IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 10, Issue.5, May 2022 E-ISSN: 2347-2693

Privacy Preserving Data Aggregation and Data Integrity in WSN

Maharajan K.[1*], R. Vadivel[2]

[1]PG Student Department of Information Technology, Bharathiar University, Tamil Nadu India
[2]Assistant Professor Department of Information Technology, Bharathiar University, Tamil Nadu, India

**Corresponding Author: pradeemaharaj18@gmail.com*

DOI: https://doi.org/10.26438/ijcse/v10i5.5357 | Available online at: www.ijcseonline.org

Received: 23/Apr/2022, Accepted: 09/May/2022, Published: 31/May/2022

***Abstract*—**Recently, several data aggregation schemes based on privacy homomorphism encryption have been proposed and investigated on wireless sensor networks. These data aggregation schemes provide better security compared with traditional aggregation since cluster heads (aggregator) can directly aggregate the cipher texts without decryption; consequently, transmission overhead is reduced. However, the base station only retrieves the aggregated result, not individual data, which causes two problems. First, the usage of aggregation functions is constrained. For example, the base station cannot retrieve the maximum value of all sensing data if the aggregated result is the summation of sensing data. Second, the base station cannot confirm data integrity and authenticity via attaching message digests or signatures to each sensing sample. In this paper, we attempt to overcome the above two drawbacks. In our design, the base station can recover all sensing data even these data has been aggregated. This property is called "recoverable." Experiment results demonstrate that the transmission overhead is still reduced even if our approach is recoverable on sensing data. Furthermore, the design has been generalized and adopted on both homogeneous and heterogeneous wireless sensor networks.

***Keywords*—**Concealed Data Aggregation, Wireless sensor networks, privacy homomorphism encryption.

I. INTRODUCTION

Wireless Sensor Network (WSN) is widely used in many applications. B. Military site monitoring, healthcare, environmental monitoring, accident reporting, etc. WSN consists of a large number of sensors that work together. Each sensor detects targets within its radio range, performs simple calculations and communicates with other sensors. Sensors are generally limited in terms of battery power, communication, and processing power. Therefore, reducing power consumption is an important concern for WSNs. Recently, a practical solution called data aggregation was introduced. The original concept is to aggregate multiple collections of data by performing algebraic or statistical operations such as addition, multiplication, median, minimum, maximum, and mean of the dataset. Data aggregation is typically performed by the cluster head when the entire network is divided into groups called clusters.

The base station (sink) might also additionally require the most fee of all sensing records to cause the instantaneously response; thus, every cluster head selects the most fee of more than one sensing records of its cluster individuals and sends the end result to the bottom station. Communication costs are clearly reduced because only the aggregated results reach the base station. Unfortunately, your opponent can capture the cluster head. This puts the entire cluster at risk. As a result, some schemes. However, these schemes either limit the data type of the aggregate or incur additional transmission overhead. Opponents can also get sensor data for cluster members even after capturing the cluster head. Two ideas are used to completely solve the above problem. First, the data is encrypted in transit. The cluster head then aggregates the encrypted data directly without decrypting it. Based on these two ideas, a well-known approach called Concealed Data Aggregation (CDA) has been proposed. CDA provides both end-to-end encryption and in-network processing with WSN. CDA uses privacy homomorphic encryption with additive homomorphism, allowing cluster heads to perform additive operations on encrypted numeric data.

The base station receives only the aggregated results. However, there are two issues. The base station receives only the aggregated results. However, there are two issues. For example, these schemes only allow the cluster head to perform additional operations on the cipher text sent by the sensor. Therefore, if the base station wants to query the maximum value of all acquired data, these have no effect. Second, the base station cannot verify the integrity and reliability of the individual acquired data. These issues appear to have been resolved if the base station can receive all acquired data rather than the aggregated result, but this technique is data aggregation in which the base station receives only the aggregated result. Therefore, we are trying to design an approach that allows the base station to receive all the acquired data, but still reduce the transmission overhead.

This document introduces a concept called Recoverable Concealed Data Aggregation (RCDA). RCDA allows a

base station to recover all acquired data generated by all sensors, even if the data is aggregated by a cluster head (aggregator). This piece of data has two functions. First, the base station can verify the integrity and reliability of all acquired data. Second, base stations can perform arbitrary aggregation functions on them. Next, we propose two RCDA schemes named RCDAHOMO and RCDAHETE for homogenous and heterogeneous WSNs, respectively. Security analysis shows that the proposed scheme is secure under the attack model. Through experiments, we have shown that the performance of our design is reasonable and affordable. It also provides a detailed comparison with other systems.

II. BACKGROUND STUDY

R. Rajagopalan and P. Varshney [1] The wireless sensor network contains small spatially distributed sensor nodes for acquisition and data processing. These sensor nodes have very little memory and are power limited. Acquired by data aggregation A large amount of energy is consumed when acquiring, processing, and transmitting / receiving data, and the collection and collection of useful data at the sink node is performed in an energy-efficient manner. In addition, you can eliminate data redundancy and extend the useful life of your network. earn about energy-efficient tree-based data aggregation techniques in wireless sensor networks. In the tree architecture, all nodes are structured in the form of a tree containing leaf nodes and forwarding nodes. This structure can be adapted to continuous monitoring applications. Tree-based aggregation consumes less power during data transmission than cluster-based or grid-based aggregation.

H. C¸ am, S. O¨ zdemir, P. Nair, D. Muthuavinashiappan, and H. Ozgur Sanli [2] Low power and secure pattern-based data aggregation protocol (ESPDA) for wireless sensor networks. ESPDA is more energy and bandwidth efficient as the cluster head prevents redundant data from being sent from the sensor nodes. ESPDA is also secure because it does not require a cluster head to decrypt the encrypted data to perform data aggregation. In ESPDA, Clusterhead first requires the sensor node to send the appropriate sample code of the collected data. If multiple sensor nodes send the same pattern code to the cluster head, only one of them can send data to the cluster head. Therefore, ESPDA has advantages over traditional data aggregation methods in terms of energy, bandwidth efficiency, and security. Simulation results show that as data redundancy increases, the amount of data sent from the sensor node to the cluster head is reduced by up to 45% compared to traditional algorithms.

H. Sanli, S. Ozdemir, and H. Cam [3] Data aggregation is applied to extend the life of wireless sensor networks (WSNs). Some researchers consider the importance of security and propose a secure data aggregation protocol. The essence of these secure approaches is to ensure that aggregators aggregate their data in a proper and secure way. In this document, ESPDA (Energy Efficient and Secure Pattern-Based Data Aggregation) and SRDA (Secure Reference-Based Data Aggregation) operate with cluster-based WSNs and detailed security analysis different from those previously presented.

D. Westhoff, J. Girao, and M. Acharya [4] In order to save the entire energy resources of the network, the collected data needs to be integrated and aggregated on the way to the final destination. 1) Hide the captured data from start to finish, 2) provide efficient and flexible data aggregation on the network. The aggregating intermediate nodes aren't required to perform at the sensed plaintext data. Describes techniques for applying a specific class of cryptographic transformations to calculate mean and motion detection aggregate functions. Indicates that this approach is feasible for classes of "ongoing" routing protocols. Proposing a key predistribution algorithm that limits the attacker's gain explains the risk of sensor node corruption, and the robustness and reliability of the backbone to which the key predistribution and key ID-dependent "down" routing protocols are connected.

C. Castelluccia, E. Mykletun, and G. Tsudik, [5] Energy-efficient and secure pattern-based data aggregation protocol (ESPDA) for wireless sensor networks. ESPDA is more energy and bandwidth efficient because the cluster head prevents the transmission of redundant data from the sensor node. ESPDA is also secure because it does not require a cluster head to decrypt the encrypted data to perform data aggregation. In ESPDA, Clusterhead first requires the sensor node to send the appropriate sample code of the collected data. If multiple sensor nodes send the same pattern code to the cluster head, only one of them can send data to the cluster head. Therefore, ESPDA has advantages over traditional data aggregation methods in terms of energy, bandwidth efficiency, and security. Simulation results show that as data redundancy increases, the amount of data sent from the sensor node to the cluster head is reduced by up to 45% compared to traditional algorithms.

E. Mykletun, J. Girao, and D. Westhoff [6] Network data aggregation is a common technique for reducing the power consumption associated with sending data over multi-hop wireless sensor networks. They explore and investigate the applicability of additional homomorphic public key cryptographic algorithms to specific classes of wireless sensor networks. Finally, it provides recommendations for choosing the most appropriate public key scheme for different topologies and wireless sensor network scenarios.
H. Chan, A. Perrig, and D. Song [7] Sensor nodes are vulnerable to attacks that endanger the node, and security issues such as data confidentiality and integrity are very important. Recently, a robust aggregation framework called Synopsis Diffusion has been developed. It combines a multipath routing scheme with a duplicate-independent algorithm to accurately calculate aggregates (predicate count, sum, etc.) despite message loss due to node and send failures. However, this aggregation framework does not address the issue of incorrect sub aggregate values

caused by the compromised node, resulting in a large error in the aggregate calculated at the base station, which is the root node of the aggregation hierarchy. This is an important issue because sensor networks are extremely vulnerable to node compromise due to the unmanned nature of sensor nodes and the lack of tamper-proof hardware. This is an important issue because sensor networks are extremely vulnerable to node compromise due to the unmanned nature of sensor nodes and the lack of tamper-proof hardware. Through theoretical analysis and extensive simulation studies, the proposed algorithm has been shown to be superior to other existing approaches.

S. Roy, S. Setia, and S. Jajodia [8] These represent algorithms that allow the base station to safely calculate the number or sum of predicates in the presence of such attacks. Attack-resistant computational algorithms compute true aggregates by excluding the contributions of compromised nodes in the aggregate hierarchy. Extensive analytical and simulation studies show that our algorithms are superior to other existing approaches. The studies network proposed a loss-resilient aggregation framework referred to as synopsis diffusion, which makes use of duplicate-insensitive algorithms on pinnacle of multipath routing schemes to appropriately compute aggregates (e.g., predicate be counted number or sum).

H. Yu [9] they present the goal of allowing aggregate queries rather than simply detecting enemies. To this end, we propose a new tree sampling algorithm that uses sampling directly to answer aggregate queries. It uses a new set sampling technique to overcome important and well-known sampling obstacles. Traditional sampling techniques are only effective when the number or total of predicates is large. Set sampling allows you to efficiently sample a set of sensors together and determine if the sensors in the set meet the predicate (but not how many). Using set sampling as a component, tree sampling has been shown to provide the correct answer despite hostile interference, without the drawbacks of traditional sampling techniques.

W. Heinzelman, A. Chandrakasan, and H. Balakrishnan [10] Develop and analyze the Low Energy Adaptive Clustering Hierarchy (LEACH). It is a microsensor network protocol architecture that combines energy-efficient cluster-based routing and media access ideas with application-specific data aggregation to deliver superior performance in terms of system lifetime, latency, and application-perceived quality. Leach has a new distributed clustering technique that allows self-organization of many nodes, an algorithm that coordinates the cluster to rotate the position of the cluster head and distributes the power load evenly across all nodes, and distributed signal processing. It includes techniques to enable and save communication resources. Our results show that LEACH can improve system life by orders of magnitude compared to the general purpose multi-hop approach.

M. Demirbas, A. Arora, V. Mittal, and V. Kulathumani [11] FLOC, a high-speed local clustering service that divides a multi-hop wireless network into non-overlapping clusters of approximately the same size. Each cluster has a cluster head, and all nodes within a unit distance from the cluster head and some nodes within a distance of m belong to the cluster. FLOC indicates that clustering locality and fault-local self-stabilization are achieved by asserting the stretch factor. The effects of clustering and failures / changes on each part of the network are contained in up to m + 1 units. Through simulation and experimentation in real-world deployments, we analyze the trade-offs between clustering time and clustering quality and suggest appropriate parameters for FLOC to achieve fast completion times without compromising the resulting clustering quality.

S. Basagni, M. Mastrogiovanni, A. Panconesi, and C. Petrioli [12] The first simulation-based in-depth study of clustering and backbone formation techniques, one of the most representative in this area of ad hoc research. Then examine the nature of the selected protocol and assess the impact of the "degree of localization" on those operations. H. How the ability to run a protocol based solely on local information affects the overall performance of the protocol. Extensive ns2-based simulation results show that highly localized protocols are rewarded with excellent performance across all relevant metrics such as protocol duration, energy consumption, message overhead, route length, and backbone size.

III. PROPOSED METHODOLOGY

Recently, a practical solution called data aggregation was introduced. The original concept is to aggregate multiple retrieved data by performing algebraic or statistical operations such as addition, multiplication, median, minimum, maximum, and mean of the dataset. Data aggregation is typically performed by the cluster head when dividing the entire network into groups called clusters. Two ideas have been used in recent studies. First, the data is encrypted in transit. The cluster head then aggregates the encrypted data directly without decrypting it.. Based on these two ideas, a well-known approach called Concealed Data Aggregation was proposed. CDA provides both end-to-end encryption and in-network processing with WSN. CDA privacy uses homomorphic (PH) encryption with additional homomorphism, allowing the cluster head to perform additive operations on the encrypted numeric data.

A RCDA SCHEME FOR HOMOGENEOUS WSN (RCDA-HOMO)

This section proposes a recoverable hidden data aggregation scheme called RCDAHOMO for homogeneous WSNs. Structure of RCDAHOMO RCDAHOMO consists of four steps: setup, EncryptSign, aggregation, and verification. The setup procedure is to prepare and install the secret required for the OS and each sensor. When the sensor decides to send the detection data

to CH, the sensor performs EncryptSign and sends the result to CH. When CH receives all the results from the members, it activates the aggregation, aggregates what it receives, and sends the final result (aggregated ciphertext and signature) to the BS. The last method is verification. BS first extracts the individual capture data by decrypting the aggregated ciphertext. The BS then verifies the reliability and integrity of the decrypted data based on the corresponding aggregated signature.

A) RECOVERY PROPERTY

The Recovery belongings tries to offer functionalities. First, BS can verify the integrity and reliability of all collected data. The BS can then perform any aggregate operation on this data. However, in RCDAHETE, BS recovers only the individual aggregated results generated by each cluster, not all collected data. Here we show that RC DAHETE also provides these features.

1. RCDAHETE can check all detection data using HSensors. More precisely, Intracluster Encrypt procedure allows L-Sensor Lj i to send not only EKj i ðdj iÞ, but also the MAC (message authentication code) of EKj i ðdj iÞ to its cluster head Hj; therefore, Hj can verify the integrity of the data sent from its cluster members.

2. Each HSensor is loaded with some required aggregation functions prior to deployment so that the BS can instruct each HSensor to perform the specified aggregation function. For example, if the BS decides to get the sum of all the data, it assigns the BS sensor to perform an add operation. The BS can then perform the final addition when getting each result from each HSensor. Similarly, if the BS decides to perform the maximum selection operation, the BS notifies each HSensor and selects the maximum value from the discovery data of the intercluster encryption procedure.

B) SECURITY AND SCALABILITY ANALYSIS

This section shows that the proposed attack model scheme is secure. More detailed security and scalability analyzes can be found in the supplements available online. First, assume that the attacker does not endanger the sensor. The proposed scheme is secure because the ingested message is encrypted. In RCDAHOMO, each sensor encrypts the message with PBS before sending. With RCDAHETE, intra-cluster traffic is encrypted with a pairwise key. In addition, our design will generate the appropriate signature for all captured data. As a result, the attacker cannot sign the fake message without the private key, and cannot modify or insert the fake message. If your opponent has the ability to endanger the sensor, consider the following situation. An attacker could endanger the sensor and use it as a legitimate sensor. WSN's existing detection mechanism cannot detect compromised sensors that are still functioning properly. Even if the value of the fake message is within a reasonable range, it cannot be detected. Attackers can also try to manipulate the aggregated results. May generate incorrect data, modify legal messages, or impersonate other sensors. The proposed scheme is still safe against the above attacks with the signature required for each generated message. Meanwhile, at RCDAHOMO, I will explain the situation when the enemy endangers the cluster head. First, the secret decryption key is not stored in the cluster, so the aggregated ciphertext or individual ciphertext cannot be decrypted. The compromised cluster head can then selectively remove some ciphertext and signatures during the aggregation procedure.

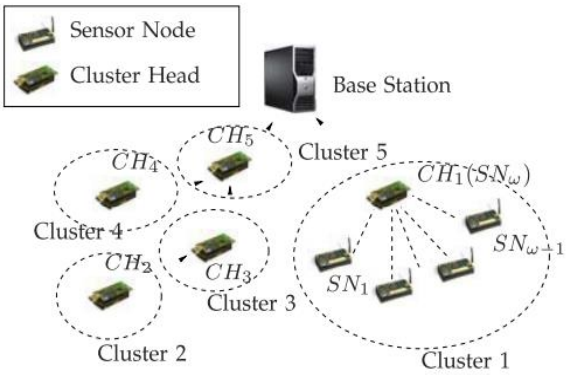

Fig 1. Example of homogeneous WSN Environment

RCDA-HETE SCHEME

Here we are trying to take full advantage of H Sensor with more computing power. You can switch the operation of L Sensor to H Sensor. In addition, HSensor is tamper-proof, allowing HSensor to partially store confidential information as needed. With these in mind, we will redesign the RCDA scheme called RCDAHETE. Using tamper-proof devices can increase hardware costs. However, in heterogeneous WSNs, most sensors are low-end sensors (LSensors). In our design, the computational cost of Lsensors has been switched to Hsensors, which makes Lsensors very cheap and simple. In fact, the overall hardware cost is reduced. RCDAHETE consists of five steps: setup, intra-cluster encryption, intercluster encryption, aggregation, and validation. The setup procedure loads the required secrets into each H Sensor and L Sensor. Intracluster encryption procedures include when the LSensor sends the discovery data to the corresponding HSensor. In the Intercluster Encrypt method, each HSensor aggregates the received data and encrypts and signs the aggregated result. Also, if the HSensor receives a ciphertext and signature from another HSensor on the routing path.

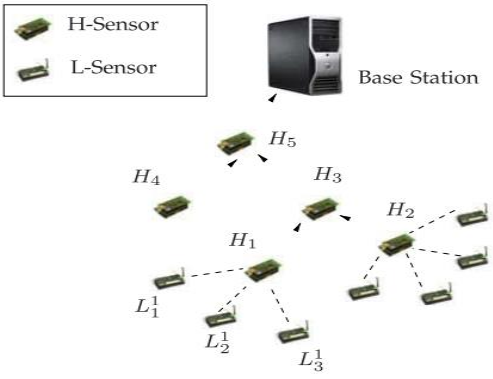

Fig 2. An Example of Heterogeneous WSN

IV. RESULT AND DISCUSSION

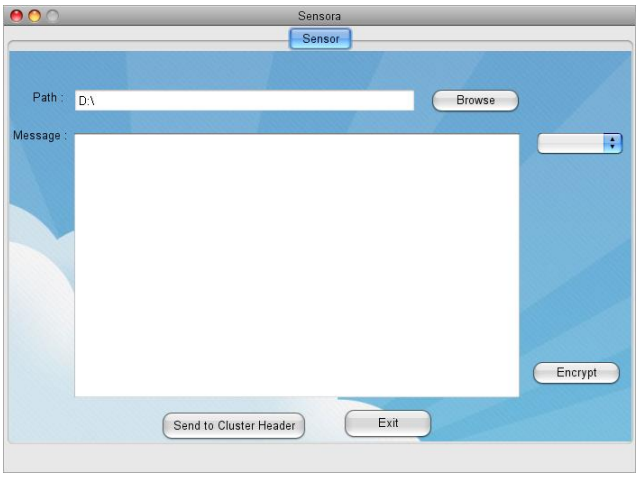

Fig 3. Sensor Data Transaction

Sensor data is the output of a device that recognizes and responds to the type of input from the physical environment. The output can be used to provide information input to another system.

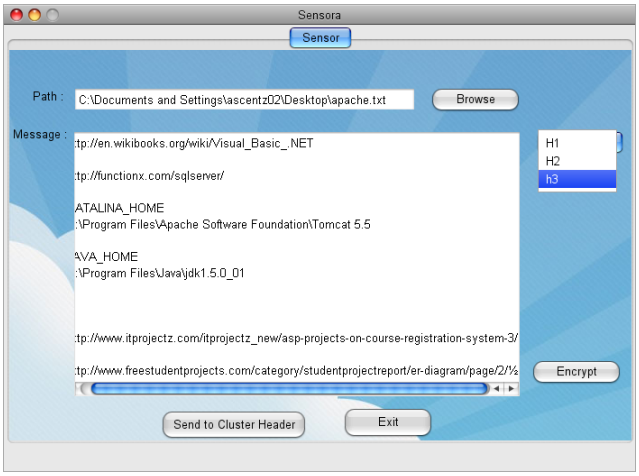

Fig 3. Selecting the Sensor

The base station cannot verify the integrity and reliability of the data by adding a message digest or signature to each capture probe. The base station can recover all acquired data even if this data is aggregated.

V. CONCLUSION

In this paper, privacy preserving data aggregation using shared key signature providing secure and authenticated data aggregation in wireless sensor network. The idea behind this scheme is the ability to share the key used in each sensors for encryption and signature verification

We have proposed recoverable concealed data aggregation schemes for homogeneous/heterogeneous WSNs. A special feature is that the base station can securely recover all sensing data rather than aggregated results, but the transmission overhead is still acceptable.

Moreover, we integrate the aggregate signature scheme to ensure data authenticity and integrity in the design. Even though signatures bring additional costs, the proposed schemes are still affordable for WSNs after evaluation.

REFERENCES

[1] R. Rajagopalan and P. Varshney, "Data-Aggregation Techniques in Sensor Networks: A Survey," IEEE Comm. Surveys Tutorials, **vol.8, no.4, pp.48-63, Oct.-Nov. 2006.**

[2] S. Madden, M.J. Franklin, J.M. Hellerstein, and W. Hong, "TAG:A Tiny Aggregation Service for Ad-Hoc Sensor Networks,"Proc. Fifth Symp. Operating Systems Design and Implementation,**2002**.

[3]J.-Y. Chen, G. Pandurangan, and D. Xu, "Robust Computation of Aggregates in Wireless Sensor Networks: Distributed Randomized Algorithms and Analysis," IEEE Trans. Parallel Distributed Systems, **vol.17, no.9, pp.987-1000, Sept.2006.**

[4] H.C¸ am, S. O ¨ zdemir, P. Nair, D. Muthuavinashiappan, and H.Ozgur Sanli, "Energy-Efficient Secure Pattern Based Data Aggregation for Wireless Sensor Networks," J. Computer Comm., **vol.29, pp.446-455, 2006.**

[5] H.Sanli, S. Ozdemir, and H. Cam, "SRDA: Secure Reference-Based Data Aggregation Protocol for Wireless Sensor Networks,"Proc. IEEE 60th Int'l Conf. Vehicular Technology (VTC '04-Fall),**vol.7, pp.4650-4654, Sep det. 2004.**

AUTHORS PROFILE

Mr. Maharajan. K received Bachelor's Degree in Computer Science in the year 2020 from K R College of Arts and Science, Kovilpatti, Tamil Nadu, affiliated to Manonmanian Sundaranar University. He is currently pursuing a Master's Degree in Information Technology from 2020 to 2022, at Bharathiar University, Coimbatore,Tamil Nadu.

Dr. R.Vadivel is an Assistant Professor in the Department of Information Technology, Bharathiar University, Tamil Nadu, India. He received his Ph.D. degree in Computer Science from Monomaniam Sundaranar University in the year 2013. M.E., Degree in Computer Science and Engineering from Annamalai University in the year 2007. B.E., Degree in Computer Science and Engineering from Periyar University in the year 2002. He obtained his Diploma in Electronics and Communication Engineering from State Board of Technical Education in the year 1999. He had published over 88 journals papers and over 45 conferences papers both at National and International level. His areas of interest include Computer Networks, Network Security, Information Security, etc.

