International Journal of Enhanced Research in Management & Computer Applications, ISSN: 2319-7471
Vol. 3 Issue 12, December-2014, pp: (44-52), Impact Factor: 1.296, Available online at: www.erpublications.com

Distributed Data Aggregation for Secure Sensor Data

Soonhwa Sung

Dept. of Computer science and Engineering College of Engineering Software Research Center (SOREC), Chungnam National University, Yuseong-gu, Daejeon, 305-764, South Korea

Abstract: To solve data aggregation problems, several algorithms have developed to operate efficient networks. However, a current data aggregation still includes a design weakness. This paper proposes to design distributed aggregation for a confidential data using Software Robot (SR) in cluster sensor networks. Software robot of smart software is based on homomorphic encryption to aggregate secure sensing data and navigates to provide efficient communications in distributed multi-layer sensor networks. Therefore, this paper suggests how to support secure data aggregation in distributed sensor networks. In an analysis, this design has improved aggregation delay and supported secure data aggregation.

Keywords: Distributed data aggregation, Homomorphic encryption, Smart software, Multi-layer sensor, Cluster Head (CH), Base Station (BS).

I. INTRODUCTION

An aggregation in sensor networks is to reduce a lot of data transmission and to reduce data redundancy and to encapsulate necessary information without requiring all of the data. Data aggregation is performed by Cluster Head (CH) which is divided into several groups known as clusters. Base Station (BS) may require the maximum value of all sensing data to activate the immediate response. Each CH selects the maximum value of multiple sensing data of its cluster members and sends the result to the BS. Therefore, communication cost is reduced because only aggregated results reach the BS. The BS only brings the aggregated results, so the usage of aggregation function is obliged and the BS cannot verify the data integrity and authenticity. Besides, an adversary can access the sensing data of its cluster members after capturing a CH. To solve the problems, the study

[1] proposed the approach that conceals sensed data end-to-end by providing efficient and flexible innetwork data aggregation. The topology-aware key-predistribution provides the best achievable security without resistant device and ensures the application of concealed data aggregation for reverse multicast traffic. However, the key-predistribution is imposed much computing cost because data is dynamically created and removed.

[2] performed a simple and provably additive homomorphic stream cipher which allows efficient aggregation of encrypted data. The advantages are the influence of compromising a sensor is actually reduced. The disadvantage of the process is rekeying operations for each sensor cause the scheme to be impractical. In addition, a synchronization mechanism should be provided for
the scheme.

[3] suggested a data aggregation scheme based on addition homomorphic public-key encryption. It looks like more secure since every sensor stores only public key. The adversary cannot steer the same attack through compromising only one sensor because he can impersonate other legal sensors to send the forged cipher texts to the CH with the same public key.

[4] proposed that the mobile agent will have the capability of travelling from the sensor node via the aggregator and the forwarder to a reader. The mobile agents detect malicious nodes and the algorithm is easy to implement. However, It has the overhead of deploying mobile agents and a significant attention due to their ability to travel in the network independently and a predefined path.

An attacker can still access the sensing data of its cluster members after capturing a Cluster Head (CH). Base Station (BS) sends data which is aggregated to form a cipher text, when encrypting a group key and a cipher key to produce a cipher text. Attackers can collect the cipher text, then find whether the data has an attacker. If an attacker is inside the text, then he analyzes the text and sends back to a user. If an attacker is not present in the text, then an adversary decrypt the data and send a user. An attacker wants to send the forged message to cheat the BS even though he does not know the secret key. To solve above problems, data are encrypted during transmission and CH aggregates encrypted data without decryption. The concept of data aggregation that the BS obtains only aggregated results will be changed. For this, BS must recover each sensing data generated by all sensors even if these data have been aggregated by CH. Hence, a scheme needs that BS can find out the integrity and authenticity of all sensing data and execute any aggregation functions on them. Meanwhile, it is important to support distributed computing in which multiple authorized network users can simultaneously and directly program sensor nodes without BS.

Therefore, this paper proposes a confidential data aggregation using software robot for distributed programming. The rest of the paper is organized as follows. Section 2 surveys the related works, Section3 formulates ID-based multi-layer sensor networks, Section 4 describes a software robot for distributed confidential data aggregation, Section 5 analysis, Section 6 concludes the paper.

Related Works

For a secure and efficient data aggregation, distributed reprogramming protocol for Wireless Sensor Networks (WSNs) has proposed. There has been a lot of research on secure reprogramming, and many interesting protocols have been proposed in recent years [5-9]. However, all of them are based on the centralized approach which assumes the existence of a BS, and only the BS has the authority to reprogram sensor nodes. Unfortunately, the centralized approach is not reliable because, when the BS fails or when some sensor nodes lose connections to the BS, it is impossible to carry out reprogramming. Moreover, there are WSNs having no BS at all, so the centralized approach is not applicable. The centralized approach is inefficient, weakly scalable, and vulnerable to some potential attacks along the long communication path. The previous study [10] for secure and distributed reprogramming proposed to map the identity and reprogramming privilege of an authorized user into a public and private key pair. One of advantages of distributed reprogramming is that different authorized users may be assigned different privileges of reprogramming sensor nodes. This is particularly important in large-scale WSNs owned by an owner and used by different users from both public and private sectors [11, 12]. However, in [13], a design weakness exists in the user preprocessing and an adversary can easily impersonate any authorized user to carry out reprogramming. Therefore, to solve a design weakness, this paper proposes a distributed data aggregation based on homomorphic cryptosystem using SR for a secure and efficient programming.

ID-based Multi-layer Sensor Networks

The main idea in IBC (Identity-Based Cryptography) is to eliminate a public key that is derivable from some known aspect of a user's identity, such that public key directories are unnecessary [14]. This scheme supports ID-based multi-layer sensor networks with the efficient routing function [15] in Fig. 1. It consists of three layers which are composed of public (first layer), Sensor Key Translation (SKT: second layer), and confidential (final layer) layers dynamically are moved in a cluster [16]. The public layer contains the raw data and a simple ID. ID services include ID issue, renewal, and revocation. The serviced ID acts as a proxy candidate. The second layer network has SKT which translates a public key into a private key. The translation supports symmetric homomorphic encryption scheme [17] in a confidential layer because symmetric homomorphic encryption scheme need not decryption keys. SKT provides to apply symmetric homomorphic encryption scheme where someone who does not have decryption keys needs to perform arithmetic operations on a set of cipher texts. The final layer, confidential layer, operates the encrypted-data aggregation without pre-installing keys for verifications.

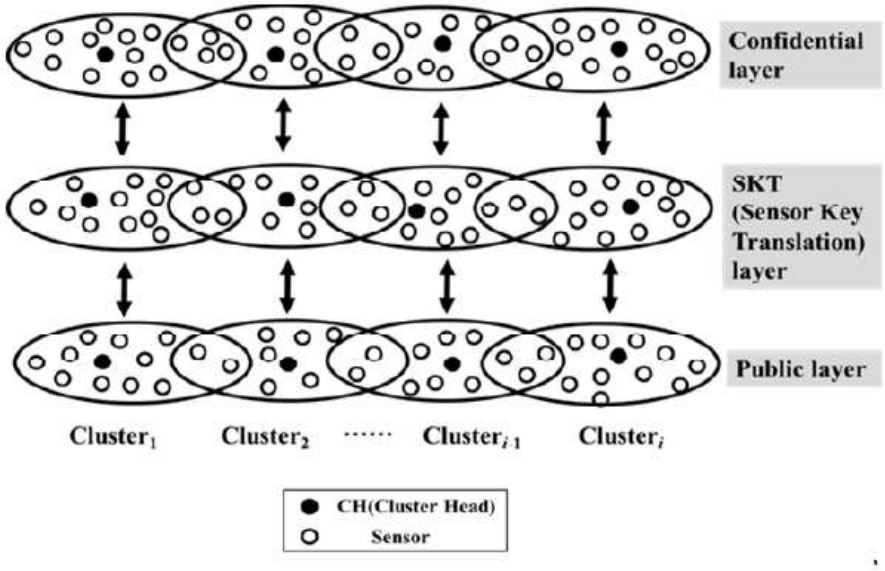

Figure 1. ID-based multi-layer sensor networks

The public layer contains the raw data and a simple ID. ID services include ID issue, renewal, and revocation. ID assignment for proxy candidates is enforced by *K*-proxies algorithm. The serviced ID acts as a proxy candidate with ID issue, ID renewal, and ID revocation. The *K*-proxies algorithm for ID assignment is as follows:

A. ***K*-proxies Algorithm:** Generate *k* request to discover *k* proxies

1**: Define**:
2**:** *w, x, y, z*: four end-nodes to set up ID assignment
3**:** *v*: proxy candidate
4**:** ID1: ID for the node *w*
5**:** ID*i*: ID for the node *x*
6**:** ID−1: ID for the node *y*
7**:** ID−*i*: ID for the node *z*
8**:** ID*self*: ID for itself
9**:** *Nx*: 1-hop neighbors of any node *x*
10**:** *Re*: request to set up ID assignment
11**:** *K*: number of proxies that must be found
12**:** ACK: acknowledge a node to be true
13**: Proxies(***k***)**: executed at randomly selected node *w, x, y,* or *z*
14**: for** *i* = 1 to *k* **do**
15**:** randomly select a node in *Nx* and send *Re*
16**: end for**
17**: if** receive positive ACK from node *v* **then**
18**:** register *v* as a proxy
19**: end if**
20**: Check**(*Re*): executed at all nodes receiving *Re*
21**: if** *Re* is not seen before **then**
22**: if** ID*self*∩ID*w* is not empty **then**
23**: if** ID*self*∩ID*x* is not empty **then**
24**: if** ID*self*∩ID*y* is not empty **then**
25**: if** ID*self*∩ID*z* is not empty **then**

26**:** register itself as a proxy for nodes *w*, *x*, *y*, and *z*
27**:** send back positive ACK to nodes *w*, *x* , *y*, and *z*
28**:** exit the procedure
29**: end if**
30**: end if**
31**: end if**
32**: end if**
33**: end if**
34**:** randomly select a neighbor other than the sender to forward *Re* [15].

This paper specially exposes the confidential layer involving distributed cipher computation. The distributed cipher computation in the confidential layer operates to support an efficient and secure data aggregation. To the distributed cipher computation, each sensing data in a cluster is based on homomorphic encryption.

B. Homomorphic Encryption

A homomorphic encryption is a special class of encryption function which allows the encrypted data to be operated on directly without requiring any knowledge about the decryption function. Suppose $EK(\cdot)$ is an encryption function with key K and $DK(\cdot)$ is the corresponding decryption function, then $EK(\cdot)$ is homomorphic with the operator $\circ$, if there is an efficient algorithm such that: $(EK(x), DK(y))=E(x \circ y)$ Given $EK(x)$ and $EK(y)$, there exists a computationally efficient Add algorithm such that : $EK(x+y)=\text{Add}(EK(x), EK(y))$. This implies that $EK(x+y)$ can be found easily from $EK(x)$ and $EK(y)$ without knowing the values for x and y[18].

Let p be a prime, $m_i<p$ the message to be encrypted, and K, k_i two random secret keys with K?0 and k, $k_i<p$. It is defined encryption as

$c_i=E(m_i, K, k_i , p)=(K \cdot m_i + k_i) \bmod p$
and decryption as

$m_i =D(c_i, K, k_i , p)=[(c_i - k_i) \cdot K_{-1}] \bmod p$

where K_{-1} is the multiplicative inverse of K modulo p.(K_{-1} always exists since p is prime)

Now consider two cipher texts c_1 and c_2 corresponding to plaintexts m_1 and m_2, respectively. Observe that we can compute the encryption of SUM $m_1+ m_2$ as

$c_1 + c_2 =E (m_1, K, k_1, p) + E (m_2, K, k_2 , p)$
$=[K \cdot (m_1 + m_2) + (k_1 + k_2)] \bmod p$
$=E(m_1 + m_2 , K, k_1 + k_2 , p)$

which can be decrypted using keys K and k_1+k_2 as

$m_1 + m_2= D(c_1 + c_2 , K, k_1 + k_2 , p)$

In general, $\sum_{i=1}^{N} m_i$ can be extracted from $\sum_{i=1}^{N} c_i$ using keys K and $\sum_{i=1}^{N} K_i$ in the decryption function [16].

The next section explains SR which encourages an efficient and secure data aggregation.

Software Robot for Distributed Confidential Data Aggregation

In several data aggregation schemes based on homomorphism encryption [17], CH can exactly aggregate the cipher texts without decryption and BS only fetches the aggregated result. However, they have the problem which BS cannot confirm the data integrity and confidence.

Table 1. Notation

Notation	Decription
WSN	Wireless Sensor Network
CH	Cluster Head
BS	Base Station
SR	Software Robot
KGC	Key Generation Center
K_{priv}	Private Key
K_{pub}	Public Key
M_{priv}	Master Key
S_{priv}	Session Key

To solve the problem, BS supports distributed confidential data using smart software. To support distributed data for authorized network users, smart software can confirm simultaneously and directly confidential sensor nodes without BS. Smart software, named "Software Robot(SR)", navigates sensor networks with Particle Swarm Optimization (PSO) algorithm [18]which is a population-based optimization technique inspired by the movements of a flock of birds or fishes. [19] has been proposed by minimizing total number of modular multiplication, so that faster exponentiation can efficiently implemented which in turn progress the time requirements of the large number of encryption or decryption. In Fig. 2, SR consists of decryption of encrypted data aggregation, verification using authentication protocol, CH joint channel, and secure aggregation monitoring.

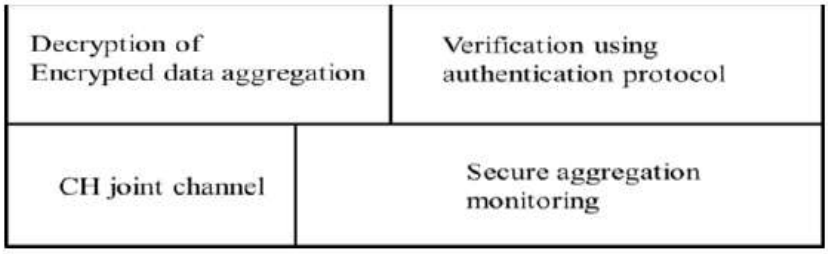

Figure 2. SR Functions of Smart Software

SR is managed by Key Generation Center (KGC) and communicates among BSs. It navigates to confirm the aggregated data security because BS cannot verify the data integrity and confidence. After navigating, it checks the data integrity and confidence of each sensing data using ID-based Node Authentication Protocol based on Sung et al.'s scheme [15]. Fig. 3 shows SR navigation protocol for confidential sensor nodes. The KGC generates a public key (*K*pub) and a private key (*K*priv) by a pseudorandom binary sequence generator and the cluster master key (*M*priv) to produce a session key (*S*priv) for the private keys (*K*priv) of the nodes in a cluster. SR sends the cluster master key (*M*priv) to CH and forwards a session key (*S*priv) for the private keys (*K*priv) of the nodes in a cluster to it. CH encrypts the sensing data by the session key (*S*priv) and sends it to BS. BS forwards to decrypt cipher texts to SR. SR decrypts and verifies them. After verification, SR approves and notifies it to KGC.

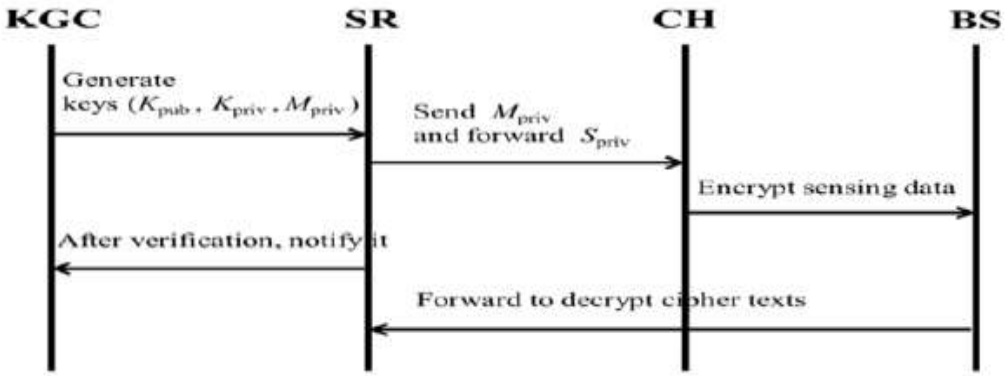

Figure 3. SR Navigation Protocol for Node Authentication

BS only aggregates data and forwards to SR for data confidence in Fig. 4. SR can evaluate and verify the trust of each sensor node. Therefore, SR can reduce time for aggregating legal sensor because BS only aggregates data and does not operate to confirm secure data aggregation.

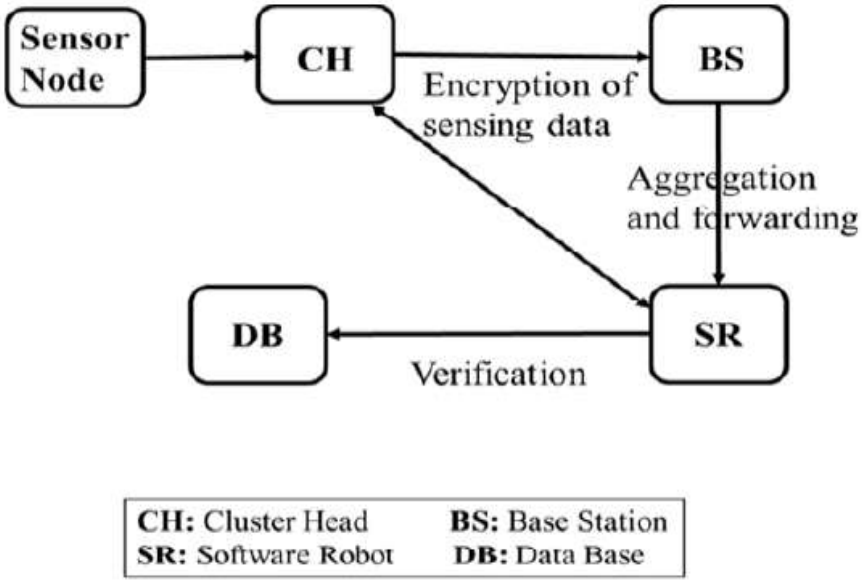

Figure 4. SR Association Relation

In addition, It can recover each sensing data generated by all sensors with PSO algorithm as followings: This scheme consists of four phases: set up, encryption, aggregation, verification

1) Set up: SR generates the key pairs for each sensor in a cluster. The key generation procedure considering the *K*-proxies algorithm for ID assignment is based on Sung et al.'s scheme.
2) Encryption: Each sensor in the CH shares a pairwise key. In this step, each sensor senses the data and encrypts it. Thus, SR sends the cipher text to another sensor in the same cluster.
3) Aggregation: CH collects all data from the cluster members and aggregates the cipher text pair from each sensor. After aggregation function of data summation, CH sends the aggregated results to BS. As soon as BS receives the aggregated results, SR instead of BS joins in the verification of them.
4) Verification: While receiving cipher text pair, SR can recover and verify each sensing data using ID-based Node Authentication Protocol based on Sung et al.'s scheme.

In Fig. 5, a cluster is comprised of CH and BS with many sensor nodes. BS aggregates data using PSO algorithm, and communicates with CH if it needs SR. CH communicates with each other about the necessity of PSO algorithm. These CH scan communications are executed periodically. After CH Scanning, CH decides whether it forwards to another CH or BS in a same cluster. When CH has forwarded to BS in a same cluster, BS recovers each sensing data generated by all sensors in a same cluster. When CH has forwarded to another CH, another CH supports distributed reprogramming in which multiple authorized users can simultaneously and directly reprogram sensor nodes except BS.

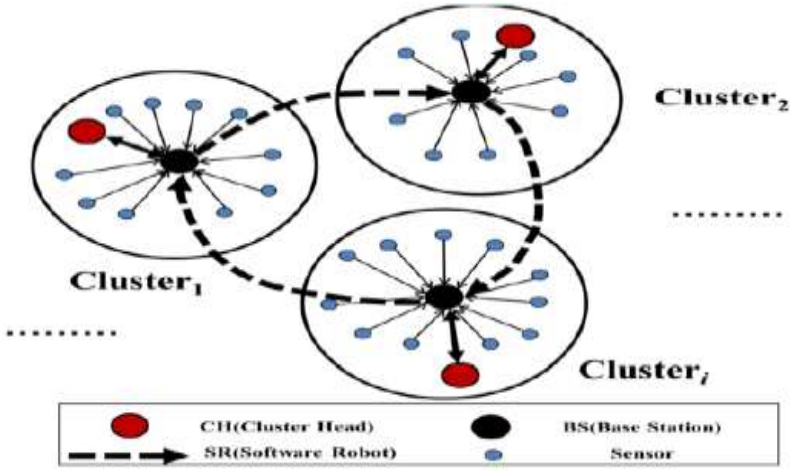

Figure 5. SR Navigator with Data Confidence

Analysis

This paper proposes a simple and secure homomorphic cryptosystem that allows efficient aggregation of a distributed encryption data with multi-layer. The proposed scheme followed current security recommendations for constrained network environments about the cryptographic primitives [17]. Concealed Data Aggregation in Multiple Applications (CDAMA) is the scheme that provides Concealed Data Aggregation (CDA) between multiple groups. CDAMA has two limitations when it aggregates multiple applications that shared in WSN. First, CDA requires data privacy and lower communication overheads. Second, aggregation of multi-application is still hard even if aggregation of cipher texts is possible, because the decryption cannot extract the aggregated results from a mixed cipher text [20]. To solve the problems, U. Sathya et al. [21] proposed enhanced data security using symmetric key algorithm. The proposed keys are used to provide security in data aggregation and separate the cipher text when it stored and retrieve from database. However, U. Sathya et al. scheme exist a system design weakness which attackers intercept keys while the data are sent to BS after aggregation. Therefore, to solve the problems which have U. Sathya et al. scheme, this paper proposes the distributed data aggregations system using SR of smart software with homomorphic encryption. In Fig. 6, comparing with U. Sathya et al. scheme, the proposed scheme has good performance evaluation. The processing delay indicates the execution time for sensors to produce cipher and decipher texts during the transmission to aggregate sensing data . The proposed scheme has less processing delay time than U. Sathya et al. scheme. The proposed scheme reduces communication overhead because SR which communicates CH with BS can securely navigate aggregated results. Moreover, the proposed scheme does not need to perform additional computations to verify the certificates because it is constructed by the ID-based concept.

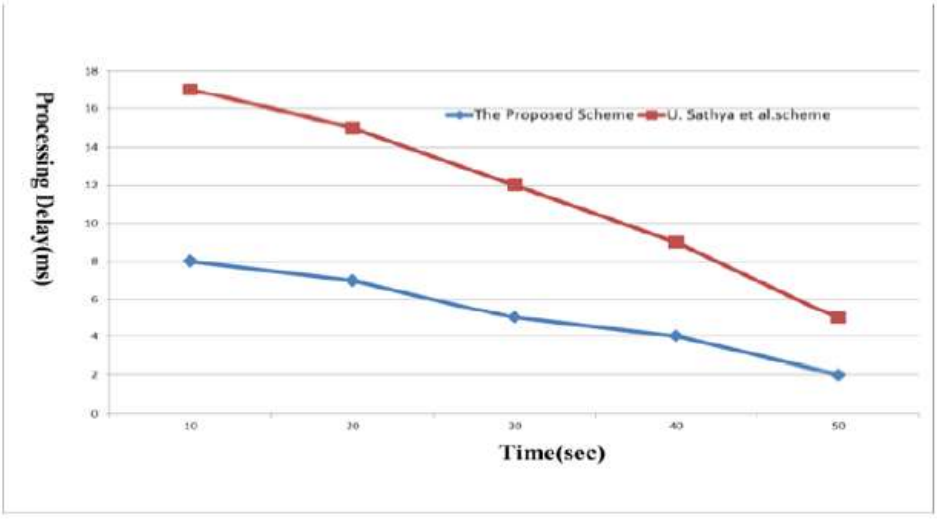

Figure 6. Processing Delay

Table 2 describes security comparison of the proposed scheme and U. Sathya et al. scheme. In user impersonation attack, to impersonate as the legitimate sensing data, an attacker attempts to make a forged sensing data. However, the attacker cannot impersonate as the legitimate sensing data by forging the aggregation request data because the attacker cannot compute the aggregation request data without knowing the private key of a sensor. In insider attack, an adversary wants to send the forged sensing data to cheat BS even though she does not know the secret key in U. Sathya et al. scheme because the scheme generates keys after aggregation. However, in the proposed scheme, an adversary cannot send the forged sensing data to cheat BS because SR using homomorphic encryption navigates BS. In mutual authentication, the proposed scheme supports mutual authentication because after SR navigation in sensor networks, SR verify a sensing data and the data stored in DB approve it, then a confidential data aggregation has completed. However, U. Sathya et al. scheme does not support mutual relations which encryption of sensing data in CH communicates with decryption of the aggregation data in BS.

Table 2. Security Comparison of the proposed Scheme and U. Sathya et al.scheme

Security Features	The Proposed Scheme	U. Sathya et al.scheme
User Impersonation Attack	Impossible	Possible
Insider Attack	Impossible	Possible
Mutual Authentication	Possible	Impossible

Acknowledgment

The author appreciates Prof. Cheong Youn, Eunbae Kong and Jaecheol Ryou for their helpful research supporting.

Conclusion

Data aggregation in WSNs is implemented to reduce data transmission and to summarize necessary information without requiring all the data. To solve some current data aggregation problems, this paper has proposed the distributed data aggregation using smart software, SR, which navigates to confirm secure aggregated data because BS cannot verify the data integrity and confidence. SR of smart software encourages the distributed BS to aggregate secure sensing data. The secure distributed data aggregation resists against attacks to cheat BS because SR based on homomorphic encryption navigates to provide secure sensing data in a cluster. In addition, the distributed data aggregation reduces the communication overhead due to SR functions. Therefore, the proposed scheme improves a security and efficiency of the distributed data aggregation.

References

[1]. Dirk Westhoff, Joao Girao, and Mithun Acharya, "Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks: Encryption, Key Distribution, and Routing Adaptation," IEEE Transaction on Mobile Computing, vol. 5, No. 10, pp.1417-1431, Oct. 2006.
[2]. C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient Aggregation of Encrypted Data in Wireless Sensor Networks," Proc. Second Ann. Int'l Conf. Mobile and Ubiquitous Systems, pp.109-117, July 2005.
[3]. E. Mykletun, Joao Girao, and Dirk Westhoff, "Public Key Based Cryptoschemes for Data Concealment in Wireless Sensor Networks," Proc. IEEE Int'l Conf. Comm., vol.5, pp.2288-2295, June 2006.
[4]. Bhavna Arora Makin and Devanand Padha, "Secure Data Aggregation Using Mobile Agents in Wireless Sensor Networks," Oriental Journal of Computer Science & Technology, vol. 3(1), pp.149-153, 2010.
[5]. P.K. Dutta, J. W. Hui, D. C. Chu, and D. E. Culler, "Securing the Deluge Network Program System," Proc. IPSN, 2006, 326-333.
[6]. Y. Law, Y. Zhang, J. Jin, M. Palaniswami, and P. Havinga, "Secure Rateless Deluge: Pollution-resistant Reprogramming and Data Dissemination for Wireless Sensor Networks," EURASIP J. Wireless Communication Networks, 2011(1), pp. 1-21, 2011.
[7]. C. Parra and J. Garcia-Macias, "A Protocol for Secure and Energy-aware Reprogramming in WSN," Proc. IWCMC, pp.292-297, 2009.
[8]. S. Hyun, P. Ning, A. Liu, and W. Du, Seluge, "Secure and DoS-resistant code Dissemination in Wireless Sensor Networks," Proc. IPSN, pp.445-456, 2008.
[9]. D. He, S. Chan, C. Chen, and J. Bu, "Secure and Efficient Dynamic Program Update in Wireless Sensor Networks," Secure Communication Networks, 5(7), pp.823-830, 2012.

[10]. D. He, C. Chen, S. Chen, and J. Bu, "SDRP: A Secure and Distributed Reprogramming Protocol for Wireless Sensor Networks," IEEE Trans. Ind. Electron., 59(11), pp.4155-4163, 2012.
[11]. (2011) Geoss. [Online] Available: http://www.epa.gov/geoss/
[12]. (2012) NOPP. [Online] Available: http://www.nopp.org/
[13]. D. He, C. Chen, S. Chan, J. Bu, and L. T. Yang, "Security Analysis and Improvement of a Secure and Distributed Reprogramming Protocol for Wireless Sensor Networks," IEEE Trans. Ind. Electron., 60(11), pp.5348-5354, 2013.
[14]. "Introduction to Identity-based Cryptography", CSEP 590 TU, [Online]Available: https://courses.cs.washington.edu /.../ youngblood_csep590tu_final_paper.pdf
[15]. Soonhwa Sung and Jaecheol Ryou, "ID-based Sensor Node Authentication for Multi-layer Sensor Networks," Journal of Communications and Networks (JCN), vol.16, No.4, August 2014.
[16]. Soonhwa Sung, "Confidential Aggregation for Wireless Transmissions," The International Conference on Information Networking 2014(ICOIN 2014), pp.390-394, Feb. 10-12, 2014.
[17]. M.K. Sandhya and K. Murugan, "Secure data aggregation in wireless sensor networks using privacy homomorphism", CCIS 2011, Advances in Networks and communications, vol. 132, pp.482-490, 2011.
[18]. Kennedy J. and Eberhart R, "Particle Swarm Optimization," Proceedings IEEE International Conference on Neural Networks, pp.1942-1948, 1995.
[19]. Arindam Sarkar, and J. K. Mandal, "Swarm Intelligence based Faster Public-Key Cryptography in Wireless Communication (SIFPKC)," International Journal of Computer Science & Engineering Technology (IJCSET), vol.3, No. 7, pp.267-273, July 2012.
[20]. Z. Shelby, K. Harke, and C. Bormann, "Constrained Application Protocol(CoAP)," draft-ietf-core-coap-18(WiP), IETF, 2013.
[21]. U. Sathya Rekha and Mrs P. Hemalatha, "Enhancing Data Security in WSN using Symmetric Key Algorithm," IOSR Journal of Computer Enfineering (IOSR-JCE) e-IISN: 2278-0661, p-ISSN: 2278-8727 Vol. 16, Issue 1, Ver. VII, pp 60-64, Feb. 2014.