Novel Clustering Techniques in Wireless Sensor Networks – A Survey

Review Paper

T.C. Swetha Priya
Department of Computer Science and Engineering,
JNTUH University College of Engineering, Science and Technology,
Hyderabad, Telangana, India
tcswetha3552@gmail.com

R. Sridevi
Department of Computer Science and Engineering,
JNTUH University College of Engineering, Science and Technology,
Hyderabad, Telangana, India
sridevirangu@jntuh.ac.in

Abstract *– A study of Wireless Sensor Networks has been growing tremendously these days. Wireless Sensor Networks play a major role in various fields ranging from smart homes to health care. WSN's operate independently in remote places. Because of tiny size of the nodes in such type of networks, they have a limited number of resources in terms of energy and power. Basically, sensor networks can be classified into flat and cluster based Wireless Sensor Networks. But, Clustering based Sensor Networks play a major role in reducing the energy consumption in Wireless Sensor Networks. Clustering also focuses on solving the No.s that arise during transmission of data. Clustering will group nodes into clusters and elects Cluster Heads for all clusters in the network. Then the nodes sense data and send that data to cluster head where the aggregation of data will take place. This paper focuses on various novel clustering techniques that improve the network's lifetime.*

Keywords*: Wireless Sensor Network, Clustering, Fuzzy clustering, Optimal Cluster Size, Adaptive Clustering, Quad Clustering, K-Means, Dragonfly Algorithm*

1. INTRODUCTION

Due to recent advancements in technology, Wireless Sensor Networks have a great demand in the market. Their versatility has increased the demand for this technology. A Wireless Sensor network is made up of a huge number of less cost sensor nodes which are of small size and limited power [1]. These nodes have the capability to communicate over shorter distances. They have limited resources in terms of storage, bandwidth, energy and computational abilities. When a node fails or becomes inactive there is no chance for replacement of that node because of their constrained amount of energy. Because the sensors are operated with battery, the lifetime of a network and depletion of energy are the major challenges in this type of networks.

Basically, the sensor networks are classified into flat Wireless Sensor Networks and cluster based Wireless Sensor Networks. The flat Wireless Sensor Networks are traditional networks in which all the nodes will take the role to generate data and route data. But this type of network is not suitable for large-sized networks because the nodes will flood the data into the network. This will generate duplicate data causing a lot of overhead during communication of data. So, the traditional flat WSN works well in small networks. The sensor nodes and base station will be randomly placed in the sensor network area. The main responsibility of these nodes is sensing the data from the environmental conditions of a particular application and forwarding that data to Base Station. This Base Station may be in a place that is distant to the nodes. So, the data that is received from nodes may be repeated or redundant leading to energy depletion. To overcome this, clustering technique is introduced in Wireless Sensor Networks. The Clustering method guarantees increase in the network lifetime. Clustering is one of the best methods for large sensor networks. Clustering is partitioning the network into sub-networks [2]. It helps to manage the network easily and it will be easy to extend the network. Each cluster consists of a Cluster-Head, sensor nodes and a Base Station. The cluster heads are selected through any of the various clustering algorithms based on several factors such as residual node energy, distance etc. The sensor nodes will gather data and send to the cluster head. The cluster heads act as a local aggrega-

tor and sends this data to Base Station. This data can be sent in either a single-hop or multiple-hop manner [3]. The communication in cluster based networks can be of two types: Inter cluster communication and Intra cluster communication. The features that make the cluster based Wireless Sensor Network an effective choice are data fusion, load balancing, energy efficiency, maintenance of cluster, secure transmission of data and better network lifetime.

This paper provides a comprehensive overview of various novel approaches to perform clustering. This paper provides a study of the cluster-based techniques in Wireless Sensor Networks. This section deals with the introduction of Wireless Sensor Networks and the significance of Clustering in Wireless Sensor Networks. The rest of the paper deals with providing a brief overview about Clustering schemes and comparison of clustering methods. The conclusion is given in the last section of this paper.

2. CLUSTERING TECHNIQUES

2.1. LOW ENERGY ADAPTIVE CLUSTERING

In this type of clustering [1, 2], the cluster heads are selected in such a way that each node present in the sensor network gets an opportunity to be a cluster head. The cluster heads consume more energy than that of cluster members. So, we need to distribute the load among all sensor nodes so that a node's energy do not get depleted after some amount of time because of frequent selection of nodes as Cluster Head. This clustering involves 2 phases: 1. Setup and 2. Steady phase. In the first phase, formation of clusters and selection of Cluster heads based on the number of times that a particular node has been selected as cluster head is done. If the probability of a node is less than the selected threshold value then that node is selected as a Cluster Head. In the second phase, the data that is gathered by Cluster heads will be sent to the Base Station. One disadvantage with this method is the node's initial energy will not be considered during the cluster head selection. Also, the nodes which already became Cluster heads are again chosen as Cluster heads causing depletion of energy of such nodes. This type of clustering will not be appropriate for larger networks.

2.2. EVENT TO SINK DIRECTED CLUSTERING

This type of clustering [4] will provide high efficiency by reducing consumption of energy. When an event is discovered by a node, a report is sent to the sink node. This data which is collected by the nodes is sent to cluster head to avoid data redundancy. In this type of clustering, unnecessary cluster formation is avoided by allowing formation of clusters only on the occurrence of an event. The movement of data is minimized within a cluster because the formation of clusters is in the direction where an event occurs. The selection of Cluster Heads will be done from upstream whereas non-cluster sensor nodes will be selected from the nodes that are downstream. The data flow will be unidirectional. The node delay in LEACH is almost double that of Event to sink directed clustering.

2.3. LOAD BALANCED CLUSTERING

In this type of clustering [5], an assistant node is chosen from all the nodes and this node helps cluster head in performing the aggregation and processing of data. This selected node helps in the transmission of data to Base Station. The received data is processed by Cluster head and transmits it to the selected assistant node. This assistant node forwards this data to Base Station. But the drawback with this type of clustering is that the flow of data among sensor nodes is not constant. So, the nodes that are nearer to Base station will receive more data compared to nodes that are far away. So there is a chance for depletion of energy of closest nodes.

2.4. ENERGY EFFICIENT HIERARCHICAL CLUSTERING

In this type of clustering [6, 7], each sensor node has the ability to decide upon whether it can be a Cluster head (CH) or not. If a node chooses to be a cluster head, then it announces its presence to all adjacent nodes. Now, this Cluster head is known as Volunteer Cluster Head. All sensor nodes which are not present within k-hop distance away from Cluster Head will retrieve the data from CH. If this advertisement message is received by any of the nodes other than the Cluster Head then those nodes will become the members of Cluster.

2.5. WEIGHT BASED CLUSTERING

In this type of clustering [8], the parameter for the selection of Cluster Head is "Weight". This can be computed based on several aspects such as the number of times that a particular node has been selected as cluster head, distance between the cluster heads and nodes, remaining node energy. The node weight is computed at each periodic interval of Clustering process. Here the cluster formation will be done in a manner in which less energy is consumed in a network. The node with highest remaining energy is chosen to be a Cluster Head. This means that the nodes with less energy will be excluded from being Cluster Heads. If any node has energy greater than that of remaining energy then that node will become a Cluster Head whereas the other nodes will be cluster members. Here the threshold value is set based on remaining energy. So, if any nodes' energy falls below this threshold value, then that node will be treated as a "dead node". Because of this, each and every node present in the network will send a message about its presence after periodic intervals of Clustering process. This method is best in comparison with other clustering methods because there are a smaller number of dead nodes.

2.6. CLUSTER HEAD ROTATION SCHEME

In this method [9, 10], there is a chance for a sensor node to go to inactive mode for an arbitrary amount of time. Then the node announces itself as a Cluster Head (CH) and transmits this announcement message to all sensor nodes present within the scope of the network. Then the adjacent nodes that are within the reach sends their responses about the selection of cluster head and based on their opinion the Cluster head will decide whether to remain as a Cluster Head or to change its role as a member of cluster. Here cluster head nodes will be rotated by giving an opportunity to the next nodes based on previous information collected at an initial point of time. The information that is collected initially includes a node identifier; its remaining energy and number of times it got a chance to become CH [11]. This information is collected in the form of packets. In this manner, Cluster Heads are rotated in a timely manner.

2.7. SELF-ORGANIZED ENERGY CONSCIOUS CLUSTERING (SECC)

This type of clustering has 2 steps [12, 13, 14]. In the first step, node energy of all nodes and neighbor node distance is computed. Here, the average distance of nodes is compared and the node within same distance gets combined and forms a cluster. Here, the clusters are formed in an energy-conscious manner based on the Threshold value. Whereas the nodes which have less value than that of Threshold goes to inactive state and further they do not involve in any of the sensor related activities.

2.8. ENERGY AWARE FUZZY CLUSTERING ALGORITHM (EAFCA)

In this algorithm [13-20] [21], first the nodes are deployed and then the computation of distance between the sensors is done. In this method, a signal is sent from sink node to all remaining nodes. Now, by considering the strength of the signal that is received, the distance from sink to sensor nodes is computed. Then, some cluster heads are picked from the entire network. Then, the threshold value is calculated and this value has to be transmitted to all nodes present in the network. Now, the sensor nodes will generate an arbitrary number and this is compared with the threshold value that it has received. The node with highest random generated number than threshold will declare itself as a Cluster Head (CH) whereas the remaining nodes will be normal sensor nodes. This method adapts creation of 2-hop clusters. It considers the factors like remaining node energy, degree of a node at 2-hop distance and CH position also employs a fuzzy logic for selection of a node as a cluster head permanently [22]. The node degree plays a major role in the selection of permanent cluster head. It means that a node with higher node degree becomes a permanent Cluster Head. This indicates that a node will be either a Cluster node member or a Cluster Head. So, the Cluster heads will make periodic announcements to the nodes to join. So, the nodes will decide to join the clusters that are nearer to them. If the distance to the two clusters is same, then the nodes from which the announcement to join is received first will be considered and the node will join in that cluster. After formation of Clusters, the information will be gathered by sensor nodes and will be forwarded to the Cluster heads. These cluster heads are responsible for aggregation of data and sending the information to sink.

2.9. ENERGY EFFICIENT RECURSIVE CLUSTERING (EERC)

This type of clustering [7, 23] is an event-oriented clustering approach. The Clusters are formed only upon the event occurrence. Here, the clusters are generated in a recursive manner and the clusters are formed only for reducing energy consumption while transmitting data. This method follows 2 steps in formation of clusters after deployment of nodes. In the first step, the Euclidean distance between two nodes is calculated. Then, based on that value clusters are formed resulting in the first level of clusters. Because this method follows recursive clustering, the first level of clusters is again partitioned to form the second level of clusters after considering the distance between nodes. After the formation of Clusters, the cluster head selection is done based on Round Robin Scheduling. In this Scheduling, the turn-around time is computed and a node with minimum value and highest energy will be selected as a Cluster Head. Now this computation of turn-around time is done after every 2 rounds. Then the node with minimum value is selected as a Cluster Head. Now the data is gathered and is sent to Head node where the aggregation of data will be done and finally data is sent to Base station by multi hop routing. This method reduces the consumption of energy and extra overhead of redundant data by aggregation.

2.10. ADAPTIVE DISTRIBUTED CLUSTERING ALGORITHM (ADCA)

In this type of clustering [24] [25], Similarity value is measured. The sensor nodes that are adjacent to each other may have a chance of producing same data. So, the Similarity will be high for such sensor nodes. Also, there is a chance that the data generated at consecutive time intervals from the same sensor nodes will be similar. The data redundancy and generation of data can also be limited through this approach helping in reducing the energy. This method follows two steps: 1. Formation of Cluster and 2. Adaptive sleep duty cycle. In the first phase, the sink node will analyze the rate at which the information is generated and Similarity of information from sensor nodes. Based on the analyzed information, the nodes will form groups leading to Clusters. The clusters that are formed will not be of same size. Now, based on the remaining node energy and connectivity of nodes the Cluster Heads will be chosen. In the second phase, a

comparison is made between the rate at which the data is generated by member nodes and the minimum level of threshold. Then the nodes with a low threshold will be assigned a duty cycle for a specific amount of time period. This sleep or awake time period has to be informed to all sensor nodes. A fair distributed scheduling is done to limit the consumption of energy. The Cluster heads gather data from the nodes and checks if there is any data that is similar. If any deviation in the data is identified then this has to be reported to sink node. Then the sink will decide whether to change sleep-awake duty cycle or whether to perform re clustering process based on the necessity. In this way some part of the periodic monitoring work will be shared by Cluster heads instead of depleting the energy of sink node alone.

2.11. IMPROVED ENERGY AWARE DISTRIBUTED UNEQUAL CLUSTERING

In this type of clustering, the random deployment of nodes is done and the nodes will have varying energies. [24, 26, 27]. The Base Station is positioned far from the Sensor Network area. Then after deployment of nodes, a signal is broadcasted to all nodes. Based on the strength of signal received the approximate distance from nodes to the Base Station is estimated. This method works in the form of rounds. A signal will be broadcasted to all the nodes by the Base Station after the deployment of nodes. Now based on the strength of the signal received, the nodes will calculate the distance. Each round involves 2 phases: 1. Setup and 2. Steady phase. The first phase comprises of 3 steps: 1. Neighborhood information Collection, 2. Cluster Head Selection and 3. Formation of Cluster head. After the formation of clusters, the second stage starts where the data transmission take place. In the steady state, there will be a number of major slots. After completion of one major slot, the rotation of cluster heads will take place. This version of extended EADUC helps in reduction of Cluster overheads and in balancing consumption of energy.

2.12. HIERARCHICAL CLUSTERING

This type of clustering [24, 28] works in 2 layers. The selection of Cluster heads will be done in the first layer and data routing is done in the second layer. Here the selection of cluster head is of major concern as it aims at reducing the consumption of energy and delay [29]. This method considers 3 parameters: energy remaining in nodes, distance from the Base Station and node degree. The weight of each node will be computed based on these 3 factors. Finally, the Cluster head will be the node which has maximum weight.

2.13. TWO LEVEL HIERARCHY FOR LOW ENERGY ADAPTIVE CLUSTERING HIERARCHY

In Hierarchical clustering [25], unbalanced Clustering is employed. But if there is no balancing in clustering, there is a chance for more consumption of energy during data collection by some of the cluster Heads. So to overcome this the clusters must be balanced by employing various levels. In this method, all the levels are connected that is first level Cluster Heads are connected to second level cluster heads which in turn gets connected to all the member nodes of the Cluster. Here the first level will summarize the data and second level will aggregate data whatever is received from the first level which will reduce the packets sent over the network and also reduces depletion of energy level of nodes.

2.14. MOBILE SINK-BASED APPROACH

Generally, the nodes are assumed to be static and are deployed randomly. But the sink node is always mobile. So, the burden will be on one mobile sink node [30]. To overcome this, in this method two mobile sink nodes will be deployed [31]. After completion of every half round these sink nodes move in anti-clockwise direction. The sensor network area is partitioned into equal areas and each area consists of unequal clusters [32]. At the beginning, the center node is chosen as Cluster Head. After completion of each round, the sensor node which is closest to the center of sensor network area and having highest remaining energy will be chosen as a cluster head. As the sink nodes are mobile, a routing mechanism is also required for transmission of data to the current location of sink. The new path information will be maintained by only a single Cluster head [33].

2.15. CLUSTERING USING FUZZY LOGIC

In Fuzzy Logic [13-19, 34], the 2 variables that are taken as input are: Remaining Energy and Required energy. This type of clustering works in 2 stages. In the first step, based on the strength of signal received every node will compute the energy that is needed to send 'n' number of bits to the Base Station. Here, a Timer is set by all nodes that is conversely proportional to an output variable "chance". If the value of "chance" is higher than it has more possibility for being a Cluster Head. A countdown timer is set by all the nodes. If this value becomes zero, then that node will become a CH. Then the distance is computed and based on the distance the formation of clusters will happen. In the second step, Time Division Multiple Access scheduling is used by Cluster Head and the cluster members will also be informed about this schedule.

2.16. CLUSTERING USING NODE RELEVANCE

This clustering [35] focuses on calculation of adjacency matrix with binary values specifying only the nodes that are adjacent to each other. This method is appropriate for nodes that are stationary because this reduces frequent announcements from nodes about their occurrence to Cluster Heads. Then the selection of Cluster Heads will be done in several levels based on factors such as number of times a node has acted as Cluster Head, association of nodes and node advertise-

ments. This must be repeated before every round. This method selects Cluster Heads for several rounds at a time by using Relevance index (R-Index). By examining the structure of the network, the nodes' R-index will be computed from adjacency matrix. The distance will be calculated between Cluster heads. The minimum distance and the adjacency matrix will be chosen to perform clustering. In this manner, the Cluster heads for multiple rounds will be computed at once and if suddenly a node fails, then adjacency matrix will be recomputed and the same process is repeated.

2.17. DISTRIBUTED FUZZY LOGIC BASED ENERGY AWARE AND COVERAGE PRESERVING UNEQUAL CLUSTERING (DECUC)

In this type of clustering [18, 35-36], a distributed method is used in the formation of clusters. The selection of clusters will be done by sensor nodes but not the sink node. This method employs 3 algorithms. In the first algorithm, the fuzzy logic inputs that are used are distance between nodes to sink, Energy level and Area coverage. Now, the degree of Fuzzy output is determined and the node with high chance or output value will be chosen as center of Cluster. Now, sensor nodes will transmit data to this cluster center where the data aggregation is done and the data that is aggregated will be sent to Sink. In the next round, the cluster centers will be selected again. In the second algorithm, election will not be there. Suppose if any Cluster Center exits the network before this round, then one of the nodes with highest degree from the previous round will be selected as new Cluster center. Here, the nodes may be operating in different algorithms in multiple rounds. The nodes that have not achieved the threshold level of energy in the second algorithm will still be in the same cluster whereas the nodes which have reached threshold will proceed to the second algorithm. Each cluster head must calculate a different threshold value. In some rounds, the value of the threshold should be reduced to half to reduce unwanted elections. The high threshold will be generated by the cluster heads with less value of chance. These nodes will be allowed to enter into the third algorithm. This clustering follows an adaptive threshold to make it significant among other methods. Now, the nodes that are in the third algorithm will have to reselect the cluster head. As the Cluster head selected during the first round will be the best, the next cluster head that is chosen must be in a good condition with respect to the energy and must be at a distance nearer to cluster center. Now the inputs that are given to the third algorithm of fuzzy logic method will be the remaining energy and distance to cluster center. Now after defuzzification, the output of inference engine is taken and sensor node must compare this value with adjacent nodes' value. The one with highest chance is chosen as cluster center. Now after these subsequent selections, the nodes in third algorithm will move to first and the sensors in second algorithm will move to third if the required threshold is reached. Otherwise they remain in same algorithm. Here, the selection of cluster centers will be reduced in different levels and also it controls the information passing in the network thereby reducing the energy.

2.18. OPTIMAL CLUSTER SIZE SELECTION-BASED CLUSTERING PROTOCOL USING FUZZY LOGIC (OCSSP)

In the clustering methods that involve fuzzy logic [37] the determination of output variables for the selection of Cluster Head involve multiple input variables. A different combination of input variables will generate different output variables called as "chance" which are used in the cluster head selection. So, a new clustering algorithm that involves multiple input and multiple output variables will be used to increase the lifetime of a network. To achieve better performance of the network, selection of inference rules will play a major role. After completion of several rounds, the rules will be decided. A two-fold restriction is imposed on the cluster size to reduce the load on the entire network. This clustering computes output variable values based on varying factors of nodes. Here, the cluster head selection will be done in first step and data transmission is done in second step. The cluster members will sense information from the environment and the cluster heads will in turn collect data from these member nodes. The Fuzzy inference system will be used in this type of clustering. This method involves several rounds. Each round has 2 steps: 1. Setup and 2. Steady phase. In the first phase, the Base Station implements fuzzy inference system for the selection of cluster head. Each Cluster head will provide space for a restricted number of members. The fuzzy Inference system is used for efficient supervision of cluster members onto which cluster head is appropriate to join. In a particular round, any node that has highest value of chance is selected as Cluster Head. Then, in the second step through this cluster head the transmission of data will be done to the Base Station.

2.19. QUAD CLUSTERING BASED ON K-MEANS ALGORITHM

This type of clustering [38] adapts K-Means clustering algorithm. The K-Means clustering is one of the unsupervised Clustering techniques. In sensor networks, this technique is employed for the cluster formation in a particular area. The whole network area is divided into 'n' clusters. The clusters have to be formed by iteratively computing the Euclidean distance between the selected data points. Now, the data points will be assigned to nearest cluster center. After computing the mean of the selected data points of corresponding clusters then a new cluster center has to be found. If there is a change in the assignment of clusters, then only this process has to be repeated for all the newly formed clusters. This clustering method involves three steps. In the first step, by using K-Means algorithm the clusters will be formed. The re clustering process must include deter-

mining the cluster's centroid position and inter-node Euclidean distances. Then the nodes' with small Euclidean distance and high residual energy will be chosen as Cluster head. In the second step, the entire network is partitioned into 4 parts. This partition is to improve energy efficiency and network coverage. Now, each partition of the network will be working in isolation. In the third step, each quadrant has a separate cluster Head. Now, again K-Means clustering algorithm must be implemented on each quadrant to further partition this cluster into 4 parts. Hence it is known as Quad Clustering. Before the final cluster head in a quad cluster is selected, the cluster head energy has to be compared with selected threshold. When compared to a full cluster, these quad clusters will utilize half of the total energy by reducing the energy consumption of network.

2.20. DRAGONFLY ALGORITHM FOR HIERARCHICAL CLUSTERING

The dragonfly based clustering [39, 40] comes under biologically-inspired method because it makes use of the way the dragonflies moves. This helps to achieve good optimization in terms of finding the better cluster head. This algorithm calculates the number of nodes that are alive and the lifetime of a network. The nodes are assumed to be present at arbitrary positions and then the clusters are formed using LEACH method. At the beginning, for the selection of cluster head the nodes with highest energy remaining will be considered. Further, the dragon fly algorithm will select cluster head. In this method, first energy and distance variables have to be initialized to locate CH node. Now the fitness function for all the Cluster heads has to be computed and the Cluster head with a least value will be selected as Cluster Head. This data has to be sent to all members of cluster heads. Then the data gathered by member nodes will be aggregated and sent to Base Station. The dragon flies form sub groups and move in multiple directions. They update their location based on factors such as structure, alignment, similarity and disturbances. To indicate the movement of dragon flies, a step vector has to be calculated. With the help of random walk concept, they determine the neighbors and solution to step vector is found. Then by utilizing the nearest distance and level of energies, a fitness value is calculated. The nodes with least fitness value will be chosen as Cluster Head. In this method, the best Cluster heads are the ones that are closer to the adjacent Cluster heads and the Base station because this will save energy.

2.21. ENERGY EFFICIENT SLEEP AWAKE AWARE NODE SCHEDULE CLUSTERING

This type of clustering [41-43] makes use of node scheduling concept. Here, the nodes will be in either sleep or active mode at the time of transmission of data among nodes. This method implements the concept of pairing to improve the lifetime of a network. The main reason for this method is to overcome the redundant data. When pairing of nodes is done then the nodes may sometimes sense similar data and forward that to the Cluster Head. Because of coupling, only one node will be active at a time by switching off another node temporarily during communication. It works on uniform network. This works well when the sensor nodes are static and do not require any transmission of data between them. This reduces the redundant transmission of data and increases the lifetime of sensor network. Here, first the data that is retrieved from various sources have to be segregated and a common pattern must be obtained and the result has to be displayed.

2.22. NODE OVERHAUL SCHEME FOR ENERGY EFFICIENT CLUSTERING

In LEACH uniform size clustering approach, the selection of cluster head will depend on probability. At the beginning of formation of clusters, the nodes will join with the cluster heads that are nearby. Because of applying the probability approach in this method, there is a chance of clusters being in various sizes. All the nodes will be having a cluster head. Based on the second-best choice, re clustering is performed for obtaining uniform clusters by moving some of the nodes from a larger cluster to other clusters [44]. To perform this operation, first the largest cluster among all the clusters must be found and then the distance must be computed between the node members of cluster and remaining cluster heads. From this the second-best choice of cluster head is obtained. The nodes having less distance to remaining cluster heads will be assigned to second best cluster heads whereas nodes closer to boundary of clusters will be assigned to remaining cluster heads. This process has to be repeated until equal sized clusters are formed. Then, a Time Division Multiple Access schedule has to be created by all cluster heads present in the sensor network. Then the data has to be transmitted in the assigned time schedules. During transmission of data, aggregation of data has to be done by Cluster Heads and then it is forwarded to the Base station. After every round, re clustering is performed to turn the roles of cluster heads. This method helps in balancing the network load by creation of uniform clusters.

2.23. DISTANCE AND ENERGY CONSTRAINED K-MEANS CLUSTERING SCHEME (DEKCS)

This type of clustering [45-47] aims at improving the performance of the network and selection of cluster head. This method uses k-Means Clustering to perform clustering. This method chooses the proximity principle that means the nearest node that is closest to the greatest number of nodes in the network. This makes sure that all the nodes in every cluster must be nearer to the cluster head so that less amount of energy will be consumed during transmission of data. The K-Means algorithm [47] divides the unlabeled multi-dimensional data into k number of clusters. The number of clusters has to be determined before itself. This method tries to reduce the distance between the nodes within a cluster and the

cluster center thereby increasing the distance between clusters. Here, the clusters that are formed will be in such way that all nodes are 1-hop away to cluster Head. For the number of clusters to be optimal, an elbow method is used. This method computes intra cluster sum of squares for the multidimensional data. After clustering, cluster head must be selected. During the selection of cluster head, the position of cluster has to be considered because the proximity principle is used in this clustering. The cost function is used to compute the Euclidean distance between a node that is selected and remaining cluster node members. From this, the nearest node that is closest to the greatest number of nodes in the network will be chosen as Cluster Head. The chosen cluster head node remaining energy level must be above the threshold value that is set. This is done to avoid dead nodes and to not make the network disconnected. If any node has energy less than the value set for threshold then it has to be made as an "Ordinary Node". The nodes that are dead will be taken out of the network. This process has to be repeated so as to ensure that the network is always connected. Here, a radio frequency link is used to gather data from cluster heads.

3. COMPARISON OF CLUSTERING TECHNIQUES

The comparison of various clustering techniques is summarized as shown in Table 1.

Table 1. Comparison of Clustering Techniques

Clustering Approach	Strengths	Weaknesses
Low Energy Adaptive Clustering [1]	• Depletion of node Energy • It balances the energy consumption of the entire network • It extends the lifetime of the network	• It is not appropriate for larger networks • The node's initial energy will not be considered during the cluster head selection
Event to sink directed clustering [4]	• It avoids the formation of unnecessary clusters in the network. • Maintenance overhead is reduced	• Clustering depends on selection of sink nodes.
Load Balanced Clustering [5]	• Flow of data among sensor nodes is not constant • Balances consumption of energy	• Energy is balanced only by restricting the size of nodes • Depletion of energy of closest nodes
Energy Efficient Hierarchical Clustering [7]	• Easy implementation • Minimize total energy • Suitable for large network	• Not suitable for less number of nodes
Weight Based Clustering [8]	• Less number of dead nodes • Low energy sensors will never be selected as Cluster Heads.	• Large number of nodes may be elected as cluster heads.
Cluster Head Rotation Scheme [9]	• Balances energy of nodes • Suitable for sensing in environments where we need to sense for longer period of time	• Not suitable for network with less number of nodes.
Self Organized Energy Conscious Clustering [12]	• Conserves energy • Lifetime of network is improved	• Do not support scalability and mobility of nodes
Energy Aware Fuzzy Clustering Algorithm [13-21]	• Aggregation of data • Lifetime of network is improved • Enhances throughput	• Do not support real time data
Energy Efficient Recursive Clustering [7, 23]	• Reduces the consumption of energy and extra overhead of redundant data	• Repeatedly Clustering need to applied causing overhead.
Adaptive distributed Clustering Algorithm [24, 25]	• Reduces depletion of energy by avoiding redundant data	• Do not support mobility.
Improved Energy Aware Distributed Unequal Clustering [24, 26-27]	• Reduces Cluster overheads • Balances consumption of energy	• Clusters formed are of uneven size.
Hierarchical Clustering [29,34]	• Reduces the consumption of energy • Improves lifetime of network.	• Imbalances consumption of energy
Two Level Hierarchy for Low Energy Adaptive Clustering Hierarchy [25]	• Reduces depletion of energy of nodes • Provides proper load distribution among sensors.	• Lowers overall energy consumption.
Mobile Sink - based Approach [30]	• Traffic load is reduced by using Data aggregation • More suitable for large-scale deployed networks • It prolongs the lifetime of network • Solves the energy hole problem	• Energy consumption problem is not resolved completely.
Clustering using Fuzzy logic [13-19]	• Avoids redundant data • Increases lifetime of network.	• It reduces number of messages
Clustering using Node Relevance [35]	• Balances node overhead • Suitable in remote and inaccessible areas.	• Size and timing of clustering is determined partly.
Distributed Fuzzy Logic based Energy aware and Coverage preserving Unequal Clustering [18]	• Provides energy efficiency • Controls the information passing in the network thereby reducing the energy	• Forms unequal sized clusters.
Optimal Cluster Size Selection-based Clustering protocol using Fuzzy Logic [37]	• Increases the lifetime of a network	• Clusters formed are of uneven size

Quad Clustering based on K-Means Algorithm [38]	• Reduces the consumption of energy	• It do not support multiple clusters.
Dragonfly Algorithm for Hierarchical Clustering [39]	• Saves Energy	• Optimization is not present
Energy Efficient Sleep Awake Aware Node Schedule Clustering [41]	• Improve the lifetime of a network by reducing energy utilization. • Supports more transmissions than usual.	• Optimization reduces flow of data.
Node Overhaul Scheme for Energy Efficient Clustering [44]	• Increases the lifetime of a network and also the node death rate.	• Do not work for unequal sized clusters.
Distance and Energy constrained K-Means Clustering Scheme [45-47]	• Increases the energy of network.	• Do not concentrate on coverage of nodes and consistency.

4. CONCLUSIONS

In this paper, the novel clustering techniques have been studied. This study reveals that Clustering algorithms follow various parameters such as distance, mobility, node position, etc. in the formation of clusters. Also, it has been found that each Clustering algorithm has its own advantages and disadvantages. So, these techniques have to be applied based on the requirement. This paper will provide a beneficiary source to all the researchers working in the domain of sensor networks.

5. ACKNOWLEDGEMENT

I would like to thank my supervisor Dr. R. Sridevi, Professor, Department of CSE, JNTU Hyderabad for providing her valuable timely suggestions in the preparation of this article and in guiding me in all aspects to carry out my research.

6. REFERENCES

[1] L. Yadav, Ch. Sunitha, "Low Energy Adaptive Clustering Hierarchy in Wireless Sensor Network (LEACH)", International Journal of Computer Science and Information Technologies, Vol. 5, No. 3, 2014, pp. 4661-4664.

[2] I. Daanoune, B. Abdennaceur, A. Ballouk. "A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks", Ad Hoc Networks, Vol. 114, 2021.

[3] M. Adnan, L. Yang, T. Ahmad, Y. Tao, "An Unequally Clustered Multi-hop Routing Protocol Based on Fuzzy Logic for Wireless Sensor Networks", IEEE Access, Vol. 9, 2021, pp. 38531-38545.

[4] A. Bereketli, O. B. Akan, "Event-to-Sink Directed Clustering in Wireless Sensor Networks", Proceedings of the IEEE Wireless Communications and Networking Conference, 2009, pp. 1-6,

[5] B. Baranidharan et al. "DUCF: Distributed load balancing Unequal Clustering in wireless sensor networks using Fuzzy approach", Applied Soft Computing, Vol. 40, 2016, pp. 495-506.

[6] B. Jan, H. Farman, H. Javed, B. Montrucchio, M. Khan, S. Ali, "Energy Efficient Hierarchical Clustering Approaches in Wireless Sensor Networks: A Survey", Wireless Communications and Mobile Computing, Vol. 2017, 6457942, 2017, 14 pages.

[7] N. Merabtine, D. Djenouri, D.-E. Zegour, "Towards Energy Efficient Clustering in Wireless Sensor Networks: A Comprehensive Review", IEEE Access, Vol. 9, 2021, pp. 92688-92705.

[8] R. Tandon, B. Dey, S. Nandi, "Weight based clustering in wireless sensor networks", Proceedings of the National Conference on Communications, 2013, pp. 1-5.

[9] S. Pradhan, K. Sharma, "Cluster Head Rotation in Wireless Sensor Network: A Simplified Approach", International Journal of Sensor and Its Applications for Control Systems, Vol. 4. 2016, pp. 1-10.

[10] M. C. M. Thein, T. Thein, "An Energy Efficient Cluster-Head Selection for Wireless Sensor Networks", Proceedings of the International Conference on Intelligent Systems, Modelling and Simulation, 2010, pp. 287-291.

[11] T. Jagannadha Swamy, G. Ramamurthy, P. Nayak, "Optimal, Secure Cluster Head Placement Through Source Coding Techniques in Wireless Sensor Networks", IEEE Communications Letters, Vol. 24, No. 2, 2020, pp. 443-446.

[12] M. Bala Krishna, M. N. Doja, "Self-organized energy conscious clustering protocol for wireless sensor networks", Proceedings of the 14th International Conference on Advanced Communication Technology, 2012, pp. 521-526.

[13] H. Bagci, A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks", Applied Soft Computing, Vol. 13, No. 4, 2013, pp. 1741-1749.

[14] Z. Ullah et al. "Energy-Efficient Harvested-Aware Clustering and Cooperative Routing Protocol for WBAN (E-HARP)", IEEE Access, 2019.

[15] R. Logambigai, A. Kannan, "Fuzzy logic based unequal clustering for wireless sensor networks", Wireless Networks, Vol. 22, 2016, pp. 945-957.

[16] S. Rana, A. N. Bahar, N. Islam, J. Islam, "Fuzzy Based Energy Efficient Multiple Cluster Head Selection Routing Protocol for Wireless Sensor Networks", Computer Network and Information Security, Vol. 4, 2015, pp. 54-61.

[17] S, Mao, C, Zhao, Z, Zhou, Y, Ye, "An improved fuzzy unequal clustering algorithm for wireless sensor network", Proceedings of the 6th International ICST Conference on Communications and Networking in China, 2011, pp. 245-250.

[18] N. Mazumdar, H. Om, "Distributed fuzzy logic based energy-aware and coverage preserving unequal clustering algorithm for wireless sensor networks", International Journal of Communication Systems, 2017.

[19] M. Mirzaie, S. M. Mazinani, "AFLCH: Self-adaptive unequal fuzzy based clustering of heterogeneous sensors in wireless sensor networks", Proceedings of the 28th Iranian Conference on Electrical Engineering, 2020, pp. 1-5.

[20] J. Hu, Y. Pan, T. Li, Y. Yang, "TW-Co-MFC: Two-Level Weighted Collaborative Multi-view Fuzzy Clustering Based on Maximum Entropy", Proceedings of the Seventh International Conference on Advanced Cloud and Big Data, 2019, pp. 303-308.

[21] M. Singh, Gaurav, S. Soni, V. Kumar, "Clustering using fuzzy logic in wireless sensor networks", Proceedings of the 3rd International Conference on Computing for Sustainable Global Development, 2016, pp. 1669-1674.

[22] Z. Zeng-Wei, W. Zhao-Hui, L. Huai-Zhong, "An event-driven clustering routing algorithm for wireless sensor networks", Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004, pp. 1802-1806.

[23] I. Akila, S. V. Manisekaran, R. Venkatesan, "Modern Clustering Techniques in Wireless Sensor Networks", InTech Open2017.

[24] N. Kalla, P. Parwekar, "A Study of Clustering Techniques for Wireless Sensor Networks", Satapathy, Smart Computing and Informatics. Smart Innovation, Systems and Technologies, Vol 77. Springer, Singapore, 2017.

[25] V. Loscri, G. Morabito, S. Marano, "A two-levels hierarchy for low-energy adaptive clustering hierarchy (TL-LEACH)", Proceedings of the IEEE 62nd Vehicular Technology Conference, 2005, pp. 1809-1813.

[26] B. Gong, L. Li, S. Wang, X. Zhou, "Multihop Routing Protocol with Unequal Clustering for Wireless Sensor Networks", Proceedings of the ISECS International Colloquium on Computing, Communication, Control, and Management, 2008, pp. 552-556.

[27] U. Hari, B. Ramachandran, C. Johnson, "An Unequally Clustered Multihop Routing protocol for Wireless Sensor Networks", Proceedings of the International Conference on Advances in Computing, Communications and Informatics, 2013, pp. 1007-1011.

[28] S. Bandyopadhyay, E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks", Proceedings of the Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies, 2003, pp. 1713-1723.

[29] S. G. Neogi, A. A. Bhaskar, P. Chakrabarti. "Energy Efficient Hierarchy-based Clustering Routing Protocol for Wireless Sensor Networks", International Journal of Computer Applications, Vol. 95, No. 13, 2014, pp. 1-8.

[30] J. Wang, Y. Yin, J.-U. Kim, S. Lee, C.-F. Lai, "A Mobile-Sink Based Energy-Efficient Clustering Algorithm for Wireless Sensor Networks", Proceedings of the IEEE 12th International Conference on Computer and Information Technology, 2012, pp. 678-683.

[31] T. K. Jain, D. S. Saini, S. V. Bhooshan, "Increasing Lifetime of a Wireless Sensor Network Using Multiple Sinks", Proceedings of the 11th International Conference on Information Technology: New Generations, 2014, pp. 616-619.

[32] S. Arjunan, S. Pothula, "A survey on unequal clustering protocols in Wireless Sensor Networks", Journal of King Saud University - Computer and Information Sciences, Vol. 31, No. 3, 2019, pp 304-317.

[33] B. Tang, J. Wang, X. Geng, Y. Zheng, J. Kim, "A Novel Data Retrieving Mechanism in Wireless Sensor Networks with Path-Limited Mobile Sink", International Journal of Grid and Distributed Computing, Vol. 5, No. 3, 2012.

[34] J.-S. Lee, C.-L. Teng, "An Enhanced Hierarchical Clustering Approach for Mobile Sensor Networks Using Fuzzy Inference Systems", IEEE Internet of Things Journal, Vol. 4, No. 4, 2017, pp. 1095-1103.

[35] M. Mirzaie, S. M. Mazinani, "MCFL: an energy efficient multi-clustering algorithm using fuzzy logic in wireless sensor network", Wireless Networks, Vol. 24, 2018, pp. 2251-2266.

[36] S. Gajjar, A. Talati, M. Sarkar, K. Dasgupta, "FUCP: Fuzzy based unequal clustering protocol for wireless sensor networks", Proceedings of the 39th National Systems Conference, 2015, pp. 1-6.

[37] P. K. Mishra, S. K. Verma, "OCSSP: Optimal Cluster Size Selection-based Clustering Protocol using Fuzzy Logic for Wireless Sensor Network", Proceedings of the IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation, 2020, pp. 1-4.

[38] B. Kumar, U. K. Tiwari, S. Kumar, "Energy Efficient Quad Clustering based on K-means Algorithm for Wireless Sensor Network", Proceedings of the Sixth International Conference on Parallel, Distributed and Grid Computing, 2020, pp. 73-77.

[39] B. Pitchaimanickam, "Dragonfly Algorithm for Hierarchical Clustering in Wireless Sensor Networks", Proceedings of the 5th International Conference on Intelligent Computing and Control Systems, 2021, pp. 192-197.

[40] S. Mirjalili. "Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems", Neural Computing and Applications, Vol. 27, No. 4, 2016, pp. 1053-1073.

[41] T. Shah, N. Javaid, T. N. Qureshi, "Energy Efficient Sleep Awake Aware (EESAA) intelligent Sensor Network routing protocol", Proceedings of the 15th International Multitopic Conference, 2012, pp. 317-322

[42] V. Kumar, V. K. Sinha, "Energy Efficient Node Schedule Clustering with Optimal Cluster in Wireless Sensor Network", Proceedings of the 2nd International Conference on Advances in Computing, Communication Control and Networking, 2020, pp. 363-367.

[43] M. Ashourian, M. Gheisar, A. H. Talkhoncheh, "An Improved Node Scheduling Scheme for Resilient Packet Ring Network", Majlesi Journal of Electrical Engineering, Vol. 9, No. 2, 2015, pp. 43-50.

[44] J. Singh, S. S. Yadav, V. Kanungo, Yogita, V. Pal, "A Node Overhaul Scheme for Energy Efficient Clustering in Wireless Sensor Networks", IEEE Sensors Letters, Vol. 5, No. 4, 2021, pp. 1-4.

[45] K. G. Omeke et al. "DEKCS: A Dynamic Clustering Protocol to Prolong Underwater Sensor Networks", IEEE Sensors Journal, Vol. 21, 2021. pp. 9457-9464.

[46] G. Y. Park, H. Kim, H. W. Jeong, H. Y. Youn, "A Novel Cluster Head Selection Method based on K-Means Algorithm for Energy Efficient Wireless Sensor Network", Proceedings of the 27th International Conference on Advanced Information Networking and Applications Workshops, 2013, pp. 910-915.

[47] M. Bidaki, R. Ghaemi, K. Tabbakh, "Towards Energy Efficient k-MEANS Based Clustering Scheme for Wireless Sensor Networks", International Journal of Grid and Distributed Computing, 2016, pp. 265-276.

[48] R. Logambigai, A. Kannan, "Fuzzy logic based unequal clustering for wireless sensor networks", Wireless Networks, Vol. 22, 2016, pp. 945-957.