International Journal of Security and Its Applications
Vol. 9, No. 6 (2015), pp. 43-52
http://dx.doi.org/10.14257/ijsia.2015.9.6.05

Energy Efficient, Integrity Preserving and Secure Data Aggregation Technique for Wireless Sensor Network

Shrijana Pradhan[1] and Kalpana Sharma[2]

Department of CSE, SMIT[1, 2]
Sikkim Manipal University
sriz_p@yahoo.com, headcs.smit@gmail.com

Abstract

Wireless sensor network consists of several resources constrained sensor nodes that are positioned in distant and un-navigable areas to monitor and sense various environmental conditions in the areas of interest. The sensor nodes arrange themselves into a network with the help of routing protocols. The sensor nodes instead of communicating the information individually to the base station, the information is sent to the transient cluster head, further it is communicated to the base station after computation. This is done to reduce the energy utilization incurred by the communication over head. In order to efficiently manage the energy level of cluster head, the cluster head should be well equipped with suitable computation processes that generate appropriate quality information to be communicated to the base station. Such computational processes are referred to as aggregation techniques. This paper proposes a refined data aggregation technique that aims at reducing the energy consumption at the same time preserving integrity and increasing security of aggregated data based on statistical approaches.

Keywords: *Wireless Sensor Network (WSN), Cluster Based Approach (CBA), data aggregation, cluster, mean, variance and standard deviation*

1. Introduction

Wireless sensor network comprises of sensor nodes deployed randomly or deterministically in the network to sense the attributes of interest. Limited power supply is one of the important constraints imposed on the sensors in wireless sensor network. The main objective of the sensor nodes is to sense the data and send it to the base station. The direct transmission of the data to the base station is not appropriate because of inherent distance between them and increased communication overhead. So, an appropriate clustering technique is to be followed in order to efficiently cluster sensors into group related by factor such as distance around a randomly or deterministically selected cluster head. The cluster head receives information from the sensors, aggregates and then communicates to the base station.

- Data Aggregation:

Data aggregation is the technique of collecting and processing the data with the aim of reducing the energy consumption in order to enhance the network life time.

Data aggregation can be performed on data sensed by the sensor nodes based on different approaches as detailed below.

- In-Network Data Aggregation

This technique is suitable for recording events occurring in a fixed region of space in the environment. The sensor network environment is parted into a set of pre-defined regions. Each region has to be observed and the observation inside the region is to be

ISSN: 1738-9976 IJSIA

communicated to the base station. Sensor device sends data to one sensor device inside its region. This sensor node is called as data aggregator. The aggregator sends the critical information to the sink node after processing. When any sensor nodes in the region detect an event, then the sensor nodes transmit their signal strength to its neighbors. If the neighbor has signal strength more than the sender, then the sender remains silent and stops its transmission of packets. Else, it waits for receiving packets from other sensors. After receiving all packets, if the sender has the highest signal strength, it becomes data aggregator and all other sensor nodes stop detecting the event. Finally the data packets are routed to the sink nodes.

- Grid Based Data Aggregation

It is well suited for mobile environments in which the time duration for an event at a particular region is very small. In this scheme, one sensor node is chosen as the data aggregator based on geographical position with respect to the base station or the center of the grid. The sensor nodes must have the information about the data aggregator inside their grid. During event detection, the sensor nodes send their information to the respective data aggregator. After collecting information from all sensor nodes, the data aggregator sends only the critical information to the base station. Thus, Grid-based scheme reduce traffic making it sure that only most important information is sent to the end nodes hence increasing throughput. End to end response time increases as the data packets exchanged between the sensors and their aggregator inside the grid fall in the critical path. But it increases congestion due to exchange of increased number of packets as compared to in-network scheme and its performance degrades when used in environment that is immobile in nature and highly localized.

- Hybrid Data Aggregation

This approach takes the best of both approaches i.e., In-Network scheme and Grid-Based Scheme for data aggregation. The sensor nodes are initially configured based on In-Network Scheme. When an event is detected by the sensor nodes, it first tries to identify the sensor node that holds highest signal strength, i.e. the sensor having the most critical and complete information about the event is selected as in case of In-Network Scheme. Further, each sensor node maintains a record of its past events as well as its corresponding signal strengths. While detecting the event, the sensor checks for the previous entries in its record table and try to identify if the event is mobile or stationary. If the event is found to be localized, then the In-Network Scheme is followed corresponding to which data aggregator is selected. Else, if it discovers a movement in the event, then it attempts to send information to the default aggregator which can be the sensor that lies close to the center of the grid and base station.

- Tree based Data Aggregation

In tree based data aggregation approach the aggregation is performed by building a tree such as minimum spanning tree in which the root is considered to be the base station and the leaves as the sensor node. Leaf nodes in the tree i.e. the sensor node transfers its data to its predecessor which performs the aggregation function and subsequently through a series of iterations forwards the data to the root node which acts as a base station.

- Cluster based Data Aggregation

In the cluster based data aggregation approach the entire network is divided into clusters. Every cluster has multiple nodes and among them a coordinator is selected known as the cluster head. Two-tier organization is used in clustering where cluster heads form the higher tier and the sensor nodes form the lower tier. The cluster head performs

the data aggregation function on the data collected from the nodes of the cluster and then forwards it to the base station after performing aggregation.

- Polynomial Regression based Secure Data Aggregation:

In the polynomial regression based secure data aggregation the sensed data is represented in the form of polynomial functions by the sensor nodes. The coefficient of the polynomial function is sent secretly by the sensor nodes to the data aggregator instead of transmitting the original data. Using these coefficients data aggregation is performed and the data is extracted by the Base Station from the aggregated result.

2. Literature Survey

WSN [4] has several sensor nodes with limited resources deployed in the area of interest. These nodes are equipped with sensor devices capable of sensing parameters of interest from the deployment area. The acquired data values are used for concluding upon quality decisions for initiating necessary action. Energy management of nodes is a crucial concern in WSN wherein energy is required both for communication and computation. Comparatively greater amount of energy is utilized for the purpose of communication compared to that of computation. To prevent this deployed nodes are grouped into clusters headed by a cluster head. The cluster head receives the data values from the sensor nodes and aggregates it and communicated it to the base station.

Data aggregation processes [5] plays a crucial role in conserving the energy supply *i.e.*, the life line of the nodes by reducing the number of communication instances required. The aggregation process may or may not reduce the data size depending on the need of the application *i. e.*, aggregation may be lossless or lossy. In lossless aggregation all the data items are communicated at one communication instance where as in case of lossy either of the following are used minimum, maximum, mean, mode or median to name a few. Aggregation may also be refined by considering semantic correlation in addition to spatial correlation and temporal correlation.

The selection of aggregation process [9] is dictated by the routing mechanism used in the network which is further governed by of network topology. These are tree based, cluster based, multi-path based and hybrid based approaches. The effective use of routing protocol can be enhanced by association an efficient aggregation process. In addition aggregation process should also emphasize on security to preserve data integrity, should be energy efficient to increase network life time, should reduce latency, minimize communication overhead and should increase accuracy of the aggregated data.

In this work [6] a query based aggregation techniques has been proposed and implemented. Based on TAG (Tiny Aggregation) this technique derives query specific aggregated value from the data set. The queries are controlled by quantified values specified on it and the mode for data selection. The proposed technique efficiently reduces the energy consumption but the process is to greater extent constrained by the limitation of TinyDB and has a considerably low loss tolerance.

Encryption and decryption techniques [7] can be used for preserving the integrity of aggregated data values but incurs greater computation overhead leading to non- optimal consumption of energy and resources. This overhead can be nullified by improving information processing during aggregation.

Project Evaluation Review Technique (PERT) [1] generates the best candidate time stamp for project execution taking into influence the optimistic time, the pessimistic time and the likely time required for same. An adaptation of the same can be used for generating candidate for aggregation.

DADMA (Data Aggregation and Dilution by Modulus Addressing) [3] technique has significantly reduced the number of data packet to be transmitted by almost 60 % but this technique works efficiently for situations where distributed dataset can be created with the knowledge of local dataset maintained by individual sensor nodes.

3. Proposed Methodology

In this approach, the sensors nodes are first organized into equalized clusters with approximately same number of sensors in the cluster. This is done in order to efficiently manage the energy depletion rate of different clusters and increase network life time. The clustering mechanism generates clusters by arbitrarily selecting initial cluster head and assembling nodes into cluster head till a predefined limit and then sharing the other sharing sensor nodes with neighboring clusters in order to create equalized clusters. Upon formation of clusters, in the individual clusters, successive cluster head are selected depending on the residual energy and the ability of the cluster head to sense predefined number of node for nominating it as cluster head.

Cluster head on receiving the values from the sensor nodes aggregates the values following steps stated below.

Step 1: Store the reading communicated by the sensors in a data structure

Step 2: Sort the data structure so as to arrange the data element in desired order

Step 3: Determine the mean (μ), the variance (*v*) and the standard deviation (σ) using the flowing formulae

- Calculates mean (μ)

$$\mu=\frac{1}{N}\sum_{i=1}^{N} xi \qquad \text{(equation 1)}$$

- Calculates variance(v)

$$v = \frac{1}{N}\sum_{i=1}^{N}(xi - u)\text{^}2 \qquad \text{(equation 2)}$$

- Perform standard deviation (σ) on the received data

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(xi - u)\text{^}2} \qquad \text{(equation 3)}$$

N=no of nodes, xi=sensed value, u=mean

Step 4: Determine the step size for verifying within how many instances of σ the values in the data structure are distributed and store this values in variable min step and max step. Here min step and max step represent the extent to which the values of the sensor reading vary from the mean in terms of deviation. This step is mandatory for determining the iteration count required while determining aggregation candidates.

Step 5: Divide the data structure into subsets taking into consideration

$$(\sigma * i)+\mu \text{ and } (\sigma *(i+1)) +\mu-1$$

where in 'i' varies from min step to max step.

Step 6: From the individual subset determine the minimum value, mean value and the maximum value.

Step 7: Determine the aggregated value by taking into consideration the influence of these values for a subset. Here the final aggregated data set is determined by evaluating one value by taking influence of minimum, mean and maximum value for each subset. This process is based on PERT techniques in software engineering, where execution time for the task is determined taking into influence of the best time, worst time, average time and instances of same.

Aggregated value= (instances $_{minimum}$ * minimum + instances $_{mean}$* mean + instances $_{maximum}$ * maximum)/ Total number of instances

The refinement in the proposed technique is that instead of having the data set being represented by one value that is either minimum or mean or maximum or mode, here we have a set of values that appropriately represent the data set considering influence of all possible values in the data set. The added advantage of this technique is that by assigning different number of instances we can derive different set of candidates based on the need of sensor network application.

Security is one key aspect of WSN. Here the data aggregators are often vulnerable to external attacks. These attacks include infestation of data set with redundant values to alter the result of aggregation process leading to communication of erroneous information to the base station.

The infestation may be placing values that are,

- less than the minimum value in the data set to be aggregated, leading to erroneous minimum and mean. In such case, minimum and mean cannot be used as basis for selecting candidate as representative for aggregation process.
- more than the maximum value in the data set to be aggregated, leading to erroneous maximum and mean. In such case, maximum and mean cannot be used as basis for selecting candidate as representative for aggregation process.
- approximately equal to the mean in the data set to be aggregated, leading to erroneous mean. In such case, mean cannot be used as basis for selecting candidate as representative for aggregation process.

Such attacks tremendously reduce the integrity of information being communicated. The proposed technique efficiently addresses to issues related to security, by not considering either of the values such as mean, minimum and maximum but generating aggregated values taking into consideration influence of all the values present in the data set.

The efficiency of the implemented process was also assessed against the ability of the cluster based technique (reference LEACH protocol). The implementation was done using Castalia framework of Omnet++ networking tool kit.

The procedure is explained with the help of an example as follows,

- **Simulation Inferences from Proposed Aggregation Technique:**

Step 1:

Table 1 here represents the reading of the sensors communicated to the cluster head. Here readings of 100 sensors are taken into consideration. An aggregated set of candidates is to be generated for the same in order to reduce the amount of information exchange taking place between the cluster head and the base station.

Table 1. Reading from the Various Sensors

Data Set:									
46	30	32	40	6	17	45	15	48	26
4	8	21	29	42	10	12	21	13	47
19	41	40	35	14	9	2	21	29	16
31	1	45	43	34	10	29	45	11	42
39	38	16	14	42	13	16	14	39	1
27	0	45	12	8	16	47	42	24	19
39	33	16	41	40	28	15	29	40	33
3	29	35	22	33	37	36	18	10	8
36	10	42	42	17	15	16	18	6	29
8	9	11	47	5	31	25	40	40	1

Step 2:

The information communicated to the cluster head is sorted for ease of further computation. Table 2 represents the sorted dataset.

Table 2. Sorting of Sensor Reading for Reducing Computation Complexity

Data Set after Sorting:									
0	1	1	1	2	3	4	5	6	6
8	8	8	8	9	9	10	10	10	10
11	11	12	12	13	13	14	14	14	15
15	15	16	16	16	16	16	16	17	17
18	18	19	19	21	21	21	22	24	25
26	27	28	29	29	29	29	29	29	30
31	31	32	33	33	33	34	35	35	36
36	37	38	39	39	39	40	40	40	40
40	40	41	41	42	42	42	42	42	42
43	45	45	45	45	46	47	47	47	48

Step 3 and **Step** 4:

After sorting the sensor reading, mean and standard deviation of the sensor readings are determined. This is done in order to determine the extent of the variation as represented in Table 3

Table 3. Mean and Standard Deviation

Mean	24.94	**Standard Deviation**	14.083
Min Step	-2	**Max Step**	2

Step 5 and **Step** 6:

Table 4. Subsets of the Data Set

Sub Set 1									
0	1	1	1	2	3	4	5	6	6
8	8	8	8	9	9	10	10	10	10
Min		0	**Mean**		5.95	**Max**		10	
Sub Set 2									
11	12	12	13	13	14	14	14	15	15
15	16	16	16	16	16	16	17	17	18
18	19	19	21	21	21	22	24		
Min		11	**Mean**		16.5	**Max**		24	
Sub Set 3									
26	27	28	29	29	29	29	29	29	30
31	31	32	33	33	33	34	35	35	36
36	37	38	39	39	39				
Min		26	**Mean**		32.5	**Max**		39	
Sub Set 4									
40	40	40	40	40	41	41	42	42	42
42	42	42	43	45	45	45	45	46	47
47	47	48							
Min		40	**Mean**		43.1	**Max**		48	

Step 7:

Table 5. Aggregated Values

Final Agg. Data:	5.63	16.81	32.53	43.42

While evaluating the final aggregated data, the ratio of instances with respect to minimum: mean: maximum was taken as 1:4:1.

Simulation Inferences for Performance Assessment of the Proposed Technique in Comparison with CBA:

The simulation was done taking into consideration 50, 100, 150, 200, 250, 300, 350, 400 and 450 nodes and various parameters were assessed.

The first and the most significant parameter assessed was Consumed Energy

Table 6. Consumed Energy (measured in Joule)

NODE	PROPOSED APPROACH	CBA
50	55.7	55.7
100	56.3	57.4
150	53.3	53.9
200	55.2	58.2
250	58.6	60.8
300	46	47.7
350	46.4	57.2
400	47.8	49.1
450	48	49.7

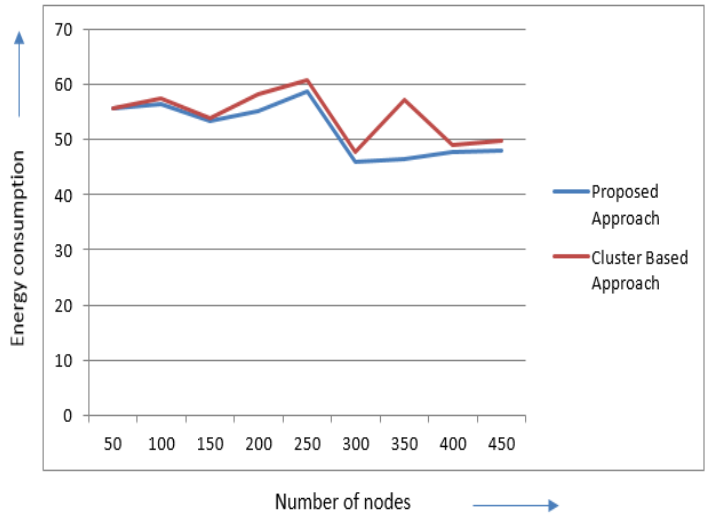

Figure 1. Consumed Energy Assessment

There is approximately a reduction of 4.6% on total energy requirement compared to that of CBA which is a significant achievement.

The second parameter assessed was application level latency,

Table 7. Application Level Latency (measured in milli-second)

NODE	PROPOSED APPROACH	CBA
50	92	19
100	96	40
150	1.8	1
200	59.5	61.25
250	88	111
300	190	73
350	58	37
400	28	23
450	8	7

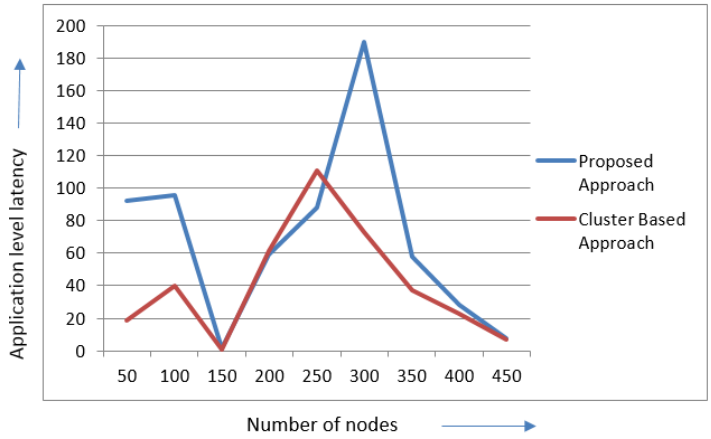

Figure 2. Application Level Latency

There is an increase of latency approximately by 40% compared to that of CBA which is high; this is as a consequence of greater computation involved while performing aggregation.

The Third parameter assessed was network layer packet size,

Table 8. Network Layer Packet Size (measured in bytes)

NODE	PROPOSED APPROACH	CBA
50	140	140
100	56	56
150	108	109
200	3	13
250	119	143
300	117.6	130.66
350	56	133
400	90.22	101.65
450	48	91.91

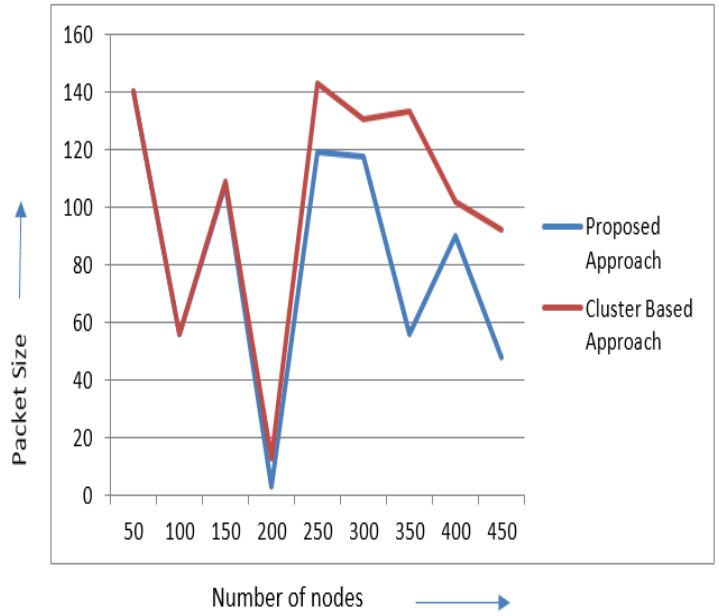

Figure 3. Network Layer Packet Size

There is a reduction of packet size by 20% compared to that of CBA which is high appreciable; this is as a consequence of the aggregation performed by the proposed technique.

4. Conclusion

In wireless sensor network, replenishment of energy for sensor node is a big challenge as these nodes once deployed may not be physically accessible; in addition transmission of large data packet increases communication overhead leading to increased consumption of energy. To address these challenges of WSN, this work implements an efficient technique based on mean, variance, standard deviation refined by concept of PERT.

The proposed technique reduced energy consumption by approximately 4.6% and data transmission by approximately 20% by compromising on latency due to increased computation.

One of the significant achievements of this technique is that it efficiently handles concerns related to security by taking into consideration influence of all possible values while deriving the aggregated set.

The future motivation would be to reduce the latency as well which would increase the efficiency of the proposed technique.

References:

[1] D. Wayne and P. E. Cottrell, "Simplified Program Evaluation and Review Technique (PERT)", Journal of Construction Engineering and Management, **(1999)**.
[2] B. H. Wendi, A. P. Chandrakasan and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks", IEEE Transactions On Wireless Communications, vol. 1, no. 4, **(2002)**.
[3] E. Cayirci, "Data Aggregation and Dilution by Modulus Addressing in Wireless Sensor Networks", IEEE Communications Letters, vol. 7, no. 8, **(2003)**.
[4] R. Rajagopalan and P. K. Varshney, "Data aggregation techniques in sensor networks: A survey", Electrical Engineering and Computer Science, vol. 22, **(2006)**.
[5] E. Fasoloy, M. Rossiy, J. Widmer and M. Zorzi, "In-network Aggregation Techniques for Wireless Sensor Networks: A Survey", Wireless Communications, IEEE, vol. 14, **(2007)**.
[6] N. S. Patil and P. R. Patil, "Data Aggregation in Wireless Sensor Network", IEEE International Conference on Computational Intelligence and Computing Research, **(2010)**.

[7] K. Maraiya, K. Kant and N. Gupta, "Wireless Sensor Network: A Review on Data Aggregation", International Journal of Scientific & Engineering Research, vol. 2, no. 4, **(2011)**.
[8] P. V. Ujawe1 and S. Khiani, "Review on Data Aggregation Review on Data Aggregation Techniques for Energy Efficiency in Wireless Sensor Networks in Wireless Sensor Network", International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 7, **(2014)**.
[9] A. Tripathi, S. Gupta and B. Chourasiya, "Survey on Data Aggregation Techniques for Wireless Sensor Networks", International Journal of Advanced Research in Computer and Communication Engineering, vol. 3, no. 7, **(2014)**.

