

# Tree Based Data Aggregation Algorithm to Increase the Lifetime of Wireless Sensor Network

P.Parthasarathy<sup>1</sup>, R.Karthickeyan<sup>2</sup>

Asst. Professor, Department of ECE, Maharaja Prithvi Engineering College, Avinashi, Tamilnadu, India<sup>1</sup>

PG Scholar, Sri Krishna College of Technology, Coimbatore, Tamilnadu, India<sup>2</sup>

**Abstract:** Wireless sensor networks are becoming increasingly important in recent years due to their ability to detect and convey real time information for many civilian and military applications. The most important challenge faced by WSNs is battery-limited lifetime of the network. Physically replacing batteries is infeasible in most real-life deployments of WSNs. Longer lifetime of the devices can be achieved by reducing the delay and the energy consumption. Hence this project aims to reduce energy during delay on collection of data, and hence by this process gathering information by sensor nodes is achieved. There are strong needs to develop wireless sensor networks algorithms with optimization priorities biased to aspects besides energy saving. In order to minimize delays in the data collection process of wireless sensor networks, clustering is used. This exploits use of clustering for reducing the delay during data collection process. The two network formation algorithms, centralized top-down and decentralized bottom-up clustering is considered and analyzed for shortening delays in the data collection process for Single and Multiple-cluster network structure. Cluster data fusion and routing algorithms are considered with data packets generated by sensor nodes.

**Keywords:** Centralised and decentralised approach, multiple cluster data fusion

## I. INTRODUCTION

A wireless sensor network (WSN) is a wireless network consisting of spatially distributed autonomous devices that use sensors to monitor physical or environmental conditions. These autonomous devices, or nodes, combine with routers and a gateway to create a typical WSN system. The distributed measurement nodes communicate wirelessly to a central gateway,

which provides a connection to the wired world where it can collect, process, analyze, and present measurement data. To extend distance and reliability in a wireless sensor network, it can use routers to gain an additional communication link between end nodes and the gateway. Advances in wireless communications and miniature, low-power, and low-cost sensors are enabling the deployment of large-scale and collaborative wireless sensor networks (WSNs). To increase the network lifetime two clustering algorithms are proposed such as top-down approach and bottom-up approach that will provide a better performance. Strong adaptability, comprehensive sensing coverage, and high fault tolerance are some of the unique advantages of wireless sensor networks. Wireless sensor networks consist of large amounts of wireless sensor nodes, which are compact, light-weighted, and battery-powered devices that can be used in virtually any environment. Because of these special characteristics, sensor nodes are usually deployed near the targets of interest in order to do close-range sensing. The data collected will undergo in-network processes and then return to the user who is usually located in a remote site. Most of the time, wireless sensor nodes are located in extreme environments, where are too hostile for maintenance. Sensor nodes must conserve their scarce energy by all means and stay active in order to maintain the required sensing coverage of the environment. Much prior work has focused on conserving energy by clustering. A network with clustering is divided into several clusters. Within each cluster, one of the sensor nodes is elected as a *cluster head* (CH) and with the rest being *cluster members* (CM). The cluster head will collect data from its cluster members directly or in a multihop manner. By organizing wireless sensor nodes into clusters, energy dissipation is reduced by decreasing the number of nodes

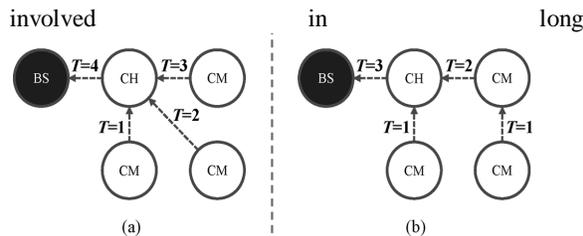


Fig.1(a) Data collection in a two-hop network and (b) Data collection in an improved multihop network. (A dashed arrow represents the existence of a data link and the direction of the arrow shows the direction of data flow)

distance transmission. Clustering provides a significant improvement in energy saving. In sensor networks with cluster, it is common for a cluster head to collect data from its cluster members one by one. Let  $T$  be the average transmission delay among nodes. Referring to the situation shown in Fig.1 (a), a base station will take  $4 \times T$  to collect a complete set of data from the network. By transforming the network into a multihop network, as shown in Fig.1 (b), it can be shown that the time needed by the base station to collect a full set of data from the network can be reduced to  $3 \times T$ . In the modified network, apart from requiring a shorter delay in data collection, cluster members will need smaller buffers to handle the incoming data while waiting for the belonging cluster head to become available.

The aim of this project is to investigate the characteristics of a delay-aware data collection network structure in wireless sensor networks. Two algorithms for forming such a network structure are proposed for different scenarios. The proposed algorithms are operating between the data link layer and the network layer. The algorithms will form networks with minimum delays in the data collection process. At the same time, the algorithms will try to keep the transmission distance among wireless sensor nodes at low values in order to limit the amount of energy consumed in communications.

## II. RELATED WORKS

In order to reduce energy, clustering is considered in this paper. The data collected by each sensor is communicated through the network to a single processing centre that uses the data. Clustering sensors into groups such that sensors communicate information only to cluster heads and then the cluster heads communicate the aggregated information to the processing centre, saving energy

Rank	1	2	...	$\log_2 N - 1$	$\log_2 N$
No. of nodes	$\frac{N}{2^1}$	$\frac{N}{2^2}$	...	$\frac{N}{2^{(\log_2 N - 1)}}$	$\frac{N}{2^{(\log_2 N)}}$

TABLE I

Cluster members rank distribution in the proposed network structure with network size  $N = 2^P$ , where  $N = 1, 2, \dots$

and bandwidth. The cost of transmitting a bit is higher than a computation; therefore, it may be beneficial to organize the sensors into clusters. Cluster-based control structures provide more efficient use of resources in wireless sensor networks. Clustering can be used for 1) Transmission management 2) Backbone formation 3) Routing Efficiency. Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. A loose definition of clustering could be “the process of organizing objects into groups whose members are similar in some way”

Intensive researches on clustering algorithms are found for energy efficiency in sensor network. Few of the important algorithms are Heinzelman *et al.* proposed a clustering algorithm called LEACH [2]. In networks using LEACH, sensor nodes are organized in multiple-cluster 2-hop (MC2H) networks (i.e., cluster members cluster head base station). Using the idea of clustering, the amount of long distance transmissions can be greatly reduced. Lindsey and Raghavendra proposed another clustering algorithm called PEGASIS [3], which is a completely different idea by organizing sensor nodes into a single chain (SC) network. In such networks, a single node on the chain is selected as the cluster head. By minimizing the number of cluster heads, the energy consumed in long distance transmission is further minimized.

## III. THE PROPOSED NETWORK STRUCTURE:

The proposed network structure is a tree structure. To deliver the maximum data collection efficiency, the number of nodes  $N$  in the proposed network structure has to be restricted to  $N = 2^P$ , where  $P = 1, 2, \dots$ . It will be shown in a later part that such restriction can be relaxed by giving up some performance. Each cluster member will be

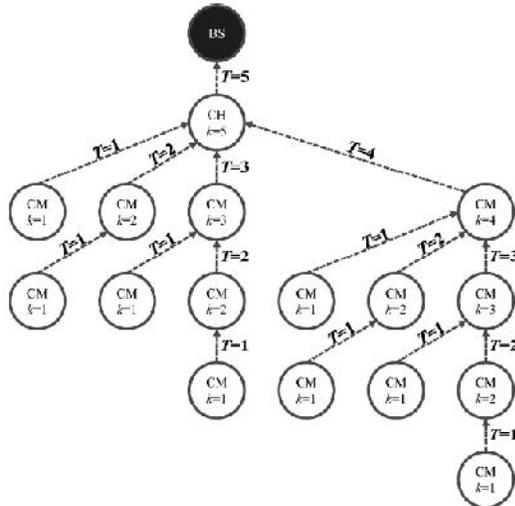


Fig. 2 Proposed network structure with network size  $N = 16$ .

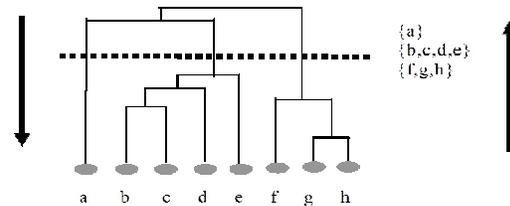
given a rank, which is an integer between 1 and  $P$ . A node with rank will form  $K-1$  data links with  $K-1$  nodes, while these nodes are with different ranks starting from 1, 2 ... up to  $K-1$ . All these  $K-1$  nodes will become the child nodes of the node with rank  $K$ . The node with rank  $K$  will form a data link with a node with a higher rank. This higher rank node will become the parent node of the node with rank  $K$ . The cluster head will be considered as a special case. The cluster head is the one with the highest rank in the network. Instead of forming a data link with a node of higher rank, the cluster head will form data link with the base station. By following this logic, the distribution of the rank will follow an inverse exponential base-2 function, as shown in Table I.

#### IV. HIERARCHICAL CLUSTERING

A cluster tree can be formed using either a top-down or a bottom-up approach as in figure 4.1. In top-down approach, a designated root node first forms its own cluster. It then selects some of its neighbours (as CHs) to form their own clusters, which in turn causes some of their neighbours to form the next level of clusters, and the process continues until the entire sensor field is covered.

The cluster tree is formed by keeping track of the parent-child relationship among CHs, and it is guaranteed to be connected as new child CHs are selected from neighbours of existing CHs. The IEEE 802.15.4 cluster tree follows this concept.

#### TOP-DOWN approach



#### BOTTOM-UP approach

#### TOP DOWN APPROACH:

Top-Down Starts with one large cluster consisting of all the data points and then recursively splits the most-‘appropriate’ cluster. top-down method starts by considering the entire data as one cluster and then splits up the cluster(s) until each object forms its own cluster. The root node, one of the sensor nodes or a resourceful base station, initiates the cluster formation process by sending a broadcast indicating its interest to form a cluster. Nodes that hear the broadcast become members of the cluster by sending an acknowledgment (ACK) to the root node. Coverage of a cluster depends on whether a single-hop or a multihop cluster is formed by forwarding the broadcast through neighbours.

The parameter hops max is used to provide a bound on the number of hops to a cluster member from the CH. After receiving ACKs from neighbours, the root node selects several new child CHs from subset of the nodes closer to the cluster boundary. Overlap among clusters can be reduced by selecting child CHs from nodes that are outside the cluster boundary. The root node then requests those selected child CHs to form their own clusters in turn by sending a unicast message to each of them. New CHs form their own clusters and then select the next set of child CHs and the process continues. Meanwhile, the cluster tree is formed by each CH keeping track of its parent and child CHs. The algorithm continues until all the possible clusters are formed.

The top-down approach is a kind of centralized control algorithm. In this approach, the base station is assumed to have the coordinates of all sensor nodes in the network. The whole approach is going to be executed at the base station. At the end of the optimization

process, the base station will instruct the sensor nodes to establish the essential data links and form the appropriate network structure.

For networks with nodes  $N = 2^P$ , where  $P=2, 3, \dots$ , the proposed network structure can be constructed according to the following algorithm.

1) The algorithm starts with considering the whole network as a fully connected network. The connection degree of a wireless sensor node is telling the number of data links associated with such node. A node with connection degree of 3 implies that such a node has formed three data links with three other nodes.

For a network of nodes  $N = 2^P$ , where  $P=2, 3, \dots$ , each node will begin with degree equal to  $N - 1$ . The nodes will form the set  $\hat{H}_{s=1}$ . Set  $b = N/2$ .

2) Select nodes from set  $\hat{H}_s$  to form set  $H_{s+1}$ , such that  $\sum_{i,j \in H_{s+1}} d_{ij}^2$  is maximized. Here  $d_{ij}$ , denotes the geographical distance between node  $i$  and node  $j$ . The rest of the nodes from  $\hat{H}_s$  will form set  $\hat{H}_{s+1}$ . The algorithm will then remove all connections (data links) among nodes within  $H_{s+1}$ . Set iterators  $s \leftarrow s + 1$  and  $b \leftarrow b/2$ .

3) Repeat step 2 until  $b < 2$ . Set  $r = 2$ .

4) Nodes with degree  $N - r$  form set  $L$ . Nodes with degree greater than  $N - r$  form set  $U$  such that set  $L$  and set  $U$  are of the same number of nodes. Connections among nodes in the two sets are reduced until each node in set  $L$  is only connected to a single node in set  $U$ . Here, data links are removed according to their distance. Details of the optimization method are given in the later part of this section. After reducing the number of connections, set  $r \leftarrow r \times 2$ .

5) Repeat step 4 until  $r = N$ .

The algorithm works in such a way that it will find a better way to form the cluster as in figure 4.3 and it has to communicate with the base station for sending the data from the cluster head that formed as a result of algorithm.

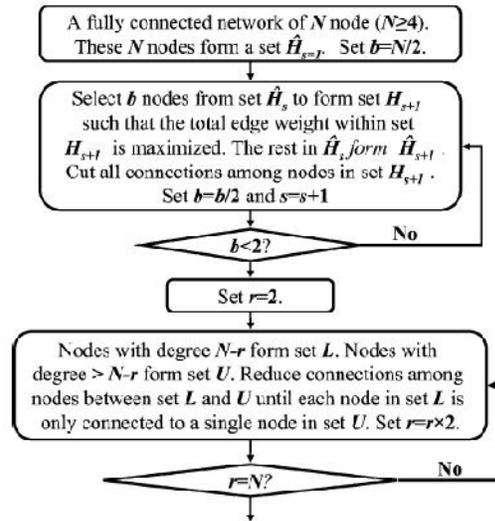


Fig: 3 Network formation of the proposed network structure using centralized Top-down approach ( $N \geq 4$ )

**BOTTOM-UP APPROACH:**

In bottom-up approach, individual clusters are formed independently and later combined together to form a higher-level structure. Bottom-Up Starts with single-point clusters and then recursively merges two or more of the most-'appropriate' clusters. Bottom-up method starts with each object in the data forming its own cluster, and then successively merges the clusters until one large cluster is formed, which encompasses the entire dataset. sequential algorithms for top-down and bottom-up hierarchical clustering.

The operation of the bottom-up approach is to join clusters of the same size together. The bottom-up approach is, when comparing with the top-down approach, more scalable. It can be implemented in either centralized or decentralized fashion. Specifically, a decentralized bottom-up approach can be described as follows.

1) Each node is labelled with a unique identity and marked as level. The unique identity will only serve as an identification which has no relation with sensor nodes' locations and connections. Here, is a function which represents the number of nodes in a cluster. For a cluster of  $i$  nodes, its  $W$  value is equal to  $\log_2 i$ . Since nodes are disconnected initially (i.e., no data link exists among wireless sensor nodes), these  $N$  nodes can be considered as  $N$  level-0 clusters. Within each cluster, one node will be elected as the sub-cluster

head. We denote SCH ( $w$ ) as a sub cluster head of a level-  $w$  cluster. In the bottom-up approach, a SCH can only make connection (i.e., setup a data link) with another SCH of the same level. Since there is only 1 node in each cluster, all nodes begin as SCH (0). The dimensions of the terrain  $(t_x, t_y)$  are provided to the sensor nodes before deployment.

2) Each SCH performs random back off and then broadcasts a density probing packet (DPP) to its neighbouring SCHs which are within a distance of  $r_{dp} = (t_x^2 + t_y^2)^{1/2} n$ . Note that the size of a DPP is much smaller than that of a data packet. A SCH can use the number of received DPP, together with the dimensions of the terrain, to estimate the total number of nodes ( $N_{est}$ ) in the network. A SCH will use the  $N_{est}$  to adjust its communication distance  $r_{com}$ .

3) Each SCH will do a random back off and then broadcast an invitation packet (IVP) to its neighbours within  $r_{com}$ . The IVP contains the level and the identity of the issuing SCH. A SCH will estimate the distances to its neighbouring SCHs using the received signal strength of the IVPs received. A SCH will also count the number of IVPs received. If the number of IVPs received has exceeded a predefined threshold or a maximum duration has been reached, a SCH will send a connection request (CR) to this nearest neighbour. If both SCHs are the nearest neighbour of each other, a connection will be formed between these two SCHs.

4) Once they are connected, the two SCHs and their belonging level- clusters will form a composite level-  $w+1$  cluster. One of the two involved SCHs will become the chief SCH of the composite cluster. The chief SCH will listen to the communication channel and reply any CR from lower levels with a rejecting packet (RP). When no more CR from lower levels can be heard, the chief SCH will start to make connection with other SCHs of the same level.

5) If a RP is received, a SCH will send a CR to its next nearest neighbour in its database. If such neighbour does not exist, the SCH will increase its  $r_{com}$ . The SCH will then broadcast a CR using the new  $r_{com}$ . Upon receiving the CR, a SCH of the same level will grant the request if it is still waiting for a CR.

6) If no connection can be made within a period of time, either all neighbours of the same level are unavailable or all CRs have been rejected, the SCH will increase its  $r_{com}$  and broadcast the CR again. This process repeats as long as  $r_{com} < \sqrt{t_x^2 + t_y^2}$ . If  $r_{com} < \sqrt{t_x^2 + t_y^2}$  the SCH will make connection with the base station directly.

7) The above processes continue until no more connection can be formed. In the bottom-up approach, the communication distance is defined as Equation Number [1]

$$r_{com} = \frac{\sqrt{t_x^2 + t_y^2}}{\alpha - \beta - w}, \quad \beta + w < \alpha. \quad [1]$$

Here  $\beta$ , is a constant which is set to 0 initially. Parameter  $\alpha$  is the estimated maximum rank of a node in the network, which is expressed as Equation Number [2]

$$\alpha = \lceil \log_2(N_{est}) \rceil + 1 \quad [2]$$

Initially, all SCHs are with  $w = 0$  and  $\beta = 0$ . Therefore, the SCHs will start broadcasting their IVPs with

$r_{com} = (\sqrt{t_x^2 + t_y^2})(\alpha - 0 - 0)^{-1}$ . If a SCH has made a connection with another SCH, its level will be increased by 1 (i.e.,  $w = 1$ ). After that, the chief SCH of the composite cluster will broadcast its IVP with  $r_{com} = (\sqrt{t_x^2 + t_y^2})(\alpha - 0 - 0)^{-1}$ . The  $r_{com}$  is

designed to be increased with  $w$  because when SCHs are paired up to form composite clusters, the average separation among composite clusters will be increased. It is more energy efficient to start the broadcasting with a longer communication range. However, if no connection can be made, a SCH will increase its  $\beta$  by one. This will increase  $r_{com}$ , which can facilitate the searching of available SCHs. A SCH will increase its  $r_{com}$  by incrementing  $\beta$  until a connection can be made. The sum of  $\beta$  and  $w$  is defined to be smaller than  $\alpha$  to ensure  $r_{com}$  is upper bounded by the diagonal of the sensing terrain.

In step 3 of the bottom-up approach, a SCH will send a CR to its nearest neighbour if the number of received IVPs has exceeded  $N$ . Here,  $N$  is the expected number of IVPs to be received which is expressed as Equation number [3]

$$N = \left\lceil \frac{\eta \pi r_{com}^2 - 1}{w} \right\rceil \quad [3] \quad \text{Parameter}$$

$\eta$  is the density of the network which can be estimated using the obtained before. When being implemented in a decentralized control manner, the above algorithm may end up with multiple composite clusters if the number of nodes is not equal to  $2^p$ , where  $p = 1, 2, \dots$ . SCHs of these composite clusters will communicate with the base

station directly. By virtue of pairing up composite clusters of same sizes, the algorithm will end up with composite clusters of completely different sizes. Considering the base station as the root of the network, the number of time slots required by the base station to collect data from all sensor nodes is given in Equation number [4]

$$t(N) = \lceil \log_2 (N + 1) \rceil \quad [4]$$

In contrast, the above algorithm can also be carried out at the base station as a centralized control algorithm. The base station is again assumed to have the coordinates of all sensor nodes in the network. When the number of nodes is not equal to  $2^P$ , where P = 1, 2..., dummy nodes can be virtually added in the calculation process, depending on the application. If a single cluster is required, dummy nodes should be virtually added to fulfil the requirement of  $N = 2^P$ , where P = 1, 2.... However, if multiple clusters can be formed, dummy nodes are not essential. When dummy nodes are virtually added, these dummy nodes will have infinite separations with the real nodes and with themselves. Note that whenever a real SCH is connecting with a dummy SCH, the real SCH will always be the chief SCH of the composite cluster. This is to ensure the removal of dummy nodes at the end of the calculation process will not partition the network.

**V. RESULT ANALYSIS:**

**DATA GATHERING TIME**

The graphical result in figure 4 shows that the number of time slots needed for collecting data packets from nodes. The plots for both top-down (centralized) and bottom-up (decentralized) method. As can be seen from the figure that the time slots needed remains the same for both the methods.

The average data collection time for the top down and bottom up approaches are plotted and table 2 with the values of number of nodes and the corresponding time slots with the range of values the time slots increased with increase in number of nodes. Hence by this the data gathering is done successfully by the algorithms.

Figure 4 is plotted with Y Axis as time slots and X Axis as number of node hence the time for data gathering is plotted which shows a slight variation between top-down and bottom-up approach

TOP-DOWN ALGORITHM		BOTTOM-UP ALGORITHM	
Number of nodes	Number of time slots	Number of nodes	Number of time slots
10	5	10	5
20	7	20	7
30	7	30	7
40	8	40	8
50	8	50	8
60	8	60	9

Table 2: Data gathering time



FIG 4: Data gathering time.

**AVERAGE LIFE TIME**

The average life time of the top-down and bottom-up approach are plotted and the values for the plot are given in table 3

TOP-DOWN ALGORITHM		BOTTOM-UP ALGORITHM	
Number of nodes	Number of rounds to form cluster	Number of nodes	Number of rounds to form cluster
10	170	10	150
20	150	20	130
30	135	30	130
40	125	40	110
50	125	50	110
60	120	60	110

TABLE 3: Number of Nodes vs. cluster formation time

The values of number of nodes and number of rounds in the table 3 gives the information that as the

number of node increases the number of rounds decreases so that the energy utilization is reduced greatly.

It is shown in Figure that the average life time of the nodes with respect to the no. of rounds of data collections. It is observed that ,the top-down approach is having an advantage over the bottom-up approach due the fact that it is centralized where as the bottom-up is de-centralized.



Fig 5: Average life time to form cluster

## VI.CONCLUSION

In this paper, Tree Based data aggregation Algorithm to Increase the Lifetime of Wireless Sensor Network cluster is implemented using NS2 and formation algorithms are proposed and analyzed. The network formation can be implemented in either centralized or decentralized manner. Two network formation approaches, namely top-down and bottom-up are derived to provide optimized results for networks with different sizes. The proposed network algorithm is shown to be the most efficient. The proposed network algorithm can greatly reduce the data collection time by forming the clusters and the cluster head only will transmit the data and the energy is more efficiently saved. While keeping the total communication distance and the network lifetime at acceptable values. In the proposed work, it is observed that the top-down approach is slightly better in terms of cluster formation time, which may also increase the lifetime of the network as the data transmission is only from cluster head to the base station.

## REFERENCES

[1] J. N. Al-karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: A Survey," IEEE Wireless Commun.Mag., vol. 11, no. 6, pp. 6–28, Dec. 2004.

[2] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application- specific Protocol architecture for wireless micro sensornetworks," IEEE Trans. WirelessCommun, vol. 1, no. 4, pp.660–670, Oct. 2002.

[3] S. Lindsey and C. S. Raghavendra, "PEGASIS: Power-efficient gathering in sensor Information systems," in Proc. IEEE Conf. Aerosp., Big Sky, MT, USA, Mar. 2002, vol. 3, pp. 1125–1130.

[4] "Efficient Clustering for Improving Network Performance in Wireless Sensor Networks" Tal Anker1, 2, Danny Bickson1, Danny Dolev1 andBracha Hod1

[5] "Energy-Efficient MAC Protocol for Wireless Sensor Networks - A Review" Smriti Joshi & Anant Kr. Jayswal,International Journal of Smart Sensors and Ad Hoc Networks (IJSSAN) ISSN No. 2248-9738 Volume-1, Issue-4, IEEE 2012

[6] "Online Incremental Clustering with Distance Metric Learning for High DimensionalData" Proceedings of International Joint Conference on Neural Networks, San Jose, California, USA, July 31 – August 5, IEEE 2011