

An Energy Efficient Hybrid Clustering Algorithm for WSNs to Increase Network Lifetime

Kanchan Deep Nagi¹ Gagandeep Singh²

¹Research Scholar ²Assistance professor

^{1,2}Department of Computer & Science Engineering

^{1,2}CGC, Landran

Abstract— Wireless Sensor Networks (WSNs) consist of hundreds of small and cost effective sensor nodes. Sensor nodes are used to sense the physiological or environmental parameters (e.g. 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.

Key words: WSN, network lifetime, clustering, cluster number prediction

I. INTRODUCTION

Over last some decades, the evolution in computers has been happened at a large in the case of processing power and at the same time decreased in term of size and also in price. In our society's daily activities the computers would participates abundant role. From past few years, computers are becoming so small and so cheap, in both economically and theoretical pont of view, this is the one of the great revolution has been done. Wireless sensor network are beginning to become a reality and receiving a great amount of attention recently so it plays an important role to improve our lives in every filed of aspect and also an important area of research. They help us to expands our knowledge in research to get the things accurately monitor, study, and control objects and environmental factors such as geological surveys, ssecurity surveillance and many more. In wireless sensor networks large no. of sensor nodes in a field connected with a sink node to transmit information about events to satellite. Sensor nodes is bacially sense data, store data, route data and processes data, so sensor nodes is also perform an important role in wireless senser network.

In wireless communication the sensor nodes send and receive the information to the system so the use of WSNs in many environments (likes border protection, disster areas, intelligent house control, health realted areas and many mores) to detect and monitoring sensve information. Wireless sensor networks are able to detect weather forecasting, track the tanks in battlefield, measures the traffic percentage on road, monitor environmental pollutants, for the power detection WSNs id helpful in

tracking Sunrays. In WSNs the sensor collect the solar energy which is cable of produce electricity.

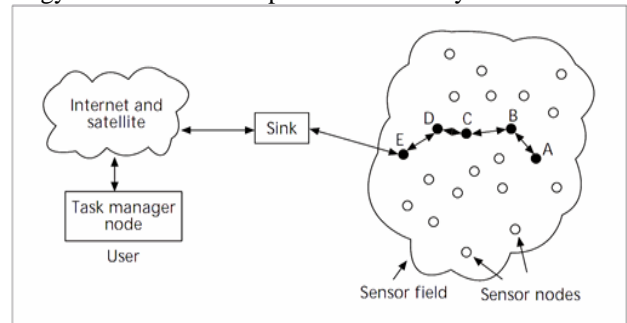


Fig. 1: The WSN clustering example in real-time scenario.

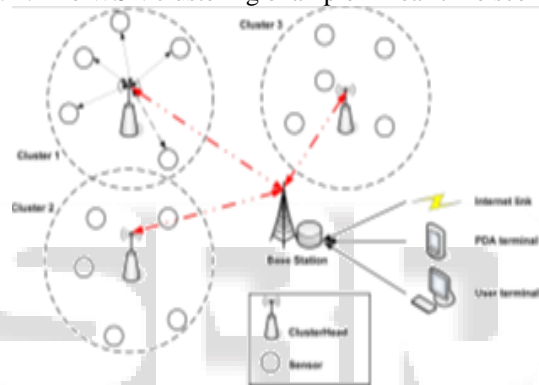


Fig. 2: The General Sensor Network Architecture

Wireless sensor networks consist of hundreds of tiny sensor nodes each sensor nodes form a group to detect data and retrieve data within the system so that the WSNs become more scalable and also improve the energy efficient, that groups which receive and retrieve data is basically form a cluster, in simple words when the large sensor nodes network is divided in to small units nodes then that unit node is known as 'Cluster'. Every cluster is managed by node cluster head (CH) and other nodes are referred as cluster nodes. Cluster nodes do not communicate directly with the sink node. Cluster head will aggregate the data, received from cluster nodes and transmits it to the base station. which minimizes the energy consumption and number of messages communicated to base station.

A. Sensor Node:

Sensor node is the important component of WSN because of its multiple role features. It senses data, stores data, routes data and processes data.

B. Clusters:

Clusters are small manageable units which simplify tasks such a communication.

C. Cluster Heads:

Clusterheads are the leader who organizes a cluster activities. It collects data from several sensor nodes and then

aggregates those data and also organizes the schedule of a cluster for communication with BS.

D. Base Station:

Base station is a central component which collects data from several nodes distributed at different locations. The deployment of base station is also a critical issue of WSN. It acts as an intermediate between the network and end-user.

E. End User:

The data in a sensor network can be used for a wide-range of applications. Therefore, a particular application may make use of the network data over the internet, using a PDA, or even a desktop computer. In a queried sensor network (where the required data is gathered from a query sent through the network). This query is generated by the end user. The clustering phenomenon, plays an important role in not just organization of the network, but can dramatically affect network performance.

II. LITERATURE REVIEW

Chen, Jiejie [1] have worked on improving the life time of Wireless Sensor Networks by Using Fuzzy c-means Induced Clustering. The authors have worked upon a new algorithm named FCM (Improving Life Time of Wireless Sensor Networks by Using Fuzzy c-means Induced Clustering) for the WSN clusters. The authors have divided all nodes into Q clusters by the Fuzzy c-means algorithm and chose the cluster head according to the node's energy and the distance between the center position of each cluster and base station in the first round. Then fixed clusters have been used in the proposed scenario. As the new round begin, the choose of cluster head node is based on the current energy value and the distance to the base station in a cluster. During the data transmission, the communication took the single-hop in the cluster and multihop in inter-cluster. The simulation results showed that the algorithm was rational and efficient.

Park, Geon Yong [2] et. al. have proposed and developed a novel cluster head selection method based on k-means algorithm for energy efficient WSNs. In this paper the authors have proposed an efficient cluster head selection method using K-means algorithm to maximize the energy efficiency of wireless sensor network. It is based on the concept of finding the cluster head minimizing the sum of Euclidean distances between the head and member nodes.

Mohan, Chander, [3] and Ashok Kumar have developed a heterogeneous fuzzy based clustering algorithm for WSNs. The authors have put their effort to handle the problem of heterogeneity in the WSN networks. In this paper, the authors have develop a clustering scheme using energy prediction for appropriate cluster head selection to optimize the energy of wireless sensor node. In this work, we introduced three different types on sensor nodes on the basis on their initial energy level. In Heterogeneous Fuzzy Based Clustering Protocol (HFCEP) firstly, eligible cluster head candidates are selected using Fuzzy Inference Engine (FIS) using inputs as Residual Energy (RE) and Predicted Residual Energy (PRE) and output as the Chance to be a Cluster Head (CCH), then from the eligible candidates set cluster heads are elected using probabilistic approach including remaining energy of every node and initial energy of node.

Said, Jihed Eddine, [4] Lutful Karim, Jalal Almhana, and Alagan Anpalagan have worked together on the heterogeneous mobility and connectivity-based clustering protocol for wireless sensor networks. The authors have addressed the problem of the inter-connectivity of the WSN nodes. In this paper, the authors have introduced a velocity-based clustering (or clustering the mobile nodes) algorithm and implement a relay placement technique in order to maintain seamless network connectivity. Simulation results show that the packet loss rate of the proposed algorithm is much lower than the existing LEACH and HEED clustering protocols.

Peng, Jian [5] et. al. has developed an energy-efficient prediction clustering algorithm for multilevel Heterogeneous WSNs. This algorithm enables the nodes to select the cluster head according to factors such as energy and communication cost, thus the nodes with higher residual energy have higher probability to become a cluster head than those with lower residual energy, so that the network energy can be dissipated uniformly. In order to reduce energy consumption when broadcasting in clustering phase and prolong network lifetime, an energy consumption prediction model is established for regular data acquisition nodes.

Tandon, Ravi, Biswanath Dey, and Sukumar Nandi [6] have done their research upon weight based clustering in wireless sensor networks. The authors have identified the potential drawback of clustering schemes in the case of heterogeneous WSNs. In a heterogeneous sensor network, electing sensor nodes with high energy which lie within the same transmission radius may be better than electing a single advanced node as cluster head. The authors have worked upon the weighted cluster head election method, called Weight Based Clustering for Heterogeneous sensor Networks (WBCHN) in this research project.

III. RESEARCH METHODOLOGY

An important component of any research design for the development of a Geocast energy efficient WSN routing protocol is a thorough theoretical and practical understanding and knowledge of the influencing factors which are drawing the specific protocol functions. There a literature study was conducted to understand and investigate the factors that influence the functioning and design of wsn Geocast routing protocol. The literature survey also covers the study of available WSN Geocast routing protocols, in context of understand and identify the common problems faced by WSN Geocast routing protocols. The proposed protocol will be designed and developed by taking specific functional and designs requirements into account and to take on the common Geocast routing protocol flaws detected in the conducted literature study. The protocol will be designed, developed, simulated and analyzed in the NS2 Simulator.

IV. EXPERIMENTAL DESIGN

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.

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.

At first, the literature on the WSN clustering protocols was studied. All of the WSN processes were studied in detail to find the loopholes in the existing processes. The flaws were found in k-means and LEACH protocols. The protocols were found with various loopholes which could be mitigated by combining the two. Then the algorithm flow was reviewed and refined whenever the change/improvement was required. Afterwards, the algorithm is programmed using NS2 simulator because its flexibility and a large number of pre-programmed functions. The NS2 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 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.

Algorithm 1: Elbow Angle and Threshold Method to determine number of cluster from the given data

- (1) Begin with the minimum possible number of clusters in the given dat
- (2) Increment the value of cluster number
- (3) Measure the cost of the optimal quality solution
- (4) When cost of the solution drops up to extreme level
- (5) Declare the number of cluster as K

- (6) Return K

On the basis of 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 to be 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.

Algorithm 2: K-Means Clustering

- (1) Input dataset with WSN node positions
- (2) K equals number of clusters calculated by elbow threshold method
- (3) Algorithm determines the random data points equal to the cluster number
- (4) The algorithm evaluates the distance of each data point from all of the the random initial data points.
- (5) The point is added to the cluster with the lowest distance
- (6) Evaluate all points with method from 4 to 5 until last point.
- (7) Return the cluster data

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 Ptimes, 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.

Algorithm 3: LEACH

- (1) Each sensor elects itself to be cluster head at the beginning of a round & Nodes that have not already been cluster heads recently, may become cluster heads
- (2) Probability of becoming a cluster head is set as a function of nodes' energy level relative to the aggregate energy remaining in the network & Average energy of nodes in each cluster X Number of nodes in the network
- (3) Each cluster head node broadcasts an advertisement message (ADV) using CSMA MAC Protocol where the message consists of the nodes' ID and a header that distinguishes it as an ADV message
- (4) Each non-cluster head node determines its cluster/cluster head that requires minimum communication energy where the Largest signal strength, minimum transmit energy for communication
- (5) Each node transmits a join-request message (REQ) using CSMA MAC Protocol where the message consists of node's ID and cluster head ID
- (6) Each cluster head node sets up a TDMA schedule and transmits it and ensures that there is no

collision in data messages, radio components can be turned off at all times except during transmit time

- (7) Once the cluster head receives all data it performs data aggregation, the resultant data is sent from cluster head to BS (a high energy transmission)
- (8) Uses transmitter based code assignment to reduce inter-cluster interference
- (9) Cluster head senses the channel before transmission

Existing Algorithm is the WSN clustering algorithm based on K-means. The K-means has been implemented on WSN for clustering and has been compared to LEACH protocol which is considered best among the existing WSN clustering protocols. The Existing algorithm flow has been give below:

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 WSNs. LEACH creates and maintains the clusters to improve the network lifetime of WSNs. 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 an 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.

Algorithm 4: Proposed Hybrid Algorithm (Cluster Evaluation & K-Means & LEACH)

- (1) Begin with the minimum possible number of clusters in the given dat
- (2) Increment the value of cluster number
- (3) Measure the cost of the optimal quality solution
- (4) When cost of the solution drops up to extreme level
- (5) Declare the number of cluster as K
- (6) Return K
- (7) Input dataset with WSN node positions
- (8) K equals number of clusters calculated by elbow threshold method
- (9) Algorithm determines the random data points equal to the cluster number
- (10) The algorithm evaluates the distance of each data point from all of the the random initial data points.
- (11) The point is added to the cluster with the lowest distance
- (12) Evaluate all points with method from 4 to 5 until last point.
- (13) Return the cluster data
- (14) Each sensor elects itself to be cluster head at the beginning of a round & Nodes that have not already been cluster heads recently, may become cluster heads
- (15) Probability of becoming a cluster head is set as a function of nodes' energy level relative to the aggregate energy remaining in the network & Average energy of nodes in each cluster X Number of nodes in the network

- (16) Each cluster head node broadcasts an advertisement message (ADV) using CSMA MAC Protocol where the message consists of the nodes' ID and a header that distinguishes it as an ADV message
- (17) Each non-cluster head node determines its cluster/cluster head that requires minimum communication energy where the Largest signal strength, minimum transmit energy for communication
- (18) Each node transmits a join-request message (REQ) using CSMA MAC Protocol where the message consists of node's ID and cluster head ID
- (19) Each cluster head node sets up a TDMA schedule and transmits it and ensures that there is no collision in data messages, radio components can be turned off at all times except during transmit time
- (20) Once the cluster head receives all data it performs data aggregation, the resultant data is sent from cluster head to BS (a high energy transmission)
- (21) Uses transmitter based code assignment to reduce inter-cluster interference
- (22) Cluster head senses the channel before transmission.

V. RESULT ANALYSIS AND COMPARISON

The project is developed using NS-2 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.

```

cluster0 = 0
cluster1 = 0
cluster2 = 0
cluster3 = 0
cluster4 = 0
cluster5 = 0
cluster6 = 0
cluster7 = 0
cluster8 = 0
cluster9 = 0
cluster10 = 0
cluster11 = 0
cluster12 = 0
cluster13 = 0
cluster14 = 0
cluster15 = 0
Agent 11 sends: call neighbors() and cluster()
agentNumber = 11, cluster = 000
NS-2: 1278 - done
Agent connects to existing sim instance. Starting a new sim...
Agent 11: 1279 - done
  
```

Fig. 1: The results of the simulation printed on the command line continued

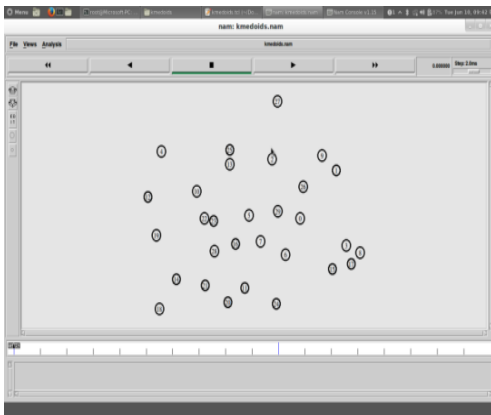


Fig. 2: The mobile nodes are in off Mode.

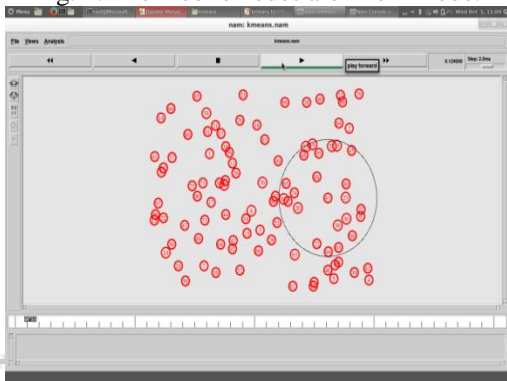


Fig. 3: The mobile nodes powering on

The cluster formation is performed on wireless sensor network consisted of 100 nodes. The nodes have communicated to each other prior to the cluster formation process. The K-means algorithm has been used in this simulation for the purpose of WSN clustering. The simulation is programmed to create three clusters. The cluster formation process starts with the nodes exchanging information with each other about their physical location by propagating X-axis and Y-axis. The algorithm forms the clusters on the basis of their location co-ordinates. After the completion of the process of cluster formation, the process of cluster head selection begins. The cluster head is selected by comparing the area of the cluster with the location of the node. The node found very near to the center of the cluster is selected as cluster head. The simulation contains 100 nodes in an area of 1000 x 800 meters. Each node has a communication or transmission radius of 250 meters under this simulation.

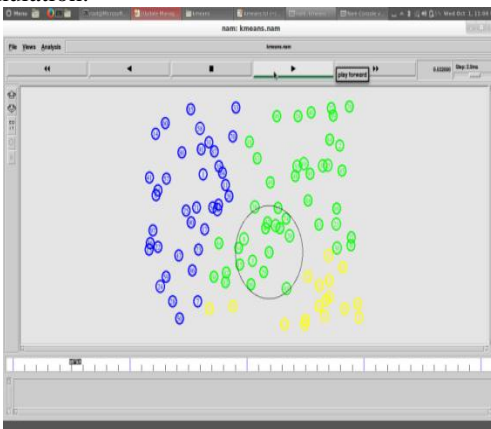


Fig. 4: The nodes after clustering with k-means algorithm

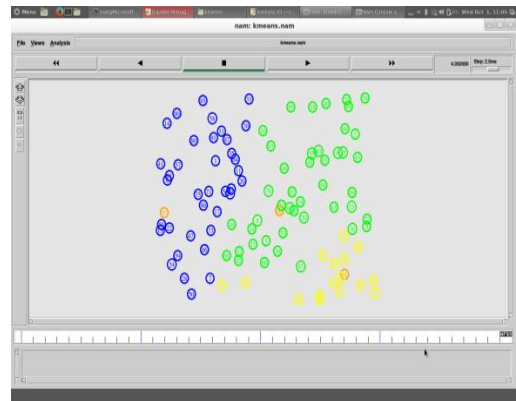


Fig. 5: Cluster Head shown in Orange color

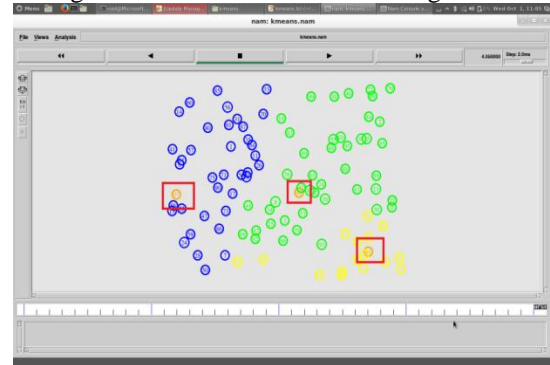


Fig. 6: Cluster Head shown in Orange color marked in boxes

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.

Cluster Number	Elapse Time (ET)
2	0.19
3	0.024
4	0.22
5	0.025

Table 1: Elapsed time by K-Means 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 100 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. The total time for all the process has been also recorded. The total time has been recorded as 0.3465 seconds.

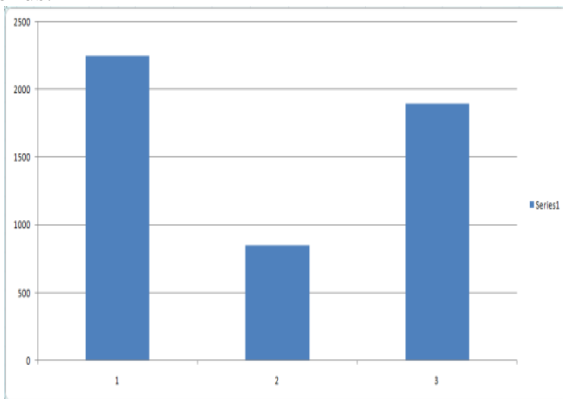


Fig. 7: Total energy of the clusters at the time of initial stage

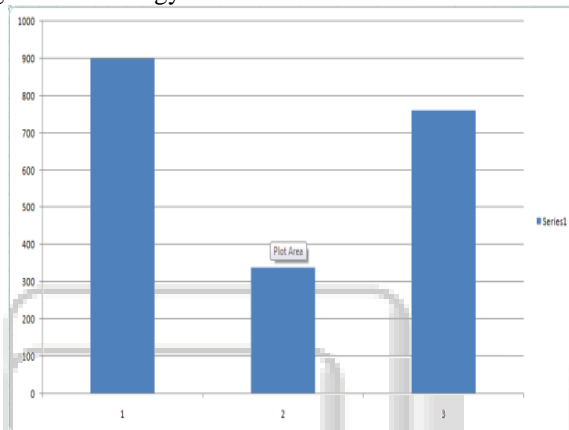


Fig. 8: Remaining energy of the clusters after the end of the simulation

VI. CONCLUSION

In this research, a new clustering algorithm for WSNs has been developed using K-Means algorithm. The inspiration for the proposed project has been derived from an earlier implemented clustering algorithm based on K-Means data clustering algorithm. The major objectives behind the development of the new algorithm were to enhance the WSN network lifetime and for the balanced centroid selection in the cluster. Balanced Centroid is the centroid in the cluster on a central location on the basis of density of the nodes irrespective to the cluster area center point. The N number of nodes are initially evaluated for the cluster formation. The distance between the nodes is calculated using Euclidean distance. The nodes are divided into K number of clusters. For each cluster K, its N number of nodes are evaluated using distance formula again to find the centroid by initializing the first stage of centroid election with random co-ordinates. Then the balanced CH (cluster head) is elected by the algorithm for a cluster K with N number of nodes. The proposed solution is based on the cluster algorithm K-means, in which centroid is selected on the basis of means technique of statistics. K-means is an algorithm uses partitioning around means which in real select the centroids around the actual center of the clusters which is given by X-center (X_c) and Y-center (Y_c) respectively. The proposed algorithm for the WSN clustering has performed well with the WSN simulations in NS2. It has formed the clusters in the actual transmission range

specifications in real time, it means the nodes out of the coverage or near to the outer circle from one point if is inside the strong transmission range of the other, then will be selected the other cluster only. It may perform in that cluster better than the earlier circle, whereas K-means generally make those mistakes. Euclidean distance is used to compute the distance between the two nodes under this simulation. The cluster on which the results were tested contained 100 no. of nodes. The number of nodes can vary in the simulation. Overall recorded results have proved that k-means have performed better than k-means clustering over WSNs. K-Means have outperformed K-Means algorithm for clustering proposed in the base paper. Hence it is also proved that the proposed algorithm has also outperformed LEACH and other protocols compared in the base paper with K-Means.

VII. FUTURE WORK

In the future, this research can lead to several other developments. The algorithm can undergo time analysis for the development of a fast WSN clustering algorithm based on the existing algorithm. The Individual nodes, Cluster Heads, Base stations, Network Governors or Wireless Sensor Network Monitoring applications used in WSNs can also use k-means algorithm for the network data clustering and analysis for the detection of several network attacks.

REFERENCES

- [1] Chen, Jiejie. "Improving Life Time of Wireless Sensor Networks by Using Fuzzy c-means Induced Clustering." In World Automation Congress (WAC), 2012, pp. 1-4. IEEE, 2012.
- [2] Park, Geon Yong, 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." In Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on, pp. 910-915. IEEE, 2013.
- [3] Mohan, Chander, and Ashok Kumar. "Heterogeneous fuzzy based clustering protocol." In Image Information Processing (ICIIP), 2013 IEEE Second International Conference on, pp. 601-606. IEEE, 2013.
- [4] Said, Jihed Eddine, Lutful Karim, Jalal Almhana, and Alagan Anpalagan. "Heterogeneous mobility and connectivity-based clustering protocol for wireless sensor networks." In Communications (ICC), 2014 IEEE International Conference on, pp. 257-262. IEEE, 2014.
- [5] Peng, Jian, Tang Liu, Hongyou Li, and Bing Guo. "Energy-Efficient Prediction Clustering Algorithm for Multilevel Heterogeneous Wireless Sensor Networks." International Journal of Distributed Sensor Networks 2013 (2013).
- [6] Tandon, Ravi, Biswanath Dey, and Sukumar Nandi. "Weight based clustering in wireless sensor networks." In Communications (NCC), 2013 National Conference on, pp. 1-5. IEEE, 2013.