Evenly Distributed Clusters over a Target Area of Wireless Sensor Network

Kaushlendra Kumar Sinha, Maheshwari Prasad Singh

Abstract**:** ***Wireless sensor networks (WSNs) finds wide applications in various fields. The most important problem faced by these networks is low lifetime. These are generally battery powered devices with ability to communicate with each other. Networks should be designed so that load is equally distributed among the nodes. In WSNs maximum load is on nodes being cluster heads, so for proper load distribution various nodes should get chance of becoming cluster head. Further entire network should have proper connectivity that is clusters should be evenly distributed throughout the network. To achieve this paper discusses algorithm to get dominating sets in a fully connected network. Dominating sets ensure that either a node is a cluster head or is adjacent to a cluster head. This leads to even distribution which may increase the lifespan of entire network. Not much attention has been given to even distribution of clusters. WSNs consists of spatially distributed nodes over a target area with sensing and communication facility. Purpose of these nodes is to study the entire area and communicate their observation to the central base station. This work presents an idea to form evenly distributed clusters. Even distribution in necessary for proper load sharing and prolonging the life of network. It needs much more emphasis than given to it. Further Ranking methodology has been discussed to rank the dominating sets based on certain parameters. This ranking methodology is used to determine which dominating set should become cluster heads ensuring even distribution. Ranking methodology comes into play if more than one dominating sets are obtained. These may be used where it is difficult for humans to physically visit on a regular basis. High lifetime of network ensures less physical presence of humans.***

Index Terms**:** ***Cluster head (CH), Dominating set (DS), Wireless sensor network (WSN).***

I. INTRODUCTION

WSNs play a vital role in the actual operation of many applications such as health monitoring, smart transportation, war zone surveillance, weather forecast, satellite communication etc. Its key features are low powered radio and sensing capability. These make it an alternate of human presence, especially in places or conditions where human presence is not viable.

Clustering is one of the methods used for prolonging the lifetime of WSNs, involving grouping of nodes into clusters and selecting CHs for all the clusters. Group of nodes and CH is called cluster. CHs collect the data from respective cluster's nodes, aggregates them and forwards to base station. One of the major challenges in WSNs is to select appropriate CHs. Generally, Cluster head (CH) is the one which consumes maximum amount of energy. So, CHs should be changed from time to time to manage energy consumption.

Revised Manuscript Received on July 05, 2019.

Kaushlendra Kumar Sinha, Department of Computer Science and Engineering, NIT, Patna, India.

Maheshwari Prasad Singh, Department of Computer Science and Engineering, NIT, Patna, India.

WSNs have many issues like coverage, security, energy-efficiency, localization, etc. Among them energy-efficiency is the major issue, as nodes are battery operated. Depending upon the requirements, WSN may be deployed in the unattended area or harsh environment where human presence is not feasible.

There are many algorithms for selection of CHs. Most of them do not deal with even distribution of clusters. This paper presents algorithm which mainly deals with the even distribution of clusters over given target area. This algorithm makes sure that each node will be the part of one of the clusters.

WSNs commonly use radio model for communication. This model is discussed in [8], [7], [9], [12]. This model explains how energy is consumed in our radio communication. Radio model uses low frequency radio waves for communications. These are low powered waves not requiring much energy thus allowing much longer time to these battery powered devices.
Given equation shows energy consumed in transmitting l-bit data to distance d.

$$E_{TX(l,d)} = \begin{cases} l.E_{elec} + l.\varepsilon_{fs}.d^2, d < d_c \\ l.E_{elec} + l.\varepsilon_{mp}.d^4, d \geq d_c \end{cases} \quad (1)$$

Where E_{elec} is energy dissipated due to circuitry, $d_c = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$, ε_{fs} is the free space coefficient and ε_{mp} is multipath coefficient. Energy consumed in receiving l-bit data at distance d is given in Equation (2).

$$E_{RX(l,d)} = l.E_{elec} \quad (2)$$

Here, energy consumed in receiving is not dependent on the distance. In radio model while transmitting data, energy is consumed in both radio circuit as well as in communication whereas while receiving energy is consumed in radio circuitry only.

II. RELATED WORK

LEACH[8],[7] which selects CH based on random variable does not makes sure that nodes are evenly distributed. This leads to the formation of orphan nodes. Orphan nodes are the ones which do not have a CH even though it is connected to the net-work. **O-LEACH[9]** discusses about various methods to deal with orphan nodes but totally avoiding orphan nodes is not mentioned in it. It leads to the working through gateway formation for orphan nodes. Here orphan nodes communicate through a gateway. This gateway could be either a node among

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

themselves (bunch of nearby orphan nodes) or a nearest node which is a part of cluster. This however does not totally solve the problem as there could be many nodes not having cluster. Gateways thus formed may have too many nodes to communicate with which will have adverse effect on lifespan. **EHA-LEACH[12]** gives the concept of neighboring nodes in selection of CH in a nearby area but energy harvesting nodes sensor nodes comes with various complex calculations in every step. **LEACH-MAC[3]** also comes with various complex calculations. LEACH-MAC does not talk about the case of uneven distribution neither does it clarify what to do under these situations. The clustering scheme used in [2] uses three parameters in formation of CHs. These parameters are remaining energy, degree of nodes and distance from cluster centroid. Now based on these parameters and applying fuzzy logic every node is given a probability of it becoming a CH. More number of parameters involved in formation of CH increases complexity. Also even distribution is not mentioned in this scheme. **LEACH-DCHS[6]** and **ME-LEACH[4]** modifies the threshold calculation criteria using the energy of the nodes. Formula uses remaining energy criteria for CHs selection. This scheme also leaves scope for uneven distribution. **SLEACH[5]** includes security features in LEACH using encryption and decryption. It adversely affects the energy of the network. To deal with energy efficiency problem caused by SLEACH another scheme **A-LEACH[1]** is introduced. However basic part remains that of LEACH causing random selection of nodes does not change. In **LEACH-GA[10]** the probability parameter is improved but threshold calculation parameter remains the same. **MOD-LEACH[11]** uses two types of signal amplification for inter and intra cluster communication. Low strength signals are used for intra cluster and high strength signals are used inter cluster communications. This allows nodes to save huge amount of energy. Overall not much discussion or work has been done on uneven distribution of clusters.

III. Problem & Solution

Uneven distribution may have adverse effects on lifespan of networks. And hence, this paper formulates the problem as "Formation of clusters to cover entire target area, so that none of the sensor nodes are isolated in the network."

The network is considered as a graph. Each node acts as a vertex of graph and communication between nodes acts an edge between them. This graph can be represented as an adjacency list in memory. This adjacency list consists of the nodes with which that node can communicate directly.

Now, next approach is to find the Dominating Sets (DSs) such that no two adjacent nodes belong to the same set. Each set consists of nodes with at least one hop distance from each other. Creating Dominating set signifies that every node has a neighbor which is a CH. In this way, this work obtains sets of nodes at one hop distance.

Further, after obtaining the Dominating sets the task is to rank these sets based on three parameters namely remaining power of nodes, degree of nodes and average distance of neighbors. Based on these parameters nodes would be given points and then points of every dominating sets will be calculated. The Dominating set getting the highest point will be selected.

IV. PROPOSED CLUSTERING & RANKING METHODOLOGY

Clustering algorithm gives DSs and Ranking methodology is used to find the most suitable set from these DSs.

A. Proposed Clustering Algorithm

Symbols used are given below:
L(i): Adjacency list for the node *i*.
NODE: The list of all the vertex V of the graph.
C[j]: List of nodes belonging to set *j*
UNODE: List of uncolored nodes and nodes not a part of set *j*

Input: Graph (adjacency list)
Output: Sets of non-neighbor nodes

Pseudo Code:

```
1. NODE  <- set of all nodes
2. For all node i ε NODE such that |L(i)|=0 delete i from NODE.
3. While (NODE! = empty)
4.    UNODE <-NODE
5.    While (UNODE!= empty)
6.       Find i ε UNODE such that |L(i)| is max
7.       C[j] <-C[j] U i
8.       NODE <-NODE-i
9.       UNODE <- UNODE-i
10.      UNODE<- UNODE-L(i)
11.   End while
12.   j=j+1
13. End while
```

Working:

Firstly elements of **NODE** is copied to another list named **UNODE**. Secondly, the vertex with highest degree is colored. Colored vertex and its adjacent ones are deleted from **UNODE**. Colored vertex is deleted from **NODE**. Colored vertex is added to set **C[j]** where j denotes the set number. This is repeated unless all nodes are deleted from **UNODE**. Once all nodes are deleted from **UNODE** then we move to first step. Before moving to First step value of **j** is incremented. First step runs until all nodes are deleted from **NODE**.

Line 1 shows that all the nodes are a part of the nodes. Line 2 deletes all the zero degree nodes (nodes with no neighbor) form NODE as our approach will not consider isolated nodes as the part of network. So, at any time **NODE** will represent a set of all the uncolored nodes. **UNODE** at the Line 4 stores all the uncolored nodes. **C[j]** used in line 7 maintains a set of all the nodes colored j. In Line 8 colored nodes are deleted from **NODE**. In lines 9 and 10 colored nodes as well as it immediate neighbor is deleted from **UNODE**. This ensures that no two neighbor nodes get the same color. Lines 11, 12 and 13 are loop running conditions. Symbol j used in Line 12 maintains the number of sets formed.

Now this approach will give one or more than one dominating sets. So, which among these should be chosen as cluster heads. This remains a problem. To overcome this a methodology to rank these sets has been discussed.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

B. Methodology to Rank Obtained Sets

In this part the sets obtained in the clustering algorithm will be ranked. Here, every node will be given points. Here, nodes having probability of higher energy consumption on being cluster head will be given less points. So, this would help in avoiding such nodes from becoming Cluster head. Based on these points every set obtained in clustering algorithm will be given points. The Dominating set having highest point will selected as CHs. This methodology does not ensures that a particular set will be ranked highest. There may be situations where more than one set will have highest points. In such case any of these sets can be selected. Given below are the parameters on which the probability of nodes are set.

Each node is given probability of becoming CH based on [2]:

- Remaining battery Power of Sensor (RPS)
- Number of Neighbor Nodes (D3N)
- Average distance from neighbors (DCC)

The above features are linguistically divided into three parts as discussed ahead. These linguistic parameters is used to categorize the nodes based on the above three features. This division enables us to give the probability to every parameters.

The term sets for each input linguistic parameter [2]:

- T(RPS) = {Low (lo), Middle (mi), High (hg)}
- T(D3N) = {Few (fw), Medium (me), Many (mn)}
- T(DCC) = {Near (nr), Moderate (mo), Far (fr)}

Total seven probabilities have been defined ranging from very weak to very strong. In this work suitable points have been given to each probability. These points helps in generalizing the nodes and cluster. Points are used as they can be used to average the sets. Probability can't be added and made to give points.

Specified probability and points given accordingly to every node [2]:

- Very Weak (vw) - 1
- Weak (w) - 2
- Little Weak (lw) - 3
- Medium (md) - 4
- Little Strong (ls) - 5
- Strong (s) - 6
- Very Strong (vs) – 7

Table 1. Table to give point to Nodes [2]

Rule	RPS	D3N	DCC	Prob	Point
1.	lo	fw	fr	vw	1
2.	lo	fw	mo	w	2
3.	lo	fw	nr	w	2
4.	lo	me	fr	w	2
5.	lo	me	mo	w	2
6.	lo	me	nr	w	2
7.	lo	mn	fr	vw	1
8.	lo	mn	mo	vw	1
9.	lo	mn	nr	vw	1
10.	mi	fw	fr	w	2
11.	mi	fw	mo	lw	3
12.	mi	fw	nr	md	4
13.	mi	me	fr	lw	3
14.	mi	me	mo	md	4
15.	mi	me	nr	ls	5
16.	mi	mn	fr	md	4
17.	mi	mn	mo	ls	5
18.	mi	mn	nr	s	6
19.	hg	fw	fr	lw	3
20.	hg	fw	mo	md	4
21.	hg	fw	nr	ls	5
22.	hg	me	fr	md	4
23.	hg	me	mo	ls	5
24.	hg	me	nr	s	6
25.	hg	mn	fr	ls	5
26.	hg	mn	mo	s	6
27.	hg	mn	nr	vs	7

Table 1 shows the points nodes gets based on the factors mentioned above. Point of each set is obtained by averaging the points obtained by its nodes. The DS getting the highest point is selected. In case of more than one DS set having same highest point any set with highest point can be selected. If more than one set have same maximum point then algorithm will randomly select any set with maximum point. This algorithm acts as a tiebreaker among dominating sets. This ranking methodology may give more than one dominating sets with same highest point. In real life also it sometimes gets difficult to decide the best ones based on certain fixed parameters.

V. RESULTS AND ANALYSIS

This section gives the complete analysis of result obtained using the Clustering algorithms and ranking methodology. Snapshots of results have been provided wherever required. Brief explanations of snapshots has been provided. Here, two cases have been discussed. Subsection A discusses the general output from the algorithm. Subsection B discusses the special case where more than one dominating sets get same highest point.

A. Case 1

For a sample network given in Figure 1 we obtain a colored graph as shown.

Adjacency list of ten nodes:

N[0]= {6 9} degree 2
N[1]= {2 3 5 7 8} degree 5
N[2]= {1 5 8} degree 3
N[3]= {1 5} degree 2
N[4]= {7} degree 1
N[5]= {1 2 3 6 7} degree 5
N[6]= {0 5 9} degree 3
N[7]= {1 4 5 8} degree 4
N[8]= {1 2 7} degree 3
N[9]= {0 6} degree 2

Sets obtained on running Clustering algorithm are:

Green = {1 4 6} Orphan nodes=0
Blue = {0 5 8g}
Orphan nodes=1
Yellow = {2 3 7 9}
Orphan nodes=0

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Point obtained by each nodes:

Node0 -3, Node3 -3, Node4 -3, Node9 -3, Node2 -4

Node6 -4, Node8 -4, Node1 -5, Node5 -5, Node7 -5

Now on averaging the points obtained by each nodes in the set we obtain the points of each set in Figure 1.

Points obtained by each set:

Green=4

Blue=4

Yellow=3.75

Now, among Dominating Sets Green and Yellow, Green has more point so

Green colored set will be selected.

B. Case 2

This shows the case where more than one set will have same highest rank.

Sample graph shown in Figure 2.

Points of each sets in Figure 2 are:

Green -3.5

Blue -3.5

Yellow -3.25

Pink -3.5

Grey -3

Here, both Dominating Sets Green and Blue have same points so now anyone of them can be randomly selected by algorithm.

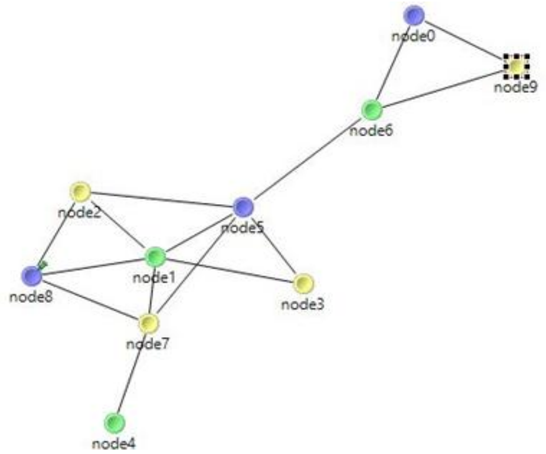

Fig 1. Sample network

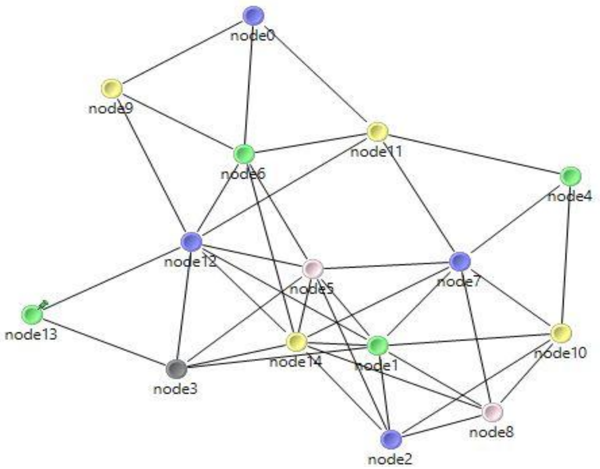

Fig 2. Blue and Green represent dominating sets.

VI. CONCLUSION & FUTURE WORKS

This approach gives clusters over the entire target area of network. This enables proper communication over the entire area concerned. It makes sure that every node is member of one of the clusters. This enables proper load sharing. Discussed algorithm may give more than one dominating set. Such cases have more than one set of cluster distribution but each will ensure even distribution. Even distribution has not been discussed much elaborately in other algorithms. Most of the other algorithms have high chances of giving uneven distribution. This results in either nodes being cluster less or having high number of nodes directly communicating to the base station. Both these conditions are undesired. Even distribution guarantees proper load sharing among nodes. It causes networks to have longer lifetime. Very few parameters have been considered which makes the algorithm less complex compared to other algorithms.

Ranking methodology part is like a contest between more than one dominating sets of node. Like any other contest here we get winner. In some cases we may get more than one winners as we see in case 2 of section V. In reality also sometimes it gets difficult to choose the best ones. More work can be done in ranking part of methodology by selecting other parameters than required. Other parameter selection may lead to different rank given to various dominating sets. Ranking however can be same of any two entities by considering any parameters. Security based approach can also be incorporated as these networks are highly vulnerable to outside attacks.

REFERENCES

1. Md Solaiman Ali, Tanay Dey, and Rahul Biswas. (2008). Aleach: Advanced leach routing protocol for wireless microsensor networks. In Electrical and Computer Engineering, 2008. ICECE 2008. International Conference on, pages 909–914.
2. Junpei Anno, Leonard Barolli, Arjan Durresi, Fatos Xhafa, and Akio Koyama. (2008). A cluster head decision system for sensor networks using fuzzy logic and number of neighbor nodes. In Ubi-Media Computing, 2008 First IEEE International Conference on, pages 50–56. IEEE.
3. Payal Khurana Batra and Krishna Kant. (2016). Leach-mac: a new cluster head selection algorithm for wireless sensor networks. Wireless Networks, 22(1):49–60.
4. Jing Chen and Hong Shen. (2007). Meleach an energy-efficient routing protocol for wsns. Chinese Journal of Sensors and Actuators, 9:035.
5. Adrian Carlos Ferreira, Marcos Aur´elio Vilac¸a, Leonardo B Oliveira, Eduardo Habib, Hao Chi Wong, and Antonio A Loureiro. (2005). On the security of cluster-based communication protocols for wireless sensor networks. In International Conference on Networking, pages 449–458. Springer.
6. MJ Handy, Marc Haase, and Dirk Timmermann. (2002). Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In Mobile and Wireless Communications Network, 2002. 4th International Workshop on, pages 368–372. IEEE.
7.] Wendi, B Heinzelman, Anantha P Chandrakasan, and Hari Balakrishnan. (2002). An application-specific protocol architecture for wireless microsensor networks. IEEE Trans- actions on wireless communications, 1(4):660–670.
8. Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. (2000). Energy-efficient communication protocol for wireless microsensor networks. In System sciences, 2000. Proceedings of the 33rd annual Hawaii international conference on, pages 10–pp. IEEE.
9. Wassim Jerbi, Abderrahmen Guermazi, and Hafedh Trabelsi. (2016). O-leach of routing protocol for wireless sensor networks. In Computer Conference on, pages 399–404. IEEE.
10. Jenn-Long Liu and Chinya V Ravishankar. (2011). Leach-ga: Genetic algorithm-based energy-efficient adaptive clustering protocol for wireless sensor networks. International Journal of Machine Learning and Computing, 1(1):79

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

11. Danish Mahmood, Nadeem Javaid, Shaharyar Mahmood, Shaima Qureshi, Atif M Memon, and Tariq Zaman. (2013). Mod-leach: a variant of leach for wsns. In Broadband and Wireless Computing, Communication and Applications (BWCCA), 2013 Eighth International Conference on, pages 158–163. IEEE.
12. Chaowei Tang, Qian Tan, Yanni Han,Wei An, Haibo Li, and Hui Tang. (2016). An energy harvesting aware routing algorithm for hierarchical clustering wireless sensor networks. KSII Transactions on Internet & Information Systems, 10(2).

AUTHORS PROFILE

Kaushlendra Kumar Sinha received his B.E. degree from Sir MVIT, Bangalore and MTech from NIT Patna. He is currently pursuing his Ph.D. in Computer Science and Engineering Department from NIT Patna. His research interests includes Wireless Sensor Networks and Machine Learning.

Dr. M. P. Singh is currently working as an Associate Professor with the Department of Computer Science and Engineering in National Institute of Technology Patna. His research area includes Social Media, Wireless Sensor Networks, Fuzzy set and Artificial Intelligence.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication