International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.5., 2015
(Sept.-Oct.)

RESEARCH ARTICLE

ISSN: 2321-7758

K-MEANS BASED CLUSTERING FOR ENERGY EFFICIENT SENSOR NETWORK

ASHISH BANSAL[1], HARNEK SINGH[2]

[1]PG Student, Dept of ECE,RAYAT INSTITUTE OF ENGINEERING AND INFORMATION TECHNOLOGY; ROPAR

[2]Assistant Professor; Dept of ECE ,RAYAT INSTITUTE OF ENGINEERING AND INFORMATION Technology

ASHISH BANSAL

ABSTRACT

The wireless sensor network are the networks of things used for the data collection for various applications. Some of the popular examples of the wireless sensor networks are weather data collection, avalanche studies, crop monitoring, activity monitoring, etc. Sensor hubs keep running on battery without direct power supply. Consequently, vitality productivity turns into the real issue in WSNs to make WSNs keep running for more. Since these remote sensor hubs keep running on batteries and they convey a restricted battery life. Clustering is the main operation performed by the sensor networks to arrange the sensor networks in the smaller groups to reduce the amount of data flooding, route updates and convergence data, which directly affects the power consumption. A number of sensor networks clustering scheme has been proposed already. In this paper, the clustering scheme has been proposed for the improvement in the existing clustering model. The simulation has been performed utilizing the MATLAB environment and the proposed methodology have been contrasted and existing k-means approach as far as clustering quality and cluster head determination quality estimation. The outcomes have demonstrated the viability of the clustering method of the proposed algorithm.

Keywords—K-means clustering, WSN clustering, energy efficient WSNs, cluster head selection, geocasting clustering protocol.

1. INTRODUCTION

Today's in the market of rapid growth of computers the processing power are increased unexpectedly but the price and size of computers have greatly reduced which encourages the use of computers very much. The latest technologies have made vast advancements in computers era and also enhance the use of computers in our daily activities. In recent years, from the economic point of view, the single-purpose desktop computers having sensors embedded in them are highly used due to cheapness in prices and reduction in size of computers. 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 1.1

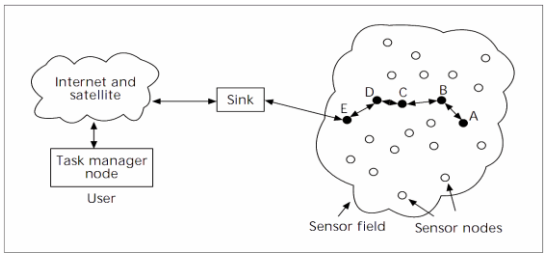

Figure 1.1: Sensor nodes scattered in a sensor

This figure 1.1 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 WSNs 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, WSNs can dynamically adapt its topology. After the deployment of sensor nodes in a sensor field, they autonomously find the neighbour 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 WSNs 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. WSNs 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 WSNs 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.

The major limitation in the WSNs are Limited Energy (Energy consumption), Network Lifetime, Application Dependency, Cluster formation and CH selection. Synchronization, Data Aggregation, etc.

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. LITERATURE REVIEW

Sonam Palden Barfunga et al. [3] proposed energy efficient routing protocol which is hierarchical and based on clustering. The author investigated the protocol in which the CH is selected by BS. There are two stages of selection procedure. First stage consisting tasks such as listing of all nodes which have to become CH. For listing nodes, this protocol evaluates the various parametric measures of nodes like relative distance between the candidate and BS. Also these parameters involve remaining energy level, how many sensor nodes (neighbours) the candidate node have, and how many times the candidate node has become the CH. In this, CH creates two schedules: SLEEP based transmit and TDMA based transmit. Seal Sarkar et al. (2012) "A Trust Based Protocol for Energy-Efficient Routing in Self-Organized MANETs" introduced an energy consumption model through which energy of all sensor nodes can be calculated. Also the author adopted a trust module based energy efficient routing protocol. Trust module is used to calculate the value of routing metric. Experiments would be conducted to compare the proposed protocol on the basis of routing overhead. This comparision shows that it increases ratio of delivery of packets and uses less energy. A number of experiments have been

conducted by the author. Result of those experiments show that the established route increases ratio of delivery of packets and reduces the delay-time. Said BEN ALL et al. (2010) "HABRP an energy efficient protocol" have proposed an energy efficient routing protocol for heterogeneous wireless sensor network. In this, the author worked on reducing the failure factor of sensor nodes. In this, the author introduced protocol for prolong the time interval before the first nodes' death so that there will be an increase in the lifetime of an heterogeneous WSNs. Also, the protocol elected some high-energy nodes(NCG) as "cluster heads" to collect the information about the cluster members and forward this data to Gateways. XU Jiu-qiang et al. (2011) "Study on WSN Topology Division and Lifetime" has proposed the algorithm for discovering and computing the connected key nodes. The additional mobile nodes are introduced to enhance the topology connectivity in WSN. A path planning algorithm is also proposed so that the lifetime prolonged and reduces the effects caused by connected key nodes. According to simulation done by author, the research has done to solve the problem involves in the division of topology caused by connected key nodes. The network topology has enhanced and increased the lifetime of network by using the technique of discovering the connected key nodes. William B. Davis et al. (2002) "Graphical Model Theory for Wireless Sensor Networks" proposed a junction tree algorithm is an information processing algorithm based on graphical model theory. The junction tree algorithm can be instantiated in a various applications useful for WSNs.

3. DESIGN AND IMPLEMENTATION

WSNs 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.

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 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 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.

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

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* $\boldsymbol{n}_i$, *i = 1,…,k, for example, to k random* $\boldsymbol{x}^t$
8. Repeat
9. For each rotation, assign it to the closest group and represented by n_i

For all $\boldsymbol{x}^t$ *in X*

$CH_i^t \leftarrow 1$ *if* $|| \boldsymbol{x}^t - \boldsymbol{n}_i || = min_j || \boldsymbol{x}^t - \boldsymbol{n}_j ||$

$CH_i^t \leftarrow 0$ *otherwise*

10. For each m_i, recalculate it based on the examples that are currently assigned to it.

For all $\boldsymbol{n}_i$, *i = 1,…,k*

$\boldsymbol{n}_i \leftarrow$ *sum over t* $(CH_i^t \boldsymbol{x}^t)$ / *sum over t* (CH_i^t)

11. Until m_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
        - 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
    - Largest signal strength, minimum transmit energy for communication

iii. Each node transmits a join-request message (REQ) using CSMA MAC Protocol
    - The message consists of node's ID and cluster head ID

iv. Each cluster head node sets up a TDMA schedule and transmits it
    - 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

4. 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.

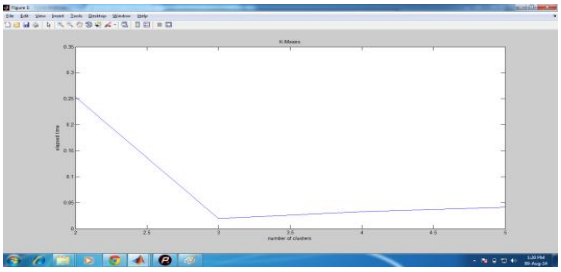

FIGURE 1: Elapsed time graph for different number of clusters

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

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.

Table 2

Number of Clusters	Davies-Bouldin Index (DB)	Calinski-Harabasz Index (CH)	Krzanowski and Lai Index (KL)	Criterian 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

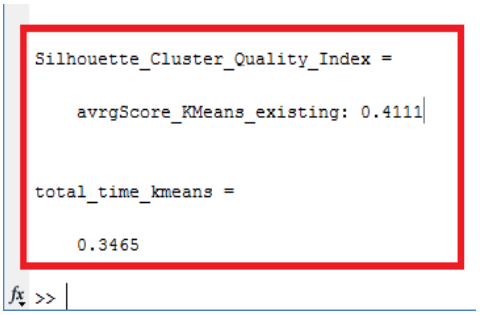

FIGURE 2: Elapsed Time and Silhouette Cluster Quality Index for Proposed clustering scheme based on K-Means Table

2: The statistical analysis of k-menas.

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

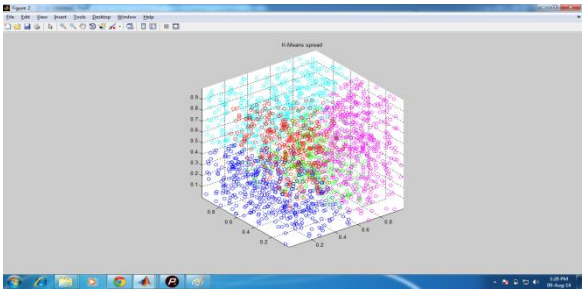

FIGURE 3; Spread cluster data of K-Means result for five clusters

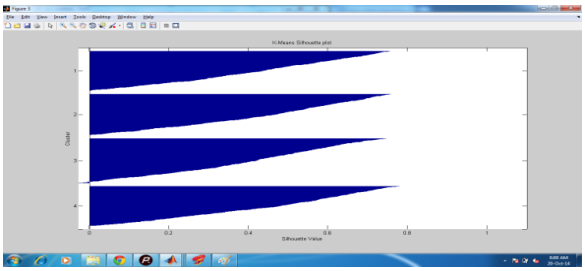

FIGURE 4: Silhouette Plot of K-Means clustering data for five clusters

MATLAB is having many build 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 & Criterian Error. Silhouette plotting has been used to plot the quality of the cluster sets that effective are the cluster created by our proposed Hybrid algorithm.

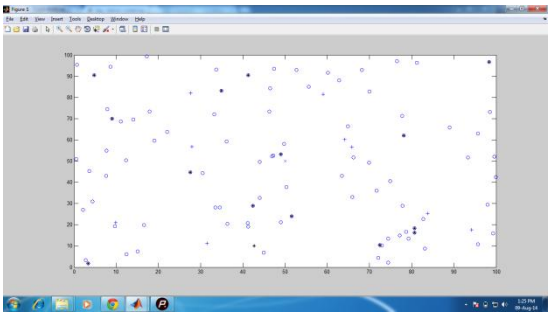

FIGURE 5: The clustering of the proposed Algorithm

5; CONCLUSION

The research project has been implemented in the MALTAB 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 & Criterian 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.

6. **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", AINAW, vol. 1, pp. 910-915, IEEE, 2013.

[2]. Sonam Palden Barfunga, Prativa Rai, Hiren Kumar Deva Sarma, "Energy Efficient Cluster Based Routing Protocol for Wireless Sensor Networks, ICCCE, vol. 1, pp. 603-607, IEEE 2012.

[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.