ISSN 2320-9984

Volume 5 Issue 1 | March 2017 ISSN: 2320-9984 (Online)

International Journal of Modern Engineering & Management Research

Website: www.ijmemr.org

# A Review on Clustering Algorithms in Wireless Sensor Networks

**Akanksha Nema**
*M. Tech. Research Scholar*
*Takshshila Institute of Engineering & Technology*
*Jabalpur (M.P.), [INDIA]*
*Email: skyakankshanema@rediffmail.com*

**Deepak Agrawal**
*Assistant Professor*
*Department of Computer Science Engineering*
*Takshshila Institute of Engineering & Technology*
*Jabalpur (M.P.), [INDIA]*
*Email: deepakagrawal@takshshila.org*

***Abstract**—The use of remote sensor frameworks (WSNs) has grown gigantically in the latest decade, pointing out the basic necessity for versatile and imperativeness efficient guiding and data party and gathering traditions in relating tremendous scale circumstances. To help mastermind lifetime in Wireless Sensor Networks (WSNs) the routes for data move are picked to such an extent that the total essentialness exhausted en route is restricted. To help high adaptability and better data aggregation, sensor center points are consistently gathered into disjoint, non covering subsets called gatherings. Packs make different leveled WSNs which join compelling utilization of limited resources of sensor center points and subsequently expands orchestrate lifetime. The objective of this paper is to show a diagram on clustering counts declared in the written work of WSNs. This paper presents logical arrangement of imperativeness capable clustering estimations in WSNs.*

***Keywords:**—Clustering algorithms, Energy efficient clustering, Network lifetime, Wireless sensor networks.*

## 1. INTRODUCTION

A wireless sensor network is a collection of nodes organized into a cooperative network [4]. Each node consists of processing capability (one or more microcontrollers, CPUs or DSP chips), may contain multiple types of memory (program, data and flash memories), have a RF transceiver (usually with a single omni-directional antenna), have a power source (e.g., batteries and solar cells), and accommodate various sensors and actuators. The nodes communicate wirelessly and often self-organize after being deployed in an ad hoc fashion. Systems of1000s or even 10,000 nodes are anticipated. Such systems can revolutionize the way we live and work.

As sensor networks have limited and non -rechargeable energy resources, energy efficiency is a very important issue in designing the topology, which affects the lifetime of sensor networks greatly.

## 2. CLUSTERING

The grouping of sensor nodes into clusters has been widely pursued by the research community in order to achieve the network scalability objective. Every cluster would have a leader, often referred to as the cluster-head (CH). Although many clustering algorithms have been proposed in the literature for ad- hoc networks, the objective was mainly to generate stable clusters in environments with mobile nodes. Many of such techniques care mostly about node reachable and route stability, without much concern about critical design goals of WSNs such as network longevity and coverage. Recently, a number of clustering algorithms have been specifically designed for WSNs. These proposed clustering techniques widely vary depending on the node deployment and bootstrapping schemes, the pursued network architecture, the characteristics of the CH nodes and the network operation model. A CH may be elected by the sensors in a cluster or pre-assigned by the network designer. A CH may also be just one of the sensors or a node that is richer in resources. The cluster membership may be fixed or variable. CHs may form a second tier network or may just ship the data to interested parties, e.g. a base-station or a command center.

### *The Clustering Problems*

Assume that N nodes are dispersed in a field. Our goal is to identify a set of cluster heads which cover the entire field. Each node $u_i$, where $1 \leq i \leq N$, is then mapped to exactly one cluster $c_j$. where $1 \leq j \leq N_c$, and $N_c$ is the number of clusters ($N_c \leq N$). The node can directly communicate with its cluster head (via a single hop).

The following requirements must be met:

1. Clustering is totally appropriated. Every hub autonomously settles on its choices in view of neighborhood data.
2. Clustering ends inside a settled number of emphasis (paying little mind to organize measurement).
3. At the finish of every Tc. every hub is either a bunch head or a non-head hub (which we allude tn as normal hub) that has a place with precisely one group.
4. Clustering ought to be effective as far as handling intricacy and message trade.
5. Cluster heads are all around conveyed over the sensor field.

### *Clustering Algorithms*

### *Single-Level Clustering Algorithm*

Each sensor in the network becomes a clusterhead (CH) with probability *p* and advertises itself as a cluster head to the sensors within its radio range. We call these clusterheads the volunteer clusterheads. This advertisement is forwarded to all the sensors that are no more than *k* hops away from the clusterhead. Any sensor that receives such advertisements and is not itself a clusterhead joins the cluster of the closest clusterhead. Any sensor that is neither a clusterhead nor has joined any cluster itself becomes a clusterhead; we call these clusterheads the forced clusterheads. Because we have limited the advertisement forwarding to *k* hops, if a sensor does not receive a CH advertisement within time duration *t* (where *t* units is the time required for data to reach the clusterhead from any sensor *k* hops away) it can infer that it is not within *k* hops of any volunteer clusterhead and hence become a forced clusterhead. Moreover, since all the sensors within a cluster are at most *k* hops away from the cluster-head, the clusterhead can transmit the aggregated information to the processing center after every *t* units of time.

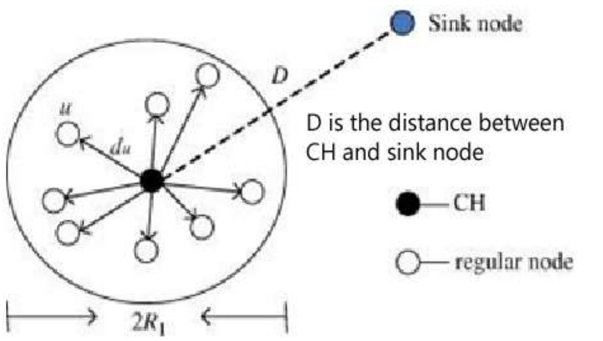


*Figure. 1. A level-1 cluster with radius $R_1$ and number of nodes $N_1$. In the clustering scheme, CH locates in the center of this cluster.*

The limit on the number of hops allows the cluster-heads to schedule their transmissions. Note that this is a distributed algorithm and does not demand clock synchronization between the sensors. The energy used in the network for the information gathered by the sensors to reach the processing center will depend on the parameters *p* and *k* of our algorithm. Since the objective is to organize the sensors in clusters to minimize this energy consumption, we need to find the values of the parameters *p* and *k* of our algorithm that would ensure minimization of energy consumption. We observe[3] that the algorithm leads to significant energy savings. The savings in energy increases as the density of sensors in the network increases.

### *Hierarchical Clustering Algorithm*

In previous section, we have allowed only one level of clustering; this algorithm allows more than one level of clustering. Assume that there are *h* levels in the clustering hierarchy with level 1 being the lowest level and level *h* being the highest. In this clustered environment, the sensors communicate the gathered data to level-1 clusterheads (CHs). The level-1 CHs aggregate this data and communicate the aggregated data or estimates based on the aggregated data to level-2 CHs and so on. Finally, the level-h CHs communicate the aggregated data or estimates based on this aggregated data to the processing center. The cost of communicating the information from the sensors to the processing center is the energy spent by the sensors to communicate the information to level-1 clusterheads (CHs), plus the energy spent by the level-1 CHs to communicate the aggregated information to level-2 CHs, …, plus the energy spent by the level-h CHs to communicate the aggregated information to the information processing center.

## 3. ALGORITHM

The algorithm [3] works in a bottom-up fashion. The algorithm first elects the level-1 clusterheads, then level-2 clusterheads, and so on. The level-1 clusterheads are chosen as follows. Each sensor decides to become a level -1 CH with certain probability$p_1$ and advertises itself as a clusterhead to the sensors within its radio range. This advertisement is forwarded to all the sensors within $k_1$ hops of the advertising CH. Each sensor that receives an advertisement joins the cluster of the closest level-1 CH; the remaining sensors become forced level-1CHs.

Level-1 CHs then elect themselves as level-2 CHs with a certain probability $P_2$ and broadcast their decision of becoming a level-2 CH. This decision is forwarded to all the sensors within $k_2$ hops. The level-1 CHs that receive the advertisements from level-2 CHs joins the cluster of the closest level-2 CH. All other level-1 CHs become forced level-2 CHs. Clusterheads at level 3,4,...H are chosen in similar fashion, with probabilities $P_3$,$P_4$,...$P_h$ respectively, to generate a hierarchy of CHs, in which any level-i CH is also a CH of level (i-1), (i-2),…,1.

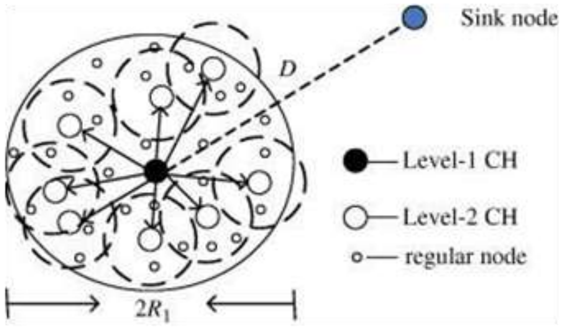


*Figure 2. A level-1 cluster with radius $R_1$. Each level-2 cluster is approximated as the circle-shaped region, and CH locates inthe center of that cluster. D is the distance between level-1 CH and sink node. In this level-1 cluster, the number of level-2 CHs is denoted as $N_{CH2}$.*

The experimental results[3] of this algorithm shows that in networks of sensors with higher communication radius, the distance between a sensor and the processing center in terms of number of hops is smaller than the distance in networks of sensors with lower communication radius and hence there is lesser scope of energy savings. The energy savings with increase in the number of levels in the hierarchy are also observed to be more significant for lower density networks. This can be attributed to the fact that among networks of same number of sensors, the networks with lower density has the sensors distributed over a larger area. Hence, in a lower density network, the average distance between a sensor and the processing center is larger as compared to the distance in a higher density network. This means that there is more scope of reducing the distance traveled by the data from any sensor in a non-clustered network, thereby reducing the overall energy consumption.

Since data from each sensor has to travel at least one hop, the minimum possible energy consumption in a network with n sensors is *n*, assuming each sensor transmits 1 unit of data and the cost of doing so is 1 unit of energy. The density of sensors and their communication radius. Hence, if one chooses to store the numerically computed values of optimal probability in the sensor memory, only a small amount of memory would be needed.

### Low Energy Adaptive Clustering Hierarchy (LEACH)

LEACH [5] minimizes energy dissipation in sensor networks due to constructing clusters. This protocol does not consider node's residual energy in the clustering process.

LEACH operation is done in two phases, setup phase and steady state phase. In the setup phase, a sensor node selects a random number between 0 and 1. If this number is less than the threshold T(n), the node becomes a CH. T(n) is computed as:

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

where r is the current round; p, the desired percentage for becoming CH; and G, is the collection of nodes that in the last 1/p rounds have not been elected as a CH. After electing CHs, every CH announces all sensor nodes in the network that it is the new CH. When each node receives the announcement, it chooses its desired cluster to join based on the signal strength of the announcement from the CHs to it. So, the sensor nodes inform their appropriate CH to join it. Afterwards, the CHs based on a TDMA approach assign the time slot to each node so that a member can send its data to its CH in this period. The sensor nodes can initiate sensing and transmitting data to the CHs during the steady state phase. The CHs also aggregate data received from the nodes in their cluster before sending these data to the BS via a single hopfashion.

### Hybrid Energy Efficient Distributed Clustering [HEED]

Younis and Fahmy [1] proposed an iterative clustering protocol, named HEED. HEED is different from LEACH in the way CHs are elected. Both, electing the CHs and joining to the clusters, are done based on the combination of two parameters. The primary parameter depends on the node"s residual energy. The alternative parameter is the intra cluster "communication cost". Each node computes a communication cost depending on whether variable power levels, applied for intra cluster communication, are permissible or not. If the power level is fixed for all of the nodes, then the communication cost can be proportional to node degree, if load distribution between CHs is required, or 1/node degree, if producing dense clusters is required. In this approach, every regular node elects the least communication cost CH in order to join it. On the other hand, the CHs send the aggregated data to the BS in a multi hop fashion.

HEED periodically selects cluster heads according to a hybrid of their residual energy and secondary parameter, such as node proximity to its neighbors or node degree. HEED does not make any assumptions about the distribution or density of nodes, or about node capabilities, e.g., location-awareness. The clustering process terminates in O(1) iterations, and does not depend on the network topology or use. The protocol incurs low overhead in terms of processing cycles and messages exchanged. It also achieves fairly uniform cluster head distribution across the network. A careful selection of the secondary clustering parameter can balance load among cluster heads. HEED outperforms weight-based clustering protocols in terms of several cluster characteristics. HEED prolongs network lifetime, and the clusters it produces exhibit several appealing characteristics. HEED parameters, such as the minimum selection probability and network operation interval, can be easily tuned to optimize resource usage according to the network density and application requirements. HEED can also be useful in multi-hop networks if the necessary conditions for connectivity (the relation between cluster range and transmission range under a specified density model) hold to the network density and application requirements. HEED can also be useful in multi-hop networks if the necessary conditions for connectivity (the relation between cluster range and transmission range under a specified density model) hold.

HEED (hybrid energy-efficient distributed clustering) is an iterative clustering protocol that uses information about the nodes" remaining energy and their communication costs in order to select the best set of cluster head nodes. During the clustering process, a sensor node can be either a tentative cluster head, a final cluster head, or it can be covered (meaning that it has heard an announcement message from a final cluster head node). At the beginning of the clustering phase, a node with higher remaining energy has a higher probability $CH_{prob}$ of becoming a tentative cluster head. If the node becomes a tentative cluster head, it broadcasts a message to all sensor nodes within its cluster range to announce its new status. All nodes that hear from at least one tentative cluster head choose their cluster head nodes based on the costs of the tentative cluster head nodes. For this purpose, the authors in[7]define the average reachability power(AMRP), which is a cost metric used to "break ties" in the cluster head election process. The AMRP of a node u is defined as the mean of the minimum power levels required by all M nodes within the cluster range to reach the node u.

$$\mathrm{AMRP}(u) = \frac{\sum_{i=1}^{M} \mathrm{Min}Pwr(i)}{M}.$$

During each iteration, a node that is not "covered" by any final cluster head can elect itself to become a new tentative cluster head node based on its probability $CH_{prob}$. Every node then doubles its $CH_{prob}$ and goes to the next step. Once the node's $CH_{prob}$ reaches 1, the node can become a final cluster head, or it can choose its cluster head as the least cost node from the pool of final cluster head neighbors. If the node completes HEED execution without selecting its final cluster head, then it considers itself uncovered and becomes a final cluster head for the upcoming round. Once the clusters are formed, all sensors send their data to the cluster head, where the data are aggregated into a single packet. The cluster head nodes form a network back-bone, so packets are routed from the cluster head nodes to the sink in a multi-hop fashion over the cluster headnodes.

### *Hausdorff Clustering*

It includes three sections. To start with, hubs sort out themselves into a few static bunches by the Hausdorff grouping calculation [7] in light of hub areas, correspondence proficiency, and system network. Second, bunches are shaped just once, and the part of the group head is ideally planned among the bunch individuals. We plan the most extreme lifetime group head planning as a whole number programming issue and propose a covetous calculation for its answer. Third, after group heads are chosen, they frame a spine system to intermittently gather, total, and forward information to the base station utilizing least vitality (cost) directing. This technique can fundamentally protract the system lifetime when contrasted and other known strategies.

With expanding hub thickness, the vitality utilization per hub increments on the grounds that there is more requirement for the neighborhood trade of messages and for the radio channels to contend. WuLi and Hausdorff bunching are believed to be exceptionally vitality productive.

The duration of WAF clustering is much longer than that of the others because WAF requires nodes to sequentially communicate through a ranking order. In addition, WAF requires significant information exchanges between potential gateways and cluster heads for backbone formation. The clustering times for the other three protocols are all modest. It should be noted that Hausdorff clustering is initiated by one node and extended to other nodes (i.e., sequential operation). However, the clustering procedure is carried out only once. Overall, Hausdorff clustering offers a good compromise between the conflicting requirement of smaller number of clusters, energy consumption per node, and average clustering time.

## 4. CONCLUSION

In this paper we have examined the current state of proposed clustering protocols. In wireless sensor networks, the energy limitations of nodes play a crucial role in designing any protocol for implementation. In addition, Quality of Service metrics such as delay, data loss tolerance, and network lifetime expose reliability issues when designing recovery mechanisms for clustering schemes. These important characteristics are often opposed, as one often has a negative impact on the other.

## REFERENCES:

[1] O. Younis, S. Fahmy, “HEED: A Hybrid, Energy-Efficient, Distributed clustering approach for Ad Hoc sensor networks,” IEEE Transactions on Mobile Computing, vol. 3, no. 4, pp. 366–379, 2014.

[2] X. Zhu, L. Shen, and T. P. Yum, “Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks,” IEEE Trans. Vehicular Technology, vol. 58, no. 2,2016.

[3] S. Soro, W. Heinzelman, “Cluster head election techniques for coverage preservation in wireless sensor networks”, Ad Hoc Networks, vol. 7, no. 5, pp. 955–972, July 2015.

[4] J. Hill, R. Szewczyk, A, Woo, S. Hollar, D. Culler, and K. Pister, System Architecture Directions for Networked Sensors, ASPLOS, November 2015.

[5] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, “Energy- Efficient Communication Protocol for Wireless Micro-sensor Networks,” Proc. of the Hawaii International Conference on System Sciences, pp. 1–10, January 2015.

[6] S. Bandyopadhyay and E. J. Coyle, “An energy efficient hierarchical clustering algorithm for wireless sensor networks,” in Proc. IEEEINFOCOM, Apr. 2013, vol. 3, pp. 1713–1723.

[7] X. Zhu, L. Shen, and T. P. Yum, “Hausdorff Clustering and Minimum Energy Routing for Wireless Sensor Networks,” IEEE Trans. Vehicular Technology, vol. 58, no. 2, 2015.