Research on Application of Improved Smart Data Aggregation Algorithm in Wireless Sensor Network

Wang Pu, Fang Liying, and Yu Lei
Department of college of Electrical Information and Controlling Engineering, Beijing University of Technology, Beijing, China

***Abstract*—Aiming at the existing problems such as high shard collision rate and so on of Smart algorithm in the data aggregation of the wireless sensor network, this paper puts forward the improved Smart data aggregation algorithm. The algorithm in this paper adds optimization factors from five aspects, and the addition of the various optimization factors makes the accuracy be improved, the communication traffic of the scheme also be effectively reduced at the same time, which indirectly prolongs the survival time of the network. The experimental results show that the improved algorithm of this paper can improve the accuracy and security of the data aggregation.**

***Index Terms*—Wireless Network; Link; Small Data Factor; Node**

I. INTRODUCTION

The wireless sensor network is an important part of the Internet of things. The wireless sensor network is composed of a large number of sensor nodes, and the energy, computing, storage, and communication ability of each node are limited. The wireless sensor network has its corresponding application in many fields, and the environmental monitoring is one of them, it specifically includes perceived exercise, temperature testing, and humidity and so on. In order to save the energy and the communication bandwidth, and reduce the sending of the original data, the sensor nodes need to collaborative process the collected original data in the network, and data aggregation is one of the methods for processing within the network [1-3]. Due to the open nature of the wireless sensor network, the data are easy to be captured and intercepted in the process of data aggregation. If the attacker breaks the wireless link or captures the nodes in the network, the data in the network is exposed. The privacy protection technology of data aggregation in the wireless sensor network usually uses the methods such as encryption, or data perturbation and so on to protect the original data, under the condition of ensuring the correct result of data aggregation [4-7].

Wireless sensor network is an important part of the Internet of things. A wireless sensor network is a self-organized network which is composed of a large number of sensor nodes. These sensor nodes are placed in a variety of realistic environments, such as some environments which is difficult for the humans to visit, and they are used to collect the specified information, making it convenient for people to analyze and manage the information [8-10]. Wireless sensor network has its corresponding application in many fields, and environmental monitoring is one of them, which specifically includes the sensorimotor, the test of temperature, as well as the humidity and so on.

Security problem is one of the concerns of the data aggregation in the wireless sensor network, and the privacy-preserving data aggregation algorithm can be divided into the unencrypted and encrypted data aggregation algorithm. The early researches mostly focus on the unencrypted data aggregation algorithm, which realizes the hiding of the original data by using the modification operation of the data, but does not have an ideal privacy protection. And the scheme based on the encrypted privacy-preserving data aggregation can be divided into the end-to-end encryption and hop-by-hop encryption. The end-to-end encryption is to build the secure link between each node and the Base Station (BS) node, the private data of each node is encrypted before being sent upward, obtain the BS node of the encryption package and then use the keys negotiated with the nodes to extract the real data, so as to make the intermediate nodes to be transparent in the process of communication. However, the average end-to-end encryption is unable to implement the data aggregation, which makes the nodes near the BS have a too big energy loss due to the too frequent execution of the sending operation, thus the communication efficiency is relatively low. In order to solve this problem, the privacy-preserving scheme proposed by Castelluccia and others introduced the homomorphic encryption technology, which realizes the end-to-end aggregation encryption and makes the data can be directly aggregated without decrypting [11]. In the hop-by-hop encryption scheme, each node is decrypted after receiving the aggregated data packet, and aggregates with the original data, after the encryption the node aggregates upward, and the whole encryption and decryption process is based on some kind of key distribution scheme mentioned above. In terms of the hop-by-hop encryption, Yi y and others studied the data aggregation algorithm based on the hop-by-hop encryption. Due to the hop-by-hop encryption does

© 2014 ACADEMY PUBLISHER
doi:10.4304/jnw.9.7.1875-1882

through the intermediate decryption process during the aggregation, and compared with the end-to-end encryption the privacy protection of the hop-by-hop encryption is reduced [12-13]. Although the end-to-end aggregation encryption is more efficient, Feng T M put forward some problems need to be solved in the process of the end-to-end aggregation encryption. In addition, Wang C and others proposed the aggregation algorithm that gives consideration to both intrusion detection and privacy protection [14-16].

The PDA privacy-preserving data aggregation scheme proposed by He W B and others studies the additive aggregation function sum, which specifically includes two algorithms, the Cluster-based Private Data Aggregation (CPDA) and the Slice Mix Aggregate (SMART). The former is the data aggregation algorithm based on the clustering, and its calculated process is complex and has a large amount of calculation. The SMART algorithm is closely associated with this paper. The privacy protection of the scheme depends on the amount of the shards, usually the effect of privacy protection cab be achieved when there is $S \geq 4$. The greater the value of J is, the better the privacy protection will be, but it will make the amount of the shards sent by each node increases, and the communication traffic will increase accordingly.

Bista R and others proposed a privacy-preserving scheme which is based on the plural algebraic expression. Perrig A and others studied the non-linear aggregation functions, such as max, min and so on, rather than the sum aggregation function, and put forward the privacy-preserving scheme under the corresponding situation. Because the SMART algorithm divides the sensory data into J pieces and then sends the shards, it causes the collision lost problem of the data, and eventually results in the decreasing of the accuracy of the data aggregation. Therefore, to make clear the relationship between the number of the shards with the collision rate and the aggregation accuracy has always been the problem researched and thought by people, and it is also the starting point of this paper.

Aiming at the problem of low accuracy of the SMART scheme, this paper adds multiple optimization factors in the original algorithm to produce a series of new algorithms, and compared with SMART these algorithms have a relatively great improvement in accuracy. At the same time, the communication traffic of the algorithm added in the local factor is relatively reduced, which indirectly prolongs the survival time of the network. In terms of privacy protection, it can be strengthened by using the methods of homomorphic encryption or hop-by-hop encryption. This paper will carry out the performance analysis of the algorithm added in the factors from four aspects, the privacy protection, the communication traffic, the receiving rate of shards and the accuracy.

This paper mainly has development and innovation works in the following aspects:

(1) The wireless sensor network is an important part of the Internet of things, and to add in the privacy protection mechanism at the same time of data aggregation is one of the means of the privacy protection and security of the wireless sensor network. Aiming at the existing problems such as high shard collision rate and so on of Smart algorithm in the data aggregation of the wireless sensor network, this paper puts forward the improved Smart data aggregation algorithm. The algorithm adds in multiple optimization factors to form the new privacy-preserving data aggregation algorithm of the sensor network. The key point of these algorithms is to improve the accuracy and security of the data aggregation, and the algorithm added in the local optimization factor forms a new privacy protection scheme at the same time of guaranteeing high accuracy, which effectively reduces the communication traffic, thus reduces the energy expenditure of the nodes and indirectly prolongs the survival time of network.

(2) In order to further validate the correctness and effectiveness of the improved Smart algorithm proposed in this paper, the simulation experiments in the aspects of privacy protection, communication traffic, receiving rate of shards, accuracy and so on are carried out. The simulations adopt the TOSSIM software under the TinyOS, and summary the comparisons between adding each factor with the Smart algorithm in terms of performance. The TAG algorithm does not have privacy protection, therefore, compared with PEPDA algorithm it obviously has certain advantages in the aspects of communication traffic, receiving rate of shards and so on.

II. Model of Wireless Sensor Network

The sensor network is abstracted as a connected graph $s(a,b)$, among which a is the set of the sensor nodes, $|a|=b|$ is the number of the sensor nodes, and edge b is the communication link between the sensor nodes.

Define that the aggregation function is $s(u)=t\left(r_1(u)+r_2(u)+...+r_b(u)\right)$, and $r_i(u)=(i=1,2,...,b)$ is the data collected by node i in t time. The typical integration functions are sum, average, max and so on. This paper only discusses the sum function, and it is marked as follows:

$$s(u)=\sum_{i=1}^{b} f(r_i(u)) \tag{1}$$

As shown in figure 1, because other functions can be simplified as the sum function.

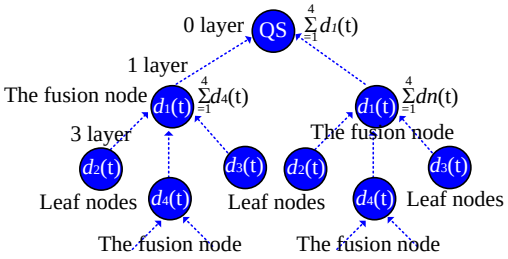


Figure 1. The schematic diagram of data aggregation

© 2014 ACADEMY PUBLISHER

Define that the time needed for the whole data aggregation is Epoch Duration, and it can be divided into three time slices, the building, the complicity and the aggregation. The building time slice executes and establishes the aggregation tree operation. In the complicity time slice, each node divides and sends its own data, and in the aggregation time slice, each node carries out the data aggregation along the aggregation tree.

Like SMART algorithm, this paper uses the method of random key distribution to carry out the encryption and decryption of the data aggregation. The method includes the following three steps:

1) Produce a key pool with k keys, and randomly select k keys from the key pool.

2) Each node determines the neighbor nodes that with the same keys with itself by exchanging information. Two nodes with the same key marked as k_i share a secure link.

3) For two neighbor nodes that do not have the same keys, a secure link formed by multiple hops link can be set up.

III. IMPROVED SMART ALGORITHM

A. Smart-L Algorithm

If the collision rate is stable, from the perspective of reducing the losses caused by the shard collision, the accuracy can be improved by using the method of restricting the size of the data. The algorithm added in the small data factor is shorthanded as SMART-L. The algorithm reduces the losses caused by collision within the complicity time slice to improve the accuracy.

Algorithm 1: the SMART-L algorithm.

1) The building preparation stage. Use the TAG algorithm to build the data aggregation tree, and each node $i(i=1,2,...,b)$ randomly select a node set Si within h hops.

2) The data collaboration stage. Assume that the network is to collect the human body temperature, and the original data collected by each sensor node has certain upper and lower bound, which is respectively set to 35 and 43. Then the designed small data factor F= (upper bound-lower bound) / $(s-1)$, namely $r=9/(s-1)$. Each node randomly selects $(s-1)$ target nodes from the node set Si, calculates to produce $(s-1)$ shards and sends them to the target nodes. The calculation method of the size of each shard is that Slice $i=r(w,w\in 0,1)$, in which i shows the i_{th} shard.

3) The data aggregation stage. Each node uses the TAG algorithm to carry out the data aggregation.

Statement:

(a) The adding of the factor will not cause the additional overhead. The difference between the SMART -l algorithm and SMART algorithm is mainly embodied in the calculation of the size of the shard, set the value of J to 3, and then the new algorithm is as follows:

$$Slice_i = 4\times w, w\in 0,1 \quad (2)$$

In which i shows the i_{th} shard.

And the original SMART algorithm is as follows:

$$Slice_i = D_s^{remained}\times w, w\in 0,1 \quad (3)$$

Among which, i shows the i_{th} shard, and $D_s^{remained}$ represents the remaining amount of data of the s_{th} node.

(b) The adding of the small data factor can reduce the losses caused by collision. In SMART-L algorithm, if the shard is missing due to the collision, then the lost data range of the node for sending the shard is (0, 9), namely the maximum loss of data is 9. While in the original SMART scheme, the range of the original data is (37, 45), then the maximum loss of data is 45, which is more than five times than the new algorithm. The SMART – L algorithm will be able to reduce at most 6/7 of the loss of data of the SMART algorithm.

B. Smart-PN Algorithm

The algorithm added in the positive and negative factors is shorthanded as the SMART-PN (SMART - Positive and Negative). The algorithm also tries to reduce the losses caused by collision within the complicity time slice. Before the specific introduction of the SMART-PN algorithm, the data aggregation situations of the following kinds of shards are observed.

As shown in figure 2, each case takes three nodes as an example. Each node divides the private data into three pieces, and sends two pieces to the other nodes. Assume that the original data collected by each node is 10, all the shards are lost, the data in the direction of out-degree is the value of shard sent by the node, and the data in the direction of in-degree is the value of shard received by the node, then the situations are as follows:

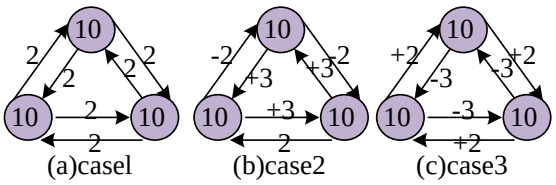


Figure 2. The situations of multiple shards

$$Case1.\ Accuracy\ p_{case1} = \frac{(10-5)\times 4}{10\times 4} = 0.7 \quad (4)$$

$$Case2.\ Accuracy\ p_{case2} = \frac{[10-4-(-2)]\times 4}{10\times 4} = 0.96 \quad (5)$$

$$Case3.\ Accuracy\ p_{case3} = \frac{[10-(-4)-2]\times 4}{10\times 4} > 1 \quad (6)$$

Accuracy

It can be obtained by observing case1 and case2 that if send the negative shards and there is collision occurs, then the final aggregation value may increase, and the precision is improved. Because if the negative shards are missing during the complicity period, it is equivalent that

© 2014 ACADEMY PUBLISHER

the overall value of the data aggregation is made up, and eventually the accuracy is improved.

It can be obtained by observing case3 that if the $\sum$ negative shard | is greater than the $\sum$ positive shard, then the precision may be distorted. The SMART-PN algorithm is given based on the above analysis.

Algorithm 2: the SMART-PN algorithm.

1) The building preparation stage. It is the same with the first step of the SMART-L algorithm.

2) The data collaboration stage. Each node randomly selects $(s-1)$ target nodes from the node set Si, calculates to produce $(s-1)$ shards and sends them to the target nodes. The calculation method of each shard is as follows: set that there are N nodes, each node is divided into J pieces, and the original data collected by each node is $t_i=(i=1,2,...b)$. The $(s-1)$ shards are set to $slice_1, slice_2, slice_{s-1}, w\in(0,1)$, and among which there are:

$$\begin{aligned} &Slice_1 = t_i \times r = t_i \times (-1)^{1-1} \times r^1 = t_i \times r^1, \cdots\cdots Positive \\ &Slice_2 = t_i \times r \times (-r) = t_i \times (-1)^{2-1} \times r^2 = -t_i \times r^2, \cdots\cdots Negative \\ &Slice_3 = t_i \times r \times (-r) \times (-r) = t_i \times (-1)^{3-1} \times r^3 = t_i \times r^3, \cdots\cdots \\ &Slice_{j-1} = D_i \times (-1)^{j-2} \times r^{j-1} \end{aligned} \quad (7)$$

3) The data aggregation stage. It is the same with the third step of the SMART-L algorithm.

C. Smart-CPN Algorithm

The algorithm added in the compensation factor is shorthanded as SMART-CPN (SMART- Compensation), and the algorithm is based on the SMART-PN algorithm. The SMART-CPN algorithm also hopes to reduce the losses caused by collision, but within the aggregation time slice it improves the accuracy by compensating the losses caused by collision.

Algorithm 3: the SMART-CPN algorithm.

1) The building preparation stage. It is the same with the first step of the SMART-L algorithm.

2) The data collaboration stage. Each node randomly selects $(s-1)$ target nodes from the node set Si, calculates to produce J-1 shards and sends them to the target nodes. The calculation method of shard of each node is the same with the SMART-PN algorithm. If each node receives the shard bag sent to it by other nodes within the data collaboration period, then it will input the source address of the shard bag to the "ACK sending table", that is to say, it uses the "ACK sending table" to record the source addresses of the shards received by the node.

3) The ACK sending stage. Each node sends the ACK data packet to the target node according to its own recorded "ACK sending table". When the stage is complete, each node can calculate the "shard sending loss rate" through the number of the received ACK data packets. If there is J-1 shards need to be sent, the number of the received ACK data packets is *ackBack*, and then the shard loss rate of the node is as follows:

$$R_i = \frac{s-1-ackBack}{s-1} \quad (8)$$

4) The data aggregation stage. Each node intelligently carries out a certain amount of compensations during the data aggregation according to its own "shard sending loss rate", and the amount of compensations w_i is as follows:

$$t_i \times (w-w^2) \times \frac{s-1-ackBack}{s-1} \quad (9)$$

Di is the number of the original data collected by node i. That is to say, in the aggregation stage before carrying out the data aggregation with the father node, node i firstly adds the amount of compensations w_i.

D. Smart-T Algorithm

The algorithm added in the random shard factor is shorthanded as SMART-T (SMART-Time slot randomized). The algorithm carries out the optimization of the algorithm from the perspective of reducing the collision rate.

It can be known from figure 3 that within the collaboration time slice of the SMART algorithm, each node almost sends the shard at the same time, the specific sending time is to send the shard 1 at the beginning of the collaboration time slice, and send the shard 2 at the half moment of the collaboration time slice, which leads to large collision probability and shard lost, so as to affect the aggregation result, and ultimately affect the accuracy.

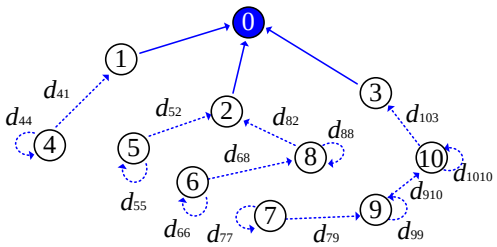


Figure 3. The contrast diagram of the shard sending moment

E. P-SMART Algorithm

The P-SMART algorithm added in the local factor is also to improve the accuracy from the perspective of reducing the collision rate, but it specifically reduces the collision rate through decreasing the communication traffic. It is noticed that in the SMART algorithm, node set Si is determined in the building time slice, and set w_i for the number of nodes in the node set w_i. When the number of the shards J is 3 and the number of the nodes is 60, there are will be 24 nodes whose value of w_i is less than 2 in average. $w_i \prec 2$ shows that the node can not find two target nodes to transmit the according shard, namely the node will send the two shards to the same destination node. Theoretically, each node has two out-degree links, while in fact the two out-degree links point to the same destination node. The nodes that fail to satisfy the needs of $w_i \geq J-1$ are defined as the "shard

© 2014 ACADEMY PUBLISHER

failure nodes". It can be known from figure 4 that when there is $J \geq 11$, almost all the nodes are the shard failure nodes. The P-SMART algorithm is to make the shard failure nodes fail to carry out the data fragmentation, and be sent to the neighbor node as a whole data.

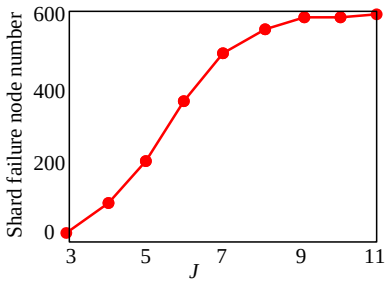


Figure 4. The variation tendency of shard failure nodes

1. The building preparation stage. Use the TAG algorithm to build the data aggregation tree, and carry out the classification of the nodes at the same time. For the nodes with $w_i \geq J-1$, they are divided into the node set Tp. And accordingly, the nodes with $w_i \leq J-1$ are divided into the node set t_r.

2) The data collaboration stage. Each node w in the node set T_w randomly divides the data into J pieces, among which one piece stays in node w itself, and the residual $J-1$ pieces randomly select $J-1$ nodes from the set Sp, and they distributively share the encrypted keys and send the keys to the selected nodes.

3) The data aggregation stage. It is the same with the third step of the SMART-L algorithm.

IV. Experimental Simulations and Analysis

A. Experimental Environments and Settings

This section mainly evaluates the performance of the algorithm added in various kinds of optimization factors mentioned above from four aspects, the privacy protection, the communication traffic, the receiving rate of shard and the accuracy, and carries out the comparison analysis with the TAG algorithm and the SMART algorithm. The simulation uses the TOSSIM software under the TinyOS, and the specific environment configurations of the network are as follows: random distribution of 608 nodes in the area of 406 x 406, the background noise is -105DBM, and the Gaussian white noise is 4DB.

B. Results Analysis

1) Privacy Protection

In the SMART algorithm, the private data of each node is divided into J pieces, and the node sends J-1 shards to other nodes. Only when the listener obtains all the $J-1$ out-degree links of the node and all the in-degree links can the listener grasps the node.

For a node, the probability of its collected data being cracked and stolen is recorded as P (q), among which q shows the probability of the links between the nodes being cracked, if there is $y \approx poverhear$, then there is a formula as follows:

$$P(y) = y^{j-1} \sum_{k=0}^{d_{\max}} P(indgree = k) y^k \qquad (10)$$

Among which, $t_{\max}$ represents the largest out-degree value in the network, and $p(in\deg ree = k)$ is the f probability of the nodes with the in-degree of k.

In addition to the P-SMART algorithm, the privacy protection of the SMART-l, SMART-PN, SMART - CPN and SMART-T algorithm is the same with the SMART algorithm, and it complies with the above formula (9).

For the P-SMART algorithm, its privacy protection is from two aspects.

1) In the P-SMART algorithm, the nodes are divided into the node set T_w and F_r. The node set T_p inherits the shard feature of the SMART algorithm, whose privacy protection complies with the above formula (9). While for the node set Fr without division, its privacy protection is in accordance with the following equation:

$$P(y) = \sum_{k=0}^{t_{\max}} P(indgree = k) y^k \qquad (11)$$

Then for all the nodes, their privacy protections are between formula (9) and formula (10).

2) In the node set Fr without division, the privacy protection of each node can be further enhanced through the end-to-end encryption of the homomorphic encryption, or just through the general hop-by-hop encryption. At the same time, the BS nodes have the possibility of involving in the fragmentation, but they may be unable to obtain the real data even if they know about the keys of the end-to-end encryption. Moreover, because the nodes in the node set T_p involve in the fragmentation, while the data of the nodes in the node set Fr do not involve in the fragmentation, but the nodes in the node set Fr will also receive the shards sent from the nodes in the node set T_p, which ensures the privacy protection of the data aggregation, that is to say, the privacy protection will also be affected by formula (8).

Because the TAG algorithm does not have the privacy protection strategy, the privacy protections of the above factors are better than that of the TAG algorithm.

2) Communication Traffic

The communication traffic can be divided into the communication traffic of the shard sending and the communication traffic of the data aggregation. Because the root node does not need to carry out the data aggregation upward, therefore the communication traffic of the data aggregation is N-1. When the number of the nodes in the network N is given, the communication traffic of the data aggregation is a fixed value. Thus, the communication traffic of the shard sending is mainly discussed, this is because the SMART algorithm wants to strengthen the additional fragmentation overhead brought by the privacy protection, and only the additional communication traffic of the SMART-CPN algorithm is

© 2014 ACADEMY PUBLISHER

the sum of the communication traffic of the shard sending and the communication traffic of the ACK sending. Because the TAG algorithm does not have the additional communication traffic of the fragmentation, therefore, there is no comparison with the communication traffic of the TAG algorithm.

Figure 5 (a) shows the communication traffics of the SMART, P-SMART and SMART-CPN algorithm, when the amount of fragmentations J is 3 and the number of nodes N is 603. Because the SMART-L, SMART-PN and SMART-T algorithm will not change the shard number of each node of the SMART algorithm, their communication traffics are consistent with that of the SMART algorithm, and the theoretical value should be $1231 = N(s-1)$, but due to the effects of the real network environments, the theoretical value can not be achieved. Compared with the SMART algorithm, the communication traffic of the P - SAMRT algorithm is decreased, but not obvious, and it is mainly due to the influence of J. Figure 5 (b) shows the changing curve of the communication traffic of the P-SAMRT algorithm, when the number of shards J is from 3 to 11. It can be found from the curve that the communication traffic increases firstly, then reduces, it achieves the peak when there is $s = 6$, and is less than 1200 after there is $s \geq 6$. The theoretical communication traffic of the algorithm is not $N \times (s-1)$ any more, but $L = |T_p| \times (s-1)$. It can be known from figure 4 that with the increasing of J, the value of $|T_p|$ reduces, therefore, the above tendency of increasing firstly then decreasing of L is existed. The additional communication traffic of the SMART-CPN algorithm compared with the SMART algorithm is the additionally added communication traffic used for sending ACK. Figure 5 shows the communication traffic of the algorithm.

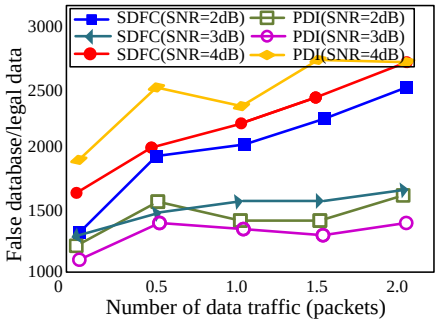


Figure 5. The communication traffic

3) Receiving Rate of Shard

The collision rate of shard is equal to 1-receiving rate of shard, and they have effects on each other. Thus, it is also very important to study the receiving rate of shard, and the receiving rate of shard is defined as follows:

$$p_r = \frac{\sum_{i=1}^{N} \text{Number of shards actually received by each node}}{\sum_{i=1}^{N} \text{Number of shards actually sent by each node}} \quad (12)$$

Figure 6 (a) shows the receiving rate of shard of SMART, SMART-T, P-SMART and TAG algorithm, when the amount of fragmentations J is 3 and the number of nodes N is 600. Because the SMART-L, SMART-PN and SMARTCPN algorithm carry out the optimization from the perspective of reducing the losses caused by collision, and these algorithms do not change the collision rate, so the receiving rates of shard of the above three kinds of algorithms are the same with that of SMART algorithm, which is shown in figure 6 (a). Compared with the SMART algorithm, the receiving rate of shard of the SMART-T algorithm increases obviously, and the receiving rate of the TAG algorithm is the highest, which is mainly because the SMART-T algorithm has certain privacy protection, so as to sacrifice the accuracy of the receiving rate of some shards. The improvement of the receiving rate of shard of the P-SMART algorithm is not obvious, this is because when there is $s = 4$, the value of $|F_r|$ is not large, and the corresponding reduction of shard is not obvious, thus the receiving rate of shard is not improved greatly. However, it can be known from figure 6 (b) that with the increasing of s, the receiving rate of shard improves significantly.

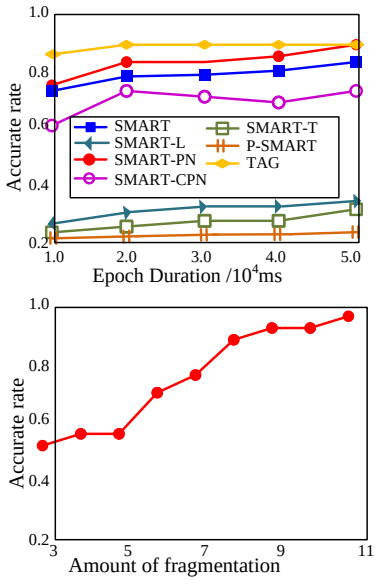


Figure 6. The receiving rate

4) Accuracy

Accuracy is the key optimization performance of the various factors in this paper. Due to the research of this paper is based on the sum function, and the accuracy is defined as follows:

$$p_c = \frac{\text{Data of root nodes after aggregation}}{\sum_{i=1}^{N} \text{Original data}} \quad (13)$$

Figure 7 (a) is the contrast figure in the aspect of accuracy performance between the 5 kinds of algorithms added in all kinds of optimization factors and the SMART and TAG algorithm. It can be known from the figure that compared with the SMART algorithm, the

© 2014 ACADEMY PUBLISHER

accuracies of the SMART-L, SMART-PN, SMART-CPN and SMART-T algorithm all have a certain degree of improvement, and the accuracy of the SMART algorithm is basically the same with that of the P-SMART algorithm, which is about 0.45. When there is $s=4$, the P-SMART algorithm does not have obvious improvement in the aspect of accuracy, but as shown in figure 7 (b), the P-SMART algorithm is very sensitive to the value of s, and with the increasing of J, the accuracy is significantly improved. When the value of s is 8, the accuracy curve tends to be gentle.

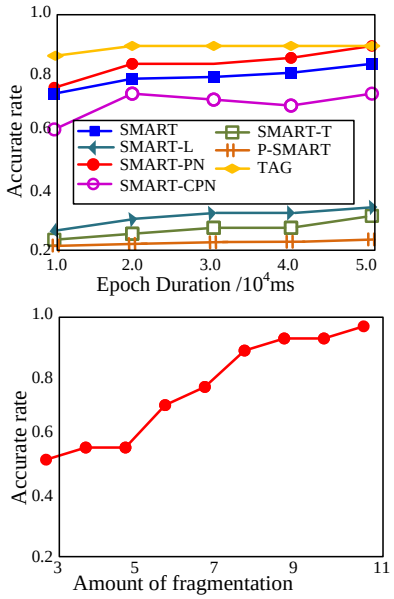


Figure 7. The accuracy

For the P - SMART algorithm, its communication traffics are all bigger than that of the SMART algorithm when the value of J is from 4 to 6, but due to the value of $|F_r|$ is respectively 93, 216 and 359, the random distribution of the nodes without fragmentation whose numbers can not be ignored in the network reduces the probability of the collision between the data of the shards, which explains why the receiving rate of shard and the accuracy still maintains monotonically increased, under the situation that the communication traffic has a peak value when there is $s=7$.

5) Algorithm with Multiple Factors

Record that SMART-LPN is the algorithm combined by the small data and the positive and negative factors; SMART-CLPN, the algorithm combined by the small data, the positive and negative factors and the compensation factor; SMART-CLPNT, the algorithm combined by the small data, the positive and negative factors, the compensation factor and the random shard factor; P-SMART-CLPNT; the algorithm contains all of the five optimization factors.

Figure 8 (a) shows the accuracy of each algorithm when adds the small data factor, the positive and negative factors, the compensation factor, and the compensation factor one by one, with the amount of fragmentations J is 3 and the number of nodes N is 608. Each accuracy curve has different degree of volatility, but it does not affect the change tendency of the curve. It can be known from the figure that with the increasing of the added factors, the accuracy of each algorithm improves step by step. The addition of the local factor makes the final accuracy of the P-SMART-CLPNT algorithm be further improved, which is shown in figure 8 (b). When there is $s \geq 8$, the precision curve of the algorithm tends to be gentle. And when there is $s=8$, the accuracy is 0.9231, the communication traffic of shard is 330, the receiving rate of shard is 0.9938, the privacy protection has the double security of the SMART scheme and the end-to-end aggregation encryption, and the total amount of the outside communication traffic is 330 (the communication traffic of shard) + 329 (the communication traffic of ACK) $\approx$ 660, which is far less than the communication traffic 1198 of the SMART algorithm when there is $s=4$, thus the energy cost of the node is effectively optimized.

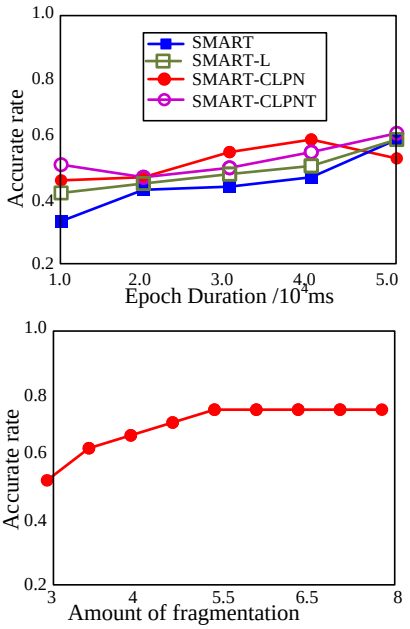


Figure 8. Accuracy of the algorithm with the mixed factor

TABLE I. COMPARISONS BETWEEN THE ALGORITHMS ADDED IN EACH FACTOR AND THE SMART ALGORITHM

Types	Privacy Protection	Communication Traffic	Receiving Rate of Shard	Accuracy
Random	The same as SMART	1041	0.7612	0.4324
Local	Introduced in End-to-end	1052	0.8532	0.8821
Small Data	The same as SMART	1076	0.9326	0.8973
Positive and Negative	The same as SMART	1098	0.9641	0.9023
Compensation	The same as SMART	1187	0.9832	0.9198
All of the Factors	Introduced in End-to-end	1198	0.9938	0.9226

The comparisons between the algorithm added in each factor and the SMART algorithm is summarized from

© 2014 ACADEMY PUBLISHER

table 1 in terms of performance. Because the TAG algorithm does not have the privacy protection, therefore compared with the PEPDA algorithm the TAG algorithm has certain advantages in the aspects of the communication traffic, the receiving rate of shard and so on.

V. CONCLUSION

In this paper, the five aspects of the SMART algorithm are optimized, and the adding of the optimization factor makes the accuracy be improved. The optimization factors are added in from five aspects, and the addition of the various optimization factors makes the accuracy be improved, the communication traffic of the scheme also be effectively reduced at the same time, which indirectly prolongs the survival time of the network. And it is also a way to construct new algorithm by including multiple factors. The experimental results show that the improved algorithm in this paper can improve the accuracy and security of the data aggregation.

REFERENCES

[1] X. L. Bai, C. L. Zhang. Low-Connectivity and Full-Coverage Three Dimensional Wireless Sensor Networks. *The 10th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2009).* ACM 2009, 978-1-60558-531-4

[2] X. L. Bai, C. L. Zhang. Full-Coverage and k-Connectivity (k =14,6) Three Dimensional Networks. *IEEE INFOCOM*, 2009, 978-1-4244-3513-5.

[3] DI Kaichang, GE Zhijiang. A brief review of Spirit's six years of Mars roving and scientific discoveries. *Journal of Remote Sensing*, 2011, 15(4) pp. 651-658.

[4] S. Duquennoy, G. Grimaud, J. Vandewalle. Smews: Smart and Mobile Embedded Web Server. *International Conference on Complex, Intelligent and Software Intensive Systems (CISIS'09).* 2009 pp. 571-576.

[5] B. Yan, G. W. Huang. Supply Chain Infomation Transmission Based on RFID and Internet of Things. ISECS International Colloquium on Computing Communication, *Control and Management* 2009 pp. 166-169.

[6] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects Across Visual Sensor Networks. *Journal of Multimedia*, Vol 7, No 1 (2012) pp. 2-8

[7] W. D, William, Phillips, R. Sankar. An Internet Overlay Architecture for Global Scale Wireless Sensor Networks. *Wireless Telecommunications Symposium*, 2010 pp. 1-6.

[8] J. Kulik, W. Heinzelman, H. Balakrishnan. Negotiation-based Protocols for Disseminating Information in Wireless Sensor Networks. *ACM Wireless Networks*, 2012, 8(2, 3) pp. 169-185.

[9] K. Sohrabi, J. Gao, V. Ailawadhi, et al. Protocols for Self-Organization of A Wireless Sensor Network. *IEEE Personal Communications*, 2000, 7(5) pp. 16-27.

[10] S. Lindsey, C. S. Raghavendra. PEGASIS: Power-Efficient Gathering in Sensor Information Systems. *In: Proc of the IEEE Aerospace Conf, Montana, USA, March* 2011.

[11] T. Melodia, D. Pompili, I. F. Akyildiz. On the Interdependence of Distributed Topology Control and Geographical Routing in Ad Hoc and Sensor Networks. *IEEE JSAC*, 2005, 23(3) pp. 520-532.

[12] W. H. Liao, Y. C. Tseng, A. P. Sheu. GRID: A Fully Location-Ware Routing Protocol for Mobile Ad Hoc Networks. *Journal of Telecommunication Systems*, 2011, 18(l) pp. 37-60.

[13] W. H. Liao, Y. C. Tseng, K. L. Lo, Sheu J P. GeoGRID: A Geocasting Protocol for Mobile Ad Hoc Network Based on GRID. *Journal of Internet Technology*, 2010, l(2) pp. 23-32.

[14] Y. Zou, K. Chakrabarty. Sensor Deployment and Target Localization Based on Virtual Forces. *Proc. of the IEEE INFOCOM New York, USA: IEEE*, 2009 pp. 1293-1303.

[15] S. DAI, C. TANG, S. QIAO, et al. Optimal Multiple Sink Nodes Deployment in Wireless Networks Based on Gene Expression Programming. *2010 Second International Conference on Communication Software and Networks*, 2010 pp. 355-359.

[16] Y. GU, Q. WU, X. CAI, B. Jeremy. On Efficient Deployment of High-End Sensors in Large-Scale Heterogeneous WSNs. *Mobile Adhoc and Sensor Systems, 2009. MASS'09. IEEE 6th International Conference on*, 2009 pp. 912-917.

© 2014 ACADEMY PUBLISHER