INTEGRITY VERIFICATION PROTOCOL BASED ON PRIVACY HOMOMORPHIC DATA FUSION

SreeRanganayaki*[1], Prof. A. Ramesh Babu[2]
Research Scholar(Ph.D), [1,2]Department of Computer Science,
Chaitanya Deemed to be University
Email Id: sreeranganayaki5@gmail.com, rameshadloori@gmail.com

Abstract: Data integration security in sensor networks can prevent effectively problems such as privacy and data usurpation, and achieve efficient data transmission. This results in a completeness verification protocol based on the integration of private homogeneous data. The agreement uses homogeneity to cryptographically guarantee data privacy; Random nodes were used to test the completeness of node aggregation to check that the nodes are able to transmit each data group faithfully. Through the analysis done theoretically but the comparison is done with simulation results comparison with existing methods, the algorithm performance is checked, and is concluded that the proposed method is able to detect data integrity in network data transmission process, and can able to achieve high accurate data and better protection privacy protection.

Key words: Data Fusion, Integrity, wireless sensor networks

I. Introduction

Wireless sensor networks (WSN) have the characteristics of energy limitation and data eccentricity, and data aggregation technology can achieve more efficient use of network resources and node energy by merging and summarizing data from different nodes, and improve the efficiency and accuracy of data collection [1]. In practice, sensor nodes are located in environments that are subject to both internal and external attacks and Security issues [2], including channel eavesdropping and data tampering. Therefore, when designing a data aggregation strategy, you should fully consider the basic characteristics such as data privacy and integrity.

Encryption is often employed to improve the privacy of data. JiaGuo et al. [3] give a review of secure data aggregation in WSN, divide the encryption methods of WSN secure data aggregation into hop-by-hop encryption and end-to-end encryption, and describe the specific aggregation process. In the SIA[4] protocol, a Merkle-hash tree is first established, and the 3 steps of aggregation-commit-proof are used to ensure the integrity and venerability of aggregation results. However, the solution adopts hop-by-hop encryption, and the aggregation node needs to encrypt and decrypt the data. He et al. [5] proposed a privacy protection algorithm SMART based on data slicing, but it is sensitive to data loss. Encryption is often employed to improve the privacy of data. JiaGuo et al. [3] give a review of secure data aggregation in WSN, and divide the encryption methods of WSN secure data aggregation into hop-by-hop encryption and end.

Encryption to the end, and describe its specific aggregation process. In the SIA[4] protocol, a Merkle-hash tree is first established, and three steps of aggregation-submission-proof are used to ensure the integrity and verifiability of aggregation results, but the scheme adopts hop-by-hop encryption, and the aggregation node needs to encrypt and decrypt the data. He et al. [5] proposed a privacy protection algorithm SMART based on data slicing, but it is sensitive to data loss.

To solve the above problems, this paper proposes an integrity verification protocol IVP based on privacy Homomorphic data fusion. The protocol adopts Homomorphic to ensure encryption privacy data, and uses nodes detected randomlyfor integrity check of node aggregation to check node aggregation node transmits packet data faithfully. The results show that the protocol can be implemented while realizing data privacy protection and integrity detection, Efficient data aggregation.

II. System model

2.1. Network model

A wireless sensor network consists of 3 types of nodes: base station base station (BS), aggregation node aggregation node (A) and source node source node (S). The network structure is shown in Figure 1, assuming that there is only one base station in this network model, and each node can calculate, send, and receive data. Build a tree structure rooted by the base station BS: BS sends the aggregation request and finally obtains the aggregation result; An aggregation node collects the data sent by its child nodes, fuses it with its own data, and sends it to its parent node. The source node collects the data in the monitoring environment and sends it to its parent node.

2.2. Key assignment

This paper adopts a random key allocation mechanism [8], the basic idea of which is that all nodes randomly select several keys from a large key pool to form a keychain, and neighboring nodes with the same key between the key chains can establish secure links. There are three stages in the formation of a random key allocation mechanism: the preallocation phase, discovery shared key, and phase establishment path key.

i. Key pre-allocation stage: first generate a large key pool with K keys and key identification, then randomly select different K keys to form a keychain, and then assign different keychains to different nodes.

ii. Shared secret discovery phase: Each node must discover the nodes with which it has a shared secret, and only nodes with a shared secret are considered connected.

iii. Path key establishment stage: If there is no shared secret between the two nodes, the link key can be established through the path where the shared secret exists. The probability that any 2 nodes can share the same key is

$$P_{\mathrm{Connect}} = 1 - \frac{((K^2-(k+1)!)^2}{(K-2k+1)!K!} \qquad (1)$$

In the same network, the attacker, as a node, also uses the same random key allocation mechanism, and also needs to randomly select k keys from this key pool with K keys, and the probability of the attacker obtaining the key is

$$2 \ _{\mathrm{Compromised}} \qquad (2)$$

In normal circumstances, $P_{Compromised}$ is a very small number.

III. Integrity Verification Protocol Based on Privacy Homomorphic Data Fusion

This paper proposes an integrity verification protocol (IVP) based on Homomorphic privacy data fusion, which can be roughly divided into four parts: system initialization, data encryption and decryption, data aggregation, and data detection.

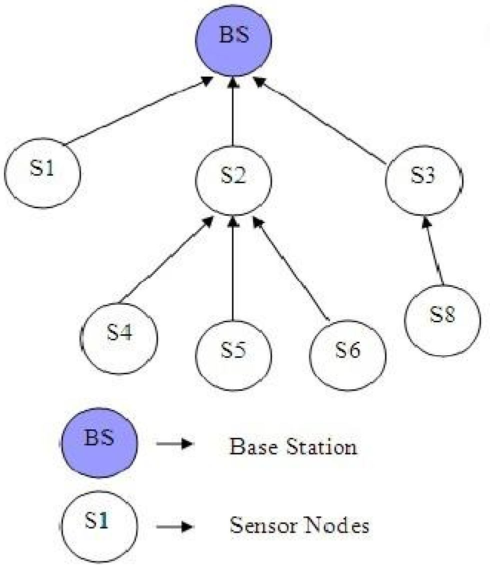

Figure 1: Aggregation tree network structure

3.1. System initialization

3.1.1. Build the aggregation tree

The basic process for building an aggregation tree is as follows: BS first sends a query request to a nearby node (let's say N0), and the N0 node broadcasts the query request to the entire wireless sensor network hop-by-hop. When all nodes in the network receive this query request, a data aggregation tree rooted in BS is generated, and the specific construction method of the aggregation tree is the same as TAG[9].

3.1.2. Selection of detection points

In order to check the integrity of data during transmission, this paper checks the integrity of node data by selecting a detection point to verify faithful node aggregation and data transmission packet.

If list of detection points in the fixed network in advance, when certain detection number points are attacked, the attacker's attack behavior will not be detected, which is not secure. Therefore, this paper adopts a random detection node selection method (DNS) to reduce the risk of detection nodes being attacked and enhance the security of the network. This selection method mainly includes 2 steps: node initialization and detection node confirmation.

i. Node initialization. Before selecting the detection point, each node loads 2 functions: a one-way hash function F(ID, x) and a mapping function fp(y). where ID is the ID of the sensor node, p is a predefined probability value, the value range of the function fp(y) is (0, 1), and when y is within the range of the F function, the value of *fp* is mapped to 1 with probability p and 0 with probability ($1-p$).

ii. Confirm the detection node. The source node generates a random number r for each piece of data, which is transmitted hop-by-hop to the base station. When the aggregation node in the transmission process receives this data, it checks the value of *fp*(F(*ID, r*)), and if the value is equal to 1, the node is selected as the detection point and generates a confirmation that the data is passed in the direction of the source node. When a child node receives the acknowledgment data sent by its parent node, the data is verified: whether the data comes from a real detection point (where r is a random number and ID' is the ID of the detection point by calculating whether *fp*F(*ID', r*)) is 1; Use the message authentication code of the data to verify whether the content of the data has been tampered with.

iii. The flow of the algorithm is shown in Figure 2. The advantage of this method of checking point selection is its randomness and dynamics. The randomness of the detection node selection makes it impossible for the attacker to determine the specific detection node, and each aggregate node may become the detection node; However, the dynamic nature of detection node selection makes the list of detection nodes uncertain, and attackers cannot predict the next detection node.

Since the detection node is randomly selected, there may be a situation where no node in the network is selected as the detection node, and only the base station of the entire network is the detection point, that is, it becomes end-to-end detection. This situation does not affect the correct operation of the detection node selection scheme.

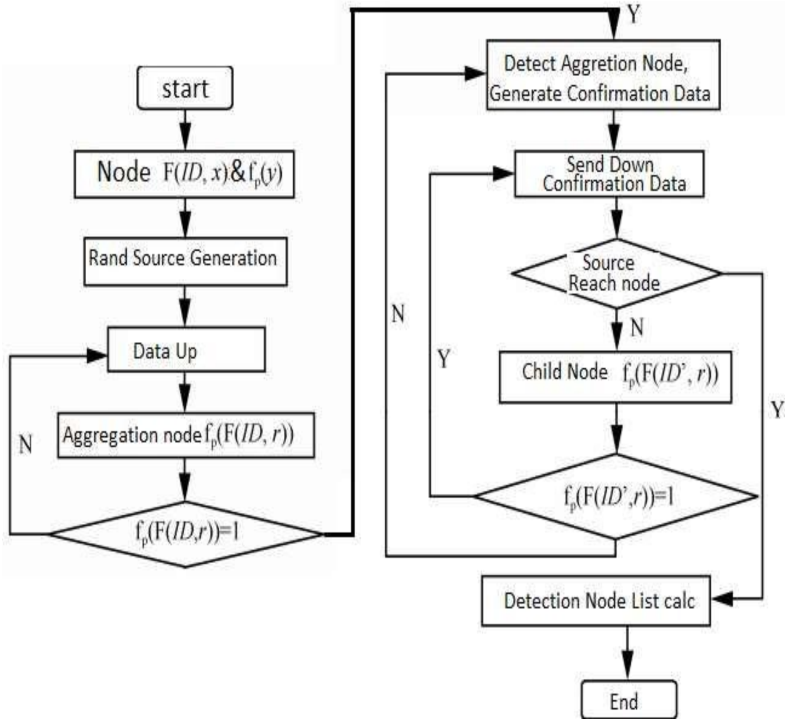

Figure 2: Checkpoint selection process

3.2. Data encryption and decryption

In this paper, homomorphic encryption technology is used [2] to encrypt the data collected by the source node from the monitoring environment. Where the common parameter d, the integer g, and the key $k=(r^2, g')$ are known to all nodes.

The following describes how to encrypt and decrypt it.

Encryption: Randomly divides the plaintext into d ciphertexts, and the encryption formula is

$$(\ ) = (\ _{1}{}^{1} \quad ,\ _{2}{}^{2} \quad , \ldots ,\ _{d}{}^{d} \quad) \qquad (3)$$

Decryption: by $a_1 r^1$mod gcalculates the scalar product of the i-th coordinate to restore $a_i$mod g, for a its is defined as

$$_{k}(\ _{k+1}(\ )) = (\sum_{i=1}^{2d+1} 2\ _{i}) \quad ' \qquad (4)$$

Algorithm:

1) The data is encrypted at the source node.

   With ‘l’ Source nodes considered for 1≤i≤l si= [1+Ek(Si)]

2) Aggregation nodes receive the si, which is transmitted from the source
3) The data is aggregated at Aggregation nodes.

   $_i$ • Here ‘i’ is considered as the child for every individual

   aggregation nodes for 1≤i≤l a $=\sum S^2$

4) Here the checkers are enabled for every detection nodes.
5) Aj node is considered as aggregation node and it is considered as node for detection
6) Then the child node Aj transmits data to the nodes of next aggregation
7) Aj receive the data from nodes of next aggregation childs.
8) The aggregate results are checked at nodes Aj aggregate and next aggregation node.
9) If the checkers are not matched, then
10) The encrypted need to be dropped and need to be informed accordingly
11) else the data need to be moved to the nodes of next aggregation
12) else the next aggregation node sends the aggregated data to next node
13) Repeat 5 to 13 steps until BS receive the data.
14) The data is decompressed and at BS and the results are checked accordingly.

Figure 3: Proposed algorithm

3.3. Data aggregation

It can be seen from the nature of homomorphic encryption [10] that the data encrypted by homomorphism can be directly aggregated at the aggregation node, and the aggregation node does not need to encrypt and decrypt the data.

Data aggregation includes sum, average, count, maximum and minimum value aggregation, etc., since other aggregation functions can be calculated by converting into summation functions [11], so this article only considers the summed data aggregation function and is defined as

$$(\ ) = \sum_{i=1}^{N}[1 + {}_{i}(\ )]^{2}$$

3.4. Data detection

The Homomorphic encryption method can ensure the end-to-end privacy of data during network transmission, but it cannot guarantee the correctness of data during transmission. In addition to detecting the final restored results aggregation, the data detection part of the IVP protocol also uses randomly selected detection points in the network to verify the aggregation results of the aggregation nodes in the network to achieve more efficient data transmission.

Based on the above four parts, this paper proposes a data fusion algorithm PROPOSED that can detect integrity, which uses the homomorphic encryption method at the source node to encrypt data to ensure privacy. In the process of data transmission, the randomly selected detection points in the network are used to detect the aggregation calculation results of the aggregation nodes. Verify the restored aggregate results at the base station. The entire algorithm process is shown in Figure 3.

IV. Performance analysis and simulation results

This paper analyzes the performance of proposed algorithm from the aspects of privacy protection, data integrity, communication overhead, and accuracy. The NS2 platform was selected for simulation experiments, and the network configuration environment was as follows: 800 nodes were randomly distributed in a 250 m×250 m area, the node transmission distance was 40m, the background noise was −99.0 dBm, the white Gaussian noise was 6 dB, and the data transmission rate was 2Mbit/s.

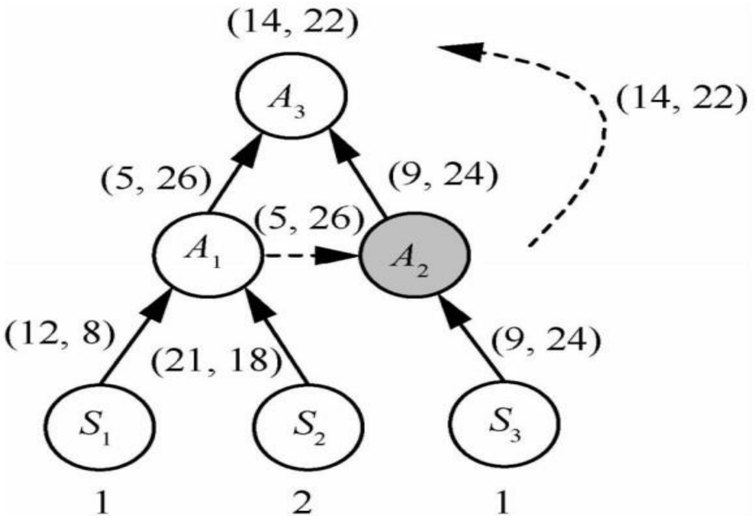

Figure 4: Network section node using PROPOSED process

4.1. Privacy Protection

The most common attack in WSN is eavesdropping attacks, where an attacker obtains information by trying to recover an intercepted ciphertext. The encryption processing in this paper adopts random division, that is, the result of ciphertext is probabilistic, so it can resist known plaintext attacks and known ciphertext attacks, and can well ensure data privacy. The randomness of the detection node selection also makes the detection node list not easy to be obtained by the attacker.

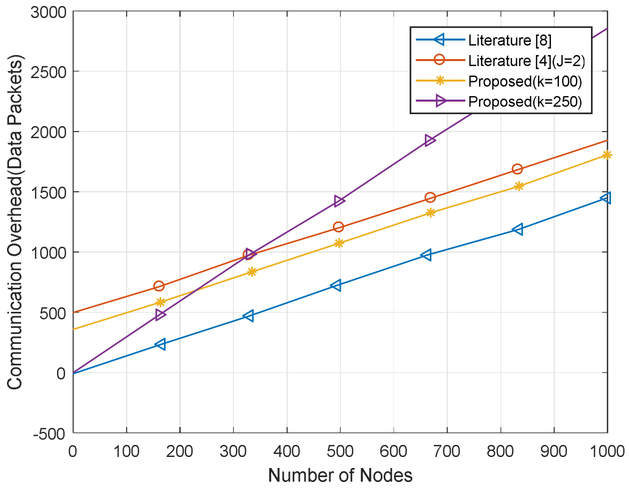

Figure 5: Communication overhead for Literature1, literature2, and proposed algorithm

4.2. Data integrity

Data integrity means that the message received by the receiver is consistent with the sender, and the information has not been tampered with or forged by the attacker. The Homomorphic encryption method can detect aggregation correctness results by restoring data (decryption) at the base station, but does not have the ability to check the integrity during data transmission.

Take the network structure of Figure 4 as an example, where A3 is the detection node, and the aggregation results of A4 are verified at A3. If the data received at A4 is tampered with (7, 68), the aggregation result of A4 will become (16, 4), and the aggregation result of the detection node is (16, 28) and the aggregation result of A4 is inconsistent, and A4 discards the data and tells A1, A2 and A3. It can be seen that the PROPOSED algorithm can detect whether the integrity of the data is broken, and improve the integrity detection ability of the network in the data transmission process.

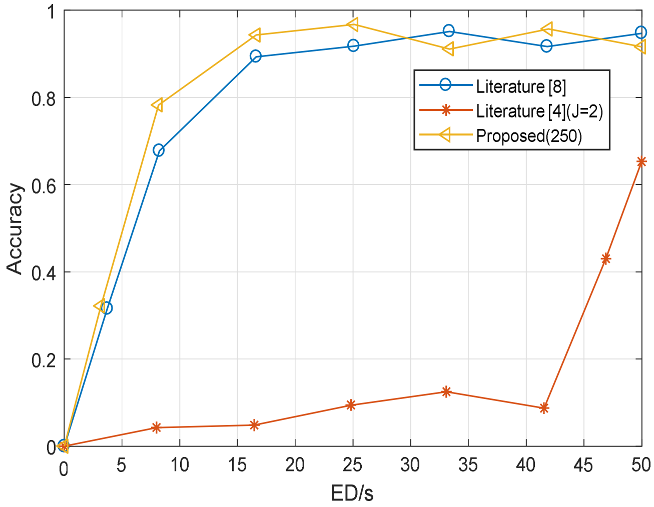

Figure 6: Accuracy comparison for Literature1, literature2, and proposed algorithm

4.3. Communication overhead

The communication overhead of a node is directly related to the energy consumption of the node. Theoretically, if N represents the number of nodes in the WSN, the data communication overhead of the literature[8] algorithm is O(N), the communication cost of the literature[4] algorithm is (J−1)N (J is the number of shards in the Literature[4] algorithm), and the proposed algorithm has 3 more communication costs of the detection node than the TAG algorithm, so the communication cost of the PROPOSED algorithm is O(N) 4k (k is the number of detection nodes in the WSN).

Figure 5 shows the comparison of the communication overhead of the literature[4], literature[8] and proposed algorithms, and the communication overhead of the three algorithms increases with the increase of nodes, and the results are consistent with the theoretical analysis.

4.4. Precision

Defines the accuracy of WSN data aggregation as the ratio of the actual data aggregation results to the sum of the data collected by the node. Ideally, there is no data loss during transmission, and the data aggregation results of the literature[4], literature[8] and proposed algorithms should be 100% accurate. However, due to the conflict of network wireless channels, node failure, and delay of data processing, the transmission information in practical applications is lost, which affects the accuracy.

Figure 6 compares the accuracy of the three algorithms literature[4], literature[8] and proposed algorithms, and the simulation is performed for changes in Epoch Duration. Since the proposed algorithm detects the data aggregation results during the transmission of data, the accuracy of the algorithm is higher than that of the literature[8] algorithm, and it can be seen from Figure 6 that its results are consistent with the theoretical analysis.

V. Conclusion

In a wireless sensor network, an attacker can exploit compromised nodes to eavesdrop on and tamper with data information. This paper proposes an IVP protocol based on privacy

Homomorphic data fusion, which realizes data privacy protection based on Homomorphic encryption, and verifies the aggregation results of nodes in the network based on random detection nodes, so as to detect whether nodes transmit data packets normally. Experimental results and theoretical analysis show that the algorithm.

It can detect the integrity of data during transmission over the network, and can have better privacy and high accuracy on the basis of lower traffic.

References:

1. He, Tian, et al. "AIDA: Adaptive application-independent data aggregation in wireless sensor networks." *ACM Transactions on Embedded Computing Systems (TECS)* 3.2 (2004): 426-457.
2. Li, Hongjuan, Kai Lin, and Keqiu Li. "Energy-efficient and high-accuracy secure data aggregation in wireless sensor networks." *Computer Communications* 34.4 (2011): 591-597.
3. Yousefpoor, Mohammad Sadegh, et al. "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: A comprehensive review." *Journal of Network and Computer Applications* 190 (2021): 103118.
4. Yang, Yi, et al. "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks." *ACM Transactions on Information and System Security (TISSEC)* 11.4 (2008): 1-43.
5. Chen, Yingwen, et al. "A blockchain-based medical data sharing mechanism with attribute-based access control and privacy protection." *Wireless Communications and Mobile Computing* 2021 (2021): 1-12.
6. Boldyreva, Alexandra, et al. "Order-preserving symmetric encryption." *Advances in Cryptology-EUROCRYPT 2009: 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings 28*. Springer Berlin Heidelberg, 2009.
7. Jose, Josna, S. Manoj Kumar, and Joyce Jose. "Energy efficient recoverable concealed data aggregation in wireless sensor networks." *2013 IEEE International Conference ON Emerging Trends in Computing, Communication and Nanotechnology (ICECCN)*. IEEE, 2013.
8. Polite, Khandys A. "PPDA: Privacy Preserving Data Aggregation in Wireless Sensor Networks." (2004).
9. Chan, Haowen, et al. "On the distribution and revocation of cryptographic keys in sensor networks." *IEEE Transactions on dependable and secure computing* 2.3 (2005): 233-247.
10. Talele, Ajay K., Suraj G. Patil, and Nilkanth B. Chopade. "A survey on data routing and aggregation techniques for wireless sensor networks." *2015 International Conference on Pervasive Computing (ICPC)*. IEEE, 2015.
11. Li, Baiyu, et al. "Securing approximate homomorphic encryption using differential privacy." *Annual International Cryptology Conference*. Cham: Springer Nature Switzerland, 2022.
12. Lyu, Lingjuan, et al. "Towards fair and privacy-preserving federated deep models." *IEEE Transactions on Parallel and Distributed Systems* 31.11 (2020): 2524-2541.