International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 04 | April-2016 www.irjet.net p-ISSN: 2395-0072

A Survey of different Clustering Algorithm in Wireless Sensor Network

Mahesh A. Kasar[1], Prof. S. D. Chavan[2]

[1]PG Student, Dept. of Electronics and telecommunication, D. Y. P. I. E. T, Pimpri, Pune
[2] Professor, Dept. of Electronics and telecommunication, D. Y. P. I. E. T, Pimpri, Pune

---***---

Abstract - *In recent years, wireless sensor networks have gained a very much importance due to their potential ability in providing solutions in various areas like health care, defense, environment, industry, surveillance and transport. Wireless sensor network are made out of a large number of low cost sensor node which are densely separated in distributed environments. The information collected by the sensors should be transmitted to a base station. Hierarchical clustering protocols have been used towards the above directions. Also, they can greatly contribute to overall system scalability, lifetime and energy efficiency. One of the most important issues in this type of networks is limited energy. Clustering is suitable method for increasing network lifetime that protects the limited sensor resources by energy saving. In this paper, different clustering algorithms which are mostly used in wireless sensor network have been explained. Also the different clustering algorithms are classified by using various approaches.*

Key Words: **wireless sensor network, sensor nodes, clustering algorithms**.

1. INTRODUCTION

Wireless sensor networks (WSNs) consist of sensor nodes connected to each other through wireless communication protocols. WSNs have many advantages such as easy random deployment, low-cost and small-size. WSNs consist of hundreds or thousands of tiny sensor nodes that are scattered in the sensor field which is the area in which the sensor nodes are deployed. Deployment of sensor nodes can be either in random fashion such as in disaster situation where for example sensor nodes are dropped from an airplane or in planned manner such as deployment of WSNs for fire alarm systems. The Base Station (BS) or sink is responsible for forwarding the desired information from WSN to the main controller through the internet via satellite or other wireless technology. Most of WSNs have fixed sensor nodes and BS. Each sensor node can collect and route the data either to other sensor nodes or back to the sink node via the path between them.

Naturally, grouping of sensor nodes into clusters has been widely adopted by researchers to satisfy different objectives and achieve high energy efficiency and prolong network life time in large scale WSN environments. The corresponding hierarchical routing and data gathering protocols give cluster based organization of the sensor nodes in order that data fusion and aggregation are possible, so that a significant energy is saved. In the hierarchical network structure each cluster contains a leader, which is also called cluster head (CH) which perform data fusion and data aggregation operation and several common sensor nodes are members. The cluster formation is two level processes where the CH nodes form the higher level and the cluster-member nodes form the lower level. The sensor nodes transmit their data to the corresponding CH nodes. The CH nodes collect data from all member nodes and transmit them to the base station either directly or through the intermediate communication with other CH nodes which is shown in fig 1.1. However, because the CH nodes send all the time data to BS, they naturally spend energy at higher rates. We have to periodically re-elect new CHs in each cluster which gives a common solution in order to balance the energy consumption among all the network nodes.

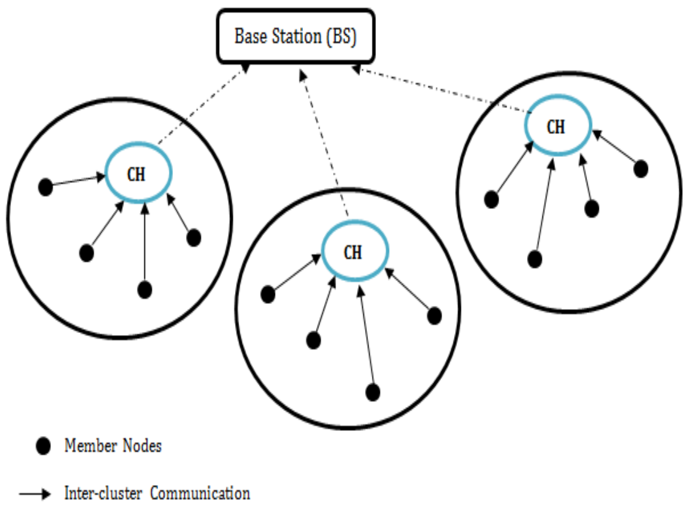

Fig 1.1 Data communication in a clustered network

The BS is the data processing point for the received data from the sensor nodes and end user access data through BS. It is generally fixed and at a far distance from the sensor nodes. The CH nodes generally act as gateways between the sensor nodes and the BS. The CH performs common function for all the nodes in the cluster which aggregate the data before sending it to the BS. The CH is the sink for the cluster nodes and the BS is the sink for the CHs. The whole structure containing member nodes, CH and BS forms the entire wireless sensor network.

1.1 Energy Consumptions Factors in WSNs

WSNs are usually deployed in harsh environments such as space, forests and battlefields. Therefore, it is difficult to physically access the wireless sensor nodes after deployment. In many cases, it is impossible to change or recharge the depleted sensor node battery. Therefore, maintaining battery life is the most important things in WSNs because it increases the useful network lifetime. During sensing, communication and processing, energy is consumed from the battery. Sensor nodes wastes energy due to reasons outlined in, namely:

A. Collisions: Collision occurs when two sensor nodes within the same coverage area transmit packets at the same time. Therefore, the two packets interfere with each other at the receiving sensor node which cannot distinguish between them. However, collisions also happen if one sensor node transmits packets before the current transmitting sensor node finishes its transmission. This is called a partial collision.

B. Idle Listening: Basically, each sensor node in WSNs can be in active, idle or sleep modes. In active mode, a sensor node consumes energy in transmitting or receiving data. In idle mode, the sensor node consumes energy to listen to the channel. In the sleep mode, the sensor node sleeps and turns off the radio transceiver. The sensor node in the idle state listens to the channel status to initiate transmission or to wait for traffic from other sensor nodes. Thus, sensor nodes consume energy in channel listening.

C. Protocol Overhead: These are the control packets used by different communication protocols. For example, Carrier Sense Multiple Access/Collision Avoidance (CSMA/CA) use control packets to manage channel access and reduce the collisions. On the other hand, the collision-free protocols such as Time Division Multiple Access (TDMA), the control traffic are less.

D. Over-emitting: Energy is wasted when a sensor node sends packets although the receiver is not ready to receive them.

2. Design challenges of clustering in WSNs

To lower energy consumption within a cluster, hierarchical routing is best way, performing data aggregation and data fusion which decrease the number of transmitted message to the BS. On the other hand, a single-tier network can cause the gateway to overload with increase in sensor density. Such overload may cause latency in communication and tracking of events. Also, single-tier architecture is n of sensors covering a wider area of interest because sensors are typically not capable of large haul communication. Hierarchical clustering is very useful for applications which are having thousands of nodes. Scalabity gives load balancing and efficient resource utilization. Applications are requiring efficient data aggregation.

First, clustering can create the route setup within the cluster. It affects on the size of the routing table stored at the individual node. It can save communication bandwidth because it limits the inter-cluster interactions to CHs and avoids exchange of messages among sensor nodes. Also, clustering can settle the network topology at the level of sensors and thus reduces overhead. Sensors only care for connecting with their CHs and not affected by changes at the level of inter-CH tier. The CH implement optimized strategies to enhance the network operation and increases the battery life of the individual sensors and the network lifetime. A CH schedules activities in the cluster so that nodes switch to the low-power sleep mode and reduce the rate of energy consumption. Further, sensors are busy in a round-robin order and the time for their transmission and reception are determined so that the sensors reties are avoided, redundancy in coverage is limited, and medium access collision is prevented. WSNs also present particular challenges in terms of design and implementation.

In WSNs, however the limited battery power, transmission range, processing hardware and memory used, etc. of the sensor nodes make the energy efficiency and the scalability factors even more important. Moreover, the challenge of prolonging network lifetime under the above restrictions is difficult by using only traditional techniques. Some additional important factors in the design process of clustering algorithms for WSNs are given below:

A. Cluster formation: The CH selection and cluster formation procedures generate the best possible clusters. However they save the number of exchanged messages low and the total time complexity remain constant and also independent to the growth of the network. Application robustness is having high priority and the designed protocols have to produce a variety of application requirements when designing clustering and routing protocols for WSNs.

B. Secure communication: As in traditional networks, the security of data is having equal importance in WSNs too. For military applications, the ability of a WSN clustering scheme to save secure communication is more important.

C. Synchronization: Slotted transmission schemes such as TDMA allow nodes to regularly schedule sleep intervals to minimize energy usage. Such schemes require corresponding synchronization mechanisms and the effectiveness of this mechanisms is considered.

D. Data aggregation: Because this process makes energy optimization possible it is a fundamental design challenge in many sensor network schemes nowadays.

3. Classification of clustering algorithm:

There is different parameter present in clustering. On basis of these parameters, clustering is classified. These parameters are given below:

A. Number of clusters (cluster count): In most recent probabilistic and randomized clustering algorithms, formation process of cluster and the CH election occurs naturally to variable number of clusters. However, the set of CHs are preplanned and thus the number of clusters is set in advance. The number of clusters is a critical parameter with reference to the efficiency of the total routing protocol.

B. Intracluster communication: In some clustering, the communication between a sensor and its CH is to be direct (one-hop communication). However, multi-hop intracluster communication is required, when the communication range of the sensor nodes is limited or very large sensor nodes and the number of CHs is prepared.

C. Nodes and CH mobility: For stationary sensor nodes and stationary CHs, it normally led to stable clusters with easier intracluster and intracluster network management. On the other hand, if the CHs or the nodes themselves are considered to be mobile, the cluster membership for each node dynamically change; forcing clusters to change over time and probably need to be continuously maintained.

D. Nodes types and roles: In some proposed network mode, the CHs are assumed to be prepared with more computation and communication resources than others. In most common network models all nodes have the same capabilities and just subsets of the deployed sensors are designated as CHs.

E. Cluster formation methodology: For time complexity and regular CH sensor node, clustering is executed in a distributed manner without coordination. One or more coordinator nodes are used to partition the whole network off-line and control the cluster membership.

F. Cluster-head selection: The leader nodes of the clusters (CHs) in algorithms are previously assigned. In most cases however, the CHs are picked from the deployed set of nodes either in a probabilistic or based on residual energy, connectivity etc.

G. Algorithm complexity: In some algorithms the fast termination of the executed protocol is one of the primary design goals. Thus, the time complexity of most cluster formation procedures proposed is constant. In some protocols, however, the complexity time has depended on the total number of sensors in the network.

H. Multiple levels: To achieve better energy distribution and total energy consumption, the concept of a multi-level cluster hierarchy is introduced.

4. Probabilistic Clustering Approaches

Clustering algorithms are developed to meet the network requirements such as network lifetime, energy efficiency, scalability, energy consumption. Probabilistic clustering protocol plays an important role in WSNs. LEACH, EEHC and HEED are probabilistic clustering algorithms and their main objective is to reduce the energy consumption and prolong the network lifetime. LEACH, EEHC, and their extensions follow a random approach for CH election whereas others like HEED and similar approaches follow a hybrid probabilistic methodology.

Low Energy Adaptive Clustering Hierarchy (LEACH) is probably the first dynamic clustering protocol which addressed specifically the WSNs needs, using homogeneous stationary sensor nodes randomly deployed, and it serves as the basis for other improved clustering protocols for WSNs. It's a hierarchical, probabilistic, distributed, one-hop protocol, with main objectives

(a) To improve the lifetime of WSNs by trying to evenly distribute the energy consumption among all the nodes of the network and
(b) To reduce the energy consumption in the network nodes.
It forms clusters based on the received signal strength and also uses the CH nodes as routers to the BS. All the data processing such as data fusion and aggregation are local to the cluster.

Another significant probabilistic clustering algorithm is Energy Efficient Hierarchical Clustering-EEHC. The main objective of this algorithm is to address the shortcomings of one-hop random selection algorithms such as LEACH by extending the cluster architecture to multiple hops. It is a distributed, k-hop hierarchical clustering algorithm aiming at the maximization of the network lifetime.

Another improved and very popular energy-efficient protocol is HEED (Hybrid Energy- Efficient Distributed Clustering). HEED is a hierarchical, distributed, clustering scheme in which a single-hop communication pattern is retained within each cluster, whereas multi-hop communication is allowed among CHs and the BS. The CH nodes are chosen based on two basic parameters, residual energy and intracluster communication cost. Residual energy of each node is used to probabilistically choose the initial set of CHs. On the other hand, intracluster communication cost reflects the node degree or node's proximity to the neighbor and is used by the nodes in deciding to join a cluster or not. Thus, unlike LEACH, in HEED the CH nodes are not selected randomly. Only sensors that have a high residual energy are expected to become CH nodes. Also, the probability of two nodes within the transmission range of each other becoming CHs is small. Unlike LEACH, this means that CH nodes are well distributed in the network.

On the basis of the probabilistic nature of LEACH, several other protocols are developed aiming at better energy consumption and overall performance. First, the LEACH-C and the LEACH-F protocols are proposed which introducing slight modifications to the initial LEACH cluster formation procedure. LEACH-C is a centralized version of LEACH, in the sense that the responsibility of the cluster creation is transferred to the BS. LEACH-F is also a centralized protocol and is based initially on the same global clustering scheme as in LEACH-C. The main difference lies on the fact that all clusters are fixed once when they are formed, thus reducing the overhead of cluster formation in the network.

5. Graph based clustering protocol:

Graph based clustering approach is a distributed multi-hop hierarchical clustering algorithm which also efficiently extends to form a multi-level cluster hierarchy. Any node in the WSN can initiate the cluster formation process. The algorithm proceeds in two phases, “Tree Discovery” and “Cluster Formation.” The tree discovery phase is basically a distributed formation of a Breadth-First-Search (BFS) tree rooted at the initiator node. Each node, *u*, broadcasts a signal once every *p* units of time, carrying the information about its shortest hop distance to the root, *r*. A node *v* that is neighbor of *u* will choose *u* to be its parent and will update its hop distance to the root, if the route through *u* is shorter. The broadcast signal carries the parent ID, the root ID, and the sub tree size. Every node updates its sub tree size when its children sub tree size change. The cluster formation phase starts when a sub tree on a node crosses the size parameter, *k*. The node initiates cluster formation on its sub tree. It will form a single cluster for the entire sub tree if the sub tree size is less than 2*k*, or else, i will form multiple clusters. The cluster size and the degree of overlap are also considered.

In Figure 2, the multi-level hierarchy is shown. This approach has a time complexity of *O(n)*, it is balanced clustering and it handles dynamic environments very well.

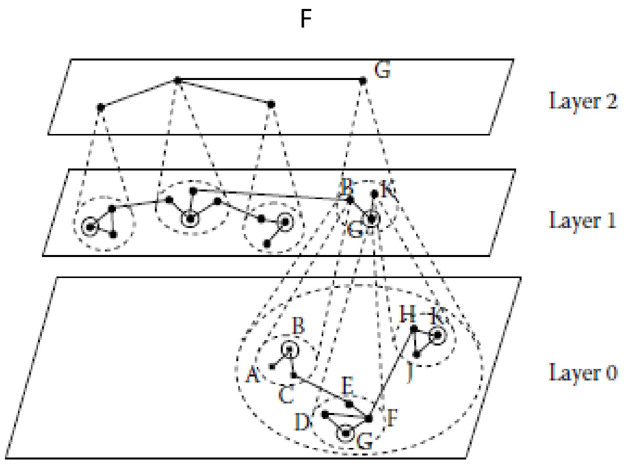

Figure 2. A multi-level clustering hierarchy

6. Weight based clustering hierarchy:

Weight-based protocol is proposed where the CH election process is based on the computation of a combined weight Wv for each node, which takes into account several system parameters such as the node degree, the transmission power, mobility, and the remaining energy of the node:

$$Wv = w1Tv + w2Dv + w3Mv + w4Pv$$

The combined weight is calculated and broadcasted by each node. The node with the smallest weight in its neighborhood is chosen as a CH.

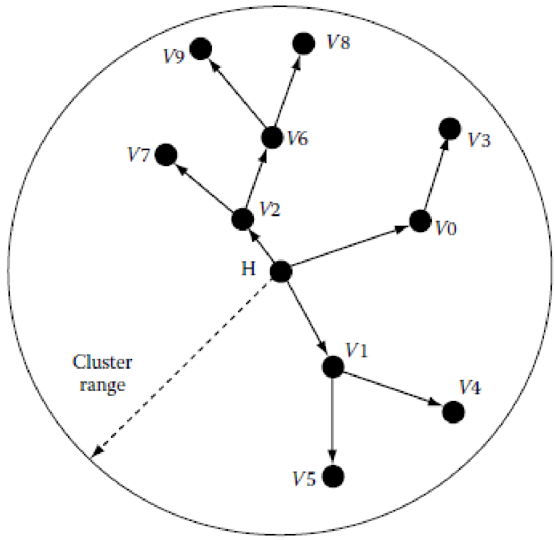

Figure 3. DWEHC multi-hop intracluster topology.

This is a non-periodic procedure for CH election. This algorithm attempted to provide better load balancing through reduced number of sensors in a cluster but the requirement of clock synchronization limits its applications. Given the node’s knowledge of the distance to its neighbors, it can assess whether it is better to stay a first-level member or become a second-level one reaching the CH over a two-hop path. Figure 3, shows the structure of the intracluster topology. Compared to HEED, the DWEHC algorithm generates more well-balanced clusters as well as to achieve significantly lower energy consumption in intracluster and intercluster communication.

7. Biological Inspired Clustering Hierarchy:

ANTCLUST is a swarm intelligence-based clustering algorithm. ANTCLUST is a model of an ant colonial closure to solve clustering problems. In colonial closure model, when two objects meet together they recognize whether they belong to the same group by exchanging and comparing information about them. In the case of a WSN, initially the sensor nodes with more residual energy become CHs. Then, randomly chosen nodes meet each other, exchange information, and clusters are created, merged, and discarded through these local meetings and comparison of their information. Each node with less residual energy chooses a cluster based on specific criteria, like the residual energy of the CH, its distance to the CH, and the cluster size. Energy efficient clusters are formed that gives large lifetime of the WSN.

6. Conclusion:

Clustering in WSNs is having high interest. In this paper, the main characteristics of the most protocols are proposed. Grouping nodes into clusters leads to hierarchical routing and data gathering protocols is the most efficient approach to support scalability in WSNs. The hierarchical cluster structures facilitate the efficient data gathering and aggregation independent to the growth of the WSN and generally reduce the total amount of communications as well as the energy spent. The main objective of most of the clustering protocols is on how to prolong the lifetime of the network and how to make a more efficient use of the resources, such as battery power. Furthermore, the combined need for fast convergence time and minimum energy consumption lead to appropriate fast distributed probabilistic clustering algorithms which quickly became the most popular and widely used in the field. Finally, several additional issues should be further studied. The most challenging issues include the development of a generic method for finding the optimal number of clusters in order maximize the energy efficiency, the estimation of the optimal frequency of CH rotation to gain better energy distribution, keeping the total overhead low, the efficient support of nodes and CHs mobility as well as the support of mobile sinks, the incorporation of several security aspects, the further development of efficient recovery protocols in case of CHs failure, etc.

7. References:

[1] Sumedha Sirsikara, Kalyani Wankhedeb, Comparison of Clustering Algorithms to Design New Clustering Approach, ELSEVIER International Conference on Advances in Computing, Communication and Control (ICAC3'15).

[2] Babu N V, Puttamadappa. C, Bore Gowda S B, "Energy Efficient Two Hop Clustering for Wireless Sensor Networks", IJCSNS International journal of Computer Science and Network Security, Vol. 13 No.9, September 2013.

[3] Kanika Sharma, Chandni, "Comparative analysis of routing protocols in Wireless sensor Network", International Journal of Computer Science and Information Technology & Security (IJCSITS), Vol. 3, No.1, February 2013.

[4] S.R.Boselin Prabhu, S.Sophia, "A Survey of Adaptive Distributed Clustering Algorithms for Wireless Sensor Networks", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.4, November 2011.

[5] M. Aslam, N. Javaid, A. Rahim, U. Nazir, A. Bibi, Z. A. Khan, "Survey of Extended LEACH-Based Clustering Routing Protocols for Wireless Sensor Networks", IEEE 2012.

[6] A. Ray and D. De "Energy Efficient Clustering Hierarchy Protocol for Wireless Sensor Network," in Proc. ICCIA, pp.1-4, Dec. 2011.

[7] O.Younis, S.Fahmy, HEED: A hybrid energy-efficient distributed clustering approach for Ad Hoc sensor networks, IEEE Transactions on mobile computing 3(4) (2004) 366-379.

[8] S.Bandyopadhyay, E.Coyle, An energy efficient hierarchical clustering algorithm for wireless sensor networks, in: Proceedings of the 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003), San Francisco, California, April 2003.