

A Modified clustering for LEACH algorithm in WSN

B.Brahma Reddy
ECE, VBIT, Hyderabad,
India

K.Kishan Rao
Vaagdevi College of Engineering,
Warangal, India

Abstract—Node clustering and data aggregation are popular techniques to reduce energy consumption in large Wireless Sensor Networks (WSN). Cluster based routing is always a hot research area in wireless sensor networks. Classical LEACH protocol has many advantages in energy efficiency, data aggregation and so on. However, determining number of clusters present in a network is an important problem. Conventional clustering techniques generally assume this parameter to be user supplied. There exist very few techniques that can solve the problem of automatic detection of number of clusters satisfactorily. Some of these techniques rely on user supplied information, while others use cluster validity indices. In this paper, we proposed a rather simple method to identify the number of clusters that can give satisfactory results. Proposed method is compared with classical LEACH protocol and found to be giving better results.

Keywords—Clustering Index; LEACH; Wireless Sensor Networks; Energy optimization; Network lifetime

I. INTRODUCTION

Node clustering and data aggregation are popular to reduce energy consumption in large Wireless Sensor Networks (WSN). Clustering in WSN is the process of dividing the nodes of WSN into groups, where each group agrees on a central node, called the Cluster Head (CH), which is responsible for gathering the sensory data of all group members, aggregating it and sending to Base Station (BS). Cluster based routing is always a hot research area in wireless sensor networks. Classical LEACH protocol has many advantages in energy efficiency, data aggregation and so on. However, determining number of clusters present in a network is an important problem. Conventional clustering techniques generally assume this parameter to be user supplied. There exist very few techniques that can solve the problem of automatic detection of number of clusters satisfactorily. Some of these techniques rely on user supplied information, while others use cluster validity indices which need additional computation time. There are several indexes such as Dunn's, PBM, Davis-Bouldin, Global and Mahalanobis distances, SVD entropy, Krzanowski and Lai, Hartigan, Silhouett, Gap Statistic proposed by earlier authors for cluster validity. We define the optimal clustering as the one which gives data transmission from the cluster members to CH and subsequently from CH to BS incurs the minimal energy or maximize total transmissions. In this paper, we proposed a simple method to identify the number of clusters that can give increased number of transmissions.

Rest of the paper is organized as follows: Section II explains a few cluster validity indexes; Section III highlights related work done by other authors; Section IV describes the

network model used; Section V presents simulation results and analyses; finally Section VI concludes observations.

II. CLUSTER VALIDITY INDEXES

Assignment to clusters relies on a distance measure; in the case of genetic algorithms, the criterion function of the optimization is called the fitness function. Here we present some measures which we tested with our algorithm. As a general guideline, these measures should favor for minimal differences between points within the cluster (intra-cluster, DCH) and maximal differences between points of different clusters (Inter-cluster, DBS).

In the original LEACH protocol, the probability corresponds to the number of desired CHs in the network. Additional metrics such as remaining node energy can also be used to change the clustering properties. LEACH divides the whole network into several clusters, and the run time of network is broken into many rounds. In each round, the nodes in a cluster contend to be cluster head according to a predefined criterion. In LEACH protocol, all the sensor nodes have the same probability to be a cluster head, which makes the nodes in the network consume energy in a relatively balanced way so as to prolong network lifetime. However, number of clusters may vary in each round. Because of this reason, network lifetime can not be defined in terms of rounds. Better definition for network lifetime is in terms of number of transmissions. Each sensor node n decides independently of other sensor nodes whether it will claim to be a CH or not, by picking a random s between 0 and 1 and comparing s with a threshold function value $T(n)$ based on a user-specified probability p . If $s \leq T(n)$ then the node claims to become CH. The threshold is defined as follows [1]:

$$T(n) = \begin{cases} \frac{p}{1-p(r \bmod \frac{1}{p})} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases}$$

Where G is the set of nodes that have not been CHs in the last $1/p$ rounds.

In the proposed algorithm, number of clusters is fixed and predetermined using cluster validation criteria. And each round one of the cluster members become cluster head. Thus, there is one packet transmission per round from each CH. To understand the components of the energy consumed, we separate the cost into intra and inter cluster energy consumption. Intuitively, as cluster size grows, energy consumption inside the cluster also grows. At the same time energy consumption from cluster heads to the base station drops significantly, since fewer CHs are present. For finding

optimal cluster combination we validate certain characteristics given below:

$$\text{Average Cluster distance } d_{CH} = \frac{\sum_{i=1}^n d(i,CH)}{n}$$

$$\text{Average BS distance } d_{BS} = \frac{\sum_{i=1}^k d(i,BS)}{k}$$

Conditions

1. Distance to CH must be less than distance to BS

$$\frac{d_{CH}}{d_{BS}} \leq 1 \quad (1)$$

2. Energy consumed must be less for route via CH compared to direct to BS

$$d_{CH}^2 + d_{BS}^2 \leq d_{DirBS}^2 \quad (2)$$

3. None of cluster ‘i’ must be closer to CH of cluster ‘j’
 $d_{BSi} - d_{BSj} \geq (d_{CHi} | d_{CHj})$ (3)

4. Energy consumed by CH for a transmission
 $\left(\frac{n}{k} - 1\right) \cdot E_{RX} + \left(\frac{n}{k} - 1\right) \cdot E_{DA} + E_{elec} +$

$$\varepsilon_{fs} \cdot d_{BS}^2 = E_{BS}$$

5. Energy consumed by a cluster member to forward data to CH

$$E_{elec} + \varepsilon_{fs} \cdot d_{CH}^2 = E_{CH}$$

Where n in number cluster members nodes and k is number of clusters. E_{RX} , E_{elec} , E_{DA} , ε_{fs} are Receive energy, Transmit energy, Data aggregation energy and free space loss respectively.

$$\sigma_{BSi} = \sqrt{E[BS_i^2] - E^2[BS_j]} \text{ where } j \neq i \text{ (Cluster } i) (i = 1 \dots k) \text{ ---Standard deviation in remaining energy of different Clusters. It must be close to zero. Global best (4)}$$

$$\sigma_{CHi} = \sqrt{E[CH_i^2] - E^2[CH_j]} \text{ where } j \neq i \text{ (Cluster } i) (i = 1 \dots \text{cluster-size) ----- Standard deviation in remaining energy of nodes in a cluster. It must be close to zero. Local best (5)}$$

Equations 1, 2 and 3 ensures clustering saves energy. Equations 4 and 5 give Global best and local best position of CHs.

Automatic determination of number of clusters present in a wireless sensor network has been a challenging problem to the researchers. There are two aspects of this problem: i) finding number of clusters and ii) finding the clusters themselves. Majority of the existing techniques assume the number of clusters as an input parameter to be supplied by the user. One of the most common techniques is k-means algorithm [1]. The k-means algorithm is a simple partitioned clustering algorithm. The objective of this algorithm is to partition the given data set S containing N data elements $\{x_1, x_2, \dots, x_N\}$ into k clusters, $\{C_1, C_2, \dots, C_N\}$ such that

$$C_i = \emptyset \text{ for all } i$$

$$C_i \cap C_j = \emptyset \text{ if } i \neq j$$

$$C_1 \cup C_2 \cup \dots \cup C_k = S$$

However, the unanswered question is which partition represents the best clustering solution. This question may be answered if we perform some test for the tendency of clustering of the concerned data set before clustering it.

In this paper, we used the cluster area as input to determine the number of clusters. Group of nodes within radius of clustering area are formed as a separate cluster. If any node is falling in more than one group, then the node will be retained in the cluster where the node is closer. Thus a node is ensured to present only in one cluster. In order to ascertain the characteristics of the clusters, we used Dunn’s Index, PBM Index and Davis-Bouldin Index.

- A. **Dunn’s index** [2]. For any k -partition Dunn defined the following index:

$$V_D = \min_{1 \leq i \leq K} \left\{ \min_{1 \leq j \leq K, j \neq i} \left\{ \frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq K} \Delta(C_k)} \right\} \right\}$$

Where

$$\Delta(C_k) = \max_{x,y \in C_k} \{d(x,y)\}$$

$$\delta(C_i, C_j) = \min_{x \in C_i, y \in C_j} \{d(x,y)\}$$

Larger values of V_D corresponds to good clusters and the number of clusters that maximize V_D is taken as the optimal number of clusters. $d(x,y)$ is the distance between node x and node y .

- B. **PBM Index**: It is defined as

$$PBM(K) = \left(\frac{1}{K} \frac{E_1}{E_K} D_K \right)^2$$

Where K is the number of clusters, E_K and D_K represent sum of within cluster dispersions and maximum between cluster separations respectively.

- C. **Davis-Bouldin (DB) Index**: This index is a function of the ration of sum of within-cluster scatter to between-cluster separation.

$$DB = \frac{1}{K} \sum_{i=1}^K R_{i,qt}$$

$$R_{i,qt} = \max_{j,j \neq i} \left\{ \frac{S_{i,q} + S_{j,q}}{d_{ij,t}} \right\}$$

Where $S_{i,q}$ is the q th moment of the points in cluster C_i with respect to their mean? And it is a measure of the dispersion of the points in that cluster. $d_{ij,t}$ is the Minkowski distance of order t between the centroids that characterize clusters C_i, C_j . The objective is to minimize the DB index for achieving proper clustering.

III. RELATED WORK

This paper is not the first to analytically evaluate clustering techniques. In a very recent [3], the author addresses problem in terms of number of sensors in an optimal cluster. Jamshid Shanbehzadeh, Saeed Mehrjoo, Abdolhossein Sarrafzadeh [4] proposed a hybrid GA-PSO based clustering algorithm that improved the lifetime of WSN effectively. In [5] O.A.Mohamed Jafar and R.Sivakumar carried a survey on ant-based clustering algorithms. Akramul Azim, Mohammad Mahfuzul Islam [6] introduced relay nodes to act as cluster heads and decrease the probability of premature death of original nodes. Jin Wang, Xiaoqin Yang, Yuhui Zheng, Jianwei Zhang, Jeong-Uk Kim [7] proposed energy-based clustering algorithm and demonstrated improved energy efficiency. GRASP based algorithm was proposed for cluster formation problem by Victor de Oliveira, Matos, Jose Elias C.Arroyo, Andre Gustavo dos Santos, Luciana B.Goncalves [8]. In [9] S.Rao Rayaoudi proposed a novel approach based on intelligent water drops algorithm to solve economic load dispatch problem. Liu Ban-teng, Chen You-rong, Zhou Kai, Jingyu Hua [10] proposed Boolean sensing model based on Poisson point process to identify the function of the rate of coverage and the node density in unit area, and then calculates the total number of nodes in the region, next uses the greedy strategy of the Prim algorithm to find a spanning tree with the maximum weight, and constructs a approximate solution for the minimum connected dominating set. Malay K Pakhira [11] proposed Visuval Assessment of Tendency based algorithm for automatic determination of number of clusters identification. Anna Forster, Alexander Forster, Amy L. Murphy [12] presented an experimental analysis for optimal cluster sizes. Benjamin Auffarth [13] presented a genetic algorithm that is fast and able to converge on meaningful clusters for real-world data sets and discussed cluster validity criteria. Jianguo SHAN, Lei DONG, Xiaozhong LIAO, Liwei SHAO, Zhigang GAO, Yang GAO [14] presented another improved version of LEACH protocol to extend life cycle of the network.

IV. SIMULATION MODEL

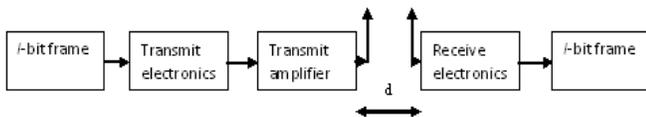


Fig.1. Radio energy model

We define WSN to be a two-dimensional network graph $G = (V, E)$ where V and E are set of Vertices (Wireless sensor nodes) and Edges (Transmission links) respectively. There is a link from sensor u to sensor v if and only if v is located in u 's transmission range, then v is called neighbor of u . All nodes are equipped with adjustable transmit power. Initially, transmit power level is fixed to transmission range R . However, when acting as CH, transmit power can be adjusted to communicate with BS. We assume clusters are circles with a transmission range of R . Nodes can communicate to all their neighbors, defined as those nodes whose distance is less than R . Energy is spent when a node sends as well as receives a packet. Energy

model of the network is shown in Fig1. Two-tier model is considered in this paper. Cluster members communicate with CH and CH aggregates the data and communicates with BS. We compute clusters and its members as given in Algorithm below.

Algorithm to computer clusters

```
Input :  $G = (V, E), R$   
 $C = (C_1, C_2, \dots, C_k) = \emptyset$   
repeat till all  $V_i$  are covered  
  for  $i = 1$  to  $k$   
    Select  $V_i$  next highest node – degree  
     $C_i = C_i \cup V_i \cup$  all one – hop neighbors of  $V_i$   
  endfor  
endrepeat  
repeat for all  $C_i, i = 1..k$   
  for all members of  $C_i$   
    if  $d(\text{member}, C_i)$  is minimum retain  
    else remove from  $C_i$   
  endrepeat  
endrepeat  
Output:  $C$ , cluster groups and members in each group
```

V. SIMULATION RESULTS

In this section, we evaluate the performance of our proposed algorithm through the simulations with respect to classical LEACH algorithm. We compare our proposed algorithm with LEACH based on three performance metrics: i) spread of the dead/alive nodes at each round; ii) Number of packets transmitted at FND(First-node dead), HND (Half nodes dead), AND (All Nodes Dead); iii) Cluster indices of Dunn's, PBM and DB. The reference network of our simulations consists of 100 nodes distributed randomly in an area of $100 \text{ m} \times 100 \text{ m}$. The BS is located at position (50, 50). Here we use the typical values $E_{elec} = 50 \text{ nJ/bit}$, $\epsilon_{fs} = 10 \text{ pJ/bit/m}^2$ and $\epsilon_{mp} = 0.0013 \text{ pJ/bit/m}^4$. As noted previously, the cluster heads are responsible for aggregating their cluster members' data. The energy for data aggregation is set as $E_{DA} = 5 \text{ nJ/bit/signal}$. The initial energy of all nodes set to 0.1 J. Every node transmits a 4000-bit message per round to its cluster head. Here we assume every node having knowledge of other node position and compute distance from BS. We normalize the distance of each node to longest node distance from BS. We computed Cluster indices for R value ranging from 0.1 to 0.9 and found to be optimum for 0.3. With this we assumed $R=0.3$ for comparison of the proposed algorithm with LEACH. p is set to 0.1 (about 10% of nodes per round become cluster heads) for LEACH. For the proposed algorithm, one of the live cluster members with highest remaining energy becomes CH in each round. Thus there is one packets transmission for every round from each cluster in the proposed algorithm. In LEACH, there can be a round without CH and thereby no packet transmission. Hence, we use number packet transmissions to BS instead of number of rounds as a measure of network lifetime. Table1 gives the comparison of Indexes values for both LEACH and proposed algorithm. As mentioned earlier in this paper, numbers of cluster heads vary

in each round for LEACH. For comparison purpose we have taken number of cluster heads of first round in LEACH protocol. But in case of proposed algorithm, number of cluster heads is same in each round. So, all Indexes computed for LEACH are related to first round only. Fig 2 gives the comparison of Packets transmitted to BS at various stages. Fig 3 gives the comparison of live nodes with respect to rounds. Here, it may be observed that though the number of rounds increases in case of LEACH, total number packets transmitted do not increase. This happens because in some rounds there may not be any CH. But in case of proposed algorithm, there will be packet transmission in every round from every cluster unless all nodes of a cluster die.

TABLE I. Comparison of Cluster Validity Indexes

Index	LEACH	Proposed algorithm
Dunn's	0.1148	0.4253
PBM	0.00007	0.0006
DBI	1.8644	2.3431
Number of clusters	9	12

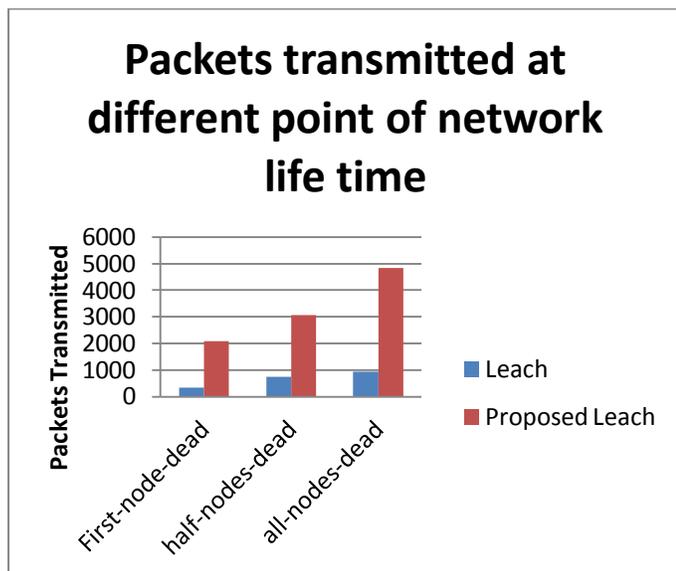


Fig.2. Number of packets transmitted

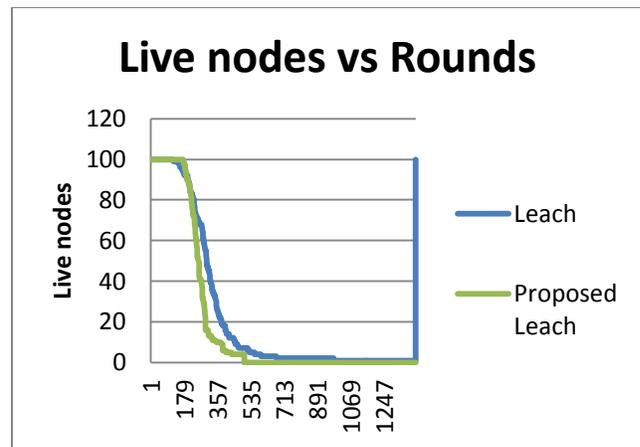


Fig.3. Number of live nodes vs Rounds

VI. CONCLUSION

In this paper, we considered a well known energy efficient clustering algorithm for WSNs called LEACH algorithm and proposed a new clustering algorithm for improved network lifetime. In this new approach, we used fixed number of clusters and cluster number is computed based on validity tests. This makes the optimum number of clusters than in LEACH algorithm and hence increases network lifetime. The simulation results show that the proposed algorithm can make the remaining energy more uniform throughout the network at every round. In the present work we assumed E_{DA} , E_{elec} to be constant. However, they can also influence the energy consumption per round. In future work, we can improve our algorithm for including variation due to E_{DA} , E_{elec} condition.

References

- [1] J.B. Ball and D Hall, "Some methods of classification and analysis in multivariate observations", Proc. of fifth Berkeley symposium on mathematical and probability, pp-281-297, 1967
- [2] T.Havens, J.C. Bezdek, J.M.Keller, and M.Popescu, "Dunn's cluster validity index as a contrast measure of VAT images", Proc. ICPR, Tampa, Florida, 2008
- [3] Wang,D, "An energy-efficient cluster head assignment scheme for hierarchical wireless sensor networks", International journal of Wireless Information Networks, 15(2), 2008, pp-61-71
- [4] Jamshid Shanbehzadeh, Saeed Mehrjoo, Abdolhossein Sarrafzadeh, "An Intelligent energy efficient clustering in wireless sensor networks", Proc. of International multi conference of engineers and computer scientists, Vol-1, March 16-18, 2011, Hongkong.
- [5] O.A.Mohamed Jafar and R.Sivakumar, "Ant-Based clustering algorithms: A brief survey", International journal of computer theory and engineering, Vol-2, No-5, October, 2010, pp-787-796
- [6] Akramul Azim, Mohammad Mahfuzul Islam, " A relay node based hybrid low energy adaptive clustering hierarchy for wireless sensor networks", International journal of energy, information and communications, vol-3, Issue-3, August, 2012, pp41-53
- [7] Jin Wang, Xiaoqin Yang, Yuhui Zheng, Jianwei Zhang, Jeong-Uk Kim, "An Energy-based Clustering Algorithm for wireless sensor networks", International Journal of Future Generation Communication and Networking Vol. 5, No. 4, December, 2012, pp-89-98

- [8] Victor de Oliveira, Matos, Jose Elias C. Arroyo, Andre Gustavo dos Santos, Luciana B. Goncalves, "AN Energy Efficient Clustering algorithm for wireless sensor networks", IJCSNS International journal of computer science and network security, Vol-12, No-10, October 2012, pp-6-15
- [9] Rao Rayaoudi, "An Intelligent Water Drop Algorithm for Solving Economic Load Dispatch Problem", International journal of Electrical and Electronics Engineering 5:1, 2011, pp-43-49
- [10] Liu Ban-teng, Chen You-rong, Zhou Kai, Jingyu Hua, "The research of wireless sensor networks optimization algorithm based on control", Third International Symposium on Information Processing, 2010, pp-420-422
- [11] Malay K Pakhira, "Finding number of clusters before finding clusters", Procedia Technology 4 (2012) 27-37, published by Elsevier Ltd.
- [12] Anna Forster, Alexander Forster, Amy L. Murphy, "Optimal cluster sizes for wireless sensor networks: An experimental analysis", ADHOCNETS 2009 LNICST 28 pp-49-63, 2010
- [13] Benjamin Auffarth, "Clustering by a genetic algorithm with biased mutation operator", 2010 IEEE WCCI CEC. Barcelona, Spain, July 18-23, 2010. doi:10.1109/CEC.2010.5586090.
- [14] Jianguo SHAN, Lei DONG, Xiaozhong LIAO, Liwei SHAO, Zhigang GAO, Yang GAO, "Research on Improved LEACH Protocol of wireless sensor networks", PRZEGLĄD ELEKTROTECHNICZNY Vol 2013, No 1b

AUTHORS PROFILE



VBIT.



B. Brahma Reddy obtained his B.Tech from JNTU and M.Tech from IIT, Madras in 1980 & 1982 respectively. He has worked for Indian Institute of Science, Indian Telephone Industries, National Informatics Centre, DishnetDSL, Reliance Infocomm for nearly 25 years. Past 6 years he is working as Professor in JNTU affiliated college. Currently he is pursuing his Doctoral programme and working for

Prof. K. Kishan Rao obtained his B.E., M.E., degrees both in Distinction from O.U. College of Engineering, Osmania University in 1965 and 1967 respectively and Ph.D. from IIT Kanpur in the year 1973. Joined Regional Engineering College, Warangal in 1972 and Retired as Principal in 2002. He has Guided 3 Ph.D. Candidates, 76 M.Tech. Dissertations, 5 Research Projects and published about 68 Technical Papers in National and international Journals and Conferences. He is a Senior Member of IEEE, Life Member of IETE, Life Member of ISTE and Life Member of APSA. His Research Interest are Wireless & Mobile Communications, Adaptive Digital Signal Processing and OFDMA Networks. Presently working as Director and Adviser to Vaagdevi Group of Technical Institutions, Warangal.