# ERROR DIAGONISIS IN ENERGY EFFICIENT DATA AGGREGATION TECHNIQUES

Sanjog Mahiman[1], Prof.Sudha Gupta[2], Chetan Ambekar[3]

Department of Electronics, K.J. Somaiya College of Engineering, Mumbai[1]

Department of Electronics, K.J. Somaiya College of Engineering, Mumbai[2]

Department of Electronics & Telecommunication, K.J. Somaiya College of Engineering, Mumbai[3]

*Abstract*— **Sensor networks are collection of sensor nodes which co-operatively send sensed data to base station. As sensor nodes are battery driven, an efficient utilization of power is essential in order to use networks for long duration hence it is needed to reduce data traffic inside sensor networks, also reduced amount of data need to send to base station. The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offer an increasingly Sensor nodes need less power for processing as compared to transmitting data. It is preferable to do in network processing inside network and reduce packet size. One such approach is data aggregation which attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. Wireless sensor networks have limited computational power and limited memory and battery power, this leads to increased complexity for application developers and often results in applications that are closely coupled with network protocols. In this paper, a data aggregation framework along with the algorithm on wireless sensor networks is presented. The aim of the proposed work consist a detailed study regarding the error which can occur by the data aggregation**

## I. INTRODUCTION

With advance in technology, sensor networks composed of small and cost effective sensing devices equipped with wireless radio transceiver for environment monitoring have become feasible. The key advantage of using these small devices to monitor the environment is that it does not require infrastructure such as electric mains for power supply and wired lines for Internet connections to collect data, nor need human interaction while deploying. These sensor nodes can monitor the environment by collecting information from their surroundings, and work cooperatively to send the data to a base station, or sink, for analysis.
The main goal of data aggregation algorithms is to gather and aggregate data in an energy efficient manner so that network lifetime is enhanced. Wireless sensor networks (WSN) offer an increasingly attractive method of data gathering in distributed system architectures and dynamic access via wireless connectivity. There are many techniques through which data aggregation takes place as [9, 10] aggregating on the basis of tree structures, [8] aggregating multi-hop networks. [5] Efficient techniques for data aggregation are explained which can be implemented.

The intelligent way to combine and compress the data belonging to a single cluster is known as data aggregation in cluster based environment. Dissemination is required for effective data handling [3]. There are some issues involved with the process of clustering in a wireless sensor network. First issue is, how many clusters should be formed that could optimize some performance parameter. Second could be how many nodes should be taken in to a single cluster. Third important issue is the selection procedure of cluster-head in a cluster [7]. Another issue is that user can put some more powerful nodes, in terms of energy, in the network which can act as a cluster-head and other simple node work as cluster-member only. There are also some problems related to diffusion of the data while conceiting the data [4, 6]. The paper consists of a case in which the



ISSN: 2278 – 909X All Rights Reserved © 2014 IJARECE

various types of errors which can occur during data aggregation are studied. A comparative study of the network is done and error percentage in the received signal is calculated.

The rest of the paper is organized as follows. Section II contains a brief analysis of various issues which can arise in a network. A brief review is provided of related work is discussed in section III. Section IV consists of the brief analysis of the algorithm. Section V consists of simulation results, conclusion in section VI.

## II. PROBLEM DEFINITION

Data aggregation protocols aims at eliminating redundant data transmission and thus improve the lifetime of energy constrained wireless sensor network. In wireless sensor network, data transmission took place in multi-hop fashion where each node forwards its data to the neighbour node which is nearer to sink. Since closely placed nodes may sense same data, above approach cannot be considered as energy efficient. An improvement over the above approach would be clustering where each node sends data to cluster-head (CH) and then cluster-head perform aggregation on the received raw data and then send it to sink. Performing aggregation function over cluster-head still causes significant energy wastage. In case of homogeneous sensor network cluster-head will soon die out and again re-clustering has to be done which again cause energy consumption.

## III. RELATED WORK: DATA AGGREGATION

Data aggregation is a process of aggregating the sensor data using aggregation approaches.[1] The general data aggregation algorithm works as shown in the below Figureure. The algorithm uses the sensor data from the sensor node and then aggregates the data by using some aggregation algorithms such as centralized approach, LEACH (low energy adaptive clustering hierarchy), PEGASIS [2] etc. aggregated data is transfer to the sink node by selecting the efficient path.
This is an address centric approach where each node sends data to a central node via the shortest possible route using a multi hop wireless protocol. The sensor nodes simply send the data packets to a leader, which is the powerful node. The leader aggregates the data which can be queried.
Each intermediate node has to send the data packets addressed to master from the slave nodes [1]. So a large number of messages have to be transmitted for a query in the best case equal to the sum of external path lengths for each node.

In-network aggregation **[8]** is the global process of gathering and routing information through a multi-hop network, processing data at intermediate nodes with the objective of reducing resource consumption (in particular energy), thereby increasing network lifetime. There are two approaches for in-network aggregation: with size reduction and without size reduction. In-network aggregation with size reduction refers to the process of combining & compressing the data packets received by a node from its neighbours in order to reduce the packet length to be transmitted or forwarded towards sink. In-network aggregation without size reduction refers to the process merging data packets received from different neighbours in to a single data packet but without processing the value of data.

In the tree-based approach **[9, 10]** perform aggregation by constructing an aggregation tree, which could bea minimum spanning tree, rooted at sink and source nodes are considered as leaves. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and therein the aggregation done by parent nodes.
In cluster-based approach, whole network is divided in to several clusters. Each cluster has a cluster-head which is selected among cluster members. Cluster heads do the role of aggregator which aggregate data received from cluster members locally and then transmit the result to sink.

## IV. ALGORITHM DESIGN:

We assume three cases for temperature sensing consisting of a network of random nodes. The cases will include fixed amount of nodes for sensing high, average and low temperature. We have considered the case of a series of 3n nodes where n is a natural number greater than equal to 3.
A detailed and energy efficient algorithm is defined for data aggregation consists of separate analysis for the aggregation of the nodes providing information regarding the average, high and low temperature. Consider m nodes in a network which are divided to 3 parts each has n nodes considered for the detection of high, low and average temperatures.
A network of n nodes randomly distributed in a differentiated area. Each has the tendency to provide the temperature related information to the cluster head.
Consider the action of sensing nodes, they are not always on. They work on the basis of a threshold temperature, if the surrounding temperature is more than threshold temperature then only it will send the data to the nodes representing high temperature.



ISSN: 2278 – 909X All Rights Reserved © 2014 IJARECE

Similarly the second pair of n nodes are provided to send the information only when the temperature recorded is less than the defined threshold value.
The third pair of n nodes are always on as they provide the average temperature of the surrounding and each and every instant.
The data reaching to the cluster head from the high temperature nodes is recorded and analysed by the following procedure.
A set of 6 readings for each set of n nodes are taken. It is assumed that the readings are initialized by zero in order to get better error efficiency. N readings are sorted into descending order of their temperature. Average of first two reading are taken and kept in another table. Then, next information is allowed to enter the table from top at the same time first reading is shifted to next position and the last reading is omitted (following the first in first out principle). In the similar manner average of the first two temperatures are taken and in the similar way, the table is filled.
The table consists of n rows, each and every time the rows are sorted in the descending order and the highest term (first element) of the sorted array is taken out and sent to the base station.
There is always a probability of error in the temperature which is sent to the base station. The error generated in this aggregation should be nil, but practically it is not.

Error diagnosis:
Let the matrix formed by the actual temperature readings be$f_n$ . After sorting it in descending order it becomes$M(f_n)$. The first two elements are averaged to get

$$\text{Max(n)} = \frac{M(f_n) + M(f_{n-1})}{2}$$
(1)

Max(n)is also an array which is updated each and every time having first in first out principle. First two elements are again averaged. The result is a single quantity which is sent to the base station.

$$w = \frac{\text{Max (n)+Max (n−1)}}{2}$$
(2)

The error produced in the temperature will be

$$errMax = \frac{w - f_n(z)}{f_n(z)}$$
(3)
Here z refers to respective node which is having the actual highest temperature.

A similar kind of process takes place in minimum temperature nodes. A set of 6 readings for each set of n nodes are taken. It is assumed that the readings are initialized by zero in order to get better error efficiency. N readings are sorted into ascending order of their temperature. Average of first two reading are taken and kept in another table. Then, next information is allowed to enter the table from top at the same time first reading is shifted to next position and the last reading is omitted (following the first in first out principle). In the similar manner average of the first two temperatures are taken and in the similar way, the table is filled.
The table consists of n rows, each and every time the rows are sorted in the ascending order and the highest term (first element) of the sorted array is taken out and sent to the base station.
There is always a probability of error in the temperature which is sent to the base station. The error generated in this aggregation should be nil, but practically it is not.

**Error diagnosis:**
Let the matrix formed by the actual temperature readings be$g_n$ . After sorting it in ascending order it becomes$L(f_n)$. The first two elements are averaged to get

$$\text{Min(n)} = \frac{L(g_n) + L(g_{n-1})}{2}$$
(4)

Min(n)is also an array which is updated each and every time having first in first out principle. First two elements are again averaged. The result is a single quantity which is sent to the base station.

$$w = \text{Min(n)} + \text{Min(n − 1)}$$
(5)

The error produced in the temperature will be

$$errMin = \frac{w - g_n(z)}{g_n(z)}$$
(6)
Here z refers to respective node which is having the actual lowest temperature.

In the case of average temperature reading of all the n nodes are taken and averaged. Similarly the value of the average is stored in an array of two rows. At each and every time the value sent to the base station will be the average reading of these two rows.

Error diagnosis:
Let the matrix formed by the actual temperature readings be$A_n(n)$ . Take the average of all the elements of $A_n$.

$$Av_n(w) = \left(\frac{1}{n}\right).\sum_{k=0}^{n} A_n\,(k)$$
(7)
$Av_n$is the average matrix, w is number of elements in $Av_n$. Average of w elements of $Av_n$is taken and


ISSN: 2278 – 909X All Rights Reserved © 2014 IJARECE

this value is sent to the base station as an average temperature.

$$Avg = \left(\frac{1}{w}\right).\sum_{k=0}^{w} Av_n\,(k)$$
(8)

The error produced in the temperature will be

$$errAvg = \frac{Avg - A_n(n)}{A_n(n)}$$
(9)

Here z refers to respective node which is having the actual average temperature.

**V. RESULTS:**

As discussed above we have taken a set of 3n nodes in a network. The nodes in this network are arranged in certain way that the nodes are close to each other. The graph shows the data which is sent and the error introduced in the system due to aggregation with respect to the number of nodes used in the aggregation. The maximum temperature denotes the aggregated data of total nodes allowed taking part in aggregation and the error responds about the change in the maximum temperature which is generated and the corresponding data which is sent.

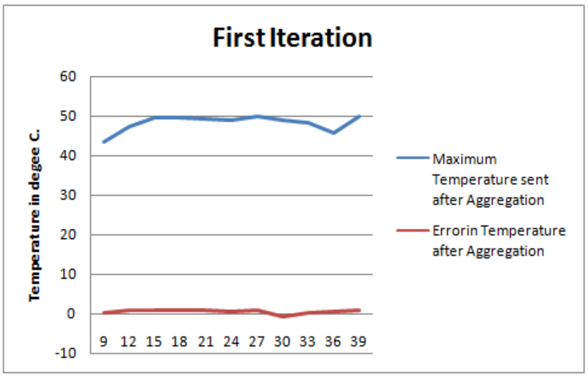


Figure 1.

The result is for first iteration in Figure 1. The error generated is less enough to affect the actual data. Also the maximum data shown in the graph doesn't have much difference from the actual data which is being generated.

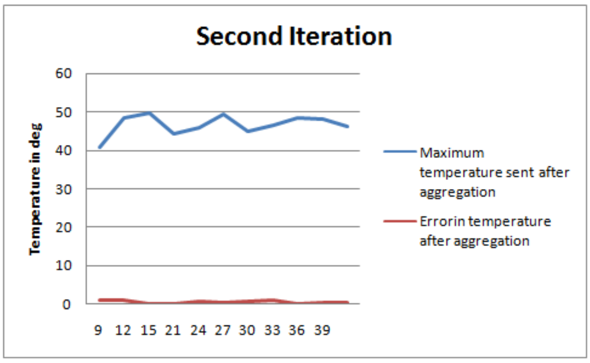


Figure 2.

Figure 2. shows the readings after second iterations. From the two iterations it is clear that the error is quite less to hamper the originality of the data. The single maximum temperature detected for n nodes matches with the original temperature with a slight error which can be accepted in turn to save transmission and receiving of more than 90%.
Figure 3. is for minimum temperature data sent on the behalf of nodes on the x axis. The green line on the graph shows the variation in the actual minimum temperature of the respective nodes to the temperature data sent.

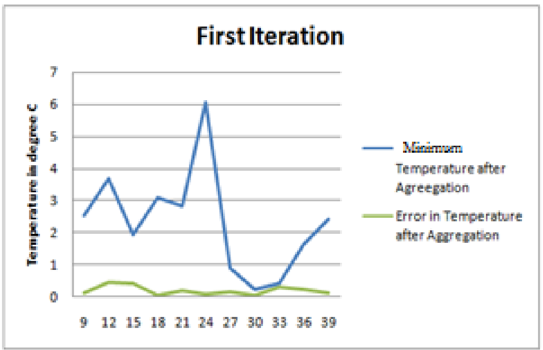


Figure 3.

In this case the error graph is comparable to the temperature the numbers of nodes are around 30-33. But allowing this error in the readings can provide a energy saving of more than 90% which is quite worthwhile.

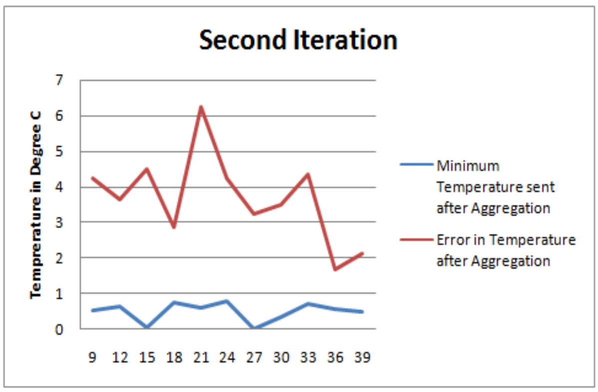


Figure. 4


ISSN: 2278 – 909X
All Rights Reserved © 2014 IJARECE

Figure 4. shows the second iteration of the minimum temperature data sent to the destination. In this case error is less as compared to the first iteration.

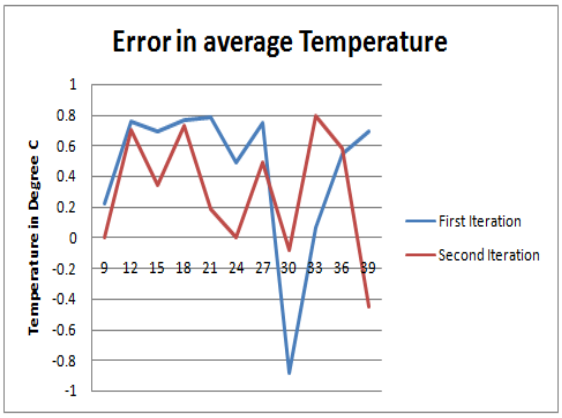


Figure. 5

Figure 5 shows the change in the error signal generated after sending the aggregated average temperature to the destination. The error is always then 1% in all the cases of high, low and average temperature. From the above simulation, it is clear that the as the nodes increases the error will increase but at the same time the % saving of energy will increase.

**APPLICATION:**

This type of data aggregation can be used in the areas which are highly energy vulnerable and the places where some sought of small error are allowed for a high saving of energy.

**VI. CONCLUSION:**

This paper consists of a detailed study about the data aggregation technique that can be implemented in the areas which are highly energy vulnerable. Also it provides an analysis regarding the error which can be generated during the aggregation. According to the results shown above the error generated is less than 1-2% in all the cases. At the cost of this error a high amount of transmitting, amplification, receiving energy can be saved. The algorithm is slightly complex, as this algorithm can be extended to make it more energy efficient and decrease its complexity.

REFERENCES:

[2]. S. Lindsey and C. Raghavendra, "PEGASIS: Powerefficient gathering in sensor information systems," in 2010 IEEE International Conference on Computational Intelligence and Computing Research Proceedings of IEEE AerospaceConference, vol. 3, Mar. 2002, pp. 1125–1130.

[3] M. Lee, and S. Lee, "Data Dissemination for Wireless Sensor Networks", in Proceedings of the 10 th IEEE International Symposium on Object and Component- Oriented Real-Time Distributed Computing (ISORC'07).

[4] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva, "Directed Diffusion for Wireless Sensor Networking", IEEE/ACM Transactions on Networking, Vol. 11, no. 1, Feb 2003.

[5]. H. Cam, S. Ozdemir, P. Nair, and D.Muthuavina shiappan, "ESPDA: Energy-Efficient and Secure Pattern-based Data Aggregation for Wireless Sensor Networks", in Proceedings of IEEE Sensor- The Second IEEE Conference on Sensors, Toronto, Canada, Oct. 22-24, 2003, pp. 732-736.

[6] Chalermek Intanagonwiwat, Ramesh Govindan, and Deborah Estrin, "Directed diffusion: a scalable and robust communication paradigm for sensor networks", (MobiCom 2000) pp 56-67.

[7] K. Dasgupta, K. Kalpakis, and P. Namjoshi, "An
Efficient Clustering-based Heuristic for Data Gathering and Aggregation in Sensor Networks", IEEE 2003.

[8]. E. Fasolo, M. Rossi, J. Widmer, and M. Zorzi, "In- Network Aggregation Techniques for Wireless Sensor Networks: A Survey", IEEE Wireless communication 2007.

[9]. M. Lee and V.W.S. Wong, "An Energy-aware Spanning Tree Algorithm for Data Aggregation in Wireless Sensor Networks," IEEE PacRrim 2005, Victoria, BC, Canada, Aug. 2005.

[10] M. Ding, X. Cheng, and G. Xue, "Aggregation tree construction in sensor networks," in Proc. of IEEE VTC'03, Vol. 4, Orlando, FL, Oct. 2003.



ISSN: 2278 – 909X All Rights Reserved © 2014 IJARECE