WSN Data Transmission Algorithm Based on Spatial Data Aggregation

https://doi.org/10.3991/ijoe.v13i12.7884

Liang Hu, Huayi Yin(✉), Chaoqun Hong
Xiamen University of Technology, Xiamen, Fujian, China
Huayi@xmut.edu.cn

Abstract—In this paper, considering the high energy consumption, loss of network lifetime and data leak in transmission during the data aggregation of the wireless sensor network, we propose an improved spatial data aggregation algorithm. Through comparison with traditional data aggregation algorithms, we verify the feasibility and rationality of the proposed algorithm and obtain the following conclusions: the proposed algorithm carries out node sensing and data aggregation within a certain area based on multiple dynamic routes. The calculation process does not require encryption and decryption, and is not affected by network topology, so it can better address the data aggregation problems in the dynamic change of network structure. Compared with other traditional data aggregation algorithms, the proposed algorithm has the advantages of low traffic, low energy consumption in data transmission, low probability of data leakage and high transmission accuracy. In data aggregation, 3 slices is the optimal quantity.

Keywords—wireless sensor network, data aggregation, data security, slice, node

1 Introduction

Wireless sensor network (WSN) is a new communication mode that has been increasingly popular in recent years. Its biggest feature is the use of wireless communication to carry out data transmission through a large number of sensing nodes, which greatly reduces the consumption of cable communication fiber. So far, it has been applied in transportation, environmental monitoring, national defense, electronic communications and other fields (Culler, 2004; Tan, et al, 2008; Parmar and Jinwala, 2016; Akyildiz, Sankarasubramaniam and Cayirci, 2002).

At present, the wireless sensor network technology is still far from mature. There are technical problems such as high energy consumption, loss of network lifetime and data leak in transmission, making data transmission and aggregation costly and easily leading to information leakage. Therefore, it is imperative to work out a new type of data transmission and aggregation algorithm with lower energy consumption, higher transmission precision and better privacy protection. So far, many researchers have

started relevant studies (Nasser and Chen, 2007; Ming and Vincent, 2007; Yahya and Benothman, 2009).

Traditional data privacy protection and data aggregation algorithms are mainly based on network topology (tree and cluster types). The sensor nodes of the above algorithm are subject to serious losses and susceptible to environmental impacts during service. The new research worked on data aggregation based on route algorithm (Wei and Yang, 2013; Lee, Kim and Chang, 2014), but few took data security into account (Conti et al, 2009). During the calculation process, frequent encryption and decryption operations increase the overall operation time (Zhou, Yang and He, 2014; Yang et al, 2008). In addition, in some special circumstances, hackers can use forged data, identity or Trojan attacks to shorten the lifetime of the wireless sensor data transmission system, increase energy consumption and even paralyze the entire network (Chan, 2007; Kumar and Dutta, 2015; Shakshuki Malik and Denko, 2008; Tang et al, 2011; Ozdemir and Xiao, 2011).

In this paper, considering the high energy consumption, loss of network lifetime and data leak in transmission during the data aggregation of the wireless sensor network, we propose an improved spatial data aggregation algorithm. Through comparison with the traditional data aggregation algorithms, we verify the feasibility and rationality of the proposed algorithm.

2 WSN spatial data aggregation algorithm

2.1 Relevant work and preliminaries

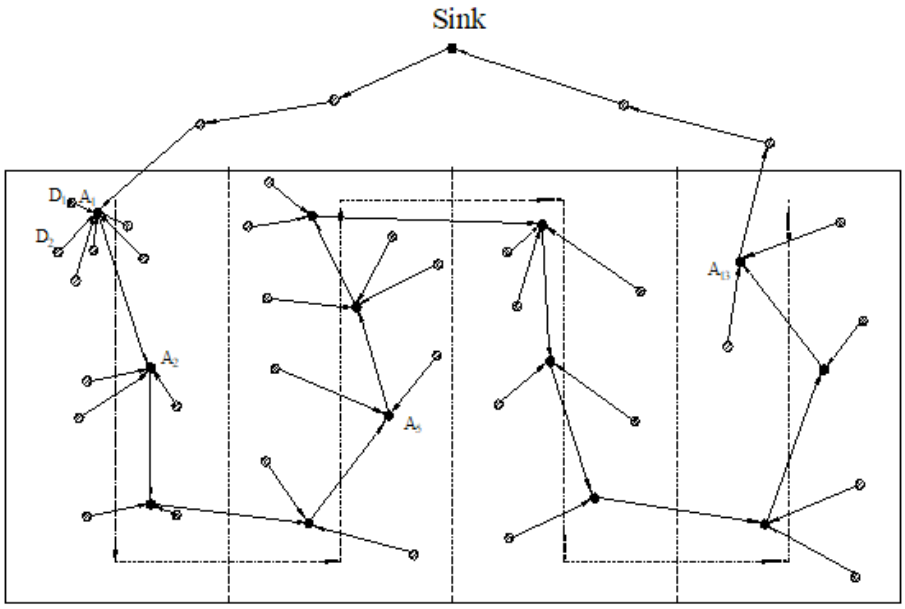


Fig. 1. Algorithm flow based on the itinerary route

The core purpose of data aggregation of the wireless sensor network is to simplify data, eliminate duplicated data, reduce redundant data and traffic to ultimately achieve energy saving and prolong network operation time. The traditional spatial data aggregation algorithms mainly adopt the tree-type network and node cluster topology structures, both of which are root nodes of Sink. They aggregate data through nodes and then transmit the encrypted or disturbed aggregation structures back to Sink. Their shortcoming is high energy consumption. In this paper, we make improvements based on the traditional data aggregation algorithms. Data aggregation mainly carries out node sensing within a certain area based on multiple dynamic routes. The algorithm process is shown in Figure 1.

The curved path in the figure represents the ideal route for spatial data aggregation. and the width of the two dotted lines is the width of the data aggregation route. The node sensing algorithm based on multiple dynamic routes has the following steps: first, it sends the system data aggregation request to the regional initial node (A_1) through Sink, and then aggregates the sensed data ((D_1 and D_2) in the node according to the route aggregation method, transmits them along the ideal route, determines the next aggregation node (A_2) on the path and sends another aggregation request from the system to A_2 to aggregate the sensed data around A_2. The subsequent data aggregation is carried out in the same way until all data in the area are aggregated. At last, it sends the aggregation results back to Sink and completes a single aggregation task in the area. According to existing studies, when the spacing between two routes is less than about 0.9 times the radio range, all the node sensed data in the area can be aggregated.

In the process of data transmission, the attack patterns that the wireless sensor network may encounter mainly include eavesdropping attacks and capturing of sensor nodes. Therefore, the main objectives of data privacy protection are to ensure the privacy of sensed data, reduce the transmission energy consumption and improve data aggregation accuracy. A good data aggregation algorithm can reduce the number of eavesdropping and collusion attacks against data, minimize the extra overheads and has great scalability.

The security of adjacent nodes in the transmission channel is guaranteed by the encryption algorithm. In this paper, we use the random secret key allocation algorithm. Supposing the probability of the two nodes having the same secret key is P_{com}, and that the probability that the third transmission node still has the same secret key is P_{ovhe}, the calculation method is as follows:

$$\begin{cases} P_{com} = 1 - \dfrac{\left((K-k)!\right)^2}{(K-k)!K!} \\ P_{ovhe} = k/K \end{cases} \tag{1}$$

2.2 Aggregation algorithm design and performance analysis

On the basis of the previous research, in this paper, we propose an improved spatial data aggregation algorithm (ISDAA). This algorithm is divided into five stages: initialization, route design, aggregation request, data aggregation and aggregation completion. Initialization means selecting k secret keys from the key collection to establish node channels; and route design involves setting the communication radius, communication routes and route width.

The data aggregation phase is the core part of the algorithm. The initial aggregation node receives the aggregation command, and through the transmission of aggregation nodes in the area, the last aggregation node in the area will also receive the aggregation command. Through the data slicing command, the sensed data are protected. Reorganization operation is carried out in the latter stage of data transmission. The distance between nodes is not fixed. If the node spacing is large, the link quality will be poor, but if the node spacing is small, the communication channels will conflict with each other. Therefore, the traffic should be reduced and the transmission process should be simplified. The detailed scheme is shown in Figure 2. The simplified algorithm is based on the ring virtual token. While receiving the aggregation task, the initial aggregation node also receives the reference line information. Through calculation, the algorithm obtains the transmission slicing information. Based on the angle between the aggregation node and the sensed data, it carries out the slicing operation and transmits the slices to the next node, and then reorganizes the data in the transmission process, which means it has to carry out two rounds of transmission.

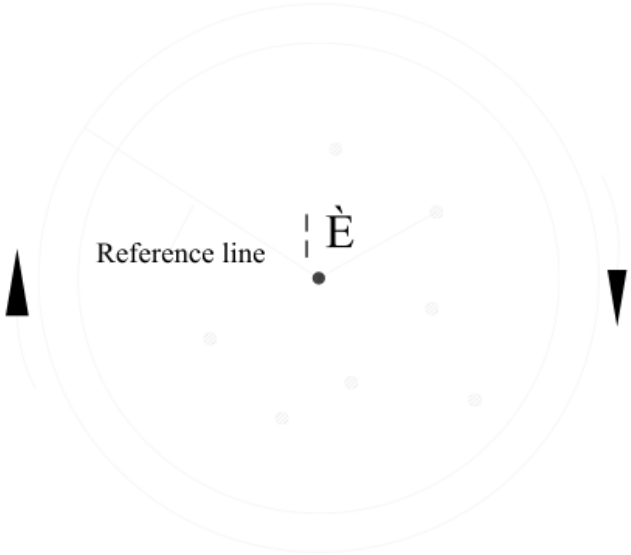


Fig. 2. Time assignation based on the itinerary route

Suppose the sensed data in the transmission process is divided into J slices, that the smaller the value of J is, the smaller the energy consumption and traffic will be, and that the value of J should meet the constraint conditions as shown in Formula (2):

$$\begin{cases} J_i \geq 1 \\ J_i + J_i^r \geq J_{\sec} \end{cases} \tag{2}$$

J_i and J_i^r are the number of the calculated sensed node and other sensed node data slices, and J_{sec} is the relevant parameter.

All the slices are reorganized and transmitted to the target aggregation node. The whole process can be expressed in Formula (3):

$$\begin{cases} \mathrm{Re}c \rightarrow A_i : v_i^d = v_i^n + d_i \\ Agg \rightarrow v_A^{agg} = \sum_{i=1}^{n} \left(v_i^d - d_i \right) + v_A \\ Tra \rightarrow v_A^d = v_A^{agg} + d_{A_i A_{i+1}} \end{cases} \tag{3}$$

where, Rec stands for receiving; Agg means aggregation; Tra means transmission; v represents slice.

We further analyze the security performance of the proposed algorithm. In the aggregation and slicing of the sensed data in the area, the probability of slice data leakage P_i^D and the probability of aggregation node data leakage P_i^A are respectively:

$$\begin{cases} P_i^D = P_r^{J_i - 1} \times P_r^{J_i^r + 1} = P_r^{J_i^r + J_i} \\ P_i^A = P_r^{n_i^D} \times P_r^{J_i^r + 1} = P_r^{J_i^r + n_i^D} \end{cases} \tag{4}$$

where, P_r is the probability of obtaining an arbitrary slice.

The traffic of data aggregation in the area can be calculated with Formula (5):

$$T_s = 0.5 \sum_{i=1}^{N} \left(\overline{e} + 1 \right) \cdot n_i^{SLN} \tag{5}$$

e is the length of the encrypted data; and n_i^{SLN} is the number of nodes. There should be at least one communication channel between adjacent nodes. Through complex calculation, the total traffic in the area can be converted as follows:

$$T_{all} = N \times B_A + \overline{d} \times \left(2 \sum_{i=1}^{D_{mm}} J_i + A_{mm} \right) \tag{6}$$

According to Formula (7), we can see that traffic is related to the proportion of sensing nodes and aggregation nodes in the area, and also has something to do with the number of slices divided from the sensing nodes. In order to enhance communica-

tion security, data aggregation is carried out for z times in each cycle. Then data aggregation traffic each time is

$$T_{ave} = \left(T_s + z \times T_{all}\right) / z \tag{7}$$

The energy consumption by data aggregation of wireless sensors is mainly concentrated in the communication and computation stages. The energy consumption of the sensing nodes E_D and that of the aggregating nodes E_A are respectively:

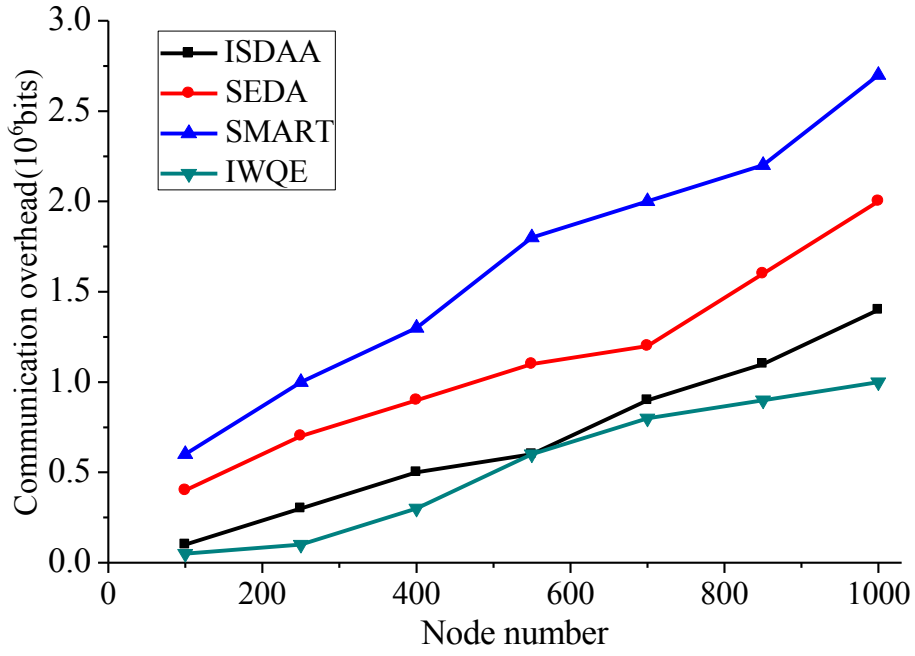
$$\begin{cases} E_D = \sum_{i}^{D_{mm}} \left[J_i \times \bar{d} \times e_T + \left(J_i^r \times \bar{d} + B_A \right) \times e_R \right] \\ E_A = \sum_{i=1}^{A_{mm}} \left[B_A \times e_T + \left(n_i^D + 1 + J_A^r \right) \times \bar{d} \times e_R \right] \end{cases} \tag{8}$$

3 Simulation test and verification analysis

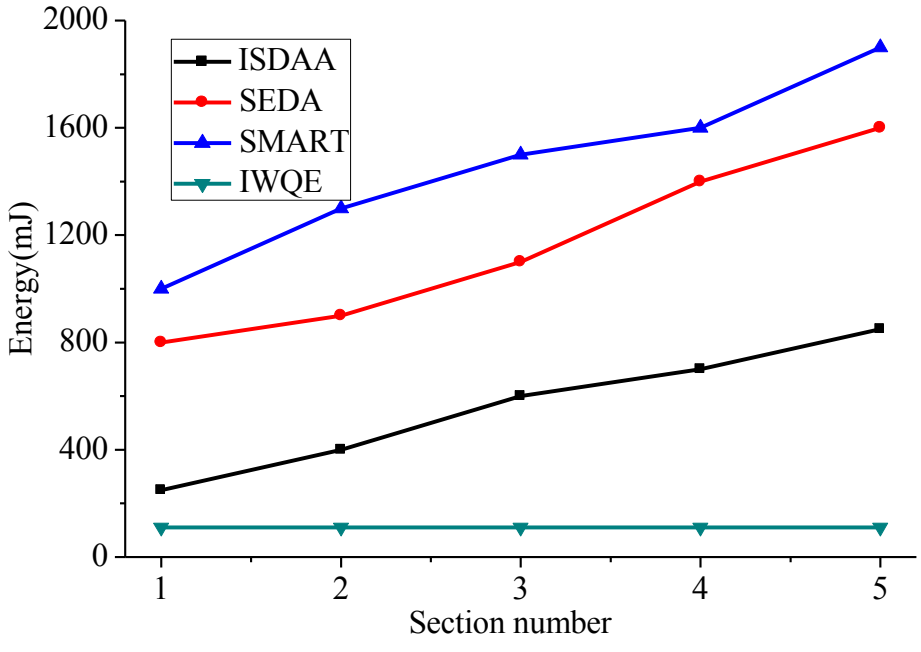
In order to verify the rationality and advantages of the WSN data aggregation algorithm proposed in this paper, we compare the proposed algorithm with the secure encrypted data aggregation (SEDA), itinerary based window query execution (IWQE) and pattern-code-based security data aggregation algorithm (ESPDA) in terms of communication quality, communication energy consumption, communication accuracy and privacy protection. The test environment for all these 4 algorithms is as follows: Core-i5; 16G memory; Win7 operating system; software environment Matlab; and experimental data set Inter Data. In order to improve the calculation accuracy, each calculation result is the average value after 30 rounds of calculation and the wireless sensor network topology generated in each round is random.

Figure 3 shows the traffic changes in four kinds of algorithms along with the number of nodes (Figure 4a) and number of slices (Figure 4b). From the figure, we can see that the traffic of SMART algorithm is much greater than those of the other three algorithms, because in the calculation process, the SMART algorithm must slice all sensed data before carrying out the transmission and encryption; the traffic of the IWQE algorithm is the smallest, but the algorithm does not consider the privacy of data transmission, which can easily lead to data leakage. The algorithm proposed in this paper has small traffic under different number of nodes and slices, and with the increase of N and J, the traffic of this algorithm decreases much faster than those of the other ones, proving the superiority of the algorithm.

Figure 4 shows how the energy consumptions of the four algorithms change along with the number of nodes (Figure 5a) and number of slices (Figure 5b). From the figure, it can be seen that, similar to traffic, the energy consumptions of the four algorithms, from high to low, are: the SMART algorithm, the SEDA algorithm, the improved algorithm proposed in this paper and the IWQE algorithm. With the increase of nodes, the consumption of the proposed algorithm is reduced faster than those of

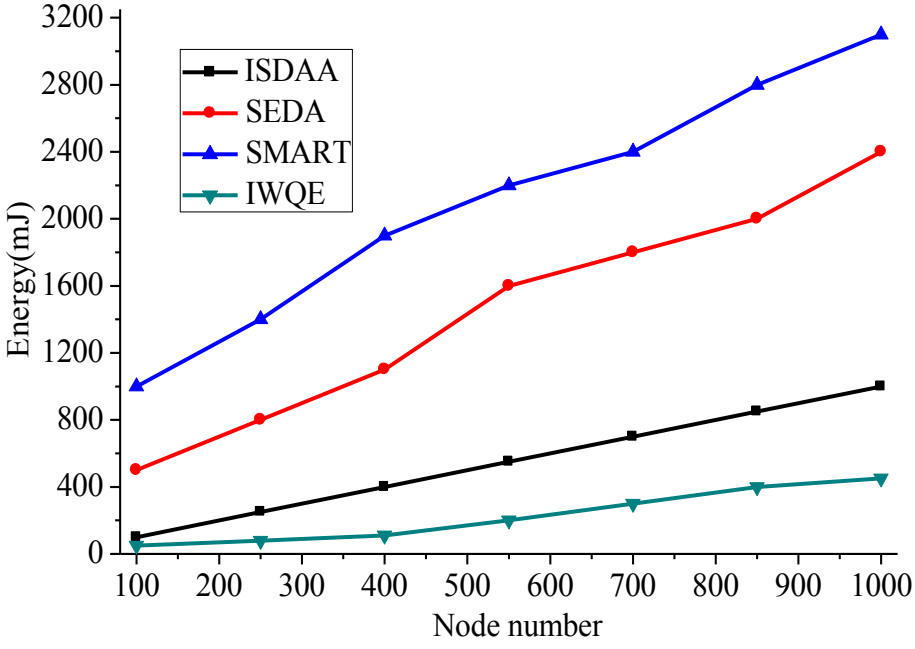


(a)

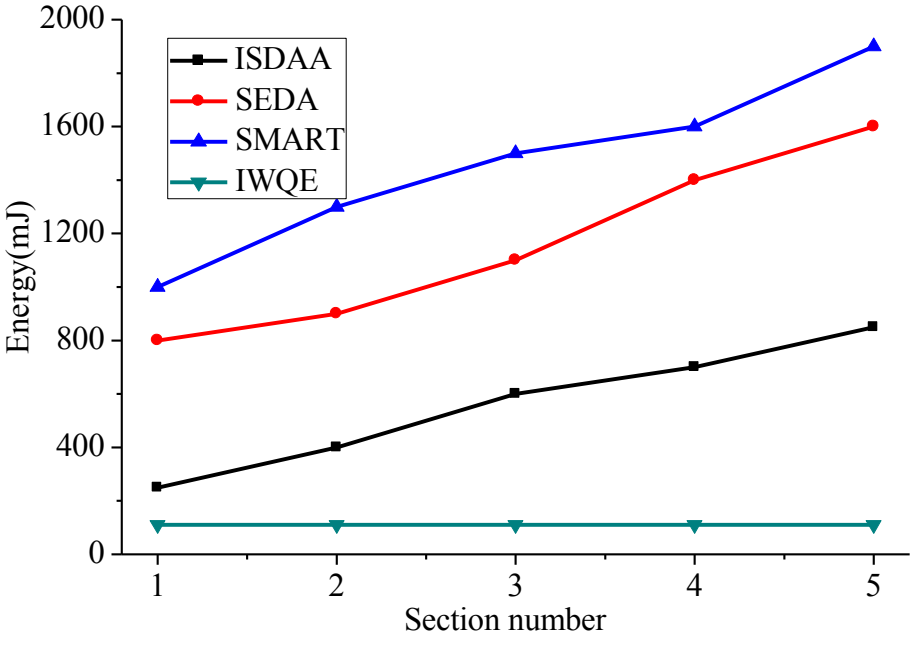


(b)

Fig. 3. Change of communication capacity with different data aggregation algorithm



(a)



(b)

Fig. 4. Change of energy consumption with different data aggregation algorithm

the SMART algorithm and the SEDA algorithm. When the number of nodes reaches 1000, the energy consumption of this algorithm is 32.2% and 41.67% of the SMART algorithm and the SEDA algorithm, respectively; and with the increase of slices, this algorithm saves more energy than the SMART algorithm and the SEDA algorithm, and the energy consumption of the IWQE algorithm is not affected by the number of slices.

Figure 5(a) shows the probabilities of the 4 aggregation algorithms having sensed data leakages in a certain area. It can be seen from the figure that when the probability of link level damage is small (<0.04), the SMART algorithm, the SEDA algorithm and the proposed algorithm all have small probabilities of data leakage, indicating that the three algorithms all have high security; when the probability of link level damage is >0.04, the data leakage probabilities of the SMART algorithm and the SEDA algorithm are significantly increased. The data leakage probability of the IWQE algorithm is maintained at a high level. This is because the IWQE algorithm does not consider the degree of access of the nodes, leading to varying privacy of different nodes. As a result, the average privacy of the sensed data is low. From Figure 5(a), we can see that the degree of access and the number of slices are important parameters to measure the security of the algorithm.

Figure 5(b) shows the sensed data leakage probabilities of the improved algorithm proposed in this paper under different numbers of slices. As can be seen from the figure, when the number of slices is small, the security of the sensed data is poor, and with the increase of node slices, the probability of data leakage is significantly reduced. When the number of slices reaches 4, the security of the data aggregation process can be guaranteed.

In the data aggregation process, when there is no data, energy or communication loss, the data aggregation accuracy can reach 100%. But in actual calculation, due to network interferences and data processing delays, data can be easily lost in the transmission process. The data aggregation accuracy is defined as follows:

$$P_c = \frac{AR}{\sum_{i=1}^{N} d_i} \tag{9}$$

AR represents the data aggregation result. Figure 6(a) shows the aggregation accuracy of the four aggregation algorithms over a given time. As can be seen from the figure, the longer the aggregation period is, the greater the aggregation accuracy will be. This is because a longer aggregation period can reduce the collision between the transmitted data and allows transmission of the data in the area to be completed within the period. As the proposed algorithm does not need the data decryption operation in the process of aggregation, greatly reducing the computation time, the aggregation accuracy of this algorithm can reach the maximum when the aggregation period is longer than 35s.

Figure 6(b) shows the accuracy of the proposed algorithm under different numbers of slices. As can be seen from the figure, the smaller the number of slices is, the higher the accuracy of the algorithm. This is because when there are many slices, the system needs to transmit more data packets, and accordingly there is a greater probability of data leakage and collision in the transmission process. From Figure 5 and 6, it can be seen that 3 slices is a proper quantity for the data aggregation of this algorithm.

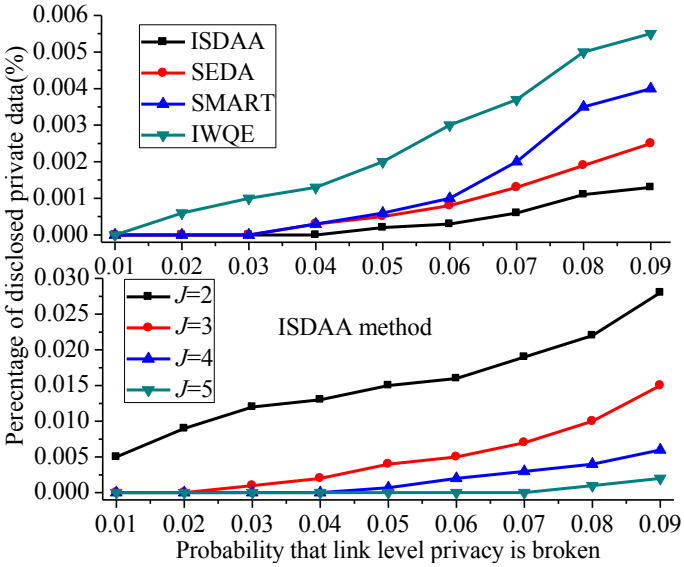


Fig. 5. The change of privacy comparisons with different algorithms and sections

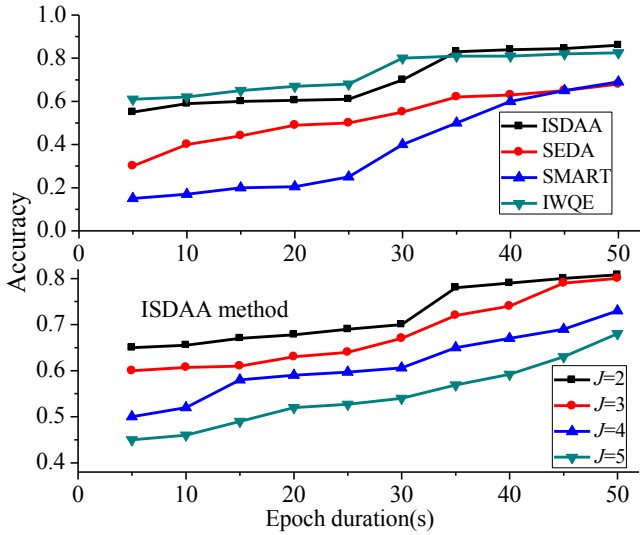


Fig. 6. The change of aggregation accuracy with different algorithms and sections

4 Conclusions

In this paper, considering the high energy consumption, loss of network lifetime and data leak in transmission during the data aggregation of the wireless sensor network, we propose an improved spatial data aggregation algorithm. Through comparison with traditional data aggregation algorithms, we verify the feasibility and rationality of the proposed algorithm and obtain the following conclusions:

1. The proposed algorithm carries out node sensing and data aggregation within a certain area based on multiple dynamic routes. The calculation process does not require encryption and decryption, and is not affected by network topology, so it can better address the data aggregation problems in the dynamic change of network structure.
2. Compared with other traditional data aggregation algorithms, the proposed algorithm has the advantages of low traffic, low energy consumption in data transmission, low probability of data leakage and high transmission accuracy. In data aggregation, 3 slices is the optimal quantity.

5 Acknowledgements

Supported by Natural Science Foundation of P. R. China (No.61503316, No.61562033), Fujian Province Science and Technology Plan of External Cooperation Project (No.2016I0015), Foundation of Fujian Educational Committee（No. JAT160358), The Fujian Provincial High School Natural Science Foundation of China（No. JZ160472), Xiamen Science and Technology Plan of University Innovation Project (No.3502Z 20153020).

6 References

[1] Culler, D. (2004). Overview of sensor networks. COMPUTER -LOS ALAMITOS, 37(8), 41-49. https://doi.org/10.5772/49376

[2] Yan, T., Gu, Y., He, T., & Stankovic, J. A. (2008). Design and optimization of distributed sensing coverage in wireless sensor networks. Acm Transactions on Embedded Computing Systems,7(3), 33. https://doi.org/10.1145/1347375.1347386

[3] Parmar, K., & Jinwala, D. C. (2016). Concealed data aggregation in wireless sensor networks: A comprehensive survey. Computer Networks,103(C), 207-227. https://doi.org/10.1016/j.comnet.2016.04.013

[4] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. Computer Networks the International Journal of Computer & Telecommunications Networking, 38(4), 393-422. https://doi.org/10.1016/S1389-1286(01)00302-4

[5] Nasser, N., & Chen, Y. (2007). Seem: secure and energy-efficient multipath routing protocol for wireless sensor networks. Computer Communications, 30(11–12), 2401-2412. https://doi.org/10.1016/j.comcom.2007.04.014

[6] Ming Lu, Y., & Vincent, W. S. W. (2007). An energy-efficient multipath routing protocol for wireless sensor networks: research articles. International Journal of Communication Systems, 20(7), 747-766. https://doi.org/10.1109/VTCF.2006.505
[7] Yahya, B., & Benothman, J. (2009). Reer: robust and energy efficient multipath routing protocol for wireless sensor networks. 1-7. https://doi.org/10.1109/GLOCOM.2009.5425587
[8] Wei, L. I., & Yang, G. (2013). Energy-saving data aggregation algorithm for protecting privacy and integrity. Journal of Computer Applications, 33(9), 2505-2510. https://doi.org/10.3724/SP.J.1087.2013.02505
[9] Lee, H., Kim, T. H., & Chang, J. W. (2014). An efficient data aggregation scheme for protecting the integrity of sensitive data in wireless sensor networks., 280, 207-214. https://doi.org/10.1007/978-3-642-41671-2_27
[10] Conti, M., Zhang, L., Roy, S., Pietro, R. D., Jajodia, S., & Mancini, L. V. (2009). Privacy-preserving robust data aggregation in wireless sensor networks. Security & Communication Networks, 2(2), 195–213. https://doi.org/10.1002/sec.95
[11] Zhou, Q., Yang, G., & He, L. (2014). A secure-enhanced data aggregation based on ecc in wireless sensor networks. Sensors, 14(4), 6701-6721. https://doi.org/10.3390/s140406701
[12] Madden, S., Franklin, M. J., Hellerstein, J. M., & Hong, W. (2002). Tag: a tiny aggregation service for ad-hoc sensor networks. Acm Sigops Operating Systems Review, 36(SI), 131-146. https://doi.org/10.1145/844128.844142
[13] Yang, Y., Wang, X., Zhu, S., & Cao, G. (2008). Sdap:a secure hop-by-hop data aggregation protocol for sensor networks. Acm Transactions on Information & System Security, 11(4), 1-43. https://doi.org/10.1145/1380564.1380568
[14] Chan, H., Perrig, A., Przydatek, B., & Song, D. (2007). Sia: secure information aggregation in sensor networks. Journal of Computer Security, 15(1), 69-102. https://doi.org/10.3233/JCS-2007-15104
[15] Kumar, M., & Dutta, K. (2015). A survey of security concerns in various data aggregation techniques in wireless sensor networks. 41(3), 191-201. https://doi.org/10.1007/978-81-322-2009-1_1
[16] Shakshuki, E., Malik, H., & Denko, M. K. (2008). Software agent-based directed diffusion in wireless sensor network. Telecommunication Systems, 38(3-4), 161-174. https://doi.org/10.1007/s11235-008-9102-4
[17] Tang, J., Dai, S., Li, J., & Li, S. (2011). Gossip−based scalable directed diffusion for wireless sensor networks. International Journal of Communication Systems, 24(11), 1418−1430. https://doi.org/10.1002/dac.1224
[18] Ozdemir, S., & Xiao, Y. (2011). Integrity protecting hierarchical concealed data aggregation for wireless sensor networks. Computer Networks, 55(8), 1735-1746. https://doi.org/10.1016/j.comnet.2011.01.006

7 Author

Liang Hu, Huayi Yin, and **Chaoqun Hong** are with the School of Computer and Information Engineering, Xiamen University of Technology, Xiamen, Fujian 361024, China (Huayi@xmut.edu.cn).

Article submitted 25 October 2017. Published as resubmitted by the authors 27 November 2017.