A New Approach to Secure Data Aggregation protocol for Wireless Sensor Network

Mukesh Kumar Jha [1]
Computer Science & Engineering Department
NIT Hamirpur (HP), India

T. P. Sharma[2]
Computer Science & Engineering Department
NIT Hamirpur (HP), India

***Abstract*— In this paper, we have proposed a secure data aggregation protocol for wireless sensor networks (WSNs) that is robust to deceitful nodes. The goal of this protocol is to guarantee the essential security needs (like source authentication, data confidentiality & data integrity) as well as to achieve low communication overhead and be fitted with various aggregation functions (like sum, average, max, min etc.). To achieve these security needs, it uses symmetric encryption and message authentication code (MAC). Encryption ensures data confidentiality while message authentication code ensures authentication and data integrity. An anomaly detection algorithm is used to detect the anomaly or outliers and thus prevent the deceitful- corrupted data from being contributed to the final aggregated results. Simulation results show that our protocol enhances the security of the aggregated data considerably in WSNs.**

***Keywords*— Wireless Sensor Network, Data aggregation, security**

I. INTRODUCTION

A wireless sensor network (WSN) is a collection of a large number of sensor nodes that have limited computation, communication and power resources. Due to the limited resources, the amount of data transmission should be minimized such that the lifetime of the sensor nodes and bandwidth utilization of the network can be improved. Due to this, the concept of data aggregation has come into the picture. Data aggregation is the process of combining the data coming from various sources and enroute them after removing redundancy such as to improve the overall network lifetime [1]. The in-network processing is done on the aggregator node. The aggregator node aggregate the data received from its child node as per the required aggregation function (like min, max, average, sum etc.) and send the aggregated result to the other high level aggregated node or sink. But in hostile environment these aggregated result should be protected from the various type of attacks in order to achieve data confidentiality, data integrity and source authentication. So security is necessary to be employed with data aggregation.

Recently various data aggregation protocols [2-11] have been proposed to remove the redundancy in the transmitted data so as to decrease to the amount of data transmission which saves a considerable amount of energy and bandwidth. But, these protocols do not provide the security means to the aggregated data. In many situations, it is necessary to protect the aggregated data from various types of attacks. In this paper, we have proposed a secure data aggregation protocol that achieves the security requirements of the aggregated data.

II. RELATED WORK

In Secure DAV[12], cluster key establishment (CKE) protocol is used to establish the secret cluster keys in the WSN. These secret cluster keys are used for the partial signature generation on the aggregated data. Elliptic Curve Cryptography (ECC) is used for the secure key management because it has smaller key size and faster computation. After that, a Secure DAV protocol is used which guarantees that the sink does not accept the altered data for an upper bound of t compromised sensor within a cluster where $t < n/2$ where n is the number of nodes in the cluster. Custer-head computes the average on the sensors data within the cluster and sends this average to all the sensors. Sensor nodes then compare this average with its own data and if the difference between these two is less than a threshold then it generates the partial signature using shared secret key and sends it to the cluster-head. Cluster-head then generates the full signature after combining partial signatures from all the sensors within the cluster and then sends this full signature along with the average reading to the sink. Sink having possession of public key then verifies this signature. Merkle Hash tree is used to check the integrity of the sensor node's readings. Secure DAV can be applied only to average aggregation function and have a high communication overhead.

In [13], the author presents a mechanism to find out the misbehaving nodes. In this protocol, sensed data is not aggregated at the immediate next hope rather it is aggregated on the second hop. This protocol guaranties data integrity and source authentication but it does not provide data confidentiality.

In [14], cryptographic operation is required only when any cheating activity is detected. A secure aggregation tree (SAT) is built with the topological constraint for the detection and prevention for cheating. The SAT is built in such a way that the child is able to listen all the incoming data from its sibling to its father so that the child node can observe the behaviour of its father, then the cheating activity of any non-leaf (aggregator) node can be detected. If the aggregated result from an aggregator is uncertain then a weighted voting scheme is introduced for taking the final decision about whether the aggregator node is cheating. If cheater

aggregator node is found then a local recovery scheme is employed which rebuild the SAT such that the cheater node is removed from the tree. It does not provide data confidentiality.

In SELDA [15], to develop trustworthiness for environments and neighbouring nodes, action of the neighbouring nodes are observed by the sensor nodes. Aggregators consider sensor node's reading received using the web of trust to enhance the reliability of aggregated data. If any aggregator is under the denial-of- service attack, then it can be detected using the monitoring mechanism. It ensures data integrity and source authentication but it does not provide data confidentiality.

III. System Model

A. Assumptions

Here we make following assumptions:

- WSN consisting of a large number of resource constraint sensor nodes.
- There exists a powerful fixed base station (BS).
- The clusters are static i.e. are formed at the start of the network.
- Cluster heads (CHs) work as an aggregator.
- All sensor nodes are immobile.

B. Network Model

Figure 1 shows the network model used. Various symbols and terms used are shown in Table I. All sensor nodes are immobile. Links between two sensor nodes is considered bidirectional. There is only single channel for communication between sensor nodes.

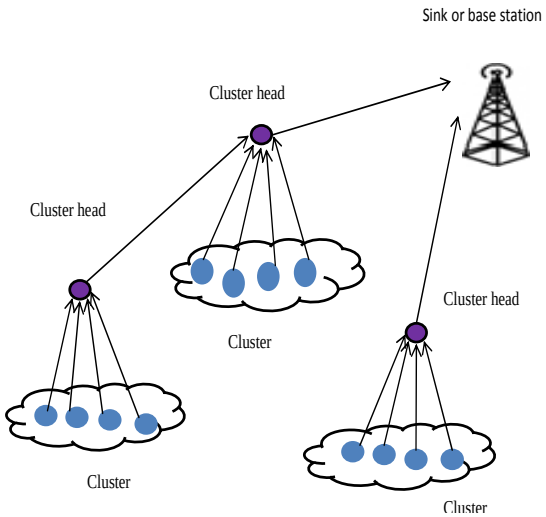

Figure 1. Network model

TABLE I
Notations used in this Paper

r	Random number
u	Regular sensor nodes
R_u	Reading of sensor node u
CH	Cluster head
BS	Base station
E	Encryption
$K_{\{BS,u\}enc}$	Shared pairwise key between base station and sensor node u used for encryption
f_{ag}	Aggregation function
u_{id}	Id of the sensor node u
MAC	Message Authentication Code
$K_{\{BS,u\}MAC}$	Key used to calculate MAC by the base station
CH_c	Child cluster head
CH_p	Parent cluster head
Agr	Aggregated reading
CH_{id}	Id of cluster head
$\|CH_i\|$	Number of sensor nodes in cluster i
OD	Outlier detection
q	Query
|	Concatenation

IV. Proposed Protocol

Base station (BS) starts the sensing process by sending a broadcast message to those sensor nodes which are located in the area of interest. For authenticated broadcast of query, we have used µTESLA [16] with some modifications. Sensor nodes then report back with their readings to the BS through aggregator. Aggregator then processes the received readings of sensors. In addition to aggregation process, it also identifies the anomaly or outlier sensor nodes by using anomaly detection algorithm [17]. It then reports back to the next level aggregator or BS with aggregated reading, outlier count & outlier sensor's ids.

A. Query Dissemination

In process of query dissemination from BS to the network, sensor should have the knowledge about aggregation function which is used for the aggregation of sensor's readings. Every sensor nodes have their distinct private key shared with the BS which is computed by taking hash on the master key of BS (K_B) with their respective ids. In addition to this, each sensor node shares pairwise key with their children which is used for encryption. The format of query packet sent by BS to the aggregator looks as follows:

BS $\longrightarrow$ u: E ($K_{\{BS,u\}enc}$, f_{ag} |q |r | BS) | MAC ($K_{\{BS,u\}MAC}$, f_{ag} | q | r | BS)

B. Transmission of sensor nodes reading to the BS

Transmission of sensor nodes readings to the base station can be done in three phase: Sensor node to cluster head, child cluster head to parent cluster head, and cluster head to base station.

- Sensor node to cluster head
  Sensor nodes send their readings to their cluster head. The packet sent by the sensor nodes to the cluster head includes ids of the sensor nodes, readings, random number. The packet format transmitted by sensor node to cluster head is like:

u → CH: E ($K_{\{u,CH\}enc}$, R_u | r | u_{id})| MAC($K_{\{u\}MAC}$, R_u|r | u_{id})

- Child cluster head to parent cluster head
  Upon the reception of readings from its cluster members, cluster head performs anomaly detection algorithm. Thus, it finds the outlier and drops the readings of outliers. Cluster head then aggregates the readings and sends the aggregated reading along with outlier ids, outlier count to the parent cluster head. The packet format sent by the child cluster head to the parent cluster head is like:

  CH_c → CH_p : E($K_{\{CH_c,CH_p\}enc}$,Agr|r | CH_{cid} |outlier count| outlier ids)|MAC($K_{\{CH_c\}MAC}$, Agr| r| CH_{cid} |outlier count | outlier ids)

- Cluster head to base station

  When cluster head receives readings from all of its children, it first runs anomaly detection algorithm to filter out the anomaly or outlier readings. After that it aggregates the readings of its child nodes according to the specified aggregation function in query packet and then it finally sends the aggregated readings with outlier ids and count to the base station. The packet format sent by cluster head to base station is like:

CH → BS: E($K_{\{CH,BS\}enc}$,Agr|r | CH_{id} | outlier count| outlier ids)| MAC($K_{\{CH\}MAC}$, Agr | r | CH_{id} | outlier count| outlier ids)

Proposed Algorithm
{
// Sensor nodes send their encrypted readings and MAC to the CH (CH works as an aggregator) //
for j=1 to | CH_i |
{
Send {E ($K_{\{u_j,CH_i\}enc}$, R_j | r |u_j)| MAC ($K_{\{u_j\}MAC}$, R_u|r |u_j)};
}
// cluster head runs anomaly detection algorithm and find outlier readings, outlier ids & outlier count and then filters out the outlier readings.
Set |CH_i | = | CH_i |- outlier count;

// cluster head aggregates readings & sends aggregated reading to BS//
for j=1 to | CH_i |
if (f_{ag} = = Average)
Set avg_i = avg_i + (R_j/ | CH_i |);
Send {E ($K_{\{CH_i,BS\}enc}$, avg_i | r |CH_i| outlier ids| outlier count)| MAC ($K_{\{CH_i\}MAC}$,avg_i | r |CH_i| outlier ids| outlier count)};
Else if (f_{ag} = = sum)
Set sum_i = 0;
for each sensor j in cluster i do
sum_i = sum_i + (R_j) ;
Send {E ($K_{\{CH_i,BS\}enc}$, sum_i | r |CH_i| outlier ids| outlier count)| MAC ($K_{\{CH_i\}MAC}$,sum_i | r |CH_i| outlier ids| outlier count)};
Else if (f_{ag} = = minimum)
list[j]= rand();
min_ value = list[1];
for(j=1;j<= | CH_i |, j++)
if(min_value>list [j])
then min_value= list[j];
Send {E ($K_{\{CH_i,BS\}enc}$, min _$value_i$ | r |CH_i| outlier ids| outlier count)| MAC ($K_{\{A\}MAC}$,min _$value_i$ | r |CH_i| outlier ids| outlier count)};
else if (f_{ag} = = maximum)
list[j]= rand();
max_ value = list[1];
for(j=1;j<= | CH_i |, j++)
if(max_value<list [j])
then max_value= list[j];
Send {E ($K_{\{CH_i,BS\}enc}$, max _$value_i$ | r |CH_i| outlier ids| outlier count)| MAC ($K_{\{CH_i\}MAC}$,max _$value_i$ | r |CH_i| outlier ids| outlier count)};

else if (f_{ag} = = median)
for(j=0; j<=| CH_i |; j++)
for(n=j+1; n<=| CH_i |; n++)
{
if(a[j]>a[n])
{
temp=a[n];
a[n]=a[j];
a[j]=temp;
}
}
if(| CH_i | %2= =0)
median= (a[| CH_i | /2]+a[(| CH_i | /2) + 1])/2;
else
median= a[| CH_i | +1 /2];
}
Send {E ($K_{\{CH_i,BS\}enc}$, $median_i$ | r |CH_i| outlier ids| outlier count)| MAC ($K_{\{CH_i\}MAC}$,$median_i$ | r |CH_i| outlier ids| outlier count)};

Now BS performs anomaly detection algorithm to filter out outliers and then aggregate received readings.

A. Simulation Environment

In this section, we evaluate the performance of our proposed secure data aggregation protocol on OMNET-4.0 [18] simulator. In order to check the performance of our proposed protocol, we take following metrics:

- Outlier success rate: It is defined as the rate at which outliers successfully detected in the wireless sensor network.
- *Accuracy improving rate*: This metric gives the rate of improving the accuracy in presence of deceitful sensor nodes.
- *Average of total transmission energy consumed per node*: It shows the average of total energy consumption at each node due to transmission of packet in the presence of outliers in wireless sensor network.
- *Communication Overhead:* Communication overhead is represented in the form of number of packets transmitted in the network.

B. Simulation Results and discussion

To find out more reliable and accurate results, we executed our proposed protocol with different number of deceitful nodes.

a. Outlier successful detection rate

It can be defined as the number of detected spiteful node over the actual number of spiteful nodes in the WSN. Figure 1 show that the successful outlier detection rate at standard deviation (SD) = 9.97 and up to SD<5. From Figure 1, it is clear that our proposed approach performs better than SDAP in successful detection of outliers at different SD (at SD=9.97, SD<5).

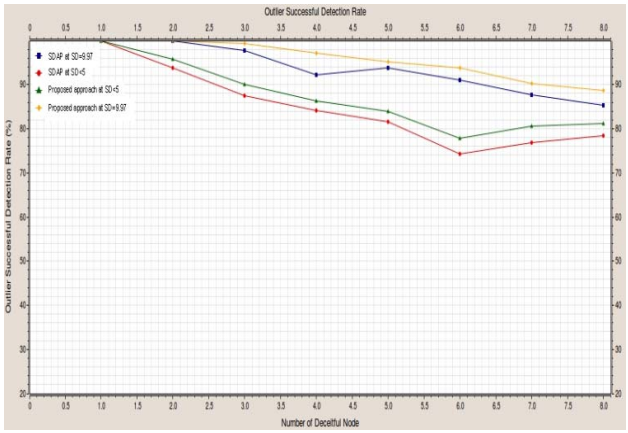

Figure1. Outlier successful detection rate vs. number of deceitful sensor nodes

b. Accuracy improving rate

Accuracy improving rate is used to measure about with how much efficiency the anomaly detection algorithm performs. The standard deviation in the presence of spiteful sensor nodes is larger than the standard deviation in the absence of spiteful sensor nodes (after filtration of spiteful sensor nodes). Hence, accuracy improving rate can be calculated by dividing the difference of SD in the presence and absence of spiteful sensor nodes by the SD in absence of spiteful sensor nodes. From Figure 2, it is clear that, our proposed approach performs better than SDAP in terms of accuracy improving rate at different SD. From Figure 2 it is also clear that as number of spiteful sensor nodes increases, the accuracy improvement rate of proposed approach also enhances.

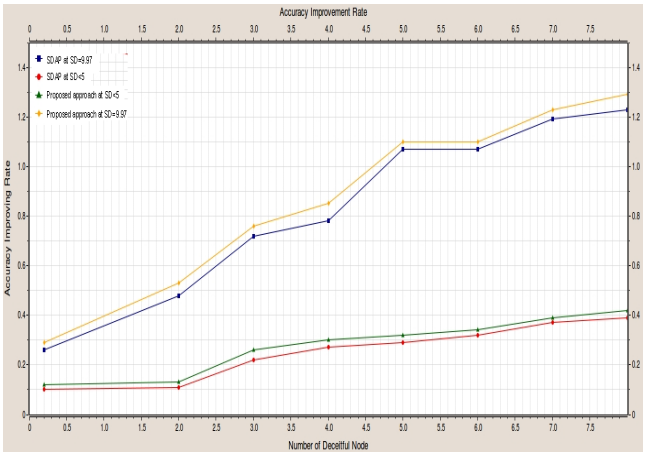

Figure 2. Accuracy improving rate vs. number of deceitful sensor nodes

c. Average of total transmission energy consumed per node

Figure 3 shows that as number of testified outliers increases, average of total transmission energy consumed per node also increases. The reason behind the increment in average of total transmission energy consumption per node with increment of number of outliers is that, as number of testified outlier sensor nodes increase, outlier ids & outlier counts have to include in the packet. Thus packet size increases which consumes more transmission energy.

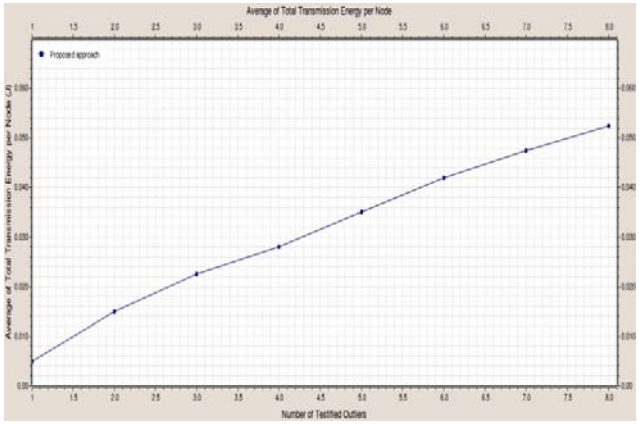

Figure 3. Average total transmission energy consumed per node versus number of testified outliers

d. Average of total transmission energy consumed per node

Figure 4 shows the communication overhead (in terms of number of messages) of our proposed approach.From Figure 4 it is clear that the performance of proposed approach degrades in terms of communication overhead with the increase in number of nodes. The reason behind this performance degradation is that as number of sensor nodes increase, the number of messages generated also increase.

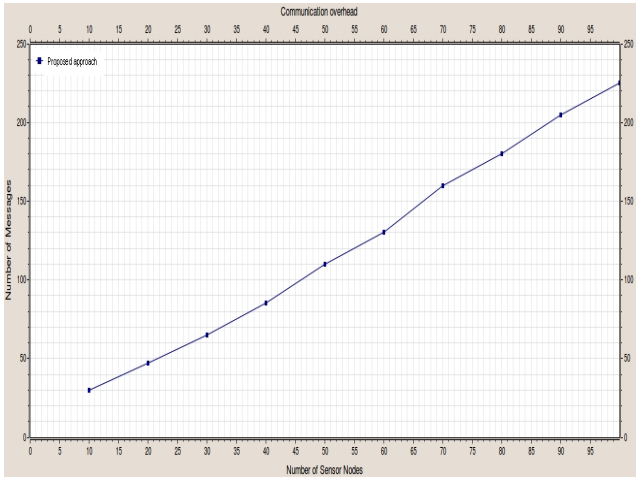

Figure 4. Communication overhead of proposed protocol

VI. CONCLUSION

Our proposed approach is based on detection and filtration deceitful sensor nodes with their sensed readings in wireless sensor networks. It uses outlier detection algorithm to detect and filter out the outlier sensor nodes. It provides high outlier detection rate due to the use of distributed approach. It uses MAC for data authentication and data integrity. In order to provide confidentiality, it uses symmetric encryption. It uses pairwise shared key for the purpose of encryption. Simulation results show, our proposed approach achieves high outlier detection rate, accuracy improvement rate and average of total transmission energy consumed per node.

REFERENCES

[1] Suat Ozdemir, Yang Xiao, "Secure data aggregation in wireless sensor networks: A comprehensive overview", Computer Networks, vol. 53, Elsevier, pp. 2022–2037, 2009

[2] C. Intanagonwiwat, R. Govindan, D.Estrin, J.Heidemann,F.silva, "Directed diffusion for wireless sensor networking," in: proceeding of IEEE/ACM Transactions on Networking, Vol.11, 2003, pp. 2-16.

[3] B. Krishnamachari, D. Estrin, S. Wicker, The impact of data aggregation in wireless sensor networks, in: Proceedings of the 22nd International Conference on Distributed Computing Systems Workshops, 2002, pp. 575–578.

[4] R. Cristescu, B. Beferull-Lozano, M. Vetterli, On network correlated data gathering, in: Proceedings of the 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 4, 2004, pp. 2571–2582.

[5] B. Zhou et al., A Hierarchical Scheme for Data Aggregation in Sensor Network, IEEE ICON 04, Singapore, 2004.

[6] S. Lindsey, C. Raghavendra, K.M. Sivalingam, Data gathering algorithms in sensor networks using energy metrics, IEEE Trans.Parallel Distrib. Sys. 13 (9) (2002) 924–935.

[7] O. Younis, S. Fahmy, HEED: a hybrid, energy-efficient distributed clustering approach for ad hoc sensor networks, IEEE Trans. Mobile Comput. 3 (4) (2004) 366–379.

[8] Y. Yao, J. Gehrke, The Cougar approach to in-network query processing in sensor networks, ACM SIGMOD Rec. 31 (3) (2002) 9–18.

[9] S. Chatterjea, P. Havinga, A dynamic data aggregation scheme for wireless sensor networks, in: Proceedings of the Program for Research on Integrated Systems and Circuits, Veldhoven, The Netherlands, 2003.

[10] P. Popovski et al., MAC-Layer Approach for Cluster-Based Aggregation in Sensor Networks, IEEE IWWAN 04, Oulu, Finland, 2004.

[11] S. Pattem, B. Krishnamachari, R. Govindan, The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks, ACM/IEEE IPSN04, Berkeley, CA, 2004.

[12] A. Mahimkar, T.S. Rappaport, "Secure DAV: a secure data aggregation and verification protocol for wireless sensor networks", In Proceedings of the 47th IEEE Global Telecommunications Conference (Globecom), November 29–December 3, Dallas, TX, 2004.

[13] L. Hu, D. Evans, Secure aggregation for wireless networks, in: Proceedings of the Workshop on Security and Assurance in Ad Hoc Networks, Orlando, FL, 28 January 2003.

[14] K. Wu, D. Dreef, B. Sun, Y. Xiao, Secure data aggregation without persistent cryptographic operations in wireless sensor networks, Ad Hoc Networks 5 (1) (2007) 100–111.

[15] S. Ozdemir, Secure and reliable data aggregation for wireless sensor networks, in: H. Ichikawa et al. (Eds.), LNCS 4836, 2007, pp. 102–109.

[16] A. Perrig, R. Szwezyk, A. Woo, S. Hollar, D.Culler, and J. Tygar, "SPINS: security protocols for sensor networks," in mobile computing and networking, 2001, pp.189-199.

[17] S.Rajasegarar, C. Leckie, M. Palaniswami, James C. Bezdek, "Distributed anomaly detection in wireless sensor network," in proceeding of IEEE conference on mobile computing and networking, USA, 2006.

[18] http://www.omnetpp.org

[19] Y. Yang, X. Wang, S. Zhu, G. Cao, "SDAP: a secure hop-by-hop data aggregation protocol for sensor networks", In Proceedings of the ACM MOBIHOC'06, 2006.