Journal of Network Computing and Applications (2017) 2: 1-6
Clausius Scientific Press, Canada

An Improved Data Fusion Method IICKPAD for Privacy Protection in Wireless Sensor Networks

ZHANG Shaomin[a, *], LI Ran[b], WANG Baoyi[c]

School of Control and Computer Engineering, North China Electric Power University, Baoding, 071003, China

a.email: zhangshaomin@126.com,
b.email:1612418752@qq.com, c.email:wangbaoyiqj@126.com

Keywords: Wireless sensor networks, data fusion, privacy protection, integrity

Abstract: Aiming at the privacy and security of data acquisition and monitoring in wireless sensor networks, an improved data fusion method, IICKPDA, is proposed for privacy protection in wireless sensor networks. The method is based on the ICKPDA algorithm, which optimizes the deadline between the same layer and the different layers，which can be randomly selected by the adjacent nodes of the data fragments to be transmitted, and also the non-leaf nodes which only the data is not collected to improve the characteristics of the data. The improved algorithm avoids the redundancy of the intermediate fusion process and protects the collected monitoring data, which improves the data fusion precision and the fusion efficiency of the intermediate fusion node. Theory and related experiments demonstrate the feasibility of the method.

1. Introduction

In general, nodes in wireless sensor networks [1]-[4] have the characteristics of computing power, storage, limited energy and low communication bandwidth, and it is difficult to transmit a large amount of sensor data. Data fusion operation can be different nodes of the data to merge and aggregate, and reduce network traffic and network energy consumption.

In the real world, the sensor nodes are placed in an untrustworthy real environment, the wireless channel is easy to be captured or eavesdropped, and the privacy data transmitted in the network may also be stolen or tampered with during the fusion process. Therefore, the wireless sensor network in the privacy protection, data integrity testing and other aspects of security there are some problems.

Domestic and foreign scholars have made some research on the data fusion method in wireless sensor networks. The TAG[5] (Tiny AGgregation) algorithm proposed in is a typical data fusion technique used in wireless sensor networks. But it does not provide privacy protection. In [6], a key and allocation scheme is proposed to encrypt the nodes and prevent the external nodes from stealing the internal sensitive data. At the same time, the encryption and decryption increases the integration cost. In [7] and [8], a data fusion privacy protection method based on homomorphic encryption is proposed, which is end-to-end encryption and cannot meet the security requirements

within the network. [9] proposed the SMART algorithm, which is fragmented by its own data, and then sent to the neighbors, the integration; but this way the energy consumption and accuracy is not high. In [10], the paper proposes a low energy consumption privacy data fusion algorithm for the problem of large energy consumption in [9]. In [11], the accuracy of the algorithm is improved by adding various optimization factors to the problem of low precision [9].

But for the literature [7] - [11] are based on privacy protection data fusion methods, they do not involve data integrity testing. [12] proposed the IPDA algorithm, which uses data redundancy to carry out integrity identification. In [13], a method of data integrity detection using complex numbers is proposed, but integrity testing is susceptible to damage. [14] proposed a new method of integrity-based data fusion based on complex numbers, but it only focuses on integrity testing and does not involve privacy protection. In [15], the ICKPDA algorithm is proposed to ensure that internal privacy data is not stolen by external attacks using the key distribution mechanism in [6]. It implements the data privacy protection of the internal nodes in the fragmentation technology before the data fusion, and realizes the integrity detection of the data by using the correlation between the binary data.

In order to improve the data fusion accuracy and the fusion efficiency of intermediate fusion nodes, this paper presents a new data fusion method for privacy protection in wireless sensor networks. IICKPDA (Improve-Integrity-CheceKing Private Data Aggregation).

In the first section, carries on the theoretical modeling to it, introduces the security encryption method; The second section improves the algorithm to form the new algorithm; The third section uses the simulation experiment result to carry on the new algorithm Assessment and analysis; Section 4 summarizes this article.

2. Theoretical Basis and Modeling of IICKPDA Method

Assuming that there are n devices in the wireless sensor network that need access, the sensor nodes of each device need to send packets to the central node without packet loss. Each device corresponds to a sensor node, the data transmitted by the sensor nodes in the network is denoted as $d_1(t)$, $d_2(t)$, $\cdots$, $d_n(t)$.Specific network model reference [15], as follows.

In this paper, the wireless sensor network model is represented by the connected graph $G(V,E)$, where $v(v \in V)$ is the network node, $e(e \in E)$ is the communication link, and $N = |V|$ is the number of nodes in the network. In this paper, the wireless sensor network has three types of nodes, such as the management node QS (Query Server) node, the intermediate fusion node (Aggregator) and the leaf node (Leaf).

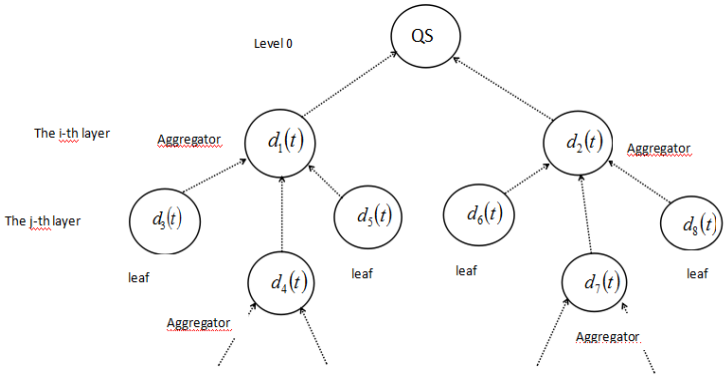

Figure 1 Schematic diagram of data fusion tree

The data fusion function of this paper is taken as an example of the sum function.

3. Data Structure and Algorithm Description of IICKPDA Method

This section details the improvement and optimization of the ICKPDA algorithm.

3.1 Improvement of Non-leaf Node Data

In the third stage of [15], the function of collecting data and transmitting data is first for all leaf nodes and fusion nodes. The difference is that the fusion node also adds a fusion data to the leaf node Features. This allows the QS node to receive more complete data and more accurate queries. Specific improvements are as follows:

1) preparation phase and 2) sensing data conversion phase see [15];

3) Data segmentation stage: the node in the 2) step after the calculation of the data is recorded as $D_i = \langle d_{i1}, d_{i2} \rangle$, the leaf node in the slice delay time slice randomly selected S_i in the adjacent node j to send the fragment, sent the fragmentation of encrypted data to the node j .Node i subtracts P_{ij} from its own data. The neighbor node j accepts the slice P_{ij}, decrypts the fragment and fuses the slice. Cycle step 1) -3) until all the required node data is acquired.

In order to reduce the effect of data loss on the precision of fusion, the data slice is randomly generated by node i within the bounded data range, see equation (1).

After splitting the data $f_i = D_i - P_{ij}$ in the node i, the node j decrypts the received encrypted data piece and adds it to its own data, the node data is $f_j = D_j + P_{ij}$.

4) data fusion phase and 5) integrity testing phase see [15].

The IICKPDA method utilizes the function of transmitting data in the literature [15] with the non-leaf node only receiving data in the data fragmentation stage and the non-leaf nodes and leaf nodes. This ensures that the collection phase as much as possible to avoid missing data collection, making the final fusion results more accurate.

3.2 Optimization of The Infinite Loop Between The Same Layer May Occur

The dead cycle between the same layers that may exist in [15] is shown in Figure 2. If node B randomly selects neighbor node C in the node set to send data fragments, which is $B \xrightarrow{P_{bc}} C$, the same that is $C \xrightarrow{P_{ce}} E$, $E \xrightarrow{P_{ef}} F$, $F \xrightarrow{P_{fe}} E$, $E \xrightarrow{P_{ec}} C$, $C \xrightarrow{P_{cb}} B$.And node B selects node C again, and the data fragment sent at this time is denoted by P'_{bc} ($P'_{bc} \neq P_{bc}$), so it is circulated. This is the case in the literature [15].

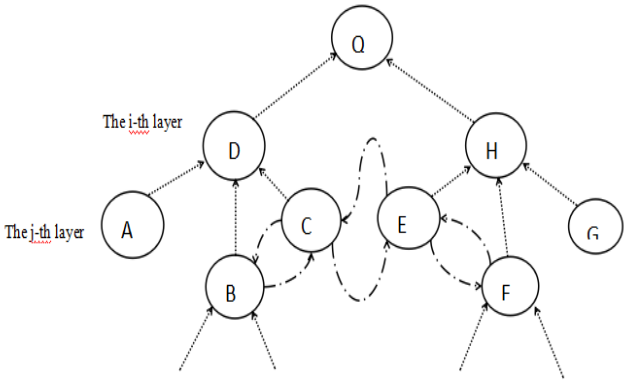

Figure 2 May appear in the same layer between the dead cycles of the schematic diagram

To solve the possibility of the same layer between the dead loops, you need to add a condition.

When the node chooses the neighbor node that sends the data fragment in the node set S, it can not select the neighbor node to which the data fragment is transmitted, so that it can not be unreachable. That is, assuming that node i and node j are at the same layer, and node i is selected at node i to send data fragments. Before node j selects adjacent node to send the slice, add a restriction condition, that is, select other neighbor nodes other than i (A parent node or other sibling node other than i) to satisfy this condition before allowing this node to send data fragments.

3.3 Optimization of The Infinite Loop Between The Different Layers May Occur

The dead cycle between the same layers that may exist in [15] is shown in Figure 3 . If node B randomly selects neighbor node C in the node set to send data fragments,which is $B \xrightarrow{P_{bc}} C$, the same that is $C \xrightarrow{P_{cd}} D$, $D \xrightarrow{P_{da}} A$, $A \xrightarrow{P_{ab}} B$ .And node B selects node C again, and the data fragment sent at this time is denoted by $P_{bc}^{'}$ ($P_{bc}^{'} \neq P_{bc}$), so it is circulated. This is the case in the literature [15].

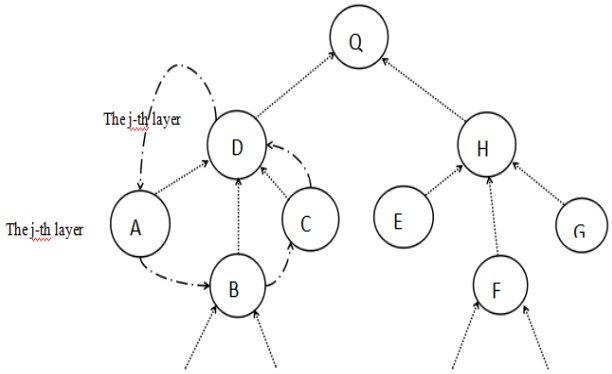

Figure 3 May appear between the different layers of the cycle of the cycle diagram

When the node chooses the neighbor node to send the data fragment in the node set, the node needs to satisfy the node with the selected neighbor node in the same layer (adjacent sibling node) or node is lower than the adjacent node to be selected. When node i selects the neighbor node j that sends the data fragment in the node set, the node i needs to satisfy $\deg_i \leq \deg_j$ (where $\deg_i$ and $\deg_j$ respectively represent the number of nodes i and node j in the data fusion tree).

4. IICKPDA Method Performance Analysis

In this paper, Tiny OS system under the TOSSIM simulation software for simulation experiments. The network configuration environment is: 600 nodes are randomly distributed in the area of 400 m×400 m . The standard is indoor environment. Wireless channel is the symmetry. Gaussian white noise is 4dB.The background noise is -105.0 dBm, and node transmission distance is 50 m. The data transfer rate is 1Mbps, and sensitivity is -108.0 dBm.

4.1 Accuracy Experiment

In this paper, the accuracy of data fusion is measured by the ratio of the fusion results of the simulation experiments to the sum of the actual perceived data of the nodes.

The accuracy of this algorithm is compared with that of IICKPDA and ICKPDA, ESPART [10], as shown in Figure 4, with different delay (node time delay for data fragmentation and fusion operation, indicated by ED).

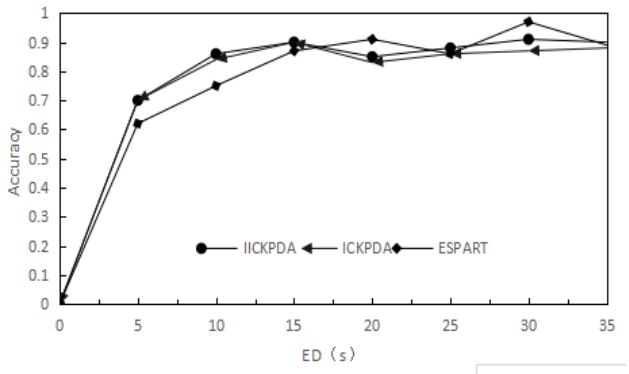

Figure 4 IICKPDA algorithm accuracy

The analysis shows that the ICKPDA algorithm reduces the accuracy of the proposed algorithm by reducing the range of the sending nodes, reducing the number of slices, limiting the slice size of the data, and improving the accuracy of the proposed algorithm. The accuracy of the improved algorithm is better than that of the other two algorithms degree. The IICKPDA method is improved on the basis of the ICKPDA algorithm, and its node data is more accurate, by increasing the sum of the root nodes after fusion, making it as close as possible to the original data, and thus improve the accuracy. The results show that the accuracy of the IICKPDA method is slightly higher than that of the ICKPDA algorithm.

4.2 Privacy Protection Analysis

The IICKPDA privacy protection is basically equal to ICKPDA privacy protection, the use of random encryption mechanism [6] for encryption. The private seed s is a global variable that is shared by each node in a particular query task and is kept secret to the external node. In order to increase the security, the query information whose value is published with QS is changed randomly in each query.

Let q be represented the probability of the private factor and the failure of the wireless channel, $d_{\max}$ is represented the maximum degree of the node, and $P(\deg = k)$ is represented the probability that the node degree is k.

IICKPDA method data is eavesdropped as follows:

1) For external attacks, the probability of data being tapped See (3):

$$P(q) = q^2 \sum_{k=1}^{d_{\max}} P(\deg = k) \cdot q^k \quad (3)$$

2) For the internal node, the probability of data being tapped See (4):

$$P(q) = q \cdot \sum_{k=1}^{d_{\max}} P(\deg = k) \cdot q^k \quad (4)$$

Because the TAG [5] algorithm has no privacy protection mechanism, the IICKPDA method and the ICKPDA algorithm have higher privacy protection than the TAG algorithm.

4.3 Data Integrity Analysis

The data integrity of the improved algorithm in this paper is consistent with that of ICKPDA in [15]. Data integrity testing is achieved using the correlation between binary data. The QS node gets the dummy data as $SM_D = \langle SM_{d_1}, SM_{d_2} \rangle$, and records the number of nodes participating in the fusion, and calculates $SM_D = \langle SM_{d_1}, SM_{d_2} \rangle$ as the sum of the private seeds of the fusion node, and obtains the true fusion data, as $SM_1 = SM_{d_1} / m - SM_s$ and $SM_2 = (SM_{d_2} - SM_s) / m$. In the practical application of

the error, within the allowable range of error, the sum of the two should be equal, so as to verify the data integrity.

5. Conclusion

In this paper, IICKPDA is based on the ICKPDA algorithm to improve, first ICKPDA algorithm in the leaf node to collect data, the data fragmented, and select the neighbor node to transmit data. Non-leaf nodes receive data only. Changing it to leaf nodes and non-leaf nodes has the ability to collect data and transfer data to it. This ensures that the collection phase as much as possible to avoid missing data collection, making the final fusion results more accurate. Secondly, the ICKPDA algorithm is optimized to improve its dead loop problem, which avoids the redundancy of the intermediate fusion process and improves the fusion efficiency of the intermediate fusion nodes.

References

[1] Culler D, Estrin D, Srivastava M. Guest editors' introduction: overview of sensor networks[J]. IEEE Computer 2004; 37(8): 41-49.

[2] Zhang L, Zhang H q Mauro Conti, et al. Preserving privacy against external and internal threats in WSN data aggregation}J}. Telecommunication Systems, 2013, 52(4): 2163-2176.

[3] Hu Wei, Zhao Wenhui. Data Fusion Approach for the Energy Internet Based on Time Window and the Adaptive Weights[J]. Journal of Systems Management, 2016,05: 907-913.

[4] Ozdemir S and Xiao Y. Secure data aggregation in wireless sensor networks: a comprehensive overview[J]. Computer Networks, 2009, 53(12): 2022-2037.

[5] Madden S, Franklin M J, and Hellerstein J M. TAG: a tiny aggregation service for Ad-hoc sensor networks[C].Proceedings of the 5th Symposium on Operating SystemsDesign and Implementation, Boston, 2002: 131-146.

[6] Eschenauer L and Gligor V D. A key-management scheme for distributed sensor networks[C]. Proceedings of the 9th ACM Conference on Computer and Communications Security, Washington, 2002: 41-47.

[7] Girao J, Westhoff D, and Schneider M. CDA: concealed data aggregation for reverse multicast traffic in wireless sensor networks[C]. Proceedings of the 40th IEEE International Conference on Communications, Seoul, 2005: 3044-3049.

[8] Westhoff D, Girao J, and Acharya M. Concealed data aggregation for reverse multicast traffic in sensor networks: encryption, key distribution, and routing adaptation[J].IEEE Transactions on Mobile Computing, 2006, 5(10):1417-1431.

[9] He W, Liu X, Nguyen H, et al.. PDA: privacy-preserving data aggregation in wireless sensor networks[C]. Proceedings of the 26th IEEE International Conference on Computer Communications(INFOCOM), Anchorage, 2007: 2045-2053.

[10] YANG Geng, WANG An-qi, CHEN Zheng-yu, et al. A Energy-Saving Privacy-Protection Data Algorithm[J]. Journal of Computers, 2011, 34 (5): 792-800.

[11] YANG Geng, LI Sen, CHEN Zhengyu, XU Jian, YANG Zhen.High. High-Accuracy and Privacy-Preserving Oriented Data Fusion Algorithm in Sensor Networks [J]. Journal of Computers, 2013,01: 189-200.

[12] He W,Nguyen H,Liu X, et al. iPDA: an integrity- protecting private data aggregation scheme for wireless sensor networks [C] //Military Communications Conference. San Diego, CA:[s. N.]，2008: 1-7.

[13] Bista R, Yoo H K, and Chang J W. A new sensitive data aggregation scheme for protecting integrity in wireless sensor networks[C]. Proceedings of the 10th IEEE International Conference on Computer and Information Technology,Bradford, 2010: 2463-2470.

[14] ZHAO Dan, YANG Geng. A Complex Field - based Integrity - protecting Data Aggregation Algorithm[J]. Journal of Computer Technology and Development, 2012, (08): 150-154 + 158.

[15] Zhou Qiang, Yang Geng, Li Sen, Chen Lei. An Integrity-checking Privacy Data Aggregation Algorithm[J]. Journal of Electronics & Information Technology, 2013, (06): 1277-1283.