Analysis of Wireless sensor network Multi-sensor Data Fusion Security Method

Weijun Lei

School of Information Engineering, Xi'an University, Xi'an, Shaanxi, China

Keywords: WSN; Data fusion; Security analysis; Homomorphism; Digital signature

Abstract: Wireless sensor network (WSN) is an important foundation for the Internet of Things. With the rapid development of information technology, WSN is more widely used. Due to the increasing number of sensors and the number of applications in WSN, the data sensed by multiple sensors is fused and processed. The requirements are constantly increasing, and at the same time, higher requirements are raised for security issues in data fusion. This paper analyzes the security method of WSN data fusion, which is helpful for further research on WSN data fusion security technology.

1. Introduction

With the rapid development of information technology, the Internet of Things is being widely used. Wireless sensor network (WSN) is an important foundation of the Internet of Things technology. Due to the increasing number and types of sensors in WSN, the difficulty and requirements for fusion processing of multi-sensor data are also increasing, especially for security issues in data fusion put forward higher requirements. In recent years, researchers from various countries have proposed many different methods of secure data fusion based on the existence of WSNs and different forms of external security threats and attacks, and according to the data fusion security goals. For example, the privacy protection of data fusion requires that the data collected by the nodes can reach the aggregation node safely, and important information transmitted by the nodes cannot be leaked to neighboring nodes, and the data after fusion must also be kept secret. In response to this situation, the traditional method is to encrypt the transmitted plaintext data in ciphertext form using a password, and only those who know the decryption key can identify the ciphertext and restore the ciphertext to plaintext. After the ciphertext data transmitted by the child nodes collected by the sink node is decrypted and fused, the fused value is encrypted and then transmitted upward. Once the attacker intercepts the encryption key, it is easy to obtain the real data information transmitted. In addition, in the process of fusion, it is necessary to directly expose the plaintext data to the sink node. If the sink node is captured, the transmitted sensitive data will also be leaked. It is a great safety hazard. At the same time, this method requires frequent encryption and decryption of data, which easily causes network overhead.

In recent years, researchers have proposed many WSN data fusion security protection methods to ensure the safety and accuracy of data fusion results.

2. Analysis of WSN Data Fusion Security Methods

2.1 Data security fusion method based on homomorphic encryption mechanism

The homomorphic encryption mechanism is an end-to-end encryption method that can be directly processed on the ciphertext. WSN intermediate sensor nodes do not need to encrypt and decrypt data. This method can achieve various summations and products. Fusion operation, to ensure the confidentiality of data. Because this method operates directly on the ciphertext, it effectively reduces the cost of calculation, extends the network lifetime from the surface, and guarantees end-to-end data security. At the same time, it may also be possible to perform distributed computing on untrusted nodes without exposing private information.

The basis of homomorphic encryption is the form of private homomorphism [1]. Private homomorphism is a form of encryption exchange that allows direct manipulation of ciphertext. The

  DOI: 10.25236/icbdem.2020.075

idea of this method is based on algebraic operations.

Assume that the encryption and decryption functions are E_{K1}, D_{K1}, The transmitted plaintext data set is $\{M_1, M_2, \cdots M_n\}$, α and β represent operations, if:

$a(E_{k1}(M_1), E_{k1}(M_2), \cdots E_{k1}(M_n)) = E_{k1}(\beta(M_1, M_2, \cdots, M_n))$ and $D_{k2}(a(E_{k1}(M_1), E_{k1}(M_2), \cdots E_{k1}(M_n)) = \beta(M_1, M_2, \cdots, M_n))$, This family of functions is called a secret homomorphism.

J. Girao and others proposed the CDA algorithm [2]. The advantages of the CDA algorithm are:

(1) Random segmentation of node data, so that the same data is encrypted with the same key, and the ciphertext obtained will not be the same, so the privacy of data fusion will be better guaranteed.

(2) The algorithm can also realize end-to-end secure transmission of data between nodes and base stations.

(3) This method saves the aggregation node from the encryption and decryption process, and uses simple modular addition to achieve data fusion, which can save a lot of overhead.

The disadvantages of this method are as follows:

(4) Since the data of each node is divided into multiple parts for transmission, the transmission overhead is also increased;

(5) Due to the use of a symmetric key system, the algorithm has poor security scalability.

Bahi et al. Proposed a secure data fusion algorithm based on elliptic curve end-to-end encryption [3]. This method allows complex operations on the ciphertext. It uses a smaller key field to distinguish ciphertexts between the same text. The direct operation of the ciphertext makes its security higher than that of the existing cryptosystem of WSN. And greatly reduce the transmission and computing overhead of the system.

In summary, the homomorphic encryption algorithm not only protects the confidentiality of the data, but also relieves the communication pressure of the sink nodes. However, it has disadvantages:

(1) The suitable fusion operation is relatively simple. At present, it is only suitable for the average operation and variance in addition fusion, and it is not suitable for the fusion operation of finding the maximum value and the median value.

(2) Cannot effectively verify the integrity of the data fusion results.

(3) The transmission of the node ID increases the transmission overhead of the network.

2.2 Monitoring and reputation mechanism security data fusion method

The secure data fusion algorithm WDA [4] proposed by Du et al. Uses peer-to-peer mutual supervision to achieve integrity verification. This algorithm divides the network into clusters, and sets supervising nodes around the cluster head. Each node in the cluster should send its own node information to its cluster head together with its corresponding supervising node. Next, the cluster head and the supervising node respectively fuse the received data. The supervising node then uses the key shared with the base station to generate a MAC value and sends the MAC to the cluster head. The cluster head sends the MAC value to the base station together with the fusion result. The base station only needs to compare the MAC value of the supervising node with the MAC value calculated by the fusion result to determine whether the fusion result is correct. This method has good robustness, but the premise is that the cluster head node and all supervised nodes cannot be captured, and this method needs to occupy multiple nodes, so that the probability of data leakage is increased, and it cannot be completely guarantee the privacy of the transmitted data.

In 2008, Ozdemir improved the SELDA algorithm, and proposed the RDAT [5] algorithm. Each node in the algorithm calculates its own functional reputation value, and uses the functional reputation value of each node as each task. Indicators for evaluation. The algorithm uses the trusted data fusion value selected by the aggregation function reputation to ensure the security of data fusion. In addition, the algorithm uses a multipath transmission method based on the reputation of routing functions for data transmission, which reduces communication overhead.

Vu et al. Also proposed THIS [6], a security fusion algorithm based on trust supervision mechanism. This algorithm uses a child node to supervise the parent node to ensure the accuracy and reliability of the final data fusion result. This algorithm reduces the dependence on the network

topology, but because its supervision mechanism is only the child node's estimation of the data fusion result sent by the cluster head, it also reduces the accuracy of the data fusion.

In 2011, Bohli et al. proposed a more flexible data fusion scheme [7], which uses the quality of fusion information (QOI) as a defense mechanism, which can better protect the integrity of the data and can resist large Strong range attack. In this method, it is considered that it is necessary for every WSN user to measure the result of data fusion with the QOI value. It can evaluate the quality of the data and detect errors or attacks. Even if some important parts of WSN nodes are controlled by the attacker, this method can identify and mitigate the impact of the attack.

2.3 Hidden real datasecurity data fusion method [8]

In 2003, Cam et al. proposed an energy-efficient pattern code-based secure data fusion protocol ESPDA. This method utilizes pattern code technology to reduce redundant data information transmitted by nodes and achieve data fusion security. The fusion process is not a fusion of real data, but a fusion of pattern codes, which hides the real data information. The node first transmits the mode code to the cluster head to remove the redundant node information. The node information after removing the redundant information uses the homomorphic encryption mechanism to transmit data to the base station. The cluster-head node in this method does not perform encryption and decryption operations, which guarantees the privacy of transmitted data to a certain extent.

The privacy protection method PDA proposed by He et al. In 2007 adopted scrambling and data segmentation technology to protect the confidentiality of private data. The method includes two forms, one is a privacy protection data fusion method CPDA based on the scrambling technology, and the other is a privacy protection data fusion method SMART based on the segmentation and recombination technology. The CPDA method first mixes the disturbed data with the perceptual data of the nodes to hide the real data, and then uses the algebraic nature of the polynomial to obtain the fusion result. Let the total number of nodes in the cluster be n, s_0 is the cluster head node, $s_1 \ldots s_n$ for other nodes in the cluster, at the same time, $s_i (0 \le i \le n)$ Generate n private random numbers $r_1^i, \ldots, r_n^i$ and calculate the disturbance data $V_j^i = v_i + r_1^i x_j + r_2^i (x_j)^2 + \ldots r_n^i (x_j)^n$, among them $0 \le j \le n$. Then, the final value of s_j is s_i and use the shared key with s_j encrypted value of shared key k_{ij} to V_j^i . After the nodes in the cluster exchange data one by one, node $s_j (0 \le j \le n)$ decrypt the perturbation data it receives before summing it to get:

$F_j = \sum_{i=1}^{n} V_j^i = v_{sum} + r_1 x_j + r_2 (x_j)^2 + \ldots + r_n (x_j)^n (v_{sum} = \sum_{i=1}^{n} v_i, r_k = \sum_{i=1}^{n} r_k^i (0 \le k \le n))$, s_j then broadcasts the combined data F_j to the cluster head s_0 . At this time, the cluster head node s_0 obtains a system of line equations consisting of equation F_j (0≤ j≤ n) with G as the coefficient matrix, among them $G = \{1, X, \ldots, X^n\} (X = \{x_0, x_1, \ldots, x_n\}^T)$, because data items in $X = \{x_0, x_1, \ldots, x_n\}^T$ are different from each other, so G is a full rank matrix. s_0 can find the trace $U = \{v_{sum}, r_1, \ldots, r_n\}^T$ through $U = G^{-1} F$, where $F = \{F_0, F_1, \ldots, F_n\}^T$. In this way, the cluster head node s_0 can obtain the accurate summation aggregation result v_{sum} without knowing the true data values of other nodes.

(1) When the size of the cluster is large, in order to obtain the final data fusion result, the cluster head needs to solve an inverse matrix of order m, which requires a large computational overhead.

(2) Every two nodes in the cluster need to exchange information. When the size of the cluster is large, the communication overhead also becomes very large, so it is not suitable for large-scale network structures.

In 2009, Li improved the SMART method and proposed CACR. The improvement of the method is mainly reflected in the following two points:

(1) Data division using the Chinese remainder theorem. This method does not rely on the value of the security threshold, and the security overhead is smaller than SMART's addition division.

(2) The TAG algorithm is used to establish multiple topological structures, and the big difference is the data fusion tree, and the vertices of the tree maintain a certain distance and are different.

Compared with the SMART method, CACR not only effectively reduces data collisions and collisions in the network, but also reduces the risk of attackers capturing nodes around the base station and stealing data fusion results. CACR performs better than SMART in terms of total network transmission overhead.

In 2011, H.Li and Yang Geng proposed the EEHA algorithm and the ESPART algorithm, respectively. Both of these algorithms are improvements to the SMART algorithm in terms of reducing data traffic and improving the accuracy of the fusion result.

The EEHA algorithm first uses the TAG algorithm to build a data fusion tree, and then divides the data of the leaf nodes in the fusion tree into multiple small pieces for transmission. Because the number of leaf nodes in the fusion tree is much smaller than the number of intermediate nodes, compared with SMART algorithm, this algorithm greatly reduces the data communication volume, reduces the network energy consumption, and also improves the accuracy of the fusion result. .

In the same year, the KIPDA algorithm proposed by Groat et al. Is a typical non-encrypted method for protecting data privacy. It adds masquerading data without encrypting node information, so that the privacy of the data is protected to a certain degree, and privacy protection is achieved. The maximum / minimum non-linear fusion and SUM fusion for privacy protection can be extended. Because this method does not require key distribution and encryption and decryption operations, it also saves node computing and communication overhead to a certain extent, and saves energy consumption of the entire network. However, they cannot distinguish between redundant data, and add disguised data to disturb the original data, which increases the computational overhead to a certain extent. The privacy protection capability of this method is relatively weak.

2.4 Digital signature secure data fusion method

In 2004, Mahimkar et al. Proposed a digital signature-based integrity data fusion algorithm Secure DAV suitable for clustered WSN [9]. In this algorithm, the nodes in the same cluster share the cluster key, and Use the elliptic curve algorithm to partially sign the real data collected by the nodes, and then pass these data to the cluster head. The cluster head then averages the received data and broadcasts the obtained average to other nodes in the cluster. The node needs to compare the data collected by itself with the fusion data of the cluster head node to form a difference. When the difference between the two is less than a fixed threshold, the signature of this node must be modified. Finally, the cluster head integrates each part of the signature, integrates it into a complete signature, and passes it to the base station. The base station can determine whether the fusion result is correct and valid by verifying the signature. This algorithm can achieve the unification of the three goals of confidentiality, integrity, and authentication. However, the process of data signing and verification consumes excessive computing and communication overhead of the system, and the fusion operation suitable for this method is relatively single. Therefore, only data fusion operations for averaging are supported.

In 2006, Y. Yang and others also proposed a digital signature-based secure fusion algorithm SDAP [10]. The algorithm is based on a tree structure. In the fusion tree, high-level nodes have a higher trust level than lower-level nodes. . And SDAP algorithm also uses the principle of processing separately to reduce the trust of high-level nodes. That is, the topology tree is first randomly divided into several logical subtrees of the same size. Each logical subtree has fewer low-level nodes, and the data collected is correspondingly less. In this case, when the high-level nodes are when captured, there will be fewer security threats. In this algorithm, the fusion data submitted by the fusion points in the logical subtree to the base station must be authenticated by digital signatures. After receiving the data, the base station must verify these signatures to achieve data confidentiality, integrity, and originality. Authentication service.

In 2011, Li et al. proposed an efficient and reliable method for secure data fusion based on identity authentication [8]. This solution uses the methods of signature fusion, batch authentication, and signature amortization to reduce communication overhead, reduce digital signatures and verification operations, and has a better fault tolerance rate.

3. Conclusion

With the rapid development of information technology, the Internet of Things, artificial intelligence, big data, cloud computing and other technologies have become more and more widely used. If the Internet realizes the interconnection between people, then the Internet of Things will achieve the connection between things. WSN, as an important foundation for realizing IoT, must also continue to improve its functions. Multi-sensor data fusion, especially the security of multi-sensor data fusion, has attracted more and more people's attention. There are some studies on accuracy, and some methods and algorithms have been developed, but these methods still cannot keep up with the requirements of the rapid development of WSN for data fusion security. Therefore, through the analysis and understanding of existing WSN data fusion security methods, continue to deepen and improve the research on WSN multi-sensor data fusion security algorithms, methods and schemes, and promote the security application of IoT on the basis.

Acknowledgements

This research was supported by the Xi'an Science and Technology Plan Innovation Fund Project Xi'an University Special Project (No: 2019KJWL26) ,Xi'an Science and Technology Plan Project (No:GXYD16.1),National University Student Innovation Training Project (No: 201911080025) and Shaanxi Key Laboratory of Surface Engineering and Remanufacturing.

References

[1] Rivest R., Adleman L., Dertouzos M. (1978) On Data Banks and Privacy Homomorphism. Foundations of Secure Computation. New York, Academic Press, 169-179.

[2] Westhoff D., Girao J., Acharya M. (2006) Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks, Encryption Key Distribution and Routing Adaptation. IEEE Transaction on Mobile Computing, 5(10): 1417-1431.

[3] Bahi J., Guyeux C., Makhoul A. (2010) Secure Data Aggregation in Wireless Sensor Networks, Homomorphism Versus Watermarking Approach.//Proc. of Conference on Ad Hoc Networks. Canada, ADHOCNETS Press, 344-358.

[4] Wu D., J Deng., Y.S. Han. (2003) A Witness-Based Approach for Data Fusion Assurance Wireless Sensor Networks[C]//Proc. of IEEE Global Telecommunication Conference. Washington, IEEE Computer Society Press, 1435-1439.

[5] Ozdemir S. (2008) Functional Reputation Based Reliable Data Aggregation and Transmission for Wireless Sensor Networks. Computer Communications, 3941-3953.

[6] Vu H., Mittal N., Venkatesan S. (2007) THIS, Threshold Security for Information Aggregation in Sensor Networks.//Proc. of the 4th International Conference on Information Technology. Washington, IEEE Computer Society Press, 89-95.

[7] Bohli J.-M., Verardi D., Papadimitrators P. (2011) Resilient Data Aggregation for Unattended WSNS.//Proc. of the 36th IEEE Conference on Local Computer Networks, 994-1002.

[8]Zhang Y. (2013) Research on Secure Data Aggregation Scheme in Wireless Sensor Networks. Master Thesis, Nanjing University of Posts and Telecommunications.

[9] Mahimkar A., Rappaport T. S. (2004) Secure DAV, a Secure Data Aggregation and Verification Protocol for Wireless Sensor Networks.//Proc. of the 47th IEEE Global Telecommunications Conference, November 29–December 3, Dallas, TX.

[10] Yang Y., Wang X., Zhu S., et al. (2008) SDAP, A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks. ACM Transactions on Information System Secure, 11(18): 1-43.