INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
*(Volume 4, Issue 2)*
Available online at: www.ijariit.com

# Comparative analysis of clustering algorithms in wireless sensor networks

*R. Janani*
*jantechjan@gmail.com*
*Thiagarajar College of Engineering, Madurai, Tamil Nadu*

*B. Jayaranjani*
*b.jayaranjani@gmail.com*
*Thiagarajar College of Engineering, Madurai, Tamil Nadu*

## ABSTRACT

***In recent years, Wireless Sensor Networks (WSN) gained popularity in sensing information from an interesting area. Day by day new technologies is emerging with innovative features, which make this field active for development and research oriented work. Since the available energy of sensor nodes are limited and hard to renew, energy supervision is critical for nodes and network lifetime in wireless sensor networks. In this paper, clustering algorithms like LEACH, K-means, Fuzzy c means are applied in WSN scenario for sensing applications to reduce energy consumption. We propose modifications on the above-mentioned algorithms by selecting cluster heads (CH) based on the residual energy of the sensor nodes. Simulation results show that the proposed approach produces better cluster head energy efficiency.***

**Keywords:** *Cluster head, Residual energy, Sensor nodes.*

## 1. INTRODUCTION

Nowadays an efficient design of a Wireless Sensor Network has become a leading area of research. There are many extensive applications of wireless sensor networks such as agricultural monitoring, surveillance, industrial monitoring and many more[5][6][7][8].A Wireless Sensor Network is a self-configuring network of small sensor nodes communicating among themselves using radio signals, and deployed in quantity to sense, monitor and understand the physical world. In Wireless sensor network, the network is embedded in environment. Nodes in the network are equipped with sensing and actuation devices to influence the environment. Sensor nodes process information and communicate it wirelessly. A general sensor node consists of sensing unit, processing unit, communication unit and power unit [10] as shown in Fig.1.It may also have application dependent additional components such as a location finding system, a power generator and a mobilizer. Generally the sensor nodes are deployed in rural areas where providing continuous power supply by transmission lines is difficult and needs high maintenance [2]. So the battery is used as a power source for the sensor nodes. In order to have enhanced network lifetime, it is important to reduce energy consumption of the sensor nodes. Data transmission and routing comparatively consumes more energy. Applying clustering techniques in routing provides reasonable impact on energy consumption.

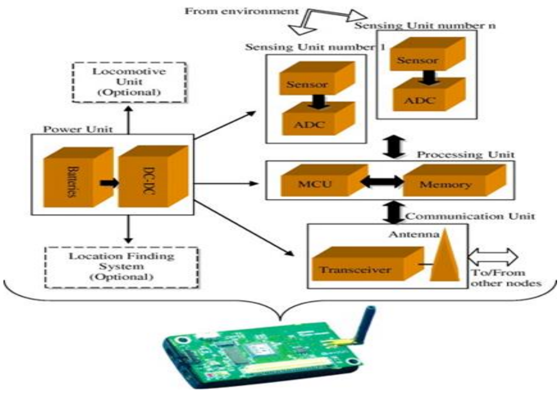


**Fig.1**

We use Zigbee technology to transmit the sensed data wirelessly between nodes as Zigbee shows a prompt decrease of Bit error rate even for less Signal to Noise Ratio [9]. This indicates that even in the presence of high noise strength almost equal to the signal strength, the system can trace the required signal eliminating the errors. This gives a very robust performance. In this paper, we look at the improvement on energy consumption of the sensor node clustering problem. We discuss LEACH (Low Energy Adaptive Clustering Hierarchy), K-means and Fuzzy c means algorithms and their drawbacks. We propose modifications on these algorithms to overcome the shortcomings. Simulation is performed in Netsim v 10.2 and their corresponding results are produced in Results and Discussions.

## 2. NETWORK MODEL

Let us consider the scenario that there are N sensor nodes distributed uniformly in an M × M region. We make some assumptions about the sensor nodes and the underlying network model:

• These sensor nodes and the BS are all stationary after deployment. This is typical for wireless sensor network applications.

• There is only one BS located far from the sensing field.

• Sensors are homogeneous and have the same initial energy.

• The BS knows the geographical information of all sensor nodes.

• To avoid the delay time too much, we use one-hop communication. The CHs directly communicate with the sensor nodes or BS.

## 3. SYSTEM ENERGY MODEL

The radio energy dissipation model in [1] has been used to estimate the energy consumption of the nodes in radio communications. The energy dissipated in radio hardware to transmit L bit packet to a distance d is,

$$ET_x(L,d) = \begin{cases} L\,E_{elec} + L\,efs\,d^2, d < d_0 \\ L\,E_{elec} + L\,emp\,d^4, d \geq d_0 \end{cases} \qquad (1)$$

Where $d_0$ denotes distance threshold and $E_{elec}$ represents the energy consumption per bit in radio electronics. efs and emp are the energy dissipation factors in the power amplifier for free space and multipath fading channel respectively.

## 4. LEACH PROTOCOL

LEACH is a routing protocol for Wireless Sensor Network expanded as Low Energy Adaptive Clustering Hierarchy. Generally, there are three steps in cluster routing protocol: the generation of cluster heads, the formation of clusters and the communication among clusters. The first two steps can be merged into one that is the establishment of clusters and the communication among clusters. Thus, LEACH protocol algorithm contains the set-up of clusters and stable data transmission. In the first stage, LEACH adopts equal probability method, selecting cluster heads in a circle and random manner and distributing the energy of the whole network evenly on each node. Therefore, LEACH algorithm reduces the energy consumption and prolongs the lifetime of the network. The executive process of LEACH is periodical, and each period includes the establishment of clusters and data transmission. The specific process is as follows:

During the set-up stage [4] of clusters, nodes will generate a number randomly between 0 and 1(including 0 and 1). If the random number is smaller than the threshold T(n), then the node will be a cluster head in this round. The calculation method of T(n) is based on the following formula:

$$T(n) = \begin{cases} \frac{p}{1-p[r\,mod\left(\frac{1}{p}\right)]}, N \in G \\ 0\ , Otherwise \end{cases} \qquad (2)$$

**p** represents the percentage of cluster nodes accounting in the total number of nodes, that is the probability of nodes becoming cluster heads;
**r** refers to the current number of rounds ;
**N** is the total number of nodes;
**G** is the set of nodes that did not become cluster heads in the 1/p round.

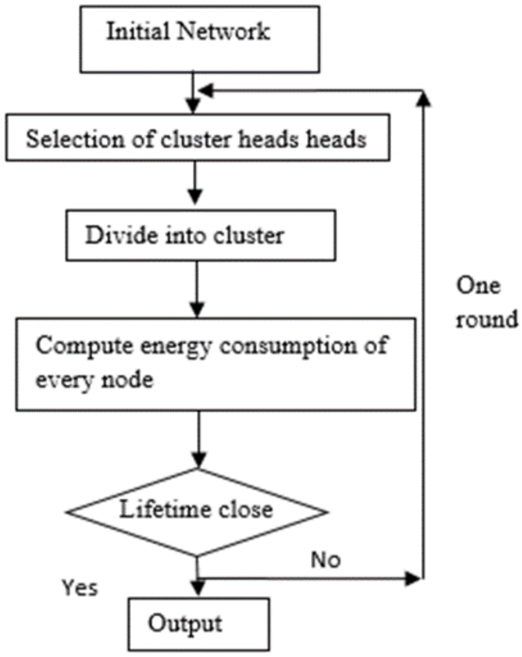


**Fig.2 LEACH**

Nodes that selected as cluster heads then send the information that it is a cluster head to its neighbor nodes, and the remaining node will choose the cluster that it will join according to the strength of the broadcast signal it receives and inform the related cluster head. After that, cluster heads create a TDMA, a timing gap created for each node in this cluster, and send this timing gap to them in form of broadcast. Thus, each node can send data in its own timing gap, while in the other timing gap, the node will enter into a sleep state, hence saving energy. During the phrase of stable data transmission, member nodes (non-cluster nodes) in the cluster will transfer the monitored data to related cluster head in its given time gap.At the end of each round, cluster heads and clusters will be re-elected, which need certain energy. In order to reduce the overhead of the system, duration of the stable stage in each round is much longer that for the establishment of the cluster. As for the cluster head, it always maintains communication status to receive the data from the nodes in its cluster at any time. Once received all data from its member nodes, then the cluster head will process the data such as data fusion to lower down the redundant data. Finally, the cluster head transmits the fused data to its own cluster head or the base node; as for non-cluster nodes, they send the data in its own time gap and during the other time, they turn off their wireless communication module to conserve energy.

## 5. K-MEANS

K-means is a prototype based algorithm that alternates between two major steps, assigning observations to clusters and computing cluster centers until a stopping criterion is satisfied. K-means algorithm is based mainly on the Euclidean distances. So here the central node collects the information about the node id, position and residual energy of all nodes and stores this information in a list in the central node. After getting this information from all nodes it starts performing the clustering K-means algorithm [3].

1) If we want to cluster the nodes into ‘k’ clusters, take ‘k’ number of centroids initially at random places.

2) Calculate the Euclidian distance from each node to all centroids and assign it to centroid nearest to it. By this ‘k’ initial clusters are formed Suppose there are N nodes are given such that each one of them belongs to Rd. The problem of finding the minimum variance clustering of this nodes into k clusters is that of finding the k centroids { $m_j$ }k j = 1 in Rd such that,

$$\left(\frac{1}{N}\right) * \sum \left(\min_j d^2(Xi, m_j)\right), for\ i = 1\ to\ N \quad (3)$$

Where d($Xi$ , $m_j$) denotes the Euclidean distance between $Xi$ and $m_j$. The points {j}k ,i = 1 ar e known as cluster centroids or as cluster means.

3) Recalculate the positions of centroids in each cluster and check for the change in position from the previous one 4. If there is a change in position of any centroid then go to STEP 2, else the clusters are finalized and the clustering process ends.

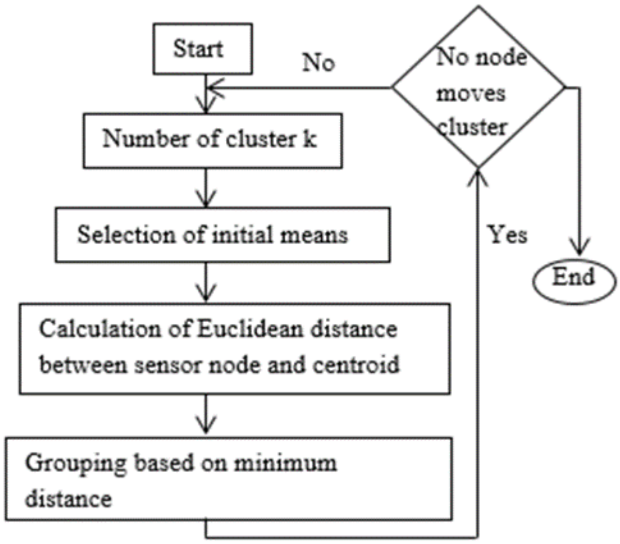


**Fig.3 K-means algorithm**

## 6. FUZZY C MEANS

FCM clustering protocols are centralized clustering algorithms, the base station computers and allocate sensor nodes into clusters according to the information of their location and the cluster head is assigned to the node having the largest residual energy. We consider a network of N sensor nodes which is partitioned into c clusters: C1, C2, ..., Cc. The purpose of the cluster formation in this protocol is to minimize the following objective function:

$$J_m = \sum \sum u_{ij}^m \, d_{ij}^2 \tag{4}$$

Where uij is node j's degree of belonging to cluster I dij is the distance between node j and the center point of cluster i. The degree uij of node j respected to the cluster is calculated and fuzzified with the real parameter m > 1 as below:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} (d_{ij}/d_{kj})^{2/(m-1)}} \tag{5}$$

The distance between the sensor node and the center point is Euclidean distance. By achieving minimization of the spatial distance, the energy balance among sensor nodes is optimized.

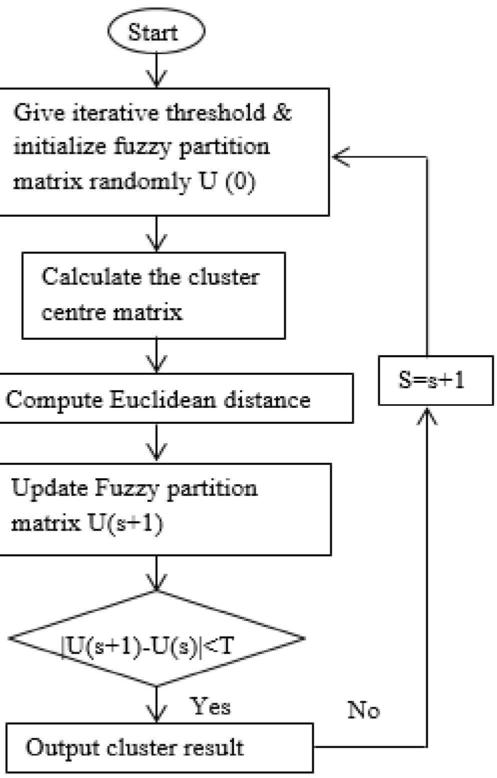


**Fig.4 Fuzzy c means algorithm**

## 7. LIMITATION

In LEACH, the CHs are randomly elected and making it possible for a minimal energy node to become CH. K-means and Fuzzy c means elects CH only based on distance. All of the above three algorithms do not consider the residual energy of the sensor node on selecting cluster head.

## 8. PROPOSED WORK

All of the above algorithms do not consider the Residual energy of the sensor node. If a node with minimum energy is elected as a cluster head, then in a short time, the nodes will be dead leading to the network failure. To avoid this issue, in all of the above algorithms, the Residual energy of the sensor nodes is considered to elect the cluster head.

a) LEACH-R: When selecting cluster heads, LEACH protocol selects cluster heads according to the random number the node generates and the threshold, while the threshold did not take the remaining energy into account, which may result in the node with low energy to be a cluster head, thus bringing premature death to clusters and affecting the lifetime of network. We introduce the remaining energy factor based on LEACH protocol, that is:

$$\boldsymbol{T(i)} = \begin{cases} \frac{P_i}{1-P_i*(r*mod\left(\frac{1}{P_i}\right))}, N \in G \\ 0, otherwise \end{cases} \quad (6)$$

There into,

$$\boldsymbol{P_i} = \frac{(E_i - E_r)^2}{E_R} \quad (7)$$

$E_i$ : Rethe sidual energy of each node in i round;

$E_r$: Average energy of rest nodes in the i round;

$E_R$: Tothe tal residual energy of rest nodes in the i round;

Calculating formula of the average energy of rest nodes Er:

$$E_r = E_R * \frac{(1-\frac{r_i}{r_{max}})}{N} \quad (8)$$

$r_i$ : The current round; $r_{max}$ : Maximum rounds of network simulation;

b) MODIFIED K-MEANS & MODIFIED FUZZY C MEANS: In k-means and fuzzy c means, clustering is performed as shown in Fig.3 & Fig.4 respectively. Inclusive of distance, we also consider the residual energy of the sensor node to elect cluster heads and we add the following variations for k-means and fuzzy c means. The essential parameters are a) the distance from a node to the cluster head b) the residual energy. The objective is to minimize the objective function $O(d_{sCH}(s_i), R_i)$ to decide node $s_i$ in the cluster to be CH:

$$\boldsymbol{O(d_{sCH}(s_i), R_i) = d_{sCH}(s_i) \times \exp\left(\frac{1}{R_i}\right)} \quad (9)$$

Where $\boldsymbol{R_i} = \frac{E_i(r)}{\overline{E(r)}}$ , r denotes the current round, $d_{sCH}(s_i)$ is the distance from the sensor node $s_i$ to the cluster center. $\overline{E(r)}$ is the average residual energy.

$$\overline{E(r)} = E_T(r)/N \quad (10)$$

N is the total number of sensor nodes.

$$E_T(r) = E_{init-}(r-1)E_{rnd} \quad (11)$$

$$E_{init} = \sum_{i=1}^{N} E_i \quad (12)$$

Where $E_{init}$ the initial is total energy of the network and $E_i$ is the initial energy of each node.

Using Function $O(d_{sCH}(s_i), R_i)$ we can get nodes with more energy and near the cluster center t,o be CHs. When the final CHs are picked, the BS notifies sensor nodes about this, and then the communication begins.The total energy dissipated in a round [2] is

$$E_{rnd} = L(2N\,E_{elec} + NE_{DA} + Kemp\,d_{bs}^4 + Nefsd_{sCH}^2) \quad (13)$$

Where K is number of clusters, $E_{DA}$ is the energy dissipated in the CH to base station.

By these modifications, the CH has relatively high residual energy and quite short distance to the cluster centers in each cluster. The CHs are selected with the least value about function $O(d_{sCH}(s_i), R_i)$. Let us analyze (8), Function $O(d_{sCH}(s_i), R_i)$ is directly proportional to $d_{sCH}(s_i)$ and $\exp(1/R_i)$. So the distance to the cluster center and the residual energy can affect the value of function $O$. i.e.

$$\boldsymbol{d_{sCH}(s_i)}\downarrow \ \mathbf{exp(1/}\boldsymbol{R_i}\mathbf{)}\downarrow => \boldsymbol{O(d_{sCH}(s_i), R_i)}\downarrow \qquad (14)$$

We discuss the following three scenarios.

Case 1: In Cj, $s_i$ is the nearest node to $\bar{u}_j$and it has the maximal residual energy where $C_j$ is the current cluster and $\bar{u}_j$ is the centroid of all the nodes $\in C_j$ . $s_i$gets the least value about function $O$. So it is selected as the CH and it is the most appropriate one as the CH.

Case 2: In Cj, the node $s_1$ and $s_2$ are closed to $\bar{u}_j$. The distance from nodes $s_1$ and $s_2$ to $\bar{u}_j$ are much at one. The residual energy of $s_1$ is more than $s_2$'s. So $s_1$ gets the lesser value about Function $O$ and it has high probaba ility as the CH.

Case 3: In Cj, the residual energy of nodes $s_1$ and $s_2$ are much at one. The distance from $s_1$ to $\bar{u}_j$is less than the distance from $s_2$ to $\bar{u}_j$. So $s_1$ gets the lesser value about Function $O$ and it has high proba ability as the CH.

Therefore, after we look at the above three cases closely, we can draw the conclusion that, when we select the CHs using Function $O$, we are able to pick up those nodes with more energy and being near the cluster center to act as the CHs.

## 9. SIMULATION ANALYSIS

The simulation is carried out in 50*50 sq.m monitoring square area, with 64 sensor nodes and the base station is located at the mid of right side of the square area. The transmission time is set to be 300ms. All of the clustering methods mentioned above is applied to achieve clustering among the sensor nodes. The simulation experiment is done with NETSIM v.10. 2 interfaced with MATLAB 2014a. The experimental parameters are shown in Table-1. Fig.5 shows the comparative result on energy consumption of LEACH-R, K-means, Fuzzy c means, modified K-means and modified Fuzzy c mean. Table-2 indicates that LEACH-R has maximum energy consumption and modified K-means offers minimum energy consumption.

**Table-1: Simulation parameters**

| PARAMETERS | VALUES |
|---|---|
| Network size | $50*50m^2$ |
| Location of the sink node | 47,16 |
| Number of nodes | 64 |
| Number of clusters | 4 |
| Physical and MAC layer | Zigbee |
| Sensor range | 100m |
| Link mode | Half Duplex |

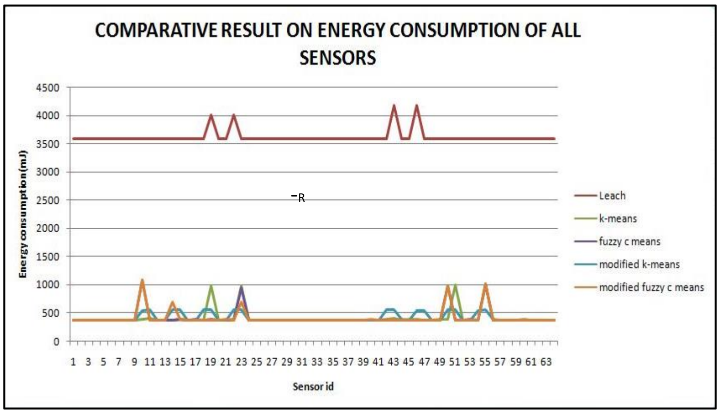


**Fig.5.The comparative result of energy consumption of all sensors**

**Table-2: Inference from Fig.5**

| PROTOCOL | MAXIMUM ENERGY CONSUMED(mJ) |
|---|---|
| Leach-R | 4177.163 |
| K-means | 972.72 |
| Fuzzy c means | 988.1 |
| Modified K-means | 539.45 |
| Modified Fuzzy c means | 983.6 |

Fig.6 shows the average cluster head energy efficiency of LEACH-R, K-means, Fuzzy c means, modified K-means, modified Fuzzy c means. On simulating the WSN environment as in Table-1 without applying clustering protocols, we obtained maximum energy consumption of 5578.4 mJ. With this reference, Table-3 is generated which illustrates that modified K-means produces a better result.

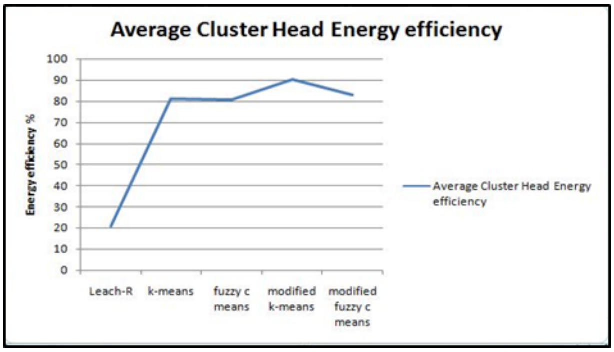


**Fig.6.Average cluster head energy efficiency**

**Table-3: Inference from Fig.6**

| **Protocol** | **Average CH energy efficiency %** |
|---|---|
| Leach-R | 20.707 |
| K-means | 80.94122 |
| Fuzzy c means | 80.6175 |
| Modified K-means | 90.01479618 |
| Modified Fuzzy c means | 82.80413712 |

## 10. CONCLUSION

In this paper, we propose improvisation in LEACH, K-means and Fuzzy c means to increase the network lifetime. From Fig.6, we could observe that the modified K-means algorithm provides higher energy efficiency of 90% when compared to a WSN setup without the application of clustering protocol. Modified Fuzzy c means also provides considerable energy efficiency of 83%. As a future work, we are planning to implement modified K-means algorithm in precision agriculture using WSN.

 

## 11. REFERENCES

[1] W.Heinzelman, A. Chanrakasan, H. Balakrishnan,"Energy-efficient communication protocol for wireless micro-sensor networks," Proc. of the 33rd Hawaii International Conf. on System Sciences, vol. 2, pp. 1-10, Jan 2000.
[2] Hiren Patel, Vipul Shah, "A Review on Energy consumption and conservation techniques for Sensor node in WSN," IEEE SCOPES,2016.
[3] P.Sasikumar, Sibaram Khara, "k-MEANS clustering in wireless sensor networks," IEEE 2012 Fourth International Conference on Computational Intelligence and Communication Networks.
[4] Alka Singh, Shubhangi Rathkanthiwar, Sandeep Kakde, "LEACH Based-Energy Efficient Routing Protocol for Wireless Sensor Networks," IEEE ICEEOT, 2016.
[5] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," Comput. Networks, vol. 38, no. 4, pp. 393–422, 2002.
[6] M. Ahlawat, "Wireless Sensor Network-A Theoretical Review," Rgsociety.Org, vol. 1, no. 2, pp. 11–19, 2013.
[7] C. Buratti, A. Conti, D. Dardari, and R. Verdone, "An Overview of Wireless Sensor Networks Technology and Evolution," Sensors, vol. 9, no. 9, pp. 6869–6896, 2009.
[8] D. Puccinelli and M. Haenggi, "Wireless sensor networks: applications and challenges of ubiquitous sensing," Circuits Syst. Mag. IEEE, vol. 5, no. 3, pp. 19–31, 2005.
[9] Santoshkumar, Udaykumar R.Y," Development of WSN System for Precision Agriculture," IEEE ICIIECS, 2015.
[10] D. Gupta, "A Review on Wireless Sensor Networks," Netw. Complex Syst., vol. 0603, no. 1, pp. 18–23, 2013.

 