International Journal of Information Security and Privacy
Volume 15 • Issue 1 • January-March 2021

Data Privacy Protection Algorithm Based on Redundant Slice Technology in Wireless Sensor Networks

Peng Li, Nanjing University of Posts and Telecommunications, China
Chao Xu, Nanjing University of Posts and Telecommunications, China
He Xu, Nanjing University of Posts and Telecommunications, China

ABSTRACT

In order to solve the problem that the privacy preserving algorithm based on slicing technology is incapable of dealing with packet loss, this paper presents the redundancy algorithm for privacy preserving. The algorithm guarantees privacy by combining disturbance data and ensures redundancy via carrying hidden data. It also selects the routing tree that is generated by the CTP protocol as the routing path for data transmission. Through division at the source node, the method adds hidden information and disturbance data. This algorithm uses hidden data and adds perturbation data to improve the privacy preserving. Nonetheless, it can restore the original data when data are partly lost. According to the simulation via TOSSIM (TinyOS simulator), in the case of partial packet loss, the algorithm can completely restore the original data. Furthermore, the authors compared accuracy of proposed algorithm, probability of data reduction, data fitting degree, communication overhead, and PLR. As a result, it improves the reliability and privacy of data transmission while ensuring data redundancy.

KEYWORDS

Collection Tree Protocol, Disturbing Data, Network Security, Privacy Protection, Slice Technology

INTRODUCTION

Wireless sensors, as an important carrier within a wireless sensor network, have several shortcomings including a potential for eavesdropping at nodes, limited energy, weak computing capacity, and a high probability of packet loss. The main task of a wireless sensor is to collect and transmit data, and these sensors are often placed in uninhabited areas, proposed by Conti et al. (2013) and Acharya et al. (2005). The data obtained by the sensors may face security threats such as being physically captured, attacked and tampered with by attackers, leading to leaking of private information, proposed by Fan et al. (2012). This is particularly an issue in important areas such as the military or medical field. Because incalculable damage can occur if the attacker obtains a wiretap to intercept sensitive data in the data link layer, and leaks some key information or tampers with the sensitive data through the

DOI: 10.4018/IJISP.20210101.oa1

wiretap. Therefore, it is important that large-scale wireless sensor network applications study and resolve wireless sensor network data privacy concerns, proposed by León et al. (2009).

The wireless sensor network is a randomly distributed network deployed on a large scale, and the data is transmitted by wireless channel. Vinodha and Anita (Vinodha & Anita, 2018) mentions such information transmission methods inevitably occur some security drawbacks. The attacker may perform data interception and stealing on the wireless sensor node through the data link layer. Therefore, the wireless sensor network needs to be protected via privacy technology.

Privacy protection technologies are mainly divided into data privacy protection technology and location information privacy protection technology in the wireless sensor networks. Location privacy protection technology is to prevent attackers from obtaining the target location through communication mode monitoring and analysis. While data privacy protection technology is to prevent attackers from eavesdropping on sensor nodes through the link layer to obtain effective information. This paper mainly concerns on achieving the security protection of the transmission data in the sensor network based on the privacy protection technology.

The current data privacy protection protocols cannot effectively deal with packet loss problem, some of the protocols use retransmission to deal with packet loss, which leads to a negative impact to the sensor networks. Therefore, the original data can be transformed, and some additional data information is increased, so that when it comes to packet loss, the packet can be recovered by using additional data information without retransmission, proposed by Emimanothaya and Babu (Emimanothaya & Babu, 2017).

For some cases where the wireless sensor networks scenario is unstable, such as Hua et al. (2018) the network topology varies. It is necessary for a node to send data to the target node as much as possible, overcoming the probability of data loss. In addition, the privacy protection protocols are effective in confronting monitoring, traffic analysis, data tampering, and replay attacks in wireless sensor networks.

In order to solve the problem that the privacy preserving algorithm based on slicing technology is incapable of dealing with packet loss, this paper presents the linear redundancy algorithm for privacy preserving. A model for a data privacy algorithm is proposed, namely Mixed-Slice private protocol (MS), which is based on slicing technology for the sensitive issue of packets. Under circumstances where the destination node would lose some of the data packets, the entire data can still be restored by a linear redundancy algorithm. The proposed algorithm shows better data accuracy, Data reduction probability and data regression degree compared to the current researches.

The remainder of the paper is organized as follows. Section 2 introduces related work. In Section 3, the system model of data privacy algorithm based on slicing technology is proposed. In Section 4, we detail the key technologies of data privacy protection. Section 5 carries out some comparative experiments and analysis. Section 6 summarizes the paper.

RELATED WORK

Currently, wireless sensor network data privacy protection techniques can be divided into three main classes: Data privacy protection protocol based on segmentation, homomorphic encryption, and disturbance technique. Some typical data privacy protocols will be elaborated on in the following paragraphs.

1. Data privacy protection protocol based on segmentation.

The main idea of data privacy protection protocol based on segmentation is to slice up the original data into separated segments proposed by He et al. (2007) and then choose multiple transmission paths to the destination node for transmitting, in order to achieve

privacy protection, proposed by Sorniotti et al. (2007). For instance, the main idea of SMART (Showail et al., 2014) (Slice Mix AggRegaTion) is to slice the original data into separate segments. The implementation process mainly consists of three steps of slicing, mixing, and aggregation. The SMART algorithm protects data privacy by obtaining only part of the data together through intermediate nodes. Its advantage includes a smaller amount of calculation and better privacy protection,while the disadvantage includes a larger communication overhead, sensitivity to data loss, and poor data integrity, proposed by Li et al. (2009) and Zhang et al. (2014). At this point, YANG et al. (2011) proposed the ESPART (Energy-Saving Privacy-preserving data AggRegaTion) protocol. ESPART is considered to be an improvement to the SMART protocol. The advantages of ESPART protocol include a smaller chance of collision when sending data, and higher privacy protection, proposed by Ameen et al. (2012). The disadvantages include a high communication latency and lack of data integrity, proposed by Wang et al. (2017) and Yu et al. (2015).

2. Data privacy protection protocol based on homomorphic encryption.

Homomorphic encryption is a type of cryptography technique which is based on the mathematical concept of computational complexity theory, proposed by Yang et al. (2015). It allows computation on cipher texts, generating an encrypted result which, when decrypted, reaches the same result of the operations as if they are performed on the plaintext. Homomorphic encryption based on the data privacy protection protocol is built on such a precondition. One of the two main implementations is AHE Protocol (Additively Homomorphic Encryption) proposed by Castelluccia et al. (2009). This method is based on the principle of homomorphic encryption where the protocol does not decrypt data, but instead directly performs operations on the encrypted data. The data is finally decrypted after the operations are complete, so as to aggregate them. The advantages of the AHE Protocol include better privacy, more security, and less computation overhead, while its disadvantage lies in lack of support for integrity verification. The other one is IPHCDA Protocol (Integrity Protecting Hierarchical Concealed Data Aggregation) proposed by Ozdemir and Xiao (Ozdemir & Xiao, 2011), which is based on elliptic curve cryptography of the homomorphic encryption algorithm. The advantages of IPHCDA protocol include more privacy, integrity verification support, and higher accuracy polymerization results. However, its disadvantages include a relatively larger communication overhead and a computational overhead.

3. Data privacy protection protocol based on disturbance technique.

Data privacy protection protocol based on disturbance technique hides the real data mainly by adding random numbers to the raw data, in order to achieve the aim of protect the privacy of data. The main protocols are CPDA (Cluster-based Private Data Aggregation) Protocol and KIPDA (K-Indistinguishable Privacy-preserving Data Aggregation) Protocol. He et al. (2007) proposed the CPDA protocol, which uses the idea of adding disturbance data to the original data, in order to hide the real data. The advantage of the CPDA protocol is that it achieves accurate polymerization. However, its disadvantages are a larger computing load and more communication overhead. Groat et al. (2011) and Zheng et al. (2018) proposed the KIPDA protocol. The main idea of this protocol is to add some false data to the raw data in order to hide the original data, rather than using encryption to keep the original data together. The advantages of KIPDA protocol include support for the integrity verification protocol and a smaller computation overhead. Its disadvantages include a larger communication overhead and a higher space complexity. In addition, Liu et al. (2017) proposed ActiveTrust and the generation and distribution of detection routes are given in the ActiveTrust scheme to achieve the desired security and energy efficiency.

Comparing with SMART, ESPART, AHE, IPHCDA, CPDA, and KIPDA protocol, the performance of the privacy protection algorithm in terms of accuracy, computationality, packet loss sensitivity and privacy protection performance are listed as Table 1.

Table 1. Comparison of the various protocols

Protocol	Communication Overhead	Calculation Overhead	Accuracy	Packet Loss Sensitivity	Privacy Protection
SMART	High	Low	Medium	Yes	Medium
ESPART	Medium	Low	High	Yes	High
AHE	Medium	Low	Low	Yes	Medium
IPHCDA	High	High	Medium	Yes	Medium
CPDA	High	High	Medium	Yes	Medium
KIPDA	High	Medium	Low	No	Low

As shown in Table 1, the communication overhead of the non-encrypted privacy protection algorithm via redundant slice technology is relatively large, mainly because of the fragmentation strategy adopted, thus increasing the communication overhead, and at the same time, due to linear restoration and slicing, the computational overhead is not large. In terms of accuracy, since there are multiple slices of hidden data, multiple verifications can be performed, so the accuracy is high.

Data privacy protection protocol based on disturbance technique has a large amount of computation and large traffic. Goswami et al. (2014) proposed data privacy protection protocol based on homomorphic encryption does not support the problem of data integrity verification. As for the data privacy protection protocol based on homomorphic encryption, it is sensitive to packet loss.

In summary, all of these typical protocols have advantages and disadvantages. The disturbance technique of the data privacy protection protocol requires a large amount of computation and communication. The segmentation of the data privacy protection protocol is very sensitive to packet loss, while homomorphic encryption based on the data privacy protection protocol lacks the support for integrity verification.

SYSTEM MODEL

Overall Operations

In the MS model shown in Figure 1, the routing tree from the source node to the destination node is built together with the CTP used to generate the information transfer tree for the data transmission.

The source carries out a fragmentation operation and adds disturbance data to strengthen the data privacy protection. Figure 1 shows the network model.

Source Node Operation

Flowchart of source node working is shown in Figure 2, the source node collects a default amount of data according to the algorithm, and then it adds the disturbance data and distributes to its parent node (the transmission node).

According to Showail et al. (2014), we set the default predefined number to 100. The predefined number is determined by the stack set in the source node, which varies according to the speed of data collection.

Figure 1. Network model of data privacy algorithm

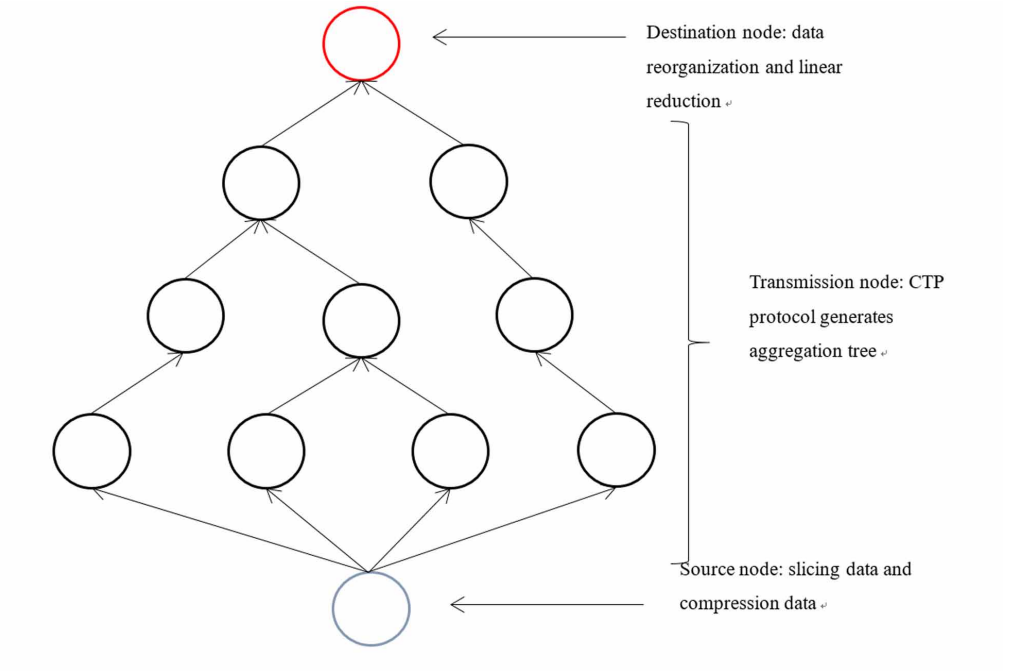
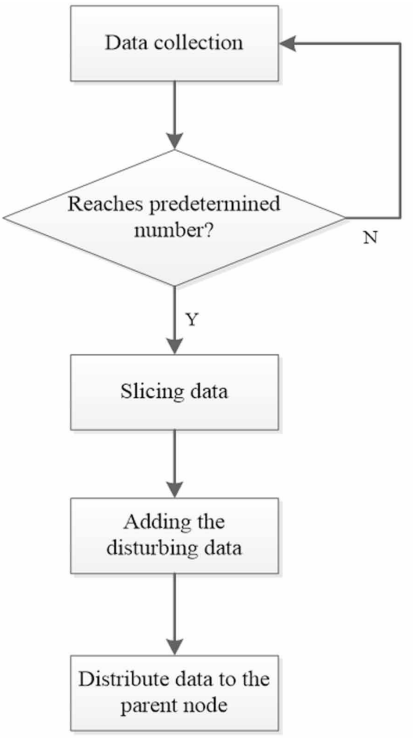


Figure 2. Flow chart of source node working



Data acquisition: The speed of data collection at the source node is comparatively faster than that of transferring data. Therefore, the energy consumption is increasing each time data is transmitted. And the collected data is not sent every time but stored in a stack. Thus, it is wiser to temporarily store some of the data in the source node, so as to dynamically adapt to the speed of transmission. The source node sends the data once the data size reaches predefined number in order to improve the efficiency of data transmission.

Data processing: The collected data is divided into pieces, and some disturbance data is then added to the data within each slice, and joined to the remainder of the slice information in each packet. All of the data packets are distributed to each parent node via the information transmission tree. The pseudo-code of the source node operation is listed in Pseudo-code 1.

Pseudo-code 1. Source node operation

```
Begin
   If(!is_getparent()) then         // No packet received from
                                       transmission node
      Waiting();
   Else
      N=Get_number_parent();       // Get the parental node number of source
      Count=Start_collectdata();
// Start collecting Data and counting the number
   If(count!=NUMBER_SLICECOUNT) then    // Count number not reach the thresh
      Collection[]=Start_collectdata();
   Else                            // Count number has reached the thresh
      pow(2,n)<=N                  //Get the entire Data array Collection[]
   A=b/MAX(pow(2,n))
   VN=pow(2,n);                    // Calculating effective nodes
   For(int i=0;i<VN;i++)
      X[i]=cal(Collection[])       // Slice the clear text
   For(int i=0;i<VN;i++)
      S[i+1]=X[i+1]+X[i+2]         // Calculating slice Data
   For(int i=0;i<5;i++)
      Add(r=Rand(0,100))           // Adding combining disturbance Data
   Send_data();
End
```

Transmission Node Operation

The main task of the transmission node is to transfer the source node data to the destination node. Both nodes use the CTP to generate the information transmission tree. 50% of the nodes are transmission nodes, or in other words, 50% of the nodes are responsible for data transmission and the remainders are for data collection.

Since a wireless sensor network might change over time, this will often lead to a node becoming invalid. So it is necessary for every node to have multiple next-hop nodes. In addition, the nodes should be able to detect whether the parental nodes are working properly at any time, in order to ensure that the information can be transmitted to the destination node effectively. The CTP is adopted by all the transmission nodes, where the Estimation Exchange Protocol Link (LEEP) broadcasts a data frame to confirm the survival of the surrounding nodes in order to ensure the survival of the parent nodes. Moreover, the transmission nodes are confirmed by the LEEP frame, and the link quality of the transmission nodes are calculated by the algorithm. Excluding the loop, the optimal path is selected

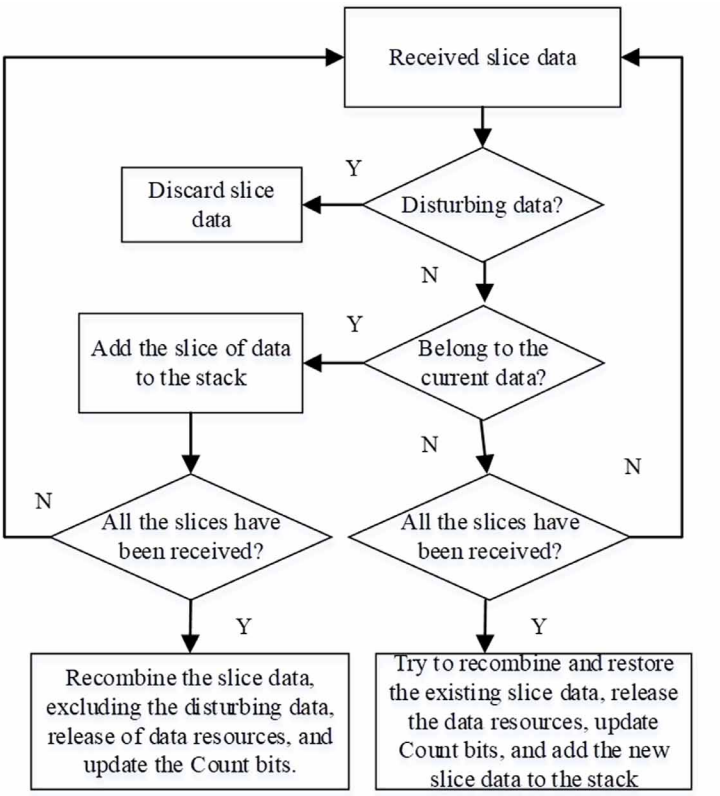
as the parent node. CTP guarantees a data delivery success rate of around 90%, proposed by Xie and Wang (Xie & Wang, 2014).

Destination Node Operation

Flow chart of destination node is shown in Figure 3:

1. When the destination node receives the data slice, either the disturbance data is abandoned, or the Count bits are read (the Count bits are used to record the number of times the data has been sliced) to determine whether it is the same as the data in the stack;
2. If not, the destination node tries to recombine and restore the existing slice data. Then it automatically releases the data resources in the stack, updates the Count bits, adds new data to the stack, and returns to wait for the next slice data. Otherwise, it simply adds the slice to the stack;
3. Then the destination node recombines the slice data, excluding the disturbing data, releases of data resources, and updates the COUNT bit, in case it calculates that all the slices have been received according to the indication of Count bits;
4. Otherwise, it continues waiting for the slide data if all the slides have not been received;
5. Finally, when the received number of data slices reaches a predetermined number, the data slices are reconstructed according to their location.

Figure 3. Flow chart of the destination node working



The pseudo-code of the destination node operation is listed in Pseudo-code 2.

Pseudo-code 2. Destination node operation

```
Begin
   If(!is_reveicve()) then        // Check if Data is received
      Waiting();
   Else
   If(is_disturbing())            // Check if combining disturbance Data
      Abandon();
      Waiting();
   Else
   If(is_thistimedata()) then
      Recieive[]=Add_data();      // Add to receiving Data
   If(isDataChange(Collection))   // Prompt data to be tampered
      Error();
   If(is_complete()) then
      Recovery_data(Recieive[]);
      Waiting();
   Else
      For(int i=0;i<VN;i++)
         X[i]=S[i]-X[i-1];
   Waiting();
End
```

DATA PRIVACY PROTECTION

Due to a large number of variable names used in the following discussion, a table is listed to summarize these variable names, as shown in Table 2.

Format of the Frame

For the privacy preserving algorithms based on slicing technology, it is sensitivity to packet loss. The current algorithms cannot effectively deal with packet loss, usually using retransmission to deal with packet loss, which exerts a negative impact on the privacy protection of wireless sensor networks. Furthermore, when the nodes are faced with dataflow sniffing, traffic analysis, data tampering, or replay attacks, wireless sensor network fails to deal with these attacks.

Therefore, the original data can be processed and additional data information can be added, so that in case of packet loss, it can be recovered without retransmission.

To sum up, for the data privacy preserving algorithm based on slice technology, the main problem is to deal with the packet loss. Therefore, a data privacy protection algorithm with redundancy mechanism is proposed in this paper.

Data is sliced at the source node, and the data slices are restructured and restored at the destination node, as shown in Figure 4.

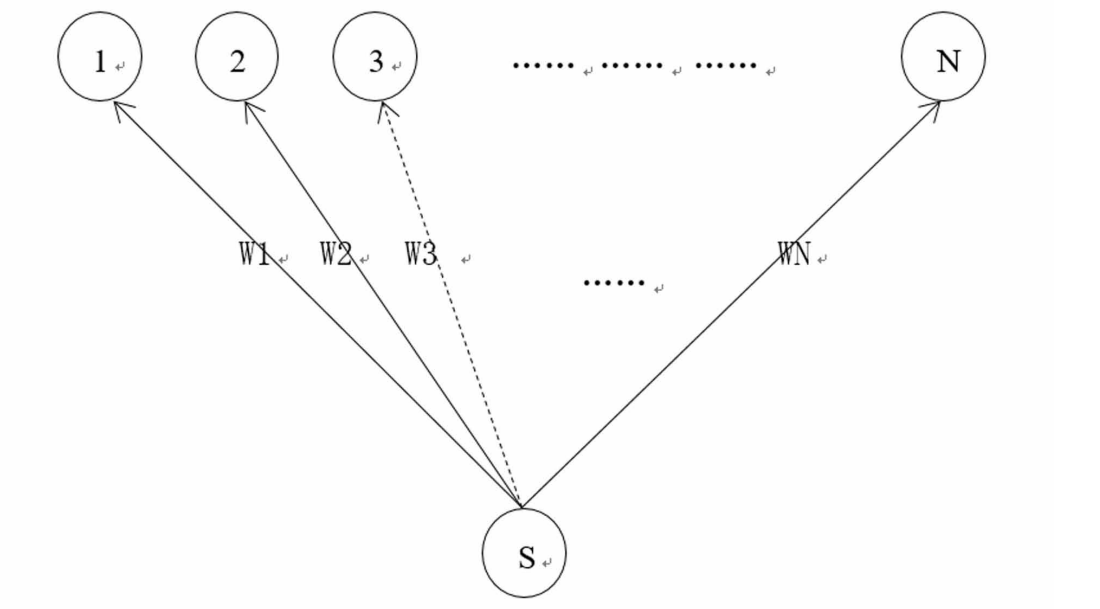
The source node S sends the data slices to each node. The dotted line represents the possible packet loss, and the sliced data is then transmitted via these nodes to the destination node through the CTP tree. Rather than every node delivering valid data, some of the nodes transmit disturbance data. Obviously, the disturbance data is recognized by the destination node and does not participate in the reconstruction process.

Assuming that the original data is B bytes, then a total of 8B bits need to be sliced.

Table 2. Meaning of the variable names

Variable Name	Meaning
B	The number of bytes of original data
b	The number of bits of original data
N	The number of parent node of the source node
A	The number of bits that are assigned to the parent node of the source node to transmit the valid data
VN	The number of valid nodes
PLR	Packet Loss Rate
HSC	Hide Slice Count
Xi	Clear text message carried by each slice data
Si	Hidden slice data information
Matrix_N	The complete linear matrix obtained from the hidden information data
Matrix_N_T	The result of the variation of the complete linear matrix
ASW	A set of known X_i solutions
CD	Clear Data(Clear data from each slice)
DD	Disturbing Data
CD_position	position of clear data

Figure 4. Source node sends the slice data to the parent nodes



Assuming that b is the number of bits per node, N is the total number of nodes, and N meets the requirements of Formula 1:

$$A = \frac{b}{MAX\left(2^n\right)}, 2^n \le N \tag{1}$$

A must be an integer, n=1,2,3…

For a maximum integer value of n, there are 2^n nodes to transmit the effective data. The number of nodes to transmit the disturbance data is N-VN.

The frame format of the source node is designed as illustrated in Figure 5. And the corresponding fields meaning is shown in Table 3. Thus, different packets can be distinguished from each other in the destination node, and whether there is data disturbance or not, and the number of slices can be determined by the union decision of tag validation bit (8 bits), count bits (16 bits), and total number of slice bit (16 bits) fields.

Figure 5. Frame format of the slicing data

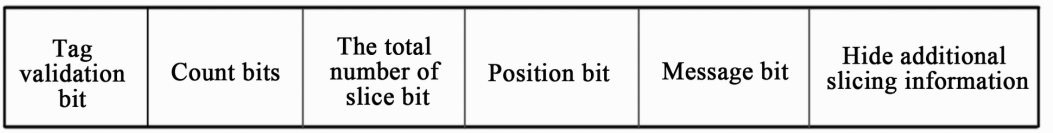


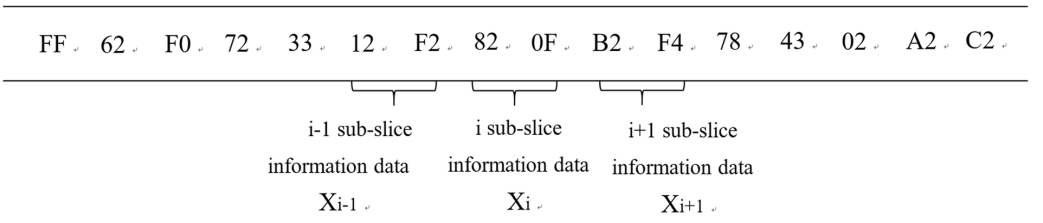
Table 3. Fields meaning of frame format

Name	Field Length	Meaning
Tag validation bit	8 bits	Some nodes transmit the disturbing data, so it is necessary to identify in the destination node, if this data is valid, it generates a 100-255 disturbing data to the tag, and otherwise it generates 0-99 disturbing data.
Count bits	16 bits	It records the sending times. When the source node sends data, the Count bit adds by 1 each time, and the Count sets to 0 when the number reaches 65536.
Total number of slice bit	16 bits	It keeps an account of the total length of data.
Position bit	16 bits	It records the position in original data.
Message bit	A bits	The information of slicing data is generated by the source node.
Hide additional slicing information	Variable length	The length is determined by the packet loss probability.

Fragmentation

The source node starts to slice the data while collecting the data, as shown in Figure 6 In the VN node, the i-th node gets effective data X_i, as well as some other parts containing hidden information. At a certain point, the packet loss rate is PLR, then it calculates HSC (Hide Slice Count) according to Formula 2:

Figure 6. Data sheet reference map



$$Min\left(HSC / VN\right) \geq PLR \tag{2}$$

HSC meets the conditions of the smallest positive integer, and the information in the slicing data meets the requirement of Formula 3:

$$S_{(i+HSC)\%VN} = X_{(i+HSC+1)\%VN} + X_{(i+HSC+2)\%VN} \tag{3}$$

Upon receiving the data, the destination node reads the Tag validation bit, discards the disturbance data. Otherwise it reads the Count bits and takes the existing data slice recognition, then it automatically releases the data resources in the stack, updates the Count bits, and adds new data to the stack. When all the data slices are received, the data is reconstructed based on the position of the data slices.

Data Reduction

The destination node regards the following two cases of incomplete data as packet loss. Firstly, it is considered if the destination node receives the next data slice ahead of time. For instance, it assumes that the latest data slice received is the 100th data slice, if the destination node has received the 101st data point in advance, then the 100th part of the data slice is considered to be lost, so the destination node begins to restore the 100th data slice. Secondly, it still assumes the 100th slice, but the 101st is not received or the 100th is not fully received. At this point, a threshold time is set in the destination node so that when it is expired, the node will restore the 100th data.

When confirming the data loss, the target node begins to restore the data. Assuming that the all of the slice data have reached the destination, we can get all the S data packets. The corresponding equation is listed as Formula 4 and a (VN * VN) linear matrix is obtained. Using transformation, we can derive that R(Matrix_N) =VN-1 and matrix listed as Formula 5:

$$Matrix\_N = \begin{bmatrix} 1 & 1 & 0 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 & 0 & 0 \\ & \vdots & & & \ddots & & \vdots & & \\ 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_{VN-3} \\ X_{VN-2} \\ X_{VN-1} \\ X_{VN} \end{bmatrix} = \begin{bmatrix} S_{VN} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_{VN-4} \\ S_{VN-3} \\ S_{VN-2} \\ S_{VN-1} \end{bmatrix} \tag{4}$$

$$Matrix\_N\_T = \begin{bmatrix} 1 & 1 & 0 & 0 & & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & & 0 & 0 & 0 & 0 \\ & \vdots & & & \ddots & & \vdots & & \\ 0 & 0 & 0 & 0 & & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ \vdots \\ X_{VN-3} \\ X_{VN-2} \\ X_{VN-1} \\ 0 \end{bmatrix} = \begin{bmatrix} S_{VN} \\ S_1 \\ S_2 \\ S_3 \\ \vdots \\ S_{VN-4} \\ S_{VN-3} \\ S_{VN-2} \\ 0 \end{bmatrix} \tag{5}$$

As the rank of the Matrix_N_T matrix is less than VN, we can get infinitely possible solutions of the matrix based on this matrix, and since each slice carries one solution, meaning that each slice offers a solution to X. And it can get Matrix_N matrix, upon receiving at least a slice of data. Assuming that the destination node does not receive the data packet X_i, the destination node reconstructs the Matrix_N according to the Formula 6, so as to restore the original data:

$$X_i = S_i - X_{i+1} \tag{6}$$

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & & 0 & 0 & 0 & 0 & 0 \\ & & \vdots & & & \ddots & & & \vdots & & \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ & & & & & & & & & & \\ & & & & & & & & & & \end{bmatrix} = \text{Matrix\_Lack} \tag{7}$$

When receiving a slice packets X_j and according to the linear transformation of Formula 7, the S set based on the packet comes to an (HSC×VN) matrix Matrix_Lack, thus it computes R(Matrix_Lack)=HSC. When the second slice is received, it computes two matrixes and combines two matrixes whose rank is VN-1 to solve for the complete data immediately according to Formula 6. For instance, when receiving the (j+1) %VN or (j-1) %VN slice data, the rank of the matrix is increased by 1, which is HSC+1, the relations between the number of receiving data and the rank of the matrix can be obtained. So, the rank of matrix meets the requirement of Formula 8:

$$HSC + n - 1 = VN - 1 \tag{8}$$

When n=VN-HSC, it solves all the solutions and Matrix_N is restored to a matrix, so that the destination node reconstructs the original data. Plus $\frac{HSC}{VN} \geq PLR$, so the destination node receives (1-PLR) *VN data in case the packet loss rate is PLR, as shown in Formula 9. As a result, it can restore the original data in the case of stable packet loss rate:

$$\left(1-PLR\right)\times VN \geq VN - HSC$$

where:

$$\frac{HSC}{VN} \geq PLR \tag{9}$$

In the case of restoring the data X_x, it sets up all known solutions for the collection, ASW= $\{X_a, X_b,..., X_m\}$, and finds the number whose distance is nearest to x in sets to reduce the time complexity:

$$Distance = Min\left|r-x\right|, Xr \in ASW \tag{10}$$

To find the nearest solution X_r in Formula 10, it calculates the X_x, when r<x:

$$X_x = (-1)^{x-r} \times X_r + \sum_{i=0}^{x-r-1} (-1)^{x-r-i-1} \times S_{r+i} \tag{11}$$

When r>x:

$$X_x = (-1)^{r-x} \times X_r + \sum_{i=0}^{x-r} (-1)^{x-r-i} \times S_{r-i} \tag{12}$$

According to the Formula 11 and Formula 12, the best time complexity for any solution is O(1), the average time complexity is O(HSN/4), and the worst time complexity is O(HSN/2). Once a solution is computed, it comes to the conclusion to all the solutions via traversing set S, thus the average time complexity of solving all the solutions is O (VN), seen as Table 4.

The MS protocol proposed is a non-encrypted redundant privacy protection algorithm based on segmentation technology with a small amount of computation. By carrying part of the hidden data information, when the target node fails to receive all the fragmented data, it is capable of restoring the whole data by the hiding fragmentation information without retransmitting, thus ensuring the small computation overhead and improving the transmission efficiency.

Table 4. Time complexity

Solving Any One Solution		
worst	average	best
O(HSN/2)	O(HSN/4)	O(1)
Solving All Solutions		
worst	average	best
O(VN)	O(VN)	O(VN)

Possibility of Restoring Complete Data

According to the characteristics of the CTP, the data transmission reliability is above 90%, so Formula 13 is concluded:

$$Min\left(HSC\right) \geq 10\% \times VN \tag{13}$$

Since the number of hidden information in each slice is 10%VN, once VN-HSN slice data have been received, it is able to restore the complete data. Actually, under some circumstance that the slice received is less than the number VN-HSN, the destination node can also reconstruct the complete data. Thus VN-HSC is the upper limit to restore the complete data, shown in Table 5.

Table 5. Possibility of destination node restore the received slice data

Number of the Slice Received	Possible
$0 \sim 1/PLR - 1$	Impossible
$1/PLR \sim VN - HSC - 1$	Possible
$VN - HSC \sim VN$	Definite

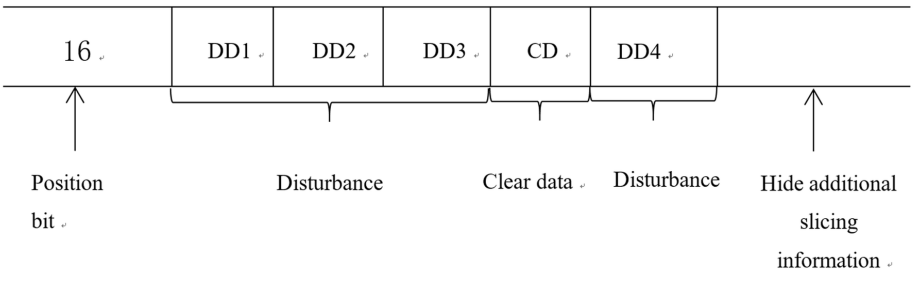
As to the lower limit for enabling reconstruction of the original data, all the recovering matrixes of Matrix_N should correspond to unique hidden slice data, seen as Formula 5. The destination node cannot restore the complete data when the number of slices received is below Formula 14:

$$\frac{1}{PLR} \tag{14}$$

Disturbing Data

Since each slice carries parts of the data, it is crucial to strengthen the privacy protection. The hidden information in "Hide additional slicing information" field is not processed. Only the information bits are disturbed, as shown in Figure 7, the data assigned to each of the slices are CD (Clear Data), meanwhile, the DD (Disturbing Data) are added.

Figure 7. Example of disturbing data



In Figure 7, it sets up a data structure for the clear data and disturbance data, and then defines an array with a length of 5, in which four elements of the array are disturbing data with only one is the clear data. The location of the CD_position data is determined by the effective position according to Formula 15:

$$CD\_position = Marker\ effective\ bit\ 5\% \tag{15}$$

When the destination node receives slice from the transmission nodes, it determines the validation of data through checking the first significant bit, reads the tag validation bit and calculates the remainder as to the effective clear data. It discards the slice if the data is invalid.

SIMULATION AND PERFORMANCE ANALYSIS

Experiments are carried out based on nesC in TinyOS operating system, with TOSSIM tools for simulation. TOSSIM is a simulation tool in TinyOS, which supports large-scale network simulation and uses the Micaz simulation platform, proposed by Lin et al. (2016), Yang et al. (2016) and Zheng et al. (2018). By specifying error at the bit level, TOSSIM can capture many causes of packet loss and noise in a TinyOS network, including missed start symbols, data corruption, and acknowledgement errors. In addition, the noise simulation data is provided by Stanford University (Lee et al., 2007). In this experiment (León et al., 2009), the network consists of 600 nodes and these nodes are randomly deployed over a 400×400-meter area. In TOSSIM simulation tool, the noise value provided by Stanford University is calibrated according to the range of 400 * 400-meter, so the range is adopted. The signal transmission range of a sensor node covers 50 meters and the data rate is 1 Mb/s. The network topology is generated by the CTP, which will be the structure with the best communication quality according to the network conditions. Finally, the experiment data are collected and analyzed with MATLAB V7.0, so as to show the result diagrams.

PRIVACY PROTECTION TESTS AND ANALYSIS

Data Accuracy

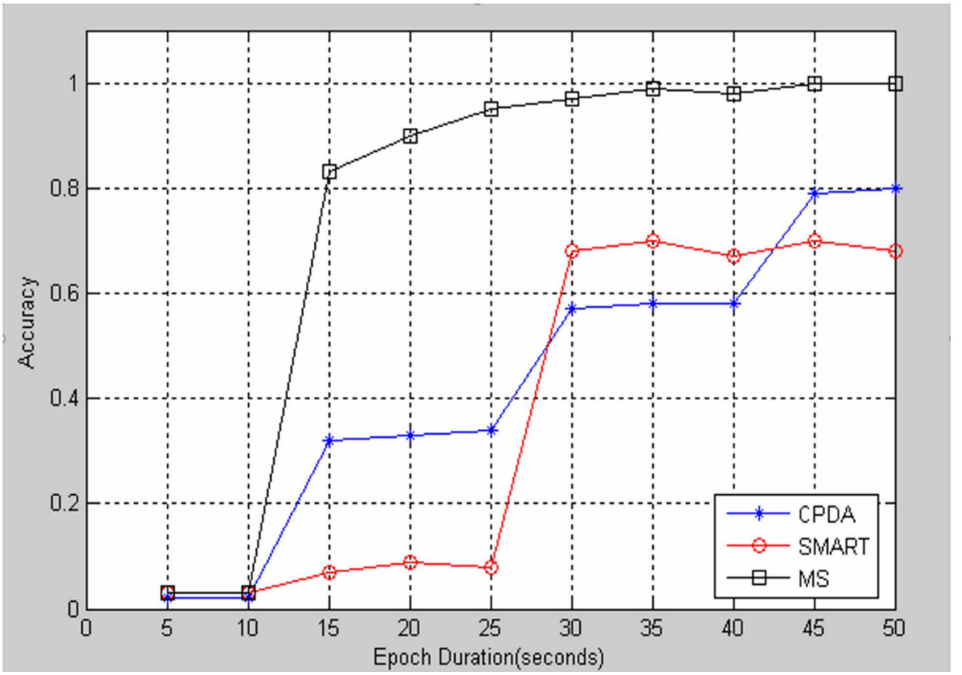
Figure 8 shows the accuracy comparison of the MS protocol, the CPDA (p_c=0.3) protocol and the SMART(J=3), proposed by León et al. (He et al., 2007; León et al., 2009). The last time is 50s. Since the comparative two protocols, namely CPDA protocol, and SMART protocol, take the number of slices defined as 100, accordingly we adopt the same slice in our proposed MS protocol, so as to make a reasonable and effective comparison with uniform standard. The experiment is carried out based on the average of 10,000 test experiments. We obtained the average of these experimental tests and finally plotted the experimental diagram of Figure 8.

Figure 8 indicates that the CTP is used by the transmission nodes. CTP has a message queuing mechanism, so the packet loss rate is relatively small. Due to the redundancy of the MS protocol, the source node has good data reconstruction in the case of packet loss. It can be seen from Figure 8 that at the beginning 10 seconds three protocols the same level of accuracy, however after 10 seconds, the accuracy of the proposed MS protocol is at least 20% higher than that of CPDA and SMART. Moreover, at 30 seconds, the accuracy of the proposed MS protocol is close to 100%. Thus, it shows that MS has better accuracy performance compared with the CPDA and SMART protocol.

Data Reduction Probability

It was stated that when no less than 90% of the slices are received, the source node can definitely reconstruct the entire data. However, when less than 90% of the slices are received, it is still possible

Figure 8. Accuracy comparison of MS, CPDA(p_c=0.3), and SMART(J=3)



to restore the entire data. It is assumed that the source node divides the data into 100 slices, with a packet loss rate setting of 10%. Our experiment is to test on each received amount of slice for 10,000 times, and calculate the probability of restoring complete data.

As shown in Figure 9, the horizontal coordinate indicates the number of data slices received, and the vertical coordinate indicates the probability that the complete data can be restored. It can be seen that when the destination node receives more than 50 slices, the probability that the source node can restore the complete data reaches approximately 100%. In fact, when the number of data slices received by the source node is between 10 and 50, the probability that the source node can restore the complete data is about 98.9%. When approximately 20 slices are received by the source node, the probability of restoration is reduced to 50%. When the source node receives less than 10 slices, it is impossible to restore the complete data. Thus, the protocol is guaranteed when the packet loss rate is less than 10%, and even when only 50% of the data packets are received, it still has a very high probability of restoring the complete data. According to the practical characteristics of the CTP, the data transmission reliability is above 90%, so that we can infer that the proposed MS protocol can almost rebuilt the slicing data at the destination node.

Data Regression Degree

Because the algorithm can restore part of the data even if it cannot receive all the data slices completely. In previous experiments, the test was the probability of restoring original data, and in this experiment, the test measures the relationship between the number of received slices and the regression degree of the original data.

It is assumed that the source node of the original data is divided into 100 slices, with a packet loss rate setting of 10%. The experiment is executed 10000 times.

Figure 9 and Figure 10 are similar in appearance but the distribution of Figure 10 is more concentrated. When the number of data slices received is more than 40 of 100 slices, the possibility

Figure 9. Relation between the number of slices received and the reduction possibility

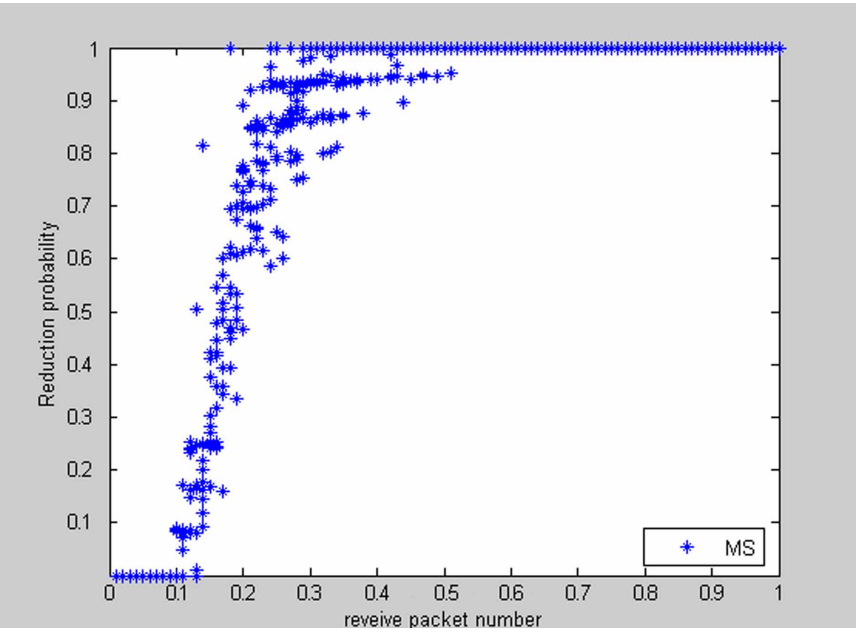
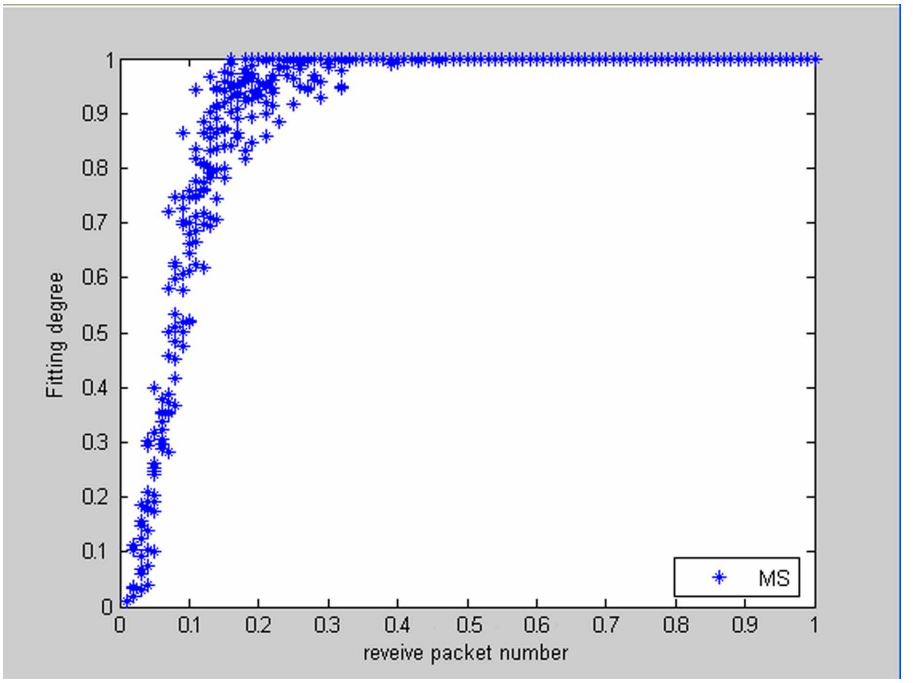


Figure 10. Relation between the number of slices and the regression degree



of restoring the original data is above 98%. The degree of regression begins to decline as the number of data slices received is less than 40. Since each slice carries 10 pieces of hidden data, the destination node may not be able to restore data when 10 successive slices have not been received and the data has not been restored from previous slices. As a result, when the number of data slices received is more than 40, it exhibits regression with an enough high degree.

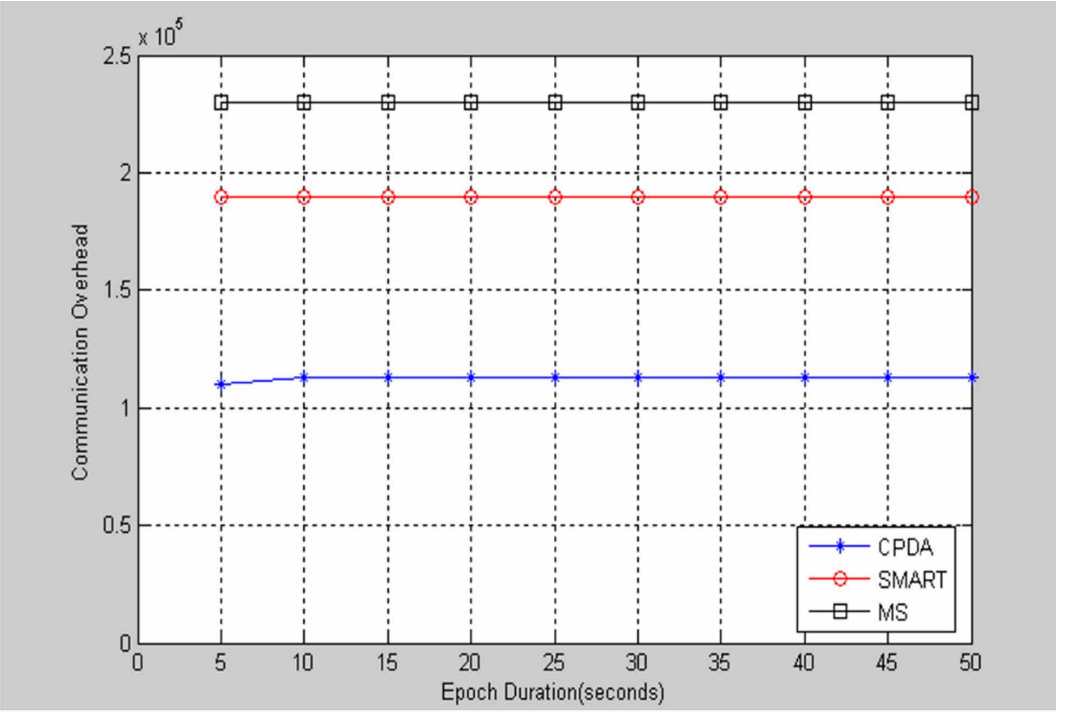
We analyze the case of the proposed MS protocol (Mixed-Slice private protocol) at different packet loss rate, and judge the performance of Data reduction probability and Data regression degree. In Figures 9 and 10. Since this feature of packet loss rate is unique to non-encrypted redundant slice privacy protection algorithms, it is unsuitable to compare it with other privacy protection algorithms. Thus, loss ratio of packet (PLR) is compared by the number of nodes.

PERFORMANCE TEST AND ANALYSIS

Communication Cost

Figure 11 makes a comparison of the communication overhead among the MS, the SMART and the CPDA protocols, proposed by León et al. (2009).

Figure 11. Comparison of MS, CPDA (pc=0.3) and SMART(J=3)



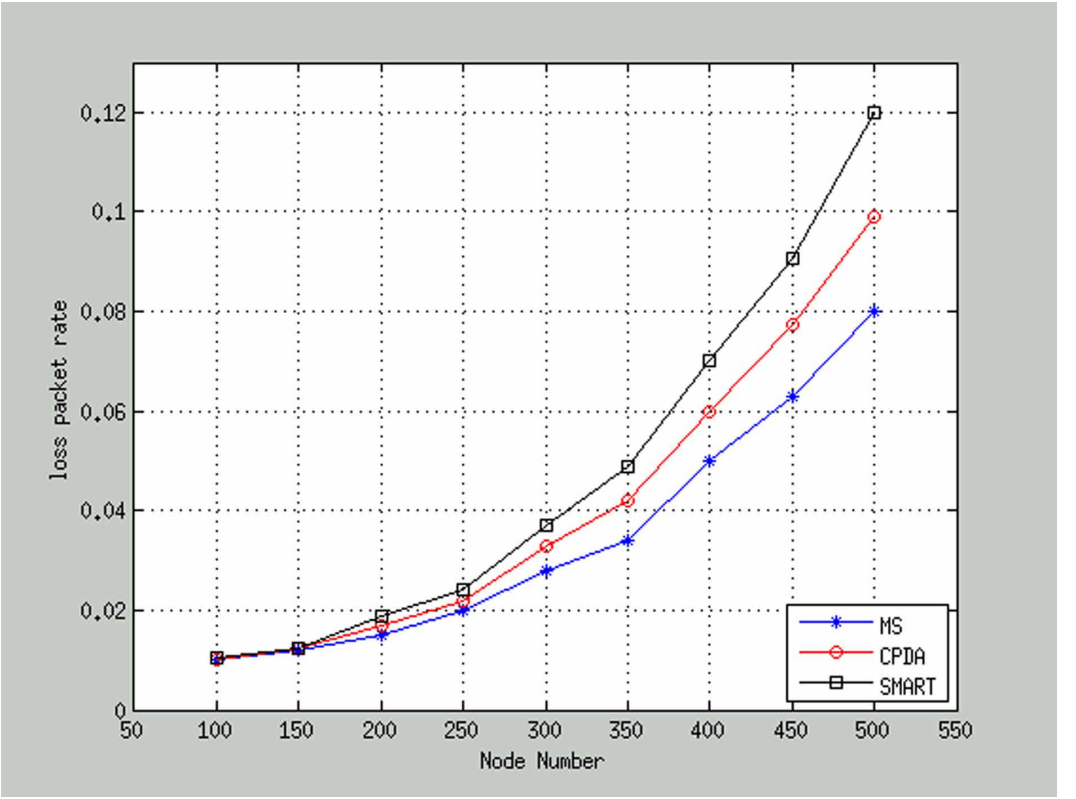
The communication overhead refers to the total bytes of packets per epoch duration time, and Figure 11 shows that the communication cost of CPDA (pc=0.3) protocol and SMART (J=3) protocol is less than MS. As to the three protocols MS, CPDA and SMART compared, communication overhead keeps stable during epoch duration time. The overhead in CDPA protocol is about 1.1×10^5 bytes, SMART protocol is about 1.9×10^5 bytes, while in MS protocol overhead reaches 2.3×10^5 bytes.

The main reason is that MS data restores only in the destination node together with some hidden information data in each slice, which leads to an increased communication overhead.

Packet Loss Rate

The packet loss rate PLR meets the Formula 2 in the paper. From Figure 12 it can be seen that as to the MS protocol proposed, when the node number is 100 the PLR is about 1%, when node number reaches 500 the PLR is about 8%. And in CPDA protocol the PLR varies from 1% to 10%, in SMART protocol the PLR varies from 1% to 12%. Figure 12 indicates that with the increasing of node number, the packet loss in the destination node significantly raises in an exponential manner.

Figure 12. Number of MS nodes and packet loss rate



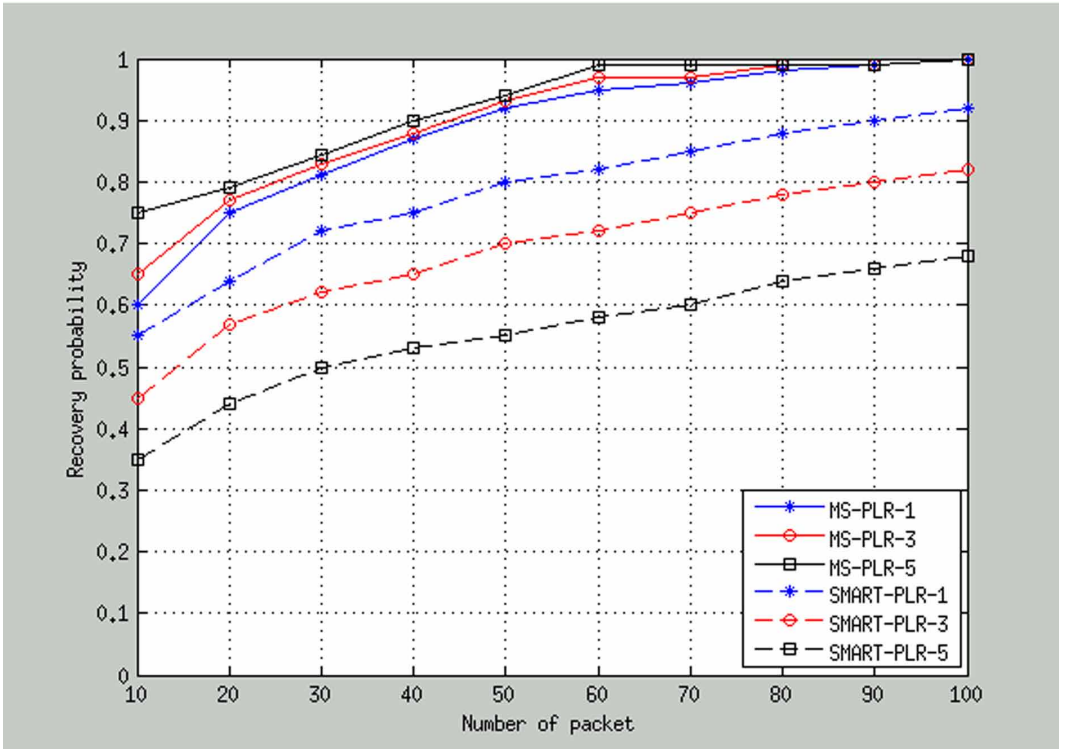
Due to the reason that destination node receives too much data packet so that collisions will increase among data packets. Thus, the packet loss rate increases significantly. The other reason is, the data is divided into slices in the source node, and collisions also increase with the increasing number of packets.

Recovery Probability

When (1-PLR) % of slices are received, the original data can probably be recovered. We perform three level experiments with 1%, 3%, 5% packet loss rate to determine the recovery probability. These processes are repeated for 10000 times.

As shown in Figure 13, we make comparison between the SMART protocol and MS protocol. The x-axis represents the number of received slices, while the y-axis represents the probability of recovering the original message. Case PLR is 1% and more than 60 slices are received, the recovery probability almost reaches 100% in MS protocol, whereas reaches 80% only in SMART protocol. When the number of received slices is between 10 and 50, the recovery probability drops following a curve, and the recovery probability in SMART protocol get worse. Case only 20 slices are received, the recovery probability remains at 80% in MS protocol and 55% in SMART protocol. The recovery probability at 5% PLR is higher than its 3% PLR counterpart.

Figure 13. Recovery probability vs. Number of packets



The performance of MS protocol proposed in this paper is about 20% better than that of SMART protocol in terms of recovery probability under the same PLR. This indicates that the MS protocol proposed is obviously superior to the SMART protocol in recovery probability aspect.

SUMMARY

This paper presents a privacy protection scheme based on slice technology in wireless sensor networks, with the main purpose of overcoming the problem of packet loss that is sensitive to privacy protection algorithms based on slice technology. MS has a very good performance in response to packet loss in order to ensure the integrity of data by carrying hidden information. At the same time, it protects the plaintext that is captured by the sensor nodes in WSN through disturbance technology to further strengthen the privacy protection. Finally, data accuracy and communication cost are compared for MS, SMART, and CPDA protocols, which is executed via simulation experiments together with

analysis of the data reduction probability, data regression degree, and packet loss rate of the MS protocol proposed in the paper.

Further research includes reducing the communication overhead and adopting a homomorphic encryption method to further enhance security. Firstly, the traffic of privacy protection algorithm based on fragmentation technology tends to be large, which leads to a large consumption of limited energy in the wireless sensor network, resulting in a decrease in the life span of a wireless sensor. Secondly, the algorithm designed in this paper can effectively deal with the situation where the intermediation nodes are manipulated. The proposed MS protocol is unsuitable to deal with the falsification of the collected original data in case the source node is being captured by the attacker. Finally, in case the data collected by the source node is very large, it will decline the accuracy of the data protection.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests regarding the publication of this paper.

ACKNOWLEDGMENT

The subject is sponsored by the National Key R&D Program of China (No. 2018YFB1003201), the National Natural Science Foundation of P. R. China (No. 61672296, No. 61602261, No. 61872196, and No. 61872194), Scientific and Technological Support Project of Jiangsu Province (No. BE2019740), Major Natural Science Research Projects in Colleges and Universities of Jiangsu Province (No. 18KJA520008), Six Talent Peaks Project of Jiangsu Province (RJFW-111).

REFERENCES

Acharya, M., Girao, J., & Westhoff, D. (2005, Apr). Secure comparison of encrypted data in wireless sensor networks. In *Proceedings - WiOpt 2005: Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (pp. 47-53). Academic Press.

Ameen, M. A., Liu, J., & Kwak, K. (2012). Security and Privacy Issues in Wireless Sensor Networks for Healthcare Applications. *Journal of Medical Systems*, *36*(1), 93–101. doi:10.1007/s10916-010-9449-4 PMID:20703745

Castelluccia, C., Chan, A. C., Mykletun, E., & Tsudik, G. (2009). Efficient and provably secure aggregation of encrypted data in wireless sensor networks. *ACM Transactions on Sensor Networks*, *5*(3), 1-36.

Conti, M., Willemsen, J., & Crispo, B. (2013). Providing source location privacy in wireless sensor networks: A survey. *IEEE Communications Surveys and Tutorials*, *15*(3), 1238-1280.

Emimanothaya, A., & Babu, R. S. (2017). Maintaining Privacy and Integrity in Two-Tiered Wireless Sensor Network Using Advanced SafeQ Protocol. *Social Science Electronic Publishing*, *1*(3), 16–23.

Fan, Y. J., Chen, H., & Zhang, X. Y. (2012). Data Privacy Preservation in Wireless Sensor Networks. *Chinese Journal of Computers*, *83*(1), 134–144. doi:10.3724/SP.J.1016.2012.01131

Geng, Y., An-Qi, W., & Zheng-Yu, C. (2011). An Energy-Saving Privacy-Preserving Data Aggregation Algorithm. *Chinese Journal of Computers*, *34*(5), 792–800. doi:10.3724/SP.J.1016.2011.00792

Goswami, S., Chandrakar, N., & Dewangan, M. S. (2014). Protecting Location Privacy in Wireless Sensor Networks against Eavesdropper. *Esrsa Publications*, *12*(10), 42–48.

Groat, M. M., He, W-B., & Forrest, S. (2011, April). KIPDA: K-indistinguishable privacy-preserving data aggregation in wireless sensor networks. In *Proceedings - IEEE INFOCOM* (pp. 2024-2032). IEEE.

He, W., Liu, X., & Nguyen, H. (2007, May). PDA: Privacy-Preserving Data Aggregation in Wireless Sensor Networks, *INFOCOM 2007.* In *IEEE International Conference on Computer Communications* (pp. 2045-2053). IEEE. doi:10.1109/INFCOM.2007.237

Hua, P., Liu, X., Yu, J., Dang, N., & Zhang, X. (2018). Energy-efficient adaptive slice-based secure data aggregation scheme in WSN. *Procedia Computer Science*, *129*, 188–193. doi:10.1016/j.procs.2018.03.033

Lee, H. J., Cerpa, A., & Levis, P. (2007). Improving wireless simulation through noise modeling. In *Proceedings of the 6th international conference on Information processing in sensor networks*. ACM.

León, Cosio, Hipólito, Nieto, & García. (2009, September). A security and privacy survey for WSN in e-health applications. In *CERMA 2009 - Electronics Robotics and Automotive Mechanics Conference* (pp. 125-130). Academic Press.

Li, N., Zhang, N., Das, S. K., & Thuraisingham, B. (2009). Privacy preservation in wireless sensor networks: A state-of-the-art survey. *Ad Hoc Networks*, *7*(8), 1501–1514. doi:10.1016/j.adhoc.2009.04.009

Lin, Y., Wang, C., Wang, J., & Dou, Z. (2016). A Novel Dynamic Spectrum Access Framework Based on Reinforcement Learning for Cognitive Radio Sensor Networks. *Sensors (Basel)*, *16*(10), 1–22. doi:10.3390/s16101675 PMID:27754316

Liu, Y., Dong, M., Ota, K., & Liu, A. (2017). ActiveTrust: Secure and Trustable Routing in Wireless Sensor Networks. *IEEE Transactions on Information Forensics and Security*, *11*(9), 2013–2027. doi:10.1109/TIFS.2016.2570740

Ozdemir, S., & Xiao, Y. (2011). Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. *Computer Networks, 55*(8), 735-1746.

Showail, A., Elrasad, A., Meer, A., Daghistani, A., Jamshaid, K., & Shihada, B. (2014). iFrag: Interference-aware frame fragmentation scheme for wireless sensor networks. *Wireless Networks*, *20*(7), 2019–2036. doi:10.1007/s11276-014-0722-1

Sorniotti, A., Gomez, L., & Wrona, K. (2007). Secure and Trusted in-network Data Processing in Wireless Sensor Networks: A Survey. *Journal of Information Assurance & Security*, *37*(2), 189–199.

Vinodha, D., & Anita, E. A. M. (2018). Secure data aggregation techniques for wireless sensor networks: A review. *Archives of Computational Methods in Engineering*, (8), 1–21.

Wang, J., Wang, F., Cao, Z., Lin, F., & Wu, J. (2017). Sink location privacy protection under direction attack in wireless sensor networks. *Wireless Networks*, *23*(2), 579–591. doi:10.1007/s11276-015-1179-6

Wenbo, H., Liu, X., & Nguyen, H. (2007, May). PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks. In *INFOCOM 2007 - IEEE 26th IEEE International Conference on Computer Communications* (pp. 2045-2053). IEEE.

Xie, , S., & Wang, , Y. (2014). Construction of Tree Network with Limited Delivery Latency in Homogeneous Wireless Sensor Networks. *Wireless Personal Communications*, *78*(1), 231–246. doi:10.1007/s11277-014-1748-5

Yang, G., Deng, X., & Liu, C. (2016). Facial expression recognition model based on deep spatiotemporal convolutional neural networks. *Journal of Central South University*, *47*(7), 2311–2319.

Yang, L. J., Ding, C., & Wu, M. (2015). An Efficient and Verifiable Privacy-Preserving Data Aggregation Scheme for Wireless Sensor Networks. *Applied Mechanics and Materials*, *742*, 7–10. doi:10.4028/www.scientific.net/AMM.742.7

Yu, C. M., Chen, C. Y., & Chao, H. C. (2015). Verifiable, privacy-assured, and accurate signal collection for cloud-assisted wireless sensor networks. *IEEE Communications Magazine*, *53*(8), 48–53. doi:10.1109/MCOM.2015.7180507

Zhang, X., Chen, H., & Wang, K. (2014, June). Rotation-based privacy-preserving data aggregation in wireless sensor networks. In *ICC 2014- 2014 IEEE International Conference on Communications* (pp. 4184-4189). IEEE.

Zheng, Z., Sangaiah, A. K., & Wang, T. (2018). Adaptive Communication Protocols in Flying Ad Hoc Network. *IEEE Communications Magazine*, *56*(1), 136–142. doi:10.1109/MCOM.2017.1700323

Zheng, Z., & Zheng, Z. (2018). Towards an Improved Heuristic Genetic Algorithm for Static Content Delivery in Cloud Storage. *Computers & Electrical Engineering*, *69*(7), 422–434. doi:10.1016/j.compeleceng.2017.06.011

Peng Li was born in 1979. He received the Ph.D. degree from Nanjing University of Posts and Telecommunications (NUPT) in 2013. Currently, he is a professor and Master Supervisor in NUPT. He has managed a number of research projects supported by the National Natural Science Foundation of China and research projects supported by the Ministry of Jiangsu Provice. His main research interests include computer communication networks (especially in Wireless Sensor Networks, Ad Hoc Network) and information security.

Chao Xu was born in 1991. He is a master student of NUPT. His research interests include software engineering and high-performance parallel computing and security.

He Xu was born in 1985. He received the Ph.D. degree from NUPT in 2012. Currently, he is an associate professor and Master Supervisor in NUPT. His main research interests include Parallel Computing technology and Information Security.