A Survey on Secure Data Aggregation Protocols in Wireless Sensor Networks

Josna Jose
Post Graduate Scholar
Dept. Information Technology,
Karunya University, India

Joyce Jose
Post Graduate Scholar
Dept. Information Technology,
Karunya University, India

Fijo Jose
Post Graduate Scholar
Dept. business administration
(Information Technology)
M G University, India

ABSTRACT

Wireless sensor network is a collection of large number of low cost resource constraint sensor nodes that are communicating using wireless medium. Sensor nodes are resource constrained in memory, sensing, communication capability, computational capability, battery power. Communication requires more power in sensor networks. One of the solutions to reduce number of bits transmitted is data aggregation. Data Aggregation is a process of aggregating data coming from different source using aggregation function to reduce redundancy in the Transmitted data. The aggregated results have great impact in accuracy and robustness of the final result get from the base station. Security is an important criterion to be considered because, wireless sensor nodes are deployed in a remote or hostile environment area that is prone to attacks easily. So data aggregation and security are essential for WSN. Many secure aggregations are proposed in wireless sensor network. But due to resource constrained nature, secure data aggregation also need new approaches. In this survey we are going to compare existing secure data aggregation protocol and their limitations and advantages.

General Terms

Data aggregation, Security, resources, WSN

Keywords

Wireless Sensor Network, Data Aggregation, Secure data aggregation

1. INTRODUCTION

Wireless Sensor Network is a collection of small sized, low cost sensor nodes, which is deployed in physical environment to gather sensing data and are communicating using wireless medium and finally it cooperatively send sensed data to base station for further processing. Sensors have many applications in military field surveillance, heath care, environmental monitor, accident report, law enforcement. Sensors are resource constrained in battery power, memory, computation, communication capabilities.

Due to the dense deployment of sensor of senor nodes in WSN, neighboring sensor nodes often have overlapping sensing ranges. Therefore it produces some similar data resulting in large volume of raw network data. Transmitting of large amount of redundant data increases the amount of data transmission and correspondingly increases amount of energy and bandwidth for the data transmission. So it is not efficient in an energy constrained wireless sensor network. One of the solutions to this is data aggregation. Data aggregation is a process of aggregating data coming from the different sources by using aggregation function (like min, max, average, sum etc.) and send the aggregated result to the other high level aggregated node. Data aggregation technique eliminate data redundancy in the transmitted data, decrease the amount of data transmission, Saves considerable energy and bandwidth, increases the robustness and the accuracy of data, increase the overall network life time.

The in-network processing is done on the aggregator node. But in hostile environment these aggregated result should be protected from the various type of attacks in order to achieve the data confidentiality, data integrity and source authentication. So security is necessary to be employed with data aggregation.

A corrupted sensor node in the hostile environment or remote area participated in the network operation and generates false sensor reading, generate forge messages, avoid legitimate messages from the aggregation result, improperly use aggregation function, and finally cause improper false aggregation result. But base station cannot understand compromised node presence from their behavior because attacker behave like in such a way that it’s incorrect result to be acceptable by base station. Hence for the secure data aggregation must need protection from eavesdroppers and prevention from intermediate nodes to access data. This makes the need of end to end encryption instead of hop by hop encryption. Various protocols have been proposed for the secure data aggregation in wireless sensor network.

Basically data aggregation protocols are divided into two types based on the topology used for aggregation. They are tree-based data aggregation protocols and cluster-based data aggregation protocols. Cluster based data aggregation reduces the latency in tree-based data aggregation by grouping the nodes. Group of nodes is called cluster. The grouping of nodes into cluster is called clustering. Cluster head perform data aggregation in cluster based data aggregation protocols whereas in tree-based data aggregation protocols, intermediate parent nodes in the path to the base station perform data aggregation.

2. SECURE DATA AGGREGATION

2.1 Primary security requirements

Achieving security in a wireless sensor network is a challenging task due to the hostile deployment and the resource constrained nature of sensor nodes. The security issues related to data aggregation are data confidentiality, availability, data Integrity, data authentication, data freshness etc [16]. These all are the security needs of wireless sensor network. Many secure aggregation protocols tried to achieve these requirements together.

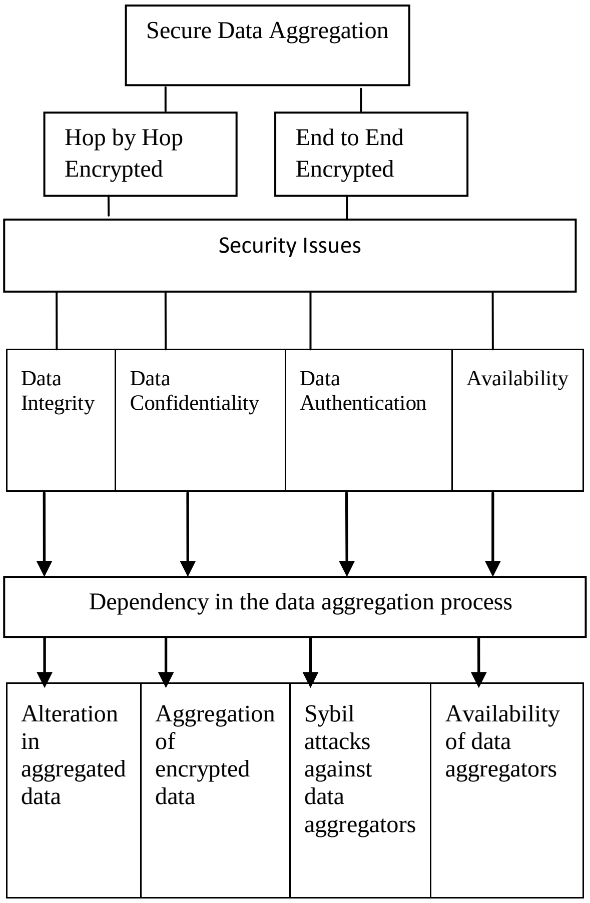
Secure Data Aggregation
Hop by Hop Encrypted
End to End Encrypted
Security Issues
Data Integrity
Data Confidentiality
Data Authentication
Availability
Dependency in the data aggregation process
Alteration in aggregated data
Aggregation of encrypted data
Sybil attacks against data aggregators
Availability of data aggregators

Fig 1: Interaction between wireless sensor network security and data aggregation process

2.2.1 Data Integrity

Ensuring that message has not been altered by malicious nodes.

2.2.2 Data Confidentiality

To keep the sensitive transmitted information from unauthorized entities. It can be achieved by aggregation of encrypted data.

2.2.3 Data Authentication

It is the verification of sender or receiver. It ensures that communicating node is the one that it claims to be.

2.2.4 Availability

It determines whether a node has the ability to use the resources and whether the network is available for the message to communicate.

2.2 Types of secure data aggregation

Hop-by-Hop data aggregation and end-to end data aggregation are two methods used for secure data aggregation [1] in wireless sensor networks.

2.2.1 Hop-by-Hop Encrypted Data aggregation:

In this, data is encrypted by the sensing nodes and decrypted by the aggregator nodes then aggregate the data and encrypt the aggregation result again. At last the sink node gets the final encrypted aggregation result and decrypts it. Aggregator nodes are vulnerable to attack because of decryption of the sensor data's in it.

2.2.2 End-to-End Encrypted Data Aggregation:

The aggregators aggregate the encrypted sensor readings without decrypting them. So intermediate aggregator node need not want to store secret information hence it provides end to end privacy between sensor nodes and sink.

Table 1: Comparison of data aggregation methods

Parameters	Hop-by-Hop Data Encryption	End-to-End Data Encryption
End to end privacy	No	Yes
Delay	Aggregation with delay	Aggregation without delay
Data Integrity	Provides Maximum Data Integrity	Provides Minimum Data Integrity
Aggregation performed on	Plain Sensor data	Encrypted Sensor data
Computational cost	Low	High
Memory requirement	High	Low
Vulnerable to attack	More to passive attack	More to Active attack
Energy Consumption	High	Low
Data Secrecy	Lesser security	High Security

2.3 Major wireless sensor network attacks

Two types of attacks in wireless sensor networks are passive and active attacks [2], [12], [13]. In passive attack, adversaries listens the transmitted packet and analyze the packet to obtain secret information. It does not make any modification in data. But in active attack, adversaries pick the data and modify the data. In active attack adversaries actively interfere the connection capturing packet.

2.3.1 Sinkhole attack

Sink is a high capability resource node. So attacker places himself in a network with high capability resources in order to confuse other nodes. As a result all data passed to attackers.

2.3.2 Sybil attack

In this attack, one node presents more than one identity in a network. It mostly affects routing mechanism.

2.3.3 Wormhole Attack

A wormhole is low latency link between two portions of a network over which attacker replays network message.

2.3.4 Hello flood Attack

In this attack, attacker broadcasts HELLO packets with high transmission power to sender or receiver. The node receiving the message assumes that the sender node is nearest to them and send packet by this node. By this attack congestion occur in network.

2.3.5 Selective Forwarding

In this attack, malicious nodes may refuse to forward certain messages and drop them, ensuring that they are not propagated any other.

2.3.6 Passive Information Gathering

An attacker with powerful resources (eg: powerful receiver well designed antenna) can pick off the data stream .strong encryption is the one of the solution to prevent from this attack.

2.3.7 Node Subversion

Capturing of one node cause the reveal of the secret information and it may cause the compromise of the whole sensor network.

3. SECURE DATA AGGREGATION PROTOCOLS

3.1 Hop by Hop Encrypted data aggregation protocols

3.1.1 ESPDA Protocol

ESPDA [3] is based on pattern codes which represents the characteristics of sensor data to perform data aggregation. In this protocol sensor node first generate and transmit the code pattern code for sensing values instead of sending sensor data. Then the cluster head identify the distinct pattern codes and request to the generator node of the pattern code to send actual data. Thereby reduces the amount of data transmitted. For providing security, the mapping interval of pattern code to the sensor data is refreshed periodically. So this method gives importance to energy, bandwidth efficiency, security.

3.1.2 SRDA Protocol

In SRDA [4], sensors send differential sensing data instead of raw sensed data by comparing raw data sensed by sensor to the reference data. So it reduces the number of bits transmitted from sensor node to cluster head. So it improves energy consumption. To increase security levels by going from lower level to higher level, SRDA uses one algorithm with security margin as adjustable parameter. Security is calculated based on number of hops from the base station. First step is the transmission of raw sensed packet in a session to cluster head by a node (leaf/cluster head) reporting to higher level cluster head. Then cluster head create reference entry for that node. Sensor node sends differential data to cluster head for subsequent readings. Finally when the session ends for a sensor node, cluster head removes the reference entry for the node from the cluster head. This method is independent of clustering scheme so this method can be applied on any level. When the reference value is greater than differential value, then the efficiency of the scheme will increase.

3.1.3 SDAP Protocol

SDAP [5] design is based on the divide-and-conquer and commit-and-attest protocol. This general purpose data aggregation protocol consists of three steps. In the first step, aggregation tree is constructed and thereby all nodes identified their parents. Then query is disseminated through the tree. In second, SDAP dynamically partition nodes in a tree into multiple group based on a novel probabilistic grouping technique which is depends on group leader selection. Then generate a group aggregates by a commitment based hop by hop aggregation. Finally in the verification and attestation step, base station identifies the suspicious group based on group aggregate and then the suspect group participates in the verification process to prove the correctness of the group aggregate. Verification includes verification of aggregation message and content of the packet. Advantages of this protocol include applicability on multiple aggregation functions, adjustable detection rate and this protocol achieve integrity, data confidentiality, Source authentication. But energy utilization and transmission overhead is high.

3.2 End to End encrypted data aggregation protocols

Most of the end to end encrypted data aggregation protocols are based concealed data aggregation. This technique provide end to end confidentiality and in network aggregation together by the application of privacy homomorphism that allow data aggregation on cipher text. Cryptographic methods that provide privacy homomorphism property are divided into two types. symmetric and asymmetric privacy homomorphism. In symmetric privacy homomorphism, symmetric key is shared among sensor node and base station. So the base station can only decrypt the data encrypted by sensor node hence end to end confidentiality is maintained. In asymmetric key based privacy homomorphism, each sensor node uses the public key of the base station to encrypt data and perform aggregation of encrypted data. The base station own the private key can only decrypt this aggregated data .So end to end confidentiality is maintained. In end to end encrypted data aggregation, data aggregators need not want to store sensitive keys.

3.2.1 CDA

Concealed data aggregation (CDA) [6],[12] is based on the symmetric additive privacy homomorphism proposed by Domingo-Ferrer [14]. In this approach, every sensor node shares a same key with the base station. So it does not guarantee privacy of individually sensed data from other sensor nodes. Because one compromised sensor leads to the decryption of every sensor data. In this approach ,each sensor node splits its data into ‘d’ parts ($d \geq 2$) and encrypt them by using common key shared with the base station and send to aggregator .Aggregator aggregate the encrypted sensor data with other sensors encrypted data because of privacy homomorphism property and finally send the aggregated result to the sink. At the sink, aggregated data is decrypted using the same key used for the encryption. Disadvantages of this technique are vulnerability to reply attack and malicious aggregation, size grow, and efficiency and also this technique do not address the problem of non-response ID.

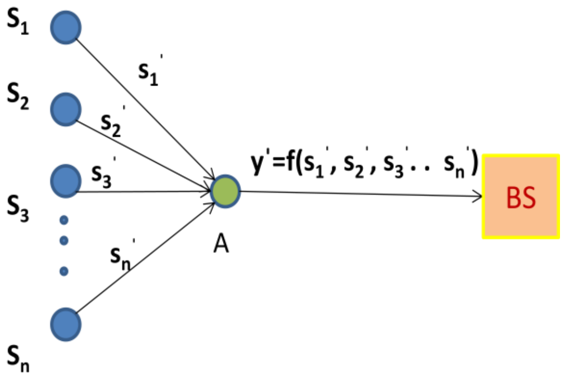

Fig 2: Operation of CDA

3.2.2 CMT-a key stream based PH

Efficient aggregation of encrypted data [7] uses symmetric additive privacy homomorphism but each sensor node shares different symmetric key per plain text with the base station. So it resist the replay attack and preserve cipher text of small size. It is a extension of one-time pad encryption with modular addition instead of Exclusive-OR operation of stream cipher. In this technique encryption is done by adding a plain text with current key modulo the length of key space. and perform aggregation of encrypted values. The decryption requires the same key used for encryption. For this sink first needed to know who all are contributed to the aggregation result. This ID problem is solved by transmission of node IDs participated in the encryption process to the sink. So it reduces the life time of WSN depending on the size of WSN and cause additional communication cost for the transmission of all participated nodes IDs and corresponding security parameters .CDA with multiple symmetric key prevent the attacking of one node to giving the secrets of other nodes. So it protects the privacy against other sensor nodes. This technique is not scalable because base station must know the keys of all aggregated packets in order to decrypt received aggregation result. This AH scheme does not protect integrity because it is vulnerable to malicious aggregation of data by adding natural numbers to the cipher text.

3.2.3 CDAP Protocol

Concealed data aggregation using privacy homomorphism [8] is based on asymmetric privacy homomorphism. So data are encrypted using public key of base station. In this concealed data aggregation, first AGGNODEs are given the public key of the base station and then network is deployed. AGGNODES are special sensor nodes that have more computational power, memory space, and battery power. Next, AGGNODE established a shared key with its neighboring sensor nodes using any random key distribution protocol. Each AGGNODE query its neighboring nodes for sensor readings in data collection phase. Then the neighboring node encrypts its data using symmetric key encryption algorithm (RC_5) along with shared key and send it to AGGNODE.Due to the symmetric key algorithm, compromised AGGNODE may reveal the secrecy to neighboring nodes. But it is local. The AGGNODE decrypts the all received data and perform aggregation. Then this aggregated data are encrypted using public key of the base station and forwarded to the sink and AGGNODEs hierarchically aggregate the encrypted data .The data are concealed from nodes in the path to the sink because of the privacy homomorphism encryption. At the sink, data are decrypted using private key of base station. Due to privacy homomorphism, CDAP have more computational overhead than hop by hop. So AGGNODESs are only allowed to encrypt and aggregate data. If the number of AGGNODEs is more, the data transmission efficiency and aggregation ability is more than hop by hop scheme.

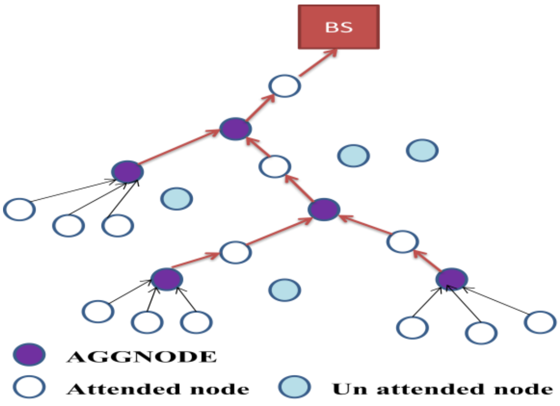

Fig 3: Operation of CDAP protocol

3.2.5 HCDA Protocol

Hierarchical concealed data aggregation for wireless sensor network [9] is based on the elliptic curve cryptography. So it is resistant to node compromise attack. This protocol provides concealed data aggregation by aggregating data of multiple sensor node group that use different public keys to encrypt their data. This is achieved by the help of group based network deployment scheme. In this, sensor nodes are divided into several groups before deployment and each group is deployed from certain location over the network so that each group covers a part of the network. Then assign a public key to each group so that the BS is able to determine and classify the data of the group based on the public key used to encrypt the data. This helps the base station to identify data of a particular region or group of the network.

3.2.6 DSSS based digital watermarking approach

In [10] end to end authentication is achieved by using digital water marking where sink can directly validate the data from the sensor node. This technique overcome the limitations of existing privacy homomorphism that provide secure in-networking process on specific aggregation functions (sum, avg etc.) by performing aggregation on cipher text. The idea of this approach is that the authentication related information is modulated as watermark and superposed on sensory data of sensor node and passed to the sink node. The intermediate nodes on the way to sink node aggregate the watermarked data without any enroots checking. When the data reached at the sink node, it authenticates the received data by validating the watermark in the data and find out data has been altered by any compromised nodes on the way to sink node. Digital watermarking is performed by visualizing the sensory data collected from whole network into images by taking snapshot at certain time in which each sensor node is visualized as pixel and it’s reading as pixel’s intensity. To balance the energy consumption, direct spread spectrum based watermarking is employed in which a sensor node embed part of the whole watermark into its sensory data. Upon reception of aggregated and watermarked data at sink, it verify the watermark and hence authenticity of the sensor data. It is a one way authentication approach from sensor node to sink.

Table 2: Comparison of different data aggregation methods based security requirements

Protocol	Source authentication	Data Confidentiality	Data Integrity
ESPDA [3]	Yes	Yes	Yes
SRDA [4]	Yes	Yes	Yes
SDAP [5]	Yes	Yes	Yes
CDA [6]	No	Yes	No
CMT [7]	No	Yes	No
CDAP [8]	No	Yes	No
HCDA [9]	No	Yes	No
DSSS-DWA [10]	No	No	Yes

4. CONCLUSION

This paper provides detailed overview of secure data aggregation protocols and we more focused on end to end encrypted data aggregation scheme that is based on privacy homomorphism. Because privacy homomorphism based data aggregation have more attention recently and it allow aggregation on encrypted data. In this end to end encrypted data aggregation, no need of keeping secret keys in aggregator nodes and also performing decryption in aggregator nodes. So it reduces the chance of attacker to get secret information. The most secure data aggregation scheme till now is concealed data aggregation scheme based on privacy homomorphism. It provide in network processing and end to end security together. But only one of the paper achieved integrity in concealed data aggregation [15] techniques. It also have some disadvantages. So integrity preserving data aggregation has research scope. One other research area is digital water-marking technique [10] that provides two way authentications.

5. REFERENCES

[1] Y.Sang and H.Shen, Secure Data Aggregation in Wireless Sensor Networks:ASurvey.

[2] Dr.G.Padmavathi, Mrs.D.Shanmugapriya, 2009, A Survey of Attacks, Security Mechanisms and Challenges in Wireless Sensor Networks, (IJCSIS) International Journal of Computer Science and Information Security.

[3] H.Cam, S.Ozdemir, P.Nair, and D. Muthuavinashiappan, ESPDA: energy-efficient and secure pattern-based data aggregation for wireless sensor networks, IEEE Sensors– The Second IEEE Conference on Sensors, Oct. 22-24, 2003, Toronto, Canada.

[4] 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, Sept. 2004.

[5] Y. Yang, X. Wang, S. Zhu, and G. Cao, SDAP: A Secure Hop-by- Hop Data Aggregation Protocol for Sensor Networks, ACM Trans. Information and System Security (TISSEC), vol. 11, no. 4, pp. 1-43, 2008.

[6] D. Westhoff, J. Girao, and M. Acharya, Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks:Encryption, Key Distribution, and Routing Adaptation, IEEE Trans. Mobile Computing, vol. 5, no. 10, pp. 1417-1431, Oct. 2006.

[7] C. Castelluccia, E. Mykletun, and G. Tsudik, Efficient Aggregation of Encrypted Data in Wireless Sensor Networks, Proc. Second Ann. Int'l Conf. Mobile and Ubiquitous Systems, pp. 109-117, July 2005.

[8] S. Ozdemir, Concealed Data Aggregation in Heterogeneous Sensor Networks Using Privacy Homomorphism, Proc. IEEE Int'l Conf. Pervasive Services, pp. 165-168, July 2007.

[9] S.Ozdemir and Y.Xiao, Hierarchical Concealed Data Aggregation for Wireless Sensor Networks.

[10] W. Zhang, Y. Liu, S.K. Das, P. De, Secure data aggregation in wireless sensor networks: a watermark based authentication supportive approach, Elsevier Pervasive Mobile Comput. 4 (2008) 658–680.

[11] J. Girao, D. Westhoff, and M. Schneider. CDA: Concealed Data Aggregation in Wireless Sensor Networks. In Proc. 40th International Conference on Communiacations, *IEEE ICC '05*, Korea, May 2005.

[12] Y.E.Aslan and E.Kayaaslan, Security in wireless sensor network, JOURNAL OF CS514 CLASS FILES, VOL.1, NO.1, JANUVARY 2008.

[13] A.Pandey and R.C Tripathi, A Survey on Wireless Sensor Networks Security, International Journal of Computer Applicationsc (0975-8887), Volume 3-No.2, June 2010.

[14] S.Peter and K.Piotrowski, On Concealed Data Aggregation for Wireless Sensor Networks.

[15] C-M Chen,Y-H Lin, Y-C Lin and H-M Sun, RCDA:Recoverable Concealed data aggregation for Data Integtrity in Wireless Sensor Networks, IEEE TRANSACTION ON PARALLEL AND DISTRIBUTED SYSTEMs,VOL.23,NO.4,APRIL 2012.

[16] S.Ozdemir and Y.Xiao, Secure data aggregation in wireless sensor networks: A Comprehensive overview, Computer Networks 53 , 2022-2037,2009.