STUDY & DESIGN OF ADVANCED DATA AGGREGATION TECHNIQUE IN WIRELESS SENSOR NETWORKS

Soumyasri S M¹, Dr. Rajkiran Ballal² and Ramakrishna Hegde³
¹²Computer Science, SDM College, Ujire, Electrical & Electronics, Marine Engineering College
²³Computer Science & Engineering, SDM Institute of Technology, Ujire.

Abstract- In today’s fast growing technology, most of the applications uses Wireless Sensor Networks(WSN) and it plays the important role to achieve the required results. The important objective while designing the WSN is to maintain the data privacy, so that neighboring nodes should not be able to get the private data and also implementing the efficient data aggregation techniques to achieve better performance in privacy preservation, less communication overhead and accuracy in intermediate data aggregation. In this paper two privacy preserving data aggregation techniques are discussed namely Cluster Based Privacy Data Aggregation(CPDA) and Slice-Mix-AggRegaTe (SMART). CPDA will have an advantage of better privacy preservation and also less communication overhead. In SMART, better privacy preservation can be achieved but there is a communication overhead.

I. INTRODUCTION

A Wireless Sensor Networks(WSN) is an Ad-hoc network consists of large number of small sensor nodes which are present in physical world to measure and analyze the environment conditions. Wireless Sensor Networks have very broad application base that includes the military and civilian field. Some of the important functionalities are Surveillance [1], tracking[2], monitoring objects[3]. The technical advancement in sensor networks changed the way people monitor, communicate and interface with the environment.

Sensors are usually small in size and designed to integrate into the real world systems. Most of the sensors are customized to suite specific applications and designed with power constrained as well as limited resources. In modern days, many heterogeneous systems should communicate with each other to accomplish the task. Since most of the systems are networked, it is very important to have efficient technique to handle all the sensors present in different systems. These sensors are suffers from restricted computation, limited user interfaces, restricted communication mechanism and limited power supply. Sensors normally provides raw data which is needed to be processed. To avoid the insufficient network bandwidth problem, only processed data can be sent over a network. Here we are referring about data aggregation. The designers of sensor network applications are mainly concerned about aggregate statistics which can be calculated over a period of time or certain regions. As a result, WSN applications receives lot of attention about data aggregation techniques.

A sensor networks are used almost all the applications in today’s life and most of the times these applications are expand to the very sensitive environments, securing the data privacy is the primary concern. For example, applications may related domestic need like getting the information about power, water usage, details on other household appliances. In these cases we will not be able to recommend to the public without the proper security mechanism. Now a days IoT (Internet of Things) are the latest technologies used every where including the domestic applications. In these cases, it may be necessary to gather a lot sensitive information in which leaking of these data may cause a serious damage to the environment. So if the system is not secure and may not guarantee a data privacy, then participating parties may not be interested in these applications. In this paper, we are discussing about how we can make these system more secure and preserving techniques for data privacy. Here we are discussing two specific
motivating applications of using Wireless Sensor Networks in which data privacy preserving is the primary goal.

- Wireless Sensor Network applications can be used in houses to collect the information about the water usage, electricity consumptions and household appliances usage with in a large neighborhood. These aggregated information may be helpful for the government and individual business for their planning. In these cases, without the proper data preserving technique, WSN may not practical. Hence we need a system which can collect aggregate data and also preserve the data privacy.

- Applications related to floor sensors, collecting the data on health information of the residents, so that based on these information health departments and insurance companies can plan accordingly. These health related data should be kept secure and privacy should be maintained.

For the above two specific applications, we will see here why preserving data privacy while collecting accurate data aggregation is important goal. Consider the scenario, the intruder trying to collect some sensitive information by intrude into the system and these sensitive information can be used to damage the individual privacy. Even in these case also, if we use proper encryption techniques on these data, intruder will not get original data, so the privacy can be preserved. Its is well know that end to end encryption is better technique provided both end parties have an agreement on encryption key. In this paper we are improving two privacy preserving data aggregation schemes called Cluster Based Private Data Aggregation (CPDA) and Slice-Mix-AggRegaTe (SMART). The goal of our work is to provide efficient data aggregation technique without compromising the privacy of the data. Our proposed work is to provide no packet loss, in both CPDA and SMART, so that WSN can obtain precise aggregation results. That means no sensor data are released to another sensors.

In CPDA scheme sensor nodes are arranged randomly into clusters. With in each cluster, our proposed technique uses algebraic properties of polynomial to calculate required data aggregate value. At the time, no nodes will have an information about other nodes. In SMART each node hides its private data by slicing into smaller pieces and these sliced data are encrypted before sending to other intermediately nodes. We evaluate these two schemes in terms of level of privacy preservation, communication overhead, accuracy of data aggregation, comparing with other aggregation scheme TAG[4], where on data privacy is provided.

II. RELATED WORKS

There has been extensive work on data aggregation schemes in sensor networks, including [4], [5], [6], [7], [8], [9]. In all these cases the assumption is that all sensors are trusted and all communications are secure. But in practical, all the sensor node are deployed in untrusted environments and all the communication channels between these node are vulnerable to attacks by the intruders. Research work in [10], [11] shows how to use the secret keys between sensor nodes to guarantee secure communications. For most of existing data aggregation technique, received data should be decrypted and aggregate the data based on the aggregation function and encrypt the aggregate result before forwarding. Because of this complexity, all these methods are very expensive and time consuming. To minimize the computational overhead, Girao et al. [12] and Castelluccia et al. [13] demonstrated using homomorphic encryption ciphers, which introduces aggregation technique without decryption involved in intermediate nodes.

In privacy-preservation domain, Huang, Wang and Borisov giving the solutions to the problem in a peer-to-peer network application in [14]. Privacy preservation has also been studied in the data mining domain [15], [16], [17], [18]. Two techniques are used. First technique is based on the data perturbation (randomization) techniques. In this techniques recovering the aggregate result is possible and also privacy is achieved up to certain extent. But this technique may not give accurate aggregation result. Another technique for privacy-preserving data mining schemes are based on Secure Multi-party Computation (SMC) techniques [19], [20], [21]. This technique is computationally expensive because of leverage of public key cryptography characteristics.
III. METHODOLOGIES

A. Sensor Networks and the Data Aggregation Model

In this proposal, a sensor network is modeled as a connected graph $G(V, E)$, where sensor nodes are represented as the set of vertices $V$ and wireless links as the set of edges $E$. The number of sensor nodes is defined as $|V| = N$. In this paper, this focuses on additive aggregation functions, that is, $f(t) = \sum_{i=1}^{N} d_i(t)$. In this case, note that, using additive aggregation functions is not too restrictive, since many other aggregation functions, including average, count, variance, standard deviation and any other moment of the measured data, can be reduced to the additive aggregation function sum [13].

B. Private Data Aggregation

In most of the wireless sensor networks applications, preserving the privacy of the data is the essential characteristic. The desirable characteristics of private data aggregation techniques are

**Privacy:** Each nodes data should be known by itself and should not be known to other Nodes and also among the compromised nodes, the attack and collusions can be handle up to some extent by using private data aggregation techniques. The attackers can attack the wireless links to reveal the private data. An effective private data aggregation techniques can be an robust to these kind of attacks.

**Efficiency:** The main goal of the data aggregation is to minimize the amount of data transmitted in sensor networks as much as possible, thus the usage of power and resources can be reduced. Bandwidth efficiency can be achieved by using data aggregation.

**Accuracy:** Aggregation of sensor data should be accurate and also remember that no other sensor nodes should know the private data of the other nodes. Accuracy is the criteria on measuring performance of the private data aggregation techniques.

C. Encryption Key Set up

Before going to the encryption key set up, we briefly present here with random key distribution mechanism [10]. There are two data aggregation technique namely CPDA and SMART – prevents the attackers from eavesdropping. In [10] key distribution scheme it consists of three stages. 1. Key pre distribution, 2. Shared key discovery and 3. Path key establishments. In the first stage last key pool of $K$ keys are and their corresponding entities are generated. From this key pool, each sensor in the sensor networks $k$ keys are randomly drawn. During second stage, each sensor node find out which neighboring sensor node shares the common key with itself by sending the discovery messages. If any two nodes share the common key, then link between those two nodes are very secure. In third stage, path key is assigned to each nodes who do not share the common keys. The Key distribution algorithm discussed above is prevents the eavesdropping over the network. The key distribution algorithm supports secure communication in large scale sensor networks and uses small number of keys. In random key distribution, any two nodes possibility of acquiring at least one common key is given by

$$P_{\text{connect}} = 1 - \frac{((K - k)!)((K - k)!)}{(K - 2k)! K!}$$

(1)

Let the probability that any other node can overhear the encrypted message by a given key be $P_{\text{overhear}}$.

$$P_{\text{overhear}} = \frac{k}{K}$$

(2)

D. Cluster Based Private Data Aggregation (CPDA)

1) **Cluster Formation:** In CPDA, initial step is to construct clusters to perform intermediate aggregation.

   - Query server triggers query with HELLO message, one of the sensor node elects as cluster leader.
   - Cluster leader communicates HELLO message to the neighbor nodes.
• Otherwise it waits for the other nodes to send the message and joins with that node by sending JOIN request.
• When this procedure is processed many times, then many clusters are created.

2) Calculation within the clusters: In this step of CPDA, intermediate aggregation with in the clusters are taken place.

3) Cluster Data Aggregation: In this step, with the help of routing tree, Data Aggregation is taken place.

E. Slice-Mix-AggRegaTe (SMART)

In CPDA technique, one drawback is data aggregation computational overhead with in the clusters. SMART technique reduces this overhead with expense of slight increase in the consumption is communication bandwidth.

Slicing : Every node \( i \) \((i = 1, 2 \ldots \ldots N)\) set of nodes \( S_i (|S_i|) \) within \( h \) hopes. For a dense WSN, we can take \( h = 1 \). Node \( i \) then slices its private data \( d_i \) randomly into \( J \) pieces (i.e., represents it as a sum of \( J \) numbers). One of the \( J \) pieces is kept at node \( i \) itself. The remaining \( J - 1 \) pieces are encrypted and sent to nodes in the randomly selected set \( S_i \). We denote \( d_{ij} \) as a piece of data sent from node \( i \) to node \( j \).

Mixing: When a node \( j \) receives an encrypted slice, it decrypts the data using its shared key with the sender. Upon receiving the first slice, the node waits for a certain time, which guarantees that all slices of this round of aggregation are received. Then, it sums up all the received slices.

Aggregation : All the nodes aggregate the data and send the result to the query server.

IV. RESULTS AND DISCUSSIONS

In this section we are discussing about how these privacy preserving data aggregation schemes are evaluated. Evaluation performed in terms of privacy preservation, efficiency and aggregation accuracy in comparison with TAG [4].

A. Privacy Preservation Efficacy:

Before proceeding for determination of privacy preservation performance, privacy metrics should be defined. In WSN, when attackers eavesdrop the communication, the private data of the sensor node can be disclosed to s nodes. There are two scenarios of privacy violation.

1. When the communication key is held by unauthorized nodes and trying to decrypt the message sent by the other nodes. According to key distribution mechanism used in this paper, the probability of having communication key by the eavesdropper and the neighbor node is given by \( P_{eaves} \) (Equation 2).

2. Many nodes are collude to collect the private data collected by node \( s \). A privacy metric \( P(q) \) is defined as the probability that the private data of node \( s \) is disclosed for a given \( q \) under either conditions above.

1. Privacy Preservation Analysis of CPDA: In CPDA, only when sensor nodes exchange the data within the same cluster, private data may be disclosed to neighboring node. Assume that cluster size \( m \), nodes need to send \( m-1 \) messages to the \( m-1 \) nodes within the cluster. It can extract private data of the member only when node knows all \( m-1 \) keys. Otherwise private data cannot be known.

2. Privacy Preservation Analysis of SMART : In the SMART scheme, a sensor node \( s \) slices its private data into \( J \) pieces and then encrypts and sends \( J - 1 \) pieces to its neighbors. It keeps one piece to itself. Because of this out degree of \( s \) is \( J-1 \) and in degree of \( s \) is number of neighbor encrypt and communicate it to the node \( s \). The eavesdropper can able to get private data only if they break the \( J-1 \) link and all incoming communication. Otherwise private data cannot be extracted at all.

Fig. 1 compares the performance of privacy preservation under CPDA and SMART. With reference to Fig.1 smaller value of \( P_c \) (probability of node independently becoming cluster leader), average cluster size is larger. Because of this performance of privacy preservation is better. But computational overhead to calculate the intermediate aggregation is larger, if the size of the cluster is larger. In SMART, the larger the value of \( J \) (the number of slices each node chooses to decompose its
private data), the better privacy can be achieved. However, a larger $J$ will also yield larger communication overhead. For both CPDA and SMART, there is a design tradeoff between the privacy protection and computation/communication efficiency.

**B. Communication Overhead**

To protect the data privacy in CPDA and SMART, data hiding and encrypted communication technique is used. By using data hiding technique, private data can be embedded into different media files. Because of this there is a communication overhead. Fig. 2 show the communication overhead TAG, CPDA where $p_c = 0.3$, and SMART with $J = 3$. Here we can see that communication overhead of CPDA is less twice than the TAG. But in SMART the communication overhead is double the value of TAG.

![Fig 1. $P_q$ of CPDA and SMART](image)

![Fig 2. Communication Overhead](image)

**C. Accuracy :**

In an ideal case when there is no data loss in communication network, then CPDA and SMART give 100% data aggregation accuracy. But practical in WSN, there is data loss over the communication network, that in turn affects the accuracy and we need calculate the accuracy metrics.
In WSN, providing accurate and efficient data aggregation is a difficult task while preserving privacy. In many civilian applications and high secure military applications, there is demand greater privacy in preserving the data. If eavesdropper can able to crack the private data easily, then many are not willing to participate in private data collection mechanism at all. To achieve greater privacy, this paper explains two efficient methods called CPDA and SMART mainly focusing on additive data aggregation functions. Based on the experiments carried out in this paper, we can see that CPDA and SMART achieves the better privacy. The communication overhead is less in CPDA compared to SMART. Computational overhead is fair in CPDA and small on SMART.

REFERENCES

[1] D. Culler, D. Estrin, and M. Srivastava, “Overview of Sensor Networks,” IEationEE Computer, August 2004.
[2] N. Xu, S. Rangwala, K. Chintalapudi, D. Ganesan, A. Broad, R. Govindan, and D. Estrin, “A Wireless Sensor Network for Structural Monitoring,” Proceedings of the ACM Conference on Embedded Networked Sensor Systems, Baltimore, MD, November 2004.
[3] Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, “Wireless Sensor Networks for Habitat Monitoring,” WSN’02, Atlanta, Georgia, September 2002.
[4] S. Madden, M. J. Franklin, and J. M. Hellerstein, “TAG: A Tiny AGgregation Service for Ad-Hoc Sensor Networks,” OSDI, 2002.
[5] C. Intanagonwiwat, R. Govindan, and D. Estrin, “Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks,” MobiCom, 2002.
[6] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, “Impact of Network Density on Data Aggregation in Wireless Sensor Networks,” In Proceedings of the 22nd International Conference on Distributed Computing Systems, 2002.
[7] Deshpande, S. Nath, P. B. Gibbons, and S. Seshan, “Cache-and-query for wide area sensor databases,” SIGMOD, 2003.
[8] Solis and K. Obraczka, “The impact of timing in data aggregation for sensor networks,” ICC, 2004.
[9] X. Tang and J. Xu, “Extending network lifetime for precision constrained data aggregation in wireless sensor networks,” INFOCOM, 2006.
[10] L. Eschenauer and V. D. Gligor, “A key-management scheme for distributed sensor networks,” in Proceedings of the 9th ACM Conference on Computer and Communications Security, November 2002, pp. 41–47.
[11] D. Liu and P. Ning, “Establishing pairwise keys in distributed sensor networks,” in Proceedings of 10th ACM Conference on Computer and Communications Security (CCS03), October 2003, pp. 52–61.
[12] J. Girao, D. Westhoff, and M. Schneider, “CDA: Concealed Data Aggregation for Reverse Multicast Traffic in Wireless Sensor Networks,” in 40th International Conference on Communications, IEEE ICC, May 2005.
[13] C. Castelluccia, E. Mykleved, and G. Tsudik, “Efficient Aggregation of Encrypted Data in Wireless Sensor Networks,” Mobiquitous, 2005.
[14] Q. Huang, H. J. Wang, and N. Borisov, “Privacy-preserving friends troubleshooting network,” in Symposium on Network and Distributed Systems Security (NDSS), San Diego, CA, February 2005.
[15] R. Agrawal and R. Srikant, “Privacy preserving data mining,” in ACM SIGMOD Conf. Management of Data, 2000, pp. 439–450.
[16] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, “Privacy Preserving Mining of Association Rules,” in Proceedings of The 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 2002.
[17] H. Kargupta, Q. W. S. Datta, and K. Sivakumar, “On The Privacy Preserving Properties of Random Data Perturbation Techniques,” in the IEEE International Conference on Data Mining, November 2003.
[18] Z. Huang, W. Du, and B. Chen, “Deriving Private Information from Randomized Data,” in Proceedings of the ACM SIGMOD Conference, June 2005.
[19] C. Yao, “Protocols for secure computations,” in 23rd IEEE Symposium on the Foundations of Computer Science (FOCS), 1982, pp. 160–164.
[20] D. Ronald Cramer and S. Dziembowski, “On the Complexity of Verifiable Secret Sharing and Multiparty Computation,” in Proceedings of the thirty-second annual ACM symposium on Theory of computing, 2000, pp. 325–334.
[21] J. Halpern and V. Teague, “Rational Secret Sharing and Multiparty Computation,” in Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, 2004, pp. 623–632.