

Clustering Techniques in Wireless Sensor Network: A Review

Lovepreet Kaur
M.Tech Scholar

Department of Computer Science and Engineering
Amritsar College of Engineering and Technology
Amritsar (India)

Sandeep Kad
Associate Professor

Department of Computer Science and Engineering
Amritsar College of Engineering and Technology
Amritsar (India)

ABSTRACT

Wireless sensor system includes hundreds to thousands of sensor nodes that helps in collecting different information including temperature, sound, area, etc. It's generally difficult to recharge or change the sensor nodes which may have confined battery capacity. Energy efficiency is therefore a key problem in sustaining the network. Certainly one of the most used alternatives to make WSNs energy-efficient is to cluster the networks. Various clustering techniques are accustomed to effectively optimize or enhance the energy of sensor nodes. In this paper we have examined the various methods of clustering which are centralized, distributed and hybrid utilized in Sensor Networks. This paper also presents a comparative study of various clustering algorithms and the issues of clustering in WSNs.

Keywords

Clustering, wireless sensor network, residual energy, multi-hop, cluster head (CH), base station (BS).

1. INTRODUCTION

Wireless Sensor Networks (WSNs) have appeared as new technologies for information gathering. The sensor units make up an accumulation of hundreds or thousands of sensor nodes implemented over a massive geographical region form a networked sensor [1]. These micro devices are small energy devices with small storage and small control power. These cheap and small-sized units have become to be possible with the newest inventions in Complementary Metal Oxide Semiconductor (CMOS) design and miniaturization techniques. Each signal node usually includes a small CPU, storage, receiver/transmitter radio and a power product [2]. The battery is an important element of a sensor. In most of the cases, it's perhaps not changeable or regular; thus it limits the period of the warning and impacts the overall purpose of network. In another area, after they're used the detectors are believed independent. Therefore, number specific treatment may be anticipated to ensure their organization. Self-configuration of these sites is needed for their performance [3]. In order to assure minimal power usage and uniform load distribution over the network, sensor nodes are arranged into clusters [4]. Clustering is one of the very effective methods in data forwarding in providing convenient framework for resource management. It may support several essential network functions within intra cluster such as channel access for cluster members and power control, along with inter clusters such as routing and code separation to prevent inter-cluster interference [1]. In order to improve the energy efficiency and reduce transmission delay, nodes are combined into several small groups named as clusters. This method of combining sensor nodes is called as clustering. For each and every cluster a selected node is known as cluster head (CH).

CH is in charge of collecting the aggregated information delivered by the other sensor node in cluster and send it to base station or sink. A cluster node is a node that has higher energy and capacity than other sensor nodes. CH offers the scalability for large counts of nodes and decreases the energy consumption. Choice of cluster head is an essential concern in developing clustering protocols [5]. Most clustering protocols use these techniques which are:

- CH with greater residual energy
- Rotation of cluster head periodically so that energy consumption of network would be balanced
- Cluster head selected on the basis of residual energy and base station communication distance [6].

2. ROUTING PROTOCOLS IN WSN

Routing protocols choose the way the nodes can communicate with one another in network. Larger amount of power or energy of network is used for transmission of data signals [7]. Routing protocols for WSN is dependent on different classification conditions such as data centric, hierarchical, location based, negotiation based, multipath based, quality of service and mobility [8]. A little description of following is given below:

2.1 Location Based

In location-based protocols sensor nodes are resolved by the way of their locations. Location information for sensor nodes is needed for sensor networks by the majority of the routing protocols to determine the distance between two specific nodes to ensure that energy usage can be estimated. Some location based protocols are: MECN, SMECN, GAF, GEAR, Span, TBF, BVGF, and GeRaF.

2.2 Data Centric

In data-centric protocols when the source sensors deliver their information to the sink, intermediate sensors can do some sort of aggregation on the data originating from multiple source sensors and deliver the aggregated data toward the sink. This technique may result in energy savings because of less transmission is needed to deliver the data from the sources to the sink. Data centric protocols are: SPIN, Directed Diffusion, Rumor Routing, COUGAR, ACQUIRE, EAD.

2.3 Hierarchical Protocols

In hierarchical structure, sensor nodes are arranged in to clusters, in which a node having lower energy can be utilized to do the sensing job and deliver the sensed data to its cluster head at small range, while a node with higher energy can be picked as a cluster head to aggregate the data from its members and ahead it to the sink. This method may not only lower the energy usage, but in addition balance traffic load and enhance the scalability. These protocols includes: LEACH, PEGASIS, HEED, TEEN, and APTEEN. [9]

2.4 Multipath-based Protocols

If data transmission is considered among the source sensors and the sink, there are two ways of routing: single-path routing and multipath routing. In single-path routing, each source sensor sends its data to the sink via the shortest path. In multipath routing, each source sensor finds the first k shortest paths to the sink and splits its load equally among these paths. The protocols are: Disjoint Paths, Braided paths, N-to-1 Multipath Discovery [9].

2.5 QoS

Along with the reducing power usage, it is important to take into account QoS requirements when it comes to delay, reliability and fault tolerance for routing in WSNs. Both fault tolerance and reliability need the implementation of greater than the required sensors so that the network can continue to work effectively and send accurate sensed data to the sink despite some sensor failures. The QoS protocols are: SAR, SPEED, Energy-aware routing [8].

3. CLUSTERING

Clustering means grouping sensor nodes geographically close into sets called "clusters". The nodes belonging in a group can execute different functions from other nodes [3]. Clustering the nodes is a well known two-layered strategy that divides the system in to two levels, and the nodes located in exactly the same layer are collected in to some clusters. To efficiently distribute the administration jobs on the set of nodes, a number of them are chose to function as the top of each and every class (cluster), which can be called as the cluster-heads (CHs). Just because a wide variety of data is furnished by the devices in WSNs are connected, the clustering employs the connection among the data, and then by aggregating them, reduces force on the device, which results in an even more efficient energy consumption. The CHs are then accountable for getting the data from typical nodes, and aggregating and moving them to the BS. A typical application of a clustered WSN is presented in "Fig 1". As revealed in the figure, the nodes in layer 1 (the standard nodes) sensing the field, generate the data, and provide them using their related CH. Then the CHs within the next layer receives these information, and subsequently perform some operations like aggregation/fusion, transmitted them to the BS in a multi-hop approach. Eventually an individual gets the information from the BS through the Internet [2]. Cluster features are divided on the cornerstone of the specifications of clusters, like the number of clusters ,cluster size ,intra-cluster communication ,inter-cluster communication The key portion of each and every clustering algorithm is the CH election, the selected CHs have a considerable impact on the clustering algorithm performance or efficiency and some traits of the CHs are: Mobility ,Node type ,Role , Clustering process ,Method ,CH selection , Algorithm complexity , Clustering nature , Clustering dynamism [2].

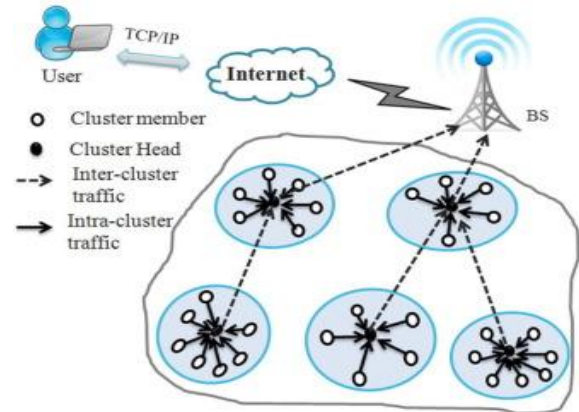


Fig 1. A typical clustered sensor network [2]

3.1 Types of Clustering Algorithm

Usually, WSNs include a large number of sensors ranging in the hundreds or even thousands. Clustering is highly effective mean for handling such large population of nodes [5]. In this part, we present and analyze some clustering algorithms based on energy efficiency for WSNs

3.1.1 Balanced Parallel K-means (BPK-means)

Tan, L., Gong, Y., Chen, G the authors have proposed a balanced parallel K-means (BPK-means)-based clustering protocol for WSN [10]. In this method, K-mean algorithm is used to cluster the sensor nodes and CHs are selected relying on the distance from the cluster centre and residual energy. It balances the clusters so the energy usage because of intra-cluster communication decreases. It just needs local communicates: each tentative CH just communicates with their topologically neighboring nodes and other tentative CH when achieves a distributed clustering scheme. Then it comes with an attractive feature of parallel computations.

Disadvantage of BPK Clustering is that it doesn't consider threshold distances in between the CH and the BS. Furthermore, by applying K-means it does consider the energy of sensor nodes for deciding the centroids. Thus a substantial quantity of energy depletion occurs for CH nodes which eventually have an effect on the network lifetime [8].

3.1.2 Load balanced clustering scheme

Another algorithm named as load balanced clustering scheme was proposed by Shujuan Jin, Keqiu Li [11]. Whole major responsibilities of a network are conducted by cluster head. That much excessive load can destroy it by consuming all of its energy. In load balanced clustering scheme an associate node is picked to simply help the cluster head to execute its data aggregation and data processing task. Associate node transfers the data to base station. Cluster head process the obtained data and transfer it to associate node. Associate node delivers this data to base station. Multi hop data transmission is employed to prevent early death of associate node.

Drawback of this scheme is that data movement among nodes is not uniform. Nodes nearer to base station get more data than nodes which are further away. Therefore nearer nodes uses more energy thus gets exhausted very soon.

3.1.3 K-means algorithm

K-means algorithm [12] is another clustering method is relies primarily on two facets for the selection of cluster head:

1. the Euclidian distances
2. residual energies of nodes

Therefore here the main node gathers the information concerning the node id, place and residual energy of nodes

and keeps these details in a list in the main node. After getting these details from all nodes it starts performing the clustering algorithm (k-mean). In K-mean algorithm, the initial centroids are selected randomly out from the input data set. Throughout each run the algorithm results in different kinds of clusters according to the different randomly chosen initial centroids. Hence random selection of initial centroids leads the algorithm into local optima [13].

3.1.4 Multi Parameter based Clustering (MPC)

Khan,A;Tamim,I,Ahmed,E [14] have proposed Multi parameter-based clustering using K-means that gives efficient clustering in WSN. It takes into account the various parameters like node energy level, Euclidean distance from base station, latency of data to reach base station in cluster formation. By varying different parameters, it can generate cluster with evenly distributed nodes, enhance intra- and inter-cluster distances and is effective at choosing less power consuming CHs.

Problem with this scheme is that, the number of preferred clusters is not calculated extensively and is thus needed to be inserted as a user input to the algorithm [14].

3.1.5 EECPK-means Midpoint Algorithm

Another method for clustering has been presented by Ray, A. and De.D, in this the authors have presented EECPK-means for WSN where midpoint algorithm can be used to enhance initial centroid selection procedure. The proposed method provides balanced clusters to eventually balance the load of cluster heads (CHs) and extends the network lifetime. Residual energy is taken into account as the parameter in along with Euclidean distance used in simple K-means algorithm for suitable CH selection. Multi-hop communication from CH nodes to BS occurs depending on their distances from BS so the CHs which is far from the sensing region does not exhaust so much and can successfully deliver the aggregate data to BS [15].

3.1.6 Low-Energy Adaptive Clustering

Low-Energy Adaptive Clustering [11] [16] [17] is among the milestones in clustering algorithms. The goal of Low-Energy Adaptive Clustering was to choose nodes as cluster heads in a way that each node gets an opportunity to become cluster head. As cluster head consumes higher energy than non cluster heads, so load is equally distributed between nodes. So that a single node does not go out of energy after a short time period just because it was often selected as cluster head. Low-Energy Adaptive Clustering includes two phases of operation

1. Set up phase
2. Steady state phase

In Set-up phase clusters are made and a cluster head is selected for each cluster. Cluster head is selected based upon a probabilistic factor. Probability of a node to become a cluster head is calculated on the basis of two factors which are as follows:

1. Number of times a node has been a cluster head.
2. Suggested total number of cluster heads for a Network.

If the value of probabilistic factor for a node is less than threshold, then it is elected as cluster head. In steady state phase all data collected by cluster heads is sent to base station. Drawback of Low-Energy Adaptive Clustering is that it does not consider initial energy as a factor to select cluster head. Therefore Nodes which have become cluster heads for same

number of time as others but have less initial energy than other nodes are probably to become lifeless prior to nodes which may have high initial energies. The algorithm doesn't work very well with large sized sensor networks because algorithm utilizes the single-hop inter-cluster technique which is not optimal for large size networks.

4. LITERATURE SURVEY

Niknam, T., Amiri, B. [18] have proposed a hybrid algorithm on the basis of the combination of ant colony optimization, fuzzy adaptive particle swarm optimization and K-means algorithm which gives better clustering that improves that network life time and energy.

Ray, A. and De.D [19] presents a clustering hierarchy protocol centered on Energy Efficient Cluster Head Selection (EECHS), which seeks to lessen energy usage within the wireless sensor network and extends the life time of the network. EECHS is great in lowering energy consumption and prolonging life time of the wireless sensor network. Protocol (EECHS) offers better performance than BEC according to number of alive nodes, residual energy and energy consumption predicted on number of round.

Tan, L., Gong, Y. and Chen, G. [10] have proposed a balanced parallel K-means (BPK-means)-based clustering protocol for WSN. In this method, K-mean algorithms are applied to cluster the sensor nodes and CHs are selected relying on their distance from the cluster centre and residual energy. It balances the clusters in a way that the energy consumption because of intra-cluster communication reduces. But, clustering doesn't consider threshold distances between the CH and the BS. Furthermore, using K-means it does consider the energy of sensor nodes for determining the centroids. Thus still a substantial amount of energy depletion occurs for CH nodes which eventually has an effect on the network lifetime.

Kotobelli, E., Zanjaj, E., Alinc, M. [20] have proposed a clustering algorithm, in which nodes are arranged in clusters and deliver their data to a shift selected cluster head. It provides an improvement on energy use with regards to member nodes, by making cluster heads static. LEACH presently includes a good energy saving technique but this modification will give an easier method towards efficiency.

Ray, A. and De.D [15], the writers have presented EECPK-means for WSN where midpoint algorithm is employed to enhance initial centroid selection procedure. The proposed method generates balanced clusters to eventually balance the load of cluster heads (CHs) and extends the network life time. It takes residual energy as the parameter along with Euclidean distance used in simple K-means algorithm for proper CH selection. Multi-hop communication from CH nodes to BS occurs depending on their distances from BS.

Khan, A., Tamim, I. and Awal, M.A. [14] have presented Multi parameter-based clustering using K-means that gives efficient clustering in WSN. It considers various parameters in cluster formation. By varying various parameters, it can create cluster with evenly distributed nodes, enhance intra- and inter-cluster distances and is capable of choosing less power consuming CHs. But, the number of preferred clusters is not determined thoroughly and thus needed to be inserted as a user input to the algorithm.

Table 1. Comparison of low energy adaptive clustering, k-means, BPK-means, Multi-parameter based clustering (MPC), Mk-means, and EECPK-means

Parameters	K-means	BPK-means	MPC	Mk-means	EECPK-means
Based on	k-means algorithm	k-means algorithm	k-means algorithm	k-means algorithm	Midpoint algorithm and k-mean algorithm
No. of clusters formed	Multiple	Multiple	Multiple	Multiple	Multiple
Balanced cluster	Yes	Yes	No	Yes	Yes
Optimum no. of CH	Yes	Yes	Yes	No	Yes
Considers minimum distance between the sensor node and CH	Yes	Yes	Yes	No	Yes
Threshold distance between CH and BS	No	No	Yes	No	Yes
Residual energy	Yes	No	Yes	Yes	Yes
Supports multi-hop	No	No	No	No	Yes
Improves network lifetime	Yes	No	Yes	Yes	Yes
Communication range of sensor nodes	No	No	No	No	No

5. CONCLUSION

In this study several clustering methodologies are analyzed and mentioned considering by which situation these algorithms are appropriate to use and where it's not giving efficient results. Some are compared on the basis of the different clustering parameters like cluster formation requirements ,distance of sensing nodes from base station, residual energy and threshold energy and calculation of the optimum number of cluster head. In wireless sensor networks confined energy is a crucial problem .Many researchers have already been performed to provide various techniques of lowering energy use in various networking environment. Clustering is one of these techniques that achieve better energy usage by splitting the entire network in to various clusters. But however there are number of dilemmas or problems to be resolved. As a future scope we can plan to consider residual energy for selection of CH, threshold distance between CH and BS for better energy use that may increase network lifetime of the network and in many approaches number of desired clusters are calculated by using number of sensors and the size of sensing region thus clusters produced are large in size in reference to the communication range of the nodes so cluster formation process can also take communication range of the nodes into account along with the size of the network in order to produce more optimal sized clusters. In such a case the cluster member can communicate more effectively with the cluster heads and thus expend lesser energies.

6. REFERENCES

[1] M.Tubaishat and S.Madria. "Sensor Networks: An overview", IEEE Potentials, pp: 20-23, 2003.
 [2] Afsar M., Mohammad H. and Tayaranim N: "Clustering in sensor networks: A literature survey "Elsevier Network and Computer Applications, pp: 198–226, 2014.
 [3] Chaudhary R., Vatta.S., "Review Paper on Energy-Efficient Protocols in Wireless Sensor Networks ", IOSR Journal of Engineering (IOSRJEN) ,pp: 2278-8719,2014
 [4] Vivek Mhatre, Catherine Rosenberg, "Homogeneous vs. heterogeneous clustered sensor networks: A Comparative

Study", IEEE International Conference on Communications (ICC 2004), 2004.

[5] Abbasi A. and Younis M. , "A survey on clustering algorithms for wireless sensor networks", Elsevier Computer Communications, pp: 2826 – 2841, 2007
 [6] Singh J.,Kumar R., Mishra A.K.," Clustering Algorithms for Wireless Sensor Networks: A Review" IEEE, pp:637978-9-3805-4416-8 , 2015
 [7] Sharma .I. Singh. R., Khurana .M, "Comparative Study of LEACH, LEACH-C and PEGASIS Routing Protocols for Wireless Sensor Network", IEEE, International Conference on Advances in Computer Engineering and Applications (ICACEA), PP: 842-846, 2015.
 [8] Patil.M, Biradar. R.C., "Survey on routing protocols in wireless sensor network", IEEE, pp: 86-91, 2012.
 [9] Goyal.D, Tripathy.M.R, "Routing protocols in wireless sensor networks: A Survey", IEEE Second International Conference on Advanced Computing & Communication Technologies, pp: 474-480, 2012.
 [10] Tan L., Gong Y. and Chen G., "balanced parallel clustering protocol for wireless sensor networks using K-means techniques", IEEE International Conference on Sensor Technologies and Applications, pp: 300–305, 2008.
 [11] Mamta, "Various clustering techniques in wireless sensor network" International Journal of Computer Applications Technology and Research, pp: 381-384, 2014.
 [12] Sasi kumar.P, Khara.S, "k- Means clustering in wireless sensor networks". IEEE Fourth International Conference on Computational Intelligence and Communication Networks, pp: 140-144, 2012.
 [13] Yong.G., Park, Kim.H., Jeong.H.W., and Youn.H.Y., "A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network"., IEEE 27th International Conference on Advanced Information Networking and Applications Workshops,pp:910-915,2013

- [14] Khan .A., Tamim.I, Ahmed .E, “Multiple Parameter Based Clustering (MPC): prospective analysis for effective clustering in wireless sensor network (WSN) using K-mean algorithm "Wireless sensor networks, pp: 18-24., 2012.
- [15] Ray A. and De D., “Energy Efficient Clustering Protocol based on K-mean (EECPK means) midpoint algorithm for enhanced network lifetime in wireless sensor network”, IET wireless sensor system, pp: 181-191, 2016.
- [16] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, “Energy-efficient routing protocols for wireless micro sensor networks,” in Proc.33rd Hawaii Int. Conf. System Sciences (HICSS), 2000.
- [17] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, application- specific protocol architecture for wireless sensor network”, IEEE Transaction on Wireless Communications, pp: 660-670, 2002.
- [18] Niknam, T., Amiri, B.: “An efficient hybrid approach based on PSO, ACO and K-means for cluster analysis”, Appl. Soft Comput., pp. 183–197, 2010.
- [19] Ray A. and De D., “Energy efficient cluster head selection in wireless sensor network”, IEEE International Conference on Recent Advances in Information Technology (RAIT’12), pp: 306–311, 2012.
- [20] Kotobelli E., Zanj E. and Alinci M., “A Modified Clustering Algorithm in WSN” International Journal of Advanced Computer Science and Applications (IJACSA’15), Vol. 6(7), 2015.
- [21] Jennifer.Y, Mukherjee. B., Ghosal.D, “Wireless sensor survey”, Elsevier Network and Computer Applications, pp: 2292-2330, 2008.
- [22] Firdaus T., Hasan .M, “A Survey on Clustering Algorithms for Energy Efficiency in Wireless Sensor Network”, IEEE International Conference on Computing for Sustainable Global Development (INDIACom), pp: 759-763,2016.