SDAF: A Secure Data Aggregation Framework for Wireless Sensor Networks

M. Sardaraz, M. Tahir, and Ataul Aziz Ikram

***Abstract*—Wireless sensor networks (WSNs) are constrained in terms of memory, computation, communication, and energy. To reduce communication overhead and energy expenditure in (WSNs), data aggregation is used. Data aggregation is a very important technique, but it gives extra opportunity to the adversary to attack the network, inject false messages into the network and trick the base station to accept false aggregation results. This paper presents a secure data aggregation framework (SDAF) for (WSNs). The goal of the framework is to ensure data integrity and data confidentiality. SDAF uses two types of keys. Base station shares a unique key with each sensor node that is used for integrity and the aggregator shares a unique key with each sensor node (within that cluster) that is used for data confidentiality. Sensor nodes calculate a message authentication code (MAC) of the sensed data using shared key with base station, which verifies the MAC for message integrity. Sensor nodes encrypt the sensed data using shared key with aggregator, which ensures data confidentiality. Proposed framework has low communication overhead as the redundant packets are dropped at the aggregators.**

***Index Terms*—Security, wireless sensor networks, data aggregation.**

I. Introduction

Wireless sensor networks (WSNs) gained popularity because of the fact that they can be used to solve the real world challenges with low cost [1]. WSNs are used in a variety of applications such as habitat monitoring [2] and target tracking [3] etc but these networks are constrained in terms of resources such as memory, communication, computation, and energy. WSNs consist of a large number of low power and low energy sensing devices called nodes. In addition to these nodes, there may be one or more powerful devices called base stations. Base station controls the network and processes the data collected by the sensor nodes. The sensor nodes sense and collect data from the environment and send it to the base station which performs further query on the collected data. The sensor nodes may be deployed in vicinity to each other as the number of nodes in the network may be very large. Due to this vicinity sensors may collect and transmit redundant data. Since the transmission of data costs much high than the computation, it is usually advantageous to organize the sensor nodes in clusters. The data are processed locally within the network and the aggregated data are sent to the base station in cluster environment. In such an environment some nodes which are called the aggregators collect the data from its neighbor nodes, process it and send the result to the base station. This technique reduces communication distance and thus energy consumption is reduced as compared to directly communication with the base station. This scenario is referred to as data aggregation in literature.

Many protocols [4]-[7] have been proposed for secure data aggregation in WSNs to reduce the communication overhead and the energy expenditure. Generally the network is divided in to a tree topology that is rooted at the base station. The sensor nodes sense data from the environment, the aggregators aggregate the data from the sensor nodes and send the data to the base station. Base station performs further query on the data.

Data aggregation reduces communication overhead significantly but it makes the security more difficult. Any compromise node can forge data or can inject false data in to the network and thus one compromised node can alter the final aggregation [8]. In general the data aggregation reduces the communication overhead but it opens new doors to the adversary and the aggregated data can easily be attacked by the adversary.

This paper presents a secure data aggregation framework (SDAF) for WSNs. This framework is applicable in cluster based networks and can provide security to the collected data. The main focus is on integrity and confidentiality of collected data. The framework uses two types of keys. One key is shared between the base station and the sensor nodes which provide integrity as the base station can verify the MAC calculated by the sensor nodes. Base station uses separate keys for each sensor node and the aggregator as well. If a node is compromised the adversary cannot obtain the keys of other nodes. The other key is shared between the aggregators and sensor nodes in that cluster. This key is used to encrypt and decrypt the data by the nodes and the aggregators to provide data confidentiality. During the tree generation the size of each cluster is kept small that an aggregator can store the keys of each sensor node in a particular cluster. The paper is organized as follows. Related work is presented in Section II. Network assumption and key setup are presented in Section III. Cluster formation and aggregator selection are discussed in Section IV. Notations and attack model are discussed in Section V and VI respectively. Proposed framework is presented in section VII. Security analysis is presented in Section VIII and the paper is concluded in Section IX.

Manuscript received September 5, 2012; revised January 13, 2013.

The authors are with the Department of Computing and Technology Iqra University Islamabad, Pakistan, (e-mail: sardaraz@hotmail.com, tahir591@hotmail.com, ata@iqraisb.edu.pk).

II. Related Work

Previous work in data aggregation has focused on how to

perform aggregation assuming a trust in the sensor nodes that each sensor node is honest. A few works have focused on secure data aggregation.

The problem of data aggregation has been studied in [8]. The authors have studied data aggregation when one node is compromised. It has been assumed that only leaf nodes sense data while the intermediate nodes only aggregate data. This protocol may be vulnerable if two consecutive nodes are compromised.

A mathematical framework which is based on evaluating security of several resilient aggregation techniques has been proposed in [9]. The paper uses mathematical measures that how much it will cause damage if an adversary compromises a number nodes in the network .The author assumed that raw data has been reached to the base station and that is why it is not really a data aggregation framework.

A secure hop-by-hop data aggregation protocol (SDAP) has been proposed in [7]. This approach is based on two principals, divide-and-conquer and commit-and-attest. SDAP uses divide-and-conquer principal to reduce damage caused by a compromised aggregator at high level. The other principal commit-and- attest is used by the base station to verify the correctness of the data. This scheme can tolerate more than one compromised nodes. However this scheme needs to send much data to ensure reasonable level of security.

Another scheme for data aggregation is ESPDA -Energy-Efficient and Secure Pattern Based Data Aggregation- for wireless sensor networks [10]. In ESPDA, cluster-head first requests sensor nodes to send the corresponding pattern code for the sensed data. If multiple sensor nodes send the same pattern code to the cluster-head, then only one of them is permitted to send the data to the cluster-head. Hence, ESPDA has advantages over the conventional data aggregation techniques with respect to energy, bandwidth efficiency and security. However no practical solution is available for compromised nodes.

III. Network Assumptions and Key Setup

A network consisting of a base station and a large number of sensor nodes is assumed. Some of the nodes are referred to as aggregators. The aggregators are same as normal nodes in terms of resources such as memory and computation etc. The aggregators are selected randomly after some time. The base station is a powerful device with sufficient computational capabilities, processing power and energy. The base station has also capable of storing large information of a dense network. In addition the based station is trusted entity and has security mechanisms and cannot be compromised. The current generation sensor nodes with limited resources, such as MICA motes [11] are assumed. The sensor nodes are deployed in an environment in which they are prone to any attack.

It is assumed that each sensor node shares a unique cryptographic key with the base station and the aggregator. Aggregators also share secret keys with the base station.

For key setup, the proposed algorithm uses an algorithm presented in [12]; any algorithm that can securely communicate keys between aggregators and sensor nodes in a cluster can be used to share keys between aggregators and sensor nodes.

IV. Cluster Formation and Aggregator Selection

For cluster formation and aggregator selection any good and relevant algorithm can be used. For cluster formation an algorithm presented in [13] is used. Algorithm in [7] is used for aggregator selection.

V. Notations Used

Following is the list of notation used in the rest of the paper.

S refers to Base Station

U, V, W, X represent normal nodes

A refers to the aggregator

K S, V refers to the shared key of base station and node V

K A, V refers to the shared key of Aggregator and node V

RV refers to the reading of node V

M1|M2 represents the concatenation of two messages M1 and M2

E (K, M) refers to the encryption of message M with key K

MAC (K, M) Represents the MAC of message m with key K

ID V Represents the unique identifier of node V

U→ V node U sends a message to node V over a single hop.

U→→ V node U sends a message to node V over multiple hops.

VI. Attack Model

A network setting with an attacker who can access the network and can compromise sensor nodes is assumed. The attacker can inject false messages in to the network or can alter the messages sent by legitimate sensor nodes. However, it is assumed that the attacker can compromise a small number of nodes in the network.

WSNs may collect sensitive data. The data received by the base station may provide basis for crucial decision. So false data may change the overall result, consequently, the overall decision will be affected. Proposed framework focuses on the integrity and confidentiality of data. Other types of attacks in which the purpose of the attacker is to disrupt the normal operation of the network are not considered. For instance, a compromised node can be used by the attacker to attack on routing protocol and thus it causes denial of service attack. The goal of the proposed framework is to defend against the attacks in which the attacker changes the final aggregation and trick the base station to accept false aggregation results. The attacker may inject false message in to the network or alter the messages sent by legitimate nodes. Proposed framework also provides confidentiality to the sensed data, the sensor nodes encrypt the sensed data during the aggregation process.

VII. Secure Data Aggregation Framework

This section presents proposed secure data aggregation

framework. The aggregation process begins at the leaf nodes. As the leaf nodes do not aggregate data, it only sends its IDs and the sensed data to its parent node. The parent node forwards the data in the encrypted format as sent by the child node to its own parent, if it is not the aggregator. The leaf node sends a MAC of the sensed data which is calculated by using the shared key with the base station. This MAC is later used by the base station to verify the contents of the data. The leaf node sends the following message to its parent node.

U→ V: IDU, 0, E (K U, A, AggU)

AggU= (RU|IDU|MACU)

MACU= MAC (K U, S, IDU|RU)

In Fig. 1 the leaf node U sends message to its parent node V. The message is encrypted with the key which is shared between node U and aggregator. The message also includes a MAC. MAC is calculated using shared key between node U and base station, base station uses this MAC to verify the integrity of the data. The flag bit is included in the message to indicate whether further aggregation is needed or not. 0 value means that the message needs further aggregation and value 1 means that no further aggregation is needed.

When a node receives the message sent by its child node, it checks the flag bit. If it is 0, it appends its own message to the message sent by its child node (if this node is not an aggregator) and forwards it to its parent leaving flag set to 0. In "Fig.1" node V is an intermediate node but not an aggregator. Intermediate node also updates its routing table i.e. the ID of the child node (sender of the message) is added. This can be used to trace the path for verification. The intermediate nodes also send MAC with the message in order to ensure the integrity of the data by the base station by verifying the MAC. The intermediate node sends the following message to its parent.

V→ W: IDV, 0, E (K V, A, AggV)

AggV= (RV|IDV|MACV| E (K U, A, AggU))

MACV= MAC (K V, S, IDV|RV)

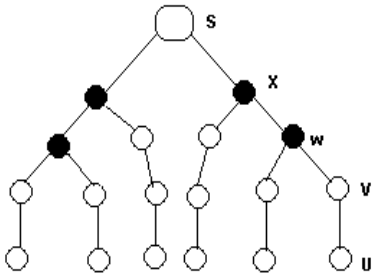

Fig. 1. An example of aggregation tree: The nodes filled with black color represent aggregators and S refers to the base station.

The intermediate node V sends its own message encrypted with the shared key between node V and the aggregator to its parent W and also the encrypted message sent by its child node. The message sent by the child node is sent without decryption as the decryption key is only known to the aggregator. It also sends its ID and the MAC of the message. MAC is calculated by using the shared key between node V and base station.

When a node that is an aggregator receives a message from its child node, it decrypts the messages encrypted by the child nodes. All the messages are decrypted using shared key with each sensor node. It compares the data values, if there are redundant messages the aggregator drops them. The aggregator updates its routing table, encrypts its own message and the messages send by the child nodes, sets the flag bit to 1 and forward it to its parent node. If the parent node itself is not the base station, it updates its routing table and forwards the message to its parent until the message reaches the base station. This message contains a flag bit set to 1 which indicates that no further aggregation is needed. In Fig.1 node X performs no further aggregation on the message send by its child node W. The aggregator sends the following message to the base station.

W→→ S: IDW, 1, E (K W, S, AggW)

AggW= (RW|IDW|MACW| AggU |AggV))

MACW= MAC (K W, S, IDW|RW)

All messages of the child nodes and the aggregator are encrypted with the key shared between the aggregator and the base station. The aggregator also calculates a MAC of its own reading by using the shared key with the base station.

When the base station receives the messages from the aggregators it uses the shared key with the aggregators to decrypt the data. To verify the integrity of the data, the base station calculates the MAC of the data for each node with the shared key of the respective nodes. If the MACs calculated by the base station and the nodes are the same, the base station accepts the data; otherwise the data are treated as corrupted.

VIII. Security Analysis

In this section security analysis of the proposed framework is presented. This approach uses two keys, one key is shared between base station and sensor nodes; whereas the other is shared between the aggregator and sensor nodes in that cluster. The shared key between the base station and the sensor nodes is used to ensure data integrity. The sensor nodes calculate MAC of the sensed data using shared key with the base station. Base station can verify data integrity by using the MAC calculated by the sensor nodes.

The other key which is shared between the aggregator and the sensor nodes is used for data confidentiality. Sensor nodes encrypt their sensed data using this key, which is decrypted only at the aggregator, the intermediate nodes cannot decrypt it and thus the chances for attacker to access the data are reduced. If the attacker compromises a single node it will not help him/her to find out keys of other nodes because each node has a unique key shared with the base station as well as the aggregator. The aggregation tree is reconstructed randomly after some time and the aggregators are selected randomly, this mitigate the attack in which a compromised node changes its aggregated value, corresponding to real abnormal event to abnormal value, thus, the base station may not notice the real event.

IX. Conclusion

In this paper a secure data aggregation framework for wireless sensor networks (WSNs) is presented. The use of separate keys makes the framework much secure. If a node is compromised the adversary cannot obtain the keys of other

nodes. The base station can verify the message contents. The data confidentiality is also provided. In future the approach can be enhanced by using the public key cryptography to achieve a high level of security.

REFERENCES

[1] F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cavirci, "A survey on sensor networks," *IEEE communication Magazine*, IEEE Computer Society, pp. 102-114, 2002.

[2] A. Mainwaring, J. Polastre, R. Szewczyk, D. Culler, and J. Anderson, "Wireless sensor networks for habitat monitoring," in *Proc. of the 1st ACM International Workshop on Wireless Sensor Networks and Applications*, pp. 88–97, 2002.

[3] T. He, P. Vicaire, T. Yan, L. Luo, L. Gu, G. Zhou, R. Stoleru, Q. Cao, A. Stankovic, and T. Abdelzaher, "Achieving real-time target tracking using wireless sensor networks," in *Proc of the 12th IEEE Real Time Technology and Applications Symposium*, IEEE Computer Society, pp. 37–48, 2006.

[4] C. Intanagonwiwat, D. Estrin, R. Govindan, and J. Heidemann, "Impact of network density on data aggregation in wireless sensor networks," in *Proc of the 22nd IEEE International Conference on Distributed Computing Systems*, IEEE Computer Society, 457–458, 2002.

[5] B. Krishnamachari, D. Estrin, and S. Wicker, "The impact of data aggregation in wireless sensor networks," in *Proc. of the 22nd IEEE International Conference on Distributed Computing Systems*, IEEE Computer Society, pp. 575–578, 2002.

[6] S. Madden, J. Franklin, M. Hellerstein, and W. Hong, "TAG: A tiny aggregation service for ad-hoc sensor networks," in *Proc. of the 5th ACM Symposium on Operating Systems Design and Implementation*, pp. 131-146, 2002.

[7] Y. Yang, X. Wang, S. Zhu, and G. Cao, "SDAP: A secure hop-by-hop data aggregation protocol for sensor networks," in *Proc. of the 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pp. 356–367, 2002.

[8] L. Hu and D. Evans, "Secure aggregation for wireless network," in *Proc. of the 2003 IEEE Symposium on Applications and the Internet Workshops*, IEEE Computer Society, pp. 384–394, 2003.

[9] D. Wagner, "Resilient aggregation in sensor networks," in *Proc. of the 2nd ACM workshop on Security of ad hoc and sensor networks*, pp. 78–87, 2004.

[10] H. Cam, S. Ozdemir, P. Nair, and D. Muthuavinashiappan, "ESPDA: Energy efficient and secure pattern based data aggregation for wireless sensor networks," in *Proc. of the 2nd IEEE International Conference on Sensors*, IEEE Computer Society, pp. 732- 736, 2003.

[11] Mica Motes. [Online]. Available: http://www.xbow.com.

[12] M. Haque, A. Saqib, and S. Hong, "An asymmetric key based security architecture for wireless sensor networks," *KSII transaction on internet and information system*, vol. 5, 2008, pp. 265-279.

[13] W. He, X. Liu, H. Nguyen, K. Nahrstedt, and T. Abdelzaher, "PDA: Privacy-preserving data aggregation in wireless sensor networks," in *Proc. of the 26th IEEE International Conference on computer communications*, IEEE Computer Society, pp. 2045-2053, 2007.

M. Sardaraz received his Master degree in Computer Science from Foundation University Islamabad. He is currently pursuing Ph.D. in Computer Science from Department of Computing & Technology, Iqra University. His research interests are WSNs, cluster and grid computing, and bioinformatics.

M. Tahir is Ph.D. (Computer Science) student at Department of Computing & Technology, Iqra University. His research interests are in parallel and distributed computing, Hadoop Mapreduce framework, bioinformatics algorithms design and analysis, Sequence alignment.

Ataul Aziz Ikram completed his B.S. in Electrical & Electronics Engineering from Middle East Technical University, Turkey in 1997. He worked as an Electrical Engineer with Bayindir Inc. for a couple of years. He completed his M.S. in Electrical Engineering in 2003, M.Phil. in 2005, and Ph.D. in 2007 from Graduate Center of City University of New York. He joined the academia in 2007. His Ph.D. research was in the area of Nanotechnology. Presently he is employed as Associate Professor at Iqra University, where he is also Head of Department of Electrical Engineering. He is also supervising undergraduate, graduate and postgraduate students in different areas of Computer Science and Electrical Engineering. His research interest spans various disciplines of electrical engineering, material science, and computing.