

## A Closer look on Clustering Protocols in Wireless Sensor Networks

\* Manpreet Kaur, \*\* Dr Vijay Kumar

\*Department of Computer Science, M. M. University, Mullana, Ambala, Haryana -133203

\*\*Computer Science, M. M. University, Mullana, Ambala, Haryana -133203

---

**Abstract:** - A wireless sensor network is basically a collection of nodes that are organized into a cooperative network. Each node consists of processing capability that contain multiple types of memory and have a transceiver, a power source and accommodate various sensors. These sensor nodes can sense, measure and collect information from the environment and, based on some local decision process, they can transmit the sensed data to the user. In our work we have studied the various cluster formation algorithms. In this study clustering of the nodes are considered with the approach of reducing energy consumption of nodes and a protocol is presented. Due to network clustering there is scalability potential in such a network. According to frequent change of cluster head nodes load distribution is performed in the cluster and eventually increase the network lifetime.

**Keywords:** - BS, CH, Clustering, Node, WSNs

---

### I. INTRODUCTION

In most wireless sensor network (WSN) applications nowadays the entire network must have the ability to operate unattended in harsh environments in which pure human access and monitoring cannot be easily scheduled or efficiently managed or it's even not feasible at all [1]. In many situations the sensor are deployed by the helicopters, by submarines when deployed under water. Moreover, considering the entire area that has to be covered, the resource constraint is one problem and the possibility of having damaged nodes during deployment is another, large populations of sensors are expected; it's a natural possibility that hundreds or even thousands of sensor nodes will be involved in the network. When the sensors are deployed in these kind of environments sensors are battery constrained and moreover their batteries cannot be recharged. Therefore, it is important to use specialized energy-efficient routing and data gathering protocols that offer high scalability so that network lifetime is preserved acceptably high in such environments.

One solution to above problem is grouping sensor nodes into clusters. It has also been widely adopted by the research community to satisfy the scalability objectives, also achieve high energy efficiency and prolong network lifetime in large-scale WSN environments. For a Cluster based architecture hierarchical and data gathering algorithms can only be used if data fusion and data aggregation is possible, thus leading to considerable energy savings. Talking about the hierarchical network structure each cluster has a leader, which is also called the cluster head (CH) and usually performs the specific tasks such as data aggregation, collection of data and sending it further to appropriate sensors while the other several common sensor nodes (SN) are called as members and are used to sense data and send information to their respective CH.

The cluster formation process eventually leads to a leveled hierarchy where the CH nodes form the higher level and the cluster-member nodes form the lower level. All the nodes in a network would occasionally transmit their data to the corresponding CH nodes. The CH nodes aggregate the data (thus decreasing the total number of packets) and transmit them to the base station (BS) either directly or through the intermediate communication with other CH nodes. The CH has to transmit the data to a larger distance as compared to it's members so obviously it will use more energy compared to the other nodes. Thus a common solution in order to balance the energy consumption among all the network nodes, is to periodically re-elect new CHs (rotating the CH role among other nodes over time) in each cluster. The BS is the data handing point for the data received from the sensor nodes, and it is from where the data is accessed by the user. It is usually considered to be fixed and on a outlying distance from the rest of the sensor nodes. The CH nodes are gateways between the sensor nodes and the BS. The function of each CH, as mentioned above is aggregating the data and then finally sending it to the BS. It can be said that the CH is the sink for the cluster nodes, and the BS is the sink for the CHs, because all member nodes will transfer the data to their respective CHs and further the CHs transfer data to the BS. Moreover, this structure formed between the sensor nodes, the sink (CH), and the BS can be replicated as many times as it is needed, creating (if desired) multiple layers of the hierarchical WSN (multi-level cluster hierarchy).

In this survey we have tried to explore the various clustering algorithms. The section 2 of this paper describes the classification of the clustering algorithm. Main classifications are probabilistic approach and non-probabilistic approach. And then finally Section 3 concludes the paper.

## II. CLUSTERING PARAMETERS

Before discussing the clustering algorithms let us discuss the clustering parameters.

1. *Number of clusters (cluster count)*: In most recent probabilistic and randomized clustering algorithms the CH election and formation process lead to variable number of clusters. In some approaches, the set of CHs are predetermined and thus the number of clusters are known. The number of clusters is usually a critical parameter with regard to the efficiency of the total routing protocol.
2. *Nodes and CH mobility*: If we consider our nodes and CHs to be stationary we are normally led to form stable clusters with facilitated intracluster and intercluster network organization. Whereas on the other hand, if we consider the nodes and CHs to be mobile, then cluster association for each node should change dynamically, thus forcing clusters to progress over time and perhaps need to be continuously monitored.
3. *Nodes types and roles*: In some proposed network models (i.e., heterogeneous environments) the CHs are assumed to be equipped with significantly more computation and communication resources than others. In most usual network models (i.e., homogeneous environments) all nodes have the same capabilities and just a subset of the deployed sensors are designated as CHs.
4. *Cluster formation methodology*: In the recent approaches, when CHs are ordinary sensor nodes and time efficiency is a chief design measure, here clustering has been performed in a distributed way without any proper coordination. Approaches used earlier followed a centralized (or hybrid) way where one or more coordinator nodes were used to divide the whole network off-line and control the cluster membership.
5. *Cluster-head selection*: The leader nodes of the clusters (CHs) in some proposed algorithms (mainly for heterogeneous environments) can be pre assigned. However in most of the cases (i.e., in homogeneous environments), the CHs are taken from the deployed set of nodes. It could be either in a probabilistic way or completely random way or may be based on some more specific criteria.
6. *Overlapping*: Several protocols also give high importance on the concept of node overlapping contained by different clusters (may be for improved routing efficiency or for quicker cluster formation protocol execution or some other reasons are also possible). However most of the algorithms try to have minimum overlapping or do not support overlapping at all.

The classification above is based on the functionality and characteristics of the sensors in the cluster, whereas the other is based on the technique that is used to actually form the cluster. In heterogeneous sensor networks (i.e., [3,10]), broadly there are two types of sensors, one are those that have higher processing capabilities and are complex and other are called as common sensors, which have lower capabilities, used to actually sense the preferred attributes in the field. Homogeneous networks are those where all nodes have the identical characteristics, hardware and processing capabilities. Also, when all the nodes have the same capabilities (homogeneous environments), a distributed cluster formation and CH election process is the appropriate technique to expand increased flexibility and rapid execution-convergence time's independent of the number of nodes of the WSN. Few approaches have been discussed using centralized or hybrid techniques (i.e., [5, 3, 5] where one or more coordinator nodes or the BS are responsible to divide the whole network off-line and control the cluster membership).

## III. CLASSIFICATION OF CLUSTERING ALGORITHMS

Broadly the clustering algorithms are classified into two categories, i.e. Probabilistic (random or hybrid) clustering algorithms and Non probabilistic clustering algorithms.

### 3.1 POPULAR PROBABILISTIC CLUSTERING PROTOCOLS

In case of probabilistic selection clustering algorithms [4–8], a prior probability is assigned to each sensor node is used to determine the initial CHs. The probabilities initially assigned to each node often serve as the primary (random) criterion in order for the nodes to decide individually on their election as CHs (in a simple, consistent, quick and completely distributed manner). The probability may be based upon the residual energy of the nodes or the communication cost among the nodes according to the requirement.

#### 3.1.1 LOW ENERGY ADAPTIVE CLUSTERING HIERARCHY (LEACH)

One of the first and most popular clustering protocols proposed for WSNs was LEACH (low energy adaptive clustering hierarchy) [4, 5]. It's an hierarchical, probabilistic, distributed, one-hop protocol, with key 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.

It forms clusters based on the received signal strength and also uses the CH nodes as gateways to the BS. All the data processing such as data fusion and aggregation are local to the cluster. In this algorithm the clusters are formed by using a distributed algorithm. In a distributed algorithm the sensor nodes make independent decisions without any centralized control. All nodes have an equal chance to become CHs to balance the energy spent per round by each sensor node. When a node initially decides to be a CH with a probability “p” then it broadcasts its decision. After its election, each CH broadcasts an advertisement message to the other nodes on receiving the message the other (non-CH) nodes decides to fit in to which cluster, by selecting the CH that can be easily reached thus using the least communication energy.

**Advantages:**

- Localised coordination scheme.
- Improved scalability for cluster formation.

**Disadvantages:**

- It cannot be used on large area networks.
- It does not guarantee good cluster heads distribution over the network.

### **3.1.2 ENERGY-EFFICIENT HIERARCHICAL CLUSTERING (EEHC)**

Another significant probabilistic clustering algorithm was earlier proposed in Ref. [6] is EEHC. It improved the shortcomings of single hop algorithm like LEACH. EEHC is a distributed, k-hop, hierarchical clustering algorithm. It aims at improving the lifetime of a network. Initially, each sensor node is elected as a CH with probability “p” and announces its election to the neighboring nodes within its communication range. The above CHs are now called the “volunteer” CHs. Next, all the nodes that are within “k”-hops distance from a “volunteer” CH, these nodes are supposed to accept the election message that could be either directly or it can even be forwarded through any intermediate node. As a result, any node that has received such CH election message and also is not itself a CH, then it becomes a member of the closest cluster. Additionally, a number of ‘forced’ CHs are elected from nodes that are neither CHs nor belong to a cluster. In particular, sometimes the election messages may not reach a node within a preset time interval  $t$ , the node becomes a “forced” CH assuming that it is not within k hops of all volunteer CHs.

**Advantages:**

- It extends the cluster architecture to multiple-hop architecture.
- It is suitable for large networks.

**Disadvantages:**

- It has large time complexity.

### **3.1.3 HYBRID ENERGY-EFFICIENT DISTRIBUTED CLUSTERING (HEED)**

Another improved and very popular energy-efficient protocol is HEED (Hybrid Energy-Efficient Distributed Clustering [7]). HEED is a distributed, hierarchical clustering design wherein a single-hop communication pattern is used within each cluster, where as multi-hop communication is allowed among the CHs and Base Station(BS). Two basic parameters are calculated for selecting the CH nodes, firstly the residual energy and secondly intracluster communication cost. On knowing the residual energy of each node we can probabilistically select the initial set of CHs. In contrast, intracluster communication cost is used by the nodes in deciding to join which cluster. Thus, not like LEACH, in this algorithm the CH nodes are not selected randomly merely only those sensors that possess a high residual energy have chance to become CH nodes.

**Advantages:**

- Cluster Heads are well distributed over the network.
- The nodes with high residual energy are expected to become CHs, thus increasing the lifetime of the network.

**Disadvantages:**

- It requires synchronization.
- Knowledge about the entire network is needed to determine reliability of network.

## **3.2 NON-PROBABILISTIC CLUSTERING**

In the category of non probabilistic clustering algorithms [9–10], more specific (deterministic) criteria for CH election and cluster formation are primarily considered, which are mainly based [9–11] on the nodes’ proximity (quantity, connectivity etc.) and also considering the information that is received from closely located nodes. In this generally formation of cluster is based on the communication of nodes with the neighbors (can be

either single-hop or multi-hop neighbors) and generally it requires more thorough exchange of messages and to some extent graph traversing may also be required, thus sometimes leads to worse time complexity than the probabilistic / random clustering algorithms. Quite the reverse these algorithms are frequently more reliable toward the direction of extracting robust and well-balanced clusters.

### **3.2.1 NODE PROXIMITY AND GRAPH-BASED CLUSTERING PROTOCOLS**

Such a node proximity-traversing-based algorithm have been earlier proposed in Ref. [10] (like Hierarchical Control Clustering ). It is a multi-hop, hierarchical clustering and distributed algorithm. It efficiently extends to form a multi-level cluster hierarchy. Any node in the network can initiate the cluster formation process. This algorithm proceeds in two phases, firstly “Tree Discovery” is done and then “Cluster Formation” is carried out. The tree discovery phase is basically a distributed formation of a Breadth-First-Search (BFS) tree rooted at the initiator node. Where each node,  $u$ , broadcasts a signal once every  $p$  units of time, it carries the information about its shortest hop distance toward the root node,  $r$ . All the neighbours  $u$  which find  $v$  closer will select  $v$  to be their parent and will then update their hop distance to the root, if the route from beginning to end till  $u$  is shorter. The signal that is broadcasted will carry its parent ID, its root ID, and also the sub tree size. For each node if the children sub tree size changes then the node will update its sub tree size. Actually the cluster formation phase will start when a sub tree on a node crosses the size parameter,  $k$ . Each node initiates cluster formation on their sub tree. If the sub tree size is less than  $2k$  then a single cluster will be formed for the entire sub tree, otherwise there will be multiple clusters formed. It is also must to consider the degree of overlap and the cluster size.

#### **Advantages:**

- These algorithms achieve balanced and stable clusters.
- Low intracluster communication cost.

#### **Disadvantages:**

- Worse lifetime as compared to probalistic algorithms.
- These algorithms also lead to increased complexity time.

### **3.2.2 WEIGHT-BASED CLUSTERING PROTOCOLS**

In addition to node proximity, some other known algorithms use a combination of metrics such as the energy remaining, amount of transmission power, etc., (or the form of equivalent combined weights) to achieve more generalized goals than single-criterion protocols. In these algorithms firstly the various weights depending on the requirement are calculated. There are various algorithms following this directive were initially borrowed from the field of mobile ad hoc networks, i.e., [12] and [13]. The weight to be considered can be any weight either a combined weight of node degree, the transmission power, mobility, and the remaining energy of the node. Thus the CHs are also choosen based on weights and further the members are also formed based on the desired weights.

#### **Advantages:**

- These provide well balanced and stable clusters.
- It also achieves better distribution of energy consumption because it considers energy while electing CHs.

#### **Disadvantages:**

- Increased communication time.
- It does no CHs rotation.

### **3.2.3 BIOLOGICALLY INSPIRED CLUSTERING APPROACHES**

In the last few years some new algorithms have also been proposed based on swarm intelligence techniques which are based on the collective behavior of social insects such as ants. They have shown excellent results in simulated experiments (compared to protocols like LEACH and HEED) with regard to network lifetime. As compared to the probalistic algorithms these algorithms have overtaken in increasing the network lifetime. In Ref. [14] the authors have proposed such a swarm intelligence-based clustering algorithm which is based on the ANTCLUST method. This method basically models an ant colony closure which is further used in solving clustering problems. In this model when two nodes meet they exchange and compare information to know whether they meet to same group or not. In WSN, initially the sensor nodes with more residual energy will become CHs autonomously. Followed by, when the randomly chosen nodes meet each other, then they exchange information among themselves, and thus clusters are formed, fused, and are made redundant through these local meetings and thus evaluating their information. Nodes with less residual energy will select a cluster based on some specific criteria, it is like the residual energy of the CH, may be its distance to the CH, or an estimate of the cluster size. Ultimately, energy efficient clusters are formed as the result in an expansion of the lifetime of the network.

**Advantages:**

- It achieves uniform distribution of CHs and the energy consumption.
- Robust and adaptive to topological changes.

**Disadvantages:**

- These cannot establish the shortest or appropriate paths before a sufficient number of agents is flooded.
- Overhead due the swarm intelligence meta-heuristic is faced by these algorithms.

#### **IV. CONCLUSION**

Generally, clustering in WSNs has been of high interest in the last decade and there is already a large number of related published works. Throughout this paper we tried to present the main characteristics of the most significant protocols that have been already proposed till now. As it was pointed out, grouping nodes into clusters, thus leads to hierarchical routing and data gathering protocols, have been accepted as the most efficient way to maintain scalability in WSNs. The hierarchical cluster structures do efficiently gather data and aggregate it independent to the increase of the WSN, generally shrink the total amount of communications plus the energy spent. As we have seen main objective of the above algorithms lie on to reduce battery consumption, make it more energy efficient and thus prolong lifetime. In some algorithms the nodes are assumed to make fast decisions (i.e., to become CHs or not) based on some probability or other local information only (i.e., on their residual energy) and usually the desired quality of the final cluster output is considered as a secondary parameter only. Another significant feature of most of these algorithms is the periodic reelection of CHs (rotation of the CH role) among all the nodes of the network. Few Clustering algorithms that adopt primary election criteria other parameters like connectivity, the nodes' proximity, or distance, can also be developed and it has also revealed good results and appropriate protocols are still in use, leading possibly to extra qualitative outputs (well balanced clusters). Though the time complexity factor of these algorithms is somewhat difficult to be kept low as in leading probabilistic/random clustering algorithms.

#### **REFERENCES**

- [1] A. A. Abbasi and M. Younis, A survey on clustering algorithms for wireless sensor networks, *Computer Communications*, 30, 2823–2814, 2007.
- [2] G. Gupta and M. Younis, Load-balanced clustering in wireless sensor networks, in *Proceedings of the International Conference on Communication (ICC 2003)*, Anchorage, AK, May 2003.
- [3] E. I. Oyman and C. Ersoy, Multiple sink network design problem in large scale wireless sensor networks, in *Proceedings of the IEEE International Conference on Communications (ICC 2004)*, Paris, June 2004.
- [4] W. R. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, Energy efficient communication protocol for wireless microsensor networks, in *Proceedings of the 11<sup>th</sup> Hawaaian Interantional Conference on System Sciences*, January 2000.
- [5] W. B. Heinzelman, A.P. Chandrakasan, and H. Balakrishnan, An application specific protocol architecture for wireless microsensor networks, *IEEE Transactions on Wireless Communications*, 1(4), 110–120, 2002.
- [6] S. Bandyopadhyay and E. Coyle, An energy efficient hierarchical clustering algorithm for wireless sensor networks, in *22nd Annual Joint Conf. of the IEEE Computer and Communications Societies (INFOCOM 2003)*, San Francisco, CA, April 2003.
- [7] O. Younis and S. Fahmy, HEED: A hybrid, energy-efficient, distributed clustering approach for Ad Hoc sensor networks, *IEEE Transactions on Mobile Computing*, 3(4), 113–129, 2004.
- [8] G. Li and T. Znati, RECA: A ring-structured energy-efficient clustering architecture for robust communication in wireless sensor networks, *International Journal Sensor Networks*, 2(1/2), 2007.
- [9] H. Zhang and A. Arora, GS3: Scalable self-configuration and self-healing in wireless networks, in *Proceedings of the 21st ACM Symposium on Principles of Distributed Computing (PODC 2002)*, Monterey, CA, July 2002.
- [10] S. Banerjee and S. Khuller, A clustering scheme for hierarchical control in multi-hop wireless networks, in *Proceedings of 20th Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'01)*, Anchorage, AK, April 2001.
- [11] S. Yi, J. Heo, Y. Cho, and J. Hong, PEACH: Power-efficient and adaptive clustering Hierarchy Protocol for WSNs, Elsevier, *Computer Networks*, 30, 2842–2852, 2007.
- [12] S. Basagni, Distributed clustering for Ad Hoc networks, in *Proceedings of Intl. Symp. on Parallel Architectures, Algorithms & Networks (ISPAN)*, pp. 310–315, 1999.
- [13] M. Chatterjee, S.K. Das, and D. Turgut, WCA: A weighted clustering algorithm for mobile ad hoc networks, *Clustering Computing*, 5, 193–204, 2002.
- [14] J. Kamimura, N. Wakamiya, and M. Murata, A distributed clustering method for energy-efficient data gathering in sensor networks, *Int. J. Wireless and Mobile Computing*, 1(2), 43–50, 2003.