HFLBSC: Heuristic and Fuzzy Based Load Balanced, Scalable Clustering Algorithm for Wireless Sensor Network

Priti Maratha1 ‧ Kapil Gupta1

Accepted: 7 January 2022 / Published online: 21 January 2022 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
In spite of the severe limitations on the resources of the sensor nodes such as memory, computational power, transmission range and battery, the application areas of Wireless Sensor Networks (WSNs) are increasing day by day. The main challenge in WSNs is energy consumption. It becomes significant when a large number of nodes are deployed. Although clustering is one of the solutions to cater to this problem, but it suffers from severe energy consumption due to the non-uniform selection of CHs and frequent re-clustering. In this paper, we propose a heuristic and fuzzy based load balanced, scalable clustering algorithm for WSNs called HFLBSC. In this algorithm, we have segregated the network into a layered structure using the area under intersection over union curve. We have selected the CHs by considering residual energy and distance threshold. We have stalled the frequent re-clustering by utilizing the decision made with the help of fuzzy logic. Our proposed scheme is capable enough to elongate the network lifetime. Simulation results confirm that on an average, HFLBSC is 32% better in terms of FND, 38% less energy consumption, 25% more alive nodes, 72% less deviation in residual energy than LEACH, FM-SCHEL, and MIWOCA.

Keywords Load balancing ‧ Load ‧ Residual energy ‧ Network lifetime ‧ Wireless sensor networks

1 Introduction

Wireless Sensor Networks (WSNs) is a collection of tiny and cheap sensor nodes [1, 2]. These can be used in several environmental monitoring and control systems such as surveillance, industrial emissions, outdoor climate, indoor temperature and management of disasters [3–5]. The pollution of the atmosphere can be monitored and managed continuously from remote locations with high precision using small sensor nodes. Sensor nodes are usually distributed and have limited resources. Sensor nodes have to transmit the data to the BS along with minimizing the redundancy in the transmitted data [6].

* Priti Maratha
niki.maratha19@gmail.com

1 Department of Computer Applications, National Institute of Technology Kurukshetra, Kurukshetra, India
1.1 Problem Statement and Motivation

Energy efficiency is one of the significant factors that should be considered when designing a WSN. In large scale networks, it is not feasible to provide a recharging facility and to store all the sensed data [7, 8]. The entire network has to perform its function with the help of power-constrained batteries. There are several power-saving strategies suggested for WSNs to enhance battery power life. Energy efficiency can be achieved with the aid of the clustering mechanism in WSN [9, 10]. Clustering is a mechanism where the sensor nodes are clustered together to accomplish a task. For cluster-based WSN, the routing process is simpler and easier, compared with non-clustered WSN.

There are several clustering techniques suggested to enhance the sensor network lifetime. Cluster heads (CHs) enable a routing protocol to transfer the data correctly from sensor nodes to BS [11]. This is the duty of a routing protocol to determine the best route from sensor nodes to BS to reduce overall transmission costs. Because non-optimal routes consume significant energy. Then, in clustered protocols, the battery of sensor nodes which are near to the sink consumed earlier in comparison to the nodes which are away from the sink (as given in Fig. 1). Also, in most of the works suggested in the literature, re-clustering is done on each round which is an energy-consuming process and purely an overhead. Along-with non-uniform distribution of CHs in the sensing region can lead to energy unbalancing issue.

1.2 Contribution

In this paper, a layered structure-based energy-efficient and unequal clustering protocol is proposed. Unequal clustering plays a good role in avoiding the hotspot problem [12, 13]. Also, to reduce the energy consumption that occurred in re-clustering, we have specified a mechanism that decides when re-clustering needs to be performed. Along with it, uniform energy consumption is ensured by uniformly selecting the CHs in the whole network. The main advantage of the routing protocol is that energy consumption spent in re-clustering has reduced to a great extent. The vital contributions of our manuscript are as follows:
(a) We have introduced a novel way and logical way of layer formation in the network to ensure the scalability by using Intersection over Union by utilizing a unique value i.e. number of nodes with a particular load on particular hop counts in Sect. 5.2. To best of our knowledge, this type of layer formation in the network is never considered in the literature for the purpose of fuzzy-logic based clustering. And the experiments backed the proposed method which is again validated by the F-Test.

(b) Along with it, uniform clustering is ensured with a new approach in Sect. 5.3 of our paper.

(c) Also, a novel idea of minimizing the frequency of re-clustering using Update Cycle is introduced in Sect. 5.5 of our paper. To achieve it, we have used the lifetime and incoming load parameters and is mentioned in the manuscript as well.

(d) In Sect. 6, we have done the statistical analysis to find the significant difference between the different arrangements of layers.

(e) Further in Sect. 7, one can notice that the results of the simulation experiments conducted to compare with existing and recent solutions have come in favor of the proposed method.

1.3 Paper Organization

The paper is organized as follows: Sect. 2 summarizes the literature on different routing protocols that are energy-efficient. Assumptions and system model of the work are briefed in Sects. 3 and 4. Section 5 includes a thorough explanation of the algorithm proposed. Statistical analysis and simulation results are portrayed respectively in Sects. 6 and 7 to demonstrate the efficiency of the proposed algorithm. Conclusion and future scope are discussed in Sect. 8.

2 Related Work

In literature, a lot of work has been done to make sensor networks as energy efficient. Clustering is the prominent solution to minimize the dissipation of energy to sustain the network alive for a longer time. In LEACH [14], there is single-hop communication between sensor nodes and the BS. Along with that, residual energy is not considered while selecting the CHs and uniform selection of CHs is not ensured. Authors in [15] presented a hierarchical routing protocol which aims to elongate the network lifetime. It reduces communication complexities. High energy nodes among the competitor nodes in a cluster are selected as CHs. Reduced energy consumption is confirmed when re-clustering is done after a specific time. A multi-level and distance-based clustering mechanism (EEMDC) proposed in [16] to make the protocol energy-efficient. The authors have proposed to divide the network area into 3 logical layers which are divided by area factor. Hop count is used to deciding the layers, however there is no concrete basis to decide the layers. Distance from BS is used basically to determine the hop count.

In conventional clustering algorithms of WSNs, residual energy and the euclidean distance between the sensor nodes and sink is not taken into consideration while selecting the CHs. An energy-efficient scheme is built in [17] that accounts these two factors in choosing the CHs. Nodes that are having less energy than total energy are not chosen as CHs. Battery power of the nodes is mainly consumed during transmission. It is minimized by choosing the relay nodes using neighboring list, whose remaining battery power is more
than a given battery threshold. But it suffers from frequent re-clustering. In [18], the main attention is given on reducing the uneven power consumption. Candidate CHs are selected from the nodes whose residual energy is greater than the average energy of the network. Thereby, requiring the information of residual energy of the neighboring nodes. A proportion value is estimated by considering the surplus energy and euclidean distance between CHs and cluster members. The least value of this proportion is used to select the CHs. This proportion is adjusted with node density, which inherently requires the re-clustering in each round. To mitigate the issues of energy consumption in WSNs, a particle swarm optimization (PSO) based clustering protocol is suggested [19]. CHs are selected using PSO based approach. The fitness value of the particles is evaluated based on the remaining energy of that node and the distance to the base station. But, re-clustering is done after every round, which is purely an overhead process and reduces network lifetime. To maximize the energy efficiency, authors in [20] proposed a protocol, named HEEMP, in which CHs are selected using the chance of election value. This value is determined using node density and remaining energy. Sink node prepares an admissible route set. This route set is the set of routes on which battery power of all the nodes is more than a particular threshold. Data transmission is done once the routes are set-up. Data is forwarded to the BS based on residual energy or distance. HEEMP has not considered the stability aspect.

Energy efficiency and network lifetime are enhanced using a fuzzy based unequal clustering algorithm (FBUCA) and a 3-level multi-hop optimized routing method [21]. Load balanced clustering has been done by doing the fuzzy-logic based clustering in which residual energy, distance to the BS, and degree centrality are considered for CHs selection. Three level multi-hop optimized routing helps in achieving data reliability. Simulation results suggest that although load balancing has been achieved, but re-clustering has been done after every round. So, there is a scope of improving the energy efficiency is still present in FBUCA. A two-level unequal clustering protocol on IoT is introduced by sharing the incoming traffic to enhance the system performance [22]. The goal is to minimize the number of CHs. To achieve the same, IoT devices are grouped optimally in unequal-sized clusters. Authors have selected the primary CH to the device having the highest node degree in the monitoring area. The secondary CHs selection process is done based on remaining energy and node degree. This process stops when all devices are attached to at least one cluster. Nevertheless, having many layers in hierarchical clustering can increase the number of hops, which further extends the delay. Then, to make the network energy-efficient and to elongate the network lifetime, an evolutionary approach and fuzzy logic-based solution is suggested [23]. Authors have used fuzzy logic to estimate the fitness value for which the fuzzy descriptors are taken as node degree, centrality, residual energy, and distance to BS. Then, invasive weed optimization (IWO) is utilized to select the candidate CHs for a round, and thereafter final head nodes are selected with the help of genetic algorithm and IWO. On the contrary, this approach also leads to more energy consumption because of frequent re-clustering in every round, and the evolutionary approach does not ensure the optimal solution.

A node rank algorithm is proposed in [24] for the systematic selection of CHs instead of random selection as done in LEACH. In this, the weight factor is calculated using the number of neighboring links, remaining energy, and distance w.r.t. the neighbors. This factor is used for selecting the CHs. Selected CHs use CSMA MAC protocol to form the clusters. Time division multiple access is used during the data transmission. Although they have not accounted uniform selection of CHs. In [25], the issue of energy consumption is handled by focusing on energy efficient hierarchical clustering methodology and packet routing technique using fuzzy c-means algorithm (EHCR-FCM). This algorithm works on
three-tier structure that depends on relative euclidean distances, remaining energy of the nodes, and centroid of the clusters and grids. To utilize the energy in an efficient manner, grid and cluster are formed in a dynamic manner. Fitness value of a node helps in the decision making whether a particular node will act as a CH or a grid head (GH). GHs are selected among the selected CHs. Data transmission strategy of GHs towards BS utilizes its distance to the BS and residual energy parameters. However, re-clustering has been done after every round which involves a lot of energy consumption.

Energy consumption for large-scale sensor fields is efficiently controlled using a fuzzy-based clustering [26]. Authors have used fuzzy logic in 4 stages: communication radius estimation, CHs selection, cluster formation, and relay node selection. The competition radius of the CHs is determined using the parameters remaining battery, distance w.r.t. BS, and node density. The chance to become a CH is determined using distance to the BS and node density. Each node determines the chance to become a CH using the input parameters residual energy and distance to the CH. Relay CHs selection is made using the input parameters delay distance, residual energy, and distance to the optimal point. However, the authors have not focused on uniform clustering. Another solution for optimal clustering and routing is suggested by [27]. In this paper, the authors have proposed a gravitational method-based clustering algorithm. Fuzzy logic is used to calculate the chance of election value. The input parameters passed to fuzzy systems are force of attraction, distance, and remaining energy. This work can be extended further by the usage of intelligent agent and minimizing frequent re-clustering.

Another approach to minimize the relay nodes’ excessive energy consumption close to the CHs is suggested in [28]. This approach is based on fuzzy logic with the non-uniform distribution. The input parameters considered for the CHs selection are residual energy of nodes and their neighbors and node degree w.r.t. each node. The authors have calculated the probability of CHs selection in a distributive way. Then, to minimize the total energy consumption, a fuzzy-based two-level hierarchical clustering algorithm is developed [29]. This objective is achieved by appropriately selecting the CHs using fuzzy logic. The input parameters that are passed to fuzzy logic are node centrality, residual energy, and node degree. The second level CHs are selected using fuzzy logic by considering residual energy, node centrality, and mobility. However, re-clustering has been done in each round.

After the literature review, we have observed that re-clustering consumes a lot of energy. And very few protocols have given the focus on reducing the re-clustering after each round. Some have tried to reduce the re-clustering but no fair idea was provided for layered structure. Also, the computational complexity of the clustering process is high. These features are mainly needed when the network is to be kept alive for a longer duration and scalability is required as well. To avoid these problems, we have proposed this technique, which reaps the benefits of layered structure and fuzzy logic. A summary of objectives, critical parameters considered, methodology used, simulation results, and weakness of the reviewed clustering protocols examined is shown in Table 1.

### 3 Assumptions

1. Sensor nodes are homogeneous in nature with an initial energy of 5 J.
2. Sensor nodes forward data in multiple hops to the BS.
3. Sensor nodes can vary their communication range with a maximum of 45 m and the range of BS is fixed as 80 m.
| Protocol          | Objective                                    | Parameters considered                                      | Methodology used                                 | Simulation results                                      | Shortcomings                                                                 |
|-------------------|----------------------------------------------|------------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------|------------------------------------------------------------------------------|
| ELDRT, 2016 [15]  | Effective Energy usage using data compression| Distance, packet size, remaining energy                    | Probabilistic based clustering and Huffman coding | Enhanced network lifetime and reduced energy consumption due to compressed packet size | Non-uniform clustering, Lot of energy consumption in re-clustering done in each round |
| EEMDC, 2015 [16]  | To prolong the network lifetime              | Distance from BS, energy                                   | Multi-level and distance-based clustering mechanism | Less energy consumption, more number of alive nodes     | Re-clustering done in each round, energy exhaustion attacks                  |
| CL-LEACH, 2016 [17]| Prolonging network lifetime by reducing energy dissipation | Remaining energy, distance between CHs and BS              | Heuristic based CHs selection                    | Less energy dissipation                                 | Re-clustering done in each round, non-optimized energy consumption           |
| Dk-LEACH, 2017 [18]| Reducing the uneven energy distribution      | Distance between the devices and w.r.t. BS, density, energy | Probabilistic CHs selection with Cost function     | Better network lifetime                                 | Non-uniform distribution of CHs                                              |
| Novel-PSO-LEACH, 2020 [19]| Mitigating the issues of energy consumption | Residual Energy, Distance between cluster member and BS | Particle swarm optimization                      | Improved network lifetime                               | More energy consumption due to re-clustering after every round, Non-uniform distribution of CHs |
| HEEMP, 2018 [20]  | Minimizing the energy consumption of the network | Node density and remaining energy                          | Centralized chance of election based CHs selection | More stable and less number of dead nodes                | Requirement of minimization of re-clustering in each round                   |
| FBUCA, 2021 [21]  | Enhancing energy efficiency and network lifetime | Residual energy, distance to the BS, and degree centrality | Fuzzy-logic based unequal clustering and three level multi-hop optimized routing | Less average energy consumption                          | CHs selection is independent of incoming load                                 |
| DLUC, 2019 [22]   | Elongating lifetime                          | Number of neighbors, residual energy                       | Heuristic, multi-channel approach                | Minimal Low energy consumption, less incoming traffic on CHs | Non-uniform distribution of CHs, Re-clustering done in each round           |
| MIWOCA, 2019 [23] | Elongating network lifetime                  | Node degree, centrality, residual energy, distance to BS  | Fuzzy logic, Invasive weed optimization, Genetic algorithm | Higher residual energy, Less number of dead nodes       | Re-clustering done in each round, standard deviation in residual energy is high |
| Protocol                  | Objective                                                                 | Parameters considered                                                                 | Methodology used                                                                 | Simulation results                                           | Shortcomings                                                                 |
|--------------------------|---------------------------------------------------------------------------|----------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|----------------------------------------------------------------|--------------------------------------------------------------------------------|
| NR-LEACH, 2017 [24]      | Minimizing the energy consumption of the network                         | Number of neighboring nodes, distance between the nodes, remaining energy              | Weighted approach usage for CHs selection                                       | Increased network lifetime with minimized delay               | More energy consumption due to re-clustering after every round               |
| EHCRC-FCM, 2021 [25]     | To minimize the energy consumption                                        | Relative euclidean distances, remaining energy of the nodes, and centroid of the clusters and grids | Hierarchical clustering methodology and packet routing technique using fuzzy c-means | Better in terms of network lifetime and number of clusters in the network | Non-uniform distribution of CHs, More energy consumption due to re-clustering in each round |
| EEFUC, 2020 [26]         | Optimizing energy consumption in data communication                       | Distance to the BS, number of alive nodes, and residual energy                         | Fuzzy-based clustering                                                           | Less number of dead nodes                                    | Total remaining energy is not much high, CHs are not distributed uniformly |
| EECGF, 2021 [27]         | Optimized routing and clustering                                           | Force of attraction, distance, and remaining energy                                     | Gravitational approach based clustering and routing                              | Improved network lifetime                                    | Non-uniform distribution of CHs, Re-clustering done in each round            |
| EEDCF, 2017 [28]         | To minimize the excessive energy consumption of relay nodes close to the CHs | Residual energy of nodes and their neighbors and node degree w.r.t. each node         | Fuzzy logic based clustering                                                     | Total residual energy is comparatively high                  | End-to-end delay is high between nodes and CHs, re-clustering done in each round |
| FM-SHEL, 2019 [29]       | Minimizing the energy consumption                                         | Residual energy, Node degree, Centrality, Mobility                                   | Fuzzy based CHs and super CHs selection                                          | Increased network lifetime, More alive nodes                  | Re-clustering done in each round, standard deviation in residual energy is high |
4. Hop-counts are calculated depending on the distance from the BS.
5. Deployment of the sensor nodes is done in a random manner and become stationary after deployment.
6. BS is deployed and fixed at some position in the network.
7. The whole network is divided into layers and nodes are assigned different layers as per some criteria. Sink is assumed to be present in layer 0 and the layer which is farthest away, assigned the maximum numeric value.
8. Communication between different layers is done using CHs. CHs in \( \ell \)th layer transmit their data to the CHs in the \((\ell − 1)\)th layer.

4 System Model

4.1 Network Model

Once the random deployment of the sensor nodes is done, they become stationary. Sink broadcasts a RQST packet to gather the details about the network. We assume that BS is present at hop count zero. To deal with the energy hole problem, the network is divided into layers on the basis of hop-counts and similarity of the load of the nodes which is quantified with the help of Intersection over Union (IoU) from the histogram of incoming load for a particular hop-counts. We have conducted many experiments to analyze how layers should be formed by considering different hop counts into one particular layer. CHs are selected initially using probabilistic approach as in LEACH algorithm using a modified threshold. CHs are selected using some heuristic approach for the region in which no CHs got selected. The re-clustering decision is taken by using the output of fuzzy logic. Also, a uniform selection of the CHs is ensured in all the layers.

4.2 Energy Model

According to energy model [14], while sending the data from node \( s_i \) to \( s_j \), energy consumption is given as:

\[
ET_{ij} = \begin{cases} 
    E_{elec} \times \kappa + \epsilon_{fs} \times \kappa \times d_{ij}^2, & d_{ij} \leq d_{TH} \\
    E_{elec} \times \kappa + \epsilon_{mp} \times \kappa \times d_{ij}^4, & d_{ij} > d_{TH}
\end{cases}
\]

where \( E_{elec} \) depicts the energy consumption to activate the electronic circuits. \( \epsilon_{fs} \) and \( \epsilon_{mp} \) depicts the energy consumption that the amplifier uses to send the control packet using free space and multi-path models respectively. Here \( d_{ij} \) depicts the euclidean distance between sender \( (s_i) \) and receiver \( (s_j) \) and \( d_{TH} = \sqrt{(\epsilon_{fs}/\epsilon_{mp})} \).

While the energy consumption during receive is given by:

\[
ER_x = E_{elec} \times \kappa
\]
5 Proposed Work

Now, we will discuss the proposed load balanced clustering algorithm named HFLBSC (heuristic and fuzzy logic-based load balanced and scalable clustering). HFLBSC is a probabilistic, distributed, and layering based approach having unequal clusters. In HFLBSC, the network is divided into layers and these layers are having unequal clusters w.r.t. each other depending on their distance from the BS. CHs selection and update cycle calculation for re-clustering are handled locally. HFLBSC utilizes the advantage provided by fuzzy logic to estimate the update cycle w.r.t. each CH. In fuzzy logic, lifetime and the incoming load of a node are passed to calculate its update cycle. It is rational to decrease the CH’s update cycle as the lifetime is decreasing. If we don’t change the update cycle as the lifetime decreases, the battery of some CHs may quickly go below usefulness criteria. This situation is taken care of in HFLBSC, as the battery power decreases. Data transmission is done among the CHs of one layer to the CHs of another layer towards the sink in a multi-hop fashion. A flow chart w.r.t. our proposed algorithm is shown in Fig. 2.

Fig. 2 Flow chart of proposed HFLBSC algorithm
5.1 Network Initialization

A set of \( n \) sensor nodes deployed in a random manner. A \( RQST(sender_id, h) \) packet is floated synchronously to get the information about the nodes present in the network where \( sender_id \) represents the node this packet came through and \( h \) is the hop value w.r.t. BS. A node receives the \( RQST \) packet from its neighboring nodes. Neighboring nodes are defined as the nodes that are one hop away. Each node sets that node as its next-hop initially from whom it receives the \( RQST \) packet first. In this way, each node will have a rough estimation of its incoming load.

5.2 Layer Construction Process

The network is divided into layers based on the incoming load on each node at a particular hop count. As we said, upon receiving the \( RQST \) packet, each node has a rough estimation of incoming load on it and each node has the information about the hop count concerning the sink node. Using this information, we find the number of nodes \( (NWLAH) \) with a particular load value \( (LD) \) at a particular hop count \( (h) \). Then, we determine the similarity of load distribution over \( j \) consecutive hop-count regions by calculating the \( IoU \) value using Eq. 1:

\[
IoU^j = \frac{\text{area of intersection of given } j \text{ functions}}{\text{area of union of given } j \text{ functions}}
\]  

(1)

The \( IoU \) value given above over \( j \) functions - \((NWLAH_h(LD), \ldots , NWLAH_{h+j-1}(LD))\) is concretely defined as given in Eq. 2:

\[
IoU^j_h = \frac{\text{area under } f_{min_{h \ldots h+j-1}}()}{\text{area under } f_{max_{h \ldots h+j-1}}()}
\]

(2)

where \( f_{min_{h \ldots h+j-1}}() \) and \( f_{max_{h \ldots h+j-1}}() \) is defined as given in Eqs. 3 and 4:

\[
f_{min_{h \ldots h+j-1}}(LD) = \min(NWLAH_h(LD), \ldots , NWLAH_{h+j-1}(LD))
\]  

(3)

\[
f_{max_{h \ldots h+j-1}}(LD) = \max(NWLAH_h(LD), \ldots , NWLAH_{h+j-1}(LD))
\]  

(4)

Lower the value of IoU dissimilar is the incoming load distribution among the nodes of two or more consecutive hop counts regions and vice-versa. So, they should be dealt differently and should not go under the same layer. However, we have conducted to confirm the same which can be seen in Sect. 6. The nodes that have hop count between \( h \) and \( h + j - 1 \) lie in the same layer only when the region’s \( IoU_h^j \) value is greater than some particular value. In this way, nodes are segregated and different layers get formed. Once the layers are constructed, the communication radius of each node is updated depending on its distance from the BS. It helps in reducing the communication distance and hence the energy consumption. Let \( CR_0 \) be the maximum communication radius which is fixed. Depending on the distance of each node from the BS \( (d_{s_i,BS}) \), we set the communication radius \( (CR_i) \) of each node \( s_i \) defined in Eq. 5 [30]:

\[
CR_i = \left( 1 - c \times \frac{d_{max} - d_{s_i,BS}}{d_{max} - d_{min}} \right) \times CR_0
\]

(5)
Here $d_{max}$ is the largest distance between any 2 nodes in the network. Value of $c$ lies in the $[0, 1]$. Once the layers are constructed and the communication radius is updated, CHs are selected in the layers. The CHs lie in $(\ell + 1)$th layer send their data to the nearest relay CHs lie in $\ell$th layer. In this way, nodes send their data to the BS.

5.3 CHs Selection Process

A cluster is composed of the nodes located in the same layer. There can be more than one cluster in a layer. The nodes lie in one layer participates in the CHs selection process. A random value between 0 and 1 is generated by each sensor node. This random value is then compared with the threshold value which is given by

$$T(s_i) = \begin{cases} \frac{b}{1-px\%} \times \left( \frac{RE_i}{IE_i} \right), & \text{if } s_i \in \hat{S}_G \\ 0, & \text{otherwise} \end{cases}$$

where $\hat{S}_G$ is the set of nodes that are not selected as CHs in last $1/p$ rounds. If the random value is less than the threshold value, then the node is elected as the CH, otherwise not. In LEACH, residual energy is not considered. It has lead to the early death of the nodes. In the proposed HFLBSC algorithm, residual energy is considered to cater to this problem. Also, in a layer, the next CH elected should be threshold distance apart from already elected CHs. In this way, a uniform selection of CHs is done in all the layers. If some nodes are left to which no CH is assigned, then CHs are selected using some heuristic approach among those nodes in which maximum residual energy is used to select the CH.

5.4 Relay CHs Selection Process

As soon as the CHs are selected, relay CHs is selected thereafter. Relay CHs are selected from the selected CHs. The CHs lie in $(\ell + 1)$th layer send their data to the nearest CHs lie in $\ell$th layer which acts as the relay node. If the incoming load on a particular relay CH in $\ell$th layer goes above a certain threshold, then the next nearest CH is selected as the relay node in the $\ell$th layer. In this way, the incoming load is balanced among the relay CHs. In a round, if the energy of any of the relay CHs goes below a threshold, then another relay CH is selected from the rest of the CHs in that layer. Once the relay CHs are selected, nodes send their data to the BS in a multi-hop fashion.

5.5 Re‑clustering Decision Using Fuzzy Logic

Re-clustering is an energy-consuming process. In the proposed HFLBSC, we have tried to reduce the number of times re-clustering is being done. Initially, when residual energy is high, then the CH can remain as the CH for a few upcoming rounds. Update cycle w.r.t. a CH tells when to re-cluster. The update cycle is to find out by passing the input variables into fuzzy logic. The input variables considered in the HFLBSC algorithm are the lifetime of each node and incoming load. The output variable found out using fuzzy logic is the update cycle ($UC$) value corresponding to each node. This variable helps in deciding when to perform re-clustering instead of doing re-clustering in each cluster in every round. Re-clustering in each cluster is decided using $UC$, estimated with the help of fuzzy logic. The update cycle depends directly on the lifetime of the
sensor node. Higher the lifetime, the higher is the update cycle. A higher update cycle
depicts re-clustering is to be done after a longer duration. The update cycle depends
inversely on the incoming load of the node. Higher the incoming load on a node, lesser
is the update cycle. Less update cycle depicts re-clustering is to be done after a short
duration. So, when we need to perform re-clustering, it depends on input variables that
are passed to the fuzzy logic i.e. lifetime w.r.t. each node and incoming load on the
node. In other words, it can be understood as given in Eqs. 7 and 8:

\[ UC \propto L \]  

\[ UC \propto \frac{1}{LD} \]  

Membership functions used for the input and output variables are trapezoidal and trian-
gular membership functions. The linguistic variables used for the lifetime variable are Very
Low (VL), Low (L), Average (A), High (H), and Very High (VH) as shown in Fig. 3. The
linguistic variables used for the incoming load variable are Very Small (VS), Small (S),
Medium (M), Large (L), and Very Large (VL) as given in Fig. 4. The linguistic variables
used for the output variable i.e. update cycle are Very High Frequent (VHF), High Fre-
quent (HF), Frequent (F), Low Frequent (LF), and Very Low Frequent (VLF) as given in
the Fig. 5. We have used the Mamdani method, which processes fuzzified input variables
and develops fuzzy rules. Inference rules used in HFLBSC are discussed in the Table 2.
When re-clustering takes place in one layer, accordingly the nodes in its next layer away from the sink, set their relay CHs. Also, the newly elected CHs select their nearest CHs as a relay node. Re-clustering in different layers occurs at the same time only when the value of \( UC \) w.r.t. current CHs are the same in both layers.

Table 2 Inference rules used in HFLBSC

| Lifetime | Incoming load | Update cycle |
|----------|---------------|--------------|
| VL       | VS            | HF           |
| VL       | S             | HF           |
| VL       | M             | VHF          |
| VL       | L             | VHF          |
| VL       | VL            | VHF          |
| L        | VS            | F            |
| L        | S             | HF           |
| L        | M             | HF           |
| L        | L             | HF           |
| L        | VL            | HF           |
| A        | VS            | F            |
| A        | S             | F            |
| A        | M             | F            |
| A        | L             | HF           |
| A        | VL            | HF           |
| H        | VS            | LF           |
| H        | S             | LF           |
| H        | M             | LF           |
| H        | L             | LF           |
| H        | VL            | F            |
| VH       | VS            | VLF          |
| VH       | S             | VLF          |
| VH       | M             | VLF          |
| VH       | L             | LF           |
| VH       | VL            | LF           |
The working steps of the proposed HFLBSC algorithm are given in Algorithm 1, 2, and 3.

**Algorithm 1** Algorithm Network Initialization and Layers Formation

**Input:** Set of sensor nodes $\hat{S}$

**Output:** Layers formed in the network

1. Begin
2. $RQST$ Packet is broadcasted into environment by $\text{sink}$.
3. Hop count of $\text{sink}$ is set as 0
4. for each node in $\hat{S}$ do
5. Next hop is initialized on first come first serve basis
6. Adjacency list of neighboring nodes is prepared
7. Hop count and tentative incoming load ($LD_i$) w.r.t. each node are estimated.
8. end for
9. Number of nodes with particular load at particular hop are determined and stored in $\text{NWLAH}$. 
10. Layers are formed in the network by finding IoU.
11. Total layers ($\text{Num\_layer}$) are calculated
12. for $\ell$ in $\text{Num\_layer}$ to 1 do
13. for each node in $\hat{S}$ do
14. Nodes are assigned to the layers using hop count and a layer-wise list of nodes ($NL$) is prepared.
15. end for
16. end for
17. Communication radius of the nodes is updated using Eqn. 5
18. End
Algorithm 2 Algorithm for CHs and Relay CHs Selection

Input: Layers information and adjacency list of sensor nodes
Output: Selected CHs and relay CHs

1: Begin
2: for $\ell$ in Num_layer to 1 do
3:     for each node in $NL_{\ell}$ do
4:         Random value is generated between 0 and 1
5:         Threshold value is calculated using Eqn. 6
6:         if random value $\leq T(s_{i})$ and node $s_{i}$ is some distance threshold apart then
7:             $s_{i}$ is chosen as the CH
8:         end if
9:     end for
10: end for
11: for $\ell$ in Num_layer to 1 do
12:     for each node in $NL_{\ell}$ do
13:         Each node will select nearest CH and become CM
14:     end for
15: end for
16: for $\ell$ in Num_layer to 1 do
17:     for each node in $NL_{\ell}$ do
18:         if no CH is assigned then
19:             $NoCHsAssigned \leftarrow$ node
20:         end if
21:     end for
22: end for
23: for each node in $NoCHsAssigned$ do
24:     Maximum energy node will be selected as CH
25: end for
26: for $\ell$ in Num_layer to 1 do
27:     for each CH in $NL_{\ell}$ do
28:         Nearest CH in $(\ell - 1)^{th}$ layer is chosen as Relay CHs
29:     end for
30: end for
31: return CHs and Relay CHs
Now, we will emphasize on the statistical analysis that we have done to find the significant difference between the different arrangement of layers. The effectiveness of different arrangements of layers is proved statistically using F-test. Then, we will discuss the various network parameters used during the simulation followed by results obtained by considering different layered approaches.

### 6.1 Layers Formation

During simulation, we have considered four different kinds of arrangements of layers using the IoU values. In 1st arrangement, layer 1 is having the nodes which are at hop count 1, 2, and 3. Layer 2 is having the nodes which are having hop count 4 and 5. While layer 3 is having the nodes which are having hop count 6 and 7. Then, the nodes having hop count 8 and 9 falls in the layer 4. Similarly three more arrangements are considered discussed in Table 3.

We have calculated weighted sum of IoU values for all the arrangements. The weighted sum of IoU for the arrangements A1, A2, A3, and A4 is 1.7146, 1.2146, 1.0128, and 0.7222. It is observed that, on an average, weighted sum of IoU for arrangement A1 is 82% more than the rest arrangements. So, A1 is best among them.

### 6.2 Statistical Analysis

We have performed F-tests to find out the substantial difference between the different arrangement of layers w.r.t. FND and HND.

### Table 3 Layers arrangements

| Arrangements | Hop-wise layers formation |
|--------------|---------------------------|
|              | Layer 1 | Layer 2 | Layer 3 | Layer 4 |
| A1           | 1, 2, 3 | 4, 5    | 6, 7    | 8, 9    |
| A2           | 1, 2    | 3, 4, 5 | 6, 7    | 8, 9    |
| A3           | 1, 2    | 3, 4    | 5, 6, 7 | 8, 9    |
| A4           | 1, 2    | 3, 4    | 5, 6    | 7, 8, 9 |
First, we have carried out the $F$-test to find whether there is a difference between all three different kinds of arrangements of layers. For better statistical analysis, we have performed $F$-test on a set of 50 random deployments w.r.t. first node death and half node death. The $F$-calculated values ($F_{cal}$) and $F$-critical value ($F_c$) obtained by considering all arrangements mentioned in Table 3 are stated in Table 4. We can observe that $F_{cal}$ w.r.t. FND and HND is 81% and 262% more than the $F_c$ value. Therefore, it can be concluded that there exists a significant difference between all the different arrangements of layers.

Second, we have carried out $F$-test to analyze which arrangement is better compared to rest. For this, we have conducted pair-wise $F$-test. The $F$-calculated values ($F_{cal}$) obtained by considering two arrangements at a time, mentioned in Table 3 are stated in Tables 5 and 6. The $F$-critical value ($F_c$) is 6.90 for $\alpha = 0.01$.

From Tables 5 and 6, we can conclude that there is a significant difference between the pairs i.e. $A_1 A_2$, $A_1 A_3$, and $A_1 A_4$. However, the pairs $A_2 A_3$, $A_2 A_4$, and $A_3 A_4$ is not having significant difference w.r.t. $F_c$ value. From the Tables 5, 6 and the $F_c$ value, we can say that the distribution of layers is best in arrangement $A_1$ and hence, layering in the network should be done with a care.

7 Simulation Results

We have simulated a homogeneous clustered WSN in which sensor nodes are deployed in a square field of size 200m $\times$ 200m using MATLAB. The number of sensor nodes deployed varied from 100 to 200 nodes. The sink is assumed to be located at (0, 50) in all the approaches. As discussed in Sect. 5, the whole network is divided into layers. To avoid the hotspot problem, we have varied the communication range of the nodes which are near to sink and away from the sink. Each sensor node can have a maximum range of 45m and a battery lifetime of 0.5J. Parameters related to network setting are given in the Table 7.
7.1 Evaluation Metrics

The following metrics are used to examine the behavior of different routing solutions in a comprehensive way:

(a) Packets Delivery Ratio (PDR): It represents the proportion of number of packets delivered to the BS and the total number of nodes in the field. The value of this ratio near to one reflects the efficiency of the network. The network can not show actually sensed region if value of PDR is low.

(b) First Node Death (FND): It is the time period (in rounds) when the very first node dies in the network.

(c) Half Node Death (HND): It is the time period (in rounds) when half of the nodes get died.

(d) Last Node Death (LND): It is the time period (in rounds) when all of the nodes get died.

(e) Energy Consumed per Round: It is the energy consumed in a single round by all the nodes in performing the major operations such as sensing, communication.

(f) Number of Alive Nodes: It is count of nodes that are alive over the period of time. i.e. the number of the nodes whose battery is not exhausted.

Figures 6 and 7 represents the number of rounds when FND, HND, and LND occurred in LEACH, FM-SCHEL, MIWOCA and different arrangements considered in proposed work when 100 and 200 nodes are considered. From the graph, it is clear that FND of HFLBSC w.r.t. A1 occurs 46%, 27%, and 23% late than LEACH, FM-SCHEL, and MIWOCA respectively. This is because of single-hop communication between sensor nodes and the sink in LEACH. In LEACH, FM-SCHEL, and MIWOCA, all the nodes have the same communication range irrespective of their distance w.r.t. BS. While in HFLBSC, intra-cluster and inter-cluster communication are accomplished using multi-hop communication. Along with it, all the nodes have different communication range w.r.t. BS. From Fig. 6, it can be observed that FND of HFLBSC w.r.t. A1 occurs 22%, 11%, and 15% late than A2, A3, and A4 respectively. We can notice that the 1st
arrangement (A1) is outperforming than the rest of the arrangements by giving FND as 846 rounds. Although, rest three arrangements (A2, A3, and A4) are doing equally good but not as much as A1. It infers node death parameters is better when the density of nodes in 1st layer is more than the layer which lies far away from the sink. Statistical analysis was done earlier also supports that there is a significant difference between different arrangements.

From Fig. 6, it is also clear that HND and LND for proposed HFLBSC algorithm w.r.t. arrangement A1 occurs 29%, 28%, 26% and 28%, 20%, 17% late than LEACH, FM-SCHEL, MIWOCA algorithms respectively. Along with it, HND and LND w.r.t. arrangement A1 is delayed 26%, 12%, 16% and 16%, 11%, 12% than other arrangements (A2, A3, A4). This is the outcome of the above-mentioned factors. Also, varied communication range and unequal clusters result in avoidance of the hot-spot problem. Thereafter, network scalability is checked by increasing the number of nodes in the network field. From the Fig. 7, we can observe that arrangement A1 outperforms the rest arrangements considered.

Figure 8 represents the energy consumption by different protocols over the rounds. Simulation results depict very clearly that energy consumption in the proposed HFLBSC is 45%, 39%, 29% less in comparison to LEACH, FM-SCHEL, and MIWOCA respectively. This is the positive effect given by multi-hop communication along with unequal and optimal re-clustering in proposed work. Network performance (Fig. 9) is equally good and similar results were observed when network is scaled over 200 nodes.
Figure 10 and 11 is providing a more clear picture of the number of alive nodes over some particular round values. Figure 10 shows till the end of network lifetime, nodes in arrangement A1 of proposed HFLBSC algorithm were able to survive 29%, 25%, 21% more time than LEACH, FM-SCHEL, and MIWOCA in terms of alive nodes. After a close analysis, we observed that energy consumption is minimum in arrangement A1 of proposed HFLBSC in comparison to others which results in more number of alive nodes in the same.
Furthermore, the standard deviation in the residual energy is very less in arrangement A1 in comparison to others (refer Fig. 12). This shows energy is dissipated in a more balanced way in arrangement A1.

Figures 13 and 14 shows the packet delivery ratio w.r.t. LEACH, FM-SCHEL, MIWOCA, and different arrangements discussed when 100 and 200 nodes are deployed into the field. We noticed that PDR at 900 rounds w.r.t. arrangement A1 of proposed HFLBSC algorithm is 7%, 5.5%, and 6% more than LEACH, FM-SCHEL, and
MIWOCA algorithms. PDR for LEACH protocol is very poor because while doing the selection of CHs, residual energy is not considered. Although PDR for FM-SCHEL and MIWOCA is more than LEACH. But in both of these, optimized re-clustering is not considered which leads to a lot of energy consumption. While in proposed HFLBSC, both of these are considered which ensures better results in proposed protocol. Multi-hop communication between the nodes and BS maintains the connectivity for a longer duration in comparison to single-hop communication. Figure 13 also showcases that PDR is almost equal for the arrangements $A_1, A_3, A_4$ in comparison to $A_2$. But we have seen that nodes remain alive for longer duration in $A_1$, so $A_1$ promises to work efficiently for longer duration comparatively.

Now, if the network lifetime is defined as the time until we are getting 50% PDR. Then also, network lifetime will be more for the proposed HFLBSC algorithm. Using proposed HFLBSC, we are getting 50% PDR till the last round in arrangement $A_1$, it implies that network lifetime of the proposed algorithm ($A_1$) is better among others.

8 Conclusion and Future Scope

In this paper, we have introduced a wise protocol named HFLBSC. The key characteristics of our proposed protocol are uniform energy consumption by the uniform selection of CHs, enhanced network lifetime due to reduced energy consumption spent in re-clustering, enhanced scalability with a layered structure. Moreover, we have utilized the fuzzy logic to estimate the update cycle by considering the parameters such as incoming traffic load on a node and lifetime of that node. This update cycle is then used to decide when re-clustering needs to be performed. Such factors differentiate this protocol from existing protocols. Further simulation results show that all the nodes drain off the battery at the same pace. Proposed algorithm results showcase that it is giving better results in terms of energy consumption, network lifetime, PDR, number of alive nodes, FND, HND, and LND. F-test confirms that there is a significant difference between different arrangements in the layered network.

Future work for the proposed algorithm can be to embed the security in the network. Security can be enhanced by adding some authentication in the sensor nodes.
Acknowledgements Priti Maratha acknowledges the support from University Grant Commission, New Delhi under National Eligibility Test-Junior Research Fellowship scheme with Reference ID-3361/(NET-JUNE 2015).

References

1. Pathak, A. (2020). A proficient bee colony-clustering protocol to prolong lifetime of wireless sensor networks. Journal of Computer Networks and Communications 2020.

2. Daanoune, I., Abdennaceur, B., & Ballouk, A. (2021). A comprehensive survey on LEACH-based clustering routing protocols in wireless sensor networks. In: Ad Hoc Networks, p. 102409.

3. Mahabal, C., & Fang, H. (2020). Smart spectrum switching and beamforming for wireless body area networks in dynamic environment. Journal of Communications and Information Networks, 5(2), 204–216.

4. Zhao, Z., Li, G., & Xu, M. (2019). An improved algorithm based on LEACH routing protocol. In: Proceedings of the 2019 IEEE 19th international conference on communication technology (ICCT). IEEE, pp. 1248–1251.

5. Visu, P., et al. (2020). Bio-inspired dual cluster heads optimized routing algorithm for wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, pp. 1–9.

6. Kortas, M., et al. (2020). The energy-aware matrix completion-based data gathering scheme for wireless sensor networks. IEEE Access, 8, 30772–30788.

7. Ullah, Z. (2020). A survey on hybrid, energy efficient and distributed (HEED) based energy efficient clustering protocols for wireless sensor networks. In: Wireless personal communications, pp. 1–29.

8. Gupta, G.P., & Saha, B. (2020). Load balanced clustering scheme using hybrid metaheuristic technique for mobile sink based wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing.

9. Wang, M., Wang, S., & Zhang, B. (2020). APTEEN routing protocol optimization in wireless sensor networks based on combination of genetic algorithms and fruit fly optimization algorithm. Ad Hoc Networks, p. 102138.

10. El Fissaoui, M., Beni-Hssane, A., & Saadi, M. (2019). Energy efficient and fault tolerant distributed algorithm for data aggregation in wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 10(2), 569–578.

11. Mehmood, A., et al. (2017). ELDC: An artificial neural network based energy-efficient and robust routing scheme for pollution monitoring in WSNs. In: IEEE Transactions on Emerging Topics in Computing.

12. Baranidharan, B., & Santhi, B. (2016). DUCF: Distributed load balancing Unequal Clustering in wireless sensor networks using Fuzzy approach. Applied Soft Computing, 40, 495–506.

13. Zhang, C., Patras, P., & Haddadi, H. (2019). Deep learning in mobile and wireless networking: A survey. IEEE Communications Surveys and Tutorials, 21(3), 2224–2287.

14. Rabiner Heinzelman, W., Chandrakasan, A., & Balakrishnan, H. (2000). Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd annual Hawaii international conference on system sciences. IEEE, p. 10.

15. Mehmood, A., et al. (2015). Energy-efficient multi-level and distance-aware clustering mechanism for WSNs. International Journal of Communication Systems, 28(5), 972–989.

16. Marappan, P., & Rodrigues, P. (2016). An energy efficient routing protocol for correlated data using CL-LEACH in WSN. Wireless Networks, 22(4), 1415–1423.

17. Ding, X.-X., et al. (2017). Dk-leach: An optimized cluster structure routing method based on leach in wireless sensor networks. Wireless Personal Communications, 96(4), 6369–6379.

18. Thiagarajan, R., et al. (2020). Energy consumption and network connectivity based on Novel-LEACH-POS protocol networks. Computer Communications, 149, 90–98.

19. Rajaram, V., & Kumaratharan, N. (2021). Multi-hop optimized routing algorithm and load balanced fuzzy clustering in wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 12(3), 4281–4289.
22. Farahani, M., & Ghaffarpour Rahbar, A. (2019). Double leveled unequal clustering with considering energy efficiency and load balancing in dense IoT networks. *Wireless Personal Communications, 106*(3), 1183–1207.

23. Sharma, R., Vashisht, V., Singh, U. (2019). Fuzzy modelling based energy aware clustering in wireless sensor networks using modified invasive weed optimization. *Journal of King Saud University-Computer and Information Sciences*.

24. Al-Baz, A., & El-Sayed, A. (2018). A new algorithm for cluster head selection in LEACH protocol for wireless sensor networks. *International Journal of Communication Systems, 31*(1), e3407.

25. Panchal, A., & Kumar Singh, R. (2021). EHR-FCM: Energy efficient hierarchical clustering and routing using fuzzy C-means for wireless sensor networks. *Telecommunication Systems, 76*(2), 251–263.

26. Phoemphon, S., et al. (2020). An energy-efficient fuzzy-based scheme for unequal multihop clustering in wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–23.

27. Selvi, M., et al. (2021). An energy efficient clustered gravitational and fuzzy based routing algorithm in WSNs. *Wireless Personal Communications, 116*(1), 61–90.

28. Zhang, Y., et al. (2017). Fuzzy-logic based distributed energy-efficient clustering algorithm for wireless sensor networks. *Sensors, 17*(7), 1554.

29. Uma Maheswari, D., & Sudha, S. (2019). Node degree based energy efficient two-level clustering for wireless sensor networks. *Wireless Personal Communications, 104*(3), 1209–1225.

30. Chen, G., et al. (2009). An unequal cluster-based routing protocol in wireless sensor networks. *Wireless Networks, 15*(2), 193–207.

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

---

**Priti Maratha** She is graduated from Kurukshetra University and Masters in Computer Applications from Panjab University Chandigarh, India. Currently, she is pursuing PhD from National Institute of Technology, Kurukshetra. Her research interests include Wireless Sensor Networks.

**Kapil Gupta** He was currently working as an assistant professor at National Institute of Technology, Kurukshetra. He received his PhD Degree from Jawaharlal Nehru University, Delhi. His research interests include Wireless Sensor Networks, Machine Learning methods and applications. He has published various papers in the journals of International repute.