Data Aggregation Framework for Clustered Sensor Networks Using Multi Layer Perceptron Neural Network

Hevin Rajesh. D, Paramasivan. B

Abstract—**In the design of secure data aggregation scheme which uses multi layer perceptron neural network for aggregation. The whole network is treated as a complex neuron system where each sensor node works as an underlying neuron and cluster head is like a central neuron. In this proposed scheme an aggregate contains not only a data value computed for the required aggregation function but also a count value indicating the number of sensor nodes involved in the aggregation operation. This aggregation function is embedded in the second hidden layer, the third hidden layer, and output layer neurons. Then, the CHs transfer the aggregated result to the base station. Finally, the base station performs the base station aggregation. After the base station has received the aggregation messages from each cluster head, it verifies the authenticity of the aggregated value in each aggregation message. This includes verifying the content of the data packet and the authenticity of each cluster head. The proposed method gives good result.**

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

I. Introduction

Sensor network consists of many nodes and they are deployed closely. The positions of the nodes are not pre planned. These sensor nodes are helpful for disaster relief operations. There are many different types of sensors are available. Sensors nodes are used for many applications [1]. Sensor networks have hundreds of sensor nodes. It can able to communicate external base station. More number of sensor node cover more geographic area and sense the environment for reading with good accuracy [2]. The sensor nodes are grouped in to various clusters. The coordinator of the cluster is cluster head. The other nodes are called member. The life time of the sensor node can be improved by clustering [3]. Sensor nodes are very significant in military and civil applications. Sensor nodes are constrained in power supply and bandwidth [4]. To save the energy the data collected by destination as aggregation [5]. The sensor nodes are suffered from restricted power and computation [6]. Sensor nodes are deployed in hostile environment, so it must be secure [7].

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

II. Proposed Scheme

The proposed Perceptron Neural Network based Data Aggregation (PNNDA) protocol is organized into three phases namely

i) Setup Phase
ii) Data Aggregation Phase
iii) Verification Phase

.

A. Setup Phase

First, PNNDA uses a novel clustering technique to partition the nodes into various clusters. In this phase, the cluster head selection and cluster formation algorithm are introduced.

The cluster head election is based on the fitness function. The fitness function gives an index to all the individual nodes on the network. The fitness is mainly designed to minimize the energy consumption to extend the lifetime of the network. The node with highest fitness value will be elected as a cluster head. The proposed Fitness Function is defined as follows.

$$fit_i = \frac{1}{E_i} + \frac{1}{NC_i} \qquad (1)$$

Here Ei is the amount of energy in the node i and NCi is the ith node centrality. The Energy of a sensor node is defined as

$$E = n * E_R + (n+1) * E_{agg} + E_T \qquad (2)$$

Where n is the number of node, is the receiving energy, Eagg is the aggregation energy and ET is the transmission energy from cluster head to sink. For a cluster head, to receive 'k' bit message from one member node, the receiving energy is used and it is formulated as

$$E_R = kE_{elec} \qquad (3)$$

Here k is the message size in bits and Eelec is the Energy consumed by the electronic circuit to transmit or receive the signal. Transmitting energy is the Energy consumed for transmitting k-bit message from cluster head to sink. Based on the threshold value, if the distance is less than the threshold value, then the transmitting energy is defined as

$$E_T(d < d_0) = kE_{elec} + k\varepsilon_{fs} d^2 \qquad (4)$$

If the distance is greater than the threshold value, then the transmitting energy is defined as

$$E_T(d > d_0) = kE_{elec} + k\varepsilon_{mp} d^4 \qquad (5)$$

Here k is the message size in bits and Eelec is the energy consumed by the electronic circuit to transmit the signal. ET is the amount of energy to transmit a k-bits packet over distance d and ER is the amount of energy for receiving the k-bits packet. εfs represents the transmitter amplifiers efficiency and channel condition for shorter distance and represents the transmitter amplifiers efficiency and channel condition for longer distance. Aggregation energy is the amount of energy required by a node in aggregating the data's received from all other nodes and also the data sensed by it. The aggregation energy is a constant and it is defined as,

$$E_{agg} = 5nJ \qquad (6)$$

The node centrality is defined as how centre the node is in the network.

$$NC_i = (n-1)\sqrt{x_m^2 + y_m^2} \qquad (7)$$

Here NC is the node centrality, (n-1) is the number of neighboring node and (xm,ym) is the position of the node. Based on these parameters, the fitness value is calculated for each node in the network. After calculating the fitness value, the probability value is defined to elect the cluster head with highest fitness value. The probability value is based on the lower level value and the upper level value of the fitness function. Here lower level value is calculated as the ratio of the sum of the fitness function values to the number of nodes.

$$Lower\,LevelValue(L) = \frac{\sum_{i=1}^{n} fit_i}{n} \qquad (8)$$

The upper level value indicates that the data points tend to be very close to the mean. The upper level indicates that the data points are spread out over a large range of values. The variance is defined as the ratio of sum of the squares of the fitness value to the number of nodes. And upper level value is the square root of the variance.

$$Upper\,LevelValue(U) = \frac{\sum_{i=1}^{n} fit_i^2}{n} \qquad (9)$$

The probability of electing the cluster head (p) is based on the nodes whose fitness values are greater than lower level value and less than the upper level value.

After electing the cluster head, the clusters are formed based on the distance function. Clustering is defined as grouping of individual nodes into clusters. Each cluster will have a Cluster Head. For transmitting the data from one member node to sink, the data's can be directly send or data's can first to the cluster heads and then to the sink. By this way, the data traffic is considerably reduced. While transmitting the data from the member node to the cluster head and then to the sink the energy consumption is also considerably reduced. The distance between each node to the each cluster heads is calculated. The average value for the above distances is calculated.

$$Average = \frac{\sum_{i=1}^{n} d(mem, CH)}{n_{CH}} \qquad (10)$$

Here d (mem,CH) is the distance between each member node to the cluster heads and nCH is the number of cluster heads. The nodes with minimum difference in distance between the each member to cluster head and the average are grouped together to form a clusters with the corresponding node as a cluster head.

$$Dis\tan ce\ Function = \sum_{i=1}^{n} d(mem, CH) - Average \qquad (11)$$

Based on the distance value, the nodes with minimum distance forms a cluster. In order to have equal number of nodes in each cluster, a steady factor is used which has only limited number of member nodes. The node that first approaches a cluster head will join with the cluster head to form clusters. Once the limit over the number of member nodes has reached the remaining nodes has to find the next nearest member node. Thus based on this the cluster formation is done. By this way of clustering, the energy efficient clusters are formed.

B. Data Aggregation Phase

Figure 1 shows the structure of neural network model which uses multi layer perceptron neural network. For WSNs, each sensor node works as an underlying neuron; CH is like a central neuron, and the whole network is a complex neuron network system.

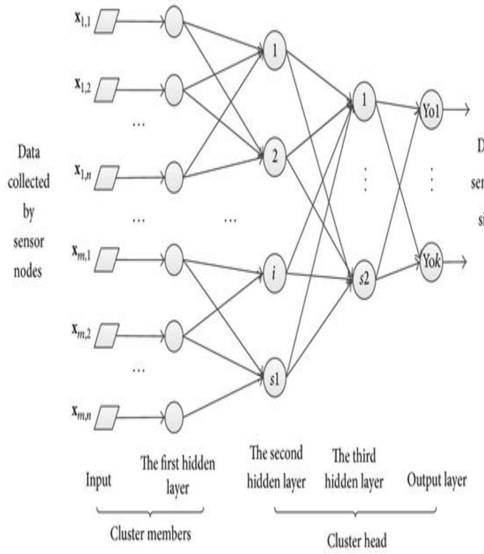

Figure 1 Multi-Layer Perceptron Neural Network for Data Aggregation

Cluster member nodes are in the input layer and the first hidden layer. CH nodes are in the second and the third hidden layer and output layer. For instance, each cluster with n members collects data of m species by different types. Consequently, this PNNDA model has n x m input layer nodes and the same number of first hidden layer neurons. The quantity of neurons in the second hidden layer, the third hidden layer, and the output layer can be adjusted by the users according to the practical application. The hidden layers have no inevitable relationship with n. It only deals with separate data of different types and need not be fully connected between the first and the second hidden layer. So do the second and the third hidden layer. But the treated data between the third hidden layer and output layer can be integrated, for they are full-connected.

In this proposed scheme, an aggregate contains not only a data value computed for the required aggregation function but also a count value indicating the number of sensor nodes involved in the aggregation operation. Here each aggregation packet contains the sender's id, an aggregated data value, and a count value to indicate how many nodes contributing to the aggregated data. In addition, a flag field of one bit is contained in each packet to show whether the aggregate needs to be aggregated further by the nodes enroute to the root. Flag value '1' means that no further aggregation is needed, whereas '0' means to be aggregated. This flag field is initialized to '0'.

C. Verification Phase

After the base station has received the aggregation messages from each cluster head, it verifies the authenticity of the aggregated value in each aggregation message. This includes verifying the content of the data packet and the authenticity of each cluster head. First, based on the cluster head id, say i, in the message, the base station finds out the individual key of the node K_i from which it decrypts the data and gets the information (i, C_i, Agg_i, MAC_i). The authenticity of the message is provided because the content format is known to the base station and it is checked by the pariwise key shared between respective nodes. After the above verification, the BS believes about the aggregate, say (c_x,Agg_x), is truly from a legitimate cluster head x. However, the base station cannot tell whether c_x or Agg_x has been modified because a compromised sensor node or the cluster head x may have modified the data, which can influence the final aggregation result at the base station. To get maximum assurance of the dependability of the aggregation result, the base station verifies the aggregation result with all the source nodes. Another term for this process is attestation because in effect the source nodes are requested to attest the validity of the aggregation result.

III. SIMULATION AND RESULTS

The performance of the proposed scheme is studied by simulating this scheme using MATLAB. Simulations were performed on 100 to 500 nodes randomly deployed in a 100m by 100m field with the base station located at coordinates (150m, 50m). The parameters used in simulation are tabulated in Table 1.

Table 1.Parameter settings for simulation

Parameter	value
Simulation time	25:150s
Initial energy of each node	2 joules
Transmission Range	2m
Energy to run the transmitter/receiver circuitry	70nJ/bit
Energy for the transmit amplifier	120pJ/bit/m^2
Size of data packet	4096 bits
Size of a control packet	20 bits
Number of compromised nodes	5:30

The protocols SDAP: A Secure Hop-by-Hop Data Aggregation Protocol for Sensor Networks [8] (Yi Yang et al 2006), READA: Redundancy Elimination for Accurate Data Aggregation[9] (Kavi Khedo et al 2010) and PRDA: Polynomial Regression based Secure Data Aggregation [10] (Suat Ozdemir et al 2011) are taken for comparison to evaluate the performance of the proposed scheme.

The performance of PNNDA was studied with respect to aggregators' energy dissipation, dropped data packet ratio, data transmission overhead, data aggregation accuracy and false data detection.

Aggregators' Energy Dissipation: It is defined as the average amount of energy spent during the aggregation process by the aggregators or the cluster heads.

Dropped Data Packet Ratio: It is the ratio between numbers of true or original data that cannot reach the BS to the total numbers of original data sent by the source to the BS.

Data Aggregation Accuracy: Message may be delayed and even dropped due to the processing time and collisions over wireless channels, so the aggregation accuracy is one of important performance metrics. The accuracy metric is defined as the ratio of the actual aggregation result collected by base station to the sum of the data sent by the all individual sensors.

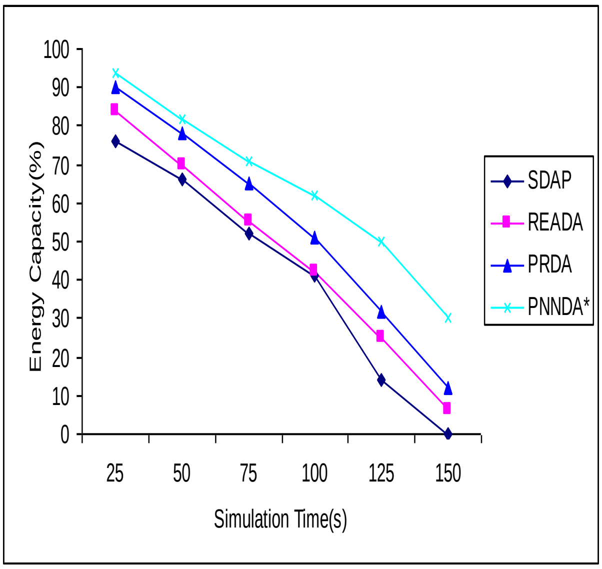

Figure 2 Aggregators' Energy Evolution

Figure 2 shows the aggregators' energy evolution for all the four protocols. Some observations are in order. First, for the simulation time less than 75s, the aggregators' remaining energy of READA is 55%. But in the case of PRDA, the aggregators' remaining energy is 65%. But compared with the PNNDA all the three remaining protocols have less aggregators' energy capacity. Second, the remaining energy level of aggregators of PNNDA is 30% even after the simulation time of 150s. Finally, the PNNDA protocol achieves 19-34% higher energy capacity for simulation time less than 100s to that of other protocols. Because the PNNDA optimizes the energy consumption during data aggregation by the energy efficient selection of cluster head procedure.

Figure 3 shows the dropped data packet ratio of the network with the presence of compromised nodes for all the four protocols. The system has been tested in the presence of 5, 10, 15, 20, 25, 30 compromised nodes. In PNNDA 14% of data are dropped in the presence of 10 compromised nodes and 42% of data are dropped in the presence of 30 compromised nodes. Thus most of the packet is delivered safely by the proposed content-based attestation process. In PNNDA, the portion of the packet contains MAC authentication information and hence the original message is prevented from the direct access of the compromised node. But in PRDA 66% of data is dropped in the presence of 30 compromised nodes because in PRDA the coefficients of the polynomials are very sensitive to the small changes of data. The SDAP protocol has higher packet drop ratio of 75% since it does not include the breadth-based attestation technique. Also SDAP cannot provide data confidentiality at data aggregators and result in latency because of the decryption or encryption process.

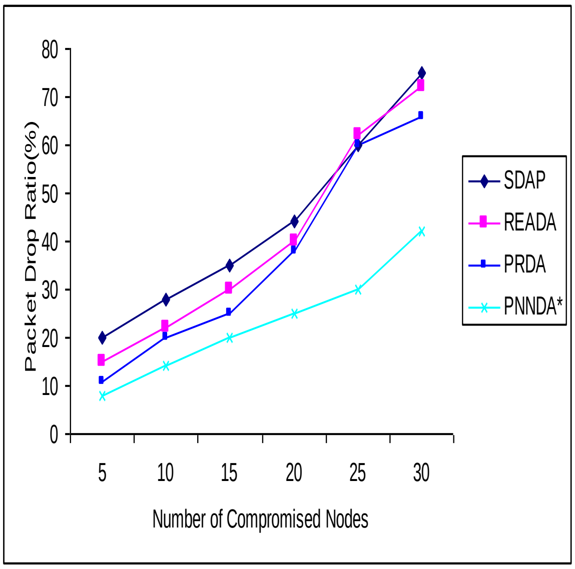

Figure 3 Packet drop ratio in the presence of compromised nodes

Figure 4 shows the accuracy metric of all the protocols. A higher accuracy value means the collected sum using the specific aggregation scheme is more accurate. The value 1.0 representing the ideal situation, where there is no data loss. The proposed PNNDA protocol maintains higher accuracy value compared to that of other protocols. Since finding the value of authenticated random challenge is hard for the attacker to perform static or dynamic analysis of the attestation procedure. But all the other protocols SDAP, READA and PRDA achieve maximum data aggregation accuracy of 0.75. Also the proposed data aggregation scheme identifies the faulty data sent by the compromised sensors to the cluster head, where the aggregation is performed. Identified faulty data eliminated before aggregation. Thus it improves the data aggregation accuracy.

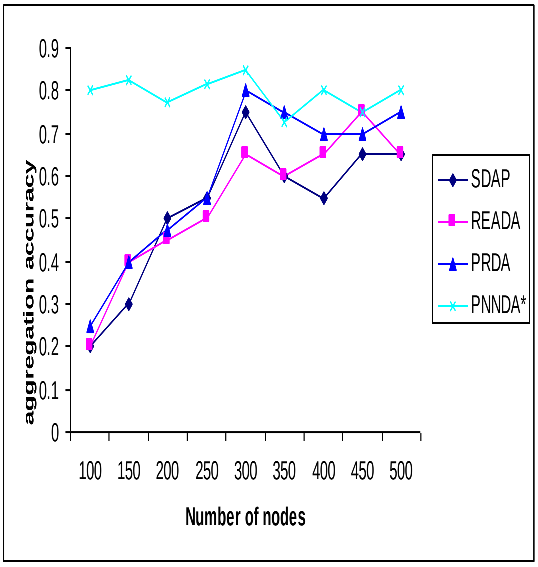

Figure 4 Data Aggregation Accuracy

IV. Conclusion

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 minimizes the packet drop ratio, increases data aggregation accuracy and filtering efficiency, reduces the key storage space and the data transmission overhead compared to other related schemes. Hence the proposed neural network approach can reduce the false data detection better in comparison to other schemes.

REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, E. Cayirci, “Wireless sensor networks: a survey”, Computer Networks 38 (2002), pp. 393–422.

[2] JAMAL N. AL-KARAKI, AHMED E. KAMAL, “Routing Techniques in Wireless Sensor Networks: A Survey”, IEEE Wireless Communications, (2004), pp. 1-23.

[3] Ossama Younis, Marwan Krunz, and Srinivasan Ramasubramanian, ” Node Clustering in Wireless SensorNetworks: Recent Developments and Deployment Challenges ”, IEEE Network, (2006), pp.20-25

[4] Kemal Akkaya , Mohamed Younis,” A survey on routing protocols for wireless sensor networks”, Ad Hoc Networks 3 (2005) 325–349.

[5] Girao, J. , Westhoff, D. , Schneider, M.,” CDA: Concealed Data Aggregation in Wireless Sensor Networks”, IEEE International conference on communications,(2005), pp. 3044-3049.

[6] Wenbo He, Xue Liu, Hoang Nguyen, Klara Nahrstedt, Tarek Abdelzaher,” PDA: Privacy-preserving Data Aggregation in Wireless Sensor Networks”, IEEE International conference on computer communications, 2007, pp. 2045-2053.

[7] Yingpeng Sang et al.,” Secure Data Aggregation inWireless Sensor Networks: A Survey”, International Conference on Parallel and Distributed Computing , Applications and Technologies,2006, pp. 315-320.

[8] Yi Yang et al,” SDAP: a secure hop-by-Hop data aggregation protocol for sensor networks”, 7th ACM international

symposium on Mobile ad hoc networking and computing, pp. 356-367, 2006

[9] Kavi Khedo, Rubeena Doomun, Sonum Aucharuz," READA: Redundancy Elimination for Accurate Data Aggregation in Wireless Sensor Networks", *Wireless Sensor Network,* 2010, pp. 300-308

[10] Suat Ozdemir, Yang Xiao," Polynomial Regression Based Secure Data Aggregation for Wireless Sensor Networks", IEEE Globecom 2011 proceedings.