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

Priti Maratha (✉ niki.maratha19@gmail.com)

National Institute of Technology Kurukshetra <https://orcid.org/0000-0003-4919-4136>

Kapil Gupta

National Institute of Technology Kurukshetra

Research Article

Keywords: Load balancing ,load ,residual energy ,network lifetime ,wireless sensor networks

Posted Date: March 17th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-306786/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Wireless Personal Communications on January 21st, 2022. See the published version at <https://doi.org/10.1007/s11277-022-09550-z>.

HFLBSC: Heuristic and fuzzy based load balanced, scalable clustering algorithm for wireless sensor network

*Priti Maratha · Kapil Gupta

Received: date / Accepted: date

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. Statistical analysis and simulation results confirm the superiority of proposed work in comparison to its competitor protocol.

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]. 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) enables a routing protocol to transfer the data correctly from sensor nodes to BS [11].

*Corresponding Author

Department of Computer Applications, National Institute of Technology Kurukshetra, India

E-mail: niki.maratha19@gmail.com

ORCID ID: <https://orcid.org/0000-0003-4919-4136>

Kapil

Department of Computer Applications, National Institute of Technology Kurukshetra, India

Fig. 1 Clustered environment in WSN

This is the duty of a routing protocol to determine the best route from sensor nodes to BS to reduce overall transmission costs. 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. In this paper, a layered structure-based energy-efficient and unequal clustering protocol are 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 rest of the paper is set out as follows: Section 2 summarizes the literature on different routing protocols that are energy-efficient. Assumptions and system model of the work are briefed in Section 3 and 4. Section 5 includes a thorough explanation of the algorithm proposed. Statistical analysis and simulation results are portrayed respectively in Section 6 and 7 to demonstrate the efficiency of the proposed algorithm. Conclusion and future scope are discussed in section 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, the distance between the nodes and BS and leftover energy is not taken into consideration during the selection of CHs in WSNs. An energy-efficient scheme is built in [17] which takes into account both factors during the selection of CHs. Sensor nodes with less battery backup than total energy are not selected as CH. Energy is mainly consumed during transmission. It is minimized by choosing the relay nodes from one-hop neighbors from the neighboring list, whose

residual energy is greater than a given energy threshold. But it suffers from frequent re-clustering. In [18], the focus is given on reducing the uneven energy 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. The proportion of the distance between CHs and non-CHs and surplus energy is estimated. The minimum value of the proportion is used to select the CHs. This proportion is adjusted with node density, which inherently requires the re-clustering in each round. A hybrid energy-efficient distributive protocol (HEED) is proposed in [19]. Multi-hop communication is used for data transmission to the sink node. Intra cluster communication cost is used for the selection of CHs. But in this algorithm, every node works individually and does not provide any information about the size and density. 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. A legitimate route set is prepared by BS. This route set is the set of routes on which all the nodes are having residual energy above a 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.

In [21], CHs are elected by considering the residual energy into consideration. It improves LEACH and LEACH-C protocols. Simulation results confirm it achieves longer stability than its competitor protocols. A node rank algorithm is proposed in [22] for the efficient selection of CHs instead of random selection as done in LEACH. In this, the weight factor is calculated using the number of links with other sensor nodes, remaining energy, and distance of out link edges. This factor is used for selecting the CHs. Cluster formation is done by selected CHs by broadcasting a CSMA MAC protocol. Data transmission is managed using TDMA slots. In [23], energy utilization is reduced to elongate the network's lifetime. One CH is selected per cluster to synchronize the transmission of data. Leftover energy and radial basis network are used for the selection of CHs. So, it gives better outcome compared to the LEACH and LEACH-C protocols.

Energy consumption for large-scale sensor fields is efficiently controlled using a fuzzy-based clustering [24]. 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 distance to the BS, node density, and residual energy. 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 [25]. In this paper, the authors have proposed a gravitational method-based clustering algorithm. Fuzzy logic is used to calculate the chance of election value.

Another approach to minimize the relay nodes' excessive energy consumption close to the CHs is suggested in [26]. 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 [27]. This objective is achieved by appropriately selecting the CHs using fuzzy logic. The fuzzy input parameters considered are residual energy, node degree, and node centrality. 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.

3 Assumptions

1. Sensor nodes are homogeneous in nature with an initial energy of $5J$.
2. Sensor nodes forward data in multiple hops to the BS .
3. Sensor nodes can vary their communication range with a maximum of $45m$ and the range of BS is fixed as $80m$.
4. Depending on the distance from the BS , the hop-counts are decided.
5. Nodes are deployed 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 ℓ^{th} layer transmit their data to the CHs in the $(\ell - 1)^{th}$ layer.

4 System Model

4.1 Network Model

Once the sensor nodes are deployed in a random manner, they become stationary. Sink broadcasts a control message i.e. $RQST$ message to get information about the network. Sink is 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 in a similar manner as in LEACH with 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], the energy consumed while transmitting data from node s_i to s_j is given as:

$$ET_{ij} = \begin{cases} \mathbb{E}_{elec} \times \kappa + \epsilon_{fs} \times \kappa \times d_{ij}^2, & d_{ij} < d_{TH} \\ \mathbb{E}_{elec} \times \kappa + \epsilon_{mp} \times \kappa \times d_{ij}^4, & d_{ij} \geq d_{TH} \end{cases}$$

where $\mathbb{E}_{elec}$ represents amount of energy required to activate the electronic circuits. ϵ_{fs} and ϵ_{mp} refers to the amount of energy required by the amplifier to send a packet of κ -bits using free space and multi-path models respectively. Here d_{ij} is the distance between sender (s_i) and receiver (s_j) and $d_{TH} = \sqrt{(\epsilon_{fs}/\epsilon_{mp})}$.

While the energy consumed during receive is given by:

$$ER_x = \mathbb{E}_{elec} \times \kappa$$

5 Proposed Work

In this section, 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

1 taken care of in HFLBSC, as the battery power decreases. Data transmission is done among the CHs
 2 of one layer to the CHs of another layer towards the sink in a multi-hop fashion. A flow chart of the
 3 proposed work is given in Fig. 2.

4 5.1 Network Initialization

5 A set of n sensor nodes deployed in a random manner. A $RQST(sender_id, h)$ message is floated
 6 synchronously to get the information about the nodes present in the network where $sender_id$ is the
 7 node this message came through and h is its hop count from the sink. A node receives the $RQST$
 8 message from its neighboring nodes. Neighboring nodes are defined as the nodes that are one hop
 9 away. Each node sets that node as its next-hop initially from whom it receives the $RQST$ message
 10 first. In this way, each node will have a rough estimation of its incoming load.

11 5.2 Layer construction process:

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

$$IoU^j = \frac{\text{area of intersection of given } j \text{ functions}}{\text{area of union of given } j \text{ functions}} \quad (1)$$

18
 19 The IoU value given above over j functions - ($NWLAH_h(LD), \dots, NWLAH_{h+j-1}(LD)$) is concretely
 20 defined as given in Eqn. 2:

$$IoU_h^j = \frac{\text{area under } f_{min_{h, \dots, h+j-1}}(.)}{\text{area under } f_{max_{h, \dots, h+j-1}}(.)} \quad (2)$$

21
 22 where $f_{min_{h, \dots, h+j-1}}(.)$ and $f_{max_{h, \dots, h+j-1}}(.)$ is defined as given in Eqn. 3 and Eqn. 4:

$$f_{min_{h, \dots, h+j-1}}(LD) = \min(NWLAH_h(LD), \dots, NWLAH_{h+j-1}(LD)) \quad (3)$$

$$f_{max_{h, \dots, h+j-1}}(LD) = \max(NWLAH_h(LD), \dots, NWLAH_{h+j-1}(LD)) \quad (4)$$

23
 24 Lower the value of IoU dissimilar is the incoming load distribution among the nodes of two or more
 25 consecutive hop counts regions and vice-versa. So, they should be dealt differently and should not go
 26 under the same layer. However, we have conducted to confirm the same which can be seen in section 6.
 27 The nodes that have hop count between h and $h+j-1$ lie in the same layer only when the region's
 28 IoU_h^j value is greater than some particular value. In this way, nodes are segregated and different
 29 layers get formed. Once the layers are constructed, the communication radius of each node is updated
 30 depending on its distance from the BS. It helps in reducing the communication distance and hence
 31 the energy consumption. Let CR_0 be the maximum communication radius which is fixed. Depending
 32 on the distance of each node from the BS ($d_{s_i, BS}$), we set the communication radius (CR_i) of each
 33 node s_i defined in Eqn. 5 [28]:

$$CR_i = \left(1 - c \times \frac{d_{max} - d_{s_i, BS}}{d_{max} - d_{min}}\right) \times CR_0 \quad (5)$$

34
 35
 36 Here d_{max} is the maximum distance between two nodes in the network. Value of c lies in the $[0, 1]$.
 37 Once the layers are constructed and the communication radius is updated, CHs are selected in the

- 1 layers. The CHs lie in $(\ell + 1)^{th}$ layer send their data to the nearest relay CHs lie in ℓ^{th} layer. In this
 2 way, nodes send their data to the BS.

Fig. 2 Flow chart of proposed HFLBSC algorithm

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. Each node generates a random number between 0 and 1. This random value is then compared with the threshold value which is given by

$$T(s_i) = \begin{cases} \frac{\hat{p}}{1 - \hat{p} \times r \bmod \frac{1}{\hat{p}}} \times \left(\frac{RE_i}{IE_i} \right), & \text{if } s_i \in \hat{S}_G \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

where $\hat{S}_G$ is the set of nodes that have not been CHs in last $1/\hat{p}$ 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 ℓ^{th} layer which acts as the relay node. If the incoming load on a particular relay CH in ℓ^{th} layer goes above a certain threshold, then the next nearest CH is selected as the relay node in the ℓ^{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 eqn. 7 and 8:

$$UC \propto L \quad (7)$$

$$UC \propto \frac{1}{LD} \quad (8)$$

Membership functions used for the input and output variables are trapezoidal and triangular 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 Frequent (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 1.

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.

Fig. 3 Input variable Lifetime

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 Message is broadcasted into environment by sink.
3: Hop count of 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 NWLAH.
10: Layers are formed in the network by finding IoU.
11: Total layers (Num_layer) are calculated
12: for  $\ell$  in Num_layer to 1 do ▷ each layer starting from farthest layer from BS
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

```

Fig. 4 Input variable Incoming Load

Fig. 5 Output variable Update Cycle (UC)

Table 1 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

6 Analysis

In this section, we will discuss the statistical analysis which 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 2.

Algorithm 3 Algorithm for Update Cycle Estimation**Input:** Set of CHs and their information**Output:** Update cycles w.r.t. CHs in all layers

```

1: Begin
2: for  $\ell$  in Num.layer to 1 do
3:   for each CH do
4:     Calculate incoming load ( $LD$ )
5:     Calculate lifetime ( $L$ )
6:     Incoming Load and Lifetime are passed to fuzzy logic to estimate Update Cycle ( $UC$ )
7:   end for
8: end for
9: return  $UC$ 

```

Table 2 Layers arrangements

Arrangments	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

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. This weighted sum is more w.r.t. arrangement A1. 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.

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 2 are stated in Table 3. Therefore, it can be concluded that there exists a significant difference between all the different arrangements of layers.

Table 3 F_{cal} values w.r.t. all arrangements for 50 random deployments

$F_{cal}-FND$	$F_{cal}-HND$	$F_c (\alpha = 0.01)$
7.0515	14.0470	3.88

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 2 are stated in Table 4 and 5. The F -critical value (F_c) is 6.90 for $\alpha = 0.01$.

Table 4 F_{cal} values w.r.t. FND (50 Nodes, $\alpha = 0.01$) **Table 5** F_{cal} values w.r.t. HND (50 Nodes, $\alpha = 0.01$)

A1	A2	A3	A4
A1	14.7703	17.7274	17.7274
A2		0.5070	0.0118
A3			0.4550

A1	A2	A3	A4
A1	41.0537	24.3755	24.3755
A2		0.6123	0.4390
A3			0.6123

From Table 4 and 5, we can conclude that there is a significant difference between the pairs i.e. A1 A2, A1 A3, and A1 A4. However, the pairs A2 A3, A2 A4, and A3 A4 is not having significant difference w.r.t. F_c value. From the Tables 4, 5 and the F_c value, we can say that the distribution of layers is best in arrangement A1 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 $200\text{m} \times 200\text{m}$ 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 section 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 6.

Table 6 Parameters List

Simulation Parameters	Value
Network area	$200\text{m} \times 200\text{m}$
Total nodes in the field	50, 100, 200
Communication Range	45m (max)
Data Packet Size	4000 bits
Initial Energy	$5J$
Energy Threshold	$0.05J$
Distance Threshold	60m
E_{elec}	50nJ/bit
ϵ_{fs}	10pJ/bit/m^2
ϵ_{mp}	0.0013pJ/bit/m^4
d_0	$\sqrt{\epsilon_{fs}/\epsilon_{mp}}$
EDA	5nJ/bit/signal

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: It represents the ratio of packets received at the sink to the total number of nodes in the field. The parameter value closer to one shows the efficiency of the network. The network can not reflect the real picture of the region being sensed if its value is low.
- (b) First Node Death (FND): It is the time period (in rounds) when the first node gets dies.
- (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 tells how many nodes are alive yet over the period of time. i.e. the number of the nodes whose battery is not exhausted.

Fig. 6 and 7 represents the number of rounds when FND, HND, and LND occurred in LEACH, FM-SCHEL, and different arrangements considered in proposed work when 100 and 200 nodes are considered. From the graph, it is clear that LEACH has the worst performance in terms of FND. FND of LEACH occurs at 578 rounds while for FM-SCHEL, it is 665 rounds. FND of HFLBSC w.r.t. A1, A2, A3, and A4 is 846, 690, 760, and 735 rounds. This is because of single-hop communication between sensor nodes and the sink in LEACH. In LEACH and FM-SCHEL, 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, we also see that the 1st arrangement (A1) is outperforming than the rest of the arrangements by giving FND as 713 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 the proposed HFLBSC algorithm for arrangement A1 has HND as 988 rounds and LND as 1130 rounds. While for LEACH, FM-SCHEL, and other arrangements (A2, A3, A4), HND is 764, 770, and 780, 881, 849 rounds and LND is 879, 970, and 970, 1011, 1002. 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.

Fig. 6 Different Node Deaths vs. Number of Rounds (100 Nodes)

Fig. 7 Different Node Deaths vs. Number of Rounds (200 Nodes)

Fig. 8 represents the energy consumption by different protocols over the rounds. Simulation results depict very clearly that energy consumption in the proposed HFLBSC is very less in comparison to LEACH and FM-SCHEL. 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.

Fig. 8 Energy Consumption vs. Number of Rounds (100 Nodes)

Fig. 10 and 11 is providing a more clear picture of the number of alive nodes over some particular round values. This figure shows till the end of network lifetime, nodes in arrangement A1 of proposed HFLBSC algorithm were able to survive than its competitors 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

Fig. 9 Energy Consumption vs. Number of Rounds (200 Nodes)

- 1 standard deviation in the residual energy is very less in arrangement A1 in comparison to others (refer
- 2 Fig. 12). This shows energy is dissipated in a more balanced way in arrangement A1.

Fig. 10 Number of Alive Nodes on particular Rounds (100 Nodes)

Fig. 11 Number of Alive Nodes on particular Rounds (200 Nodes)

Fig. 12 Standard Deviation of Residual Energy vs. Number of Rounds

- 3 Fig. 13 and 14 shows the packet delivery ratio w.r.t. LEACH, FM-SCHEL, and different arrange-
- 4 ments discussed when 100 and 200 nodes are deployed into the field. PDR for LEACH protocol is
- 5 very poor because while doing the selection of CHs, residual energy is not considered. Although PDR
- 6 for FM-SCHEL is more than LEACH. But in FM-SCHEL, optimized re-clustering is not considered
- 7 which leads to a lot of energy consumption. While in proposed HFLBSC, both of these are consid-
- 8 ered which ensures better results in proposed protocol. Multi-hop communication between the nodes
- 9 and BS maintains the connectivity for a longer duration in comparison to single-hop communication.

Fig. 13 also showcases that PDR is almost equal for the arrangements $A1$, $A3$, $A4$ in comparison to $A2$. But we have seen that nodes remain alive for longer duration in $A1$, so $A1$ promises to work efficiently for longer duration comparatively.

Fig. 13 PDR vs. Number of Rounds (100 Nodes)

Fig. 14 PDR vs. Number of Rounds (200 Nodes)

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 $A1$, it implies that network lifetime of the proposed algorithm ($A1$) 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 packet delivery ratio, network lifetime, energy consumption, 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 the University Grant Commission, New Delhi under the National Eligibility Test-Junior Research Fellowship scheme with Reference ID- 3361/(NET-JUNE 2015).

References

- [1] Aruna Pathak. "A Proficient Bee Colony-Clustering Protocol to Prolong Lifetime of Wireless Sensor Networks". In: *Journal of Computer Networks and Communications* 2020 (2020).
- [2] Ikram Daanoun, Baghdad Abdennaceur, and Abdelhakim Ballouk. "A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks". In: *Ad Hoc Networks* (2021), p. 102409.

- [3] Chinmay Mahabal, Hua Fang, and Honggang Wang. "Smart Spectrum Switching and Beamforming for Wireless Body Area Networks in Dynamic Environment". In: *Journal of Communications and Information Networks* 5.2 (2020), pp. 204–216.
- [4] Zhen Zhao, Guangming Li, and Menghui Xu. "An Improved Algorithm Based on LEACH Routing Protocol". In: *2019 IEEE 19th International Conference on Communication Technology (ICCT)*. IEEE. 2019, pp. 1248–1251.
- [5] P Visu et al. "Bio-inspired dual cluster heads optimized routing algorithm for wireless sensor networks". In: *Journal of Ambient Intelligence and Humanized Computing* (2020), pp. 1–9.
- [6] Manel Kortas et al. "The Energy-Aware Matrix Completion-Based Data Gathering Scheme for Wireless Sensor Networks". In: *IEEE Access* 8 (2020), pp. 30772–30788.
- [7] Zaib Ullah. "A Survey on Hybrid, Energy Efficient and Distributed (HEED) Based Energy Efficient Clustering Protocols for Wireless Sensor Networks". In: *Wireless Personal Communications* (2020), pp. 1–29.
- [8] Govind P Gupta and Binit Saha. "Load balanced clustering scheme using hybrid metaheuristic technique for mobile sink based wireless sensor networks". In: *JOURNAL OF AMBIENT INTELLIGENCE AND HUMANIZED COMPUTING* (2020).
- [9] Minghao Wang, Shubin Wang, and Bowen Zhang. "APTEEN Routing Protocol Optimization in Wireless Sensor Networks Based on Combination of Genetic Algorithms and Fruit Fly Optimization Algorithm". In: *Ad Hoc Networks* (2020), p. 102138.
- [10] Mohamed El Fissaoui, Abderrahim Beni-Hssane, and Mostafa Saadi. "Energy efficient and fault tolerant distributed algorithm for data aggregation in wireless sensor networks". In: *Journal of Ambient Intelligence and Humanized Computing* 10.2 (2019), pp. 569–578.
- [11] Amjad Mehmood et al. "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* (2017).
- [12] B Baranidharan and B Santhi. "DUCF: Distributed load balancing Unequal Clustering in wireless sensor networks using Fuzzy approach". In: *Applied Soft Computing* 40 (2016), pp. 495–506.
- [13] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. "Deep learning in mobile and wireless networking: A survey". In: *IEEE Communications Surveys & Tutorials* 21.3 (2019), pp. 2224–2287.
- [14] Wendi Rabiner Heinzelman, Anantha Chandrakasan, and Hari Balakrishnan. "Energy-efficient communication protocol for wireless microsensor networks". In: *Proceedings of the 33rd annual Hawaii international conference on system sciences*. IEEE. 2000, 10–pp.
- [15] Nikhil Kumar Singh, Ankit Kasana, and Vibhav Kumar Sachan. "Enhancement in lifetime of sensor node using Data Reduction Technique in Wireless Sensor Network". In: *International Journal of Computer Applications* 145.11 (2016), pp. 1–5.
- [16] Amjad Mehmood et al. "Energy-efficient multi-level and distance-aware clustering mechanism for WSNs". In: *International Journal of Communication Systems* 28.5 (2015), pp. 972–989.
- [17] Preetha Marappan and Paul Rodrigues. "An energy efficient routing protocol for correlated data using CL-LEACH in WSN". In: *Wireless Networks* 22.4 (2016), pp. 1415–1423.
- [18] Xu-Xing Ding et al. "Dk-leach: An optimized cluster structure routing method based on leach in wireless sensor networks". In: *Wireless Personal Communications* 96.4 (2017), pp. 6369–6379.
- [19] Ossama Younis and Sonia Fahmy. "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks". In: *IEEE Transactions on mobile computing* 3.4 (2004), pp. 366–379.
- [20] Mohit Sajwan, Devashish Gosain, and Ajay K Sharma. "Hybrid energy-efficient multi-path routing for wireless sensor networks". In: *Computers & Electrical Engineering* 67 (2018), pp. 96–113.

-
- [21] Salim EL Khediri et al. "A new approach for clustering in wireless sensors networks based on LEACH". In: *Procedia Computer Science* 32 (2014), pp. 1180–1185.
- [22] Ahmed Al-Baz and Ayman El-Sayed. "A new algorithm for cluster head selection in LEACH protocol for wireless sensor networks". In: *International journal of communication systems* 31.1 (2018), e3407.
- [23] Siddhi Sharma, Deepak Sethi, and PP Bhattacharya. "Artificial neural network based cluster head selection in wireless sensor network". In: *International Journal of Computer Applications* 119.4 (2015).
- [24] Songyut Phoemphon et al. "An energy-efficient fuzzy-based scheme for unequal multihop clustering in wireless sensor networks". In: *Journal of Ambient Intelligence and Humanized Computing* (2020), pp. 1–23.
- [25] Munuswamy Selvi et al. "An energy efficient clustered gravitational and fuzzy based routing algorithm in WSNs". In: *Wireless Personal Communications* 116.1 (2021), pp. 61–90.
- [26] Ying Zhang et al. "Fuzzy-logic based distributed energy-efficient clustering algorithm for wireless sensor networks". In: *Sensors* 17.7 (2017), p. 1554.
- [27] D Uma Maheswari and S Sudha. "Node degree based energy efficient two-level clustering for wireless sensor networks". In: *Wireless Personal Communications* 104.3 (2019), pp. 1209–1225.
- [28] Guihai Chen et al. "An unequal cluster-based routing protocol in wireless sensor networks". In: *Wireless Networks* 15.2 (2009), pp. 193–207.

Figures

Figure 1

Clustered environment in WSN

Figure 2

Flow chart of proposed HFLBSC algorithm

Figure 3

Input variable Lifetime

Figure 4

Input variable Incoming Load

Figure 5

Output variable Update Cycle (UC)

Figure 6

Different Node Deaths vs. Number of Rounds (100 Nodes)

Figure 7

Different Node Deaths vs. Number of Rounds (200 Nodes)

Figure 8

Energy Consumption vs. Number of Rounds (100 Nodes)

Figure 9

Energy Consumption vs. Number of Rounds (200 Nodes)

Figure 10

Number of Alive Nodes on particular Rounds (100 Nodes)

Figure 11

Number of Alive Nodes on particular Rounds (200 Nodes)

Figure 12

Standard Deviation of Residual Energy vs. Number of Rounds

Figure 13

PDR vs. Number of Rounds (100 Nodes)

Figure 14

PDR vs. Number of Rounds (200 Nodes)