Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 49 (2015) 194 – 201

Procedia Computer Science

ICAC3'15

Issues of Data Aggregation Methods in Wireless Sensor Network: A Survey

Sumedha Sirsikar[a],Samarth Anavatti[b]

[a&b]*Department of Information Technology,MIT,Pune,411038*

Abstract

In Wireless Sensor Networks (WSN) sensor nodes are deployed in a region to sense the information. These sensor nodes sense the similar information and sends it to sink node. This thing leads to redundancy at sink node. Sink node wastes most of its energy in processing redundant packets. To save the energy of node in order to prolong the network lifetime there is need to eliminate redundancy. Data aggregation is a process in which intermediate node receives multiple input packets performs aggregation and produce single output packet in the network. This process will reduce the number of redundant packets in the network. But redundancy sustains reliability. Therefore there is need to maintain redundancy but it should be up to an adequate level. In this paper we have focused on different issues in data aggregation process such as delay, redundancy elimination, accuracy and traffic load and mentioned various methods to solve those issues and then we compared some data aggregation techniques based on strategy, delay, redundancy, average energy consumption and traffic load. Further we have proposed a model based on our study which performs data aggregation at multiple levels and not only maintains the tradeoff between energy conservation and reliability but also addresses all the issues in data aggregation technique.

Peer-review under responsibility of organizing committee of the 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)

Keywords: Wireless Sensor Network; Data Aggregation; redundant data;

1. Introduction

Wireless Sensor Networks (WSN) contains large number of small sensor nodes. These nodes communicate with each other wirelessly within their radio range. Due to limited power of sensor nodes it has become necessary to reduce energy consumption at each sensor node to enhance the overall life time of wireless sensor network. This can

Peer-review under responsibility of organizing committee of the 4th International Conference on Advances in Computing, Communication and Control (ICAC3'15)
doi:10.1016/j.procs.2015.04.244

be achieved by eliminating redundancy from WSN. Because nodes waste their power in processing redundant data. Therefore redundancy removal has become solution for improving WSN's life time. Data Aggregation is the technique which gather 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 make balance in between these two, reliability and energy conservation. This paper focuses on data aggregation issues like redundancy, delay, accuracy and traffic load also we have mentioned some mechanisms to solve these issues. Further we have done comparative study of various data aggregation techniques based on Strategy, Delay, Redundancy, Average Energy consumption and Bandwidth-overhead and proposed a new model of data aggregation which solves all these issues.

The rest of the paper is organized as follows. Section 2 contains Background of data aggregation, section 3 is related work. Section 4 discusses issues in the Data aggregation techniques and methods to solve these issues. Section 5 comes out with the comparative study of all techniques. Section 6 contains our proposed model to address all the issues in data aggregation and section 7 is conclusion of our work.

2. Background of Data Aggregation

The sensors deployed in the closer region sense the same phenomena which leads to produce lot of duplicate data. This duplication of data consumes more bandwidth and energy of sensor network. In data aggregation, sensor data collected by sensor nodes are aggregated by using some data aggregation algorithms and then aggregated data is forwarded towards base station. There are different strategies for data aggregation described as follows.

2.1. Data Aggregation Strategies

There are four strategies for aggregation some of them are listed below:
Centralized Approach, In-Network Aggregation, Tree-Based Approach, Cluster-Based Approach [1].

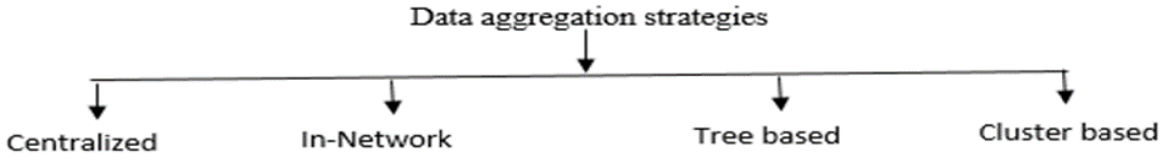

Figure 1: Data aggregation strategies

Centralized Approach: In this approach each node sends data to a central node via the shortest possible route. All the sensor nodes simply sends the data packets to a node, which is the powerful among all other nodes. This node is called as header node. This header node performs aggregation of data coming from all nodes and result of this aggregation process will be a single packet.

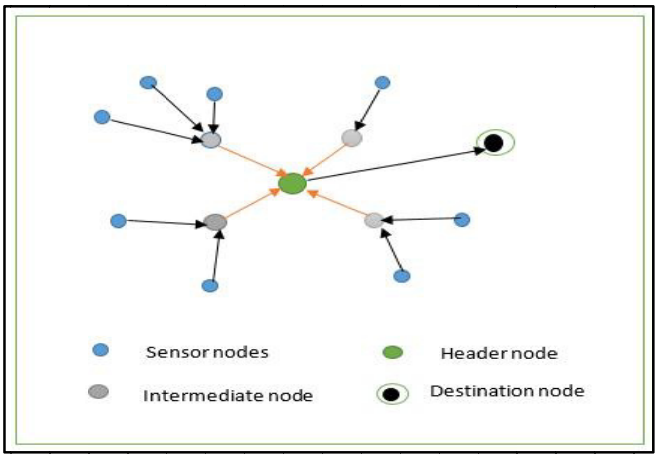

Fig 2: Centralized data aggregation

In-Network Aggregation: The aggregation is the global process of gathering and routing information through a multi-hop network and processing data at intermediate nodes with the objective of reducing power consumption. There are two approaches for in-network aggregation:

With size reduction : It refers to the process of combining & compressing the data packets received by a node from its neighbors in order to reduce the packet length that is to be transmitted or forwarded towards sink.

Without size reduction: In-network aggregation without size reduction refers to the process merging data packets received from different neighbors into a single data packet but without processing the value of data [1].

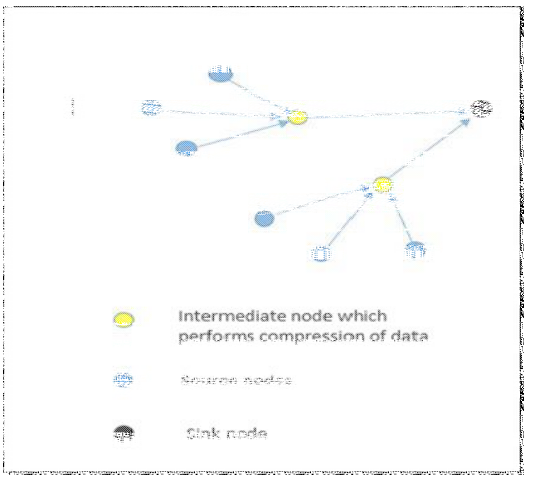

Fig 3: in – network with size reduction

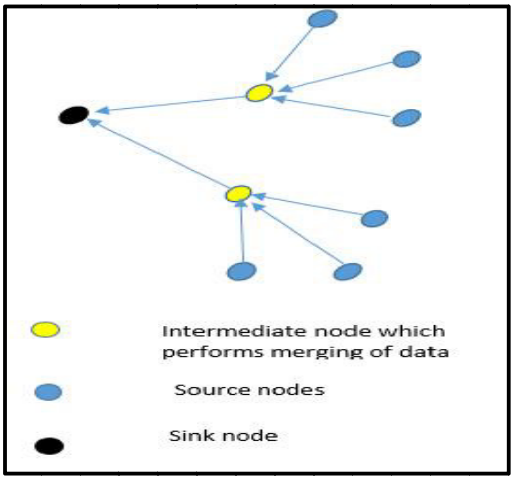

fig 4: in-network without size reduction

Tree-Based Approach: In the tree-based approach at first Data Aggregation Tree (DAT) is formed. For each data transmission minimum spanning tree is created. Each node has a parent node to forward its data. Flow of data starts from leaves nodes up to the sink and the aggregation done by parent nodes.

Cluster-Based Approach: In cluster-based approach, whole network is divided in to several clusters. Each cluster is consisting of many sensor nodes. For each cluster one header node is elected which is also called as cluster-head. The Cluster-heads performs the aggregation and then transmit the result. It will result in reducing the bandwidth overhead as total number of packets to be transmitted are less.

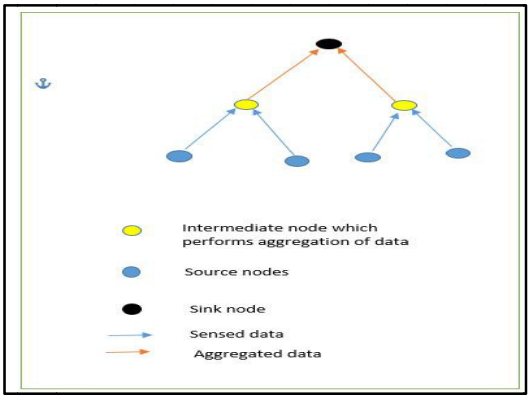

Figure 5: Tree based data aggregation

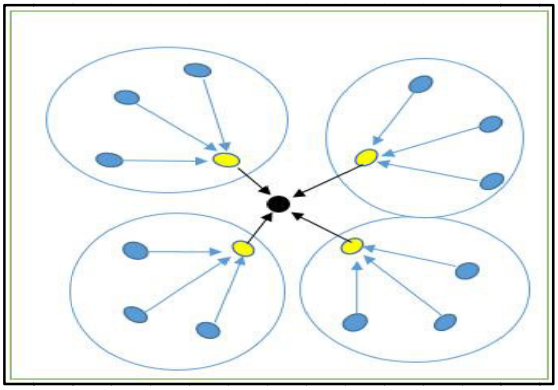

Figure 6: Cluster based data aggregation

3. RELETED WORK:

Nandini. S. Patil, Prof. P. R. Patil [1] performed comparison of performance of WSN with data aggregation and performance without data aggregation.AvidAvokh,GhasemMirjalilyPatil [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.

PrakashgoudPatil,Umakant 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.

DnyaneshwarMantri, NeeliRashmi 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 issues Gopi, GiridaraVarma 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 S.Mathapati, Siddarama. R. Patil [6] have developed a technique which performs reliable and energy efficient data aggregation and they compared its performance eRDC [11].

4. ISSUES IN DATA AGGREGATION:

Data aggregation technique removes the redundancy of the data. Each data aggregation technique has some issues like redundancy, delay, accuracy and traffic load. Because of these issues performance of data aggregation technique gets affected. In our paper, we considered some methods to solve these issues which belongs to tree based and clustering based data aggregation strategies. Following taxonomy shows issues and methods in data aggregation to solve these issues.

4.1. Redundancy Elimination:

In Wireless Sensor Networks (WSN) sensor nodes sense the same kind of data and forward it to the sink node. Sink node wastes its energy in processing this redundant data. So to improve lifetime of network there is need of redundancy elimination. Following are some methods for redundancy elimination.

Data Aggregation at intermediate nodes: Data aggregation technique takes multiple data packets as input performs aggregation and produce single packet as output. Above mentioned approach performs data aggregation not only at sink but also at intermediate nodes. This aggregation reduces redundancy and also traffic in the network. E.g. Dynamically Balanced Spanning Tree (DBST) [2]

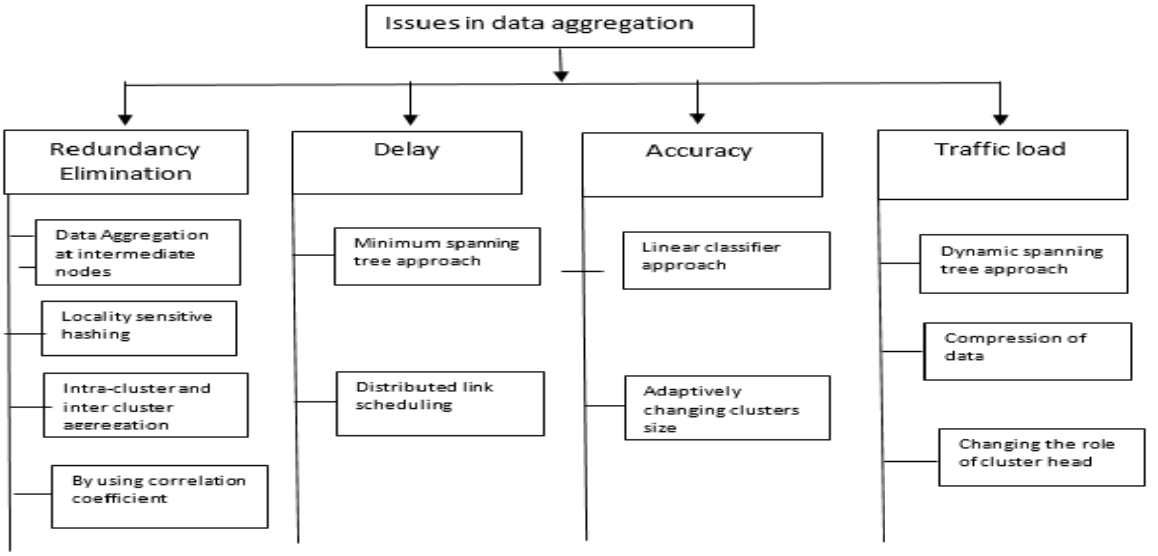

Figure 7: Metho
ds to solve data aggregation issues

Locality sensitive hashing: This method generates the hash code which is small in size than data. These LSH codes are sent to the supervisor node. When data is to be forwarded supervisor node selects one of the sensor node based on the similarity of hash code. Aggregation supervisor further maintains redundancy count for similar LSH codes. The similarity count below or equal threshold is accepted and those who have more than threshold are rejected. In this way supervisor node eliminates outliers. E.g. SVM based Data Redundancy Elimination (SDRE) [3]

Intra cluster and inter cluster aggregation: In this approach data compression is done at source nodes and then compressed data is forwarded into cluster. Redundancy inside the cluster is eliminated at the cluster head node. Intra cluster aggregation is shown in following figure 8.
After intra cluster aggregation, aggregation is done in between clusters called inter cluster aggregation. In this approach data compression is done at cluster head nodes and then compressed data is forwarded into network. E.g. Redundancy Eliminated Data Dissemination (REDD) [5], Bandwidth efficient Heterogeneity aware Cluster based Data Aggregation (BHCDA) [4].

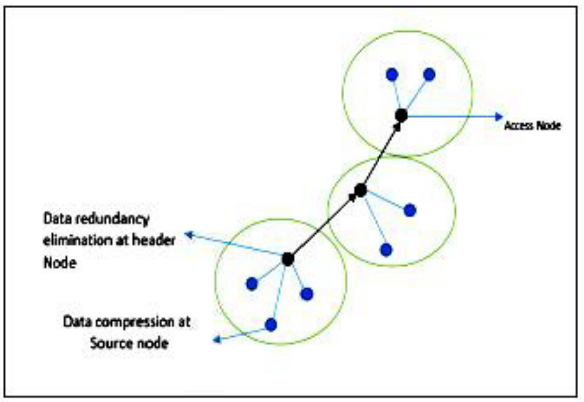

Figure 8: intra cluster aggregation

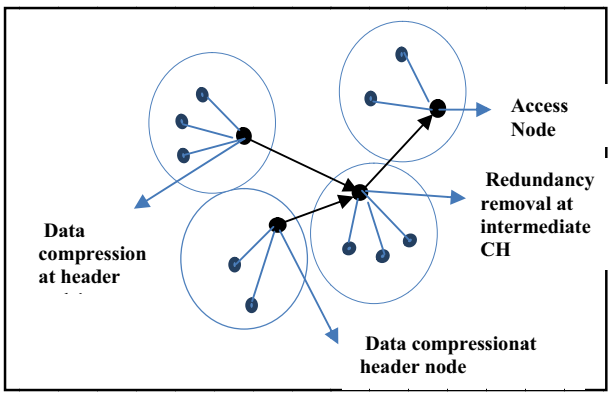

Figure 9: inter cluster aggregation

Correlation coefficient: Correlation coefficient is an integer used to represent the whole set of readings recorded by all the nodes in the sensor field. The value of the correlation coefficient (H) ranges from 1 to10. H = 1 is for strong correlation of data and as the correlation coefficient increases, the degree of correlation between data decreases. So if value of correlation coefficient is greater than 3 then both data are kept. But if the value of correlation coefficient is less than 3 then there is need of normalization of data redundancy. E.g. REDD [5]

4.2. Delay

Delay is nothing but time taken by receiver to receive a packet is higher than the time taken by the sender to send a packet. Data aggregation functions reduces redundancy but may increase delay because data from nearer sources may have to be held back at an intermediate node in order to be aggregated with data coming from sources that are farther away. In the worst case, the latency due to aggregation will be proportional to the number of hops between the sink and the farthest source. Tree based methods used to reduce delay are as follows.

Distributed data aggregation scheduling: In the distributed data aggregation scheduling algorithm the time slot is assigned to each and every node in the aggregated tree. The node will transmit their data during the assigned time slot. E.g. SDRE [3], Delay Efficient Data distributed Aggregation (DEDA) [8].

Minimum spanning tree creation: In tree based network when source node wants to send the data to destination, data may flow from various paths consumes more bandwidth and requires more delay to reach to destination. Minimum spanning tree provide a path between source to destination which has minimum weight. This path will provide fast delivery of data to destination which reduces delay DBST[2].

4.3. Accuracy

In Data Aggregation process aggregator node performs aggregation. This aggregated data is then forwarded in to the network. So if this aggregator node gets compromised or there is any compromised nodes repeatedly sending wrong data, may leads to inaccuracy. Following are some methods that validates the sensor nodes and allows only valid nodes to forward data in order to provide accuracy in aggregation process.

Linear classifier method: This method divides the hyper plane of sensor nodes in two linear partitions. One contains valid nodes and other contains invalid. This validity is checked by using redundancy count of each sensor node. This count is maintained by supervisor node based on similarity of hash code. So this method allows only valid nodes to forward the data. So accuracy is maintained [3].

Context aware system: Context aware system contains one rule engine. This rule engine contains set of rules. Data sensed by sensor nodes are checked against the rules of the rule engine. If provided context satisfies the rules then the data is transmitted else it is rejected. This system keeps accuracy intact. This approach is used in REDD [5].

Adaptively changing cluster size: To avoid false data, another approach used is changing the size of the cluster. The Cluster Head (CH) performs aggregation and sends this aggregated data to Coordinator node(CN). The CN calculates Loss Ratio (LR) which is amount of data packets forwarded to the data packets received. This value of LR is forwarded to CH. Based upon this loss ratio, the size of cluster is changed that is the number of forwarding nodes added or removed. As the size of cluster is changing adaptively outliers are get removed. Reliability is increased due to altering the cluster size before data transmission and also LR is measured at CN itself therefore, energy consumption is effectively reduced Eg. Energy Efficient Reliable Data Aggregation Technique (EERDAT) [6].

4.4. Traffic load

In wireless sensor network the responsibility of data aggregation is given to the cluster head. If this role of cluster head is fixed to one node then that node suffers from heavy traffic load. The node wastes maximum of its energy due to this traffic. There is one mechanism using which traffic load is get distributed among all nodes equally.

Dynamic spanning tree approach: In this approach for each round new aggregator node is selected based up on the residual energy and then new data aggregation tree is formed. This tree is minimum spanning tree which is

routed through this aggregator node. This approach keeps the role of aggregation function rotating between the nodes. Due to which traffic load on one node gets distributed in all nodes equally. Eg. DBST[2].

Distributing role of cluster head: The role of cluster head is distributed in other nodes based on parameters like residual energy or distance. After each round, each node calculates its residual energy and based on that the node with highest residual energy is selected as Cluster head. In this way the role of cluster head is distributed among all nodes. This method reduces load from single node and distribute it with others. E.g.BHCD [4], EERDAT[6].

5. Comparative Study of Data Aggregation Technique

In this section we have done the comparative study of data aggregation techniques. These techniques are mainly focusing on to address the above stated issues like redundancy, delay, accuracy and traffic load. Techniques we have adopted for study includes Dynamically Balanced Spanning Tree (DBST) which provides dynamic structure of tree to solve hotspot problem and improve energy conservation. SVM based Data Redundancy Elimination (SDRE) algorithm makes use of SVM for redundancy elimination [3]. Bandwidth efficient Heterogeneity aware Cluster based Data Aggregation (BHCDA) which performs both intra cluster and inter cluster aggregation to eliminate redundancy [4]. Redundancy Eliminated Data Dissemination (REDD) algorithm makes use of context aware system for validation and correlation coefficient is used to eliminate redundancy from valid data [5]. Adaptive Energy Efficient Reliable Data Aggregation Technique (AEERDAT) this algorithm changes the size of cluster and allows only valid nodes to be in the cluster in order to maintain reliability [6]. Energy Efficient and Balanced Cluster-Based Data Aggregation Algorithm for Wireless Sensor Networks (EEBCDA) and Delay Efficient Distributed Data Aggregation (DEDA) [8].
Table1 shows the comparison of techniques based on strategy, delay, average energy consumption and traffic load.

Table 1: Comparison of data aggregation techniques.

Techniques	Strategy	Delay	Redundancy	Accuracy	Avg. Energy consumption	Traffic load
DBST	Tree Based	Moderate	Less	Moderate	Less	Less
SDRE	Tree Based	Less	Less	High	Less	Moderate
BHCDA	Cluster Based	Moderate	Less	Less	Less	Less
REDD	Cluster Based	Less	Less	Moderate	Less	Moderate
EERDAT	Cluster Based	Less	Moderate	High	Less	Moderate
EEBCDA	Cluster Based	Less	Less	High	Less	Less
DEAD	Tree Based	Less	Moderate	Moderate	Moderate	Moderate

6. Proposed Model:

In Wireless Sensor Network there should be some mechanism that solves the data aggregation issues. We have proposed one model which performs data aggregation at two levels one at cluster head and another at storage node. At cluster head we are using Locality Sensitive Hash codes which shows the similarity between the data or data redundancy.We are not eliminating all redundant data here, we are keeping data redundancy up to some adequate level so that accuracy will be intact and then we will forward it towards storage node. Storage node will performs the same as cluster head did and forwards data to the base station. Storage node is a node which has highest residual energy, this storage node will get changed when another node will have residual energy more than the present storage node. In this way we can balance the load on one storage node to other nodes. This system will try to maintain tradeoff between energy efficiency and accuracy also it balances the traffic load. Following figure shows the proposed model

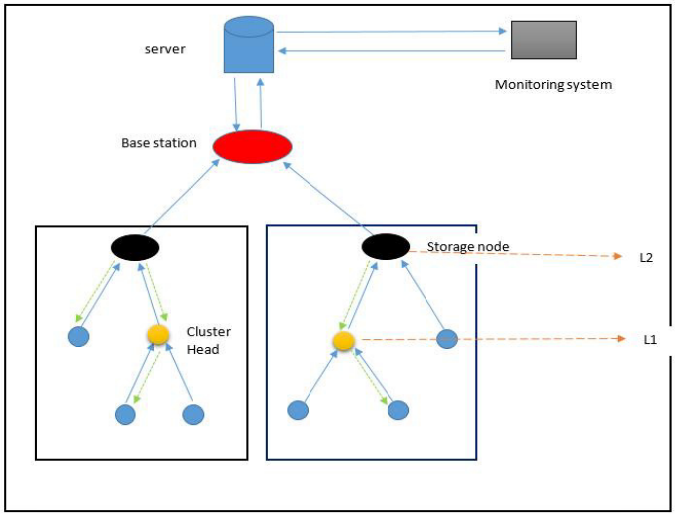

Figure 10: proposed model

7. Conclusion:

Data aggregation technique removes data redundancy from Wireless Sensor Networks (WSN) in order to enhance the lifetime of WSN. Data aggregation technique has issues like redundancy, delay, accuracy and traffic load that are need to be addressed. This work presents some methods to solve these issues and which are helpful for improving performance of data aggregation techniques. Further we did comparative study of various data aggregation techniques and based on that study we proposed one model for data aggregation. Proposed model uses multilevel data aggregation approach and tries to solve all the issue of data aggregation.

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.