Liu *et al. EURASIP Journal on Wireless Communications and Networking* (2020) 2020:19
https://doi.org/10.1186/s13638-020-1643-6

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Energy-efficient privacy-preserving data aggregation protocols based on slicing

Xiaowu Liu[1*], Jiguo Yu[2], Xiaowei Zhang[1], Qiang Zhang[1] and Can Fu[1]

Abstract

Wireless sensor networks (WSNs) have become one of the most vigorous techniques in the network domain. However, the sensor nodes of WSNs tend to become the target of attackers due to the broadcast communication mode and the unattended deployment nature. Although it can prevent the sensitive data from being compromised, Slice-Mix-AggRegaTe (SMART) needs to exchange messages frequently in a network, which put tremendous overhead on the sensor nodes with limited resources. Faced with these issues, this paper proposes an energy-efficient privacy-preserving data aggregation protocol based on slicing (EPPA) where a novel slicing mode is adopted to reduce the numbers of slices, which can significantly prevent the data from being compromised and decrease the communication overhead. Meanwhile, an enhanced scheme based on EPPA, called multi-function privacy-preserving data aggregation protocol (MPPA), is presented and it supports multiple functions in the process of data aggregation, such as max/min, count, and mean. The theoretical analysis and the simulation evaluation show that the proposed aggregation protocols demonstrate a better performance in the privacy preserving and the communication efficiency.

Keywords: Privacy preserving, Data aggregation, Data slicing, Euclidean-based decomposition

1 Introduction

Nowadays, wireless sensor networks (WSNs), seen as a popular technique, are applied into various applications, such as environmental monitoring, smart home, Internet of Things, and military battlefields [1, 2]. Many new challenges have emerged with the development of artificial intelligence and big data [3–5]. WSNs are formed by a large number of resource-constrained sensor nodes. It is almost impractical and uneconomical for each node to send its sensing data directly to base station (BS) [6–8], because the energy of the node will be exhausted in the process of data transmission and the battery capacity of the sensor node cannot meet the requirement of network application. The energy issue has become a major concern in both industrial practice and academic world [9–11]. Data aggregation (DA) [12–15] technique which is one of the essential techniques in assuring the effectiveness of WSNs can effectively overcome the energy obstacle by fusing data and decreasing redundancy in many critical applications.

Most of the nodes of a WSN are deployed in an unattended physical environment and can perceive the data from the surrounding environment, mix the data through DA technology, and then transmit them to BS through a secure channel. However, WSNs with DA may produce more secure vulnerabilities than the ones without DA technique [16]. An aggregation node (AN) in a WSN which stores a lot of perceptual data is vulnerable to attack. If ANs are compromised successfully by an adversary, the sensitive data of the whole network may be revealed. Therefore, more challenges are emerging in privacy preserving and energy effectiveness of WSNs with DA.

Although a few DA privacy-preserving protocols have been proposed in recent years [17, 18], many challenging issues remain to be conquered in resource-constrained WSNs. A cluster-based private data aggregation (CPDA) scheme was proposed [19]. In CPDA, the sensor nodes are randomly organized into clusters and all the nodes calculate aggregation values according to the algebraic properties of polynomials. This protocol ensures that private data cannot be revealed by compromising a single node. However, the complex computation of CPDA puts a huge burden on node resource. For this reason, the

*Correspondence: liuxw@qfnu.edu.cn
[1]School of Information Scicence and Engineering, Qufu Normal University, Yantai Road, Rizhao, 276826 China
Full list of author information is available at the end of the article

Slice-Mix-AggRegaTe (SMART) protocol was proposed to decrease computation overhead and promote the capacity of privacy preserving [19]. In SMART, each node hides the sensing data by cutting them into slices. After being encrypted, each slice is sent to different intermediate aggregation nodes or neighboring nodes. When an intermediate node receives the slices sent by its neighbors, the aggregation process is executed and further aggregation may be performed in the next hop node until BS arrives. SMART guarantees data privacy in the process of aggregation and the computation complexity is decreased compared with CPDA. However, it requires more communication consumption which shortens the lifetime of network and degrades the effectiveness of sensor node. Faced with this dilemma, Hua et al. proposed an energy-efficient adaptive slice-based secure data aggregation (ASSDA) scheme [20]. In ASSDA, the same data slice of a node can be sent once only in the same time slot. All the data of leaf node are sliced according to the transmitting distance and the number of receivers. ASSDA improves the efficiency of data slicing and reduces the energy consumption of nodes.

This paper presents an energy-efficient privacy-preserving data aggregation protocol based on slicing (EPPA) which adopts data decomposition to reduce the number of slices and significantly save communication costs. It is worth noting that the decomposition can protect the private data from being destroyed by collusion attack. In EPPA, only the addition aggregation function is considered and the non-addition aggregation functions are the challenging work in the current research. As an improvement to EPPA, we propose an enhanced EPPA protocol, called multi-function privacy preserving data aggregation (MPPA), which focuses on multiple aggregation functions and may improve the universality of privacy-preserving technique in WSNs.

This paper is organized as follows. Section 2 reviews the related works. Section 3 discusses the system model. Section 4 and Section 5 elaborate EPPA and MPPA in details, respectively. Section 6 evaluates the performance of our protocols. We conclude this paper in Section 7.

2 Related work

DA has become one of the most effective methods to decrease the system overhead by reducing the bandwidth occupation and improving the energy efficiency in WSNs [21]. However, the broadcast communication mode and the insecure deployment pattern of WSNs make DA a challenging work when various attacks emerge in a network. These issues have attracted more attention than ever before in both WSNs and other fields [22–25]. Many mechanisms have been verified to prevent DA from being compromised, and some privacy-preserving protocols with DA have been proposed to guarantee the privacy of sensing data. SMART is the pioneering method in slicing-based data aggregation protocols.

As the name suggests, "Slice-Mix-AggRegaTe (SMART)" is a three-step scheme for the privacy preserving of DA. It divides the sensing data into many slices and sends each slice to different destination nodes in order to hide the sensitive information in varied slices which can provide a better privacy protection for WSNs. Many researches have proved that SMART can protect the data security at a little cost of communication bandwidth occupation [19, 20]. The workflow of SMART is described as follows.

- Step 1: Slicing. For each node $i(i = 1, ..., N)$, its private data d_i can be cut into j pieces. Firstly, it randomly selects a set of nodes $J(|J| = j - 1)$ within h hops. A dense WSN takes $h = 1$, namely, node i selects one-hop node as its neighbors. Then, one of the pieces is kept in node i itself. The remaining $(j - 1)$ pieces are encrypted and sent to the nodes in set J. In general, SMART is a fixed slicing scheme which divides the private data into three slices. Take Fig. 1 as an instance, node 5 divides its data into three slices, m_{51}, m_{45}, and m_{55}. One slice (m_{55}) is reserved by node 5 itself and two slices (m_{51} and m_{45}) are sent to node 1 and node 4, respectively. We expand the idea of SMART in Fig. 1 by claiming that some nodes may cut their sensing data to more than (node 6 and node 7) or less than (node 9) three pieces. We hope to illustrate that three pieces are only one of the choices for slicing mechanism.

- Step 2: Mixing. When a node receives an encrypted slice from node i, it firstly decrypts the data using the shared key with node i. Then, it sums up all the received slices sent by other neighbors as shown in Fig. 2.

- Step 3: Aggregation. Each AN aggregates the received slices to a single data packet and transmits it to the upstream nodes following the routing path in an aggregation tree until BS arrives as shown in Fig. 3.

Although the SMART algorithm has a good performance in the protection of data privacy, there are still some limitations that affect its practicality. (1) The communication overhead is high. The communication overhead is directly related to the number of slices. In SMART, all the nodes are cut into j slices. Therefore, the total number of slices of a WSN with N nodes is $N*j$, where N represents the number of nodes. Transmitting too many slices will consume huge amounts of energy and decrease the system performance in terms of effectiveness and lifetime.

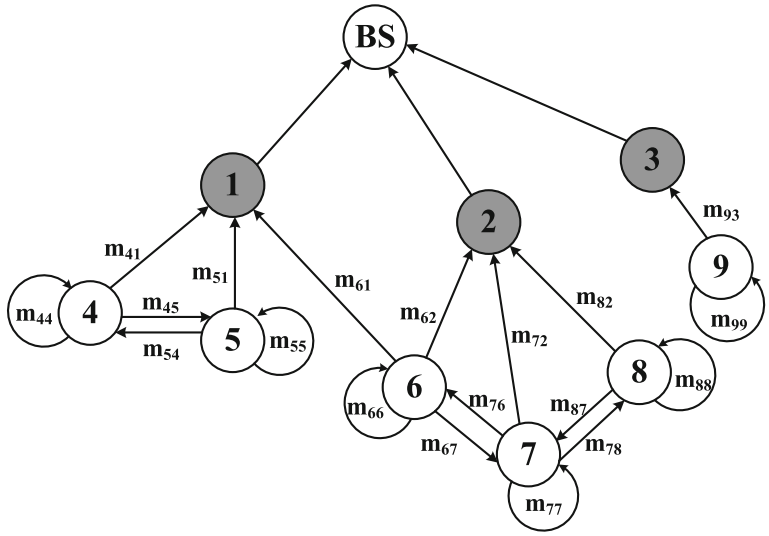


Fig. 1 Data slicing

(2) The probability of collision may arise. More slices may produce more packet transmissions which may result in the higher collision probability. This is an inevitable issue in SMART. Ultimately, this may affect many network parameters, such as delay, transmission error, accuracy, and efficiency [26–28].

According to the abovementioned limitations, SMART algorithm needs to be improved and challenging issues have been tackled by many contributions. He et al. proposed a SMART-based addition aggregation which was a fixed slicing scheme [19], and each node divided its raw data to J slices (J is a predefined number) and $(J-1)$ slices are sent to its neighboring nodes and one slice is held in itself. After all the nodes have received the data slices from their neighboring nodes, these slices were mixed and the new result is transferred to an upstream AN. In this way, attackers only eavesdrop on several incomplete data slices rather than the whole data packet. Although this scheme preserves the private data effectively, it results in a large number of message exchanging. In order to reduce the amount of information transmission, Li et al. proposed an energy-efficient and high-accuracy scheme for secure data aggregation (EEHA) [29]. There are different operations for two types of nodes that are leaf nodes and aggregation nodes. The former executes slicing and mixing operations to preserve the data privacy, while the latter is not involved in slicing and is only responsible for aggregating its private data with the slices received

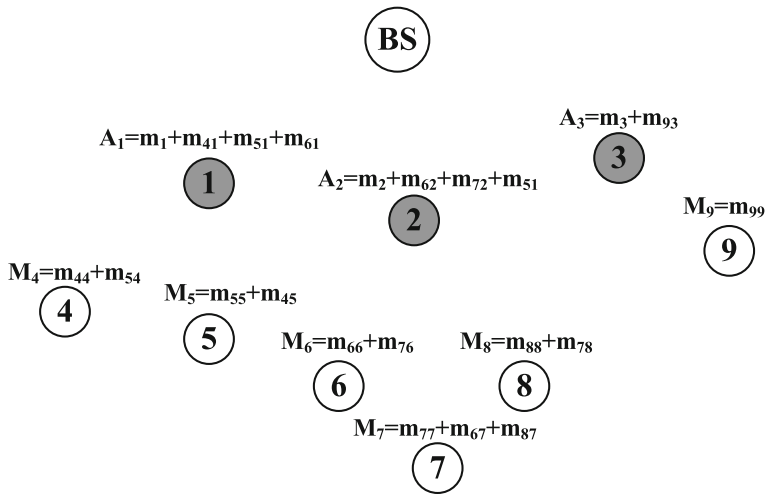


Fig. 2 Data mixing

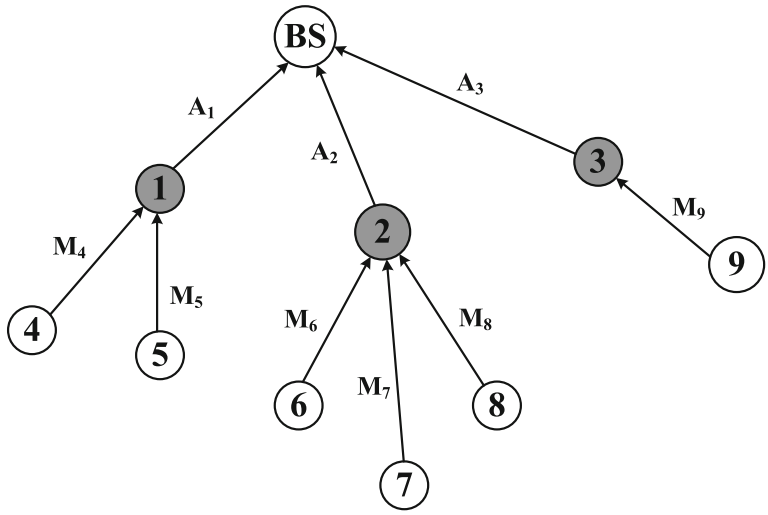


Fig. 3 Data aggregation

from leaf nodes into a new aggregated data. Compared with SMART, EEHA can dramatically reduce the communication overhead and increase the accuracy of DA at a tiny price of privacy preserving. In addition, Liu et al. proposed a high energy-efficient and privacy preserving (HEEPP) [30]. It introduced random distribution into the slicing and mixing techniques to determine the number of data slices, which resulted in the lower energy consumption and the higher security of DA. However, both EEHA and HEEPP ignore the privacy preserving in case of a collusion attack. If collusion occurs, malicious sensor nodes can easily compromise the security of network to an unaccepted level. Furthermore, all the data stored in the primary node (aggregation node) may be disclosed to adversaries, which may cause a huge damage to the entire network.

Based on the abovementioned issues, we present an EPPA protocol which uses the data decomposition technique to overcome the limitations of traditional slicing mechanisms. It can not only conceal the raw data through slicing, but also decrease the number of slices at the same time. The EPPA scheme is also capable of dividing sensitive data into two pieces by extended Euclidean, which may hide the original data well and effectively defend against collusion attacks. Besides, another enhanced protocol, MPPA, is proposed, which supports multiple aggregation functions and expands the adaptability of our scheme in a practical WSN-based system.

3 System models

In this section, we present the topology, the adversary model, and the design requirements of network.

3.1 Network model

In this paper, we employ the tree topolgy to organize sensor nodes for DA in WSNs, as shown in Fig. 4. Generally speaking, there are three types of nodes in a DA protocol: base station (BS), aggregation node (AN), and leaf node (LN). BS is trusted and has powerful computation and storage capacities. There is only one BS in a WSN. BS is regarded as the control center of the network and the final destination of aggregation result. As the root of the aggregation tree, it is responsible for broadcasting the query to other nodes and processing the aggregation results received from ANs. ANs are in charge of forwarding the query instructions sent by BS, collecting and aggregating the data from their child nodes. LN is used for data acquiring, data slicing, and data mixing. It cuts the raw data into slices, sends the slices to the neighboring nodes, and assembles the received slices to compute an intermediate mixing value. Assumed that the network size is N and each node is assigned different roles, such as BS, AN, and LN. N nodes are organized into an aggregation tree (the formation process will be described in Section 5), and the data are transferred from LN to BS along the tree.

3.2 Adversary model

Because the nodes of WSN are usually deployed in an unattended environment, the attackers can easily launch multiple attacks to destroy the privacy and integrity of aggregation results. Data packet is transmitted through the radio communication channel in a WSN. Therefore, it is easy to obtain the private information through overhearing the transmission from its neighboring wireless links. Each node can monitor the sending packet in the radio coverage and the private data may be exposed to the attackers due to the open transmission mode.

3.3 Design goal

The design goal of our scheme is to use an extended Euclidean method to achieve lower communication overhead and prolong the lifetime of network. Meanwhile, the

proposed scheme is expected to guarantee the security of private data and defend against both the collusion attack and the eavesdropping attack. Therefore, an ideal data aggregation scheme should meet the following conditions.

- Privacy-preservation: In the design of a secure data aggregation scheme, the preservation of private data is always a key security issue. Wireless links are vulnerable to the eavesdropping attack and the sensitive information is usually revealed to undesired neighboring nodes. The privacy leakage limits the application of WSNs and prevents WSNs from being applied into some critical industrial fields in which data privacy is regarded as one of the most important attributes. In order to broaden the application field of WSNs, it is necessary to ensure data privacy and network efficiency in a proposed protocol. Meanwhile, the data privacy aggregation scheme should be able to run in a reasonable way even in case of collusion attack.
- Efficiency: Sensor nodes in a network are usually equipped with irreplaceable or non-repeatable charging battery. Therefore, energy consumption is a critical issue since the emerging of WSNs, and it is also a core issue in the design of secure data aggregation protocol. Information exchange among nodes consumes most of the energy in a WSN. Therefore, how to decrease the traffic without sacrificing the function of network becomes an interesting challenge in the current study. Data aggregation is an important energy-saving technique which can reduce the data transfer as much as possible through intra-network fusion. Therefore, the effectiveness of network is promoted and the lifetime of network is prolonged.

4 Energy-efficient privacy-preserving data aggregation protocol based on smart

This section introduces an energy-efficient privacy-preserving data aggregation (EPPA) protocol based on slicing in details. EPPA consists of four steps: the aggregation tree construction, the data slicing, the data mixing, and the data aggregation. Depending on the data slicing technique, EPPA can improve the performance of privacy preserving, decrease the energy consumption, and prolong the network lifetime of WSN compared with SMART.

4.1 Aggregation tree construction

A common technique for DA is to build an aggregation tree which is formed by all the aggregation paths from sensor nodes to BS. The aggregation tree construction process is illustrated in Fig. 5. Firstly, the BS triggers a query through broadcasting a "HELLO" message as shown in Fig. 5a. When the message is received, a sensor node elects itself as AN with a probability p_c, which is a preselected

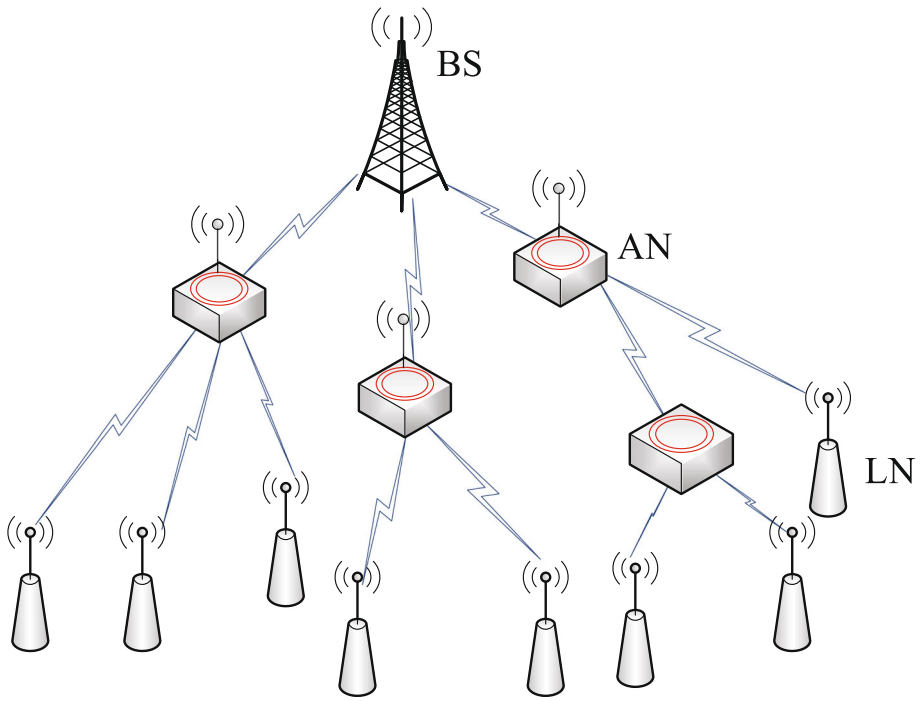
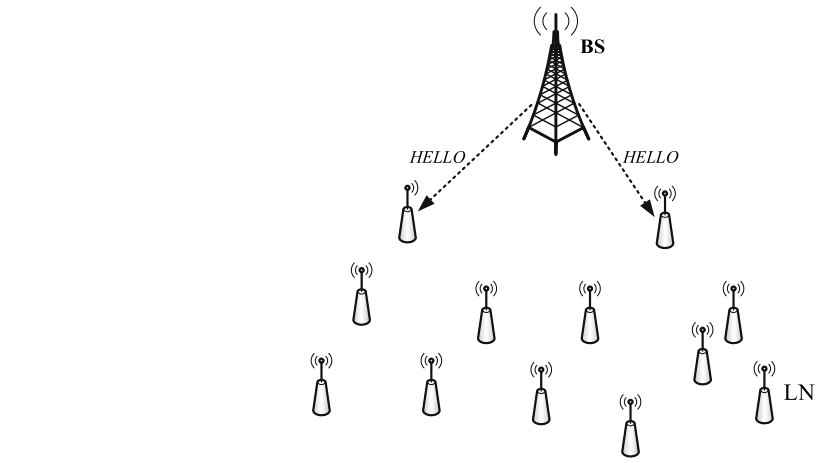
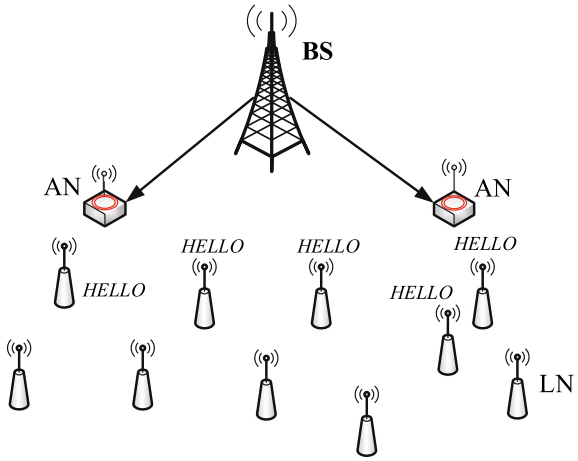


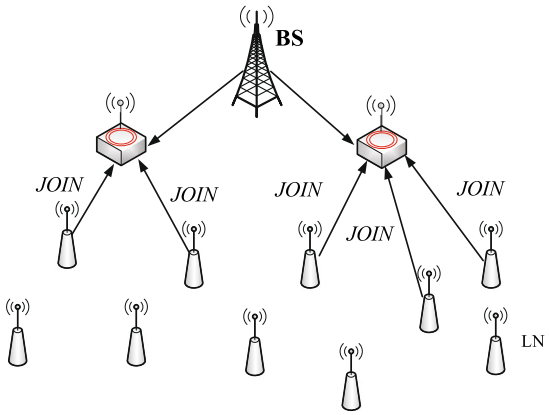
Fig. 4 Aggregation tree



(a) The BS triggers a "Hello" message.



(b) AN broadcasts the "HELLO" message to its neighboring nodes.



(c) A node becomes a children of AN.

Fig. 5 The aggregation tree construction

parameter for all the nodes. If a node becomes AN, it will forward the "HELLO" message to its neighboring nodes as demonstrated in Fig. 5b. Otherwise, the node will wait for a certain period of time to get other "HELLO" messages from its neighboring nodes. If a node receives multiple "HELLO" messages, it randomly selects an upstream node as its parent by sending a "JOIN" message as shown in Fig. 5c. This process is recursively executed until all the nodes join in the tree.

4.2 Data slicing

We use the extended Euclidean-based decomposition approach [31, 32] to optimize the slicing technique in this subsection. The original data of each node are decomposed into two parts, one is reserved by the node itself and the other is sent to its neighboring node. We will demonstrate the extended Euclidean theorem and the feasibility proof as follows.

Theorem 1 (extended Euclidean): *Two numbers a and b are non-negative and non-zero integer. $gcd(a, b)$ denotes the greatest common divisor of a and b . Therefore, there are X and Y which satisfy $gcd(a, b) = aX + bY$.*

Proof The induction method is adopted to prove the theorem. Three cases are considered.

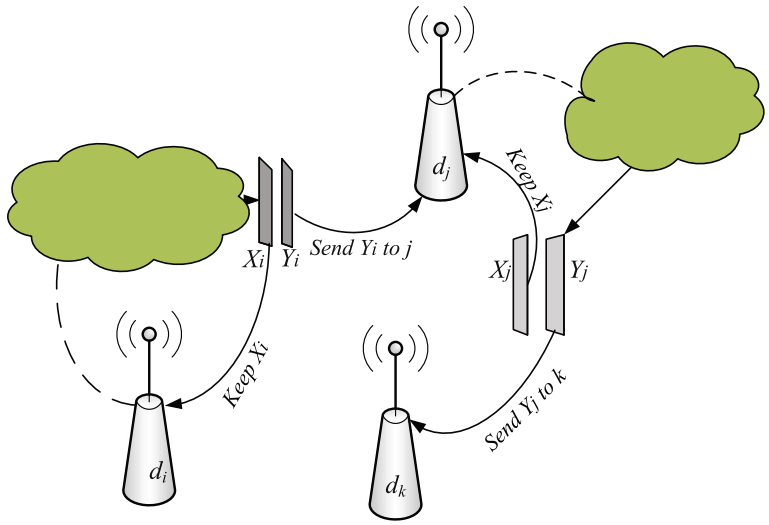


Fig. 6 Data slicing

(1) If $b = 0$, then $gcd(a, b) = gcd(a, 0) = a$. Thus, $X = 1$ and Y is an arbitrary value.

(2) If $a = 0$, then $gcd(a, b) = gcd(0, b) = b$. Then, X is an arbitrary value and $Y = 1$.

(3) If $a \neq 0$ and $b \neq 0$. Assumed that $b*X_1+a\%b*Y_1 = gcd(b, a\%b)$ has solutions X_1 and Y_1. Let $a = k*b+c$ then

(i)

$$\begin{aligned} gcd(b, a\%b) &= b*X_1 + a\%b*Y_1 = b*X_1 \\ &+ (a - k*b)*Y_1 \\ &= b*X_1 + a*Y_1 - k*b*Y_1 \\ &= a*Y_1 + b*(X_1 - k*Y_1). \end{aligned} \quad (1)$$

(ii) $gcd(b, a\%b) = gcd(a, b)$

(iii) $a*Y_1 + b*(X_1 - k*Y_1) = gcd(a, b)$.

Therefore, $X = Y_1$ and $Y = X_1 - k*Y_1$. □

According to the extended Euclidean theorem, the definition of slicing scheme is given as follows.

Definition 1 *For the sensing data of node i,* $d_i (i = 1, 2, ..., N)$*, let* $d_i = a, d_i + 1 = b$*, thus*

$$aX_i+bY_i=gcd(a,b) \rightarrow d_iX_i+(d_i+1)Y_i=gcd(d_i,d_i+1).$$

Therefore, raw data d_i can be decomposed into two slices, X_i and Y_i. Because $gcd(d_i, d_i + 1) = 1$, BS can recover the original data using Eq. (2) according to the extended Euclidean algorithm after all the slices are received.

$$d_i = \frac{(1 - Y_i)}{(X_i + Y_i)} \quad (2)$$

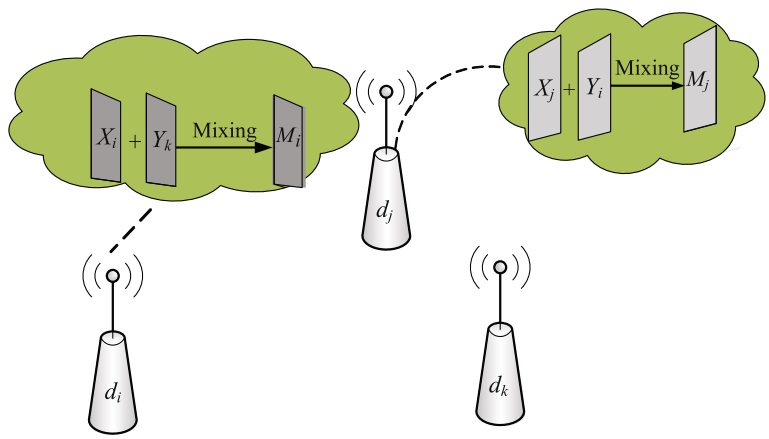


Fig. 7 Data mixing

Node i keeps slice X_i and sends slice Y_i to one of its neighboring node in order to hide the raw data. Figure 6 demonstrates the process of slicing based on the extended Euclidean decomposition.

As shown in Fig. 6, d_i and d_j are the sensing data of node i and node j, respectively. According to the extended Euclidean theorem, d_i is divided into X_i and Y_i; d_j is divided into X_j and Y_j. Node i and node j keep X_i and X_j for themselves, respectively. Slice Y_i is sent to node j, and slice Y_j is sent to node k.

4.3 Data mixing

In order to guarantee all the slices are received, the nodes wait for a certain time slot. Then, each node sums up all the received slices with their own data to get a new packet using Eq. (3). Figure 7 shows the mixing process of sensor nodes.

$$M_i = \sum_{\{j|j\in N, j\neq i\}} Y_j + X_i \tag{3}$$

ALGORITHM I. EPPA DETAILS

Input: the number of nodes, N; probability p_C
Output: aggregation data $\sum M_i$

1. **Begin**
2. BS broadcasts*HELLO*to construct an aggregation tree
3. **For**$(i = 1; i <= N; i++)$
4. **If** i becomes AN
5. forwards *HELLO*
6. **Else** send JOIN to AN
7. **End If**
8. **End For**
9. **For**$(i = 1; i <= N; i++)$
10. let $d_i = a, d_i + 1 = b$
11. According to $d_i X_i + (d_i + 1)Y_i = gcd(d_i, d_i + 1)$
12. d_i is sliced to X_i and Y_i
13. **If** $(j! = i)$
14. Y_i is sent to j
15. X_i is kept in i
16. **End If**
17. **End For**
18. **For**$(i = 1; i <= N; i++)$
19. $M_i = \sum_{\{j|j\in N, j\neq i\}} Y_j + X_i$
20. **End For**
21. AN aggregates slices to $\sum M_i$
22. **Return** $\sum M_i$
23. **End**

As shown in Fig. 7, the mixed result obtained by node j is $M_j = X_j + Y_i$.

4.4 Data aggregation

After slices are summed up, the node encrypts the new results and sends them to its parent. Each AN aggregates all the data received from its children nodes. Then, the aggregation results are transmitted to BS along the path in the aggregation tree as shown in Fig. 4.

The pseudo-code of EPPA is described in Algorithm I. Lines 1–8 describe the construction of an aggregation tree. The slicing mechanism is listed from Line 9 to Line 17. Firstly, the original data are sliced into two pieces by the proposed scheme. Then, one slice is sent to the neighboring node and the other is saved by the node itself.

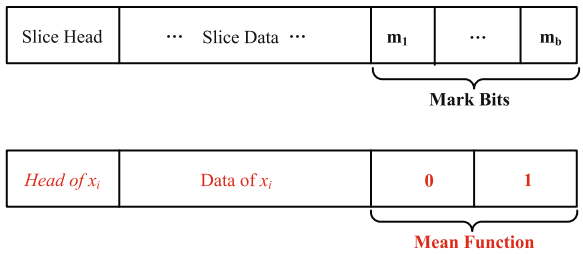


Fig. 8 Mark bits

5 Multi-function privacy-preserving data aggregation protocol

In the previous discussion, EPPA focuses on the addition aggregation function and neglects other aggregation functions, which means that its applications are restricted to the scenarios where sensing data can only be added or summed. Therefore, the enhanced EPPA, multi-function privacy-preserving data aggregation protocol (MPPA) will be discussed in this section. MPPA supports multiple aggregation functions and significantly improves the universality of slicing for WSNs.

Based on EPPA, MPPA adds some mark bits to the original slices in order to distinguish the types of aggregation functions, as shown in Fig. 8. Firstly, BS determines how many functions should be provided before a network is deployed. Secondly, it calculates the length of mark bits m according to Eq. (4).

$$\min\{2^b = f\} \rightarrow m = \lceil log_2 f \rceil \tag{4}$$

where f represents the number of aggregation functions. Thirdly, BS negotiates about the correspondences between mark bits and functions with aggregation nodes. In the data slicing process, nodes insert mark bits for each slice according to the aggregation requirement of BS. We

Table 1 Correspondence table

Flags	Functions
00	Addition
01	Mean
10	Count
11	Not used

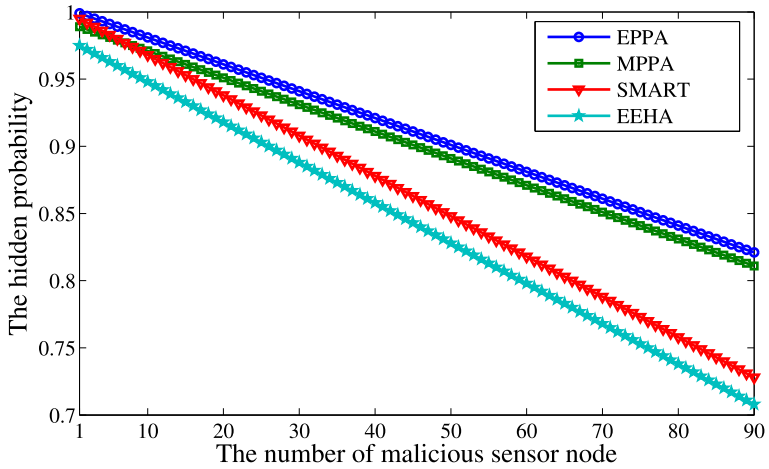


Fig. 9 Hidden probability with malicious sensor nodes

take three aggregation functions as an example (e.g., $f = 3$ in Eq. 4).

- According to Eq. (4), BS computes the number of mark bits $b = \lceil \log_2 3 \rceil = 2$. In other words, we need two bits to satisfy the requirements of three aggregation functions of BS.
- BS establishes the correspondence between mark bits and aggregation functions, as shown in Table 1. Notice that the combination "11" is left for future use if we hope MPPA to support more than three aggregation functions.
- A node selects mark bits and tags them into a slice based on the requirement of BS as formatted in Fig. 8. For example, the mark bits should be "01" if BS hopes to calculate the mean of sensing data in its query (as shown in Fig. 8).

The pseudo-code of mark bit is described in Algorithm II in which the process of adding a flag to a data slice is displayed. The slicing, mixing, and aggregation processes in MPPA are the same as those in Algorithm I of EPPA. The difference is that the data aggregation function is executed according to different mark bits.

6 Simulation results and discussion

We evaluate EPPA and MPPA through the theoretical analysis and the simulation experiment. We also compare our schemes with two typical protocols (EEHA and SMART) in terms of privacy preservation, communication overhead, and network lifetime. It should be pointed out that the number of slices is three ($J = 3$) in SMART and EEHA.

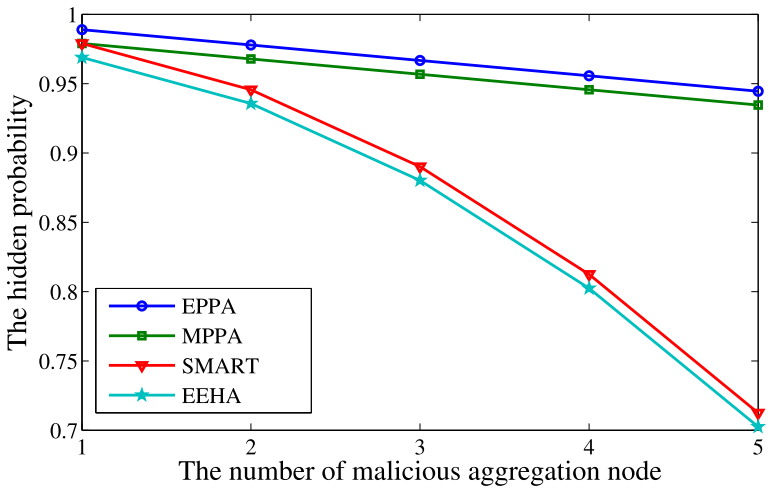


Fig. 10 Hidden probability with malicious ANs

ALGORITHM II. MARK BIT OF MPPA

Input: the number of function, f
Output: SLICING WITH FLAG
1. **Begin**
2. BS computes flags bits $m = \lceil log_2 f \rceil$
3. sends the table r to each aggregation node
4. **For**$(i = 1; i <= N; i++)$
5. selects flag according to the query of BS
6. adds the flag to slices, X_i and Y_i
7. **Return** X_i and Y_i
8. **End For**
9. **End**

6.1 Privacy-preserving analysis

This paper conducted an analysis of the exposure probability of privacy caused by collusion attacks. Assumed that there are malicious ANs and LNs in a network. After slicing is executed, Y_i is sent to node j and d_i can be recovered if node j is a malicious one which colludes with its AN. Then, the private data of node i may be revealed. There is another situation where X_i is kept by node i itself. d_i can also be recovered if node i is malicious and colludes with its AN. Considering these two cases, the hidden probability can be illustrated using Eq. (5) which expresses the ability of privacy preserving.

$$\begin{aligned} P &= 1 - P_r\Big(\big(C(A) \cap C(S_j | j \neq i)\big) \cup \big(C(A) \cap C(S_i)\big)\Big) \\ &= 1 - \frac{|A_m|}{|A|} P_r\big(C(S_j | j \neq i)\big) - \frac{|A_m|}{|A|} P_r\big(C(S_i)\big) \end{aligned} \quad (5)$$

where $C(*)$ represents the event that entity $*$ is malicious. $|A_m|$ and $|A|$ denote the number of malicious ANs and the number of ANs, respectively.

Then, the hidden probability P_p of EPPA, P_m of MPPA, P_s of SMART, and P_e of EEHA can be inferred as Eqs. (6), (7), (8) and (9), respectively.

$$P_p = 1 - \frac{|A_m|}{|A|}\left(\frac{|LN^*| - 1}{|LN|}\right) - \frac{|A_m|}{|A|}\left(\frac{|LN^*|}{|LN|}\right) \quad (6)$$

$$P_m = 1 - \frac{|A_m|}{|A|}\left(\frac{|LN^*| - 1}{|LN|}\right) - \frac{|A_m|}{|A|}\left(\frac{|LN^*|}{|LN|}\right) - P_f \quad (7)$$

$$P_s = 1 - 2 \times \frac{|A_m|}{|A|}\left(\frac{|LN^*| - 1}{|LN|}\right) - \frac{|A_m|}{|A|}\left(\frac{|LN^*|}{|LN|}\right) \quad (8)$$

$$P_e = 1 - 2 \times \frac{|A_m|}{|A|}\left(\frac{|LN^*| - 1}{|LN|}\right) - \frac{|A_m|}{|A|}\left(\frac{|LN^*|}{|LN|}\right) - P_d \quad (9)$$

In Eq. (7), P_f represents the probability that an attacker will obtain the private information based on mark bits. In EEHA, when the data of ANs are not sliced, P_a denotes the probability that the attacker will destroy the privacy through the compromised ANs. We compare the hidden probabilities of SMART, EEHA, EPPA, and MPPA with $|LN^*|$ malicious sensor nodes (leaf nodes) and $|A_m|$ malicious aggregation nodes. Obviously, as EPPA and MPPA adopt a more complex slicing technique, their privacy performances are superior to those of SMART and EEHA in the cases of malicious LNs and malicious ANs as shown in Figs. 9 and 10.

6.2 Communication overhead

The communication overhead mainly resides in transmission of data slices. In SMART, each node needs to exchange two messages in a sensing round. In EEHA, only the leaf nodes divide their data into slices and two of the slices are sent to their neighbors. Each AN needs to transmit one message for data aggregation. Therefore, the bandwidth consumption of SMART is higher than that of EEHA. Similar to EEHA, only leaf nodes need to perform the operations of slicing and exchanging operations

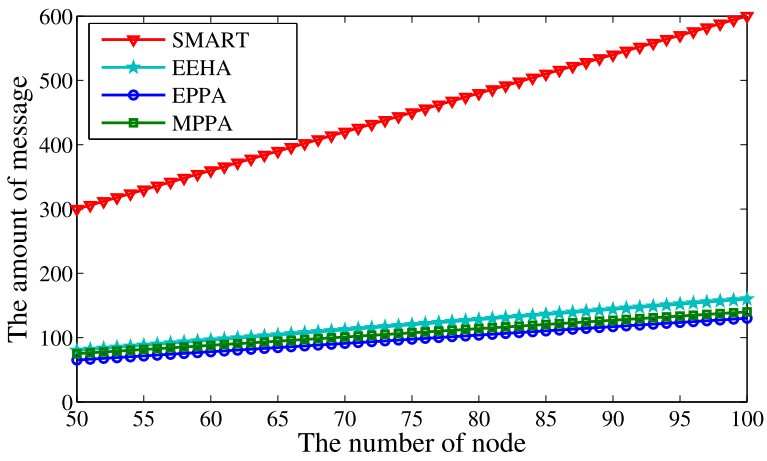


Fig. 11 Communication overhead

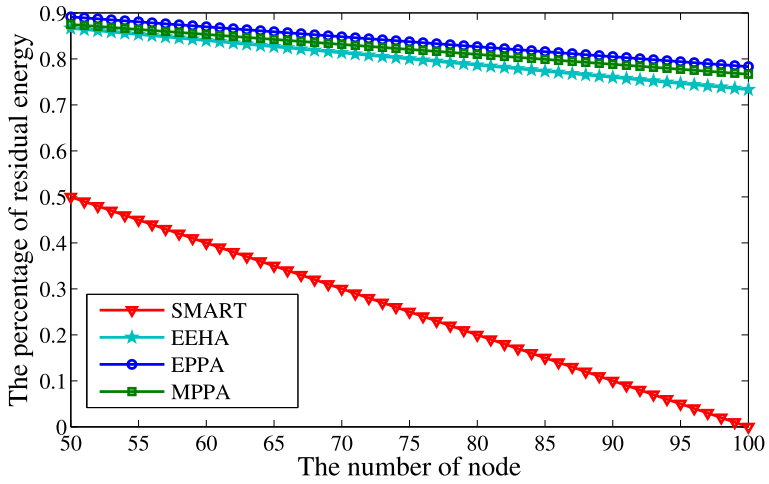


Fig. 12 Residual energy

in EPPA and MPPA. However, the sensing data are divided into two pieces and only one piece is sent to the neighbor. Obviously, the energy consumption of EPPA and MPPA is lower than that of EEHA and SMART. Hence, the communication overhead of four schemes can be formalized as Eqs. (10), (11), (12), and (13).

$$C_{\mathrm{SMART}} = N * K = 3N \tag{10}$$

$$C_{\mathrm{EEHA}} = \delta * N * K + (1-\delta) * N = 2\delta N + N \tag{11}$$

$$C_{\mathrm{EPPA}} = \delta * N * K + (1-\delta) * N = \delta N + N \tag{12}$$

$$C_{\mathrm{MPPA}} = \delta * N * K + (1-\delta) * N + b = \delta N + N + b \tag{13}$$

where δ and K represent the proportion of leaf nodes and slices, respectively. The message complexity of four schemes is shown in Fig. 11. The less the information is exchanged, the lower the energy will be consumed.

6.3 Network lifetime

Figure 12 shows the percentage of residual energy in a network as execution time escapes with which we can evaluate the network lifetime of different protocols. It is shown that SMART consumes energy much faster than other schemes. This is because there are more message exchanges in SMART in each round. Also, EPPA and MPPA have a longer lifetime in contrast with SMART and EEHA for their Euclidean-based slicing technique.

7 Conclusion

WSNs are composed of resource-constrained sensors which are usually deployed in an unattended or wild area. Therefore, energy and privacy issues are the main concerns in WSNs. The proposed EPPA scheme adopts data decomposition to replace the traditional slicing scheme in order to reduce energy consumption and prevent the private data from being compromised. In addition, an enhanced EPPA protocol called MPPA is proposed for the purpose of dealing with the problem that the typical slicing mechanism supports the addition aggregation function only and fails to meet the requirements of WSNs in many application scenarios. MPPA is characterized by focusing on multiple functions and significantly improving the universality of DA in WSNs. Simulation results show that our approaches can effectively decrease the overhead of communication and guarantee the data privacy. However, EPPA and MPPA can only be applied to integer aggregation which may produce a negative impact on data accuracy. We will make more in-depth research on decomposition so that it can be applied into more scenarios in the future.

Abbreviations

AN: Aggregation node; ASSDA: Adaptive slice-based secure data aggregation; BS: Base station; CPDA: Cluster-based private data aggregation; DA: Data aggregation; EEHA: Energy-efficient and high-accuracy; EPPA: Energy-efficient privacy-preserving data aggregation protocol; HEEPP: High energy-efficient and privacy preserving; LN: Leaf node; MPPA: Multi-function privacy-preserving data aggregation protocol; SMART: Slice-Mix-AggRegaTe; WSNs: Wireless sensor networks

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants 61672321, 61832012, 61771289, and 61373027 and the Shandong Graduate Education Quality Improvement Plan SDYY17138.

Authors' contributions

XZ, QZ, and CF are the principal contributors in terms of simulation modeling and the generation/interpretation of numerical results. In a supervising role, XL and JY formulated the research problem and contributed to the simulation modeling and the discussion of results. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China under Grants 61672321, 61832012, 61771289, and 61373027 and the Shandong Graduate Education Quality Improvement Plan SDYY17138.

Availability of data and materials

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]School of Information Scicence and Engineering, Qufu Normal University, Yantai Road, Rizhao, 276826 China. [2]School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Daxue Road, Jinan, 250353 China.

Received: 18 September 2019 Accepted: 2 January 2020
Published online: 14 January 2020

References

1. M. Hussain, P. Khan, S. Kyung, in *Proceedings of ICACT'09, vol. 1*. Wsn research activities for military application (IEEE, Phoenix Park, 2009), pp. 271–274
2. S. Wan, Y. Zhao, T. Wang, et al., Multi-dimensional data indexing and range query processing via voronoi diagram for internet of things. Futur. Gener. Comput. Sys. **91**, 382–391 (2019)
3. R. Zhang, P. Xie, C. Wang, G. Liu, S. Wan, Classifying transportation mode and speed from trajectory data via deep multi-scale learning. Comput. Netw. **2019**, 106861 (2019)
4. L. Wang, H. Zhen, X. Fang, S. Wan, W. Ding, Y. Guo, A unified two-parallel-branch deep neural network for joint gland contour and segmentation learning. Futur. Gener. Comput. Syst. **100**, 316–324 (2019)
5. Y. Zhao, H. Li, S. Wan, et al., Knowledge-aided convolutional neural network for small organ segmentation. IEEE J. Biomed. Health Inform. **23**(4), 1363-1373 (2019)
6. I. F. Akyildiz, A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
7. M. Saad, E. Fehr, N. Kamenzky, J. Schiller. Automated testing of wsn applications, (2008), pp. 157–166. https://doi.org/10.1109/icsnc.2008.30
8. Y. Jennifer, B. Mukherjee, D. Ghosal, Wireless sensor network survey. Comput. Netw. **52**(12), 2292–2330 (2008)
9. R. Wan, N. Xiong, Q. Hu, H. Wang, J. Shang, Similarity-aware data aggregation using fuzzy c-means approach for wireless sensor networks. EURASIP J. Wirel. Commun. Netw. **2019**(1), 59 (2019)
10. G. Li, Y. Wang, Automatic arima modeling-based data aggregation scheme in wireless sensor networks. Eurasip J. Wirel. Commun. Netw. **2013**(1), 85 (2013)
11. L. Qi, Y. Chen, Y. Yuan, S. Fu, X. Zhang, X. Xu, A qos-aware virtual machine scheduling method for energy conservation in cloud-based cyber-physical systems. World Wide Web J. **2019**, 1–23 (2019)
12. F. Mohamed, B. Abderrahim, S. Mostafa, Multi-mobile agent itinerary planning-based energy and fault aware data aggregation in wireless sensor networks. Eurasip J. Wirel. Commun. Netw. **2018**(1), 92 (2018)
13. A. Kemal, D. Murat, A. Savas, The impact of data aggregation on the performance of wireless sensor networks. Wirel. Commun. Mob. Comput. **8**(2), 171–193 (2010)
14. K. Katarzyna, M. Gerard, N. Hassan, Additively homomorphic encryption and fragmentation scheme for data aggregation inside unattended wireless sensor networks. Ann. Telecom. **74**(3), 157–165 (2019)
15. W. Li, X. Liu, J. Liu, P. Chen, S. Wan, X. Cui, On improving the accuracy with auto-encoder on conjunctivitis. Appl. Soft Comput. **81**, 105489 (2019)
16. Y. Lu, N. Sun, A resilient data aggregation method based on spatio-temporal correlation for wireless sensor networks. Eurasip J. Wirel. Commun. Netw. **2018**(1), 157 (2018)
17. G. Yang, S. Li, X. Xu, H. D. Z. Yang, Precision-enhanced and encryption-mixed privacy-preserving data aggregation in wireless sensor networks. Int. J. Distrib. Sens. Netw. **9**(4), 427275 (2013)
18. J. Wang, Y. Chen, Research and improvement of wireless sensor network secure data aggregation protocol based on smart. Int. J. Wireless Inf. Networks. **25**(3), 232–240 (2018)
19. W. He, X. Liu, H. Nguyen, K. Nahrstedt, T. Abdelzaher, in *Proceedings of INFOCOM'07*. Pda: privacy-preserving data aggregation in wireless sensor networks, (2007), pp. 2045–2053. https://doi.org/10.1109/infcom.2007.237
20. P. Hua, X. Liu, J. Yu, N. Dang, X. Zhang, Energy-efficient adaptive slice-based secure data aggregation scheme in wsn. Procedia Comput. Sci. **129**, 188–193 (2018)
21. A. Kemal, D. Murat, A. Savas, The impact of data aggregation on the performance of wireless sensor networks. Wirel. Commun. Mob. Comput. **8**(2), 171–193 (2008)
22. X. Xu, Y. Xue, L. Qi, Y. Yuan, X. Zhang, T. Umer, S. Wan, An edge computing-enabled computation offloading method with privacy preservation for internet of connected vehicles. Futur. Gener. Comput. Syst. **96**(1), 89–100 (2019)
23. L. Qi, X. Zhang, W. Dou, C. Hu, C. Yang, J. Chen, A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment. Futur. Gener. Comput. Syst. **88**, 636–643 (2018)
24. W. Gong, L. Qi, Y. Xu, Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment. Wirel. Commun. Mob. Comput. **2018**, 1–8 (2018)
25. Y. Xu, L. Qi, W. Dou, J. Yu, Privacy-preserving and scalable service recommendation based on simhash in a distributed cloud environment. Complexity. **2017**, 1–9 (2017)
26. Y. Hamed, Y. Hossein, A. Naser, M. Ali, Structure-free real-time data aggregation in wireless sensor networks. Comput. Commun. **35**(9), 1132–1140 (2012)
27. P. Steffen, P. Krzysztof, L. Peter, in *Proceedings of CCNC'07*. On concealed data aggregation for wsns, (2007), pp. 192–196. https://doi.org/10.1109/ccnc.2007.45
28. C. Castelluccia, E. Mykletun, G. Tsudik, et al., in *The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services*. Efficient aggregation of encrypted data in wireless sensor networks, (2005), pp. 109–117. https://doi.org/10.1145/1525856.1525858
29. H. Li, K. Lin, K. Li, Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks. Comput. Commun. **34**(4), 591–597 (2011)
30. C. Liu, Y. Liu, Z. Zhang, Z. Cheng, High energy-efficient and privacy-preserving secure data aggregation for wireless sensor networks. Int. J. Commun. Syst. **26**(3), 380–394 (2013)
31. P. Gallant, J. Lambert, S. Vanstone, in *Proceedings of Annual International Cryptology Conference*. Faster point multiplication on elliptic curves with efficient endomorphisms, (2001), pp. 190–200. https://doi.org/10.1007/3-540-44647-8_11
32. D. A. Cox, *Primes of the Form X+NY: Fermat, Class Field Theory, and Complex (Second Edition)*. (Wiley, Hoboken, 2011)

Publisher's Note