International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 04, April 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

RECENT TRENDS IN BALANCED CLUSTERING

Sonam Gupta[1] , Dr. Amit Sharma [2]

[1] M.Tech Scholar, [2] Professor, Department of Computer Science & Engineering,

Vedant College of Engineering & Technology ,Bundi ,Rajasthan,(India)

ABSTRACT

In this research Paper I am focusing on The In this paper, we consider a network of energy constrained sensors deployed over a region. Each sensor node in such a network is systematically gathering and transmitting sensed data to a base station (via cluster- heads). This paper focuses on reducing the power consumption of wireless sensor networks. Firstly, we pro-posed an Energy-balanced Clustering Routing Algorithm called LEACH-L, which is suitable for a large scope wireless sensor network. Secondly, optimum hop-counts are deduced. Lastly, optimum position of transmitting node is estimated. Simulation results show that our modified scheme can extend the network lifetime by up to 80% before first node dies in the network. Through both theoretical analysis and numerical simulations, it is shown that the proposed algorithm achieves higher performance than the existing clustering algorithms such as LEACH, LEACH-M.

I. INTRODUCTION

Wireless Sensor Networks (WSNs) can collect reliable and accurate information in distant and hazardous envi-ronments, and can be used in National Defence, Military Affairs, Industrial Control, Environmental Monitor, Traf-fic Management, Medical Care, Smart Home, etc. The sensor whose resources are limited is cheap, and depends on battery to supply electricity, so it's important for Rout- ing to efficiently utilize its power.

In this paper, we propose an advanced multiple-hop routing protocol named LEACH-L, whose features can be described as follows: when the cluster-heads are close to, they directly communicate with Base station (BS); when they are far in distance, they telecommunicate by multiple-hop way, and the shortest transmission distance is limited. The sensors in different areas use different frequencies and gaps to communicate with BS. In the last part, the authors respectively simulate LEACH [1], LEACH-M [2], and LEACH-L on the Matlab. The simu-lation experiments indicate that: LEACH-L can prolong the whole network lifetime.

Keywords: Wireless Sensor Network, LEACH, Energy Efficient, Multiple-Hop

1.1 Review of the existing Balanced Clustering Algorithms

Bannerjee and Ghosh[7] in 2006 address the need of balance in clusters along with scalability. A lot of work has been done to make the clustering algorithms, especially k-means, scalable by either summarizing the information in the memory so as to reduce the database scans required or partitioning the database into small sections and perform clustering algorithms on each section. However, not even one algorithm control the number of data points to be included in a cluster and the results therefore include some very small and some extremely large clusters. With applications needing balance in clusters, authors propose a three step strategy, an

International Journal of Advance Research in Science and Engineering
Vol. No.5, Issue No. 04, April 2016
www.ijarse.com
IJARSE
ISSN 2319 - 8354

idea different from the agglomerative strategy of removing a cluster form consideration after it has reached its desired size, FSCL technique or graph partitioning algorithms using Min-cut. The proposed algorithm samples a small part of the dataset and progresses with clustering of the sampled data. The remaining population of data is then initialized to the formed clusters and refined according to the balance constraints. The results of the proposed balanced clustering technique are better than the unconstrained clustering algorithms which prove the efficiency of the proposal.

Chen et al [8] in 2006 proposed Size-Regularized (SR-Cut) cut approach in spectral clustering. The desired size of the clusters formed is passed as an input to the algorithm. The sum of inter cluster similarity and relative the distance between the two clusters constitute the cost function that optimizes when a best cut of the graph is found, thereby making the problem NP-Complete. The NP-completeness of the problem is solved through an approximation algorithm. A multi-cluster problem can be easily solved through the proposed algorithm by applying the same recursively on datasets to be clustered. A min-cut clustering technique to regulate balance in the resultant clustering is also proposed by Chang et al [13] in 2014.

II. CLUSTERING SCHEME OVERVIEW

2.1 What Is Clustering?

In clustering, the sensor nodes in a WSN which are located geographically adjacent to each form a group and come in the same cluster following some set of rules. Inside a cluster structure, sensor nodes serve either as a cluster head or a cluster member [4]. A cluster head acts like a local coordinator for its cluster and performs functions such as transmission among different clusters, data forwarding, etc. Data collected from different nodes is aggregated at the cluster head which is then forwarded to the base station as shown in Figure 1. Advantages of clustering are: (1) it reduces energy consumption by improving bandwidth utilization (2) reduces wasteful consumption of energy by reducing overhead. Most of the algorithms try to maximize the network lifetime by equalizing energy consumption between nodes and by distributing the load between different nodes from time to time [5]. The cluster head is changed during the reformation of a cluster along with the members in it.

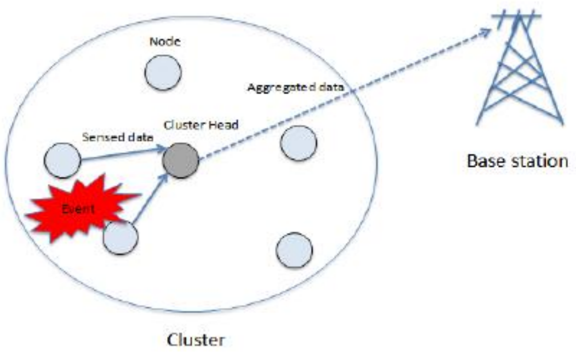

Figure1: Event detection at the cluster head

2.2 Why Do WSN Require Clustering?

Clustering is used in WSN when there are a large number of nodes involved. It increases system capacity by reusing the network's resources. Two nearby clusters cannot use the same frequency or code set [6, 7]. Cluster heads enhance the performance in case of routing as they act like a virtual backbone in intercluster routing. Clustering in WSNs is very difficult due to the inherent characteristics which differentiate these networks from other wireless networks such as mobile ad hoc networks or cellular networks [8]. It is difficult to distinguish the sensed data came from which node due to large number of nodes. Communication within one cluster as well as between different clusters can take place as combination of single hop and multi- hop as illustrated in fig. 2. In single hop communications, each sensor node directly communicates with the BS. While in multi- hop communications, nodes route their data over several hops until the data reaches BS; due to limitations in their transmission range.

2.3 Types of Clustering

Clustering algorithms differ with respect to the metrics they use for cluster control such as energy, lifetime calculations, hops, distance from the cluster head and also the type of controls such as centralized or distributed [9].

2.3.1 Hierarchical clustering (Connectivity based clustering)

In Hierarchical clustering, network is split into several clustered layers [10], and data travels from a lower clustered layer to a higher clustered layer. Data is first aggregated but as it moves from one node to another it covers greater distances that helps the data to reach the BS faster, thus reducing travel time and latency [11].

2.3.2 Data centric clustering

Data centric clustering is designed to find shortest distance between pairs of nodes [12]; it aggregates data and sends it from different recourses to a destination using named data. Since assigning global identifiers to every sensor nodes in a WSN may not be feasible due to the huge number, nodes are addressed by their locations, proximity, or capability rather than a globally unique identifier.

2.4 Challenges of Clustering

WSNs present vast challenges in terms of implementation. There are several key attributes that designers must carefully consider, which are of particular importance in wireless sensor networks.

- Cost of Clustering
- Selection of Cluster heads and Clusters
- Real-Time Operation
- Synchronization
- Data Aggregation
- Repair Mechanisms
- Quality of Service (QoS)

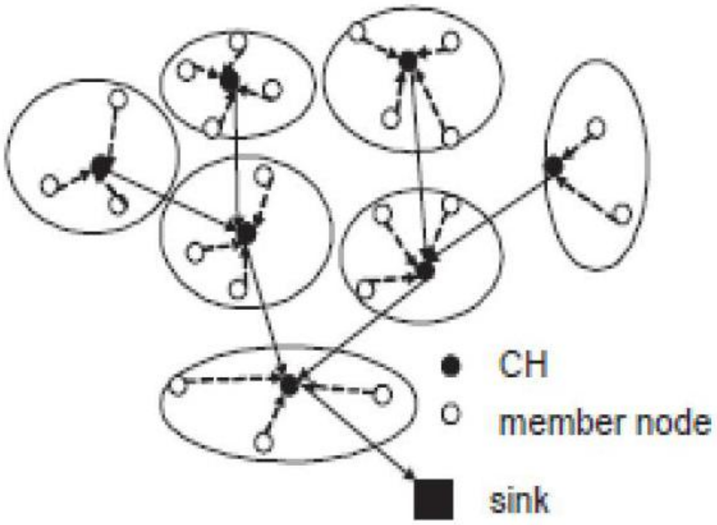

Figure 2: Network illustrating single hop and Multi-hop routing in WSN.

III. CLUSTERING PROTOCOLS

During last few years many clustering algorithms have been proposed for Wireless Sensor Networks as an efficient way for balancing the energy consumption and prolonging the lifetime of the networks [13]. Here we mention some of the recent clustering algorithms.

3.1 Equalized Cluster Head Election Routing Protocol (ECHERP)

ECHERP [14] uses an energy efficient mechanism to select a node as the cluster head. ECHERP considers the current and the estimated future residual energy of the nodes, along with the number of rounds that can be cluster heads in order to maximize the network lifetime. The protocol computes the energy consumed using the Gaussian elimination algorithm in order to minimize the overall network energy consumption at every single round. The selection of CH is done on the bases of minimization of total energy consumption; the node that does so is elected as the CH rather than the node having the highest energy left, as in many other protocols. ECHERP also adopts a multi-hop routing scheme to transfer fused data to the base station

3.2 Distance based Cluster Head Selection

This protocol [15] performs better in terms of network lifetime by balancing the energy load among all the nodes. This algorithm selects cluster head among some of the wireless sensor nodes based on net distance with base station. This technique aims to increase the lifetime of the whole network, and to increase the number of nodes, which will remain alive for the maximum period of time.

3.3 Cost Based Energy Balanced Clustering and Routing Algorithm(CEBCRA)

CEBCRA is distributed algorithm which consists of three phases namely Broadcast node, CH selection and Data Transfer. CEBCRA selects CHs amongst the normal sensor node using a weight function of the residual energy and the number of neighbors of a sensor node [16]. Then all non-CH sensor nodes join a CH, which has maximum cost value within its communication range. The cost function is the composite measure of residual energy of the CH, its distance to base station and also the distance from the sensor node to the CH. For multi-hop routing, a CH needs to relay the data to the BS through other CHs. Therefore, the other CHs are treated as relay nodes. So, a CH needs to find the best neighbor CH (relay node) for data routing such that energy consumption is minimum.

3.4 Battery Level Aware Clustering(BLAC)

BLAC considers the battery level along with density or degree of a node to choose the *cluster-head*. It aims at increasing the lifetime, thus keeping all the nodes alive to perform a specific task. Simulation results proved that BLAC increases the full network lifetime by three times than the traditional clustering schemes through balancing energy consumption of nodes and it gives high data ratio [17]. BLAC has four variants; BLAC-bg uses battery level and degree of a node; BLAC-bs combines the battery level and density of a node. BLAC-rg and BLAC-rs run in two steps. They reduce graphs before calculating the clusters. Each of these variants have their particular features and used according to the specific application.

3.5 Regional Energy Efficient Cluster Heads based on Maximum Energy Routing Protocol (REECH-ME)

The basic aim of this protocol [18] is to maximize the lifetime of a network and it especially focuses on its stability period. In this, a node having the maximum energy in a region is selected as a Cluster Head (CH) of that region for a particular round and number of the CHs in each round remains the same. Thus, this scheme enhances the desired attributes, i.e. minimum energy consumption, maximum stability period, better lifetime and throughput allot.

3.6 Energy efficient heterogeneous clustered scheme (EEHC)

EEHC [19] works on three types of sensor nodes normal nodes, advanced nodes and super nodes. There are $N \times m \times mo$ number of super nodes used with β times more energy than the normal node. Where N represents the total number of nodes m and mo are the fraction of the total number of nodes. The rest $N \times m \times (1\text{-}mo)$ number of advanced nodes which are equipped with α times more energy than the normal nodes; the remaining $N \times (1\text{-}m)$ number of normal nodes. EEHC is more effective in prolonging network lifetime as compared to LEACH.

3.7 Enhanced Developed Distributed Energy-Efficient Clustering (EDDEEC)

EDDEEC is an adaptive energy aware protocol. It dynamically varies the probability of nodes to become a CH in a balanced and efficient way in order to distribute equal amount of energy between sensor nodes. EDDEEC worked on three types of heterogeneous nodes, for probabilistic selection of CH a routing algorithm is used. [20]. EDDEEC prolongs network lifetime, longer stability period and increased number of messages sent to base station as compared to DEEC [21], DDEEC [21], and EDEEC [21].

3.8 Balanced Energy Efficient Network Integrated Super Heterogeneous Protocol (BEENISH)

BEENISH worked on four types of heterogeneous nodes as normal nodes, advanced nodes, super nodes and ultra-super nodes. In BEENISH, CH is selected based on ratio of residual energy of a node and average energy in the network field [22]. BEENISH prolongs network lifetime, longer stability period and increased number of messages sent to base station as compared to DEEC, DDEEC, and E-DEEC.

3.9 Energy Efficient Clustered Chain Based Power Aware Routing Protocol (EECCPAR)

The main goal of this protocol is to enhance the lifetime of the network and reduce the communication cost by evenly distributing the energy load among all sensor nodes [23]. This protocol divides the network into clusters and then constructs chain among cluster members and also among CHs. During data transmission CHs introduces three Thresholds i.e. MIN, MAX and Change Factor. The performance of this protocol is better then PEGASIS. This protocol doesn't suffer from clustering overhead, since clustering is not performed in each phase.

3.10 Event Driven Hierarchical Cluster based Protocol

This is an energy efficient hierarchical routing technique [24] in which cluster heads are selected based on the prediction of transmission energy via shortest possible distance to the base station for transmitting the event driven information. In this clusters of the sensor nodes are created geographically, rotate
the role of CH, and optimize the CH selection by the help of prediction of energy used for transmission in every rounds of simulation, and when event occur CH aggregates event data before transmitting it to the BS. The concept of hierarchical routing technique can be effectively used along with event driven routing to designed energy efficient routing protocol in WSN. In this algorithm, the clusters are geographically formed into different sizes to see how it could affect the network lifetime of WSN.

3.11 Distributed hierarchical agglomerative clustering (DHAC)

Hierarchical agglomerative clustering (HAC) [25] is a conceptually and mathematically simple clustering approach which uses four clustering methods, CLINK, SLINK, UPGMA, and WPGMA. All of these methods comprise three common key steps: obtain the data set, build the similarity matrix, and execute the clustering algorithm. Based on the concept of HAC, DHAC method was proposed for distributed environments by improving the HAC algorithms. The main idea behind DHAC is that a node only needs one-hop neighbor knowledge to build clusters. To apply the DHAC algorithm in WSNs, a bottom-up clustering approach is followed using simple six steps. Firstly, the qualitative connectivity data is obtained as input data set for DHAC. Secondly, the similarity matrix is built. Thirdly, the similar nodes are grouped together by executing the distributed clustering algorithm. The last three steps are cutting the cluster tree with the threshold, merging the smaller cluster, and electing the CHs. the performance of DHAC is much better than LEACH. The clustering dissipated energy of DHAC is about 4 times less than that of LEACH which predicates that DHAC can achieve much high reliability and efficiency in energy consumption at 300 rounds timing.

3.12 Density-based Energy-efficient Clustering Heterogeneous Algorithm (DECHA)

In this algorithm, the election probability of nodes to become cluster heads is evaluated [26]. With regard to the probability, there is density referring to the position information of a node, and together with its energy capacity serve as primary weighted metrics. Further evaluation is done for a better selection of cluster heads. In DECHA, the position information of nodes, define the density of each node as the number of its neighbor nodes is considered, and together with the energy capacity, regard it as an important evaluation metric for electing candidate CHs. DECHA also sets further adjustments to seek more proper CHs, thus promote both lifetime and energy-efficiency.

3.13 Low Energy Time Based Clustering Technique

This clustering technique [27] is based on the residual energy of nodes to maximize the network lifetime. Each node decides to be CH depending on its residual energy and the average energy of network's nodes. As CH performs data aggregation in order to reduce the number of packets to be transmitted to the Base Station (BS), the choice of CH must be optimal so that the compressed data at BS is reliable.

TABLE 1: Comparison on the basis of LND for Various Algorithms

Algorithm	Last node dead time(LND)		No. of nodes	Area
	LEACH	*Proposed Algorithm*		
ECHERP	After 110th round	After 145th round	500	100m × 100 m
DECHA	1195th round	4924th round	100	500m x 500m
REACH-ME	1500th round	2500th round	100	100m x 100m
DHAC	90% nodes dead till 585th second	90% nodes dead till 1120th second	Total time interval =1200 sec Node Number 100	100m x 100m

IV. CONCLUSION

Achieving balance in clusters improve the overall quality of clustering. The earlier proposals as improvements to the k-means algorithm were focused only on the initialization problem of the algorithm or in making the algorithm more scalable by reducing the number of iterations. But none brought about equal sized clusters with some of the clusters too small and the others significantly large. Equal population of data points in the resultant clusters then emerged as the need of various applications like Wireless sensor networks, document clustering, market data segmentation, online retail, image searching etc. Recently, a lot of work has been redirected towards having balance in clusters. The balance requirement is not only achieved through equal number of data points in each cluster, but also through other parameters like size or density of the cluster. This paper provides a brief review regarding the concept of balanced clustering and the note-worthy research works in this direction. Applicability of the concept to almost every field related to data mining and even others portray the importance of balance.

REFERENCES

[1] E.W.Forgy, “Cluster analysis of multivariate data: efficiency v/s interpretability of classifications”, Biometrics, Vol. 21, pp. 768–769, 1965.

[2] X. Wu et al. “Top 10 algorithms in data mining”, Knowledge and Information Systems, Vol. 14, No. 1, pp. 1-37, 2008.

[3] A. K. Jain, “Data Clustering: 50 Years Beyond K-Means”, Pattern Recognition Letters, Vol. 31, No. 8, pp. 651-666, 2010.

[4] P. S. Bradley and U. Fayyad, “Refining initial points for k-means clustering”, Proceedings of the 15th International Conference on Machine Learning, pp. 91–99, 1998.

[5] S.S. Khan and A. Ahmad “Cluster center initialization algorithm for kmeans clustering” Pattern Recognition Letters, Vol. 25, Issue 11, pp. 1293–1302, 2004.

[6] M.E. Celebi and H.A. Kingravi, “Deterministic Initialization of the K-Means Algorithm using Hierarchical Clustering” International Journal of Pattern Recognition and Artificial Intelligence, Vol. 26, Issue 7, pp. 1250018-1250034, November 2012.

[7] A. Banerjee and J. Ghosh, “Frequency-sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres”, IEEE Transactions on Neural Networks, Vol. 15, No. 3, pp. 702-719, 2004.

[8] Yixin Chen, Ya Zhang and Xiang Ji, “Size regularized cut for data clustering”, Advances in Neural Information Processing Systems, pp. 211-218, 2005.

[9] Hagen, L., Kahng, A.B.: New spectral methods for ratio cut partitioning and clus-tering. IEEE Transactions on Computer-Aided Design 11(9), 1074–1085 (1992)

[10] PS Bradley, KP Bennett and A. Demiriz, “Constrained k-means Clustering”, Redmond Microsoft Research, pp. 1-8, 2000.

[11] M. I. Malinen and P. Franti, “Balanced K-Means for Clustering”, Structural, Syntactic, and Statistical Pattern Recognition, pp. 32-41, 2014.

[12] Zhu, S., Wang, D., Li, T.: Data clustering with size constraints. Knowledge-Based Systems 23(8), 883–889 (2010)

[13] Xiaojun Chang, Feiping Nie, Zhigang Ma, and Yi Yang, “Balanced k-Means and Min-Cut Clustering”, Available at arXiv:1411.6235

[14] G Tzortzis, A. Likas, ”The MinMax k-Means clustering algorithm”, Pattern Recognition Letters, Vol. 47, pp. 2505–2516, 2014.

[15] Divya Saini and Manoj Singh, “Variance Based Clustering For Balanced Clusters In Growing Datasets”, Proceedings of the International Congress on Information and Communication Technology (ICICT-2015), Udaipur, 2015.

[16] Mohammad Hossein Bateni, Aditya Bhaskara, Silvio Lattanzi and Vahab Mirrokni, "Distributed Balanced Clustering via Mapping Coresets", Advances in Neural Information Processing Systems 27 (NIPS 2014).

[17] Buyang Cao and Fred Glover, "Creating balanced and connected clusters to improve service delivery routes in logistics planning", Journal of Systems Science and Systems Engineering, Volume 19, Issue 4, pp 453-480, 2010

[18] Christopher Tim Althoff; Adrian Ulges and Andreas Dengel , "Balanced Clustering for Content-based Image Browsing", Proceedings of the GI-Informatiktage 2011, .V. 3, 2011.

[19] Heena Chawla, Parveen Verma, "Balanced K Means Based Clustering Algorithm for Energy Efficient in Wireless Sensor Networks", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 3, Issue 11, 2014

[20] Javier Monroig Vives, "Balanced K-Means Clustering Algorithm For A Multi-Objective Partitioning In The Bike Repositioning Problem", Department of Civil and Environmental Engineering, South Kensington campus, Imperial College London.

[21] Hemant Rajput, Shekhar Sharma, Amit Mistry, Sagar Soni, Mrs. Suvarna Jadhav. "INTRANET BASED E-LEARNING USING REAL TIME TRANSMISSION PROTOCOL." International Journal of Advanced Technology in Engineering and Science 3.Special Issue No. 01 (2015): 58-65.