

# EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN

Purnima Bholowalia  
Computer Science and Engineering  
LPU, Phagwara (Punjab) India.

Arvind Kumar  
Computer Science and Engineering  
LPU, Phagwara (Punjab) India.

## ABSTRACT

WSN consist of hundreds of thousands of small and cost effective sensor nodes. Sensor nodes are used to sense the environmental or physiological parameters like temperature, pressure, etc. For the connectivity of the sensor nodes, they use wireless transceiver to send and receive the inter-node signals. Sensor nodes, because connect their selves wirelessly, use routing process to route the packet to make them reach from source to destination. These sensor nodes run on batteries and they carry a limited battery life. Clustering is the process of creating virtual sub-groups of the sensor nodes, which helps the sensor nodes to lower routing computations and to lower the size routing data. There is a wide space available for the research on energy efficient clustering algorithms for the WSNs. LEACH, PEGASIS and HEED are the popular energy efficient clustering protocols for WSNs. In this research, we are working on the development of a hybrid model using LEACH based energy efficient and K-means based quick clustering algorithms to produce a new cluster scheme for WSNs with dynamic selection of the number of the clusters automatically. In the proposed method, finding an optimum 'k' value is performed by Elbow method and clustering is done by k-means algorithm, hence routing protocol LEACH which is a traditional energy efficient protocol takes the work ahead of sending data from the cluster heads to the base station. The results of simulation show that at the end of some certain part of running the proposed algorithm, at some point the marginal gain will drop dramatically and gives an angle in the graph. The correct 'k' i.e. number of clusters is chosen at this point, hence the "elbow criterion".

## General Terms

WSN, K-Means, Elbow method, LEACH.

## Keywords

Energy, WSN, K-means, Elbow method, LEACH.

## 1. INTRODUCTION

Wireless Sensor Networks have been receiving a great amount of attention recently due to their substantial applicability to improve our lives. They aid us by extending our ability to accurately monitor, study, and control objects and environments of various scales and conditions such as human bodies, geological surveys, habitats, and security surveillance. Large no. of sensor nodes in a field connected with a sink node to transmit information about events to satellite associated is shown in Figure 1. This figure shows sensed data is delivered to the user. Suppose data is sensed by the sensor node A inside the sensor field. Since the transmission range of radio for each sensor is short, A, at first, passes sensed data to the neighbor node B. In this example, this data may be routed by the path A-B-C-D-E-Sink. Since sink is already connected

to the Internet, it can deliver sensed data to the user directly from sink. Sensor nodes in WSN can also autonomously process and cooperatively analyze sensed data inside networks so that they can prune the redundant data observed inside a network and deliver only necessary data to the user through sinks. Furthermore, WSN can dynamically adapt its topology. After the deployment of sensor nodes in a sensor field, they autonomously find the neighbor nodes and start communicating with each other in various ways, normally using multi-hop communications. In wireless communication and embedded micro-sensing technologies, the advancements encourage the use of WSN today in many environments to detect and monitoring sensitive information. Such environments include border protection, disaster areas, health-related areas, and intelligent house control and many more. WSN are here to detect and track the tanks on a battlefield, tracking the personnel in a building, measure the traffic percentage on a road, monitor environmental pollutants, detect fire and rain. Sensors contribute to electricity production, and also used in collecting the solar energy where WSN tracks the sun rays to detect the power. Now, whether the WSN are starting to become a reality in this world, but there are some limitations such as change in topology randomly, restrictions in power, limited computational resources like power, error-prone medium, energy-efficiency. The energy consumption is an important limitation of WSN which demands researcher's skills to get a way in reducing the energy consumptions by sensor nodes used in WSN.

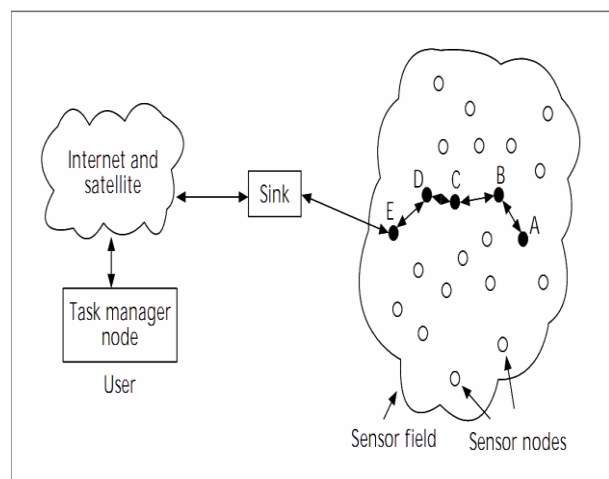


Fig. 1 Sensor nodes scattered in a sensor field

Wireless sensor network consists of number of sensor nodes clustered together to collect various environmental or other physiological data like temperature, pressure, seismic graph, sound, location, etc. Sensor nodes run on battery without

direct power supply. Hence, energy efficiency becomes the major issue in WSN to make it run for longer. Clustering means dividing of large sensor network into small manageable units called 'clusters' which do data aggregation tasks so that WSN can become more scalable and energy efficiency also can improve. To achieve better scalability and to achieve the energy efficient routing of the network, the usage of the clustering scheme for routing is an important factor in WSN. Along these advantages, clustering scheme has one issue that is how to determine the optimal number of clusters so that energy consumption can be reduced.

### 1.1 Clustering

Clustering means dividing of large sensor network into small manageable units called 'clusters' which do data aggregation tasks so that WSN can become more scalable and energy efficiency also can improve. To improve the scalability of the network and to achieve the energy efficient routing, the usage of the clustering scheme for routing is an important factor in WSN. Along these advantages, clustering scheme has more advantages like conserving bandwidth using while communication within the clusters, reduction of redundant transferring of messages between nodes in a network, and also setting-up local routes within the clusters.

## 2. REVIEW LITERATURE

There are many algorithms which has been derived previously to determine k automatically. Most of them methods are wrappers around k-means or some other clustering algorithm for fixed k. Using wrapper method splits and merges rules for centers to increase or decrease k value as the algorithm proceeds.

In [9], Pelleg and Moore work on a regularization framework for learning k and they call it X-means. Pelleg's algorithm try to find many k values and scores to each model of clustering using the so-called Bayesian Information Criterion(BIC). The following formula is used to calculate the BIC value:

$$BIC(C|X) = L(X|C) - p \cdot 2 \log n$$

where  $L(X|C)$  is the log-likelihood of the dataset X according to model C,  $p = k(d + 1)$  is the number of parameters in the model C with dimensionality d and k cluster centers, and n is the number of points in the dataset.

After calculating BIC for each model of clustering, X-means chooses the model which has best BIC score on the data. There are other scoring functions are also available aside BIC.

Some of researchers uses the method of MDL for finding best k. Bischof et al. [10] also uses a minimum description length (MDL) framework, where the description length is a measuring vale which tells us about how well the data are fit by the model. This algorithm starts with a large value for k and removes centers (reduces k) whenever that choice reduces the description length. Between steps of reducing k, they use the k-means algorithm to optimize the model fit to the data.

## 3. PROBLEM FORMULATION

As studied in literature review, we come to know that there are many clustering algorithms which requires the pre-set k value for deciding number of clusters by the user. K value has a great significance upon which the performance of algorithm greatly depends. Now a serious question occurs in a mind which is that 'which K should be chosen when there is no availability of prior knowledge about the cluster quantity?'

## 4. K-MEANS ALGORITHM

K-MEANS proposed by Geon [1] is one of the most popular clustering methods which are based on partitioning. It is very simple and fast algorithm for cluster head selection.

### 4.1 Network Model of K-Means

1. Each node has an ID number.
2. All nodes are fixed or pseudo-static.
3. All nodes are able to send the data to the BS.
4. All nodes are able to control their energy consumption.
5. The initial energy of all the nodes is same.
6. The CHs are aware of their remaining energy.
7. The sensor nodes are randomly distributed in the target area.

### 4.2 Steps of K-Means

#### Step 1: Initial clustering

K-means algorithm is executed for cluster formation with the target WSN. Assume that the WSN of n nodes is divided into k clusters. First, k out of n nodes is randomly selected as the CHs. Each of the remaining nodes decides its CH nearest to it according to the Euclidean distance.

#### Step 2: Reclustering

After each of the nodes in the network is assigned to one of k clusters, the centroid of each cluster is calculated. With the new CH in each cluster, Step 2 is recursively executed until the CH is not changed any more.

#### Step 3: Choosing the CH

After the clusters are formed, an ID number is assigned to each node of a cluster according to the distance from the centroid, assigning smaller number to the closer one. Figure 5 shows the ordering of the sensor nodes with the ID number.. The ID number of a node indicates the order to be chosen as the CH. Therefore, the ID number plays an important role in the selection of a node as CH.

---

#### Algorithm 1: K-Means

---

##### // Creation of Clusters

1. Select random K points as initial Centroid
2. **Repeat**
3. Form K clusters by assigning all points to the closest centroid
4. Recalculate Centroid for each cluster
5. **Until** The Centroid don't change

##### // Reselection of Cluster Head

1. Input on Base Station: the nodes with **ID number**
2. If Energy of cluster head < Energy of threshold
3. then
4. All nodes ← **CheckID()**
5. Current cluster head = **ChangeHeader()**
6. All nodes ← **InformMsg()**
7. **Send data to BS.**

The residual energy of the CH is checked every round to retain the connectivity of the network. If the energy of the CH is smaller than the preset threshold, the node in the next order is selected as a new CH. The newly elected CH informs other nodes of the change of the CH.

## 5. ELBOW METHOD

Elbow method is a method which looks at the percentage of variance explained as a function of the number of clusters. This method exists upon the idea that one should choose a number of clusters so that adding another cluster doesn't give much better modelling of the data. The percentage of variance explained by the clusters is plotted against the number of

clusters. The first clusters will add much information but at some point the marginal gain will drop dramatically and gives an angle in the graph. The correct 'k' i.e. number of clusters is chosen at this point, hence the "elbow criterion". The idea is that Start with  $K=2$ , and keep increasing it in each step by 1, calculating your clusters and the cost that comes with the training. At some value for  $K$  the cost drops dramatically, and after that it reaches a plateau when you increase it further. This is the  $K$  value you want. The rationale is that after this, you increase the number of clusters but the new cluster is very near some of the existing. In the fig. 1, the distortion  $J()$  goes down rapidly with  $K$  increasing from 1 to 2, and from 2 to 3, and then  $W$  reach an elbow at  $K=3$ , and then the distortion goes down very slowly after that. And then it looks like maybe using three clusters is the right number of clusters, because that's the elbow of this curve. Distortion goes down rapidly until  $K=3$ , and really goes down very slowly after that hence number of cluster needed for this data set is 3.

---

**Algorithm 1: Elbow Method to determine K of K-means**

---

1. Initialize  $k=1$
2. Start
3. Increment the value of  $k$
4. Measure the cost of the optimal quality solution
5. If at some point the cost of the solution drops dramatically
6. That's the true  $k$ .
7. End

Based on pre-evaluated cluster number, the cluster nodes start the computations and divide themselves in the clusters according to the pre-evaluation. The cluster nodes divide themselves in the pre-evaluated number of clusters using Euclidean distance calculation. The cluster formation is performed using the K-Means algorithm. K-Means algorithm has already been proved faster than LEACH algorithm in the cluster nodes. LEACH algorithm is still the most balanced and most adaptive algorithm to most of the WSN scenarios. The K-means uses the Euclidean distance on the basis of each node's position coordinates ( $X$  &  $Y$ ) in two way ground. The K-means then computes the distance formula and divide the cluster nodes in particular number of clusters.

## 6. LEACH

Heinzelman, et.al [2] introduced LEACH protocol which is a hierarchical based. It is a first hierarchical clustering algorithm for WSN. LEACH i.e. Low Energy Adaptive Cluster Hierarchy based protocol is a hierarchical clustering energy efficient routing protocol which reduces the energy consumption so that the network life can be increased of WSN. In LEACH, clusters are formed. And each cluster has its own CH which is selected randomly by selecting a random number either 0 or 1 by the node so that the energy dissipation among nodes can be balanced.

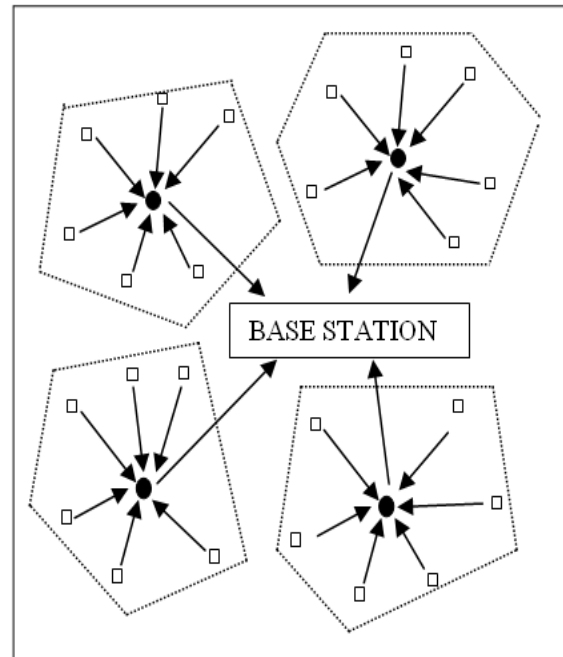
The following are the steps of each round of this protocol:

### Step 1: Advertisement phase

In this initial step of LEACH protocol, all nodes in a range are issued a notification by the eligible CH so that they can become a cluster member in respective cluster. The nodes accepts the offer if they have adequate signal i.e. RSS (Received Signal Strength).

### Step 2: Set-up phase

In this step, the nodes respond to their respective CH.



**Fig 2. Leach protocol architecture.**

### Step 3: Schedule creation

At the time, the cluster head gets response from the nodes, and then the CH makes a TDMA scheme and sends this scheme via message to its cluster members to intimate them about the schedule when they have to transmit their information to CH.

### Step 4: Data transmission

The individual sensor collects data during its time interval and then transmits to the CH. After its fixed time period schedule, the cluster members off their radio signals to reduce its energy usage. CH fuses the data it collected and passes that data (fused data) to the Base Station.

In LEACH, there are different rounds having different in each round. Once a node is chosen as CH, then it cannot become CH for next ' $p$ ' rounds. Thereafter, probability of each node for becoming CH in each round is ' $1/p$ '. At the end of each round, each non-CH node chooses the closest CH and becomes a part of that cluster for the transmission of data.

## 7. PROPOSED SCHEME

The energy efficiency and response time are the major wireless sensor network issues. In this research project, a hybrid energy efficient quick k-means and LEACH based clustering algorithm using automatic cluster number selection is being developed. The proposed algorithm will be hierarchical clustering based segmentation mechanism. The performance of the proposed protocol will be compared with that of existing k-means and LEACH through simulation experiments. The proposed protocol would be created with flexible functioning to deal with large number of nodes. The functional mechanism to decide the number of clusters dynamically will be developed to empower the proposed algorithm. The proposed protocol will produce better scalability to the protocol for dealing with very large wireless sensor networks.

The proposed algorithm is a solution to enhance the energy efficiency than the existing k-means and LEACH protocols. The proposed algorithm is the combination of K-Means clustering and LEACH protocol. The two protocols has been combined to enhance the network lifetime by making the

clustering process energy efficient and quicker. A unique algorithm is programmed to decide the number of cluster automatically. This algorithm takes place at the base station node. The base station gathers the data of all of the nodes and computes the algorithm to get the possible number of clusters. This algorithm uses the mechanism of evaluation using various distance based formulas to evaluate the number of clusters. This computation is performed on the base station because base stations are not only connected to direct supply, also very quick to perform the computations. If these computations were performed on the cluster nodes, it would choke a lot of energy on those nodes. Because the WSNs are not connected to the direct power, the battery choked once cannot be charged back. That's the reason that pre-evaluation of number of clusters is performed on the base station node. The pre-evaluation result is then forwarded to all of the nodes in the cluster.

In this research, we are working on the development of a hybrid model using LEACH based energy efficient and K-means based quick clustering algorithms to produce a new cluster scheme for WSN with dynamic selection of the number of the clusters automatically. The energy efficiency and response time are the major wireless sensor network issues. In this paper, we proposed an energy enhanced protocol which uses K-means clustering scheme for selecting the cluster heads for each cluster and hence routing protocol LEACH which is a traditional energy efficient protocol takes the work ahead of sending data from the cluster heads to the base station. But the k means requires input parameter 'k' that's no. of clusters. An inappropriate choice of k may lead to poor results. That's why we use 'Elbow method' for determining no. of clusters 'k' which gives well distributed data due to roughly equal size clusters.

The performance of the proposed protocol will be compared with that of existing k-means and LEACH through simulation experiments. The proposed protocol would be created with flexible functioning to deal with large number of nodes. The functional mechanism to decide the number of clusters dynamically will be developed to empower the proposed algorithm. The proposed protocol will produce better scalability to the protocol for dealing with very large wireless sensor networks. In this, k means and LEACH protocols are combined. Then the algorithm flow was reviewed and refined whenever the change/improvement was required. Afterwards, the algorithm is programmed using MATLAB simulator because it its flexibility and a large number of pre-programmed functions. The MATLAB environment helped us to develop the proposed algorithm quickly. The experiment results were thoroughly analyzed and compared with the existing algorithm results. The proposed algorithm is a solution to enhance the energy efficiency than the existing k-means and LEACH protocols. The proposed algorithm is the combination of K-Means clustering and LEACH protocol. The two protocols have been combined to enhance the network lifetime by making the clustering process energy efficient and quicker. A unique algorithm is programmed to decide the number of cluster automatically. This algorithm takes place at the base station node. The base station gathers the data of all of the nodes and computes the algorithm to get the possible number of clusters. This algorithm uses the mechanism of evaluation using various distance based formulas to evaluate the number of clusters. This computation is performed on the base station because base stations are not only connected to direct supply, also very quick to perform the computations. If these computations were performed on the cluster nodes, it would choke a lot of energy on those

nodes. Because the WSN are not connected to the direct power, the battery choked once cannot be charged back. That's the reason that pre-evaluation of number of clusters is performed on the base station node. The pre-evaluation result is then forwarded to all of the nodes in the cluster.

---

#### Algorithm 5: Proposed Algorithm

---

1. Initialize  $k=1$
2. Start
3. Measure the cost of the optimal quality solution
4. If at some point the cost of the solution drops dramatically
5. Cluster number incremental
  - a. Condition is true
  - b. Increment the value of  $k$  by 1
6. End
7. Choose a method to initialize the  $n_i$  to be the mean of cluster
 

*Initialize  $n_i, i = 1, \dots, k$ , for example, to  $k$  random  $x^t$*
8. Repeat
9. For each rotation, assign it to the closest group and represented by  $n_i$ 

*For all  $x^t$  in  $X$*

$$CH_i^t \leftarrow 1 \text{ if } \|x^t - n_i\| = \min_j \|x^t - n_j\|$$

$$CH_i^t \leftarrow 0 \text{ otherwise}$$
10. For each  $n_i$ , recalculate it based on the examples that are currently assigned to it.
 

*For all  $n_i, i = 1, \dots, k$*

$$n_i \leftarrow \text{sum over } t (CH_i^t x^t) / \text{sum over } t (CH_i^t)$$
11. Until  $n_i$  converge (Repeat the steps 8 to 10 until  $n_i$  converge)
12. Cluster Head Selection
  - a. Each sensor elects itself to be cluster head at the beginning of a round
  - b. Nodes that have not already been cluster heads recently, may become cluster heads
  - c. Probability of becoming a cluster head is set as a function of nodes' energy level relative to the aggregate energy remaining in the network
    - i. Average energy of nodes in each cluster  $X$   
Number of nodes in the network
13. Cluster Formation
  - i. Each cluster head node broadcasts an advertisement message (ADV) using CSMA MAC Protocol
    - o The message consists of the nodes' ID and a header that distinguishes it as an ADV message
  - ii. Each non-cluster head node determines its cluster/cluster head that requires minimum communication energy
    - o Largest signal strength, minimum transmit energy for communication

- iii. Each node transmits a join-request message (REQ) using CSMA MAC Protocol
  - o The message consists of node's ID and cluster head ID
- iv. Each cluster head node sets up a TDMA schedule and transmits it
  - o This ensures that there is no collision in data messages, radio components can be turned off at all times except during transmit time
- v. Once the cluster head receives all data it performs data aggregation
- vi. Resultant data is sent from cluster head to BS (a high energy transmission)
- vii. Uses transmitter based code assignment to reduce inter-cluster interference
- viii. Cluster head senses the channel before transmission

Based on pre-evaluated cluster number, the cluster nodes start the computations and divide themselves in the clusters according to the pre-evaluation. The cluster nodes divide themselves in the pre-evaluated number of clusters using Euclidean distance calculation. The cluster formation is performed using the K-Means algorithm. K-Means algorithm has already been proved faster than LEACH algorithm in the cluster nodes. LEACH algorithm is still the most balanced and most adaptive algorithm to most of the WSN scenarios. The K-means uses the Euclidean distance on the basis of each node's position coordinates (X & Y) in two way ground. The K-means then computes the distance formula and divide the cluster nodes in particular number of clusters.

In this proposed work, we use LEACH on each cluster computed by K-Means. Each cluster will elect its cluster head using LEACH protocol in each round. If LEACH runs for P times, it elects cluster head on each round. This means each cluster is having P percentage as desired percentage of cluster heads. Every node has a probability of 1/P of becoming a cluster head in every round. Each node after becoming cluster head once cannot become cluster head for next P rounds.

Then the LEACH protocol is used in the each cluster to find the best transmission path to transmit the sensor data to the base station to facilitate the maximum network lifetime by making the sensor nodes energy efficient. LEACH protocol (Low energy adaptive cluster hierarchy) is a TDMA based MAC protocol for WSN. LEACH creates and maintains the clusters to improve the network lifetime of WSN. LEACH is a hierarchical clustering protocol. It facilitates the connectivity of cluster nodes with the cluster head. Every node uses stochastic algorithm at each node on all rounds to determine itself a cluster head in that particular round. LEACH also facilitates the route discovery for the WSN nodes in that cluster. The cluster head after receiving the data from the other cluster nodes aggregate and compress the data and forward to the BTS or sink.

## 8. RESULTS AND DISCUSSIONS

The proposed algorithm is the unique & balanced combination of three unique algorithms. The elbow method is used to determine the number of clusters. The elbow method uses percentage of variance. The variance is defined as a function of the cluster number. The total variance is calculated by using F-test. This method is used to determine the best number of clusters to ensure the best quality clusters. The quality clustering (especially correct number of clusters) is effective to improve the energy efficiency.

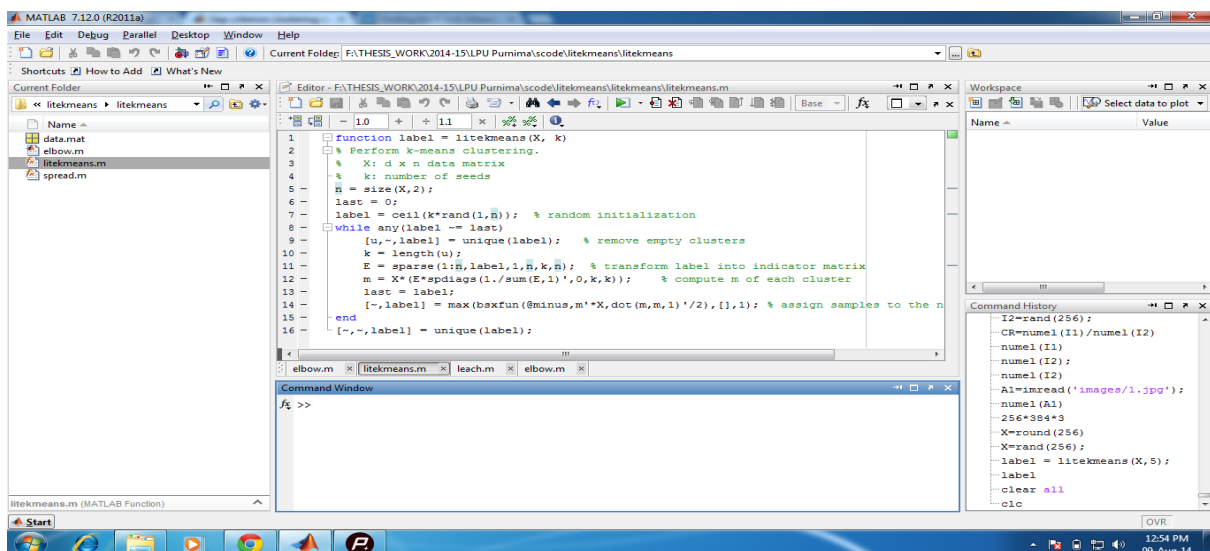
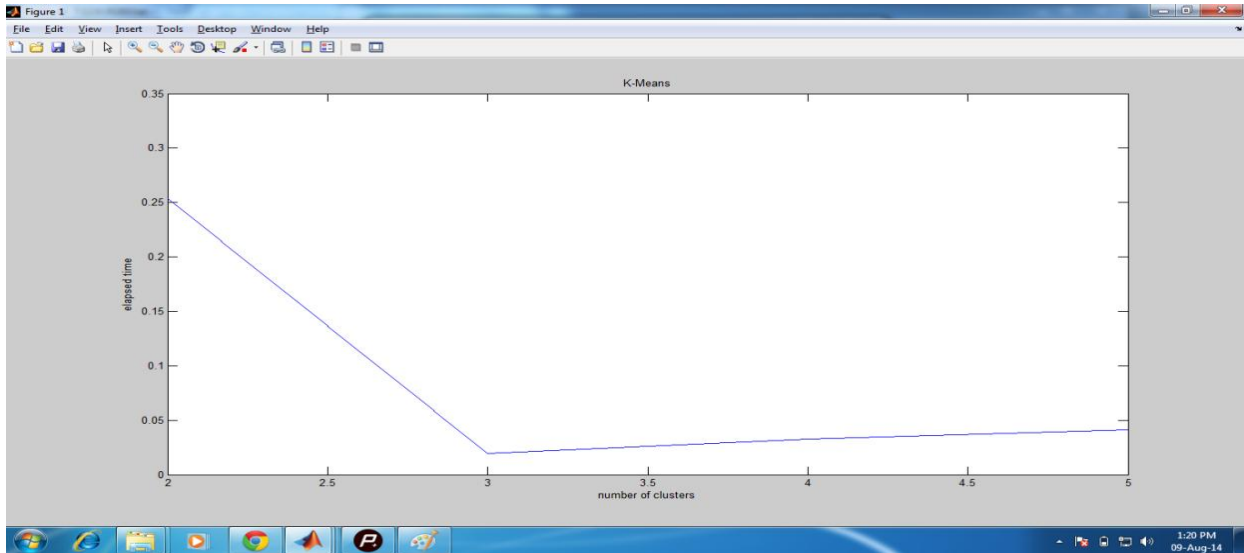


Fig 3. The MATLAB interface with K-Means code open

The project is developed using MATLAB because of its ease of interface and availability of pre-programmed necessary functions. Our hybrid algorithm is combination of K-Means and LEACH clustering. K-Means clustering and LEACH has been combined to improve the clustering task to improve the cluster quality and to make the whole process enough energy efficient to increase the WSN network lifetime. The input dataset undergoes the cluster number determination using the

elbow method. The result of elbow method is passed as K to the K-Means clustering algorithm. K-Means then create the cluster of WSN data-using X, Y and Z location coordinates. K-Means evaluate the WSN data to produce the clusters and results the cluster IDs. The clusters are then divided on the basis of cluster IDs and clustering information is forwarded to the LEACH protocol.



**Fig 4. Elapsed time graph for different number of clusters**

The K-Means algorithm has been evaluated for different number of cluster on the data. The data size taken as input dataset includes the data of 2000 nodes and stores X, Y and Z location coordinates of each node. Each location coordinates are evaluated using K-Means algorithm by using location aware clustering to divide the wireless sensor nodes in recommended number of clusters. The cluster number was recommended by latter mentioned elbow method of cluster number determination. The results of K-Means are produced quickly and effectively. This latter point has been proved in the base paper. The cluster time varies according to the number of clusters been decided to divide the wireless sensor nodes. The K-Means clustering algorithm has been proved to be way faster than the existing scheme LEACH. But the LEACH scheme is effective to produce the clusters and to transmit the data from the nodes to the base station or sink.

**Table 1. Elapsed time by K-Means for different number of clusters**

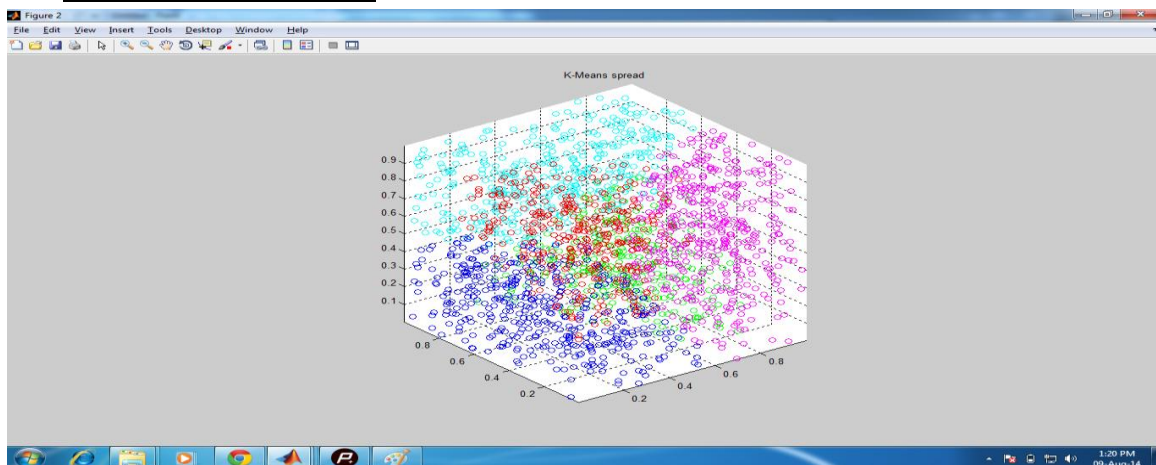
Cluster Number	Elapse Time (ET)
2	0.19
3	0.024
4	0.22

5	0.025
---	-------

**Table 2. Results in the shape of statistical errors**

Number of Clusters	Davies-Bouldin Index (DB)	Calinski-Harabasz Index (CH)	Krzano wski and Lai Index (KL)	Criterion Error
2	1.6216	683.72	589.30	185.62
3	1.3168	684.69	614.85	98.53
4	1.0629	773.09	580.75	57.62
5	0.9953	781.90	567.41	38.81

The total time for all the process has been also recorded. The total time has been recorded as 0.3465 seconds.



**Fig 5. Spread cluster data of K-Means result for five clusters**

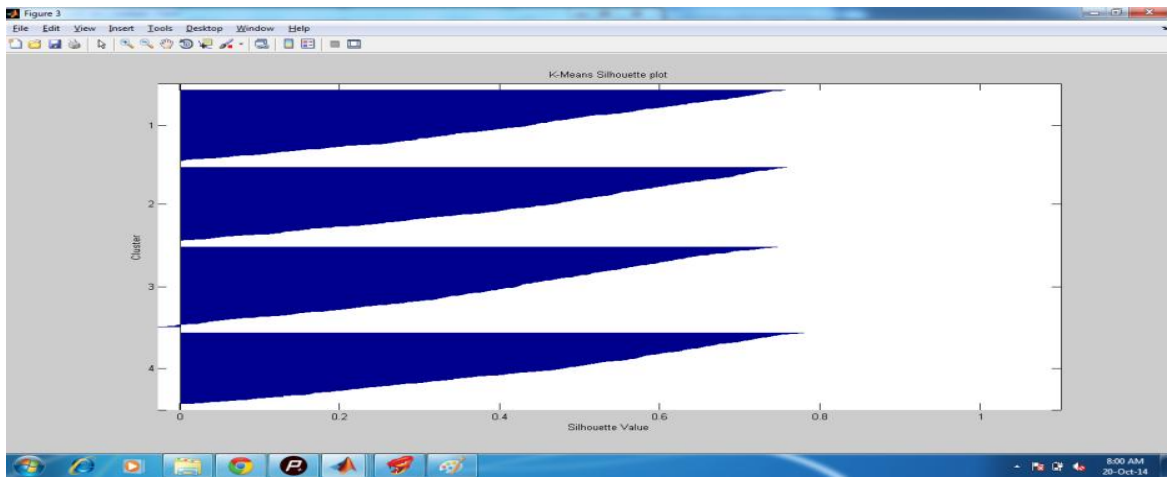


Fig 6. Silhouette Plot of K-Means clustering data for five clusters

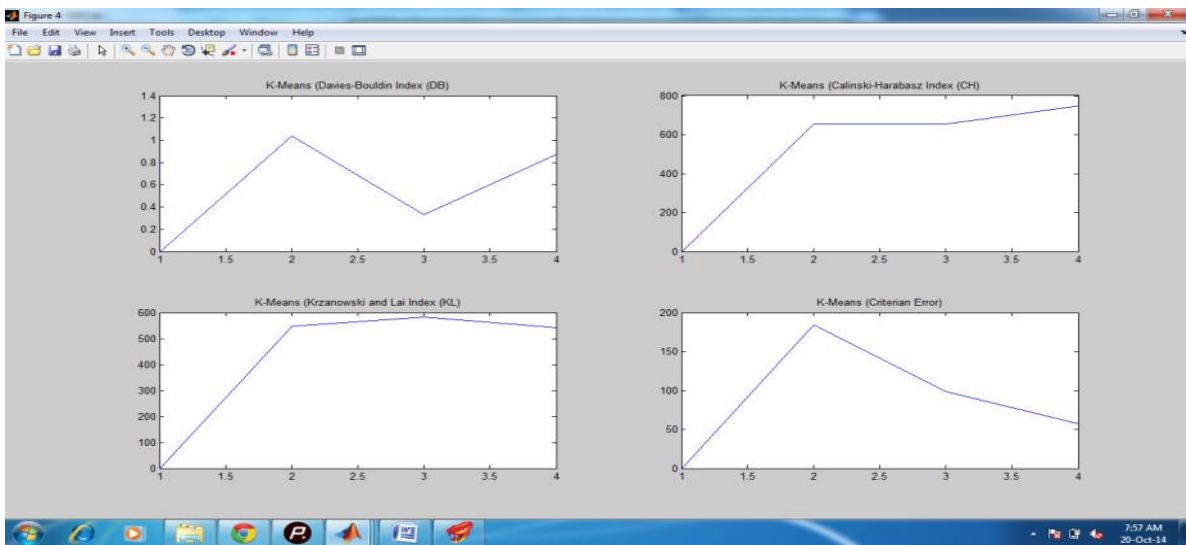


Fig 7. Statistical Errors of clustering results

MATLAB is having many built-in tools like statistical tool box, which facilitate the calculation of various types of statistical errors to measure the cluster quality. The quality of the clusters is measured in this project by using Davies-

Bouldin Index, Calinski-Harabasz Index, Krzanowski and Lai Index & Criterion Error. Silhouette plotting has been used to plot the quality of the cluster sets that are effective in the cluster created by our proposed Hybrid algorithm.

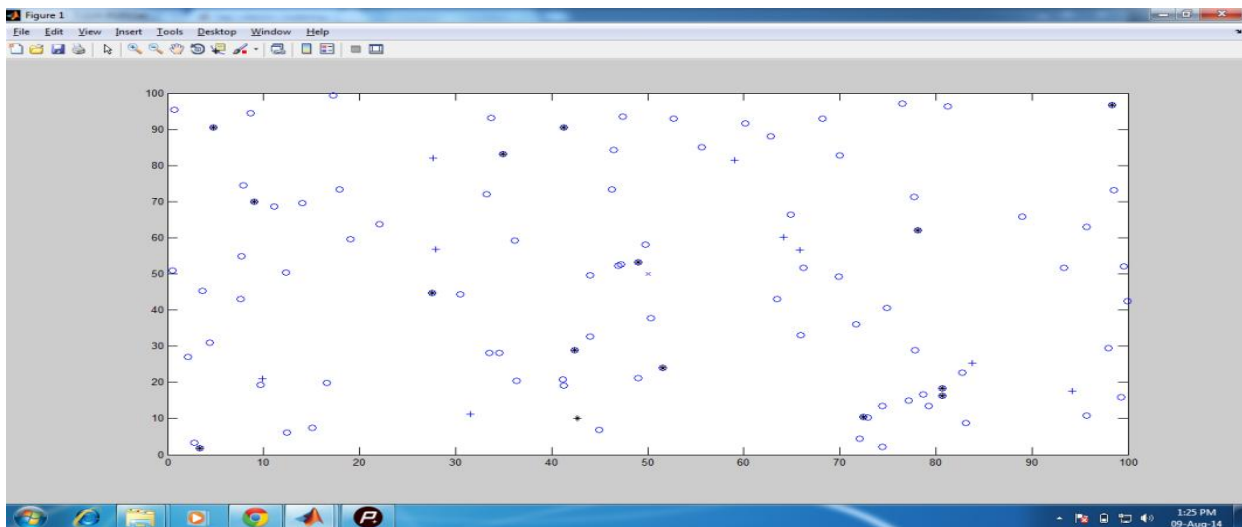


Fig 8. The clustering of the proposed Algorithm

## 9. CONCLUSION

The research project has been implemented in the MATLAB simulator. The results have been analyzed in detail. The results have proved that the new proposed method has been proved itself efficient than the existing routing scheme based on various performance parameters. The cluster quality has been analyzed by various statistical cluster quality measuring parameters. The cluster quality has been measured by Davies-Bouldin Index, Calinski-Harabasz Index, Krzanowski and Lai Index & Criterion Error. Results of all of the indices has proved that the clustering has been proved to be enough efficient to make the whole process energy efficient and improves the network lifetime. Silhouette plotting has been used to plot the quality of the cluster sets. The results prove that the wireless sensor network has been divided into the most balanced clusters and they are proved to be effective using our proposed Hybrid algorithm. It has also been proved that the clustering has been enough quick process using our proposed method than the existing processes. In future the clustering process can be improved to perform better than proposed and existing clustering schemes using improved, hybrid or new methods of WSN clustering. The scheme can be improved to produce more quality clusters with much better cluster number estimation algorithm. The cluster number estimation algorithm can also improved to produce more accurate results.

## 10. ACKNOWLEDGMENTS

I thanks to my respectful guide Asst. Prof. Arvind Kumar who have contributed towards development of the paper.

## 11. REFERENCES

- [1] Geon Yong Park, Heeseong Kim, Hwi Woon Jeong, and Hee Yong Youn, "A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network", 2013 27th International Conference on Advanced Information Networking and Applications Workshops, 20130325.
- [2] Heinzelman, W., Chandrakasan, A., and Balakrishnan, H., "Energy-Efficient Communication Protocols for Wireless Microsensor Networks", Proceedings of the 33rd Hawaiian International Conference on Systems Science (HICSS), January 2000.
- [3] John M. Shea, Joseph P. Macker, "Automatic Selection of Number of Clusters in Networks using Relative Eigenvalue Quality, IMCC, vol. 1, pp. 131-136, IEEE 2013.
- [4] Study on WSN Topology Division and Lifetime XU Jiu-qiang, WANG Hong-chuan, LANG Feng gao, WANG Ping, HOU Zhen-peng, IEEE, 2011
- [5] Hierarchical Adaptive Balanced energy efficient Routing Protocol (HABRP) for heterogeneous wireless sensor networks BY Said BEN ALL\*, Abdellah EZZATI, Abderrahim BENI HSSANE, Moulay Lahcen HASNAOUI, IEEE, 2010
- [6] Raymond Wagner, Shriram Sarvotham, Hyeokho Choi, Richard Baraniuk, "Distributed Multiscale Data Analysis and Processing For Sensor Networks", Rice University Technical Report, February 9, 2005.
- [7] Scott Briles, Joseph Arrowood, Dakx Turcotte, Etienne Fiset, "Hardware-In-The-Loop Demonstration of a Radio Frequency Geolocation Algorithm", Proceedings of the Mathworks International Aerospace and Defense Conference, May 24-25, 2005.
- [8] Sundeep Pattem, Bhaskar Krishnamachari, and Ramesh Govindan, "The Impact of Spatial Correlation on Routing with Compression in Wireless Sensor Networks," ACM/IEEE International Symposium on Information Processing in Sensor Networks (IPSN), April 26-27, Berkeley, CA 2004.
- [9] Dan Pelleg and Andrew Moore. X-means: Extending K-means with efficient estimation of the number of clusters. In Proceedings of the 17<sup>th</sup> International Conf. on Machine Learning, pages 727–734. Morgan Kaufmann, San Francisco, CA, 2000.
- [10] Horst Bischof, Ales Leonardis, and Alexander Selb. MDL principle for robust vector quantisation. *Pattern analysis and applications*, 2:59–72, 1999.
- [11] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323, 1999