for intermediate nodes. To increase the lifetime, tree-based or cluster networks force the intermediate nodes

(a sub tree node or a cluster head) to perform aggregation, i.e., to be aggregators (AG).

2.2 Attack Model

First of all, we categorize the adversary’s abilities as follows:

1. Adversaries can eavesdrop on transmission data in a WSN.
2. Adversaries can send forged data to any entities in a

   WSN (e.g., SN , AG, or BS).
3. Adversaries can compromise secrets in SN s or AGs through capturing them.

Second, we define the following attacks to qualify the security strength of a CDA scheme. Part of these attacks refer to Peter et al.’s analysis [15]. Based on adversary’s abilities and purposes, we further classify these attacks into three categories.

3 PRELIMINARIES

3.1 Privacy Homomorphic Cryptosystem

Privacy homomorphic encryption (PH) is an encryption scheme with homomorphic property. The homomorphic property implies that algebraic operations on plaintexts can be executed by manipulating the corresponding cipher texts;

EK ð Þ is the encryption with key K, DK ð Þ is the decryption with key K, and and denote operations on cipher texts and plaintexts, respectively. In general, operations and can be addition, multiplication, and so on.Similar to conventional encryption schemes, PH schemes are classified to symmetric cryptosystem when the encryption and decryption keys are identical, or asymmetric cryptosystem (also called public key cryptosystem) when the two keys are different. Symmetric PH schemes, such as Domingo-Ferrer scheme [17] or Castelluccia et al.’s scheme, usually are more competitive in terms of efficiency than asymmetric schemes. The most notable asymmetric PH schemes are based on elliptic curve cryptography (ECC). Compared with RSA cryptosystems, ECC provides the same security with a shorter key size and shorter cipher texts. A 160-bit ECC cryptosystem provides the same security as a 1,024-bit RSA cryptosystem [5]. In energy- constraint WSNs, constructing PH via ECC is more efficient. Up to now, the PH schemes on ECC include elliptic curve Okamoto-Uchiyama (EC-OU), elliptic curve Naccache- Stern, elliptic curve Paillier, and elliptic curve ElGamal Encryption Schemes (EC-EG) [2].

3.2 CDA Based on PH

Conventional hop-by-hop aggregation schemes are insecure because an adversary is able to forge aggregated results such as compromising all the AG's child nodes when he compromises the secret of an AG. To diminish this impact, PH schemes have been applied to WSNs [9], [10], [2], [12],[14]. By PH schemes, SN s encrypt their sensed readings and allow AGs to homomorphically aggregate their cipher texts without decryption. Therefore, compromising AGs earns no advantage of forging aggregated results. West off et al. [9] and Girao et al. [10] proposed CDA based on symmetric PH to facilitate the aggregation of encrypted data. In contrast to symmetric PH construction, Mykletun et al. [2] adopted public-key-based PH to construct their systems, and Girao et al. [12] extended the ElGamal PH encryption to construct an aggregation scheme. In these schemes, because all SN in a network only share a common key for encryption [9], [10], [2], [12], an adversary can forge the aggregated results by simply compromising one SN

(PRNG) and adds its messages with the key under modulation. The AG aggregates those cipher texts through modular addition. And the BS decrypts the cipher text received by modular subtraction with all the temporal keys.If an adversary tries to forge aggregated results, he must compromise all SN s. However, their scheme cannot prevent the adversary from injecting forged data packets into the legitimate data flow. In addition, key synchronization must be guaranteed because each SN must rekey after each encryption.

BGN provides additive and multiplicative homomorphism. Since the multiplicative property, based on the bilinear pairing [13], is much expensive and inefficient for SN s [21], we only utilize the additive homomorphism of BGN. In this paper, we first provide a possible application for BGN, data aggregation. Furthermore, we modify BGN to fit multigroup construction for stronger security and better applicability in CDA.

BGN is constructed on a cyclic group of elliptic curve points. Precisely, these points form an algebraic group, where the identity element of the group is the infinite point,

1 [22]. Notation ord(P) denotes the order of a point P . Supposing ord(P) ¼ q, it indicates that q is the minimum integer that satisfies q P ¼ 1. In the KEYGEN function, the order of E is equivalent to the number of points in E. The detail construction of E is depicted in Section 6.3.

The ENC function is based on point addition and scalar multiplication over points G and H. As we can see, the cipher text is composed of the message part (the scalar of the point G) and the secure randomness (the scalar of the point H). Due to homomorphic properties, the AGG function aggregates cipher texts via point addition; it is trivial to see that the scalar values of point G were added in the end, yielding the sum of the corresponding message. Conse- quently, the final result will be the form of M Gþ R H, where M is the sum of the messages and R is the sum of the randomness. The DEC function decrypts the aggregated result to obtain the plaintext value, M . Recall that the order of points G and H are different.

4 CDAMA

BGN is implemented by using two points of different orders so that the effect of one point can be removed by multiplying the aggregated cipher text with the order of the point, and then the scalar of the other point can be obtained. Based on the same logic of BGN, CDAMA is designed by using multiple points, each of which has different order. We can obtain one scalar of the specific point through removing the effects of remaining points (i.e., multiplying the aggregated cipher text with the product of the orders of the remaining points). The security of CDAMA and BGN are based on the hardness assumption of subgroup decision problem, whereas CDAMA requires more precise secure analysis for parameter selections, discussing in Section 6.2. We use CDAMA (k ¼ 2) to explain how it works in multiple groups.

4.1 CDAMA (k ¼ 2) Construction

Assume that all SN s are divided into two groups, GA and GB . CDAMA contains four procedures: Key generation, encryption, aggregation, and decryption, listing in Fig. 2. As we can see, CDAMA (k ¼ 2) is implemented by using three points P; Q, and H whose orders are q1 ;q2 , and q3 , respectively can be aggregated to a single cipher text, but the aggregated. Now, the cipher text contains only the product of G (i.e., ordðHÞ M G) such that we can apply the discrete logarithm to retrieve the value M . In fact, discrete logarithm canp be solved by Pollard's method whose efficiency is Oð ffiffiffiffiT Þ.

Now, we use a brief instance to explain how BGN works in CDA. When sensor S1 gets its sensed reading M1 , S1 performs the ENC function to encrypt M1 as cipher text C1 . After that, S1 sends C1 to its aggregator AG. Once AG received all cipher texts fC1 ; ... ;C g from its child nodes, fS1 ; ... ;S g, the AG aggregates C1 to C through executing recursive (1) AGG operations on all cipher texts re- ceived, e.g., AGG(AGG(AGG(C1 ;C2 Þ;C3 Þ C). Then, AG sends the aggregated result to the next aggregator. message of each group can be obtained by decrypting the cipher text with the corresponding SK.

Considering deployment, the private keys should be kept secret and only known by the BS. SN s in the same group share the same public key and no other entities outside the group knows the group public key. How to securely deliver the public keys to different groups of SN s will be discussed later in Section 4.4. Another major change is the decryption procedure. By performing individual decryption, the BS extracts individual aggregated results of different groups from an aggregated cipher text.

4.2 A Concrete Example

Now, we use an instance to describe how CDAMA (k ¼ 2) works. In Fig. 3, a WSN consists of six SN s and four AGs. After deployments, they form three clusters. Each SN belongs to either application A or B. Without loss of generality, sensors A1 , A2 , and A3 perform application A and keep the public key PKA ¼ ðn; E; P; H;TA Þ. The others, B1 , B2 , and B3 keep PKB ¼ ðn; E; Q; H;TB Þ. Four aggregators, AG1 to AG4 are deployed to gather messages from their child nodes. To simplify the example, we set the order of P, Q, and H to small numbers. We assume that jq1 j ¼ jq2 j ¼ jq3 j ¼ 10, e.g., ord(P) ¼ q1 ¼ 521, ord(Q) ¼ q2 ¼

523,ord (H) ¼ q3 ¼ 541, a n d n ¼ q1 q2 q3 ¼ 147; 413; 303, where jqi j is the bit size of qi . Moreover, we assume T ¼

128 and x ¼ 3 such that the maximal sensed value in both applications is at most 42 (i.e., TA ¼ TB ¼ 42).

We assume the messages of these sensors are MA1 ¼

13;MA2 ¼ 21, MA3 ¼ 10, MB1 ¼ 32, MB2 ¼ 17, and MB3 ¼

24. They are encrypted to the corresponding cipher texts. After the aggregation by the AGs, the BS receives the final aggregated result AR4 whose value is 36P þ 73Qþ

195; 121; 825H ¼ 36P þ 73Qþ 477; 385; 22H. The aggre- gated result in application A, MA ¼ M1 þ M2 þ M3 ¼ 36 can be obtained by decrypting AR4 using SKA in the following steps:

4.3 Generalization of CDAMA

CDAMA (k ¼ 2) can be generalized to CDAMA (k > 2). The paradigm of generalization uses different generators to construct different key pairs for groups. For security reasons, the order of E should be large enough. Therefore, when k becomes large, the length of cipher text will also expand. The analysis on this overhead is stated in Section 6.2. For multi-application WSNs, the SNs belonging to one specific application are assigned the same group public key. Under CDAMA, the cipher texts from different applications can be aggregated together, but they are not mixed. The ciphertexts can be integrated into a ciphertext and transmitted to the BS. The BS then individually decrypts the aggregated ciphertext to extract the aggregated value of each application.

4.4 Key Distribution

In the end of this section, we briefly address how to deliver the group public keys to SNs securely. There are two main approaches.

Conventional Aggregation Model with Multiple Groups

Interestingly, applying CDAMA to the conventional aggregation model can mitigate the impact from compromising attacks. Fig. 6 shows an example of this case. In Fig. 6, all SN s are in the same application, e.g., fire alarm, but they can be arranged into two groups through CDAMA construction. Each group could be assigned a distinct group public key. Once an adversary compromised a SN in group A; it only reveals PKA , not PKB . Since the adversary can only forge messages in group A, not group B, the SN s in group B can still communicate safely. The ideal case is that CDAMA assigns every node for its own group, resulting in the strongest security CDAMA ever offered. However, this is impractical because the size of cipher text becomes extremely large when we construct groups with a huge group number.

5 DISCUSSION

In this section, we discuss several issues in CDAMA, including efficient implementation, cipher text length, and curve selection. The first is efficient computation. Since a lot of operations in CDAMA are based on scalar multiplication on elliptic curve points, skills which accelerate scalar multiplications can enhance the performance of CDAMA.

5.1 Efficient Scalar Multiplication

In CDAMA, the efficiency of encryption and decryption depends on the performance of scalar multiplication on elliptic curves. Decryption is not considered here because a BS is considering as powerful as a workstation. Given a random elliptic point P, we calculate k P with a given integer k by scalar multiplication.

5.2 Size of Cipher texts

Size of cipher texts is another metric for performance and cost evaluation. In CDAMA, the cipher text is stored as a couple of elliptic curve affine points. If the finite field of elliptic curve is F p , the size of cipher text is jpjþ 1 bits because we only store the x-coordinates of curve points and the additional one bit for the sign of y-coordinate. On the other hand, CDAMA requires specific curves of the given order. If we construct a curve with a given order, how can we estimate the bit length of the finite field, i.e., jpj in F p ? Based on Theorem 1, the size of finite field of a curve is

5.3 Generating Suitable Curves

The main challenge of constructing CDAMA is generating the set of elliptic curve points with a given order(generating the curves with given orders). The BGN scheme adopts pairing-friendly curves (also called super singular curves) to construct their scheme because bilinear pairing is necessary under their construction [33].However, these curves do not have computational efficiency because the length of the underlying field doubles; if the given order is k-bit long, the underlying prime field requires 2k bits. In CDAMA, we select different approach because bi-linear pairing is no longer required and length of the prime field doubles based on the given order in pairing-friendly curves. To find suitable curves in CDAMA, we select Bro¨ ker's [34] approaches to generate desired curves.

6 SECURITY ANALYSIS AND COMPARISON

In this section, we analyze the security of CDAMA and other conventional schemes. More specifically, we compare CDAMA with four well-known CDA schemes: CDA [9],

[10], Castelluccia et al.'s scheme [14], Mykletun et al.'s scheme [2], and TinyPEDS [12]. In Mykletun et al.'s scheme, the authors applied several well-known public key PH schemes to WSNs. They recommended two schemes which are suitable for WSNs, EC-OU and EC- EG. Since TinyPEDS [12] is the same as the EC-EG scheme

[2], we chose TinyPEDS as a candidate. In addition to these four schemes, BGN—from which our proposed CDAMA is extended—is also analyzed. Consequently, we analyze CDA, Castelluccia et al.'s scheme, TinyPEDS, EC-OU, BGN, and CDAMA based on the attack model defined in Section 2.2. A1. Ciphertext only attack. All schemes can defend against this basic attack.

7 PERFORMANCE EVALUATION AND COMPARISON

7.1 Candidate Schemes for Comparison

We only compare the performance of CDAMA with TinyPEDS [12] and EC-OU [2] because CDA [9], [10] and Castelluccia et al.'s scheme [14] are both symmetric schemes; therefore, they are not suitable to compare with asymmetric schemes. In general, symmetric schemes are more efficient but less secure than asymmetric ones. The security properties of CDA and Castelluccia et al.'s scheme have been verified in the previous section. To make the comparison comprehensive, BGN is also covered. Consequently, we chose EC-OU over F p (jpj ¼ 1; 024-bit), TinyPEDS over F p (jpj ¼ 163-bit), BGN over F p (jpj ¼

1; 024), and CDAMA (k ¼ 2 4) over F p (jpj ¼ 768; 1; 024, and 1,280) as candidates.

7.2 Evaluation Measurements

For evaluating these schemes, three terms 1, 2, and 3 are defined. The first two terms are used in measuring the computation cost, including the encryption cost on SN s and aggregation cost on AGs. The decryption cost on a BS is not measured because BSs are always as powerful as work- stations. The final term 3 is the communication cost per application:

1. The encryption cost on SN . The computation cost of encryptions. The unit is mJ/per encryption.
2. The aggregation cost on AG. The computation cost of aggregations. The unit is mJ/per aggregation.
3. The communication cost per applications. The cipher text size required by an application. The unit is bits/per application.

7.3 Evaluation Results

To analyze the computation cost, the same metric for all schemes is required. Since TinyPEDS, EC-OU, BGN, and CDAMA are all built on elliptic curves, encryption and aggregation are based on two kinds of operations, point addition and point scalar multiplication. In elliptic curve arithmetic, two basic operations are point doubling and adding. A point adding is computing P þ Q, where P and Q are curve points. Point doubling is computing 2P. Scalar multiplication is to compute r Q, where r is a scalar. Based on point adding and doubling, scalar multiplication is accomplished by the half-and-add algo- rithm [36]. More specifically, computing r Q requires

7.4 Performance Gain of CDAMA

In the above analysis, the computation cost of CDAMA is significantly large. Although data aggregation can reduce the communication effectively, sensors must pay higher computation cost for encryption and aggregation. To argue with this point, we estimate the performance gain from the whole WSN based on CDAMA.First of all, we classify sensors in large scale WSNs to three types by their tasks: Leaf nodes, AGs, and forwarders

Leaf nodes are leaves of a formed topology (e.g., a tree); they gather information from the deployed environment and send the result back the BS via other nodes. AGs are the intermediate nodes in the topology, such as parent nodes or cluster heads; they aggregate the forwarded messages if possible. Forwarders are the nodes on the path to the BS; their main task is to forward the aggregated result to the BS without aggregation. Next, we estimate the energy consumption on different nodes.

We compare CDAMA (k ¼ 2) with a WSN without data aggregation (also called Data Forwarding Scheme, DFS). In DFS, a leaf node encrypts its sensed reading by symmetric encryption schemes (e.g., AES) and forwards the cipher text to its parent AG. AGs and forwarders just transmit the received data without any in-network processing. Both schemes (rather than hop-by-hop aggregation) provide end- to-end security, thereby avoiding the forgery of aggregated result. The result is shown in Table 3. We assume that CT and CR are the cost of receiving one bit and transmitting

1 bit, respectively. C1E is the cost of AES encryption, and C2E and C2A are the cost of CDAMA encryption and aggregation, respectively. '1 and '2 are the bit length of a cipher text of AES and CDAMA, respectively. Moreover, we evaluate the result in Table 3 by substituting variables with practical values. That is to say, we use the estimated results of energy consumption on MICAz and TelosB in [38] to analyze the performance gain. The results are shown in Table 4. We assume that the deployed topology is a three-layer cluster.

8 CONCLUSION

For a multi-application environment, CDAMA is the first distinct applications can be aggregated, but not mixed. For a single-application environment, CDAMA is still more secure than other CDA schemes. When compromising attacks occur in WSNs, CDAMA mitigates the impact and reduces the damage to an acceptable condition.

ACKNOWLEDGMENTS

This research was supported by the National Science Council under the Grants NSC 101-2917-I-564-059, NSC 100- 2911-I-002-001, and NSC 100-2218-E-007-006.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A Survey on Sensor Networks," IEEE Comm. Magazine, vol. 40, no. 8, pp. 102-114, Aug. 2002.

[2] R. Min and A. Chandrakasan, "Energy-Efficient Communication for Ad-Hoc Wireless Sensor Networks," Proc. Conf. Record of the 35th Asilomar Conf. Signals, Systems and Computers, vol. 1, 2001.

[3] B. Przydatek, D. Song, and A. Perrig, "SIA: Secure Information Aggregation in Sensor Networks," Proc. First Int'l Conf. Embedded Networked Sensor Systems, pp. 255-265, 2003.

[4] A. Perrig, J. Stankovic, and D. Wagner, "Security in Wireless Sensor Networks," Comm. ACM, vol. 47, no. 6, pp. 53-57, June 2004.

[5] L. Hu and D. Evans, "Secure Aggregation for Wireless Networks," Proc. Symp. Applications and the Internet Workshops, pp. 384-391, 2003.

[6] H. Cam, S. O¨ zdemir, P. Nair, D. Muthuavinashiappan, and H.O. Sanli, "Energy-Efficient Secure Pattern Based Data Ag- gregation for Wireless Sensor Networks," Computer Comm., vol. 29, no. 4, pp. 446-455, 2006.

[7] H. Sanli, S. Ozdemir, and H. Cam, "SRDA: Secure Reference-based Data Aggregation Protocol for Wireless Sensor Networks," Proc. IEEE 60th Vehicular Technology Conf. (VTC '04-Fall), vol. 7, 2004.

[8] Y. Wu, D. Ma, T. Li, and R.H. Deng, "Classify Encrypted Data in Wireless Sensor Networks," Proc. IEEE 60th Vehicular Technology Conf., pp. 3236-3239, 2004.

[9] D. Westhoff, J. Girao, and M. Acharya, "Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks: Encryption, Key Distribution, and Routing Adaptation," IEEE Trans. Mobile Computing, vol. 5, no. 10, pp. 1417-1431, Oct. 2006.

[10] J. Girao, D. Westhoff, M. Schneider, N. Ltd, and G. Heidelberg, "CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks," Proc. IEEE Int'l Conf. Comm. (ICC '05), vol. 5, 2005.