IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

A Survey on Privacy Preserve Methods in Data Aggregation

Harsha K.M[1*], Divya James[2]

[1,2]Dept. of Information Technology, Rajagiri School of Engineering and Technology, KTU University, Kochi, India

Corresponding Author: harsha.k.m84@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i6.755760 | Available online at: www.ijcseonline.org

Accepted: 14/Jun/2019, Published: 30/Jun/2019

***Abstract*—** Wireless Sensor network has improved their enhancement in enormous applications like monitoring environment, health application monitoring, military surveillance and also tracking of target that are of real time field. When there is communication between the nodes it affects the network lifetime which results in large consumption of energy. In order to reduce this energy consumption, Data aggregation techniques have been employed which eliminates the unnecessary data that is travelling from source node to sink node. When Wireless Sensor Network is deployed in a hostile environment the sensor node would be susceptible to node failures and also compromised by the adversary, thus it would become critical. There are different data aggregation approaches in wireless sensor network mainly used to increase the consumption of energy and also there are different approaches to preserve the data aggregation mainly used protect the data. It also preserves the various security issues such as data freshness, data integrity, data confidentiality in data aggregation.

Keywords— Wireless Sensor Network, Data Aggregation, Network lifetime, Energy consumption**.**

I. INTRODUCTION

Wireless Sensor Network is an emerging technology which consists of base station and large number of sensor nodes. The role of sensor nodes is to sense the data from the environment and gathers and transmit the data to the base station. There are various applications in WSN which are intrusion detection, habitat monitoring and health care system and used to monitor physical and environmental conditions like sound, pressure and temperature [1]. Due to employment of nodes in a random and impenetrable manner which is difficult for recharging and replacing the batteries. The main concern of WSN is to maximize the network lifetime[2].

In WSN it consists of there is a source node and base station that acts as an arbitrator between the users and the sensor network. We can inject queries into the sensor network and can gather the result from the base station, this can be done only if anyone wants the information from the sensor network. Data gathering is the way gathering the detected information from different sensor hubs which is transmitted to base station for further handling [3]. It is unsuitable for the sensor nodes to transmit the data directly to base station because the nodes are resource constraint. In extensive sensor network, the information produced is generally immense for the base station to do the processing. Subsequently, there is a need to combine and condense the information at the sensor node which eventually dispose of the excess data transmitted to the base station resulting in energy conservation. This can be effectively well versed by Data aggregation. It is the process of accumulating the data from sensor nodes to eliminate unwanted transmission and give quality data to the base station.

The rest of the paper is organized as follows: In section II data aggregation architecture, security requirement and performance measures are explained. In section III different data aggregation approaches are discussed. In section IV different privacy secure methods are explained. At last, section V is about the conclusion of the paper.

II. DATA AGGREGATION

Data aggregation normally involves the combined data from numerous sensor nodes and the aggregator node transmit the accumulated information to the base station. There are three types of nodes in WSN: sensor nodes, aggregator node and the querier node.

Regular sensor nodes sense the information from the environment and send to the aggregator hubs essentially these aggregator hubs gather information from different sensor hubs of the system, totals the information bundle utilizing a some accumulation work like sum, average,count, max min and afterward sends totals result to upper aggregator hub.It takes a lot of energy when the

correspondence between the sensor nodes, aggregator nodes and the querier nodes takes place.

Fig.1 represents the architecture of data aggregation, it is the process of the sensing the data from the environment and which is aggregated to the base station by using the data aggregation approaches. Here the aggregated data is obtained by using the data aggregation algorithm such as centralized approach, LEACH(Low Energy Adaptive Clustering Hierarchy), TAG(Tiny Aggregation) etc. This collected information is transmitted to the base station (sink node).

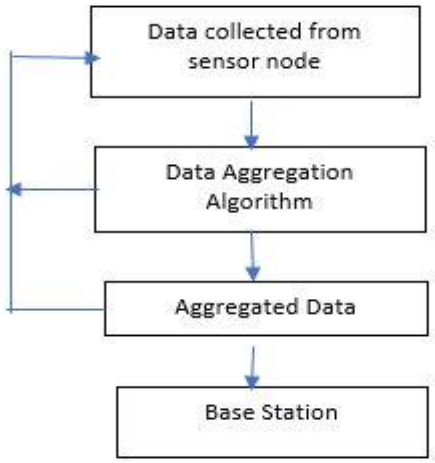

Figure 1: Architecture of Data Aggregation

A. Security Requirements of Data Aggregation

Sensor network need to satisfy a few necessities for providing a secure communication. General security prerequisites of wireless sensor network are availability, confidentiality, integrity and authentication. These necessities gives assurance against the data transmitted over the sensor network[4].

1. Data Confidentiality

In sensor network, chances of data spill is more when the information streams from many intermediate nodes. To provide the data confidentiality, an encrypted data utilized so just beneficiary decodes the data to its unique form.

2. Data Integrity

Data received by the receiver ought not be adjusted or altered is Data Integrity. Original data is changed by interloper or because of harsh environment. The interloper may change the information as indicated by its need and sends this new information to the receiver.

3. Data Authentication

It the method of affirmation that the communicating node is the one that it processes to be. It is imperative for receiver node to do the verification that the data is gotten from an authentication node.

4. Data Freshness

In this it suggests that each message transmitted over the channel is new and fresh. It ensures that the old messages can't be replayed by any node. This can be unwound by adding some time related counter to check the freshness of information.

5. Data Availability

Data Availability implies that the services are accessible all the time even if there should be an occurrence of a few assaults such as Denial of Service.

B. Design Goals of Data Aggregation

Design objectives are used to attain precise data aggregation and to preserve privacy of the data, they are as follows;

1. Latency

It implies a defer that is associated with routing, data transmission and data aggregation. It tends to be estimated as the time taken between the packet came to at the sink node and forward the packet at the source node. This factor considered as an essential factor in a few applications, for example, real time data tracking and security surveillance. In order to support the scalability of the network for the periodic applications, the data must be gathered by the data aggregation protocol from the network without considering the system estimate and outline measure.

2. Energy Efficiency

Because of the impediment in the energy resources, the WSN languishes over longer duration, so the main aim of the data aggregation techniques is to accomplish optimal depletion of energy by reducing the measure of information transmitted to sink. The correspondence between the nodes should synchronized by the data aggregation scheduling and enables the nodes to switch their radio transceivers in off state in most of the time to minimize it.

3. Data Accuracy

The data aggregation protocol diminishes the measure of transmitting data effectively. Here the aggregated value may drop some data. To ensure the precision of data there is essential to expand the nodes which are engaged with the data aggregation process. The appropriation of holding up time successfully, enables every node to receive and aggregate the sensed information of its predecessors and hence shields the precision of data.

4. Collison

The data aggregation protocol convention helps in keeping the impact by reducing the data volume that should be carried out by the network. Moreover it additionally

decreases the aggregate of the collision by data path selection and data scheduling.

III. DATA AGGREGATION APPROACHES

Performance of routing protocol depends on data aggregation process. The primary purpose is to aggregate the data so as to limit the energy utilization.So sensor nodes should route the packets dependent on the information packet content and choose the next hop in order to promote the aggregation.Fundamentally the routing protocol is isolated by the network structure,that is the cause why routing protocols is based on considered approaches.

A. Tree Based Approaches

The tree based methodology is characterizing from constructing an aggregation tree. The type of tree is minimum traversing tree, sink consider as a root and source hub consider as leaves. Information spilling of data begin from source node over to root node implies sink(base station). Disadvantage of this methodology, if there should arise an occurrence of data packet misfortune at any level of tree, the information will be lost for single level as well as for entire related sub tree too. Example is TAG (Tiny aggregation). It depends on two phases distribution phase and collection phase[5].

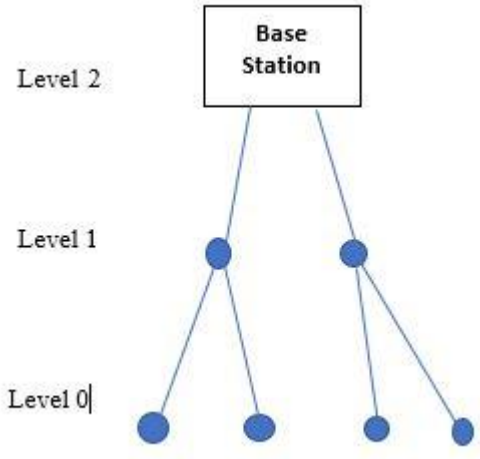

Figure 2: Tree Based Approach

B. Cluster Based Approach

It is difficult for sensors to transmit the information specifically to the sink due to the energy constraint sensor network of larger size. Here, the cluster based approach is of hierarchical approach. In this the entire network is partitioned into several clusters, where each cluster has cluster head which is chosen by the cluster members based on the high energy efficiency. Cluster-heads play the task of aggregator in which total information got from cluster group and afterward transmit the outcome to base station. The fig.3 illustrates cluster based sensor network. Here the cluster head can communicate right to the sink through long range transmission or other cluster head through multi hop. Example is LEACH protocol(Low Energy Adaptive Clustering Hierarchy).It is mainly used for applications which have constant monitoring and periodic data reporting. It consist of two phases; Cluster Setup Phase and Cluster Steady Phase.

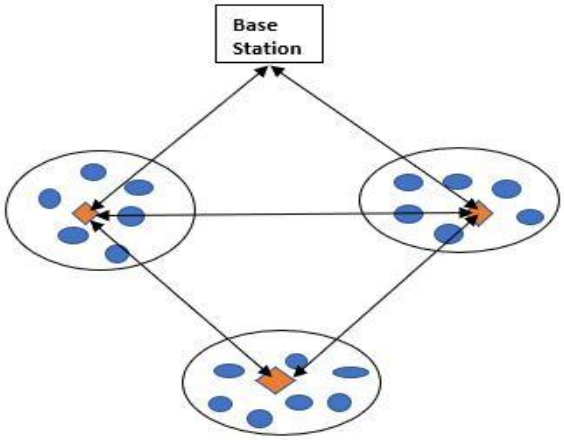

Figure 3: Cluster Based Approach

IV. PRIVACY PRESERVING METHODS

There are numerous techniques with the end goal to ensure the protection of the source data. Privacy preserving of data is mainly achieved by secure key distribution and randomized data perturbation technique. There are two types of secure aggregation protocol in WSN, that are End to End encryption and Hop by Hop encryption. In End to End encryption protection guarantees only sink node can get the data. The intermediate nodes don't have the way to learn the keys that are just shared between source and sink node. The aggregation is done by the aggregator node on the encrypted data even without the decryption. Where as in Hop by Hop encryption, it considers that all the nodes are trusted. Thus, the sharing of key with the parent prompts inconvenience in the key administration. It is exceptionally costly because of decrypting each value before aggregation [6]. In this section, following are the different methods of securing the privacy of data aggregation in WSN.

A. Preserving Data and Key Privacy in Data Aggregation

In paper[7],propose Preserving Data and Key Privacy used to attain data and key privacy protection in data aggregation. It will also reduces computational and communicational overhead. By using this technique it does not unveil the key and data to other nodes in the network when its encrypted content of the data is shared. First step is the secure key distribution, here each node has a specific secret key (Ki),

unique identifier (IDi). The sink node also has special secret key (Ni) for each node and it encrypts (Ni) with (Ki) i.e node specific key which is sent to node i in the network .Each node upon receiving already loaded with secret key (Ki),used to decrypt and get the node specific secret key (Ni) i.e Ni = D(Mi) i = 1 to n. This key Ni is known to sink and node i, this is changed for each node and for each session which ensures data freshness in the network. It employs homomorphic End to End encryption which is used to achieve data confidentiality[8].For securing the privacy of data aggregation it consists of four steps, they are;

1. Tree Construction

The WSNs consists of enormous sensor nodes. The figure 5 shows that the node S sends its id and level to all the children nodes. When a node D receives this message for the first time then node D assigns its level to be level of node S added with one and assign it to parent to be S. After this node D transmits its level and id to its children nodes hence this process is continued when all the nodes have got the message. It is arranged into a tree structure, the sink node is disseminate its secret keys to all the nodes in the network through this tree.

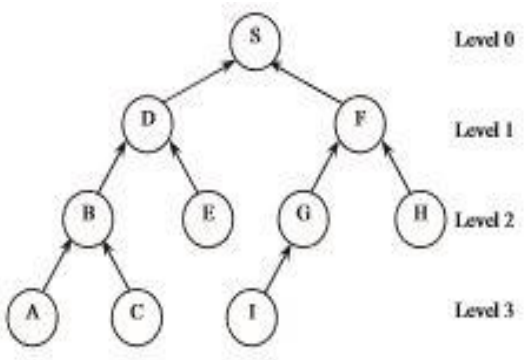

Figure 4: Tree Construction

3. Leaf Node Encryption

The encryption of data is done on the leaf node without aggregation to its aggregator node. When a leaf node A sends a packet to the parent node B, it will include the sensors identity of A, encryption of sensors reading of node A and its count value.

4. Intermediate Node Encryption

Here the data encryption is done at the source node and it can be decrypted by the sink node only. The leaf node sends its encrypted data to the intermediate node and also adds the received encoded data from all the leaf nodes and its encrypted data. When intermediate node B wants to sent encrypted data to parent node D, it will include Sensors identity of A,B and C and also encryption of sensors reading of node B and its count value is 3.

5. Decryption at the Sink Node

Here the sink node receives the aggregated data with the nodes ID and count value.It finds the node with special secret key for the IDs and subtracts all node special secret keys to get the original data. Thus this method is an efficient scheme mainly used to preserve privacy of key and data which reduces communicational overhead and energy consumption. By using the secured and strong key management the security of data and key is achieved.

B. Power Efficient Privacy Preserving Data Aggregation

In paper[8] is based on privacy preservation protocol which achieves non deferred data aggregation by doing aggregation on the encrypted data. Thus node compromise attack frequency would get reduced. The sink node would get aggregated result with reduced communication and computational overhead.

Our primary aim is to give a secure data aggregation which ensures the protection, validness and freshness of individual detected data and also the precision and secrecy of the accumulated data without displaying an immense overhead on the battery. First step is secure key distribution, in sensor network each node is assigned with common secret key (K), a node specific key (Ni) and a unique ID. The sink node has a secret key(K), a session key(Ks) and all the nodes (ID,K) pairs in the network. Here the sink node generates the session key for every session and sends to all the nodes in the network by using secret key (K). By receiving the message each sensor node decrypt and get the session key by using the secret key. Each sensor node have encryption key by XORing the session and node specific key. If a node sends a data to sink, it request the base station to get the session key and calculates the encryption key and encrypts the data. It avoids the transmission of secret key to all the nodes because each node has inbuilt secret key and therefore the security is increased.

For Power Efficient Privacy Preserving data aggregation consists of four steps, they are;

1. Construction of Aggregation Tree

The TAG protocol is mainly used to construct the aggregation tree. When base station sends request, the leaf node send its own data to its parent node for performing the aggregation . It send the partially aggregated result to upper node for futher processing.

2. Slicing

The leaf node only does the slicing operation. Each leaf node is sliced its data into m number of pieces. Then encryption is done on each slice by using encryption key generated by the node. One of the m encrypted slices is kept on the lead node itself and the rest m-1 encrypted is appended with id and send it to the neighbor nodes.

3. Mixing
When it receives all the slices, the node adds up all the encrypted slices by using the homomorphic technique. It does operation on the encrypted texts.It is used when nodes does not have the decryption key and also the intermediate node aggregate the encrypted data to their children even without decrypting their data.

4. Aggregation
Each leaf node sends the aggregated result and the encrypted slice to its parent node.After receiving the aggregated result from its child nodes,the intermediate node does the encryption on its data using its encryption key and sums up its aggregated result. This method solves the energy burden that is imposed by hop by hop encryption on the aggregation node.Thus it achieves the end to end confidentiality and in network aggregation without delay. For verifying the correctness of aggregated result without introducing overhead the integrity checking mechanism should be implemented.

C. Secure End to End Data Aggregation
In paper[9] one way to increase the network lifetime is to reduce the number of bits transmitted. It proposed a secure data aggregation scheme which provides end to end data privacy. It follows Homomorphic encryption which allows the aggregation on ciphertext. The encryption for the node I is Ci = Xi + Ki(Mod M) and decryption isXi = Ci – Ki (Mod M). Encryption is done by adding a key to the data value and decryption is by subtracting key from the aggregated value.In this protocol, we compute the cipher text for non responding node for its data to be 0 rather than sending the non responding information. The sink node that receives the cipher text is added with key of all the nodes. Sink node is obtained when subtracting keys of all sensor node aggregated data.Thus it will reduce the number of bits transmitted since it is not sending any information about the non responding nodes. The proposed protocol compute the average of the received data and also include the count of non responding nodes.

The nodes are organized into tree structure and tree levels are numbered from 0,1,2,…h where h is the height of the tree. The level 0 is the sink node.Children nodes at i+1 establish the secret keys with nodes at level i.Public key cryptosystem is used to establish the secret key.The secret key is used for encrypting the sensed data and cipher text is transmitted to the next higher level.Aggregators at level h-1 adds the responding nodes cipher text. For non responding nodes it calculates the cipher text its message value as 0 which is added to the aggregator cipher text. The message with aggregated data and the count of non responding nodes is send to h-2 level.At this level it adds all the received ciphertext and count values and send it to next level h-3.This process is repeated in all levels in the tree till the level 1. Finally at level 0 the message is decrypted by the sink node by subtracting all the keys from the cipher text and also computes the average value of the counter which provides the information regarding the non responding nodes. The proposed protocol ensures end to end data privacy with less number of bits transmitted and also ensures the best feature of end to end data privacy.

TABLE 1 : Privacy Secure methods

Sr. No	Methods	Advantages	Disadvantages
1	PDKP	Chances of Data Collison and Communication Overhead is less.	Less Packet Delivery Ratio.
2	PEPPDA	Increased Accuracy	No: of Operation is more and No integrity Checking mechanism
3	SEEDA	Less number of bits transmitted is less.	Significant Overhead.

V. CONCLUSION

A comprehensive survey on energy efficient data aggregation is done. Since most of the energy is consumed during transmitting and receiving the sensed data,data aggregation has become an important issue. By using different privacy preserve methods it is acheives authenticity,accuracy,data freshness,data confidentiality in data aggregation.

REFERENCES

[1] Taochun Wang, Xiaolin Qin, and Liang Liu, "An Energy-Efficient and Scalable Secure Data Aggregation for Wireless Sensor Networks", International Journal of Distributed SensorNetworks, Hindawi, 2013.

[2] Shirshu Varma, Uma Shanker Tiwary, "Data Aggregation in Cluster based Wireless Sensor Networks", Proceedings of the First International Conference on Intelligent Human Computer Interaction, Springer, 2009, pp 391-400.

[3] Priyanka B.Gaikwad, Manisha R.Dhage,"Survey on Secure Data Aggregation in Wireless Sensor Network",Internationa lConference on Computing Communication Control and Automation,2015.

[4] Raja Waseem Anwar, Majid Bakhtiari, Anazida Zainal, Abdul Hanan Abdullah and Kashif Naseer Qureshi,, "Security Issues and Attacks in Wireless Sensor Network", World Applied Sciences Journal, Volume 30, Issue 10, pp. 1224-1227,November 14.

[5] S. Madden et al., "TAG: a Tiny Aggregation Service for Adhoc Sensor Networks", OSDI 2002, Boston, MA,December 2015.

[6] Adrian Perrig, Robert Szewczyk, Victor Wen, David Culler, J. D. Tygar, "SPINS: Security Protocols for Sensor Networks," in ACM journal of Wireless Network.

[7] V.Akila , Dr T.Sheela , "Preserving Data and Key Privacy in Data Aggregation in Wireless Sensor Networks" Second International Conference on Computing and Communication Technology,2017.

[8] Joyce Jose, M Prince and Josna Jose, "PEPPDA: Power Efficient Privacy Preserving Data Aggregation for Wireless Sensor Networks" IEEE International Conference on Emerging Trends in Computing, Communication and Nanotechnology,2013.

[9] A.S.Poornima , B.B.Amberker. " SEEDA : Secure End-to End DataAggregation in Wireless SensorNetworks" IEEE Seventh International Conference on Wireless and Optical Communication Networks 2010.

[10] H.S.Annapurna, M.Siddappa, "Secure Data Aggregation with Fault Tolerance in Wireless Sensor Networks", IEEE International Conference on Emerging, Research in Electronics and Computer science Technology 2015.

[11] Riker, André, Eduardo Cerqueira, Marilia Curado, and Edmundo Monteiro, "A two-tier adaptive data aggregation approach for M2M group communication", IEEE Sensor Journal 16,no.3(2016).

[12] Kiran Maraiya, Kamal Kant, Nitin Gupta, "Wireless Sensor Network: A Review on Data Aggregation" ", International Journal of Scientific & Engineering Research, Volume 2, Issue 4, 2011.

[13] P.Raghu Vamsi,Krishna Kant,"Secure data aggregation and Instrution detection in Wireless Sensor Networks", IEEE International Conference on Signal Processing and Communication,2015.

[14] V.Vaidehi,R.Kayalvizhi,N.Chandra Sekar, "Secure Data Aggregation in Wireless Sensor Networks,"IEEE International Conference on Computing Sustainable Global Development,2015.

[15] Mohamed Ben Haj Frej, Khaled Elleithy, "Secure Data Aggregation Model (SDAM) in Wireless Sensor Networks," IEEE 14th International Conference on Machine Learning and Applications,2015.

