IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

ANALYTICAL STUDY ON CLUSTERING TECHNIQUES AND ALGORITHMS IN WIRELESS SENSOR NETWORKS

[1] Dr. A. Ashok Kumar,[2] S. Karthik
[1]Associate Professor,[2]Research Scholar
[1]Alagappa Government Arts College, Karaikudi, TamilNadu, India
[2]Bharathiar University, Coimbatore,TamilNadu India

Abstract: A new generation of Wireless Sensor Networks (WSNs) has increased a massive attention of researchers on its design challenges and various applications. In the remote environment, sensor nodes are randomly deployed or pre-fixed to transmit the gathered information over a distance through wireless channels. Sensors in the network are limited with energy supply, memory, communication and computational capabilities. Due to energy constraint, nodes communicate within the shorter communication range. Hence, the energy efficiency, throughput, scalability and lifetime of the network are the key issues in WSN. In order to accomplish these issues, the clustering model has been introduced. At present, a cluster-based network is an extremely vibrant research area where nodes are grouped into clusters to minimize the communication overhead, energy depletion, end-to-end delay and ease the path management. In addition, clusters provide a structured way of communication for the unstructured network and make the resources more efficiently. The objective of this paper is to review and analyze the taxonomy of different attributes of clustering. The paper also presents the optimal clustering techniques and the prominent clustering algorithms with several design and performance metrics.

Index Terms - Sensor Nodes, Clustering, Intra and Inter-Cluster, Single and Multi-hop, Cluster Head.

I. INTRODUCTION

In recent advances, Wireless Sensor Network (WSN) enables large degree coverage for many applications [1] such as Military, Environmental, Industry, Forest Fire Detection and so forth. The major issues such as deployment, topology, localization, synchronization, fault-tolerance, communication, routing, scalability, energy dissipation and network lifetime must be considered for the better performance of WSN. In general, WSN consists with minimum requirements of a Base Station (BS) and a collection of Sensor Nodes. Wireless sensors use radio frequency or infrared signals for the communication/transmission between the devices. A low-cost sensor operates at the speed of 4-8 MH_z with 4 KB RAM and 916 $MH_{z\ RF}$. The sensor network may be either a Homogeneous (uniformity) or Heterogeneous (non-uniformity) based on the application wherein, nodes perform a set of high-level tasks such as detection, data processing, scheduling and communication.

Sensors are randomly distributed or pre-assigned in the sensing area to monitor and gather the parameters such as temperature, sound, pressure etc. Sensor nodes are battery powered with a small amount of memory whereas BS has got tremendous energy, storage, computational and communicational resources [1]. Sensors neither are replaced nor recharged in the field, therefore, designing energy efficient algorithms for prolonging the sensors and network lifetime becomes an imperative motivation. In addition, routing is also a major issue in the network due to limited resources, dynamic topology and a high number of node deployments. It is important to manage carefully otherwise it may result in high energy consumption and data redundancy. There are two types of mode are used for the transmission and they are single-hop (direct link) and multi-hop (multiple links). In terms of energy conservation, single-hop may not be possible for a long range of transmission in the network whereas multi-hop may be efficient since it requires less energy at every intermediate node. But, it is important to control the overall communication overhead, path management and propagation delay in multi-hop.

In order to meet the various challenges and design issues of WSN, a standard technique has been introduced called a cluster. The method of grouping the sensor nodes is termed as clustering [6] and it is a hierarchical based most widely used model in WSN which facilitates the different objectives for better performance. The first step in clustering is, cluster generation which has to be created based on the requirements and most of the applications anticipate equal-sized clusters [5] for attaining the load balancing objective. By the efficient cluster formation, it eliminates the data redundancy and provides the uniform energy depletion among the nodes. Each cluster contains a cluster head (CH), member nodes and boundary nodes. Its major roles are defined as in the order that aggregation, data forwarding and communication among CHs. The main problem of clustering is non-uniform distribution [14] of nodes in the clusters which leads to high energy dissipation. The clusters are organized by the CHs and they are acting as coordinators. Therefore, an optimal CH election scheme [2] is needed as it aggregates the data and communicates with BS. Cluster-based network reduces the communication overhead among the nodes.

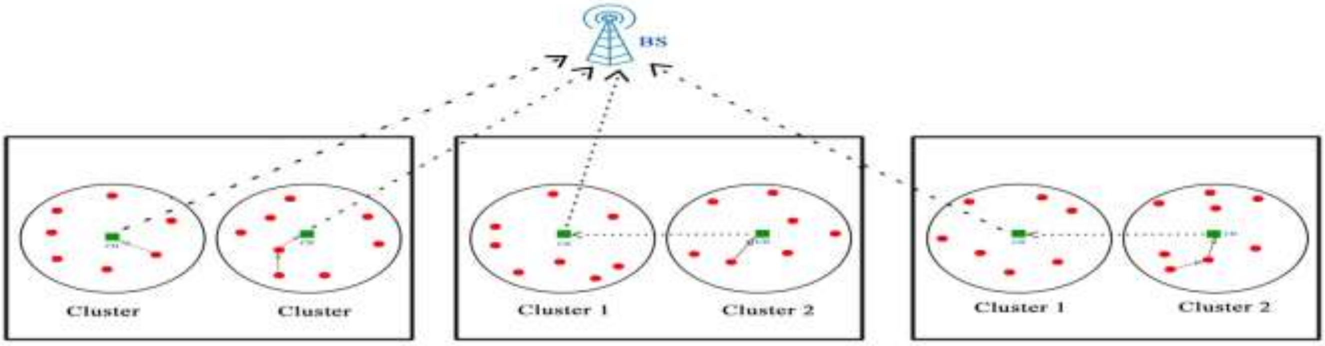

Figure 1. Intra and Inter-Cluster Communication Model.

Clustering is an effective method in data forwarding which imparts a convenient framework for resource management [12]. In the cluster, member node forwards the sensed data to the respective CH either directly or through its neighbor nodes. This process is called intra-cluster communication. CH accepts the data from its member nodes then aggregates it and forwards it to BS through its adjacent CH of another cluster. This process is called inter-cluster communication [6]. Both the communication models are shown in Figure 1. The energy of CHs may be depleted quickly as it performs various tasks than its member nodes. Hence, considering the lifetime extension of the sensors and clusters is provoked by accepting the rotation of CH role [15] within the cluster. Clustered networks are more useful with regard to attaining scalability, energy efficiency, topology stability and network lifetime.

The rest of the paper is organized as follows: Section 2 describes the sensors energy dissipation model. The clustering techniques are comprised of objectives, attributes and approaches in section 3. The eminent clustering algorithms are presented with merits and demerits in Section 4. Finally, in Section 5, the conclusion and challenges during algorithm development are given.

II. SENSORS ENERGY CONSUMPTION MODEL

Sensor tasks such as sensing, transmission/ communication and computations are the source of energy dissipation. As discussed in Section 1, energy is a valuable resource which needs to be utilized efficiently. A radio hardware model [7] is used in many clustering protocols.

There are two models have been used for energy dissipation, one is free space model E_{fs} (d^2 power loss) and another one is multi-path fading model E_{mp} (d^4 power loss). Both models are examined depending on the distance between the transmitter and the receiver. Hence, to transmit a *p*-bit packet at distance d, the energy expended for free space model $E_{Tx\text{-}fs}$ is described by:

$$E_{Tx\text{-}fs}(p,d) = E_{elec} \,.\, p + E_{fs} \,.\, p \,.\, d^2, \;\; \text{if } d < T_r$$

The energy expended for multi-path propagation $E_{Tx\text{-}mp}$ is given by:

$$E_{Tx\text{-}mp}(p,d) = E_{elec} \,.\, p + E_{mp} \,.\, p \,.\, d^4, \;\; \text{if } d >= T_r$$

The energy dissipated for receiving a *p*-bit packet by a node is defined as:

$$E_{Rx}(p) = E_{elec} \,.\, p$$

where, E_{Tx} is required energy utilization for packet transmission, E_{elec} is electronic energy dissipated to send and receive the packet, E_{Rx} is required energy utilization for packet receiving and T_r is the Transmission range for the propagation between two sensors.

The transmission range (T_r) can be defined by equating both models:

$$T_r = \sqrt{E_{fs} / E_{mp}}$$

The distance between the transmitter and the receiver is larger than T_r then, the multi-path model is used otherwise, the free space model is taken to measure the energy dissipation.

III. CLUSTERING TECHNIQUES

The basic idea behind clustering is dividing the network into small units called clusters. It is one of the suitable techniques to minimize energy consumption and maximize the network lifetime. The clustering techniques
are comprised and discussed in various aspects as follows:

A. Clustering Objectives

The clustering objective is built with the consideration of application necessities. The algorithms are proposed based on the objectives only. Some important objectives are highlighted below for the network:

Node Deployment: It deals with the deployment of nodes in the network. There are two types,

- Deterministic: The nodes are set up according to a fixed plan or predefined format. It will not change the position at any cause.
- Randomized: Nodes are placed non-uniformly in the sensing area. It supports fault tolerance but the delay may be a great challenge.

Load Balancing: Balancing [5] the clusters are a quite challenging task as they should limit the number of nodes in the clustered network. It has two important processes: (i) constructing equal-sized clusters i.e. even number of nodes per cluster (ii) uniform energy consumption among nodes. Both of them are mostly achieved by the CH.

Data Fusion and Aggregation: Data fusion is receiving or gathering the data from the sensor nodes whereas, aggregation [12] is eradicating the redundant data for the seamless communication. Both the processes are done only by CH.

Energy Efficiency: Multi-hop routing is the solution to demonstrate the network more energy efficient for a large scale network. The energy calculation for sending and receiving is already discussed in Section 2. The CH node sends the aggregated node to BS via multiple intermediate nodes but, it can send it directly if CH is capable.

Scalability: There are possibilities for appending and removing the nodes if the network is dynamic. Nodes which are deployed should ensure that they are covered in any one of the clusters. The connectivity is an important requirement in many applications. Therefore, the network must be more scalable and it can localize the route setup within the cluster.

Deadlock Prevention: Nodes which are closer to BS may be overloaded when other nodes approaching them to transmit the data. Such nodes energy may be depleted rapidly and to avoid this, load balanced clusters [15] are suggested where it maintains less number of nodes nearer to BS or through some scheduling mechanisms.

Fault Tolerance: In an unattended region, if sensor nodes are distributed it may not be possible to replace whenever they affected or failed due to energy exhaustion etc. Hence, the network must be capable to reconstruct or reconfigure the clusters. To avoid this task, cluster maintenance or overlapping may be feasible approaches.

Network Lifetime: The network lifetime depends on many factors such as energy efficiency, optimal CH and effective routing. It is a crucial task in order to satisfy these factors simultaneously within the network. The network life [15] can be extended only if the clusters are more energy efficient.

B. Clustering Attributes

Clustering scheme aims at achieving some characteristics but it directly or indirectly affects the various performance metrics. Some of the attributes are discussed below with its complexities.

Node Types and Roles: Cluster network consist of two different types of nodes namely primary and secondary. In the clusters, CHs are primary nodes and member nodes are secondary nodes. CHs may act as a sink node or relay node [2] in case of heterogeneous network. CHs major roles are data fusion & aggregation and communication to BS whereas, member nodes role is just forwarding the data and they act as routers.

Cluster Count: It is a variable to determine a total number of CHs required in the network. There is no guarantee to form an equal number of nodes per cluster if CHs are already predefined. However, a network consisting of a large number of clusters may be difficult to control.

Mobility: In most of the existing algorithms, sensor nodes and BS are stationary after the deployment. But in some applications, nodes are moving around (mobility) then the cluster topology [12] dynamically gets changed. Such nodes must be equipped with a GPS antenna to identify their locations; therefore, they are cost-effective.

Cluster Formation: Clusters are created either by CH or BS and it refurbishes a physical network into the virtual network. If clusters are formed by BS then it is a centralized approach where nodes are static while in a distributed approach, CH forms the clusters without nodes coordination and they are dynamic. There are two metrics must be considered before the cluster formation: (i) minimum number of clusters (ii) uniform cluster size.

Cluster Head Election: The performance of the network is mainly based on the CH. CHs may be elected from the set of deployed nodes by following any one of the methods such as probabilistic [11] or random or with specific parameters. In some existing algorithms, CHs are predetermined where it does not perform effectively. The parameters which are used for the CH election in various algorithms are residual energy, node degree and distance to BS.

Cluster Complexity: Clusters are formed for many objectives but those clusters itself have complexities such as computational and communication. Computational states, the number of attempts taken to complete the cluster creation that reflects on the efficiency of the algorithms. Communication complexity [13] specifies the difficulties of routing for data transmission and elimination of data replication.

Intra-Cluster Communication: A communication between node(s) and its relevant CH within the cluster is intra-cluster communication. Generally, it takes only one hop because, most of the nodes may be in transmission range with CH but, for a large scale network, it is established with multiple hops.

Inter-Cluster Communication: Sometimes, a situation arises like CHs may not have long-haul communication capability to communicate to BS. In such a case, CH approaches its adjacent CH for the same task. Both the communication modes are applicable but to increase the scalability of the network, multi-hop inter-cluster communication performs efficiently.

C. Hierarchical based Clustering Approaches

The clustering protocols are segmented into several collections and as a group attains more than one clustering objectives. Most of the clustering algorithms meet the following approaches:

Homogeneous and Heterogeneous Networks: The network construction is based on the characteristics and capabilities of sensor nodes in order to attain the requirements. A network may be composed of homogeneous or heterogeneous nodes. In homogeneous network [7], nodes are alike which means they have similar capabilities in terms of resources and CHs may be elected by following any one of the methods discussed in the CH election section. Moreover, the role of CH is rotated periodically. But, in heterogeneous [8], there are two types of nodes: one is a rich node which has high computational resources and the other one is a common usual node. A node with higher processing capacity among the rich nodes may be elected as CH and the rotation of CH may not be required.

Static and Dynamic: In the cluster, static denotes once the nodes are fixed and clusters are constructed then they never change at any cause but CH alone may be rotated periodically to gain energy efficiency. In dynamic clustering, nodes are replaced and clusters are reformed whenever topology change. This approach is better one compare to static but cost-effective.

Probabilistic and Non-probabilistic Method: As discussed earlier, the performance of the cluster-based network mainly depends upon the CH. Hence, there must be a standard and efficient scheme for CH election. For instance, if CH is elected by considering the node degree alone then the nodes which are not connected with any of the selected CHs, even though if they are in T_r then, by any chance neither become CH nor cover in the clusters. In a probabilistic manner [11], CH is selected based on the probability hence, for every round, CH needs to be changed and all the nodes get chance to become CH. In a non-probabilistic (deterministic) method, CH is elected by considering the parameters such as residual energy, distance to BS and node degree. This method is more reliable, balanced and efficient compared to probabilistic.

Centralized and Distributed: A Centralized algorithm is suitable only for small-scale applications in that CH or BS takes responsibility to construct the clusters whereas, in distributed algorithms, both the cluster formation and CH election are done by the nodes themselves. Once the CHs are elected, they form the clusters. Most of the distributed algorithms are proposed for a heterogeneous environment.

Overlapping Clusters: The process of combining the clusters is overlapping. It gives common nodes called boundary nodes which are helpful for inter-cluster communication and fault tolerance. Moreover, it is helpful in reducing the energy when one CH communicates with another CH and it minimizes the frequent CH re-election process.

Uniform and Non-uniform Clustering: Uniformity states that sensor nodes are uniformly disseminated in the clusters.

In addition, uniform energy dissipation [15] among the nodes is possible only if the clusters sizes are equal and nodes are stationary. However, even distribution of nodes may be a challenging task in the network. In non-uniform clustering [14], load balancing is not possible since cluster size varies thus leads to high energy depletion among the clusters.

IV. CLUSTERING ALGORITHMS

A number of clustering algorithms have been proposed which attains the various objectives and most of them are for energy efficiency. In this section, some eminent algorithms are classified and analyzed with different performance metrics of cluster-based WSNs.

A. Low-Energy Adaptive Clustering Hierarchy: LEACH [3] was the first clustering distributed algorithm for single-hop intra and inter-cluster communication proposed by Heinzelman et. al. The LEACH idea was energy conservation of sensor nodes by CH rotation so that, each node get chance to become CH. LEACH has two phases:

Set up phase: In this phase, the processes CH election and CH formation are performed. CH is elected by following probabilistic method. The probability takes two values: one is, number of times a node has been a CH and the second one is, suggested total number of CHs for a network. Each node chooses a random number '**T**' between the interval 0 and 1. If the random number of a node is less than a threshold value then that node is elected as CH for the current round. The random number of a node is computed as:

$$T(x) = \begin{cases} \dfrac{p}{1 - p\,(r.mod(1/p))} & \text{if } x \in G \\ 0 & otherwise, \end{cases}$$

where, x represents a node, p is the desired percentage of CHs, r denotes the current round number and G is member nodes that have not been selected as CHs in the last 1/p rounds. Once the CH is elected, it will advertise a message to other nodes to join in the cluster.

Steady-state phase: CH uses a TDMA scheduling scheme to give the time slot to the member nodes to transmit the data whereas, member nodes use CSMA to broadcast a message. Finally, CH aggregates the data and sends it to BS directly.

LEACH is not applicable if the CHs are far from the BS. Therefore, many algorithms have been proposed to improve LEACH such as MR-LEACH, TEEN, PEGASIS and many more.

Merits:

- The role of CH is rotated to utilize the energy efficiently.
- LEACH is a distributed algorithm where CH is not predetermined.
- It gives a chance to every node to become a CH.

Demerits:

- LEACH is not suitable for the large-scale network because it uses single hop communication.
- CH is elected by following probabilistic method using a random number rather it should have been considered the important parameters such as residual energy, node degree and distance.
- CH may die early because there is a chance for a node to become a CH with very less energy.

B. Hybrid Energy-Efficient Distributed Clustering: HEED is an adaptive and distributed clustering algorithm for homogeneous network proposed by Younis et al. HEED [10] differs from LEACH in many factors such as residual energy consideration for CH election, multi-hop intra and inter-cluster communication, nodes isolation avoidance and few more. The goal of HEED is like LEACH but, it does not elect the CHs randomly. HEED has three main characteristics. (i) Energy depletion varies for all the nodes (ii) CH election probability can be adjusted to ensure inter-cluster connectivity and (iii) CHs are uniformly well distributed throughout the network. In HEED, cluster construction is performed on a hybrid combination of two parameters i.e. residual energy and intra-cluster communication cost (Node Degree). Nodes having a high residual energy can become CH and this algorithm uses a probabilistic method to elect CHs.

Each sensor node estimates its probability by using the equation $CH_p = C_p * E_{resi} / E_{max}$ to become a CH. C_p expresses the number of nodes to become CHs, E_{resi} describes the residual energy of a node and E_{max} is maximum energy. CH_p should not ahead of the threshold value. The algorithm consists of a constant number of iterations for CH elections. Initially, a node would be a tentative CH, if its probability is less than 1 and a node may become a CH permanently only if CH_p its reaches 1. Once the CHs are finalized then the tentative CH turn into a member node.

Merits:

- HEED offers uniform distribution of CHs.
- It improves the network lifetime and better load balancing.
- Member nodes can directly communicate with its CH.

Demerits:

- HEED selects extra CHs unnecessarily.
- Network efficiency decreases due to a number of repeated iterations.
- Entire knowledge of the network is required to determine intra-cluster communication cost.

C. Distributed Weight-based Energy Efficient Hierarchical Clustering: DWEHC is also a completely distributed algorithm on the whole network proposed by Ping Ding et. al. Each cluster contains the minimum topology, which is locally optimal. The goal of this algorithm is to balance the cluster sizes, energy efficiency and optimize the intra-cluster topology. DWEHC [9] generates the clusters and each node finds its neighbors $N_{(s)}$. After finding the neighbors, sensor nodes calculate its weight because DWEHC sets a parameter weight in CH election. A node that has the largest weight of all its neighbors will become a temporary CH. A node can become a real CH only if a given percentage of its neighbors elect it as their temporary CH and nodes except CHs will join with the clusters as member nodes. Node calculates its weight using the given formula:

$$W_{weight}(s) = \left[\sum_{u \in N_{\alpha,c}(s)} \frac{(R-d)}{6R} \right] * \frac{E_{resi}(s)}{E_{ini}(s)}$$

where, s is a node, R describes the cluster range or radius, u is a neighboring node, d is the distance between s and u, $N_{\alpha,c}(s)$ describes the set of neighbors of a node s and E_{resi} , E_{ini} are residual energies of s. The residual energy is an important input in CH election since CH consumes much more energy to forward all the data to BS. There are two levels in the algorithm with regard to communication. In the first level, the neighbors of CH can directly communicate with its CH and in the second level, nodes which are not neighbors of CH, they can communicate to CH through their neighbor nodes. All CHs consolidate the data packets into one data packet thus reducing overhead. DWEHC provides end-to-end connectivity in inter-cluster communication.

Merits:

- The CH election is based on the parameters i.e. residual energy and node degree.
- DWEHC reduces the energy consumption in both intra and inter-cluster communication.
- The clustering iterations are less.

Demerits:

- DWEHC is not suitable for large-scale network due to single-hop inter-cluster communication.
- Though iteration processes are less, message overheads are high.

D. Distributed Energy-Efficient Clustering: DEEC [8] is a distributed multilevel clustering algorithm for heterogeneous sensor networks proposed by Li Qing et. al. The objective of this algorithm is to reduce energy consumption and increase network lifetime. There are two types of nodes in the network: advanced nodes (high energy) and normal nodes. In the two-level heterogeneous networks, E_0 is the initial energy of normal nodes and m the fraction of the advanced nodes which has 'a' times more energy than normal nodes. The advanced nodes are equipped with an initial energy of $E_0(1+a)$, and normal nodes $(1-m)N$ are equipped with an initial energy of E_0. Therefore, the total initial energy of both the nodes is computed by:

$$E_{total} = N(1-m)\, E_{0+Nm}E_0(1+a)=NE_0(1+am)$$

In the multilevel heterogeneous networks, a node s_i is equipped with an initial energy of $E_0(1+a_i)$, which is a_i times more energy than E_0. The total initial energy of the multi-level heterogeneous networks is given by:

$$E_{total} = \sum_{i=1}^{N} E_0(1+a_i) = E_0 \left[N + \sum_{i=1}^{N} a_i \right]$$

In DEEC, the CHs are elected by a probability based on the ratio between residual energy and the average energy of the network. Let n_i the number of rounds to be a CH for the node s_i based on the residual energy $E_i(r)$ at r round and it is expressed as:

$$n_i = 1/p_i = \overline{E(r)}/p_{opt}\, E_i(r) = n_{opt}(\overline{E}(r)/ E_i(r)$$

where, p_i denotes average probability to be a CH during n_i rounds, $E(r)$describes the average energy at round r of the network and $n_{opt} =1/ p_{opt}$. The nodes with high residual energy take more turns to be the CHs than lower ones. Now, the average energy of the network is computed. Each non-CH sends L bits data to the CH a round. Thus, the total energy depleted in the network during a round is equal to: $E_{round} = L\,(2NE_{elec} + NE_{DA} + k\varepsilon_{mp}d^4_{toBS} + N\varepsilon_{fs}d^2_{toCH})$, where k is the number of clusters, E_{DA} denotes energy depleted for Data Aggregation, d_{toBS} is the average distance between CH and BS and d_{toCH} is the average distance between member nodes and CH. By using those two equations of residual energy and average network energy, the optimum CH is elected.

Merits:

- It does not require any global knowledge of energy at every election round.
- DEEC can perform well in multilevel heterogeneous networks.
- It uses the average energy of the network to control the energy expenditure of nodes.

Demerits:

- The stability of the clusters is not at the expected level.
- CHs energy will deplete quickly as CHs are directly connected to the BS.

E. Energy-Efficient Unequal Clustering: EEUC is a multi-hop distributed clustering algorithm to balance the energy dissipation proposed by Chenga Li et. al [7]. In general, if CHs are situated nearer to BS then CHs will be overburdened with heavy traffic and tend to die early. To avoid this, EEUC forms the unequal clusters and have smaller cluster sizes thus nodes will consume less energy during the data forwarding. In EEUC, CH election is primarily based on the residual energy only. Every node becomes a tentative CH with the probability. Suppose, a node s_i becomes a tentative CH then it has a competition range R_{comp} which is a function of its distance to the BS. CHs are selected by localized competition also. It is expressed as:

$$s_i.R_{comp} = \left[1 - c\,\frac{d_{max} - d\,(s_i,BS)}{d_{max} - d_{min}}\right] R^0{}_{comp}$$

where, c is a constant coefficient between 0 & 1, d_{max} , d_{min} denotes the maximum, minimum distance between the node and BS, $d\,(s_i\,,BS)$ is the distance between s_i & BS and $R^0{}_{comp}$ is the maximum competition radius. The cluster size is proportional to the distance to BS. In this algorithm, the data transmission for intra-cluster is identical with LEACH. But, in inter-cluster multi-hop routing, a CH chooses a relay CH based on the residual energy and its distance to the BS. The simulation result shows that the unequal clustering mechanism really extends the network lifetime. Table 1 compares the discussed algorithms with its attributes.

Merits:

- Improves the network lifetime compare to LEACH and HEED.
- Avoids the long distance communication between CH and BS.

Demerits:

- Message complexity is high.
- Communication is not effective in both intra and inter-cluster.

Algorithm	Homogeneous / Heterogeneous	Centralized / Distributed	CH Selection Scheme	CH - Uniform Distribution	Cluster Count	Intra-Cluster Topology	Inter-Cluster Communication	Energy Efficient	Balanced Clustering	Cluster Stability
LEACH [3]	Homogeneous	Distributed	Random	No	Variable	1-hop	Direct Link	No	Average	Average
HEED [10]	Homogeneous	Distributed	Random	Yes	Variable	1-hop	Direct Link / Multi -hop	Yes	Good	High
DWEHC [9]	Homogeneous	Distributed	Random	Yes	Variable	Multi - Level	Direct Link	Yes	Very Good	High
DEEC [8]	Heterogeneous	Distributed	Residual Energy	Yes	Variable	1-hop	Direct Link	Yes	Good	Average
EEUC [7]	Homogeneous	Distributed	Residual Energy	Yes	Variable	1-hop	Multi-hop	Yes	Good	Average

Table 1. Comparison of Clustering Algorithms

V. CONCLUSIONS

In this paper, clustering techniques & algorithms along with its attributes, approaches and performance metrics are analyzed. The network lifetime and scalability objectives for a large scale WSN are achieved only by the energy efficiency metric. Moreover, it is well known that the overall performance of a cluster-based network is mainly based on the Cluster Head (CH). Therefore, the optimum CH must be chosen and the paper suggests that parameters based scheme would be the finest option than a random method for CH election. In case of a homogeneous network, the role of CH must be rotated among the sensor nodes to attain load balancing. In the clustering algorithms, energy depletion is better in intra-cluster communication as compare to multi-hop inter-cluster communication. This paper also helps to observe the vital challenges during the algorithm development such as efficient utilization of resources, equal sized clusters for uniform energy dissipation, the optimal number of clusters and CHs, efficient multi-hop routing for communication/transmission and so forth. Finally, it is concluded that though clustering is a standard technique, clustering overheads and communication complexities are still the challenging tasks in cluster-based WSN.

REFERENCES

[1] I.F.Akyildiz, W.Su, Y.Sankarasubramaniam and E.Cayirci, “A survey on sensor networks”, IEEE Communications Magazine, vol. 40, no. 8, pp. 102105, 2002.

[2] D.Jia, H.Zhu, S.Zou, and P.Hu,“Dynamic cluster head selection method for wireless sensor network,” *IEEE Sensors Journal*, vol. 16, no. 8, pp. 2746–2754, 2016.

[3] W.R.Heinzelman, A.Chandrakasan, and H.Balakrishnan, “Energy-efficient communication protocol for wireless microsensor networks,” in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS ’00)*, vol. 2, p. 10, January 2000.

[4] Y.Q.J.Yu, G.Wang, Q.Guo, and X.Gu, “An energy-aware distributed unequal clustering protocol for wireless sensor networks,” *International Journal of Distributed Sensor Networks*, vol. 2014, p. 8, 2014.

[5] R.N.Enam, R.Qureshi and S.Misbahuddin, “A Uniform Clustering Mechanism for Wireless Sensor Networks”, International Journal of Distributed Sensor Networks, vol. 2014, (2014), pp. 14.

[6] M.M.Afsar and M.H.Tayarani-N, “Clustering in sensor networks: a literature survey,” *Journal of Network and Computer Applications*, vol. 46, pp. 198– 226, 2014.

[7] C.Li, M.Ye, G.Chen, J.Wu, “An energy-efficient unequal clustering mechanism for wireless sensor networks,” in Proceedings of 2005 IEEE International Conference on Mobile Ad-hoc and Sensor Systems Conference(MASS05), Washington, D.C., pp. 604-611, November 2005.

[8] Li Qing , Qingxin Zhu and Mingwen Wang, "Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks", Computer Communications, Volume 29, Issue 12, August 2006, 2230-2237.

[9] P.Ding, J.Holliday, A.Celik, “Distributed energy efficient hierarchical clustering for wireless sensor networks”, in Proceedings of the IEEE International Conference on Distributed Computing in Sensor Systems(DCOSS’05), Marina Del Rey, CA, June 2005.

[10] 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.

[11] H.Huang, J.Wu, “A probabilistic clustering algorithm in wireless sensor networks”, in Proceedings of IEEE 62nd Semiannual Vehicular Technology Conference (VTC), Dallas, TX September 2005.

[12] X.Liu, “A survey on clustering routing protocols in wireless sensor networks,” *Sensors*, vol. 12, no. 8, pp. 11113–11153, 2012.

[13] A.A.Abbasi and M.Younis, “A survey on clustering algorithms for wireless sensor networks,” *Computer Communications*, vol.30, no.14-15, pp. 2826–2841, 2007.

[14] G.Chen, C.Li, M.Ye, and J.Wu, “An unequal cluster-based routing protocol in wireless sensor networks,” *Wireless Networks*, vol.15, no.2, pp. 193–207, 2009.

[15] Y.Liao, H.Qi, and W.Li, “Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks,” *IEEE Sensors Journal*, vol.13, no.5, pp. 1498–1506, 2013.