An Adaptive K-Means Based Method for Energy Efficiency Routing in WSN

Atul Manchanda[1] Meenu Saini[2]
[1,2]Student
[1,2]Department of Electronics & Communication Engineering
[1,2]Global Research Institute of Management and Technology, Radaur, Haryana, India

Abstract— The proposed paper is about to use some cluster algorithm such as K-Means for finding an optimal clustering scheme instead of using some random method, thus using less energy and more rounds of transmission to Base Station. For this, it will combine few parameters such as Distance as basis parameter for clustering. The proposed system is supposed to increase the overall network life time of WSN.

Key words: WSN, cluster algorithm, LEACH Protocol

I. Introduction

WSNs are battery driven and there is usually no option for recharging. Therefore an energy efficient form of communication is essential. In multi-hop networks, each node performs both as a transmitter and a receiver. As such, power failure in nodes causes significant impact on the network and force routing changes that affect the topology of the network. One of the most important design issues in WSN is scalability. The number of sensor nodes in a field may extend hundreds or thousands. As such any routing protocol should be able to work with such huge network. Concerning the routing protocols, the reduced energy resources, the scalability and the resilience arise as the main limitations in wireless sensor networks, therefore classifies the routing protocol is an important stage in the design of efficient WSN. This paper presents a unique and adaptive K-means based algorithm in which we need not to pre-decide about the number of clusters. The adaptive k-means based LEACH algorithm gives better performance.

II. Related works

Network routing protocols are in charge of routing scheme as well as maintaining the network structure in WSNs. There are three types of network structure: flat routing [1], hierarchical routing [3, 4] and location-based routing [2, 11, 12]. However, in order to focus in our area of research, we present further discussion of only flat and hierarchical routing protocols.

In flat routing protocols nodes play the same role and have similar functionality in transmitting and receiving data. In this type of network it is not possible to assign a global identifier to each node due to large number of nodes. Therefore, base station send queries to different part of the field and waits for the data from sensors in selected parts of the field. This approach is called data centric routing [24]. SPIN (Sensor Protocols for Information via Negotiation) [5] and DD (Direct Diffusion) [8] are two examples of the data centric routing protocols that save energy by data negotiation and omitting the redundant data.

The SPIN protocol aims at disseminating information among all the sensor nodes by using information descriptors for negotiation prior to transmission of the data. These information descriptors are called meta-data and are used to eliminate the transmission of redundant data in the network. In SPIN, each sensor node also has its own resource manager that keeps track of the amount of energy that the particular node has. Prior to transmission or processing data, the nodes poll their resource manager if they have enough energy or not. This allows the nodes to cut back on activities when their resources are low increasing the life of the node in the process.

Directed Diffusion is another data dissemination protocol in which attributes value pairs name the data generated by the nodes. This is a destination-initiated reactive routing technique in which routes are established when requested. A sensing task or interest is propagated throughout the network for named data by a node and data, which matches this interest, is then sent towards this node. One important feature of the data diffusion paradigm is that the propagation of data and its aggregation at intermediate nodes on the way to the request-originating node are determined by the messages, which are exchanged between neighboring nodes within some distance (localized interactions). Tasks are described by a list of attribute-value pairs that describe the task. This description is called an interest. The data, which is sent as a response to such an interest, is also named in a similar manner. The querying node is the sink node and it broadcasts its interest message periodically to all of its neighbors. All nodes have an interest cache in which each item corresponds to a different interest.

On the other hand hierarchical routing is mainly considered as two layer architecture where one layer is engaged in cluster head selection and the other layer is responsible for routing. Cluster head in hierarchical routing is the node which is responsible for collecting data from other nodes in the cluster, aggregating all data and sending the aggregated data to the base station. Creating clusters and assigning communication task to cluster heads contributes to a more scalable and energy efficient network [9]. The main goal of all the hierarchical routing protocols is to appropriately create clusters and choose cluster heads in order to reserve energy in the network. The hierarchical routing is a feasible solution for reducing energy consumption in WSNs. Within a cluster, cluster head manages the member nodes and assigns them tasks which lead to reduction in redundant data transmission. Moreover, cluster head has some responsibilities such as data collection and data aggregation from their respective cluster members. Energy consumption greatly reduced in this kind of routing method since the total data messages sent to the base station is minimized by data aggregation.

The WSN clustering protocols can be classified into two categories: probabilistic and deterministic. In probabilistic clustering protocols a node becomes a CH with a certain probability, which requires an exchange of overhead messages for the CH's election. The EEHC [14], EECS [15], and HEED [16] fall in the probabilistic class and PEGASIS [17], and TASC [18] are categorized in the deterministic class. The surveys dealing with WSN clustering protocols can be found in [13].

III. THE PROBLEM WITH EXISTING CLUSTERING ALGORITHM AND PROPOSED SOLUTION

In clustering analysis, a fundamental problem is to determine the best estimate of the number of clusters, which has a deterministic effect on the clustering results. However, a limitation in current applications is that no convincingly acceptable solution to the best-number-of-clusters problem is available due to high complexity of real data sets. Choosing an appropriate clustering method is another critical step in clustering. A large number of clustering methods are available for cluster analysis. However a fundamental problem in applying most of the existing clustering approaches is that the number of clusters needs to be pre-specified before the clustering is conducted. The clustering results may heavily depend on the number of clusters specified. It is necessary to provide educated guidance for determining the number of clusters in order to achieve appropriate clustering results. At the current stage of research, none of the existing methods of choosing the optimal estimate of the number of clusters is completely satisfactory. The gap method was recently proposed by Tibshirani, et al. [20] The main idea of the gap method is to compare the within cluster dispersions in the observed data to the expected within cluster dispersions assuming that the data came from an appropriate null reference distribution. Simulation results reported by Tibshirani, et al.[20] indicated that the gap method is a potentially powerful approach in estimating the number of clusters for a data set. Our Thesis objective will be to enhance the existing K-Means Clustering Algorithm for Color Image Segmentation. we need to find the method to find the optimal number of cluster for K-Means clustering then validate the validity of the clusters.

In [19] k-means based clustering technique is used and then CH is selected on the basis of residual energy, spatial position of nodes in the cluster. K-means clustering is a well-known separating method. In this objects are categorized as belonging to one of K groups. The outcomes of partitioning method are a set of K each object (nodes in our case) of data set belonging to single cluster. In every cluster there may be a centroid or a cluster descriptive. In the case where we reflect real-valued data, the arithmetic mean of the attribute vectors for all objects within a cluster provides an appropriate representative. K-means is a data mining algorithm which performs clustering of the data samples. The clustering is the division of a dataset into a number of groups such that similar items falls or belong to same groups. In order to cluster the file, K-means algorithm used as iterative approach. The algorithm requirement is to produce K clusters then there will be K initial means and final means after termination of clustering algorithms, in every object of dataset develops a member of one cluster. As explained already, in [46] after clustering CH is chosen based on residual energy, its distance from BS and its position in cluster. In this paper, clusters are formed first by specifying manually no. of cluster. No of cluster can't be a fixed number as it should depend on the placement of nodes. If nodes are in proximity, there is no point of creating new clusters. So we will modify above algorithm in terms of specification of clusters. We have added a module before creating number of clusters using k-means. Now it will look like this:

Step 1: Find the nodes that are alive. Nodes that are not of type dead nodes are alive.

Step 2: In alive nodes using some standard clustering algorithm such as k- means or k-medoid for spatial distribution of nodes in clusters we will create a large number of clusters.

Step 3: Merge clusters in such a way that it satisfies similarity and dissimilarity criteria between each of the existing cluster pairs in node placements.

Step 4: In the next step, from each cluster CH is chosen on the basis of surplus energy only a node is having. So our proposed scheme involves both things spatial distribution as well as energy distribution in the network architecture which may ultimately improve the network life and its quality.

IV. THE PROPOSED METHODOLOGY

To simulate LEACH, we have used random 100-node networks for our simulations with similar parameters used in [7]. We placed the BS at a far distance from all other nodes. For a 100m x 100m plot, our BS is located at (50, 200) so that the BS is at least 100m from the closest sensor node. We have used the same energy model as discussed in [19] which is the first order radio model. In this model, a radio dissipates Eelec = 50 nJ/bit to run the transmitter or receiver circuitry and Eamp = 100 pJ/bit/m2 for the transmitter amplifier. In our simulations, we used a control packet length k of 200 bits to send information from non-CH node to CH node. Size of packet length K of 6400 bits is fixed to send information from CH node to BS.

A. LEACH Protocol:

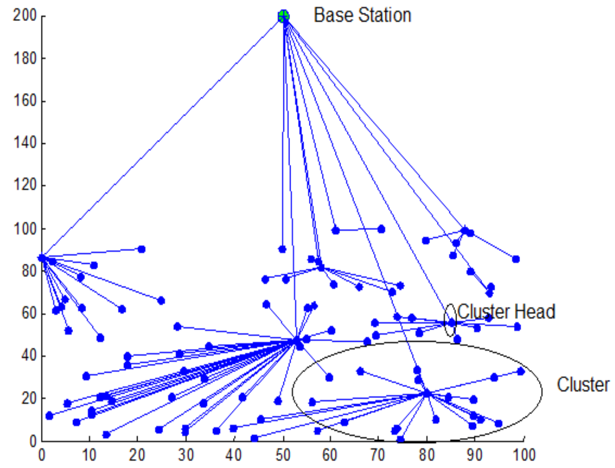


Fig. 1: Leach routing topology

LEACH (Low Energy Adaptive Clustering Hierarchy) is first proposed by Wendi B. Heinzelman of MIT. LEACH is a clustering-based protocol that utilizes randomized rotation of local cluster base station (CH) to evenly distribute the energy load among the sensors in the network [5]. LEACH uses localized coordination to enable scalability and robustness for dynamic networks, and incorporates data fusion into the routing protocol to reduce the amount of information that must be transmitted to base station. LEACH rearranges the network's clustering dynamically and periodically, making it difficult for us to rely on long lasting node-to-node trust relationships to make the protocol secure. LEACH assumes every node can directly reach a base station by transmitting with sufficiently high power.

This protocol provides a concept of round. LEACH protocol runs with many rounds. Each round contains two phases: Cluster setup phase and steady phase in which clusters are formed and then energy is dissipated respectively.

B. K-means based LEACH algorithm:

In this algorithm idea is to select cluster in such a way that their intra distance is minimum which ensures that less communication energy is consumed and WSN can run more rounds. K-means [10] is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as centers of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the k centroids change their location step by step until no more changes are done. In other words centroids do not move any more. Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centers.

The algorithm is composed of the following steps:

(1)	Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
(2)	Assign each object to the group that has the closest centroid.
(3)	When all objects have been assigned, recalculate the positions of the K centroids.
(4)	Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Fig. 4.2: K-means Algorithm

Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect. K-means is a simple algorithm that has been adapted to many problem domains. As we are going to see, it is a good candidate for extension to work with fuzzy feature vectors.

C. Adaptive K-Means LEACH Algorithm:

In this algorithm number of clusters is predetermined by using the gap method first, which is applied on the node deployment to know the optimum number of clusters. As we know that in K-means algorithm the number of clusters, K must be supplied as a parameter hence validity measure which is based on the intra-cluster distance measures which allows the number of clusters to be determined automatically. The basic procedure involves producing all the segmented images for 2 clusters up to Kmax clusters, where Kmax represents an upper limit on the number of clusters. Then our validity measure will be calculated to determine which is the best clustering by finding the minimum value for our measure. Validity measure is simply the ratio of intra cluster distance measure to the inter cluster distance, intra cluster distance is defined as the distance between a point and its cluster centre within a cluster whereas inter cluster distance is the distance between the two clusters. Although it can be proved that the procedure will always terminate, the k-means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centers. The k-means algorithm can be run multiple times to reduce this effect.

V. RESULTS ANALYSIS

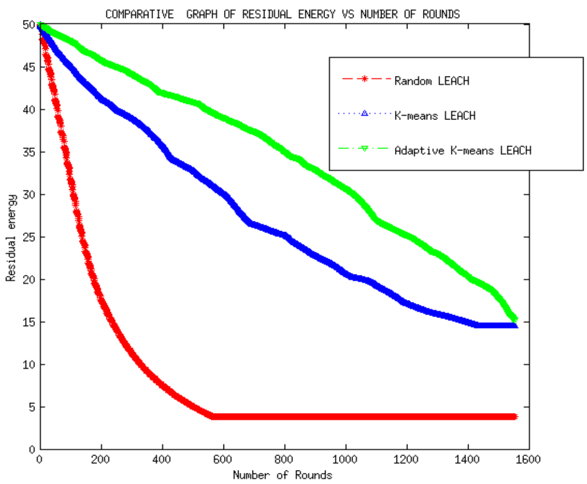
The following table no. 1 shows the results obtained from the experimentations done as per the setup explained in the previous section. Six algorithms have been implemented in this thesis. In first algorithm i.e. Random LEACH algorithm is implemented where CHs are selected randomly based on a probability function. We have taken this probability as 10%. It is further improved by using a fair distribution of energy by selecting maximum energy nodes to be CHs. In this method a fix number of CHs are selected based on the number of nodes that are living. Another modification is made in second algorithm where nodes are clustered based on inter distance by using a standard algorithm such as K-means. The third algorithm is based on k-means only. We measure algorithms' efficiency by assessing total no. of rounds up to which network survives. A network is assumed to be live if more than 25% nodes are alive with total energy greater than zero.

In the table 1 it is clearly shown that Adaptive K-means clustering based LEACH algorithms perform far better as compared to other methods if we consider the number of rounds covered by the algorithms. The k-means based algorithm performs nearly two and a half times better than random LEACH. If we consider a network, dead if 50% nodes are dead then LEACH is performing better than K-means based LEACH. If we consider 75% node criterion for network life then the Adaptive K-means LEACH algorithm performs better.

WSN Routing Algorithm	**Network Life (in rounds)**	**Rounds in which first Node Dead**	**Rounds in which 50% Node Dead**	**No of packets sent in total rounds**	**Remaining Energy after 25% node is dead (Joules)**
Random LEACH	568	110	298	10079	3.79
K-means LEACH	1439	26	694	8083	14.4745
Adaptive K-means LEACH	1554	113	1193	7978	15.9897

Table 5.1: Experimentation Results

If we closely look the figures 2 and table 1 then we can easily say that the K-means based algorithms are much better than random LEACH. In case of adaptive K-means nearly 250% network life improvement is recorded for over simple LEACH and 10% over K-means. If we compare the no of dead nodes as per our simulation results adaptive K-means LEACH seems to perform better than random LEACH, but there nodes once start dying accelerates network decay very fast. On one front random LEACH and Kmeans-LEACH algorithms are lacking i.e. network disintegration in this front. In these algorithms, first node is dead very. This is the grey area which needs to be addressed in future research. If we consider no of packets sent to BS then LEACH is clearly winner. It has sent highest no of packets to BS. This may be due to uneven size of the clusters in which lot of energy is wasted for sending the same amount and same vicinity data. We have shown these statistics in the following figures.



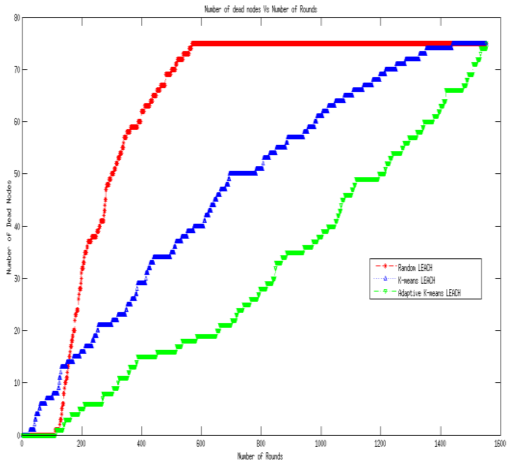


Fig. 2: Nodes Remaining Energy pattern in WSN and No of dead nodes per round

VI. CONCLUSION

The network life, no of dead nodes and no of packet sent to BS affect performance of routing algorithm in WSN. The performance of cluster based routing protocol shows some differences by varying life pattern among nodes and number of dead nodes. From our experimental analysis we conclude that adaptive k-means based LEACH algorithm gives better performance in network life overall but could not restrict early network disintegration. We have improved the network life but one thing; we have observed that node starts dying early which is an area of concern in PSO LEACH. This can be addressed by considering other parameters of nodes' characteristics such as remaining node energy in addition to distance between them while clustering them. This technique may delay early node death problem.

REFERENCES

[1] D. Braginsky and D. Estrin. Rumor routing algorthim for sensor networks. In Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications. ACM, pp. 22-31, 2002.
[2] N. Bulusu, D. Estrin, L. Girod and J. Heidemann. Scalable coordination for wireless sensor networks: self-configuring localization systems, In processing of sixth International Symposium on Communication Theory and Applications, 2001.
[3] S. Capkun, M. Hamdi, J. Hubaux, GPS-free positioning in mobile ad-hoc networks. In processing of the 34th Hawaii International Conference on System Sciences, pp. 3481–3490, 2001.
[4] Q. Fang, F. Zhao, and L. Guibas. Lightweight sensing and communication protocols for target enumeration and aggregation. In Proceedings of the 4th ACM international symposium on Mobile ad hoc networking, pp. 165-176, 2003.
[5] W. R. Heinzelman, J. Kulik, and H. Balakrishnan. Adaptive protocols for information dissemination in wireless sensor networks. In Proceedings of the 5th annual ACM/IEEE international conference on

Mobile computing and networking, pp. 174-185, 1999.

[6] W. R. Heinzelman, A. Chandrakasan and H. Balakrishnan. Energy-efficient communication protocol for wireless microsensor networks. Proceedings of the 33rd Annual Hawaii International Conference on system Sciences, vol.2, pp. 3005-3014, 2000.

[7] W. B. Heinzelman, A. P. Chandrakasan and H. Balakrishnan. An application-specific protocol architecture for wireless microsensor networks. IEEE Transactions on Wireless Communications, vol. 1, no. 4, pp. 660- 670, 2002.

[8] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann and F. Silva. Directed diffusion for wireless sensor networking. IEEE/ACM Transactions on Networking,, vol.11, no.1, pp. 2-16, 2003.

[9] J. N. Al-Karaki and A. E. Kamal. Routing techniques in wireless sensor networks: a survey," Wireless Communications, IEEE , vol.11, no.6, pp. 6- 28, 2004.

[10] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297

[11] Perrig, R. Szewczyk, J. D. Tygar, V. Wen, and D. E. Culler. 2002. SPINS: security protocols for sensor networks. Wireless Networks, vol. 8, no. 5, pp.521-534, 2002.

[12] K. Sohrabi, J. Gao, V. Ailawadhi and G. J. Pottie. Protocols for self-organization of a wireless sensor network. IEEE Personal Communications, vol. 7, no. 5, pp.16-27, 2000.

[13] O. Younis, M.Krunz, and S.Ramasubramanian. Node clustering in wireless sensor networks: recent developments and deployment challenges. IEEE Network , vol.20, no.3, pp. 20- 25, 2006.

[14] Bandyopadhyay, S. and Coyle, E.J., “An energy efficient hierarchical clustering algorithm for wireless sensor networks,” INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies, vol. 3 pp 1713–1723, 2003.

[15] M. Ye, C. Li, G. Chen, and J. Wu, “EECS: An energy Efficient Clustering Scheme in wireless Seneor networks,” National Laboratory of Novel Software Technology, Nanajing University, China, 2005.

[16] O. Younis, and S. Fahmy, “HEED: A hybrid, energy-efficient, distributed clustering approach for ad-hoc sensor networks,” IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366–379, October-December 2004.

[17] Lindsey, S. and Raghavendra, C.S.,“PEGASIS: Power-efficient gathering in sensor information systems,” IEEE Aerospace Conference Proceedings, vol. 3, pp. 3–1125, 2002

[18] Virrankoski, R. and Savvidees, A., “TASC: topology adaptive spatial clustering for sensor networks,” IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, pp. 614, 2005

[19] Jia Yunjie, Liu Ming, Zhu Song, Dong Pengtao, A clustering routing algorithm based on energy and distance in WSN, International Conference on Computer Distributed Control and Intelligent Environmental Monitoring, 2012, pp. 9-12

[20] Robert Tibshirani, Guenther Walther and Trevor Hastie, “Estimating the Number ofData Clusters via the Gap Statistic” J.R. Statist. Soc. B (2001), 63, pp. 411—423

