

# SPATIAL CORRELATION BASED CLUSTERING ALGORITHM FOR RANDOM AND UNIFORM TOPOLOGY IN WSNs

Bhavana H.T<sup>1</sup>, Jayanthi K Murthy<sup>2</sup>

<sup>1</sup>M.Tech Scholar, Dept. of ECE, BMS College of Engineering, Bangalore

<sup>2</sup>Associate Professor, Dept. of ECE, BMS College of Engineering, Bangalore

## Abstract

In Wireless Sensor Networks (WSN) sensor nodes with similar readings can be grouped such that, it is enough to report a single reading from the entire group. A representative node is selected from each cluster to do the reporting job. This helps to increase the battery life of sensor nodes. However, efficiently identifying sensor groups and their representative nodes is a challenging task. In this paper, a distributed algorithm is proposed to determine a set of representative nodes which exploits the tradeoff between data quality and energy consumption. In this paper, we group the sensor nodes based on their inherent spatial and data correlation in WSN. The proposed clustering algorithm is applied for uniform and random topology of sensor network. The results based on different metrics such as average number of clusters formed, energy consumption and average variation in cluster size are compared for both topologies.

-----\*\*\*-----

## 1. INTRODUCTION

A WSN consists of thousands of spatially distributed sensor nodes. Once deployed, the sensor nodes form a network through short-range wireless communication. They collect environmental data and send to the data processing center (Sink node). Each sensor node collects local data.

Data collection in WSNs consists of two main processes sampling and wireless communication. As the sensor nodes are battery operated their energy and communication bandwidth are limited. Fortunately most of the applications require less data accuracy. Hence it is unnecessary for all the sensor nodes to sample and report data. Energy consumption can be reduced greatly by turning off redundant sensor nodes and it will be wise choice for long term data collection.

The region in which all the sensor nodes send the readings which are similar in nature is known as Spatial Correlation region and therefore it is enough to send a single report to represent the correlation region [1]. The readings of a sensor node may be predicted from that of its nearby sensor nodes with high assurance.

This paper aims to save energy in continuous data collection applications. To do this job nodes are partitioned with similar observations into a cluster. Because of energy constraint and difficulty involved in discovering the spatial correlation pattern, spatial clustering in an energy-efficient and distributed manner is a tough work. A distributed algorithm is proposed in this paper to form the clusters based on correlation between sensor nodes. The same algorithm is applied on two network topologies uniform and random. The comparison is done between these two topologies using three different key metrics. A novel ranking method is used to select efficient representative nodes to cover entire network.

The rest of the paper is organized as follows: section II contains the related work about the spatial correlation based clustering algorithms, the system model is described in section III, section IV elaborates the algorithm, Results obtained are discussed in section V and conclusion and future work are suggested in section VI.

## 2. RELATED WORK

Clustering is a well known technology which helps in topology management in WSNs. In [2] [3] and [4] authors describes taxonomy of clustering and different energy aware clustering algorithms. These papers give a clear idea about the cluster properties, cluster head capabilities and clustering algorithms.

In [5] the author has introduced the spatial correlation concept in WSNs. It aggregates the data using joint entropy. A linear model is proposed to capture the spatial correlation. But in real world most of the systems may not be linear.

To estimate the energy dissipation for sensor networks author in [6] has proposed simple radio model LEACH.

B. Gedik et al. in [7] have proposed ASAP, a distributed clustering algorithm which groups the nodes with similar readings. ASAP can control the number of clusters but it needs a number of iterations to select efficient cluster heads to cover all nodes and distribution of cluster heads may not be so good.

EEDC proposed in [8] considers spatial and temporal correlation between nodes to group them. The spatial correlation range (user defined parameter) is used as a metric to find similarity between nodes. DCglobal proposed in [9] is a centralized algorithm which selects CH based on their data coverage range. However, EEDC and DCglobal

are centralized algorithms hence communication overhead to select cluster heads and to form the cluster is high.

Chih-Chieh Hung has extended DCglobal to distributed version known as DClocal it uses Manhattan distance to measure similarity. The DClocal uses more counters and hence it requires accurate time synchronization. Zhidan Liu et al. have proposed a distributed algorithm known as DSCC in [10]. This algorithm considers both distance and data similarity to group the sensor nodes. In this paper the algorithm used till the selection of CH is similar to that of DSCC.

Topology of nodes play an important role in resource constrained WSNs [11] [12]. Most of the clustering algorithms consider the uniform deployment of sensor networks. But in real time scenarios the sensor nodes will be random in fashion. We need to design an efficient clustering algorithm by considering random topology as well.

The efficient clustering algorithms demand less communication cost, long battery life, low cost and more data accuracy. The main problems associated with the existing spatial correlation based clustering algorithms are 1) No strict Requirement on similarity measurement between nodes. 2) Abundant communication overhead for spatial clustering. 3) Number of iterations to select enough cluster heads. 4) Algorithms are limited to uniform topology.

### 3. SYSTEM MODEL

In this paper, a WSN consisting of  $N$  sensor nodes with the Sink node located outside is considered. The sensor node  $i$  is denoted as  $S_i$  and the sensor node set as  $S = \{S_1, S_2, S_3, \dots\}$ , where  $|S| = N$ . All sensor nodes and Sink node are stationary after deployment. The initial energy of all sensor nodes is same and nodes are homogenous. Once the clustering begins the nodes can have different residual energy. The approximate distance between the nodes is estimated through received signal strength. Finally each sensor node can adjust the transmission power to change the communication range with single-hop communication between an active node and the Sink. After the cluster formation we assume a simple scheduling property where only the cluster head will report the reading and all other cluster members will be in sleep mode.

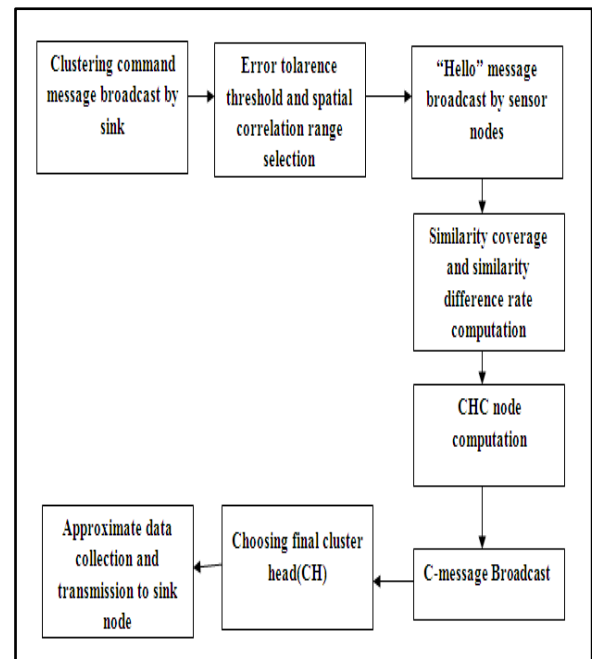


Fig: 1 System Model

The system model of the proposed algorithm is shown in figure 1. The model receives two user defined parameters namely Spatial Correlation Range ( $R_{SC}$ ) and Error Tolerance Threshold ( $\epsilon$ ) to calculate the similarity between nodes. Each sensor node adjusts the transmission power to broadcast a "Hello" message (H-msg), which contains its residual energy and recent reading serial to neighboring nodes in the range of spatial correlation. Using this information each node decides its similar nodes. Based on the local information, energy level and representative capability each node decides whether to become a cluster head candidate (CHC). Later, all CHC nodes will compete to be the final cluster head (CH) by sending the "Competition" message (C-message) which consists of only node id to its neighboring nodes. All non-CH nodes will choose an appropriate cluster to join and hence the clustering will finish.

### 4. ALGORITHM TO FIND THE CH

This algorithm ends with the selection of appropriate CH. The clustering algorithm works according to following steps.

**1) Detection of similar nodes:** The Euclidean distance is used to calculate the geographical distance between the nodes. We say that  $S_j$  is the neighboring nodes of a node  $S_i$  only if the distance between two nodes is within a spatial correlation range  $R_{SC}$ .

The Manhattan distance is a metric used to find the dissimilarity between the nodes using their recent reading serials. The dissimilarity measurement function  $f_{md}(S_i, S_j)$  is defined as

$$f_{md}(S_i, S_j) = \frac{\sum_{k=1}^q |V_k(S_i) - V_k(S_j)|}{q} \dots \dots \dots (1)$$

In Equation 1,  $V(S_i) = \{V_1, V_2... V_q\}$  and  $V'(S_j) = \{V'_1, V'_2... V'_q\}$  are the sensor reading serials of  $S_i, S_j$  respectively and  $q$  is the number of readings from each node.

If the dissimilarity measurement between nodes  $S_i$  and  $S_j$  is less than half of error tolerance threshold i.e.,  $f_{md}(S_i, S_j) \leq \epsilon/2$  and  $S_j$  is neighboring node of  $S_i$  then both nodes are said to be similar to each other. Greater  $R_{SC}$  will lead to fewer clusters but this leads to larger dissimilarity of cluster. On the other hand lesser  $R_{SC}$  results in less number of clusters which increases the communication overhead.

**2) Computation of Similarity Coverage Rate and Similarity Difference Rate:** The similarity coverage rate  $C_r(S_i)$  is the rate of coverage capacity of a node. It is calculated using equation 2. If the  $C_r(S_i)$  is high then the capacity of node to cover its neighboring node is also will be high.

$$C_r(S_i) = \frac{|SN(S_i)|}{|NBR(S_i)|} \dots \dots \dots (2)$$

The similarity difference rate  $S_r(S_i)$  gives the dissimilarity measurement between the nodes using their Manhattan distance. It is computed using equation 3.  $S_r(S_i)$  expresses the similar degree between  $S_i$  and its similar nodes. Both  $C_r(S_i)$  and  $S_r(S_i)$  fall in the range of  $[0, 1]$ .

$$S_r(S_i) = \frac{\sum_{S_j \in SN(S_i)} f_{md}(S_i, S_j)}{\epsilon} \dots \dots \dots (3)$$

**3) Computation of Representative Capability  $p(S_i)$  and Relative Energy Level  $E_R(S_i)$ :** Equation 4 gives the expression for representative capability of a node.

$$p(S_i) = \alpha * C_r(S_i) + (1 - \alpha)S_r(S_i) \dots \dots (4)$$

Where  $\alpha$  is an adjustable parameter to decide the relative importance of  $C_r$  and  $S_r$ .

The relative energy level  $E_R(S_i)$  of a sensor node  $S_i$  is calculated using equation 5.

$$E_R(S_i) = \frac{E(S_i) * (N-1)}{E(S_i) + \sum_{S_j \in NBR(S_i)} E(S_j)} \dots \dots \dots (5)$$

$E(S_i)$  is the residual energy of  $S_i$ . The ordinary sensor node becomes CHC node only if it has  $E_R(S_i) > 1$  and  $p(S_i) > \delta$ , where  $\delta$  is a parameter which controls the number of CHC.

**4) Selection of CH:** A CHC node and broadcasts competition message (C-msg) which includes ID and the size of similar node set. In the perspective of spatial correlation, node  $S_i$  defeats node  $S_j$  if and only if given conditions are satisfied. 1) They are similar nodes 2)  $SN(S_i)$  contains more nodes than  $SN(S_j)$  3) More than half nodes of set  $SN(S_j)$  are in set  $SN(S_i)$ .

The ranking assignment is performed as per the flowchart indicated in figure 2. At the expiry of  $T_{CHC}$ , each CHC node confirms its own ranking and algorithm will select the final

CHs in an iterative manner by exploiting these rankings. This CH algorithm is shown in figure 3.

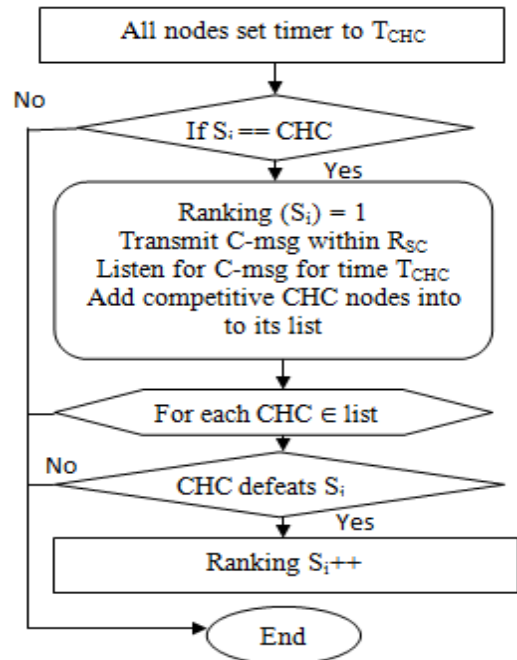


Fig: 2 Ranking assignment algorithm

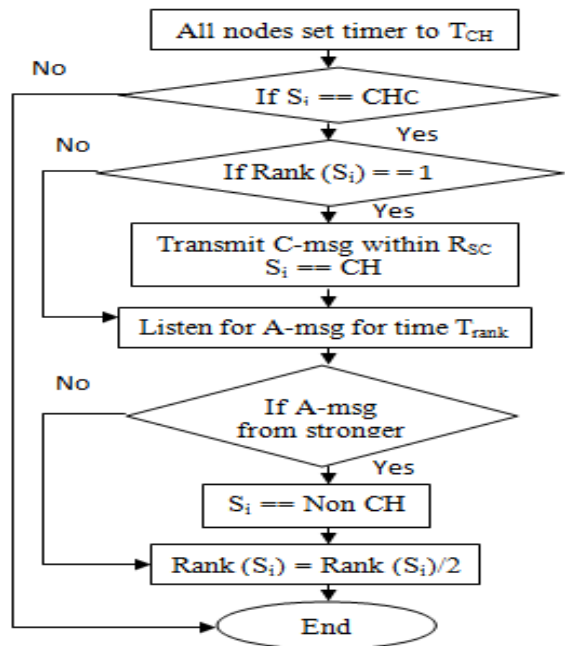
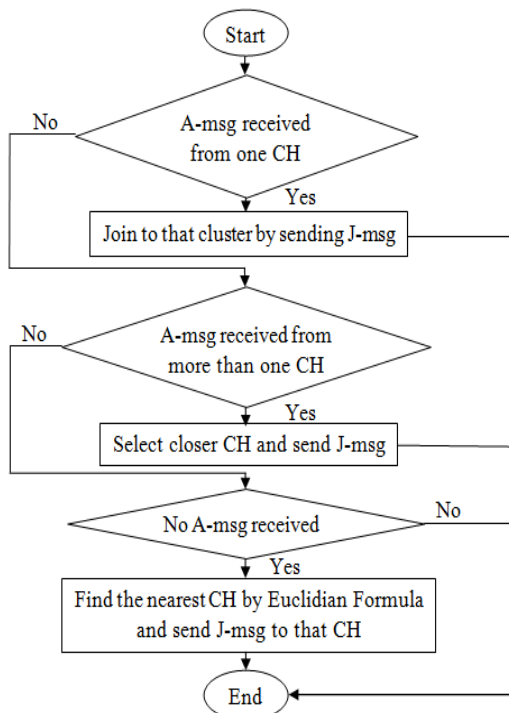


Fig: 3 CH Selection Algorithm

The CH selection phase lasts for  $T_{CH}$ . In first step loop checks whether the node is CHC or not, if so then its ranking is verified. If the ranking is equal to 1, it selects itself as a final CH and broadcasts a CH advertising message (A-msg) which only includes its node ID to neighbors in range of  $R_{SC}$ . After receiving an A-msg, any ordinary node or CHC node whose ranking is larger than 1 will add this CH to its list and consider itself as being “covered” by a CH node. At the expiry of  $T_{rank}$  every “uncovered” CHC node

reduces its ranking a half and goes to next iteration. This loop is executed till a node becomes CH or non-CH.

**5) Proposed algorithm for Formation of cluster:** The cluster formation algorithm is shown in figure 4.



**Fig: 4** Formation of cluster

In this phase if an ordinary node receives only one A-msg from any other CHC, it immediately joins that CH. In second case if any node receives A-msgs from more than one node it selects the closer node as the CH and joins to that cluster. In third case if a node does not receive any A-msg then it finds the nearest CH using Euclidean distance formula. Finally every ordinary node will join one cluster and hence the cluster formation process will end.

## 5. RESULTS OBTAINED

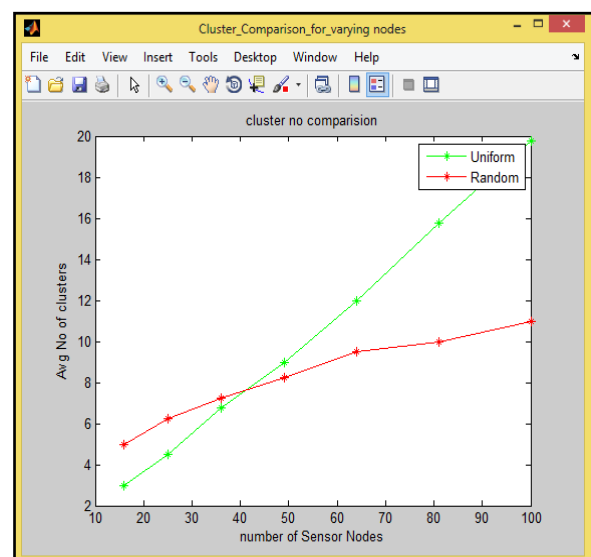
The simulation is carried out in Matlab. The two user defined thresholds spatial correlation range and error tolerance threshold are selected as 20 and 0.7 units respectively. The residual energy of nodes is taken in the range [4, 5]. The simulation is carried on both uniform and random distribution of nodes. The two threshold  $\alpha$  and  $\delta$  are set as 0.5 and 0.4 respectively. For uniform deployment the distance between immediate neighboring nodes is considered as 13 units.

Extensive simulations are carried out by varying the sensor nodes from 16 to 100. As the number of nodes varies the field size is also varied from 52X52 to 130X130 respectively. All results in this paper are average values of twenty simulations.

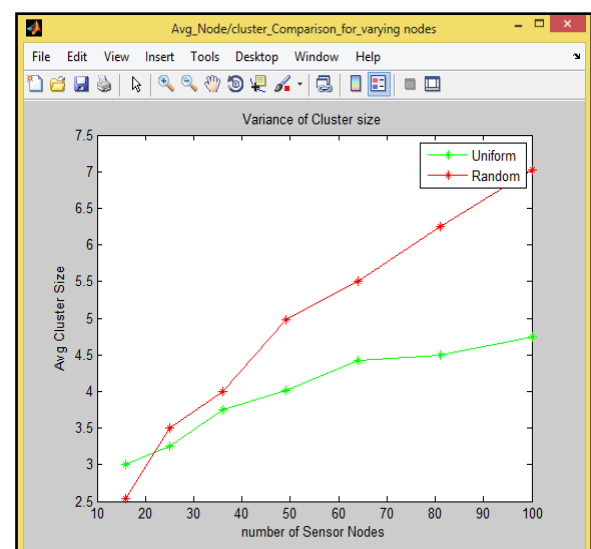
The figure 5 shown below is the comparison of number of clusters formed. This shows that when the sensor nodes are

less in number the number of clusters generated is less. But as the sensor nodes increase the cluster number rapidly changes in uniform topology. The graph for random topology gradually increases as the number of nodes increases. Less number of clusters facilitates lower power consumption as in spatial correlation based clustering more nodes can be turned into sleep mode. For the network with more number of nodes, random topology is suitable as it ends with less number of clusters when compared to uniform topology.

The variation in cluster size is shown in figure 6. The uniform topology shows less variation in cluster size. Since the number of clusters formed rapidly increases in uniform topology it can maintain the same cluster size. The cluster size variation is more in random topology, however as explained above more no of nodes can be turned into sleep mode. Hence there may not be significant effect on energy efficiency.

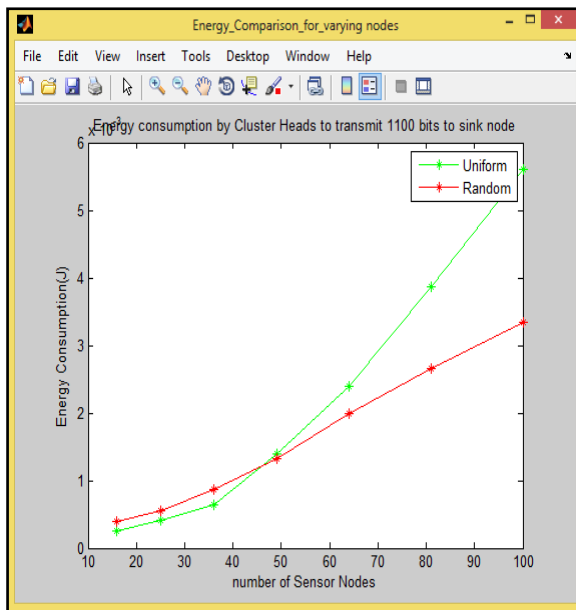


**Fig: 5** Number of Clusters Formed



**Fig: 6** Variation in Cluster Size

We consider the famous radio model LEACH to estimate the energy dissipation of sensor nodes.



**Fig: 7** Energy Consumption

The energy consumed for CH to transmit data to sink node is calculated and results are plot in the figure 7. The parameters assumed for transmission of data from one node to another are given as  $E_{elec}= 50$  nJ/bit,  $E_{fs}=10$  pJ/bit/m<sup>2</sup>,  $E_{mp}=0.0013$ pJ/bit/m<sup>4</sup>. The threshold distance  $d_0$  is varied according to field size.

The above figure clearly indicates that the energy consumed by uniform topology rapidly increases as the number of nodes increase. In random topology this curve is gradually increasing and always it is less than uniform topology. Hence for network with less number of nodes uniform topology is suitable. The random topology will be energy efficient when the sensor nodes are more.

## 6. CONCLUSIONS AND FUTURE WORK

In the proposed model the cluster formation is done with reduced communication overhead between the nodes, as the messages exchanged during clustering is very less. The computational complexity involved in selecting the efficient cluster heads is little more. However, the energy required for communication is more when compared to the computation.

The simulations are carried out for uniform and random topologies. When the sensor nodes are less in number uniform topology gives better result in all perspective. When the number of nodes is more random topology is best suited since the energy efficiency is more.

As the future work efficient scheduling algorithms can be implemented in every cluster to make few nodes to work at a time and hence more data fidelity can be acquired. Along with this clustering algorithm good routing algorithms can be integrated to get better results.

## REFERENCES

- [1]. Arthur Getis, "A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective" ISSN, vol. 4 pp.0016-7363, 2007.
- [2]. Pedro A. Forero, Student Member, IEEE, Alfonso Cano, Member, IEEE, and Georgios B. Giannakis, Fellow, IEEE, "Distributed Clustering Using Wireless Sensor Networks" IEEE journal of selected topics in signal processing, vol. 5, no. 4, august 2011.
- [3]. Vinay Kumar, Sanjeev Jain and Sudarshan Tiwari, "Energy Efficient Clustering Algorithms in Wireless Sensor Networks: A Survey", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 2, September 2011
- [4]. A. Abbasi and M. Younis, "A Survey on Clustering Algorithms for Wireless Sensor Networks," Computer Communications, vol. 30, no. 14-15, pp. 2826–2841, 2007.
- [5]. W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "An Application Specific Protocol Architecture for Wireless Microsensor Networks," Wireless Communications, IEEE Transactions on, vol. 1, no. 4, pp. 660–670, 2002.
- [6]. Sandeep Pattern, Bhaskar Krishnamachari, Ramesh Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks". IPSN, vol. 1, pp. 846. 2004.
- [7]. B. Gedik, L. Liu, and P. Yu, "ASAP: An Adaptive Sampling Approach to Data Collection in Sensor Networks," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, no. 12, pp. 1766–1783, 2007.
- [8]. L. Chong, W. Kui, and P. Jian, "An Energy-Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation," Parallel and Distributed Systems, IEEE Transactions on, vol. 18, no. 7, pp. 1010–1023, 2007.
- [9]. C. C. Hung, W. C. Peng, and W. C. Lee, "Energy-Aware Set-Covering Approaches for Approximate Data Collection in Wireless Sensor Networks," IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 11, 2011.
- [10]. Zhidan Liu, Wei Xing, Bo Zeng, Yongchao Wang, Dongming Lu, "Distributed spatial Correlation Based Clustering for WSNs", IEEE computer society, pp. 56-64, 2013.
- [11]. Quazi Mamun, "A Qualitative Comparison of Different Logical Topologies for Wireless Sensor Networks", Sensors 2012, vol.12, pp. 14887-14913, 2012.
- [12]. Divya Sharma, Sandeep Verma, Kanika Sharma, "Network Topologies in Wireless Sensor Networks: A Review", IJECT, vol.4, Issue-3, 2013.