A Homomorphic MAC-based Secure Data Aggregation Scheme for Wireless Sensor Networks

Yun Liu[1,2], Chaoran Li[3], Jing Zhang[4], Qing Liu[5]

[1] Key Laboratory of Communication & Information Systems, School of Electronic and Information Engineering, Beijing Jiaotong University, China
[2] CETC Key Laboratory of Aerospace Information Applications, China
[3] Information Engineering Department of Space Star Technology Co., Ltd., China
[4] The 54th Research Institute of CETC, China
[5] Power Construction Corporation of China, Ltd., China
liuyun@bjtu.edu.cn, sum41foxdie@gmail.com, zj_hb@163.com, liuqing@powerchina.cn

Abstract

A Wireless Sensor Network is composed of a large number of wireless sensor nodes and usually deployed in unattended environment to collect target information. At present, WSN is facing a growing range of security threats owing to its wireless and resource-constrained characteristics. Compromised sensor nodes can easily corrupt data accuracy and integrity by falsifying sensed information, selectively forwarding or misdirecting received data packets during the process of data aggregation. To solve these security problems, we propose a homomorphic MAC-based secure data aggregation scheme for WSNs (HMSDA) that can provide adequate protection of data confidentiality and integrity for wireless sensor networks. The simulation results indicate that HMSDA can effectively identify the tainted data in the process of data aggregation.

Keywords: Wireless sensor network, Secure data aggregation, Homomorphic MAC, Data confidentiality, Data integrity

1 Introduction

Wireless Sensor Networks for the low cost, convenient deployment and high automatic, are being widely used in various areas, such as military surveillance, industrial production, medical monitoring, hazardous materials transportation and etc [1]. Wireless sensor networks have some unique characters which distinguish them from the traditional networks [2-4]. One of them is the significant amount of redundant data generated by the overlapping sense ranges. In a large scale wireless sensor network, the redundant data can cause a large amount of unnecessary data traffic, bring a processing burden to the base station and a decision time delay of the system. To improve the energy efficiency of WSNs, data aggregation technique is proposed, which can effectively increase the operational efficiency and prolong the network lifetime by reducing redundant data [5-8]. Due to the openness of wireless sensor networks and their harsh working environment, the attackers can disrupt the security performance of WSNs through tampering, forgery, and other kind of malicious behaviors [9-10]. Consequently, network centers may make the wrong decision based on the distorted data aggregation results, which can cause irreparable damages. Therefore, how to ensure the data confidentiality and integrity in the process of data aggregation with low energy consumption has become one of the key research areas in the field of WSNs.

At present, the research on confidentiality and integrity protection mechanism for data aggregation in wireless sensor networks has achieved certain research results. Castelluccia et al. proposed a secure data aggregation scheme based on CMT encryption [11] and the complexity of the algorithm in the proposed scheme is low, which can reduce the computational load of sensor nodes. However, the scheme is not practical because the communication overhead of each sensor node is very large during the process of decryption. Subsequently, the researchers also proposed the AIE scheme [12], which can effectively reduce the communication overhead during data decryption and improve the energy efficiency and practicability of the scheme. Bahi et al. propose a secure data aggregation scheme based on Elliptic Curve Cryptography [13], which can effectively protect the data confidentiality in the process of data aggregation, but it does not provide the mechanism of data integrity verification. He et al. integrate the data integrity verification mechanism into the two existing secure data aggregation schemes separately, and two new schemes iPDA and iCPDA [14] are discussed.

*Corresponding Author: Yun Liu; E-mail: liuyun@bjtu.edu.cn
DOI: 10.3966/160792642018121907010

Papadopoulos et al. propose a secure data aggregation scheme SIES based on secret data sharing and homomorphic encryption technology [15]. This scheme can provide effective protection of data integrity, and achieve the high accuracy of data aggregation. However, the sizes of ciphertexts and secret keys are too large, which would increase the amount of data transmission between sensor nodes. Ozdemir et al. propose a secure data aggregation scheme DAA [16] and this scheme provides data privacy preservation through the deployment of node behavior supervision mechanism. Agrawal et al. propose a homomorphic message authentication code mechanism based on the technique of network coding [17], which has high security performance and practicability.

This paper puts forward a secure data aggregation scheme HMSDA based on the technique of homomorphic message authentication code. This scheme can optimize and improve the mechanism of homomorphic MAC proposed by Agrawal which is applied into the research field of WSNs data aggregation to provide an effective lightweight data integrity verification mechanism, and realize the additive homomorphic data aggregation.

2 The Technique of Homomorphic Message Authentication Code

In cryptography, the basic mechanism of data integrity verification can be summarized as follows: First, the sender adds a short redundant message (check code) into the plaint text with MAC function, and then sends the data to the receiver. Then, data receiver verifies the received data with MAC function. Finally, data receiver judges whether the transmitted data are damaged or not during the transmission by comparing the generated check codes. Figure 1 presents the mechanism of MAC.

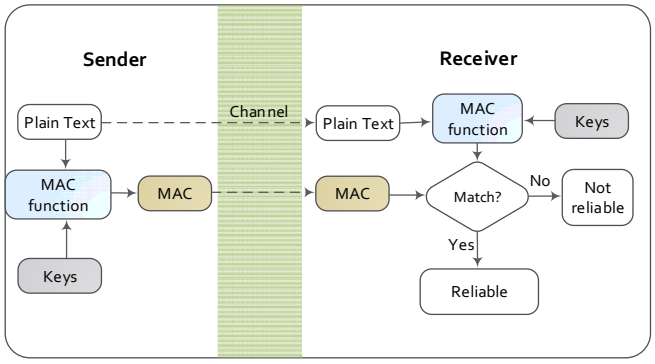


Figure 1. The mechanism of MAC

It can be observed from Figure 1 that in MAC mechanism, data are transmitted in clear text form, which means that data confidentiality and privacy cannot be protected during the process of transmission. Homomorphic Message Authentication Code (HMAC) technique can provide both data integrity verification and data privacy protection in the same time. According to the definition of homomorphism operation, the homogeneity of a homomorphic encryption function $Enc(\ )$ can be expressed as follows:

$$Enc(m_1 \oplus m_2) = Enc(m_1) \otimes Enc(m_2) \quad \textbf{(1)}$$

where m_i represents the raw data from data source i, operator $\oplus$ represents the mathematical method for manipulating plaintext, and $\otimes$ refers to the mathematical method for manipulating ciphertext. Similarly, the homogeneity of homomorphism message authenticate function $MAC(\ )$ can be expressed as follows:

$$MAC(m_1 \oplus m_2) = MAC(m_1) \otimes MAC(m_2) \quad \textbf{(2)}$$

This paper optimizes and improves the homomorphic message authentication code mechanism proposed by Agrawal [17], which suits wireless sensor network scenario better. The proposed homomorphic message authentication code mechanism will be presented detailedly in the following section.

3 Network Model

There are three kinds of nodes in the network: Base Station (BS), Aggregator node (An), Sensing node (Sn). BS is located at the root of aggregation tree and it has sufficient supplies of energy. It broadcasts data sensing order R_t to the lower network nodes, and is responsible for data decryption and data integrity verification. An is responsible for data aggregation and the transmission of data aggregation results. Sn performs data sensing operations based on the received data sensing order R_t, and transmits the encrypted raw data to An. From a hardware perspective, An and Sn are both resource-constrained wireless sensor nodes [18]. Figure 2 shows a formed aggregation tree network, where the BS broadcasted the data sensing order R_t to the lower network nodes. Figure 3 shows the process of additive data aggregation that triggered by order R_t.

4 Homomorhic-MAC Secure Data Aggregation HMSDA

In this section, we introduce the details of the homomorphic MAC mechanism which can be applied in HMSDA, and demonstrate the work process of HMSDA scheme by a data aggregation example.

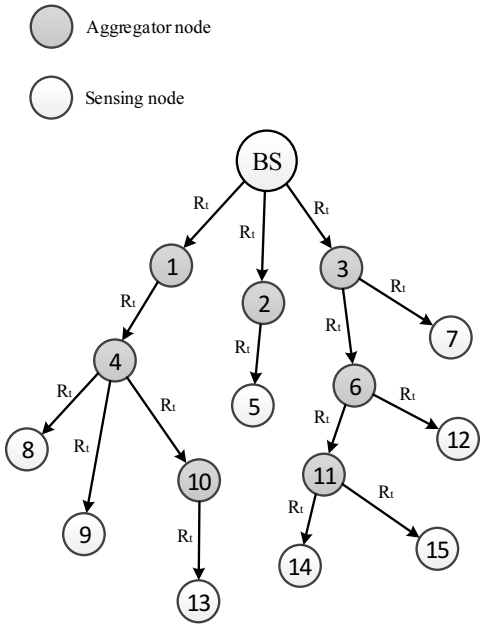


Figure 2. BS broadcasts data sensing order R_t

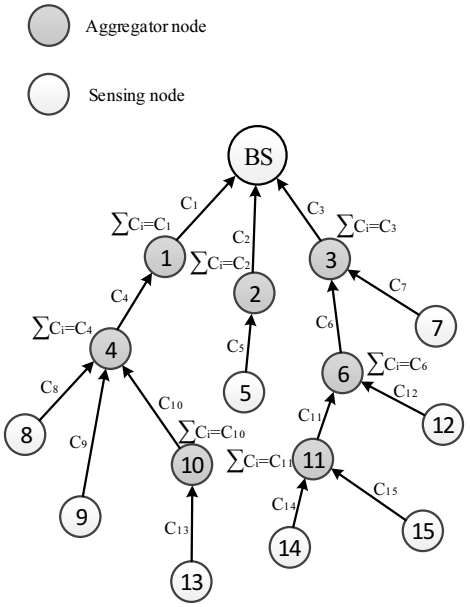


Figure 3. The process of data aggregation

4.1 The Generation of Secret Keys

We consider m_i as the raw data collected by node i, and according the previous definition of homomorphic MAC function, raw data m_i can be vectorized as the n-dimensions vector $m_i = (m_{i1}, m_{i2}, m_{i3}, ..., m_{in}), m_{ij} \in F_q$, and the space of plaintext is finite field F_q^n. We assume that ε_i represents the weight of node i, and the MAC key pairs generated by BS is (Mk_1, Mk_2). In the initial stage of network construction, the MAC key pairs, Pseudo Random Generator $(PRG)G$ and Pseudo Random Function $(PRF)F$ are preset in the BS and sensor nodes.

This paper adopts the homomorphic encryption based on prime-factorization problem and $v = a * b$, where v is the public key, and prime number a and b are secret keys. For the vectorized plaintext m_i, the encryption mechanism generates n secret key pairs (g_i, h_i), $i \in [1, n]$ for the corresponding n components of each plaintext vector.

4.2 The Algorithms of Encryption and Signature

The algorithm of signature. The algorithm substitutes variables Mk_1 、 Mk_2 、 id_i into pseudo random generator G and pseudo random function F to generate vector y in the finite field F_q and integer x_i, which are substituted into the signing function $Sign(Mk_1, Mk_2, id_i, m_i, i) = t_i$. The operation process is shown as follows:

$$\begin{aligned} &x_i + y \times m_i = t_i, \\ &y = G(Mk_1), x_1 = F(Mk_2, id_i) \end{aligned} \tag{3}$$

The symbol $*$ represents the operator of vector inner product computation, and the computation process can be illustrated as follow:

$$y * m_i = y_1 * m_{i1} + y_2 * m_{i2} ... + y_n * m_{in} \tag{4}$$

The algorithm of encryption. The cryptographic function $Enc(a, b, g_i, h_i, m_i)$ apply the preset secret keys a, b, g_i, h_i to encrypt the raw data, and the encryption process is shown in formula 5.

$$\begin{aligned} &Enc(a, b, g_i, h_i, m_i) \\ &= (g_1 \cdot m_{i1} \bmod a,\ h_1 \cdot m_{i1} \bmod b), \\ &..., (g_n \cdot m_{in} \bmod a,\ h_n \cdot m_{in} \bmod b) \\ &= (\alpha_{i1}, \beta_{i1}), ..., (\alpha_{in}, \beta_{in}) \end{aligned} \tag{5}$$

The sensor nodes integrate the cipher text C_i, information weight ε_i and homomorphic MAC t_i into data packets and transmit them to the aggregator nodes.

4.3 The Process of Aggregation

The aggregator nodes apply function $Agg(C_i)$ to perform additive aggregation to the ciphertext C_i and MAC t_i. Consider E as the set of the child node IDs, and the computation method is shown in formula 6.

$$t = \sum_{i \in E} t_i, \varepsilon_i, \quad \beta = \sum_{i \in E} C_i, \varepsilon_i \tag{6}$$

After the process aggregation, the aggregator nodes consolidate the aggregated ciphertext C_i, information weight ε_i and homomorphic MAC t_i into a data packet C_i and transmit it to the parent nodes.

4.4 The Algorithm of Decryption

BS substitutes the received data aggregation results, shared secret key pairs, the modular inverses of secret key pair and the node weight ε_i into function $Dec(C_i)$ to perform decryption. The process of decryption is shown as follows.

$$Dec(C_i) = \alpha_i^{-1} \cdot b \cdot b^{-1} + \beta_i^{-1} \cdot a \cdot a^{-1} \bmod v = m_i \quad \textbf{(7)}$$

$$m = \sum_{i \in E} Dec(C_i) = \sum_{i \in E} m_i \quad \textbf{(8)}$$

As we can see in formula 8, the decryption result m is the final aggregation result, and E is the set of the child node IDs. After the decryption, BS needs to verify the integrity of aggregation result.

4.5 The Verification of Data Integrity

In the stage of data integrity verification, the BS puts the shared secret key pairs, node weights and the decrypted data aggregation result into the signing function $Sign(Mk_1, Mk_2, id_i, m_i, i)$ again, and compares the computation result t_i' with the decrypted MAC t_i. The computation process of t_i' is shown in formula 9

$$\begin{aligned} &y = G(Mk_1), x = \textstyle\sum_{i=1}^{n} F(Mk_2, id_i) \cdot \varepsilon_i \\ &t' = x + y * \textstyle\sum_{i=1}^{n} m_i \bmod q \end{aligned} \quad \textbf{(9)}$$

If $t = t'$, then the integrity of data aggregation result remains intact, if $t \neq t'$, then the integrity of data aggregation result is compromised, and the system abandons this aggregation result.

4.6 An Example of Data Aggregation in HMSDA

In this section, we introduce the work mechanism of HMSDA by a data aggregation example. The wireless sensor network in this example is consist of one BS, one aggregator node and two sensor nodes, the architecture of the network aggregation tree is illustrated in Figure 4.

In this network, we assume that the raw data collected by sensing node 1 and 2 are $m_1 = 12, m_2 = 9$, which are vectorized into a vector space that consist of two components: $n = 2, m_{11} = 5, m_{12} = 7, m_{21} = 7, m_{22} = 2.$ The information weights of the sensing nodes are $\varepsilon_1 = 5, \varepsilon_2 = 4.$ The preset shared secret keys in the network are $v = 187$, a=17, $b = 11$, $g_1 = 6$, $g_2 = 3$, $h_1 = 9$, $h_2 = 10$. After the initialization of network parameters, the sensor nodes start the process of data encryption and homomorphic MACs generation.

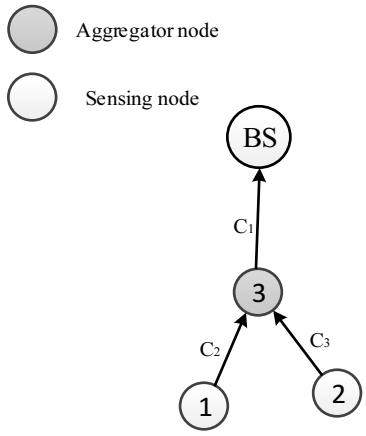


Figure 4. The architecture of the aggregation tree

Data encryption and the generation of MACs. After the collection of target data, sensing node 1 and 2 apply pseudo random generator G and pseudo random function F to generate the tags of the raw data, $x_1 = 7$, $x_2 = 3$, $y = (3,5)$, and then substitute this tags and vector m_1, m_2 into the algorithm of signature:

$$\begin{aligned} t_1 &= x_1 + y * m_1 = 7 + (3,5) * (5,7) = 57 \\ t_2 &= x_2 + y * m_2 = 3 + (3,5) * (7,2) = 34 \end{aligned} \quad \textbf{(10)}$$

In the stage of data encryption, sensor node 1 and 2 substitute the data vector m_1, m_2 and information weight ε_1, ε_2 into function $Enc(\,)$ to perform data encryption, the encryption process is shown below:

$$\begin{aligned} Enc(m_1) &= \begin{pmatrix} (m_{11} \cdot g_1 \bmod a, m_{11} \cdot h_1 \bmod b), \\ (m_{12} \cdot g_2 \bmod a, m_{12} \cdot h_2 \bmod b) \end{pmatrix} \\ &= \begin{pmatrix} (5 \times 6 \bmod 17, 5 \times 9 \bmod 11), \\ (7 \times 3 \bmod 17, 7 \times 10 \bmod 11) \end{pmatrix} \\ &= ((13,1),(4,4)) \\ Enc(m_2) &= \begin{pmatrix} (m_{21} \cdot g_1 \bmod a, m_{21} \cdot h_1 \bmod b), \\ (m_{22} \cdot g_2 \bmod a, m_{22} \cdot h_2 \bmod b) \end{pmatrix} \\ &= \begin{pmatrix} (7 \times 6 \bmod 17, 7 \times 9 \bmod 11), \\ (2 \times 3 \bmod 17, 2 \times 10 \bmod 11) \end{pmatrix} \\ &= ((8,8),(6,9)) \end{aligned} \quad \textbf{(11)}$$

After data encryption and MACs generation, node 1 and 2 separately integrate $Enc(m_1)$, t_1, ε_1 and $Enc(m_2)$, t_2, ε_2 into data packet C_1 and C_2, and send the data packets back to aggregator node 3.

Data aggregation. Aggregator 3 substitute the received data $Enc(m_1)$, $Enc(m_2)$, t_1, t_2, ε_1, ε_2 in function $Agg(\,)$ to perform additive aggregation, the process is shown as follow:

$$\begin{aligned}\beta &= \left(\varepsilon_1 \cdot Enc(m_1) + \varepsilon_2 \cdot Enc(m_2)\right)\\ &= 5\times\left((13,1),(4,4)\right) + 4\times\left((8,8),(6,9)\right)\\ &= \left((97,37),(44,56)\right)\\ t &= \sum_{i\in E} t_i \cdot \varepsilon_i \bmod 64\\ &= 5\times 57 + 4\times 34 \bmod 64 = 37\end{aligned} \quad (12)$$

In formula 13, E indicates the set of child node tag i, in this case $E=\{1,2\}$. After data aggregation, aggregator 3 integrates the aggregation results into data packet C_3 and transmits it back to BS.

Data decryption and integrity verification. Firstly, BS separately performs modular inversion on secret keys a, b, g_1, g_2, h_1, h_2 about mod a and mod b. The computation process is shown below:

$$\begin{aligned}&g_1^{-1} = 6^{-1} \equiv 3 \bmod 17, g_2^{-1} = 3^{-1} \equiv 6 \bmod 17,\\ &h_1^{-1} = 9^{-1} \equiv 5 \bmod 11, h_2^{-1} = 10^{-1} \equiv 10 \bmod 17,\\ &a^{-1} = 17^{-1} \equiv 2 \bmod 11, b^{-1} = 11^{-1} \equiv 14 \bmod 17.\end{aligned} \quad (13)$$

BS decomposes the received data packet C_3 to get the encrypted aggregation result, and then performs data decryption. The computation process is as follows:

$$\begin{aligned}\beta' &= \begin{pmatrix}\left(\alpha_1 \cdot g_1^{-1} \bmod a, \beta_1 \cdot h_1^{-1} \bmod b\right),\\ \left(\alpha_2 \cdot g_2^{-1} \bmod a, \beta_2 \cdot h_2^{-1} \bmod b\right)\end{pmatrix}\\ &= \begin{pmatrix}\left(97\times 3 \bmod 17, 37\times 5 \bmod 11\right),\\ \left(44\times 6 \bmod 17, 56\times 10 \bmod 11\right)\end{pmatrix}\\ &= \left((2,9)(9,10)\right)\end{aligned} \quad (14)$$

$$\begin{aligned}Dec(C_1) &= \alpha_1^{-1} \cdot b \cdot b^{-1} + \beta_1^{-1} \cdot a \cdot a^{-1} \bmod v\\ &= 2\times 11\times 14 + 9\times 12\times 2 \bmod 187\\ &= 53\\ Dec(C_2) &= \alpha_2^{-1} \cdot b \cdot b^{-1} + \beta_2^{-1} \cdot a \cdot a^{-1} \bmod v\\ &= 9\times 11\times 14 + 10\times 12\times 2 \bmod 187\\ &= 43\end{aligned} \quad (15)$$

As shown previously, the final aggregation result is $m = Dec(C_1) + Dec(C_1) = 96$. After this, BS performs data integrity verification by applying MACs verifying algorithm: BS substitutes data MACs t_1, t_2 and data weight ε_1, ε_2 into signing algorithm to get the MAC of received data aggregation result m. The algorithm is as follows:

$$\begin{aligned}t' &= \left(\left(\varepsilon_1 \cdot x_1 + \varepsilon_2 \cdot x_2\right) + y * m \bmod 64\right)\\ &= \left(5\times 7 + 4\times 3\right) + (3,5) * (53,43)\\ &= 37\end{aligned} \quad (16)$$

Finally, BS compares the generated MAC t' with received MAC t, and the result is $t' = t = 37$. Therefore, BS can determine the integrity of received data is intact, and the aggregation result is accepted by the data processing center. Through this example, we can see that the HMSDA scheme supports additive homomorphic aggregation, and provides efficient data integrity verification of the aggregation results.

5 Performance Analysis

In this section, we apply Tiny OS Simulator to analyze the performance of HMSDA scheme. The comparative data aggregation schemes include SIES scheme and iCPDA scheme which are integrated with the mechanism of data integrity protection. The comparative items include communication overhead, computation load and data aggregation accuracy.

5.1 Communication Overhead

Communication overhead in the network is an important factor of sensor node lifetime. Firstly, we compare the data transfer formulas of the three schemes in which we set the number of sensor nodes in the network to N, and the sizes of the data packet transmitted in the network are the same with each other, and the communication overhead of each data packet is counted as one standard data transfer unit. As discussed previously, during the process of constructing aggregation tree and data aggregation, each sensor node transmits 2 data packets in HMSDA scheme and therefore the data transfer formula of HMSDA can be expressed as follows:

$$\text{HMSDA: } DT_H = 2N \quad (17)$$

According to the literature 0, during the generation of network topology, the amount of transmitted data packet of each node in SIES scheme is 1. In the process of data aggregation, the nodes transmit the collected target data back to the upper level nodes directly, thus the amount of data transmission of each node is 1 too. Therefore, the data transfer formula of SIES can be concluded as following:

$$\text{SIES: } DT_s = 2N \quad (18)$$

According to the literature 0, in iCPDA scheme, every node needs to send 1 data packet during the phase of network topology generation. In the stage of data aggregation, the aggregator nodes need to send 5 packets, including the random number, pseudo data packet, aggregated data packet, and so on, but the sensing nodes only need to send 2 packets, i.e., the pseudo data packet and aggregated data packet. Therefore, the amount of data transmission of aggregator node is $6N \cdot p_a$ and the amount of data transmission of sensing node is $3N \cdot (1 - p_a)$. Thus, the data transfer formula of iCPDA can be expressed as

follows:

$$\text{iCPDA: } DT_i = 6N \cdot p_a + 3N \cdot (1 - p_a) = 3N \cdot (1 + p_a) \quad \textbf{(19)}$$

By comparing the data transfer formula of these three schemes, it can be found that, theoretically, the amount of data transmission of SIES and HMSDA scheme are same, and the amount of data transmission of iCPDA scheme is larger.

In this simulation, we set the number of sensor nodes in the network to 300, the range of data aggregation time interval is 5 to 40 seconds, and the sensor nodes are densely-deployed in an area of $150m * 150m$. The data packet volumes of HMSDA, iCPDA and SIES scheme are 30 bytes, 30 bytes, 32 bytes, respectively. The probability of a sensor node becoming an aggregator node p_a is set to 0.3. HMSDA, SIES and iCPDA scheme are separately simulated under different conditions of data aggregation interval time for 10 times, and we present average simulation results as follows:

As can be seen from Figure 5, during the process of increasing data aggregation time interval, the communication overhead of each scheme has remained relatively constant, and when the time interval increases to 40 seconds, the communication overheads of the three schemes have basically reached their theoretical values. Among them, the communication overhead of iCPDA scheme has reached $3.24 * 10^4$ bytes, which is the highest, and it is followed by SIES scheme, $1.812 * 10^4$ bytes. The communication overhead of HMSDA scheme is the lowest, $1.8 * 10^4$ bytes. The simulation reveals that the communication overhead of the HMSDA scheme is lower than that of SIES and iCPDA scheme.

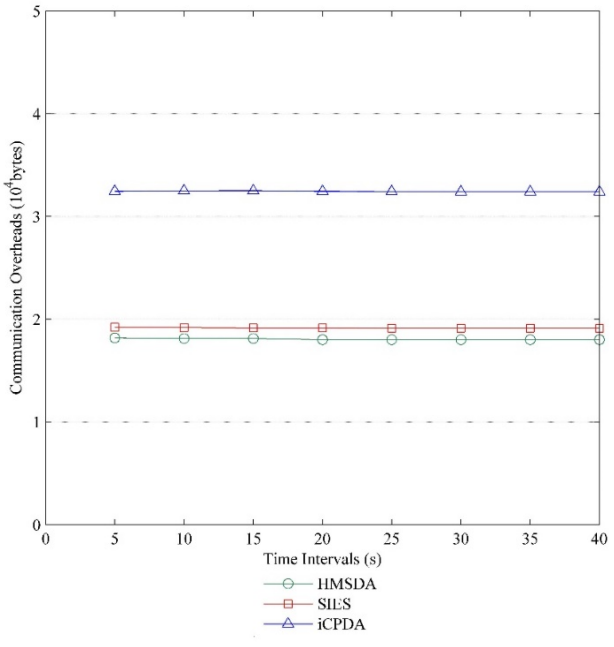


Figure 5. The comparison of communication overheads

As mentioned in previous section, the level of energy consumption of the network is proportional to its communication overhead and inversely to its lifetime. Therefore, by the theoretical and demonstration research above, the energy consumption of iCPDA scheme is the highest, which has the shortest lifetime, and the energy consumption of HMSDA scheme is the lowest, which has the longest lifetime. The level of energy consumption and network lifetime of SIES scheme is intermediate between HMSDA scheme and iCPDA scheme.

5.2 Computation Load

Considering that wireless sensor nodes are resource-constrained devices, the computation load during data aggregation has a significant influence on the lifetime and work efficiency of networks. In this part, we first carry on the comparison of the computation overhead formula of the three schemes, and then carry on the comparing simulations to the three schemes according to their computation overhead formulas. Since the three schemes all use additive data aggregation, the computation load of data aggregate operation is not included in this comparison.

According to the previous introduction, the computation load of the sensor nodes in HMSDA scheme mainly depends on the vector dimension n. Suppose that the aggregator node ratio in network is p, then the required additive computation overhead for a single aggregation of the sensing nodes is $N \cdot (1-p)(2n-1)$, multiplication-division computation overhead is $N \cdot (5n+1)(1-p)$. The required additive computation overhead for a single aggregation of the aggregator nodes is $4n \cdot Np$, and the multiplication-division computation overhead is $2n \cdot Np$. From the foregoing, the additive computation overhead formula CO_H^A and multiplication-division computation formula CO_H^M in HMSDA scheme can be concluded as follows:

$$\begin{aligned} CO_H^A &= N \cdot (1-p)(2n-1) + 4n \cdot Np \\ &= 2Np - N + 2nNp + Np \\ &= N(2n-1) + Np(2n+1) \end{aligned} \quad \textbf{(20)}$$

$$\begin{aligned} CO_H^M &= 4N \cdot Np + 2n \cdot Np \\ &= 6nNp \end{aligned} \quad \textbf{(21)}$$

In SIES scheme, sensor nodes need to perform $HM(\cdot)$ computation 3 times by applying hash function during the generation of secret key pairs. In the process of data aggregation, each node in the network needs to perform a single modular addition and a single modular multiplication, and their computation overhead can be represented by CO_{Add} and CO_{Multi}, respectively. Thus, the computation overhead formula

of SIES scheme CO_S can be expressed as follows:

$$CO_S = N \cdot \left(CO_{Add} + CO_{Multi} + 3CO_{HM} \right) \quad \textbf{(22)}$$

Obviously, the data perturbation mechanism applied in iCPDA scheme is same with that of CPDA scheme, and during a single data aggregation, non-cluster nodes need to perform modular addition 8 times, modular multiplication 9 times, data encryption and decryption twice. Compared with the non-cluster nodes, the cluster nodes need to perform one more Gauss elimination during a single data aggregation. Thus, the computation overhead formulas of non-cluster nodes and cluster nodes in iCPDA scheme CO_i^S and CO_i^A can be expressed as follows:

$$\begin{aligned} CO_i^S &= N \cdot \left(1 - p_a\right) \cdot \\ &\left(8CO_{Add} + 9CO_{Multi} + 2CO_{Enc} + 2CO_{Dec}\right) \end{aligned} \quad \textbf{(23)}$$

$$CO_i^A = N \cdot p_a \left(CO_i^S + CO_G \right) \quad \textbf{(24)}$$

The comparison of the computation overhead formulas of the three schemes reveals that the theoretical computation load of iCPDA scheme is the highest, and the theoretical computation load of SIES and HMSDA scheme are very close when vector dimension $n \le 3$. In the simulation analysis, the time consumed by a single data aggregation is regarded as the comparison norm. The p_a in iCPDA scheme is set to 0.3, and the p in HMSDA scheme is set to 0.3 too. In order to reflect the relationship between the computation load and security performance, the three schemes are separately simulated under different conditions of vector dimension n for 10 times, and the range of n is $[3,12]$. We took the average value of elapsed time as the simulation result which are shown below:

As can be seen from Figure 6, the computation load of iCPDA scheme is the highest, a single data aggregate operation requires $2.33 * 10^4$ ms. The computation load of HMSDA and SIES scheme are very close when the vector dimension n is small. But along with the increase of vector dimension n, the computation load of HMSDA scheme starts to rise slowly. A single data aggregate operation in HMSDA scheme requires $0.81 * 10^4$ ms when $n = 8$, and the single aggregation period of SIES had hold mostly steady around $0.54 * 10^4$ ms. Through the simulation results, we can see that the computation load of HMSDA scheme is obviously lower than that of iCPDA, and slightly higher than that of SIES scheme.

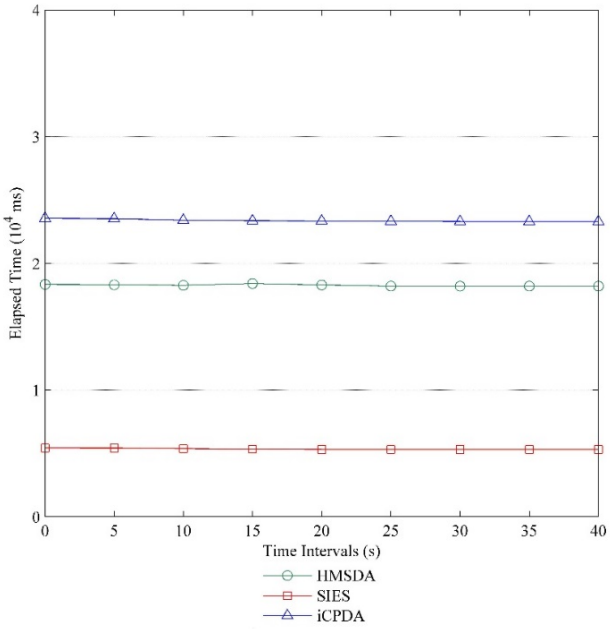


Figure 6. The comparison of computation load

5.3 The Accuracy of Data Aggregation

Data aggregation accuracy is one of the most important performance indicators of data aggregation scheme. It is measured by the actual data aggregation results to the theoretical data aggregation results. In ideal status, the accuracy of data aggregation can reach 100%, but in practical terms, the process of data aggregation will inevitably be impacted by the negative factors like transmission collisions, BERT, energy attenuation, data packet loss, and so on. Therefore, the actual data aggregation accuracy cannot reach the theoretical level.

In this section, we set the range of data aggregation time interval from 0 to 40 seconds, and the probability of a sensor node becoming an aggregator node p_a is 0.3. HMSDA, SIES and iCPDA scheme are separately simulated under different condition of data aggregation interval time for 10 times, and we took the average value of data aggregation accuracy as the simulation result which is shown as follows:

As we can see in Figure 7, with the increase of the data aggregation time interval, data aggregation accuracy of the three schemes have different degrees of improvement. That is because the increase of time interval reduced the amount of data transmission in the network per unit time, thus lowered the probability of transmission collision, packet loss, these factors eventually led to the rise of the data aggregation accuracy.

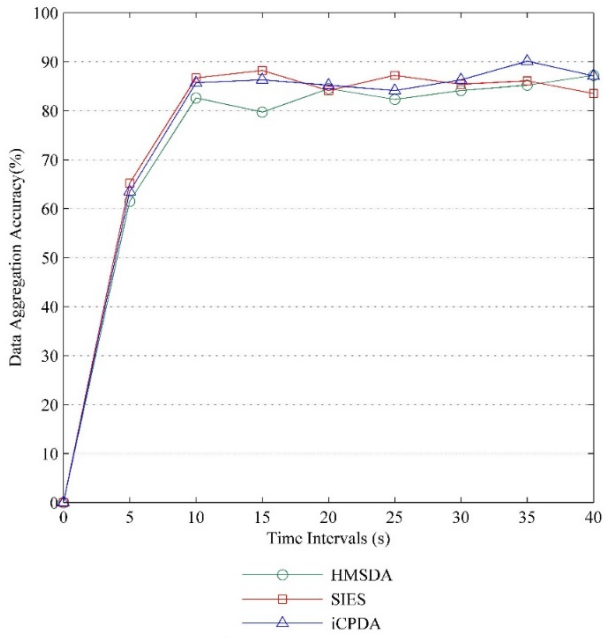


Figure 7. The comparison of data aggregation accuracy

In HMSDA and the SIES scheme, the growth trends of data aggregation accuracy are similar, HMSDA scheme has reached 88.1%, SIES can reach up to 84.3%. The growth rate of data aggregation accuracy of iCPDA scheme is relatively slow, and finally reach 79.1%.The difference of data aggregation accuracy of the three schemes is mainly caused by the amount of data transmission in the process of data aggregation. Through the above simulation results, we can see that the data aggregation accuracy of HMSDA scheme is better than SIES and iCPDA scheme.

6 Conclusions

In this paper, we analyze the existing secure data aggregation schemes which take data integrity and confidentiality into consideration and further propose a secure data aggregation scheme HMSDA based on homomorphic MAC and encryption technique. Based on the realization of end-to-end security data aggregation, the scheme provides a complete data integrity verification mechanism. As can be seen in the stage of performance analysis, compared with iCPDA and SIES scheme, HMSDA scheme has obvious advantages in data aggregation accuracy, computation load and communication overhead. In conclusion, the scheme meets the design requirements of this paper.

Acknowledgements

This research was partly supported by National Natural Science Foundation of China Under Grants 61071076, National High-tech Research And Development Plans (863 Program) Under Grants 2011AA010104-2, The Academic Discipline and Postgraduate Education Project of Beijing Municipal Commission of Education.

References

[1] Z. J. Zhang, C. F. Lai, H. C. Chao, A Green Data Transmission Mechanism for Wireless Multimedia Sensor Networks using Information Fusion, *IEEE Wireless Communications*, Vol. 21, No. 4, pp. 14-19, August, 2014.

[2] N. Xu, S. Rangwala, K. K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, D. Estrin, A Wireless Sensor Network for Structural Monitoring, *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, Baltimore, MD, 2004, pp. 13-24.

[3] T. Ko, J. Hyman, E. Graham, M. Hansen, S. Soatto, D. Estrin, Embedded Imagers: Detecting, Localizing, and Recognizing Objects and Events in Natural Habitats, *Proceedings of the IEEE*, Vol. 98, No. 11, pp. 1934-1946, November, 2010.

[4] G. Simon, M. Maróti, Á. Lédeczi, G. Balogh, B. Kusy, A. Nádas, G. Pap, J. Sallai, K. Frampton, Sensor Network-based Countersniper System, *Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems*, Baltimore, MD, 2004, pp. 1-12.

[5] J. Polastre, R. Szewczyk, A. Mainwaring, D. Culler, J. Anderson, Analysis of Wireless Sensor Networks for Habitat Monitoring, in: C. S. Raghavendra, K. M. Sivalingam, T. Znati (Eds.), *Wireless Sensor Networks*, Springer US, 2004, pp. 393-423.

[6] E. Cayirci, T. Coplu, SENDROM: Sensor Networks for Disaster Relief Operations Management, *Wireless Networks*, Vol. 13, No. 3, pp. 409-423, June, 2007.

[7] A. Milenković, C. Otto, E. Jovanov, Wireless Sensor Networks for Personal Health Monitoring: Issues and an Implementation, *Computer Communications*, Vol. 29, No. 13, pp. 2521-2533, August, 2006.

[8] C. Li, Z. Zhang, F. Xiong, Q. Liu, An Efficient and Stable Route Protocol in Wearable Body Networks, *2015 First International Conference on Computational Intelligence Theory, Systems and Applications (CCITSA)*, Yilan, Taiwan, 2015, pp. 104-109.

[9] W. Zhang, Z. Zhang, Belief Function Based Decision Fusion for Decentralized Target Classification in Wireless Sensor Networks, *Sensors*, Vol. 15, No. 8, pp. 20524-20540, August, 2015.

[10] Z. Zhang, W. Zhang, H. C. Chao, C. F. Lai, Toward Belief Function-based Cooperative Sensing for Interference Resistant Industrial Wireless Sensor Networks, *IEEE Transactions on Industrial Informatics*, Vol. 12, No. 6, pp. 2115-2126, December, 2016.

[11] C. Castelluccia, E. Mykletun, G. Tsudik, Efficient Aggregation of Encrypted Data in Wireless Sensor Networks, *Proceedings of the 2nd IEEE Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*, San Diego, CA, 2005, pp. 109-117.

[12] C. Castelluccia, Securing Very Dynamic Groups and Data Aggregation in Wireless Sensor Networks, *IEEE International Conference on Mobile Adhoc and Sensor*

Systems (MASS 2007), Pisa, Italy, 2007, pp. 1-9.

[13] J. M. Bahi, C. Guyeux, A. Makhoul, Efficient and Robust Secure Aggregation of Encrypted Data in Sensor Networks, *2010 Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM)*, Venice, Italy, 2010, pp. 472-477.

[14] W. He, X. Liu, H. Nguyen, K. Nahrstedt, A Cluster-based Protocol to Enforce Integrity and Preserve Privacy in Data Aggregation, *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops*, Montreal, Canada, 2009, pp. 14-19.

[15] S. Papadopoulos, A. Kiayias, D. Papadias, Exact In-Network Aggregation with Integrity and Confidentiality, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 24, No. 10, pp. 1760-1773, October, 2012.

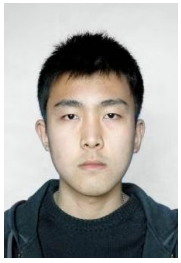
[16] S. Ozdemir, H. Çam, Integration of False Data Detection with Data Aggregation and Confidential Transmission in Wireless Sensor Networks, *IEEE/ACM Transactions on Networking (TON)*, Vol. 18, No. 3, pp. 736-749, June, 2010.

[17] S. Agrawal, D. Boneh, Homomorphic MACs: MAC-based Integrity for Network Coding, *Proceedings of the 7th International Conference on Applied Cryptography and Network Security*, Paris, France, 2009, pp. 292-305.

[18] S. Madden, M. J. Franklin, J. M. Hellerstein, W. Hong, TAG: A Tiny Aggregation Service for Ad-hoc Sensor Networks, *Proceedings of the 5th Symposium on Operating Systems Design and Implementation*, Boston, MA, 2002, pp. 131-146.

Biographies

Yun Liu is a Professor of Communication and Information Systems, Beijing Jiaotong University; Dean of Communication Engineering Department, Beijing Jiaotong University; Director of key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education; Director of Institute of Network Consensus Security, Beijing Jiao tong University; Vice-chair of Teachers and Staff Representative Committee, Beijing Jiaotong University. She is currently a Fellow of IET, UK and specialist enjoying special government allowance. In addition, she is an evaluation expert of State Scientific and Technological reward, State Natural Sciences Fund in communication, National High Technology Research and Develop men Program (HTRDP), an advanced counselor of China Tietong Company, and an advanced counselor of Beijing Municipal Office of Internet Propaganda and Management.



Chaoran Li received the Ph.D. degree in communication and information system from Beijing Jiaotong University, Beijing, China, in 2016. He is currently a system engineer for Space Star Technology Co., Ltd. His current research interests include the areas of secure data aggregation, complex networks, machine learning.

Jing Zhang received the bachelor's degree in control technology and instrument professional in 2005 and the master's degree in communication and information system from the Guilin University of Electronic Technology and the Ph.D. degree from the Beijing University of Posts and Telecommunications in 2011. Since 2011, he has been with the 54th Research Institute of CETC. His research interests are satellite communication and cognitive radio.

Qing Liu, got her master's degree in computer technology from University of Science & Technology Beijing, China, in 2009. Her currently work focuses on the related fields of information security and the cloud computing for the Power Construction Corporation of China.