Cluster Based Secure Data Aggregation Using Simulated Annealing Approach

Hevin Rajesh. D, Paramasivan. B

Abstract— **Proposed scheme used a soft computing approach simulated annealing to deal with security attacks and prevent direct access of data by a compromised cluster member or cluster head during data forwarding in WSNs. Secured cluster key distribution technique used in proposed scheme improves the security of data aggregation in WSNs. It is found that the simulation results are consistent with theoretical analysis. The objective of using simulated annealing in the proposed scheme is to generate energy efficient data aggregation. Also the key refreshing phase minimizes the security risk caused by a compromised node. Since each cluster member only has a share of the secret cluster key and the cluster key is hidden from each member.**

***Index Terms*—Aggregation, Cluster, Neural Network, Sensor.**

I. INTRODUCTION

In most sensor networks, many sensors are deployed in the geographic area. It is mainly for fault tolerance and reliability. When an event takes place, the sensing node observes the event and sends the reporting message [1].The advanced sensor network over the traditional method of monitoring. Sensor network deployment provides more economically for long term study. A set of system design requirements cover hardware of the node, sensor networks. It has the capability of remote data access and management [2].The advanced sensor nodes are extremely small and low powered. It is having programming capability [3]. To minimize the bit transmission between sensors nodes need to find an efficient encryption algorithm, which is essential for increase the battery life time [4]. Energy efficiency has been known as most important in WSN operation. Because limited energy is provided with small sensor nodes. By reducing the data redundancy can improve the energy efficiency [5]. Data aggregation is commonly used technique in wireless sensor network. When the sensor nodes are deployed in the hostile environment data confidentiality and integrity are important [6]. Data communication between nodes consume more energy [7].

II. PROPOSED SCHEME

In cluster based data aggregation protocols, cluster heads aggregates data locally and transmit the aggregation result to the base station through multi-hop forwarding or directly. The proposed Simulated Annealing based Data Aggregation (SADA) protocol is organized into different phases.

i) Cluster Organization phase
ii) Cluster Key Distribution Phase
iii) Data Aggregation Phase
iv) Key Refreshing Phase

.

A. Cluster Organization Phase

The proposed scheme reduces energy consumption during communication because in this scheme sensor nodes are grouped into number of clusters. Here, clustering is based on the distance measure. The main feature of this scheme is its simplicity in operation and its speed in computation. In this clustering approach, location information of each deployed sensor node is required to group the related sensor nodes based on the region where they are present. Location information is determined by the coordinates (x_i, y_i) of each sensor node. Clustering or grouping is performed based on the minimum distance between Cluster Head (CH) and sensor node. The steps followed by the proposed scheme in clustering are described below. Let $S = \{s_1, s_2, s_3 \ldots s_n\}$ be the sensor nodes deployed in an application specific region. The procedure followed for clustering approach is as follows.

Step 1: BS arbitrarily chooses the desired number of clusters; let it be assigned as 'k'.

Step 2 : BS selects 'k' sensor nodes randomly among 'n' deployed sensor nodes to function as cluster centers.

Step 3 : Assign the remaining nodes to their closest cluster centre.

Step 4 : Construct the cluster in such a way that each cluster should contain minimum number of nodes.

The procedure followed for determining a suitable cluster head is described below.

Step 1: Re-compute the CH using centroid method: (i) Find the sum of all x co-ordinate of sensor nodes closer to the cluster centre and divide by number of sensor nodes. (ii) Find the sum of all y co-ordinates of sensor nodes closer to the cluster centre and divide by number of sensor nodes. Then determine the centroid point (C_x, C_y).

Step 2 : Determine the sensor nodes very close to the centroid point.

Step 3 : From the determined nodes select the one which has maximum residual energy as the CH.

This algorithm does not require any other specific metrics except the location information to organize the sensor nodes into clusters and is computationally faster than other clustering methods. Thus a WSN is organized into hierarchical clusters based on the region where CH and group of nearby sensors are present.

***Hevin Rajesh. D**, Information Technology, St. Xavier's Catholic College of Engineering, ChunkanKadai, India.,*
***Paramasivan. B**, Computer Science & Engineering, Kovilpatti,India..*

ISSN: 2278 – 1323

B. Cluster Key Distribution Phase

Once the network has been organized into clusters and suitable CH for each cluster is determined, information about CH should be notified to cluster members. To represent its role as CH to its members, CH broadcast control information representing its responsibility as CH together with its identity (ID). When a cluster node (CN) receives the message from its CH, it should send a reply message representing its identity to the CH. The reply message includes the acknowledgement for the request received along with its ID. CH generates a secret Cluster Key (CK). Each cluster member only has a share of the secret cluster key and the cluster key is hidden from each member. The public key for the secret cluster key is known to all members within the cluster as well as the base station. Since the secret cluster key is hidden from all members, attacks on the cluster key are not possible. This phase includes the following steps.

Step 1: Each node will compute its private and public key pair (K_{public},$K_{private}$) and broadcast its public key to all nodes within the cluster. This key pair is different from the share of the cluster key and is used for secure communication with any other node.

Step 2: The CH chooses a randomly generated cluster key CK and forms a congruence system as follows:

$$X \equiv a_1 (\text{mod } K_1)$$

$$X \equiv a_2 (\text{mod } K_2)$$

$$:$$

$$X \equiv a_n (\text{mod } K_n) \qquad (1)$$

Where $a_i = CK \oplus K_{private}$. CH solves the system of congruence equations to obtain the value of X.

Step 3: Encrypt X to generate the partial signature Y using the following equation

$$Y = (\text{ones complement of } X) \oplus K_{public}$$

(6.2)

Step 4: The cluster-head accumulates all partial signatures from the cluster members, combines them to form a full signature and sends the full signature along with the aggregate reading to the base station. The base station who has the public key, can then verify the signature.

C. Data Aggregation Phase

Each sensor within a cluster will have its share of the secret. This share is used to generate partial signatures on the aggregate readings. In this phase the Simulated Annealing (SA) soft computing approach is used to perform data aggregation.

D. Key Refreshing Phase

The cluster key is used for ensuring the confidentiality of the cluster communication. If the same cluster key is used continuously for a long period it may result in cryptanalytic attack. Moreover, if an adversary has captured the key he would be able to break the security in communication of the whole cluster. In order to strengthen the security of cluster communication, it is essential to renew the key occasionally. Cluster key can be renewed when a new member is included or a member is removed from the cluster. Also the proposed scheme ensures forward and backward secrecy when a node enters into the cluster and at the time a node leaves from the cluster.

Add new nodes to the cluster: When a new node wants to enter into the network, it needs to report to the BS its physical location. The BS identifies the cluster nearest to the location of the new node and sends the new node the cluster information to which it should join. New node then sends the request to join the cluster along with its ID. After receiving the join request, the CH initiates the rekeying process to compute the new cluster key to enhance key refresh ness and to maintain forward secrecy.

III. SIMULATION AND RESULTS

This work is implemented over NS-2.32, the network simulator. The main objective of the simulation is to evaluate the performance of the proposed scheme with and without the presence of compromised nodes in WSNs. The parameters used in simulation are tabulated in Table 1.

Table 1.Parameter settings for simulation

Parameter	**value**
Number of nodes	100:500
Area of deployment	100x100 m2
Simulation time	10:150s
Initial energy of each node	2 joules
Transmission Range	2m
Initial temperature	20000
Base station location	(90,90)
Number of compromised nodes assigned to each cluster	5:30
Underlying MAC protocol	IEEE 802.11
Channel	Wireless
Propagation	TwoWay Ground
Network type	WirelessPhy
Queue	DropTail/ PriQueue
Antenna	OmniAntenna

The protocols CLUDDA: Clustered Diffusion with Dynamic Data Aggregation [8], LEO: Simple Least Time Energy Efficient Routing Protocol with One-Level Data Aggregation [9] and GMDA: Grey Model based Data Aggregation [10] are taken for comparison to evaluate the performance of the proposed scheme.

The performance of SADA was studied with respect to network lifetime, average energy consumption of a node, average packet delivery ratio, latency and filtering efficiency. Simulation results show that proposed scheme achieves maximum efficiency when compared with other related schemes.

Network lifetime: Network lifetime is defined in the literature in different ways. In this work, the network lifetime is considered as the time until when all the nodes in the network die out of their energy. Network lifetime depends on the average energy spent. The greater is the energy spent, the lesser is the network lifetime.

Latency: Latency is defined as the delay involved in data transmission, routing, and data aggregation. It can be measured as the time delay between the data packets received at the sink and the data generated at the source nodes.

Filtering efficiency: It is defined as the percentage of false data to be filtered out within a specified number of hops. Filtering efficiency can be based on the authentication information owned by the forwarding node to detect and drop forged MAC.

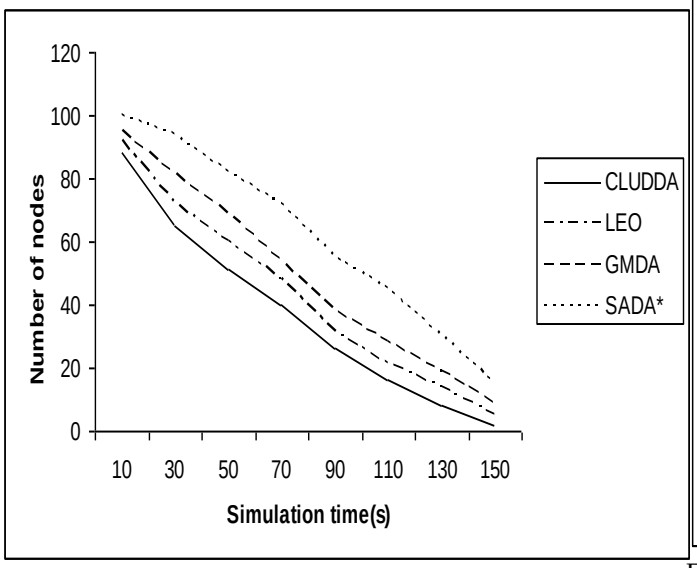

Figure 1 Network Lifetime Comparison

Figure 1 shows the network lifetime for all the four protocols. Some observations are in order. First, for the simulation time less than 50s can lead to significant improvement on the lifetime of the LEO compared with that of CLUDDA. But compared with the SADA the lifetime of the LEO is not good. Second, the number of nodes alive in the SADA is larger after the simulation time of 70s. Finally, the SADA protocol achieves 13-25% higher node alive ratio for every round compared to that of other protocols. Because the SADA optimizes the energy consumption during data aggregation and of a cluster head, the time period for which the node acts as a cluster head is pre-determined.

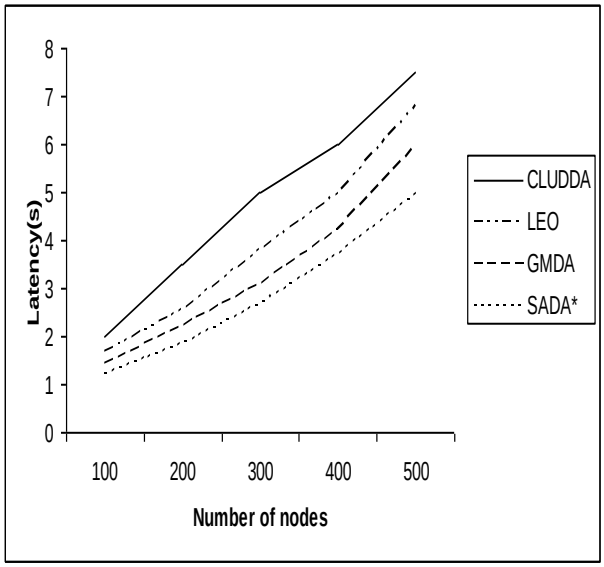

Figure 2 Comparison of Latency

Figure 2 shows the comparison of the latency or end-to-end delay of all the protocols. The proposed SADA protocol has a lower delay compared to that of the other protocols. The protocol selects the cluster heads that have the capability to send the aggregated data to the base station directly. The maximum delay is 5 seconds for a network size of 500 nodes, which is 30% lower than that of CLUDDA. Due to the combination of clustering and diffusion mechanisms, the CLUDDA has a higher delay than that of the LEO and GMDA protocols.

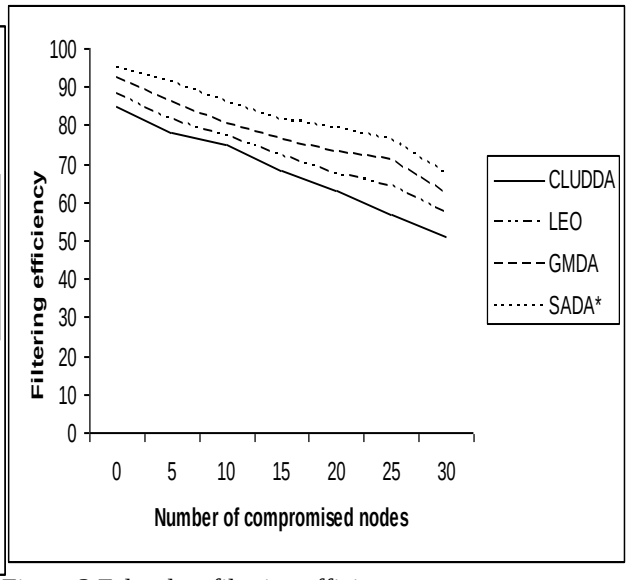

Figure 3 False data filtering efficiency

Figure 3 illustrates the filtering efficiency of SADA and related schemes CLUDDA, LEO and GMDA which can filter false data injected by compromised cluster members. SADA is tested by allowing compromised en-route cluster node with in the cluster to inject false event report. This is tested in the presence of 5, 10, 15, 20, 25, 30 compromised nodes. Simulation results show that on average SADA filters 80% of false data in the presence of 15 compromised nodes. In the presence of 20 compromised nodes, CLUDDA filters 63% of false data, LEO filters 72% of false data and CMDA filters 76% of false data. In CLUDDA filtering is based on the authenticated information maintained by the cluster heads. But in SADA, false data injected through a compromised cluster head will be detected immediately by its cluster head.

ISSN: 2278 – 1323

In this scheme each cluster head which receives the event report has the required authentication information to verify n the report. If the verification succeeds, then the report is considered for aggregation. Otherwise it is detected as a false report and the report is dropped. The false report is thus filtered by the cluster head as early as possible. Hence, it is proved that filtering efficiency of the proposed scheme is higher than related schemes.

IV. CONCLUSION

When compared with other related schemes, proposed scheme has a lot of benefits. The benefits of proposed scheme are dropping of false data as early as possible and the computation and communication overhead involved for authentication is low compared to other related schemes. The achievability of the proposed scheme is evaluated through performance analysis and simulation results. The results show that proposed scheme outperforms well compared to the existing schemes in terms of network lifetime, node energy consumption, packet delivery ratio, latency and filtering efficiency. Hence, proposed scheme can achieve enhanced security during data aggregation in a vulnerable environment. This scheme is suitable for sensitive data transmission for densely populated large scale WSNs.

REFERENCES

[1] Kyle Jamieson, Hari Balakrishnan, Y. C. Tay, "Sift: A MAC Protocol for Event-Driven Wireless Sensor Networks", Wireless Sensor Networks, Lecture Notes in Computer Science, Volume 3868, pp 260-275, 2006

[2] Alan Mainwaring et al, "Wireless Sensor Networks for Habitat Monitoring", Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, pp. 88-97,2002.

[3] Arati Manjeshwar and Dharma P. Agrawal, "TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks", Proceedings 15th International Parallel and Distributed Processing Symposium, pp. 2009-2015, 2000.

[4] Claude Castellucia, "Efficient aggregation of encrypted data in wireless sensor networks", The Second Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services, pp. 109-117, 2005.

[5] Huifang Chen et al, "Adaptive data aggregation scheme in clustered wireless sensor networks", Computer Communications, vol. 31, pp. 3579–3585, 2008.

[6] Yingpeng Sang, "Secure Data Aggregation in Wireless Sensor Networks: A Survey", Seventh International Conference on Parallel and Distributed Computing, Applications and Technologies, pp. 315-320, 2006.

[7] Hani Alzaid, "Secure Data Aggregation in Wireless Sensor Network: a survey", Proceedings of the sixth Australasian conference on Information security, pp. 93-105, 2008

[8] Website:http://webhost.laas.fr/TSF/cabernet/cabernet/ workshops/radicals/2003/papers/ Cabernet 2003_-_Chatterjea ,_Havi.pdf

[9] Sudip Misra et al, "A simple, least-time, and energy-efficient routing protocol with one-level data aggregation for wireless sensor networks", Journal of Systems and Software, Volume 83, Issue 5, pp. 852–860, 2010.

[10] Guiyi Wei et al," Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter", Computer Communications, vol.. 34, pp. 793–802, 2011.

