Energy Efficient Multilevel Data Aggregation technique for adequate redundancy removal in WSN

Samarth Anavatti
Department of Information Technology
Maharashtra Institute of Technology
Pune, India
anavattisamarth@gmail.com

Sumedha Sirsikar
Department of Information Technology
Maharashtra Institute of Technology
Pune, India
sumedha.sirsikar@mitpune.edu.in

***Abstract*— Wireless Sensor Networks (WSN) consisting of small number of sensor nodes. These sensor nodes are deployed in a region to sense the information. As the sensor nodes deployed in the same region sense similar information and sends it to sink node. This phenomenon leads to redundancy at sink node. Sink node wastes most of its energy in processing redundant packets. To save the energy of nodes processing data packets in order to prolong the network lifetime, there is a need to eliminate redundancy. Data aggregation is a process which eliminates redundant data packets in the network. The process in which intermediate node receives multiple input packets performs aggregation and produce single output packet in the network is called Data Aggregation. But redundancy sustains reliability. Therefore there is need to eliminate redundant packets to an adequate level so as to maintain reliability. In this paper we introduced one data aggregation technique which performs data aggregation at two different levels, at first level (at cluster head) we are using Locality Sensitive Hashing which will keep redundancy count for each node and node which has lesser count than threshold are chosen for aggregation, the result of this aggregation is passed to next level for aggregation (at storage node) and final aggregation result is given to base station. Also in this technique we provided session id based security mechanism to prevent false data from invalid node. This data aggregation technique will maintain tradeoff between energy efficiency and accuracy and will provide security from invalid nodes.**

Keywords- WSN, Data aggregation, redundancy, LSH Introduction

I. INTRODUCTION

Wireless Sensor Networks (WSN) consists from large number of small sensor nodes. These nodes communicate with each other wirelessly within their radio range. Sensor nodes have limited battery power and therefore it has become necessary to reduce energy consumption at each sensor node to enhance the overall life time of wireless sensor network. This can be achieved by eliminating redundancy from WSN [1]. There are two types of redundancies temporal redundancy and spatial redundancy. Spatial redundancy means multiple sources obtaining information about one region. Spatial redundancy is based on placement of the sensor nodes in sensing region. In WSN spatial density is high. Temporal redundancy also called time redundancy can be defined as performing specific action more than once, to increase the readability. Both redundancies are useful to increase reliability and accuracy [3]. But sensor nodes waste their energy in processing redundant data therefore redundancy removal has become a solution for improving WSN's life time. Data Aggregation is the technique which gathers and aggregate data in an energy efficient manner and reduces redundancy so that network lifetime is enhanced.

Though redundancy removal improves the lifetime of the WSN, redundancy is important to get accurate results in WSN. So there is a need to keep the redundancy up to an adequate level. Also there is another issue in data aggregation process which is security, nodes may get compromised and will send inaccurate data which leads to inaccurate aggregation. In this paper we proposed one data aggregation technique which contains security mechanism to provide security from invalid nodes.

WSN has wide range of application it can be used in agriculture, military, bio-diversity mapping, disaster relief operations etc. Our proposed model will also be useful to survey large geographical area. Consider an example, in agricultural field a farmer want to check that, is water supplied to all plants properly or not? This can be achieved by dividing total area of field in 4 quadrants and then deploy many sensor nodes in it. Then form clusters of all sensor nodes. These sensor nodes will sense water level in each region, and send respective readings to CH. All cluster heads will send these readings to one level above node as in our model we call it as Storage Node and result from these nodes are combined at base station and then by converting it into human readable language it is transferred to monitoring station. In this way farmer can operate water supply to his farm by sitting at home.

The rest of the paper is organized as follows, Section II contains related work, problem statement is defined in section III, Network model and mathematical model are given in sections IV and V respectively. Section VI contains proposed model of our work. Section VII is information about simulation parameters and Expected performance behavior is shown in section VIII. Conclusion is given in section IX.

II. RELATED WORK

Patil [1] performed comparison of performance of WSN with data aggregation and performance without data aggregation. Ghasem [2] proposed an approach called Dynamic Balanced Spanning Tree Approach which is improvement over fixed spanning tree approach. In fixed spanning tree there was problem of "hotspots" which is eliminated here. This work results in minimum energy consumption and also it balances the traffic load.

Kulkarni [3] developed a new technique which makes use of Support Vector Machine (SVM) to eliminate redundant data and then used LSH algorithm to eliminate outliers who may send false data. In this way this technique not only reduces energy consumption but also eliminate false data.

Dnyaneshwar Mantri, Neeli Rashmi Prasad, Ramjee Prasad [4] developed cluster based approach which works efficiently in heterogeneous environment. This approach reduces redundant transmission by performing intra cluster and inter cluster data aggregation which utilizes bandwidth efficiently and reduces energy consumption.

Sumalatha Ramachandran, Aswin Kumar Gopi, Giridara Varma Elumalai, Murugesan Chellapa [5] proposed a novice cluster based approach which uses context aware system to validate data and then eliminate redundancy of validated data. This redundancy is removed using correlation coefficient technique. Basavaraj Mathapati, Siddaram Patil [6] have developed a technique which performs reliable and energy efficient data aggregation and its comparison with erdc [11].

III. Problem Statement

WSN is consisting of sensor nodes which are tiny and have less battery power. If this battery power of a sensor node get completely exhausted then the node dies. Replacing the battery of sensor node or charging it again are not feasible solutions. So there is a need of some mechanism which will reduce the processing of sensor nodes and conserve its energy.

Data aggregation is a technique which takes number of data packets as input and produce one aggregated packet as output. Data aggregation is useful for energy conservation because in WSN sensor nodes deployed in some area senses the same information and forwards to the header node or intermediate nodes. These nodes waste their energy in processing the same packet frequently.

Data aggregation eliminates redundancy from WSN. But redundancy sustains accuracy. So there is a need to maintain the tradeoff between energy conservation and accuracy. In this paper we proposed a data aggregation approach which not only maintains the tradeoff between energy conservation and accuracy but also reduces the traffic load and provides security.

IV. Network Model

Entities of all components of our model are defined in this section. Figure:1 shows basic structure of network model. Let M be the network model which can be described as follows:

M= {N, C, CH, SN, BS}

where:

N: Number of sensor nodes

C: Cluster of sensor nodes

CH: Cluster Head

SN: Storage Node

BS: Base Station

In our model geographical area is divided into regions and there are many sensor nodes deployed in each region. Some of which are static and some are mobile in nature. Geographical region contains many clusters (C). Each cluster has coordinator node called Cluster Head (CH). Each region contains Storage Node (SN) which combines the data coming from all CH's and send it to BS. SN are static and have high initial energy.

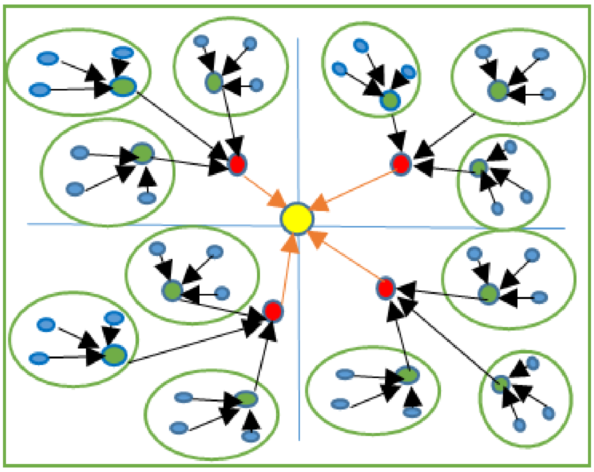

Figure 1: Network Model

V. Mathematical Modelling

Mathematical model is described in this section.

A. N = { Set of all nodes present in the network}

N = { n1, n2, n3, ….}

Set contains all sensor nodes from the network and many clusters are formed to manage these sensor nodes.

B. CH = Cluster Head

CH Є N

CH is the Cluster head which monitors the functionality of cluster.

C. SN ={ Set of all Storage nodes}

SN = {x | x Є N ʌ x ≠ CH}

Storage Node is a node which has more initial energy and power. This node has storage capacity also. It is static node and it is unique in each region.

D. Ei = Initial Energy of each node

Energy consumed by node i, initially we have defined the amount of energy required for certain events example Send, Receive, Drop etc. According to this when particular event occurs that much energy is reduced from initial energy Ici.

E. Ici= Energy Consumed by i th node

= Initial Energy Ei – (Residual energy)

F. P {Set of parameters} = { TCE,DL, PDR, AEC }

1) *DL – Delay*
2) *PDR- Packet delivery ratio*
3) *AEC– Average Energy Consumption*
4) *TCE- Total Consumed Energy*

G. Delay (DL)

$$Delay(DL) = \sum_{i=1}^{N} EndTime(i) - StartTime(i) \qquad \dots (1)$$

H. Total Consumed Energy(TCE)

$$TCE = \sum_{i=1}^{N} (Initialenergy - remainingenergy) ofnode \qquad \dots (2)$$

I. Average Energy Consumption(AEC)

$$\boldsymbol{AEC} = \sum_{i=1}^{Nt} Ici/Nt \qquad \dots (3)$$

Average Energy Consumption is the total energy consumed by nodes in the network. Here we considered only Nt nodes from N (Nt < N) because that nodes are having redundancy count lesser than threshold (Th).

VI. PROPOSED MODEL

Proposed model shown in Figure:2 is described briefly in this section. In this model, total geographical area is divided into regions and many sensor nodes are deployed in it. Clusters are formed on each region, and CH is elected. Initially the node which is closer to SN is selected as CH. After one round the new CH is selected based upon residual energy. In this way we have dynamically distributed the role of CH due to which hotspot problem will get solved [2]. We performed data aggregation at two levels to eliminate redundancy. First level is data aggregation at cluster head; in this level first of all, each node creates its LSH code and send it to CH, as acknowledgement to CH's hello packet. The CH maintains one table called match table as shown bellow.

TABLE I. MATCH TABLE

Node ID	*LSH CODE*	*Redundancy Count (RC)*
1	(1,10)	5
2	(2,10)	7

This table I contains entries for all nodes in the cluster and with their LSH code and redundancy count for each LSH code. This LSH code is nothing but combination of node id and session id. When session starts each node when it has data, will send data to CH. So there is a chance that it will send same data packet more than once since its session is alive. CH maintains similarity count also called Redundancy Count(RC) of each data packet for particular node with in same session. Value of less redundancy count and value of highest redundancy count is used for calculating threshold for redundancy count. If node have RC value greater than threshold (Th) that node is not considered in aggregation process and we make it sleep. Only those nodes have RC value lesser or equal to threshold are considered in aggregation process. As we have considered only few nodes for aggregation, we can sustain reliability and considering few nodes for aggregation will save energy consumption also. This results in reducing average energy consumption of network. Now aggregation of data from eligible nodes is done and result of aggregation of a single cluster is sent to SN.

SN node is static node in each region with high processing power and capacity. We performed second level aggregation at SN. As all CH from same region will send data to SN, to save the time and processing power of BS we performed data aggregation at SN. Result of this aggregation is given to BS.

In this model along with energy conservation we are providing security mechanism based on Session ID of node. This security mechanism will perform validation of cluster nodes. Security mechanism is as follows:

a) All the three types of nodes create their own session ID(Unique)

b) Normal Sensor Nodes will send their ID's to CH and CH will send its ID + $\Sigma all\ nodes\ SID$ to SN

c) SN will add its own ID with CH data and will generate key for that cluster

d) Key is distributed to all nodes in the cluster

Nodes have to send with this key whenever it has data to send. CH matches this value with its key. If both matches data is accepted otherwise rejected. In this way we will allow only the nodes from the cluster to send data to cluster head.

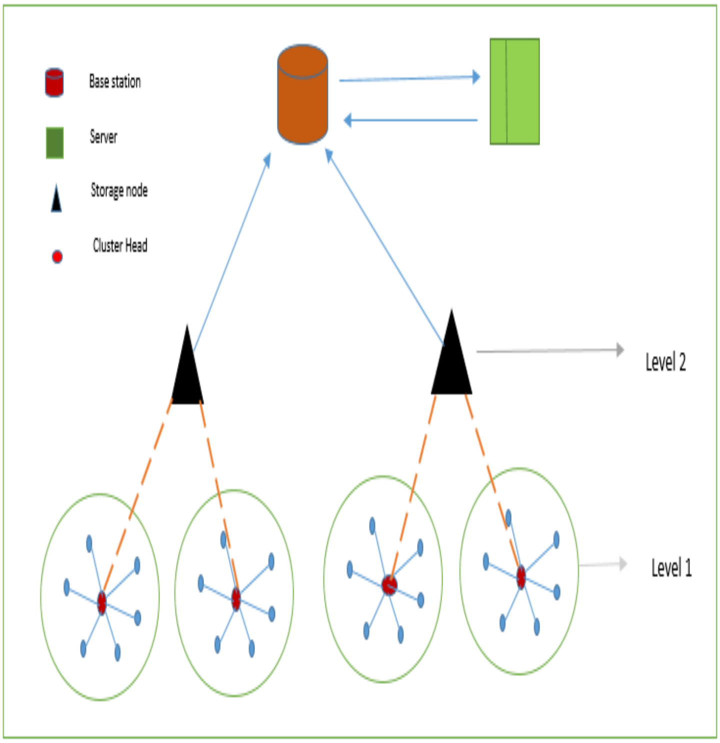

Figure 2 : Proposed Model for Data Aggregation

VII. Simulation Parameter

In this section we created a table which shows parameters and their values at the time of initialization.

TABLE II. SIMULATION PARAMETERS

Parameter	*Value*
Network size	500*500m
Initial energy	5J
Packet size	4000bits
Transmit/Receive energy	50nJ/bit
Transmission Range	50 m
Channel Type	Wireless

VIII. Expected Performance Behaviour

With the help of mathematical model described in section V, we have done theoretical calculation for residual energy and based on that performance comparison of general aggregation technique with our aggregation technique is done. To do this we have done some assumptions like initial energy =5J, initial energy of SN= 10J. Number of nodes = 100.
For simplification of calculation we assumed energy required for send, receive etc. events is 1J and energy required for aggregation is 1J. According to this ordinary node will consume 1J energy and aggregator nodes will consume 2J. Following table shows amount of residual energy after each session (Round).

TABLE III. CALCULATION OF RESIDUAL ENERGY FOR GENERAL AGGREGATION TECHNIQUE

Round Number	*Total number of nodes*	*Total Energy*	*Energy Consumed*	*Residual Energy*
1	100	500	102	408
2	100	408	102	306
3	100	306	102	204
4	100	204	102	102

TABLE IV. ALCULATION OF RESIDUAL ENERGY FOR OURAGGREGATION TECHNIQUE

Round Number	*Total Number of Nodes*	*Total Energy*	*Nodes participated in aggregation*	*Energy Consumed*	*Residual Energy*
1	100	505 J	50	52 J	453 J
2	100	453 J	60	62J	391 J
3	100	391 J	40	42J	349 J
4	100	349 J	70	72 J	277 J

In Table III and Table IV we have residual energy after 4^{th} round is more in our technique. So from this we can conclude that our technique will reduce average energy consumption and will improve network lifetime.

IX. Conclusion

Data aggregation technique removes data redundancy from Wireless Sensor Networks (WSN) in order to enhance the lifetime of WSN. But in some situations redundancy provides accuracy in WSN. So there is a need of data aggregation technique which will make balance between accuracy and energy conservation.

The proposed work gives an approach which performs data aggregation at two different levels along with that LSH is used to eliminate redundancy up to an adequate level. In this way, this technique maintains tradeoff between energy conservation and accuracy. This technique also performs validation of nodes present inside the cluster to avoid false data which will give more accurate aggregation results.

References

[1] Nandini. S. Patil, Prof. P. R. Patil, "Data Aggregation in Wireless Sensor Network," IEEE International Conference on Computational Intelligence and Computing Research, ISBN: 97881 8371 3627

[2] Avid Avokh, Ghasem Mirjalily Patil, "Dynamic Balanced Spanning Tree (DBST) for Data Aggregation in Wireless Sensor Networks," 5th International Symposium on Telecommunications (IST2010), 978-1-4244-8185-9/10 ©2010 IEEE

[3] Prakash goud Patil, Umakant Kulkarni, "SVM based Data Redundancy Elimination for Data Aggregation in Wireless Sensor Networks," 978-1-4673-6217-7/13 © 2013 IEEE

[4] Dnyaneshwar Mantri, Neeli Rashmi Prasad, Ram jee Prasad, "BHCDA:Bandwidth Efficient Heterogeneity aware Cluster based Data Aggregation for Wireless Sensor Network," ICRTIT 2011, 978-1-4673-6217-7/13 2013 IEEE

[5] Sumalatha Ramachandran, Aswin Kumar Gopi, Giridara Varma Elumalai, Murugesan Chellapa, "REDD: Redundancy Eliminated Data Dissemination in Cluster Based Mobile Sinks," ICRTIT 2011, 978-1-4577-0590-8/11 ©2011 IEEE

[6] Basavaraj S. Mathapati, Siddarama. R. Patil, "Energy Efficient Reliable Data Aggresgation Technique for Wireless Sensor Networks," International Conference on Computing Sciences, 978-0-7695-4817-3/12 © 2012 IEEE

[7] Gagarin, S. Hussain, and L.T. Yang, "Distributed hierarchical search for balanced energy consumption routing spanning trees in

Wireless Sensor Networks," J. Parallel Distrib. Comput. Vol. 70, no. 9, pp. 975-982, 2010

[8] Prakash goud Patil, Umakant Kulkarni "Delay Efficient Data Aggregation Algorithm in Wireless Sensor Network," International Journal of Computer Application (0975-8887) vol.69 No.1, May 2013

[9] A.C. Viana, A.Ziviani and R.Friedman, "Decoupling Data Dissemination from Mobile Sink's Trajectory in Wireless Sensor Networks," IEEE Communication Letters, vol. 13, no. 3, Mar. 2009

[10] Yuea Jun Weiming Zhang, Weidong Xiao, Daquan Tang, and Jiuyang Tang, "Energy efficient and balanced cluster-based data aggregation algorithm for wireless sensor networks," Procedia Engineering 29 (2012): 2009-2015

[11] Ren P. Liu, John Zic, Iain B. Collings, Alex Y. Dong, and Sanjay Jha, "Efficient Reliable Data Collection in Wireless Sensor Networks," in proceedings of IEEE 68th Vehicular Technology Conference, VTC2008, 2008.