Journal of Network Intelligence ©2017 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **2**, Number **4**, November **2017**

Clustering Formation in Wireless Sensor Networks: A Survey

Jeng-Shyang Pan
School of Information Engineering and Science
Fujian University of Technology, China
jengshyangpan@gmail.com

Trong-The Nguyen, Thi-Kien Dao
Department of Information Technology
Haiphong Private University,Vietnam
35 Danlap Road, Du Hang Kenh, Le Chan, Hai Phong, Viet Nam
vnthe@hpu.edu.vn, jvnkien@gmail.com

Tien-Szu Pan
Department of Electronics Engineering
National Kaohsiung University of Applied Sciences, Taiwan
tpan@cc.kuas.edu.tw

Shu-Chuan Chu
School of Computer Science, Engineering and Mathematics
Flinders University, Australia
jan.chu@flinders.edu.au

Received July 2017; revised November 2017

Abstract. *Clustering formation is the modern energy efficient techniques in the designing and implementing Wireless sensor networks (WSN). Clustering provides various advantages like energy efficiency, lifetime, scalability, and less delay; but it leads to hot spot problem. The proposed unequal clustering is to overcome this issue. In uneven clustering, the cluster size varies proportionally to the distance to the base station (BS). In this paper, a comprehensive survey of various clustering formation approaches with their objectives, characteristics, etc. is presented. Also, the classifications of uneven clustering methods are carried out and compared them based on various cluster properties, Cluster Head (CH) properties, and clustering process.*
Keywords: Clustering Formation, Wireless Sensor Networks, Data aggregation, Clustering, Hotspot problem.

1. **Introduction.** Fast developing in the field of integrated circuits and information technology leads to the development of cheap and compact size sensor nodes. WSN composed of a set of a vast number of sensor nodes arranged in an ad-hoc fashion to observe and interact with the physical world. Each sensor node contains four components: sensors, microcontrollers, power supply and transceivers. The sensors in the sensing unit measure the physical parameters in the real world such as temperature, pressure, humidity, vibration, acoustic signal, infrared, vehicular movement [1]. The sensed information is processed by the processing unit and forwarded to the Base Station (BS) through single hop or via intermediate nodes by the communicating unit. WSN is commonly used in

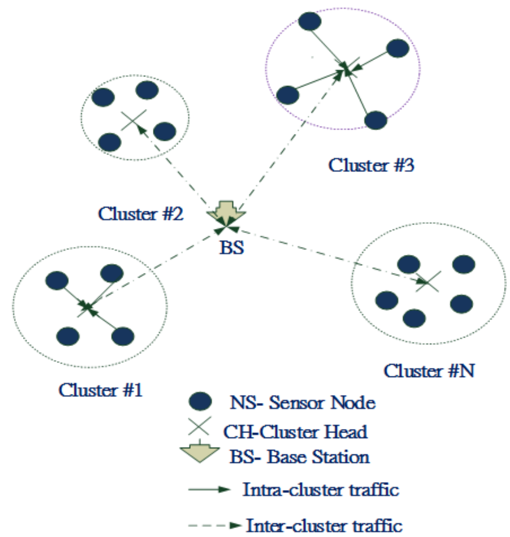

FIGURE 1. Architecture of clustering in WSN

real-time monitoring and tracking applications such as military surveillance, agriculture, disaster management, health-care monitoring, industry automation, inventory control [1] [2].

WSN is usually required to deploy in regions where human intervention is difficult or not possible. Energy consumption, bandwidth, and memory are needed to consider as a major issue in WSN design. As the sensors are deployed in the harsh environment, it is tough or not possible to replace or recharge batteries. WSN is energy constrained; transmission cost is more than sensing and processing cost [3]. Therefore, an energy efficient data transfer strategy is required to forward data from sensor nodes to BS to lengthen the network lifetime. An important energy effective method is clustering; the most important energy efficient technique where the sensor nodes are organized into groups termed as clusters. The regular nodes in the cluster are called as cluster members, and a Cluster Head (CH) is selected among them [4]. The architecture of clustering in WSN is shown in Figure 1.

There are two types of traffic in clustered WSN: data transmission within a cluster defined as intra-cluster traffic and data transmission between clusters which is termed as inter-cluster traffic. The cluster members sense the real world parameters and transmit the sensed value to its CH. CH receives and aggregates data to remove redundant data and transmit aggregated data to CH directly or via intermediate CHs. The cluster members cannot send the data directly to BS. It sends only to the CH and CH forwards it to BS. The advantages of clustering are: energy consumption is reduced by improving bandwidth utilization, reduced overhead, increased connectivity, stabilized network topology, decreased delay, effective load balancing and reduction in the size of the routing table. The CHs near the BS consumes more energy and drains out energy more quickly than the CHs farther from BS. CHs closer to BS are loaded with heavy traffic due to: intra-cluster traffic from its cluster members, data aggregation, and inter-cluster traffic from other CHs for relaying data to BS. This result to disrupt the network connectivity and coverage issues are created in the clusters closer to BS. This issue is referred as hot spot problem. To prevent the network from hot spot issue, utilizing unequal clustering techniques are for load balancing between the CHs [5]. The architecture of unequal clustering in WSN is

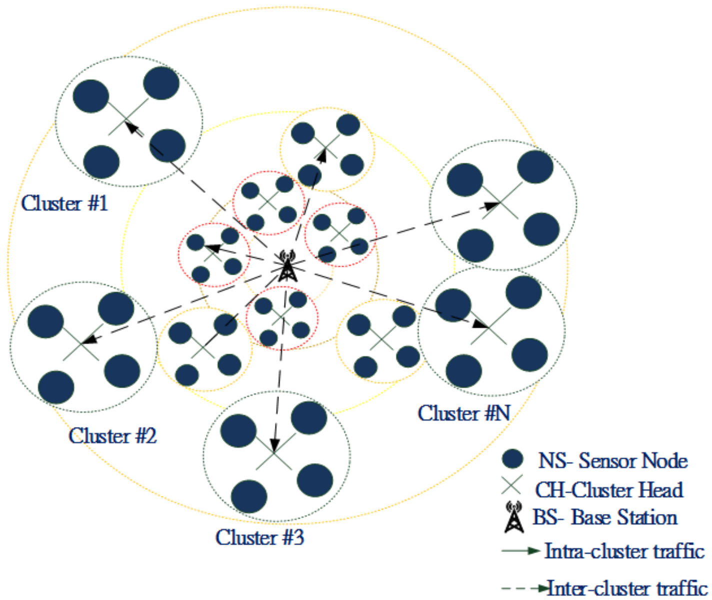

FIGURE 2. Architecture of Unequal Clustering in WSN

shown in Figure 2. Uneven Clustering reduces the size of the clusters closer to BS and the cluster size increases as the distance between the BS and CH increases. The cluster size is directly proportional to the distance of CHs from BS. When the distance to BS increases, cluster size also increases. Smaller cluster near the BS indicates less number of cluster members and less intra-cluster traffic. So, the smaller clusters consume less energy for intra-cluster traffic and concentrate more on inter-cluster traffic. Similarly, larger clusters farther from BS indicates more cluster members and spend more energy on intra-cluster traffic. As a result, it spends less energy for inter-cluster traffic and no need of spending more energy for inter-cluster routing. Unequal Clustering permits all CHs to pay the same amount of energy so that the CHs near BS spend equal energy as CHs farther from BS. So, uneven grouping eliminates hot spot problem by balancing the load efficiently.

Various clustering techniques have been presented for energy efficient WSN for recent times. Although several comprehensive surveys are given for clustering approaches, a few has performed a review on both of clustering and unequal clustering methods. Probabilistic-based unequal clustering protocols are examined, and the protocol are compared with terms of node deployment region (square or circle), mobility, location awareness and data aggregation[6]. Another review of Probabilistic-based unequal clustering protocols is presented. It compares various rules based on some nodes, energy efficiency, balanced cluster, location awareness and heterogeneity level [7]. Elaborate survey for clustering formation inspired us to perform this work. In this paper, both equal and unequal clustering objectives, characteristics, classification, merits and demerits of each technique are discussed in detail and comparison is also made based on various cluster properties, Cluster Head (CH) properties and clustering process is given.

The remainder of the paper is organized as follows. Related properties of clustering are overviewed with objectives and characteristics in Section 2. Equal Clustering algorithms are classified and explained in Section 3. Classified algorithms of unequal Clustering formation are explained and tabulated in Section 4 and the paper is concluded in Section 5.

2. **Related Properties of Clustering.**

2.1. **Clustering objectives.** There are two categories of clustering formation such as unequal clustering and equal clustering in WSN. The objectives of unequal clustering are same as equal clustering with some additional functions. The nodes are clustered in WSN with different purposes which are based on application requirement. Energy conservation and eliminating hot spot problem are the most common objectives of unequal clustering. Some of the additional objectives are explained as below.

Scalability: Sensor nodes are deployed in large numbers ranging from hundreds to thousands based on the application requirement in the real scenario. The design of routing techniques should consider the ability to work with these huge number of sensor nodes. As a sensor node in the cluster needs to transmit data to a node in another cluster, the nodes should know the details of the associated receiving cluster head (CH). The hierarchical architecture provides scalability in large-scale WSN by dividing the sensing field into various layers, and each layer is again divided into some clusters [8]. It leads to increase in scalability and reduces the size of the routing table.

Load Balancing: Load balancing plays an essential role to prolong the network lifetime. Load balancing is a crucial issue where CHs are selected from available nodes in the network [9]. Uniform load distribution among the CHs is essential to avoid hot spot problem. Unequal Clustering guarantees uniform load distribution in which all the CHs consumes the approximately same amount of energy. As a result, more energy efficient network can be easily achieved.

Fault-Tolerant: In applications, sensors are deployed in a harsh environment (e.g., sensors are dropped from a helicopter) and these nodes have an increased risk of physical damage, malfunction of nodes. Fault-tolerant nodes are important in a critical application where the loss of some sensor data results to catastrophe. Clustering is an efficient way to make a fault tolerant and secured WSN [10]. The self-organized WSN manages the fault by the process of re-clustering the network. The re-clustering process not only increases the resource burden, but It also disrupts the current operation. Re-clustering, assigning backup CH, depute CH or rotating CH results to fault-tolerant with an advantage of proper load balancing [11].

Data Aggregation/ fusion: Since a large number of sensor sense the same data in the physical environment, there is a greater chance of data redundancy. Data aggregation is an effective way of avoiding redundant data transmission, and also it reduces the number of transmissions. This technique is a signal processing method, which aggregates all received packets into an output packet. This technique amplifies the common data and suppresses the unwanted noise [12]. In WSN, CH performs data aggregation of all data received from its cluster members and forwards the aggregated data to the base station (BS) via single hop or multi-hop. Consequently, the number of transmissions and the total load of the network are also significantly reduced.

Stabilized network topology: Nodes are organized into clusters and CH is selected from each cluster; CH is responsible for any topology changes at the cluster level. CH contains the information of its cluster members like node id, location, and energy level. Managing the network topology in hierarchical architecture is better than flat architecture. When a node dies or moves to another cluster, these changes are immediately registered and informed by CH to BS and reclustering will be done to maintain the network topology effectively.

Increased lifetime: The major aim of unequal clustering improves the network lifetime as a long as possible. As the sensors are energy-constrained, maximizing the network lifetime is crucial for real-time applications. Intra-cluster communication can be reduced by selecting nodes as CHs with more neighbor nodes [13]. Clustering and routing process can also be combined to maximize lifetime [14]. Clustering prolongs the lifetime of WSN by rotating CHs properly among the cluster members, sleep modes, and cluster maintenance techniques can be properly utilized to increase the network lifetime.

2.2. **Clustering characteristics.** Various clustering characteristics are used to classify different clustering approaches. In this subsection, the three discussed aspects of clustering formation are in detail such as Cluster properties, CH properties, and Clustering process properties.

2.2.1. *Cluster properties.* The specifications of the cluster are defined as cluster properties which include: cluster count, cluster size, intra-cluster communication and inter-cluster communication.

Cluster count: The number of the cluster formed is predefined or variable based on the application requirement. In some cases, the number of clusters is 5% of a total number of the nodes deployed. In applications, the number of clusters is variable when the CHs are randomly selected.

Cluster size: The cluster size can be classified into equal and unequal size cluster. In equal clustering, the size of the cluster is same throughout the network. In unequal clustering, the cluster size is determined based on the distance to BS. The cluster size is smaller when the distance to BS is small and the size increases as the distance to BS increases.

Intracluster communication: Intra cluster communication involves the data transmission between CH and cluster member within a cluster. Based on the clustering approaches, the communication can be direct or multi-hop. For large scale WSN, multihop communication is needed for data transmission within a cluster.

Inter cluster communication: Inter cluster communication can be direct or multi-hop communication. Usually, the multi-hop mechanism is preferred for energy efficient data transmissions from CHs to BS through intermediate CHs in large scale WSN. In some applications of small-scale WSN, the communication between the CH and BS is single hop transmission.

2.2.2. *CH properties.* CH performs the following operations: Collecting data from cluster members, aggregating data and forwarding data to BS through direct or multi-hop communication.

Role: The CH receives data from its cluster members, performance data aggregation of the collected sensor data, relays the aggregated data to BS.

2.2.3. *Clustering process.* The characteristic of Clustering process is listed below.

Clustering methods: There are two methods of clustering: centralized and distributed. In centralized approaches, a central authority like BS or super nodes controls the entire operation (cluster formation, CH selection, etc.,) while distributed approaches have no central authority and widely employed in large scale WSN.

The objective of Node grouping: Various objectives of nodes grouping are already discussed in this study. e.g., Fault tolerance, load balancing, etc. Nature: The clustering process can be proactive, reactive or hybrid. The node continuously senses the data and forwards it to CH. In proactive type, the CH transmits the data to BS continuously. In reactive type, CH transmits the data whenever the sensed value crosses the predefined threshold. In hybrid cases, CH transmits the data to BS at longer regular time intervals and also when the value crosses the threshold value.

CH selection: There are three ways to select CH in WSN: probabilistic methods attribute-based method and preset type. In probabilistic approaches, CHs are selected randomly without any prior consideration. In the attribute-based method, various metrics are used to select CHs like residual energy, node degree, node centrality, expected residual energy, distance to BS, etc. In preset type, CHs are predetermined before placing the sensors in the sensing field.

3. **Classification Equal Clustering Algorithms.** The purpose of clustering methods is the prolonging the network lifetime, saving residual energy of each sensor node, and ensuring the connectivity nodes. A detailed classification of the clustering formation is reviewed based on different metrics i.e. routing protocols. Figure 3 shows three divided categories clustering techniques metrics included: based-block clustering, based-grid clustering, and based-chain clustering.

3.1. **Based-Block Clustering.** The traditional based-block clustering algorithms included LEACH [4], HEED[15], LEACH-C[16], TEEN[17], CCM, ECHA[18]. We reviewed the advantages and disadvantages of this classification as follows:

Low Energy Adaptive Clustering Hierarchy (LEACH): LEACH randomly selects CHs among the sensor nodes and rotates the role of sensors to distribute the energy load of the network evenly [4]. The organized LEACH is two phases including a setup phase and steady state phase. The setup phase, the nodes are arranged themselves into clusters, with one node acting as CH. Nodes elect as CHs with a probability. CHs broadcast their status to its member nodes in a cluster of the network. Node chooses a group as belong to whom in the system with communicating minimum energy. The turned role of the node as a CH is to next CH node in each round. A randomized rotation is carried out according to a high-energy among sensors to avoid of draining the battery of the sensors. The selected CH expressed in choosing a random number from $0to1$, with considering a threshold value $T(m)$.

$$T\left(m\right) = \begin{cases} \frac{\theta}{1-\theta\times(i \ mod \ 1/\theta)} \ , \ m \ \in S \\ 0 \ , \ otherwise \end{cases} \tag{1}$$

where θ is the desired percentage of CHs, m is the round number, i is the current round and S is the set of nodes that have not yet been selected [18]. The steady state phase, the sensor nodes can begin sensing and transmitting data to their CHds. After receiving all the sensed data, CHs aggregates it before sending it to the BS. However, LEACH uses single hop communication so it can not use in large scale network. There are the equal chance for each node to become CH, but cannot be selected as CH in the subsequent round. CHs of LEACH are elected by probability so uniform distribution cannot be ensured and cannot provide load balancing.

Low-Energy Adaptive Clustering Hierarchy-Centralized (LEACH-C)[16]: LEACH-C was created based on LEACH to organize the nodes in WSN into clusters with each group have a CH. LEACH-C divides rounds into the set-up and steady-state phases. It differs from LEACH only in that it uses a high-energy BS to finish the choice of CHs. Each round in the set-up phase, every member node transfers its sensed data to BS. BS then selects the CHs based on the energy information and broadcasts the IDs of CHs to other member nodes. This method can save the nodes with more power and more chance to become the CH in the current round. Furthermore, this phase, every sensor node needs to send its ID and energy information to remote BS to compete for the role of cluster heads, which causes power consumption to the long distance transition. However, LEACH-C has drawbacks like the following: Randomly selected CHs made the residual energy of nodes does not take into account to consider in the process, and the support of BS is needed. Reclustering is high-frequency, so it wastes a certain amount of energy.

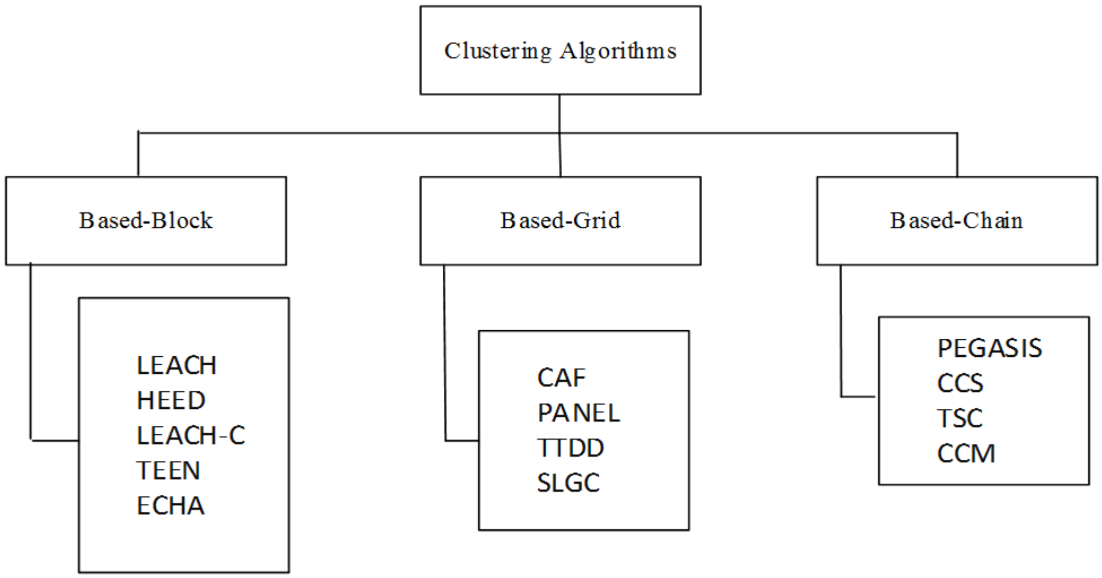

Figure 3. Classification of equal clustering algorithms

The coverage area of LEACH-C is smaller than LEACH. The distributed CHs are not uniformly, place the locating CHs can be at the edge of the cluster.

Threshold Sensitive Energy Efficient Sensor Network (TEEN)[17]: Data transmission of TEEN can be controlled by varying two thresholds. It is well suited for time critical applications. However, whenever thresholds are not met, the node will not communicate. Data may be lost if CHs are not able to communicate with each other. Every cluster changes times, and the CH broadcasts to its members.

Energy-based cluster head selection algorithm (ECHA)[18]: ECHA is an improved LEACH version to support long-lifetime in WSN. In ECHA, the number of the selected cluster heads (CHs) is optimized based on the effect of the distances from the member nodes to the CH, from the CHs to the base station (BS) and the high energy of each CHs. However, the impacts of parameters of the density of nodes and weighting probability are verified to find out the best cases for ECHA applications.

3.2. **Based-Grid Clustering.** The popular based-grid clustering algorithms are GAF[19], PANEL[20], SLGC[21], TTDD[22], etc. as shown in figure 3. The advantages and disadvantages of these methods are reviewed as below:

Geography-Informed Energy Conservation for Ad Hoc Routing (GAF) [19]: GAF converses energy by identifying nodes that are equivalent from routing perspective, and then turning off the unnecessary node, keeping a constant level of routing fidelity. GAF moderates the policy using application and system level information; nodes that source or sink data remain on and intermediates node monitoring and balance energy consumption. However, massive traffic injection and delay is not predictable in this algorithm.

Position-based aggregator node election (PANEL) [20]: PANEL supports asynchronous sensor network applications where BSs fetch the sensor readings after some delay. It is on other aggregator node election protocols. PANEL was to support reliable and persistent data storage applications. PANEL ensures load balancing, and it supports intra-and inter-cluster routing allowing the sensor to the aggregator, aggregator to aggregator, base station to the aggregator, and aggregator to BS communications. However, Cluster determined the geographic position information, and special conditions are needed which is not alway available.

SLGC[21]: prolongs network lifetime and decreases power consumption by selecting CH in the grid. SLGC has lower energy consumption compared with LEACH, due to it selects CHs in the current and next rounds based on the best condition of the parameters in the network. So, the CHs reduce the volume of controlling messages for next rounds and inform nodes for sending data into CH of the particular round. However, SLGC has significant overhead due to complex data communication.

Two-Tier Data Dissemination (TTDD) [22]: TTDD provides scalable and efficient data delivery to multiple, mobile sinks in large-scale WSN. Each data source in TTDD proactively constructs a grid structure, which enables portable sinks to continuously receive data on the move by flooding queries within a local cell only. TTDD's design exploits the fact that sensors are stationary and location-aware to construct and maintain the grid infrastructure with little overhead. However, It has a significant latency, low energy efficiency, and it also requires sensor nodes to be stationary and location aware.

3.3. **Based-Chain Clustering.** The popular based-Chain clustering algorithms are PEGASIS[23], CCS[24], TSC[25], CCM [26], etc. as shown in Figure 3. The advantages and disadvantages of these methods are reviewed as below:

Power-Efficient Gathering in SensorInformation Systems [23]: PEGASIS gather sensed information in an energy efficient manner is critical to operating the sensor network for a long period. If each node transmits its sensed data directly to the base station, then it will deplete its power quickly. PEGASIS is an improved scheme of LEACH, which is a near-optimal chain-based protocol that minimizes energy. In PEGASIS, each node communicates only with a close neighbor and takes turns transmitting to the base station, thus reducing the amount of energy spent per round.Two new schemes to minimize heat delay using CDMA and non-CDMA sensor nodes. If the goal is to reduce only the delay cost, then a binary combining system can be used to accomplish this task in about log N units of delay with parallel communications and incurring a slight increase in power cost. With CDMA capable sensor nodes, a chain-based binary scheme performs best regarding energy and delay. If the sensor nodes are not CDMA capable, then parallel communications are possible only among spatially separated nodes and a chain-based 3-level hierarchy scheme performs well. However, LEACH, and PEGASIS on energy delay using extensive simulations for different network sizes. Results show that PEGASIS perform 80 or more times better than the right plan and also outperform the LEACH.

The Concentric Clustering Scheme (CCS) [24]: CCS is a version of enhanced PEGASIS based on the concentric clustering scheme to solve efficient energy consumption in WSN. The main idea of the concentric grouping scheme is to consider the location of the BS to enhance its performance and to prolong the lifetime of the wireless sensor networks. CCS performs better than the current PEGASIS protocol by about 35%. However, it had unbalanced energy consumption, and it had a considerable delay because of the long chain.

Track-sector clustering (TSC) [25] over WSN: in TSC scheme, the network is divided into concentric circular tracks and triangular sectors. This division of tracks and sectors helps to reduce the energy consumption by minimizing redundant data transmission and providing shortest distance between head nodes and the BS. For the formation of tracks and sectors in the network, the proposed TSC requires further computations. However, all calculations for the structure of tracks and sectors are carried out in the BS just at the beginning of the network setup. Therefore, the goal of energy efficiency

Table 1. Comparison of the Equal Clustering Algorithms

Algorithms	**Consumed Energy**	**Balanced Load**	**Delayed Delivery**	**Cluster Stability**	**Scaled Network**	**Complexi-ty**
LEACH	Low	Medium	Small	Very low	Medium	Low
HEED	Medium	Medium	Medium	Medium	Medium	Medium
LEACH-C	Medium	Medium	Small	High	Medium	Medium
EECS	Medium	Medium	Small	Low	Low	High
CCM	Very low	Medium	Small	Very low	Low	Medium
TEEN	High	Good	Small	high	Very Low	Medium
ECHA	Medium	Good	Medium	Medium	Low	High
GAF	Medium	Medium	Small	high	High	Medium
PANEL	Medium	Medium	Small	High	Low	High
TTDD	Very Low	Bad	Small	High	Very low	Low
SLGC	Medium	Medium	Medium	Medium	Medium	Medium
PEGASIS	Low	Good	Small	Very Low	Medium	High
CCS	Low	Bad	Small	Low	Low	Medium
TSC	Medium	Bad	Large	Medium	Medium	Medium

is achieved without additional power consumption of the distributed nodes after system configuration process in TSC. However, the drawback of this algorithm is the unbalanced node distribution.

Chain-Cluster based Mixed (CCM)[26]: CCM used of the advantages of LEACH and PEGASIS and offered improved outcomes. Two stages of runs and a few chains in a network used as the phases. In the first step, sensor nodes in each chain transmit data to their chain head node in parallel, using an improved chain clustering process. In the second phase, all string head nodes group as a cluster in a self-organized manner, where they transmit fused data to a voted cluster head using the cluster based routing. However, CCM used Chain head selection criterion, and it has less energy consumption compared with LEACH.

3.4. **Comparison Clustering Algorithms.** Table 2 shows the comparison of the Equal Clustering Algorithms in terms of characteristic e.g., Consumed Energy; Balanced Load; Delayed Delivery; Cluster Stability, Scaled Network and Complexity.

4. **Unequal Clustering Algorithms.** In order to prevent the network from hot spot issue, unequal clustering techniques can be utilized for load balancing between the CHs [5]. Unequal clustering reduces the size of the clusters closer to BS and the cluster size increases as the distance between the BS and CH increases. The cluster size is directly proportional to the distance of CHs from BS. When the distance to BS increases, cluster size also increases. Smaller cluster near the BS indicates less number of cluster members and less intra-cluster traffic. So, the smaller clusters consume less energy for intra-cluster traffic and concentrate more on inter-cluster traffic. Similarly, larger clusters farther from BS indicates more cluster members and spend more energy on intra-cluster traffic. As a result, it spends less energy for inter-cluster traffic and no need for expenditure more energy for inter-cluster routing. Unequal Clustering permits all CHs to spend the same amount of energy so that the CHs near BS spend equal energy as CHs farther from BS. So, uneven grouping eliminates hot spot problem by balancing the load efficiently. For constructing unequal clusters, the parameter radius Rc is used to add to adjust the distances of CHs to BS [27]. The objective function of constructing unequal clusters as each CH needs to adjust their equal cluster distances:

$$D_j^{\,t+1} = D_j^{\,t} \times R_c \tag{2}$$

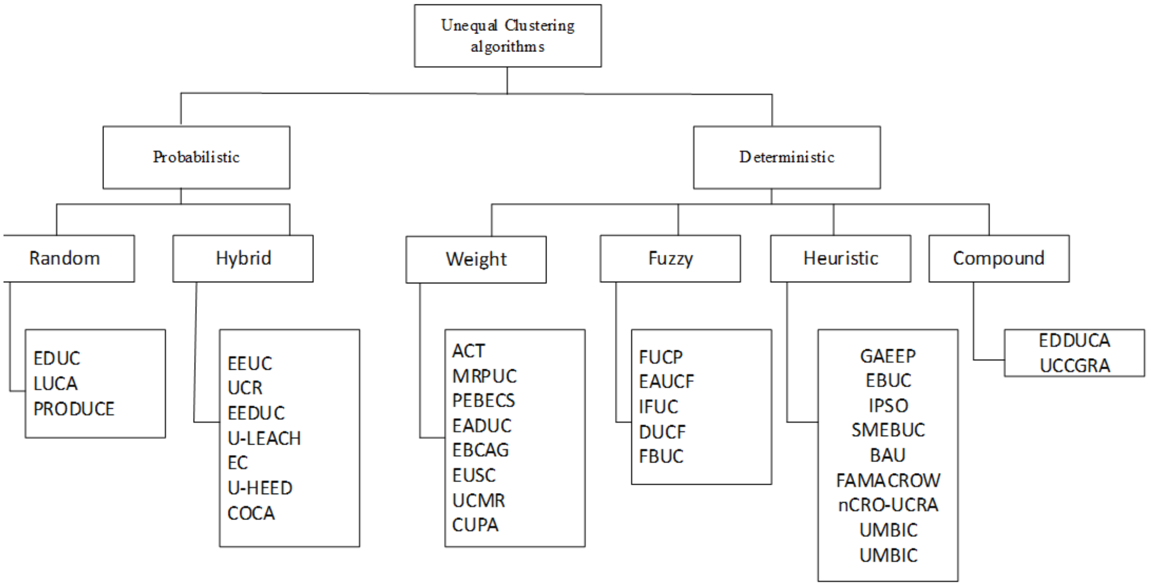

FIGURE 4. Classification of equal clustering algorithms

Where D_j is the distance from node CHj to the BS, $t+1$ is the current time as present iteration number, and R_c is a ratio adjusting parameter. This R_c is calculated as:

$$R_c = \left[1 - \alpha\frac{d_{max} - D_j}{d_{max} - d_{min}} - \beta(1 - \frac{E_r}{E_{max}})\right] \times R_{max} \tag{3}$$

where d_{max} and d_{min} are the maximum and minimum distance from the CHs in the network to the BS; R_{max} is the maximum value of competition radius; α is a weighted factor whose value is in [0, 1]; E_r is the residual energy of CH_j. Unequal Clustering algorithms are classified into categories: probabilistic, deterministic is shown in Figure 4.

4.1. **Probabilistic clustering algorithm.** The category of probability is to select CH randomly, and that is found to be simple, near optimal overhead, faster convergence, and energy efficient clustering method. Probabilistic approaches are divided into random methods and hybrid methods. Random strategies are simple which selects CHs and form randomly and achieve near optimal overhead. Hybrid methods are used to properly balance the clusters by combining random methods with some parameters like residual energy or distance to BS. Hybrid approaches are iterative based or competition based which increases the complexity concerning message and time.

4.1.1. *Random methods:* These methods are reviewed as below:
Probability Driven Unequal Clustering Mechanism for WSN (PRODUCE) : PRODUCE is a randomized unequal clustering algorithm which maximizes the network lifetime and coverage time in WSN where the density of nodes is high [28]. Coverage time is the time when the first CH runs out of battery resulting in coverage issues and lifetime indicates the time when all the nodes in the network die. It uses localized probabilities for constructing clusters of unequal size and stochastic geometry for inter-cluster routing. Unequal Clustering organizes clusters of smaller size near BS and larger size farther from BS. This makes the CHs near the BS to focus more on inter- cluster data relaying and

CHs farther from BS can concentrate more on intra-cluster communication. This probability driven unequal clustering scheme balances the energy consumption results to the maximization of network lifetime and coverage time especially in the network where the density of the nodes is high. It gives better results in terms of network lifetime when compared to EEUC.

Energy Driven Unequal Clustering (EDUC) : EDUC is a distributed algorithm which reduces the energy consumption in heterogeneous WSN. This algorithm effectively manages the energy consumption of nodes within a cluster to reduce the energy depletion. It involves a different clustering algorithm, and energy drove CH rotation method [29]. There are two phases in EDUC: cluster construction and data collection phase. The cluster construction phase includes CH competition stage and cluster formation stage. Each node serves as CH only once in the entire network lifetime. CHs prepare a TDMA schedule for its cluster members to avoid an intra-traffic collision. CHs are elected randomly, and the energy level is computed accurately for CH rotation. CHs forward the data directly to BS. This assumption of single-hop communication is not possible in many real-time applications. This method is not useful for multi-hop networks because energy threshold should be exact. Compared with LEACH and HEED, EDUC can prolong the network lifetime.

Location Based Unequal Clustering Algorithm (LUCA) : Another probability based distributed scheme is LUCA, where the cluster size varies proportionally to the distance from BS[30]. LUCA forms smaller clusters near the BS and larger clusters farther from BS. In the initialization stage, each node has a back-off timer with some random value. When the nodes receive any CH advertisement message within the time interval, it joins to the cluster. None of the CH advertisement messages are received; it selected itself as CH and advertises as CH to its neighbors. As LUCA organizes unequal clusters based on the location from BS, it uses GPS to determine its location. In LUCA, the nodes are location aware which makes it unsuitable for much real-time application and increases energy overhead.

4.1.2. *Hybrid approaches in unequal clustering algorithm.* These methods include as below.

Energy Efficient Unequal Clustering (EEUC) : EEUC is a hybrid approach widely used for periodical data gathering application in WSN [31]. To avoid hotspot problem, the network is divided into clusters of various unequal sizes, and multi-hop routing is involved for forwarding data to BS. The cluster size closer to BS is smaller which reduces the energy consumption due to intracluster traffic and save more energy for inter-cluster multi-hop routing. The cluster size is directly proportional to the distance to BS. EEUC is a distributed method where localized competition elects the CHs, and last CHs are chosen based on the residual energy of the sensor node. In multi-hop data transmission, CHs chooses the relay node using two metrics: residual energy and distance of relay CH from BS. EEUC significantly reduces the power consumption and lengthen the network lifetime when compared to LEACH and HEED and more practical than UCS.

Energy Efficient Distributed Unequal Clustering protocol (EEDUC) : EEDUC is a distributed approach can be used for periodical data gathering applications. The drawback of EEUC is overcome by EEDUC. In EEDUC, each node broadcasts the hello message based on the waiting time. The weight (waiting time) metric is computed at each node based on the residual energy and node degree; nodes with maximum weight are selected as CHs [32]. EEDUC balances the energy consumption 24% better than EEUC. EEDUC successfully maximizes the network lifetime. However, the CH chooses the adjacent CH as a relay node and without considering the residual energy and distance to BS.

Unequal Cluster based Routing (UCR) : UCR divides the network into clusters of unequal sizes to avoid hot spot problem. UCR protocol incorporates two components: energy efficient unequal clustering algorithm (EEUC) for clustering process, inter-cluster greedy geographic and energy aware routing protocol for multi-hop routing process [33]. BS broadcasts the beacon signal to all sensors to calculate its distance from every node based on the received signal strength. This is helpful for selecting appropriate power for data transmission to BS and building unequal clusters. Tentative CHs are chosen randomly, and they compete for final selection of CH. Each tentative CH has a competition range, and it is used to construct unequal size clusters. After the selection of CHs, CH broadcasts advertisement message to the network. The node joins as cluster member to the CH with higher received signal strength, and the Voronoi region of sensor node is also constructed. For multi-hop inter-cluster routing, the relay nodes are chosen based on the ratio of residual energy and energy cost of the relay paths. It achieves maximum lifetime over HEED but it prone to error and less robust because of the noise in the real environment.

Unequal LEACH: LEACH is a first developed clustering algorithm which elects CH based on probability. Due to practical difficulties in LEACH, improved LEACH with more practical setup phase is proposed. Improved LEACH constructs clusters based on Adaptive On-demand Weighting (AOW) which is a tradeoff between residual energy to total energy and competition range [34]. The clusters are unequal in size and the CHs directly forward the data to BS without any intermediate nodes. Round robin method is used for cluster head selection and the rotated time slot is predetermined. In steady state phase, cluster members use TDMA schedule to transmit data to CH and CH uses CSMA schedule to transmit the received and aggregated data to BS. After a certain period, re-clustering takes place. Improved LEACH results to more stabilized network topology and maximum lifetime compared to LEACH. This protocol is well suited for large scale WSN where the density of nodes is high.

Energy-Efficient Clustering (EC) : EC is a simple, scalable and energy aware clustering algorithm which calculates the required cluster size based on the hop distance to sink. It attains approximate equalization of network lifetime and reduces energy consumption [35]. In EC, tentative CHs are selected randomly and final CHs are elected depending on their residual energy. An energy-efficient multi-hop data collection protocol is proposed to determine the amount of energy consumption and the performance of EC. The distributed inter-cluster routing algorithm performs uniform energy distribution and produces minimal overhead due to route discovery process. EC is not dependent on the energy efficient data collection protocol. EC performs well and achieves more equalization in energy than UCR and HEED.

Energy-Balancing Unequal Clustering Protocol for WSN (EB-UCP) : EB-UCP is presented to attain maximum lifetime and uniform load balancing among all the nodes in WSN. To eliminate hot spot problem, a probabilistic approach is used to organize clusters [36]. Clusters near the BS are smaller in size for more inter- cluster routing and less intra-cluster routing. The sensing field is divided into various layers, and each layer is assigned a different probability relative to the distance to BS. The nodes nearer to BS have a higher probability which means more CH occupies smaller cluster area and less number of cluster members. This results to less energy dissipation for intra-cluster traffic and saves energy for inter-cluster routing. Tentative CHs are chosen based on the nodes residual energy to equalize the energy consumption. The multi-hop data transmission depends on unequal clustering algorithm and energy balancing layering algorithm. CH selects the relay node based on the residual energy. When two nodes have equal residual energy, CH chooses the relay node randomly from these nodes. EB-UCP achieves maximum lifetime when compared to LEACH and EEUC.

Unequal Hierarchal Energy Efficient Distributed Clustering (UHEED) : UHEED is a distributed approach which resolves the hotspot problem and maintains the same amount of residual energy in the network and maximizes the network lifetime [37]. It is the improved version of HEED. HEED uses two parameters for clustering: residual energy and node degree. The CHs use the same competition radius and distance to BS is not considered. This leads to hot spot problem. UHEED constructs various size clusters depending upon the distance of CHs to BS and smaller clusters are formed for CHs nearer to BS. This method also finds the approximate cluster size using the distance information. A competition radius formula is used to create smaller size cluster near BS. This result to reduced energy consumption in smaller clusters due to intra-cluster traffic and preserve more energy to focus on more inter-cluster traffic. This algorithm results to increased network lifetime compared to LEACH, HEED and Unequal LEACH.

Constructing Optimal Clustering Architecture (COCA) : COCA is a scalable, distributed unequal clustering scheme which investigates the logical difficulties in unequal clustering methods in homogeneous sensor network [38]. COCA constructs optimal clustering architecture, energy-aware CH rotation and routing for approximate equalization of energy consumption in the whole network. COCA uses a strategy that the number of the cluster in the unit area increases when the distance to BS decreases. It saves more energy for multi-hop routing between clusters. For CH selection, all nodes exchange residual energy information with its neighbors. The node with maximum residual energy declared them as CH. Each CH randomly choose some CH in the neighbor cluster as routing candidates and CH with highest residual energy is elected as final routing CH. It is a simple and effective protocol with reduced energy consumption and maximum lifetime twice or thrice than UCR.

4.2. **Deterministic approaches.** On contrast to probabilistic approaches, deterministic approaches use standard metric for selecting CHs. The commonly used traditional metrics are residual energy, node degree, expected residual energy, distance to BS, node centrality, etc. which are attained locally. This information is usually updated by exchanging message between its neighbors. This approach is called as deterministic approaches because the clusters with elected CHs are more controllable. This is again is classified into four categories: Weight based, Fuzzy based, heuristic based and Compound unequal clustering algorithm. In weight based approach, weight is calculated at each node based on some metrics such as residual energy, node degree, distance to BS, etc. The node with minimal weight is elected as a cluster head. Fuzzy logic is used to elect CHs in situations where uncertainties are more. The cluster head is chosen based on fuzzy input parameters. The input parameters can be residual energy, node degree, distance to BS, node centrality, etc. and the fuzzy output parameters are cluster size and probability of becoming CHs. The clustering problem in WSN is considered to be a NP-hard problem and evolutionary algorithms are suitable to produce optimal solutions for NP hard problems. In the recent years, heuristic based clustering algorithms provide optimal solution in the process of selecting CHs and cluster size. Many optimization algorithms such as Genetic Algorithm (GA), Ant Colony Optimization (ACO), Artificial Bee Colony Optimization (ABC), Particle Swarm Optimization (PSO), Bacterial Foliage Algorithm (BFA), Differential Evolution (DE), Simulated Annealing, etc are used in WSN. Each algorithm defines various metrics in fitness function to achieve better performance. Heuristic approaches are centralized, and a central authority like BS controls all operations in the network. In exceptional cases, some approaches works in distributed manner using agent nodes. The compound algorithm uses different metric like connected graph, Sierpinski triangle, etc. in clustering methods.

4.2.1. *Weight based unequal clustering algorithms.* They include as follows.

Multi-hop Routing Protocol with Unequal Clustering (MRPUC) : MRPUC is a distributed approach aims to equalize the energy consumption in all nodes and forward the data to BS with relays to reduce energy consumption [39]. MRPUC selects CH with more residual energy and the regular nodes joins the cluster as cluster members in which the CH contains maximum residual energy and lesser distance to BS. Relay nodes are chosen based on minimal energy consumption to relay data and the CHs with more residual energy. Intercluster routing tree is formed as a network backbone, and multi-hop routing is done for effectively transmitting data from CH to BS. The network lifetime is increased by 34.4% when compared to equal clustering schemes.

Partition Energy Balanced and Efficient Clustering Scheme (PEBECS) : PEBECS mitigates the hot spot problem and uniformly distributes the load throughout the network. It divides the sensing field into equal partitions and groups the nodes in the partition to clusters of unequal sizes[40]. CH selection is done by a weight based heuristic algorithm based on residual energy, node degree and distance to BS. A CH competition radius algorithm is introduced to assign number of cluster members in each cluster. The CH in the partition near to BS is smaller in size to reduce intra-cluster traffic and save energy for intercluster traffic. It effectively balances the energy consumption, improves scalability and lengthens the network lifetime significantly.

Arranging cluster size and data transmission WSN (ACT) : ACT is a distributed routing protocol which reduces the cluster size closer to BS, where the CHs spend more energy for inter-cluster multi-hop routing [41]. By this method, every CH spends the approximately same amount of energy and CHs next to BS does not drain out of energy quickly. To prolong the network lifetime, ACT partitions the network into a number of hierarchical levels. ACT calculates the cluster radius mathematically to construct unequal size clusters, and cross-level data transmission is proposed to improve the network lifetime. ACT successfully shows the energy consumptions that CHs has the same amount of energy dissipation throughout the network. **Energy Balancing Unequal Clustering Approach for Gradient-based routing (EBCAG)** : EBCAG is a distributed clustering approach which balances and reduces the energy consumption of all CHs [42]. In EBCAG, each sensor node holds a gradient value, which represents the minimum number of hop count to BS. It gives the optimal cluster radii based on the gradient value. Tentative CHs are randomly selected with a probability T (predefined threshold). Tentative CHs whose residual energy is maximum is selected as final CHs. Based on the CHs gradient value, the unequal cluster radius is constructed. The CH collects the data from its cluster members and forwards the aggregated data to BS based on the descending gradients of CH. EBCAG successfully attains energy equalization among CHs and improves the network lifetime significantly when compared to EDUC and HEED.

Enhanced Unequal Clustering Scheme (EUCS) : Existing clustering approaches selects CHs with maximum residual energy and rotates the CH periodically for uniform load distribution in the cluster. As a result, more energy is wasted for CH re-election and gathering data from its cluster members. To preserve energy in the process of reclustering and eliminating hot spot issue in multihop WSN, EUCS is proposed. It organizes the nodes into unequal size clusters [43]. In EUCS, CHs are selected based on two metrics namely, residual energy and distance to BS. Re-election takes place when the energy level of the current CH exceeds the threshold value. This minimizes the frequent re-election of

CH and eliminates overhead. For inter-cluster routing, a new parameter called threshold distance is introduced. When the distance of BS from CH is below a threshold value, it sends the data directly. In some situations where the distance is high, relay nodes are chosen based on residual energy. EUCS has improved network lifetime when compared to HEED.

Energy and proximity based unequal clustering algorithm (EPUC) : EPUC balances the energy consumption of all CHs by spatially distributing the clusters. In EPUC, the area is divided into tracks around the BS. Nodes located at the same track create clusters of same size [44]. In EPUC, small cluster are formed in the proximity of the BS. EPUC uses two CH distance metrics: distance to BS and distance to its neighboring nodes within a cluster. In CH selection process, nodes with maximum residual energy is chosen as candidates CH (CCHs) and distance metric rules are applied to select final CHs from CCHs. EPUC achieves maximum lifetime compared to UCR and EEDC.

Coverage aware and Unequal Clustering Algorithm (CUCA) : CUCA is a first coverage based distributed scheme which uses unequal clustering and single hop communication [45]. The major aim of CUCA is uniform load distribution among the CHs. For a maximum lifetime, the node whose sensing range is completely covered by the sensing range by its neighbors is given higher priority to be selected as CHs over the partially overlapped sensor nodes. The cluster radius is computed by the distance of the node from CH where the cluster radius is inversely proportional to the distance from CH. Unequal Clustering in CUCA eliminates the hotspot problem. The cluster member whose sensing area is fully covered by its neighbors is moved to sleep mode to avoid redundant data and minimizes energy consumption significantly. When the overlapped nodes are not present, CHs are selected based on partially overlapped area and residual energy. Hence, the result of the node death does not disrupt the coverage of the deployed region. Moreover, the nodes having less residual energy and the covering area is completely covered by its neighbor are moved to sleep state to save energy. CUCR algorithm achieves maximum lifetime compares to CA and HEED.

Improved Energy-Aware Distributed Unequal Clustering for heterogeneous WSN (Improved EADUC) : Improved EADUC aims to lengthen the lifetime of EADUC and avoid hot spot problem in multi-hop heterogeneous WSN. It is widely used in continuous data gathering applications [46]. It differs from EADUC is the consideration of node degree while computing competition radius in addition to residual energy and distance to BS. Node degree is included for proper energy balancing in the network. In Improved EADUC, CHs are selected based on the ratio of the average energy of neighbor nodes and residual energy of the node itself. The competition radius is calculated based on three metrics: residual energy, distance to BS and node degree. Improved EADUC uses energy as a relay metric for selecting relay nodes while EADUC uses distance to BS as a relay metric. For several rounds, same cluster setup is used which eliminates re-clustering overhead and minimizes energy consumption. Improved EADUC achieves maximum lifetime than EADUC and HUCL

4.2.2. *Fuzzy based unequal clustering algorithm.* Because of uncertainties occurring in WSN environment, a number of protocol uses fuzzy logic for making decisions effectively. Fuzzy has many advantages over classical approaches which include: low computational complexity, more flexibility, less development cost, less memory, less design time and fault tolerant. In WSN, Fuzzy logic is used for selecting CHs effectively. Fuzzy logic for CH selection uses input parameters like residual energy, distance to BS, distance from neighbors, node degree, centrality, excepted residual energy and the output parameters are CH selection probability and cluster size. Various fuzzy based approaches are proposed to attain energy efficiency in unequal clustering algorithm[14].

Energy Aware Fuzzy Unequal Clustering algorithm (EAUCF) : EAUCF is also a distributed approach used to lengthen the network lifetime like EEUC[44]. In EAUCF, tentative CHs are randomly selected, and residual energy is used to elect final CHs. EAUCF mainly focused on determining the competition radius of CHs to evenly distribute the load. EEUC uses residual energy to determine the competition radius while EAUCF uses fuzzy logic to compute the competition radius based on residual energy and distance to BS. When the tentative CH residual energy and distance to BS is high, then the competition radius is large and vice versa. It is more stable and attains maximum lifetime when compared to LEACH CHEF and EEUC.

Improved Fuzzy Unequal Clustering algorithm (IFUC) : IFUC is a distributed approach which reduces the energy consumption and lengthens the network lifetime[47]. Fuzzy logic is used for electing CHs and determining the radius of the cluster. The fuzzy input parameters are residual energy, distance to BS and node density. The output parameters are the probability of becoming CH and cluster radius. After each tentative CH determines the chance of becoming final CHs and competition radius, they compete for final CHs by exchanging messages. If a tentative CH finds other CH within its range and higher chance, it is selected as final CH. Ant Colony Optimization (ACO) finds the shortest path for inter-cluster routing (Dorigo and Gambardella, 1997).ACO is used for energy aware inter-cluster routing to provide an optimal path from CH to BS. In ACO, the next hop route node is selected based on the communication cost and degree of energy consumption along the path. IFUC provides improved network lifetime over LEACH and EEUC.

Fuzzy based Unequal Clustering Protocol (FUCP) : FUCP is a novel clustering algorithm which incorporates a novel CH selection algorithm and relays traffic distributed algorithm to eliminate hot spot problem [48]. Fuzzy Inference System (FIS) is used for selecting CHs in a distributed manner. Mamdani method is used for fuzzification. FUCP uses three input parameters for CH selection: residual energy, the centrality of the node and Link Quality Indicator (LQI). Usage of LQI in CH selection increases the reliability of WSN. The fuzzy output parameter is the probability of becoming CH. To avoid hot spot problem, FUCP determines the number of CH in 0ppp0a layered hexagon using scatter factor and distance of hexagon from BS. Node scatters factor defines the average distance of each node to its neighbor node in the same hexagon. High scatter factor represents the node are scattered from each other and requires more CHs to concentrate intra-cluster communication. FUCP achieves 40% more energy-efficient, 30% maximum lifetime and 57% more packet transmission to Distributed Energy Efficient Hierarchical Clustering. **Fuzzy Logic Based Unequal Clustering (FBUC)** FBUC is also a distributed clustering algorithm which concentrates on the method of joining cluster members with the CH [49]. FBUC is the improved version of EAUCF. Tentative CHs are selected based on a probabilistic method. After the selection of tentative CHs, fuzzy logic is used to determine competition radius. The fuzzy input parameters for calculating competition radius are residual energy, distance to BS and node degree. Residual energy and node degree is used to elect final CHs. The nodes join the CH based on the CH degree and distance to CH to effectively utilize the energy and maximize the network lifetime. It achieves maximum lifetime when compared to LEACH and EAUCF. **Distributed Load Balancing Unequal Clustering in Wireless Sensor Networks using Fuzzy approach (DUCF)** DUCF is a distributed approach for uniform load distribution which balances the load among the clusters and extends the network lifetime [50]. It uses fuzzy for CH selection and to determine the cluster size. To avoid hot spot problem, smaller clusters are formed for CHs closer to BS and larger cluster size for CHs farther from BS. The three input parameters to FIS are residual energy, node degree, and distance to BS. The probability of becoming CH and cluster size are the two fuzzy output parameters. Mamdani method is used for fuzzification and centroid method is used for defuzzification. The output parameter size restricts the number of cluster members in cluster results to proper load balancing. Multi-hop data transmission for inter-cluster routing which also reduces the energy consumption. DUCF performs well when compared to LEACH, CHEF, and EAUCF.

4.2.3. *Heuristic based unequal clustering algorithm.* They comprise as follows.

Energy Balanced Unequal Clustering for Wireless Sensor Networks (EBUC): EBUC is a centralized unequal clustering protocol commonly used in periodical data gathering applications [51]. It uses PSO algorithm to select candidate CHs and divide the network into various sized clusters. The multi-hop inter-cluster routing protocol is based on a greedy algorithm, and residual energy and distance to BS is used to choose a relay node. This algorithm reduces the rate of dead nodes and lengthens the network lifetime.

Genetic Algorithm based Energy-Efficient Adaptive Clustering Hierarchical Protocol (GAEEP) GAEEP uses Genetic Algorithm (GA) to determine the number and position of CHs to reduce the energy consumption and improve the stability of WSN[52]. The entire operation undergoes many rounds; each round consists of two phases namely setup phase and steady state phase. In the setup phase, BS runs GA and determines the optimal number of CHs and the position of CH. The inter-cluster routing from CH to BS takes place in steady state phase. When a node is located much closer to BS than any CH, the node directly transmits the data to BS. Each CH uses TDMA schedule and assign slots to its cluster members to avoid collision due to intracluster communication. To reduce inter-cluster collision, CHs uses CDMA code to reduce energy consumption. GAEEP is energy-efficient and more reliable than LEACH, SEP, ERP, LEACH-GA and DEU in both homogeneous and heterogeneous networks.

Unequal Clustering by Improved Particle Swarm Optimization: Improved Particle Swarm Optimization (IPSO) is used in Energy Balanced Unequal Clustering (EBUC) to improve the performance and maximize the network lifetime [53]. This algorithm is a distributed scheme which eliminates hot spot problem and also overcomes standard PSO issues. EBUC algorithm operates in several rounds, and each round starts with a setup phase followed by a steady state phase. Each node sends its information (node id, energy level, location, etc.) to BS and BS runs IPSO algorithm to select the best candidate CHs. Cluster formation and CH selection takes place in the setup phase. TDMA schedule is prepared by CHs during setup phase to avoid collision due to intra-cluster communication. Cluster members send data in their respective slot, and it forwards the data to BS. IPSO eliminates the standard PSO issues and number of the dead node is significantly reduced in several rounds.

Sink Mobility based Energy Balancing Unequal Clustering Protocol (SMEBUC): SMEBUC is proposed to balance the energy consumption, and it uses Shuffled Frog Leaping Algorithm (SFLA) to elect CHs and form clusters of various sizes according to residual energy [54]. To minimize the reclustering overhead, CH continuously works to find the exchange time and node weight. There are two stages in SMEBUC: cluster establishment and data transmission. N cluster heads to be selected from M nodes leads to the optimization problem. Improved SFLA algorithm is used to partition clusters of various sizes. The greedy algorithm finds the relay node for multi-hop inter-cluster routing.

Table 2. Comparison of the unequal clustering algorithms

Schemes	**CH Selection**	**Consumed Energy**	**Proactive Nature**	**Balanced Load**	**MultiHop**	**Homogeneous**
PRODUCE	Random	High	Yes	Medium	Yes	Yes
EDUC	Random	Medium	Yes	Bad	Yes	No
LUCA	Random	Low	Yes	Medium	Yes	Yes
EEUC	Hybrid	Medium	Yes	Medium	Yes	Yes
EEDUC	Hybrid	Medium	Yes	Medium	Yes	Yes
UCR	Hybrid	Medium	Yes	Medium	Yes	Yes
ULEACH	Hybrid	Very low	Yes	Medium	Yes	Yes
EC	Hybrid	High	Yes	Good	Yes	Yes
EB-UCP	Hybrid	Medium	Yes	Good	Yes	Yes
UHEED	Hybrid	Medium	Yes	Medium	Yes	Yes
COCA	Hybrid	Medium	Yes	Medium	Yes	Yes
MRPUC	Deterministic	Very Low	Yes	Bad	Yes	Yes
PEBECS	Deterministic	Medium	Yes	Medium	Yes	Yes
EADUC	Deterministic	Low	Yes	Good	Yes	Yes
ACT	Deterministic	High	Yes	Bad	Yes	Yes
EBCAG	Deterministic	Medium	Yes	Bad	Yes	Yes
EUCS	Deterministic	High	Yes	Medium	Yes	Yes
UCMR	Deterministic	Low	Yes	Medium	Yes	Yes
EPUC	Deterministic	Medium	Yes	Good	Yes	Yes
CUCA	Deterministic	Medium	Yes	Good	Yes	Yes
EADUC	Deterministic	Medium	Yes	Bad	Yes	Yes
EAUCF	Fuzzy	Medium	Yes	Medium	Yes	Yes
IFUC	Fuzzy	High	Yes	Bad	Yes	Yes
FUCP	Fuzzy	High	Yes	Medium	Yes	Yes
FBUC	Fuzzy	Medium	Yes	Medium	Yes	Yes
DUCF	Fuzzy	Medium	Yes	Medium	Yes	Yes
EBUC	Heuristic	High	Yes	Medium	Yes	Yes
GAEEP	Heuristic	Medium	Yes	Good	Yes	Yes
SMEBUC	Heuristic	Medium	Yes	Good	Yes	Yes
IPSO	Heuristic	High	Yes	Good	Yes	Yes
FAMACROW	Heuristic	High	Yes	Medium	Yes	Yes
nCRO-UCRA	Heuristic	High	Yes	Medium	Yes	No
UMBIC	Heuristic	Medium	Yes	Good	Yes	Yes
EDDUCA	Compound	Medium	Yes	Good	Yes	Yes
UCCGRA	Compound	High	Yes	Medium	Yes	Yes

To eliminate hot spot problem, sink mobile location algorithm is used. This algorithm achieves less energy dissipation and efficient balancing than LEACH and EBUCP.

Fuzzy and ant colony optimization based combined Mac, routing and unequal clustering cross-layer protocol for wireless sensor networks (FAMACROW) FAMACROW is a cross-layer hierarchical protocol which includes: CH selection, unequal cluster formation and inter-cluster multi-hop routing protocol[55]. FAMACROW consists of three phases: setup phase, neighbor finding phase, and steady state phase. During the setup phase, the nodes are built into layers. In neighbor finding phase, each nodes broadcast its details using non-persistent CSMA MAC protocol. The steady state phase includes CH selection, clustering, and data delivery. CHs are selected by fuzzy logic and ACO is used for intercluster multi-hop routing. Fuzzy logic uses three input parameters to select CHs which includes residual energy, the number of neighboring nodes and quality of communication link. The reliability is increased by the inclusion of link quality in CH selection. ACO chooses the relay node based on four parameters: residual energy, distance to BS, queue length and delivery likelihood. FAMACROW achieves 41% more energy-efficient, 63% more network lifetime, 15% more throughput than IFUC.
Novel Chemical reaction optimization based unequal clustering and routing algorithm for Wireless Sensor Networks (nCRO-UCRA) nCRO-UCRA is a distributed approach eliminates the hot spot problem by combining novel chemical reaction optimization (nCRO) paradigm with unequal clustering and routing algorithm[56]. nCRO algorithm is used for selecting CHs and cluster members joins CH based on a derived cost function. The CH forwards he data to BS using UCRA algorithm. The algorithm is developed with efficient methods of molecular structure encoding and novel potential energy functions. nCRO-UCRA achieves better performance in terms of residual energy, lifetime, number of alive nodes and convergence rate compared to CRO-UCRA.

An Unequal Multi-hop Balanced Immune Clustering protocol for wireless sensor networks (UMBIC) UMBIC uses Unequal Clustering Mechanism (UCM) and Multi-Objective Immune Algorithm (MOIA) for adjusting the energy dissipation of inter-clusters and intra-clusters [57]. UCM organizes the clusters of various size based on the distance to BS and residual energy. MOIA produces optimal clusters and routing tree is constructed for covering the entire region, ensuring connectivity and low communication cost of all nodes. CH rotation takes place when the present CH residual energy becomes lesser than the threshold energy value. UMBIC avoids the hot spot problem with less overhead and computational complexity.

4.2.4. *Compound unequal clustering algorithm.* There are algorithms as follows.

Energy Degree Distance Unequal Clustering Algorithm (EDDUCA) EEDUCA algorithm is proposed to approximate equalization of energy consumption, improves the network lifetime and also eliminates the hot spot problem [58]. There are three phases in EDDUCA: cluster formation phase, CH selection phase, and data transmission phase. In cluster formation phase, Sierpinski triangle divides the network into unequal clusters. The size of the clusters is same when the distance to the BS is same. In CH selection phase, CHs are selected based on residual energy, node degree, and distance to BS. A node weight is calculated based on the above three parameters in each cluster, and the node with minimal weight is elected as CH. Each node forwards data to CH and CH transmits to its upper cluster until it reaches the BS. It performs well and lengthens the lifetime when compared to ECLEACH.

Energy-Efficient Routing Algorithm Based on Unequal Clustering and Connected Graph in Wireless Sensor Networks (UCCGRA) UCCGRA is a distributed approach improves the energy efficiency in two methods: cluster head election and clusters routing. A voting scheme is used to construct unequal size clusters, and smaller clusters are constructed near BS to reduce intra-cluster traffic [59]. CH selection is mainly based on topology, residual energy, and transmission power. For inter-cluster multi-hop communication, Connected graph based routing is used which utilizes the geographic location of the nodes. UCCGRA distributes the load effectively, reduces energy consumption and maximizes the network lifetime than UCRA, EEUC and HEED.

4.3. **Comparison Unequal Clustering Algorithms.** Table 2 shows a comparison of some reviewed unequal clustering approaches. Various algorithms are compared based on cluster properties which include a method (probabilistic, deterministic), consumed energy, intra-cluster communication and inter-cluster communication (direct or multi-hop). The algorithms also are compared based on clustering process which includes the way of CH selection, nature (proactive or reactive). Unequal clustering protocols are classified into the classes: probabilistic, and deterministic.

5. **Conclusion.** Wireless sensor network (WSN) is an integral part of the Internet of Things (IoT), and it makes huge of devices to share data for improving the environmental user control. As sensor nodes are energy constrained, reducing the energy consumption and maximizing the network lifetime are the major research challenge in WSN. Clustering formation is the modern energy efficient techniques, but it suffers from hot spot problem and introduces connectivity issues in the network. Unequal Clustering evenly distributes the load, eliminates the hot spot problem and maximizes the network lifetime. Many advanced algorithms are for energy efficient WSN via equal and unequal grouping. In this paper, we explained various equal and unequal clustering algorithms with their objectives, characteristics, classification, merits, and demerits. All the reviewed algorithms are also compared based on different cluster properties, Cluster Head (CH) properties, and clustering process. We examined the literature of clustering formation and presented tables and discussion. The design of an appropriate equal and unequal clustering algorithms depends on the user and application requirements.

REFERENCES

[1] C. F.Garcia-hernndez, P. H.Ibargengoytia-gonzalez, J.Garca-hernandez, and J. Perez-daz, Wireless Sensor Networks and Applications: a Survey, *J. Comput. Sci.*, vol. 7, no. 3, pp. 264-273, 2007.

[2] I. F. Akyildiz, W. Su, Y.Sankarasubramaniam, and E.Cayirci, A survey on sensor networks, *IEEE Commun. Mag.* , vol. 40, no. 8, pp. 102-105, 2002.

[3] R. V. Kulkarni, A. Frster, G. K. Venayagamoorthy, Computational intelligence in wireless sensor net- works: A survey, *IEEE Commun. Surv. Tutorials,* vol. 13, no. 1, pp. 68-96, 2011.

[4] W. R. Heinzelman, A.Chandrakasan, H.Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci.,* vol. 0, no. c, pp. 3005-3014, 2000.

[5] S. Soro W. B.Heinzelman, Prolonging the lifetime of wireless sensor networks via unequal clustering, *in Proceedings - 19th IEEE International Parallel and Distributed Processing Symposium, IPDPS* 2005, vol. 2005.

[6] N. Sabor, S. Sasaki, M. Abo-zahhad, S. M. Ahmed, A Comprehensive Survey on Hierarchical-Based Routing Protocols for Mobile Wireless Sensor Networks: Review , Taxonomy , and Future Directions, *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 2818542, 23 pages https://doi.org/10.1155/2017/2818542vol.

[7] G. V. Selvi R. Manoharan, A Survey of Energy Efficient Unequal Clustering Algorithms for Wireless Sensor Networks, *International Journal of Computer Applications*, vol. 79, no. 1, pp. 1-4, 2013.

[8] L.Kleinrock F.Kamoun, Hierarchical Routing for Large Networks: Performance Evaluation and Optimization, *Comput. Networks,* vol. 1, no. 3, pp. 155-174, 1977.

[9] R.Kacimi, R.Dhaou, A. L.Beylot, Load balancing techniques for lifetime maximizing in wireless sensor networks, *Ad Hoc Networks*, vol. 11, no. 8, pp. 2172-2186, 2013.

[10] Y.Zhou, Y.Fang, Y.Zhang, Securing wireless sensor networks: a survey, *IEEE Commun. Surv. Tutorials*, vol. 10, no. 3, pp. 6-28, 2008.

[11] W. B.Heinzelman, A. P.Chandrakasan, H.Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, *IEEE Trans. Wirel. Commun.,* vol. 1, no. 4, pp. 660-670, 2002.

[12] M. Bagaa, Y. Challal, A. Ouadjaout, N. Lasla, N.Badache, Efficient data aggregation with in-network integrity control for WSN, *Journal of Parallel and Distributed Computing,* vol. 72, no. 10. pp. 11571170, 2012.

[13] T.-T.Nguyen, C.-S.Shieh, M.-F.Horng, T.-G.Ngo, T.-K.Dao, Unequal clustering formation based on bat algorithm for wireless sensor networks, *In: Nguyen VH., Le AC., Huynh VN. (eds) Knowledge and Systems Engineering. Advances in Intelligent Systems and Computing*, vol. 326. pp 667-678, 2015.

[14] T.-T.Nguyen, C.-S. Shieh, T.-K. Dao, J.-S. Wu, W.-C. Hu, Prolonging of the network lifetime of WSN using fuzzy clustering topology, *in Proceedings - 2013 2nd International Conference on Robot, Vision and Signal Processing, RVSP* 2013, 2013.

[15] O.Younis S.Fahmy, Distributed clustering in ad hoc sensor networks: A hybrid, energy-efficient approach, 23rd Annu. *Jt. Conf. IEEE Comput. Commun. Soc. INFOCOM 2004, Hong Kong*, vol. 3, no. 4, pp. 629-640, 2004.

[16] W. B.Heinzelman, A. P.Chandrakasan, H.Balakrishnan, An application-specific protocol architecture for wireless microsensor networks, *IEEE Trans. Wirel. Commun.,* vol. 1, no. 4, pp. 660-670, 2002.

[17] A.Manjeshwar D. P.Agrawal, TEEN: A Routing Protocol for Enhanced Efficiency in Wireless Sensor Networks, *in Proceedings 15th International Parallel Distributed Processing Symposium,* 2001, pp. 2009-2015.

[18] T.-T. Nguyen, T.-K.Dao, M.-F. Horng, C.-S. Shieh, An Energy-based Cluster Head Selection Algorithm to Support Long-lifetime in Wireless Sensor Networks, *J. Netw. Intell.,* vol. 1, no. 1, pp. 2337, 2016.

[19] Y.Xu, J. Heidemann, D.Estrin, Geography-informed energy conservation for Ad Hoc routing, *Proc. 7th Annu. Int. Conf. Mob. Comput. Netw. - MobiCom 01* , pp. 70-84, 2001.

[20] P.Schaffer L.Buttyn, Position-based aggregator node election in wireless sensor networks, *Int. J. Distrib. Sens. Networks*, vol. 2010, 2010.

[21] A. G.Delavar, S.Shamsi, N.Mirkazemi, J.Artin, SLGC: A New Cluster Routing Algorithm in Wireless Sensor Network for Decrease Energy Consumption, *Int. J. Comput. Sci. Eng. Appl.,* vol. 2, no. 3, p. 39, 2012.

[22] H.Luo, F.Ye, J.Cheng, S. Lu, L.Zhang, TTDD: Two-tier data dissemination in large-scale wireless sensor networks, *in Wireless Networks,* 2005, vol. 11, no. 12, pp. 161-175.

[23] S.Lindsey, C.Raghavendra, K. M.Sivalingam, Data gathering algorithms in sensor networks using energy metrics PEGASIS , IEEE Trans. Parallel Distrib. Syst., vol. 13, no. 9, pp. 924-935, 2002.

[24] S. M.Jung, Y. J.Han, T. M.Chung, The concentric clustering scheme for efficient energy consumption in the PEGASIS, *in International Conference on Advanced Communication Technology, ICACT,* 2007, vol. 1, pp. 260-265.

[25] N.Gautam, W.-I. L. W.-I.Lee, J.-Y. P. J.-Y.Pyun, Track-Sector Clustering for Energy Efficient Routing in Wireless Sensor Networks, 2009 Ninth IEEE Int. *Conf. Comput. Inf. Technol.,* vol. 2, pp. 116-121, 2009.

[26] F. Tang, I.You, S. Guo, M. Guo, Y. Ma, A chain-cluster based routing algorithm for wireless sensor networks, *J. Intell. Manuf.,* vol. 23, no. 4, pp. 1305-1313, 2012.

[27] T.-T.Nguyen, C.-S. Shieh, M.-F.Horng, T.-G. Ngo, T.-K. Dao, Unequal Clustering Formation Based on Bat Algorithm for Wireless Sensor Networks, in Knowledge Systems Engineering: Proceedings of the Sixth International Conference KSE 2014, V.-H.Nguyen, A.-C.Le, V.-N.Huynh, Eds.Cham: *Springer International Publishing*, 2015, pp. 667-678.

[28] J. H.Kim, C. S.Hussain, W. C.Yang, D. S.Kim, M. S.Park, Produce: A probability-driven unequal clustering mechanism for wireless sensor networks, *in Proceedings - International Conference on Advanced Information Networking Applications, AINA,* 2008, pp. 928-933.

[29] J.Yu, Y.Qi, G.Wang, An energy-driven unequal clustering protocol for heterogeneous wireless sensor networks, *J. Control Theory Appl.,* vol. 9, no. 1, pp. 133-139, 2011.

[30] S.Lee, H.Choe, B.Park, Y.Song, C.Kim, LUCA: An Energy-efficient Unequal Clustering Algorithm Using Location Information for Wireless Sensor Networks, *Wirel. Pers. Commun.,* vol. 56, no. 4, pp. 715-731, 2011.

[31] C.Li, M.Ye, G.Chen, J.Wu, An energy-efficient unequal clustering mechanism for wireless sensor networks, *in 2nd IEEE International Conference on Mobile Ad-hoc Sensor Systems, MASS 2005,* vol. 2005, pp. 597-604.

[32] J.Yu, Y.Qi, G.Wang, Q.Guo, X.Gu, An Energy-Aware Distributed Unequal Clustering Protocol for Wireless Sensor Networks, *Int. J. Distrib. Sens. Networks*, vol. 7, no. 1, p. 202-145, 2011.

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

[34] P.Ren, J.Qian, L.Li, Z.Zhao, X.Li, Unequal clustering scheme based LEACH for wireless sensor networks, *in Proceedings - 4th International Conference on Genetic Evolutionary Computing, ICGEC* 2010, 2010, pp. 90-93.

[35] D. Wei, Y.Jin, S.Vural, K.Moessner, R.Tafazolli, An energy-efficient clustering solution for wireless sensor networks, *IEEE Trans. Wirel. Commun.*, vol. 10, no. 11, pp. 3973-3983, 2011.

[36] J.Yang D. Zhang, An energy-balancing unequal clustering protocol for wireless sensor networks, *Inf. Technol. J.,* vol. 8, no. 1, pp. 57-63, 2009.

[37] E.Ever, R.Luchmun, L.Mostarda, A.Navarra, P.Shah, UHEED - an unequal clustering algorithm for wireless sensor networks, *Sensornets* 2012, 2012.

[38] H.Li, Y.Liu, W.Chen, W.Jia, B.Li, J.Xiong, COCA: Constructing optimal clustering architecture to maximize sensor network lifetime, *Comput. Commun.,* vol. 36, no. 3, pp. 256-268, 2013.

[39] B. Gong, L. Li, S. Wang, X. Zhou, Multihop Routing Protocol with Unequal Clustering for Wireless Sensor Networks, *in 2008 ISECS International Colloquium on Computing, Communication, Control, Management*, 2008, pp. 552-556.

[40] Y. Wang, T. L. X. Yang, D. Zhang, An energy efficient balance hierarchical unequal clustering algorithm for large scale sensor networks, *Inf. Technol. J.,* vol. 8, no. 1, pp. 28-38, 2009.

[41] W. K.Lai, C. S.Fan, L. Y.Lin, Arranging cluster sizes transmission ranges for wireless sensor networks, *Inf. Sci. (Ny)* ., vol. 183, no. 1, pp. 117-131, 2012.

[42] T. Liu, Q. Li, P.Liang, An energy-balancing clustering approach for gradient-based routing in wireless sensor networks, *Comput. Commun.,* vol. 35, no. 17, pp. 2150-2161, 2012.

[43] M. Mohamed-Lamine, New clustering scheme for wireless sensor networks. *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, Algiers, 2013, pp. 487-491. doi: 10.1109/WoSSPA.2013.6602412.

[44] M. M.Afsar M.Younis, An energy- proximity-based unequal clustering algorithm for Wireless Sensor Networks, *in 39th Annual IEEE Conference on Local Computer Networks*, 2014, pp. 262-269.

[45] N.Mazumdar H.Om, Coverage-aware Unequal Clustering Algorithm for Wireless Sensor Networks,*in Procedia Computer Science,* 2015, vol. 57, pp. 660-669.

[46] V. Gupta R.P ey, An improved energy aware distributed unequal clustering protocol for heterogeneous wireless sensor networks, *Eng. Sci. Technol. an Int. J.,* vol. 19, no. 2, pp. 1050-1058, 2016.

[47] S. Mao, C. Zhao, Z. Zhou, Y.Ye, An improved fuzzy unequal clustering algorithm for wireless sensor network, *Communications Networking in China (CHINACOM),* 2011 6th International ICST Conference on. pp. 245-250, 2011.

[48] S. Gajjar, M. Sarkar, K. Dasgupta, FAMACRO: Fuzzy ant colony optimization based MAC/routing cross-layer protocol for wireless sensor networks, *in Procedia Computer Science*, 2015, vol. 46, pp. 1014-1021.

[49] R. Logambigai A.Kannan, Fuzzy logic based unequal clustering for wireless sensor networks, *Wirel. Networks,* vol. 22, no. 3, pp. 945-957, 2016.

[50] B. Baranidharan B.Santhi, DUCF: Distributed load balancing Unequal Clustering in wireless sensor networks using Fuzzy approach, *Appl. Soft Comput. J.,* vol. 40, pp. 495-506, 2016.

[51] C. J.Jiang, W. R.Shi, M. Xiang, X. L.Tang, Energy-balanced unequal clustering protocol for wireless sensor networks, *J. China Univ. Posts Telecommun.,* vol. 17, no. 4, pp. 94-99, 2010.

[52] M.Abo-zahhad, S. M.Ahmed, N. Sabor, A New Energy-Efficient Adaptive Clustering Protocol Based on Genetic Algorithm for Improving the Lifetime the Stable Period of Wireless Sensor Networks, *Int. J. Energy, Inf. Commun.,* vol. 5, no. 3, pp. 47-72, 2014.

[53] S. Salehian S. K. Subraminiam, Unequal Clustering by Improved Particle Swarm Optimization in Wireless Sensor Network, *Procedia Comput. Sci.,* vol. 62, no. Scse, pp. 403-409, 2015.

[54] X.Fan F.Du, Shuffled frog leaping algorithm based unequal clustering strategy for wireless sensor networks, *Appl. Math. Inf. Sci.* , vol. 9, no. 3, pp. 1415-1426, 2015.

[55] S. Gajjar, M. Sarkar, K. Dasgupta, FAMACROW: Fuzzy ant colony optimization based combined mac, routing, unequal clustering cross-layer protocol for wireless sensor networks, *Appl. Soft Comput. J.,* vol. 43, pp. 235-247, 2016.

[56] P. C. Srinivasa Rao H.Banka, Novel chemical reaction optimization based unequal clustering routing algorithms for wireless sensor networks, *Wirel. Networks,* pp. 1-20, 2016.

[57] N.Sabor, M.Abo-Zahhad, S. Sasaki, S. M. Ahmed, An Unequal Multi-hop Balanced Immune Clustering protocol for wireless sensor networks, *Appl. Soft Comput. J.,* vol. 43, pp. 372-389, 2016.

[58] A. B. F.Guiloufi, N. Nasri, A. Kachouri, An Energy-Efficient Unequal Clustering Algorithm Using Sierpinski Triangle for WSNs, *Wirel. Pers. Commun.,* vol. 88, no. 3, pp. 449-465, 2016.

[59] H. Xia, R. huaZhang, J. Yu, Z. kuan Pan, Energy-Efficient Routing Algorithm Based on Unequal Clustering Connected Graph in Wireless Sensor Networks, *Int. J. Wirel. Inf. Networks,* vol. 23, no. 2, pp. 141-150, 2016.