Microsystem Technologies (2024) 30:393–409
https://doi.org/10.1007/s00542-023-05508-8

TECHNICAL PAPER

HCM: a hierarchical clustering framework with MOORA based cluster head selection approach for energy efficient wireless sensor networks

Ikkurthi Bhanu Prasad[1] · Saumitra Gangwar[1] · Yogita[3] · Satyendra Singh Yadav[2] · Vipin Pal[1,4]

Received: 15 December 2022 / Accepted: 8 July 2023 / Published online: 25 July 2023
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Energy-efficient operation is one of the prime goals of Wireless Sensor Networks (WSNs) because of the limited battery capacity and the harshness of the environment they are deployed in. Clustering techniques with multihop communication is one such technique that is suitable to achieve this goal. These approaches suffer from load balancing and hotspot issues because of the uneven energy consumption between cluster head (CH) nodes and member nodes (MNs). The Hierarchical Layer Balanced Clustering (HLBC) approach is a clustering framework that is highly effective in comparison to the current state of the art in addressing such issues. This work aims to optimize the process of CH selection and reduce intra-cluster communication distance (IACD) in the context of HLBC. For this purpose, Multi-Objective Optimization based on Ratio Analysis (MOORA) has been employed by considering three critical attributes, viz., residual energy, node centrality, and distance to relay, along with a Shannon entropy-based attribute weighting scheme for CH selection. The modified Dijkstra-based minimum spanning tree formation technique based on energy left, load, and distance to the relay node has been presented to reduce IACD and to distribute the load on MNs evenly. The proposed HCM scheme has been analyzed corresponding to two network scenarios, each for a homogeneous and heterogeneous network, based on three performance measures: node death rate, energy consumption, and network lifetime. The proposed framework has outperformed other state-of-the-art techniques for both homogeneous and heterogeneous cases in all considered scenarios. The First Node Death (FND) of EMUC, HLBC, and the proposed HCM approach are at rounds 201, 341, and 417 for scenario 1 and 254, 309, and 382 for scenario 2, respectively, for homogeneous cases. The proposed HCM protocol has achieved a percentage increase of [22–24]% in terms of FND corresponding to homogeneous network scenarios and a [23–27]% improvement in FND corresponding to heterogeneous network scenarios.

✉ Yogita
yogita@nitkkr.ac.in

✉ Vipin Pal
vipinpal@nitm.ac.in

Ikkurthi Bhanu Prasad
bhanuprasad@nitm.ac.in

Saumitra Gangwar
saumitragangwar@nitm.ac.in

Satyendra Singh Yadav
satyendra@nitm.ac.in

1 Department of Computer Science and Engineering, National Institute of Technology, Meghalaya, India

2 Department of Electronics and Communication Engineering, National Institute of Technology, Meghalaya, India

3 Department of Computer Engineering, National Institute of Technology, Kurukshetra, India

4 Department of Computer Engineering, Netaji Subhas University of Technology, New Delhi, India

1 Introduction

Effective monitoring and information analysing capabilities in harsh and critical regions has led the Wireless Sensor Networks (WSNs) for a broader range of potential applications and has gathered a lot of research interest among the researchers. A typical sensor network usually comprises hundreds or thousands of sensor nodes and, at the minimum, a sink node or base station (Mondal et al. 2022; Bhattacharjya et al. 2021; Prasad et al. 2021). The lifetime of the sensor nodes, which depends on limited battery capacity, is critical in order to maintain effective monitoring of the target area (Soni and Mallick 2017; Assari et al. 2020). After deployment, sensor nodes are deemed autonomous; their lifespan is determined by how they optimize their energy usage. The key to optimizing nodes energy usage is to design energy-efficient data communication methods. Clustering approaches are proven to be energy efficient in which the nodes in the sensor network are divided into clusters (Singh et al. 2021; Priyadarshi et al. 2018). Each cluster has a Cluster Head node (CH) and Member Nodes (MN). The task of the member nodes is to monitor the sensing region and communicate the sensed information to its respective CH node. CH nodes aggregate the received information and communicate to sink node by employing direct or multi-hop communication (Priyadarshi et al. 2020; Gangwar et al. 2023; Prasad et al. 2023).

Most of the energy consumption across the network results from sensed data communications by the sensor nodes to the sink. Multi-hop communication is proven to be an energy-efficient approach in the clustering techniques for forwarding the sensed data. But it results in uneven energy consumption across the nodes in the network, resulting in energy holes or hotspot problems (Jain et al. 2022; Elkamel et al. 2019; Prasad et al. 2020). Hence, achieving the load balancing among the nodes in an energy-efficient manner by eliminating the hotspots in the network is the prime requirement for the clustering protocols. The majority of the clustering solutions proposed in the literature tried to balance the load among the nodes in the network by forming equal-sized or unequal-sized clustering architectures. The main idea behind developing equal-sized clustering architectures is to have an equal amount of load on all the clusters (Elshrkawey et al. 2018; Pal et al. 2015a). However, these solutions still persist with hotspot problems as multi-hop communication is used for intra-cluster data communication. To overcome this problem, unequal-sized clustering solutions are proposed by having small-sized clusters formed near the base station to load balance the network (Amini et al. 2020; Baranidharan and Santhi 2016). But the problem with unequal-sized clustering architectures is that the small-sized clusters formed near the base station generate more frames in a round and have higher energy consumption compared to other clusters, leading to uneven load balancing. Additionally, they also suffer from synchronization issues. To overcome these problems, authors in Prasad et al. (2021) proposed the HLBC scheme, which is a hierarchical layer balanced solution that assimilates the benefits of both equal and unequal sized clustering solutions and eliminates their persisting issues. The HLBC scheme can still be improved in the aspects of: (i) CH selection and (ii) intra-cluster data communications.

With the prime motivation of designing energy efficient communication paradigm, the objectives of the proposed work are:

- To select the best-suitable CHs by considering multiple parameters such as node centrality, residual energy, and distance to relays for every cluster in the HLBC architecture that can help in achieving balanced energy consumption.
- To minimize the intra-cluster communication distances in each cluster in the HLBC architecture, which further improves the energy efficiency of the proposed protocol.

CHs play a crucial role in the network because they perform essential tasks like intra-cluster data gathering and inter-cluster data communication. Hence the best-suited nodes must be elected for this role among the available ones. While electing CHs for a cluster, all the essential characteristics, such as node centrality, residual energy, and distance to relays, should be considered. Multi-criteria Decision-making is a technique where a decision is made by considering multiple relevant criteria. These methods can be best suited for CH selection because the sensor nodes are evaluated based on various criteria before they are selected as CHs. The intra-cluster and inter-cluster communication distances should be minimum in a clustering protocol to make it energy efficient. At the intra-cluster level, multi-hop data communication can be used instead of direct communication. This aids in further load balancing the network and improving the network lifetime. Greedy algorithms can be employed to form routing paths in the network. At every step, greedy techniques will find the optimal solutions by considering the best available choice. While constructing Intra-cluster communication paths in a cluster, all the influential parameters that have an effect on the node's energy consumption should be taken into account.

In this paper, a Hierarchical Clustering framework with MOORA based Cluster head Selection approach (HCM) has been designed and has been implemented over HLBC protocol. To achieve the first objective of selecting the best suitable CHs in the network, MOORA based CH selection technique has been proposed. The outcome of the proposed method is the best appropriate CHs for each cluster in the HLBC architecture. The sensor nodes that have positive residual energies in a cluster contest for the role of CH and are designated as the competing alternatives. These competing alternatives are evaluated based on ratio analysis, where a multi objective function is optimized. The various criteria that were considered while assessing the alternatives are Residual Energy (RE_i), Node Centrality (NC_i) and distance to Relay (dtR_i). The weights of each criterion are dynamically calculated using the Shannon-Entropy technique. To attain the second objective, a modified Dijkstra-based Minimum Spanning Tree (MST) is built in each cluster to reduce intra-cluster communication distances. The usual distance parameter, along with the essential residual energy and load on relay node parameters, is also considered while setting up intra-cluster data communication paths among member nodes to CHs in each cluster to construct an MST.

The specific contributions of the proposed protocol are as follows:

- Proposed a MOORA-based CH selection technique where CHs are selected on the basis of multiple relevant criteria that helps in selecting optimal CH in a cluster.
- For estimating the weights of the considered criteria Shannon entropy technique is utilized, which determines the weightages of the criteria dynamically on the basis of input measures.
- Proposed a modified Dijkstra-based MST formation technique that helps to form load-balanced intra-cluster communication paths in the clusters.
- To give a statement about the scalability and the adaptability of the presented protocol, performance evaluation is done on two different network scenarios (120×120) m and (240×240) m on the basis of various parameters such as node death rate, network lifetime and energy consumption.
- The performance of the proposed protocol is assessed in both homogeneous and heterogeneous cases.

The paper organization is done as follows: in Sects. 2 and 3, the summary of related works and the preliminaries are presented. Section 4 describes the proposed protocol and its implementation. In Sect. 5, the performance assessment of the proposed scheme and comparison to other simulated protocols in both homogeneous and heterogeneous cases has been reported. Finally, in Sect. 6, we concluded the research contribhtion of the presented paper.

2 Related works

Clustering architecture based approaches have been employed successfully for energy efficiency issue in WSN. This section outlines various clustering paradigm for WSN.

LEACH (Heinzelman et al. 2000) is one of the first hierarchical clustering protocol proposed in the literature. LEACH employs a probability based CH selection technique for selecting CHs. It is sensitive to heterogeneity, and its performance is unstable in the network due to its CH selection strategy. The CH election strategy selects CHs randomly without giving any preference to nodes with higher residual energies.

To prolong the stability period of a WSN, Smaragdakis et al. (2004) proposed Stable Election Protocol (SEP) in which the sensor nodes are classified as normal nodes and advanced nodes. Advanced nodes have more energy compared to normal nodes. Weighted probabilities are computed based on the initial energy of the nodes with respect to other nodes and help in deciding the CH nodes for a round. The advanced nodes tend to become CHs more often. However, SEP is more suitable for two-level heterogeneous environments but has proven inadequate for higher-level heterogeneous networks. To address this issue, Qing et al. (2006) proposed a heterogeneous aware clustering algorithm known as the DEEC protocol, which operates well in multi-level heterogeneous environments. Uniform node energy consumption is achieved by rotating the CH role among all the nodes. Each node claims itself as CH with a probability computed using the ratio of remaining energy and the average network energy. DEEC adapts the rotating epoch, giving the nodes with high initial energy more chances to become CHs. Simulation results reveal that the DEEC protocol performs better than the SEP protocol in terms of throughput and attaining a more extended network lifetime. The authors in Pal et al. (2015b) investigated the impact on the performance of popular heterogeneous clustering approaches such as LEACH, SEP, and DEEC regarding the placement or positioning of heterogeneous nodes in a sensor network. To analyze the performance of the approaches above, the worst, best, and average case placements of heterogeneous nodes are considered.

Suniti et al. in Dutt et al. (2018) proposed the CREEP protocol for two-level heterogeneous WSN, where the number of CHs in WSN is fixed to a certain value to reduce the amount of computational overhead involved in CH selection and also to maintain an appreciable network lifetime. In every round, $K\%$ of the nodes having the highest residual energy among the total alive nodes are termed CHs. For data communication, the dual hopping technique is employed. The nodes surrounding the sink radius 'R' use direct communication. The remaining nodes employ multi-hop communication by selecting a node within the sink radius as the relay node.

Most clustering algorithms depend on the random number generated for choosing CHs. Such dependence may lead to poor performance (i.e., sensor nodes having less battery capacities get appointed as CHs) of the clustering protocols. To avoid this, the authors in Yogita et al. (2022) proposed a distributed dynamic clustering protocol for heterogeneous WSNs, which reconditions the random number generated depending upon the residual energy parameter. Reconditioning modifies the random number generation such that a small number is generated for high residual energy nodes, whereas a large number is generated for low residual energy nodes.

Benelhouri et al. (2022) investigated the effect of heterogeneity on the MGEAR protocol for prolonging the lifetime of the network. The sensor network is partitioned into various regions, and a gateway node is placed at the network's center. Sensor nodes in the first region send data directly to the sink, whereas sensor nodes near the center provide data to the gateway node. The remaining nodes are partitioned into two equal regions, and the clustering technique is employed to communicate the sensed data. The percentage of node energy to average network energy is utilized to compute the CH election probability for CH selection. Simulation results reveal balanced energy consumption among the nodes and an enhancement in the network's lifetime. Hung et al. (2020) proposed an economical energy routing technique for heterogeneous WSNs where several WSNs deployed in the same environment cooperatively transmit the relay packets generated by a WSN. For optimal route establishment, various parameters, like the residual energy of primary sensor nodes and neighbor nodes, along with the transmission direction of the packets, are considered.

In El Alami and Najid (2019), Alami et al. proposed an enhanced clustering hierarchy (ECH) approach for minimizing energy consumption in WSNs. In this approach, neighboring sensor nodes that generate redundant data are put into a sleep–wake cycle so that not all the nodes are in the wake cycle simultaneously. In each round, there will be sleeping nodes, waking nodes, and CHs. The sleeping nodes will be in sleep mode to save energy, while the nodes in waking mode will be responsible for sensing data and forwarding the data to the CH. This scheme saves sending redundant data to the base station and thus minimizes energy consumption in the network and is applicable for both homogeneous and heterogeneous networks. Hamzah et al. (2019) proposed a fuzzy method for selecting CHs and cluster formation by considering parameters like residual energy, distance from the BS, the density of neighboring nodes, compactness of the neighboring nodes, and the location suitability computed based on the average of the local consumed energy of the adjacent nodes. The selected CHs using this fuzzy model ensures that optimal CHs are chosen, and the network longevity is guaranteed by considering additional parameters that impact the energy consumption inside a network. Additionally, the Gini index is used to measure the distribution of remaining energy among the sensor nodes.

Wang et al. (2022) used a combination of a self-organizing map neural network and firefly algorithm for optimal cluster formation and CH selection. Clusters are formed inside the network using the self-organizing map neural network, after which a single node from each group will act as the initial population for the firefly algorithm. The fitness function is so designed that the clusters near the base station will be smaller than the clusters which are far away from the base station to prevent the hotspot problem. For inter-cluster routing, the ant colony optimization algorithm has been used so that CHs far from the BS avoid sending the data directly to the BS, which will incur a high energy cost and lead to the premature death of CH nodes. Another feature of this work includes an improved intra-cluster communication mechanism in which nodes do not necessarily send data to their respective CHs based on their allotted timeslot but only when they have valid data to be transmitted.

Shyama et al. (2022), proposed a fault-tolerant routing path identification with genetical swarm optimization (FTGSO) to make the network more robust and adaptive to faults. Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) are combined to form Genetical swarm optimization (GSO) that brings the best of both techniques in determining faulty nodes in a cluster as well as a fault-free route in the network. The remaining energy, coverage, communication costs, and location all play a role in choosing the CH. Additionally, GSO introduces a fault-free routing path, and a self-healing technique is used to address any network connectivity problems and restore regular system operation. Vahabi et al. (2022) proposed an energy-aware method (EAM) that exploits a greedy approach in chain-based routing to transmit data between only consecutive layers based on nodes' remaining energy and distance to the sink. The proposed method applies to only rectangular and circular networks, which is a limitation of the work as depending on the application, the deployment area might deviate from these two scenarios.

Hoang et al. (2013) presented the Harmony search algorithm cluster-based protocol (HSACP), which makes use of the harmony search algorithm to reduce overall intra-cluster distances while equitably distributing energy consumption across all nodes. The goal of the threshold-sensitive energy-efficient routing protocol (GATERP) proposed by Mittal et al. (2019), which uses genetic algorithm to choose the best CHs, is to reduce total intra-cluster and inter-cluster distances while minimizing energy consumption for both the data aggregation and data transmission phases. The Proficient Bee Colony Clustering Protocol (PBC-CP) (Pathak 2020) is a cluster-based algorithm based on an artificial bee colony that chooses energy-efficient paths to forward sensed data to the BS using multi-hop routing while taking into account residual energy, distance to the BS, and the degree of nodes. Le-Ngoc et al. (2022) proposed a distributed fuzzy clustering scheme incorporated with an improved squirrel search algorithm to achieve better network lifetime. During the clustering phase of the protocol, parameters like residual energy, node centrality, node type, and the number of neighboring nodes were considered to achieve an improved cluster formation. The optimized squirrel search algorithm (OSSA) is used to fine-tune the fuzzy logic controller and find multi-hop routing paths to the BS.

Esmaeili et al. (2022) proposed a combined fuzzy firefly algorithm and random forest (FFA-RF) that operates in two phases. The first phase is the training and testing using FFA and RF, whereas the actual application is made in the online phase for newer unseen network instances. In the first phase, networks with different configurations are formed by varying the total number of sensor nodes, network size, aggregation factor, location of BS, and network lifetime definition. Afterward, the FFA algorithm is used for clustering in these instances, and training and test data sets are generated by repeating FFA on all the cases. The test dataset is used to assess the generalizability of the trained model, while the training dataset is used to train the RF model using the bagging approach. The trained RF model is used to choose the appropriate CHs at each round in the online phase for any new WSN. PS-SFLA, which uses the SFLA or shuffling frog leaping algorithm for fuzzy multi-hop clustering protocol, was proposed by Fanian et al. (2021). The authors assessed parameters including alive nodes, packets received by the BS, and WSN lifetime while contrasting the PS-SFLA with other clustering techniques. They claimed that PS-SFLA outperforms other clustering algorithms and extends WSN lifetime by using superior fuzzy input parameters. A modified version of LEACH called LEACH-PSO is proposed by Thiagarajan (2020) that uses particle swarm optimization for CH selection. By balancing energy usage by choosing a close to ideal set of CHs, it seeks to maximize the performance of LEACH.

The clustering strategy described in Alia (2018) aims to increase WSN lifetime and minimize power usage. This strategy utilizes a decentralized fuzzy clustering method to build the WSN's infrastructure first and then automatically uses a harmony search-based algorithm to select the best clusters. Once clusters have formed, the sensor nodes transmit their sensed data to the sink nodes. The authors claimed that their method could determine the ideal number of CHs, improve data transmission to sink nodes and prolong the lifespan of the WSN. EEFCM-DE, a hybrid power-aware clustering strategy utilizing a differential evolution algorithm and FCM clustering method, was introduced by Sharma et al. (2019). In this approach, the clusters are formed using the fuzzy c-means (FCM) algorithm, and the CH is chosen using the differential evolution (DE) algorithm. The fitness of each node is determined for the CH selection using a devised fuzzy inference method. The authors demonstrated that EEFCM-DE is more energy-efficient than other clustering techniques by carrying out the necessary simulations. Gangwar et al. (2022) proposed a hierarchically structured cluster routing protocol that uses the concept of m-way trees to distribute the load among nodes inside a cluster properly. This approach creates a hierarchical structure inside a cluster where each node can have a maximum of $'m'$ child nodes. The parent nodes will collect and aggregate the data received from child nodes before forwarding it. The advantage of this approach is that it can be applied over any clustered WSN to minimize energy consumption.

Based on the literature review done for both homogeneous and heterogeneous networks, we have concluded that none of the previous works consider an approach that combines a Shannon entropy-based weight estimation with the multi criteria based decision making method for CH selection to form an balanced hierarchical layer clustering to achieve better energy efficiency and load balancing.

3 Preliminaries

This section provides information about the basic idea, network model, the radio model, and the notations used in this work.

3.1 Basic idea

The clustering protocols presented in the literature are operated across the network over rounds. Any clustering protocol operation happens until all nodes deployed over the network have exhausted their energies. In every round, the clustering protocols operate in two phases: the setup phase and the steady state phase. In the setup phase, the sensor nodes are organized into hierarchical clusters, and a node from each cluster is selected for the role of CH, while the remaining nodes act as MNs. The TDMA schedules are transmitted to each MN by the CH node, where every MN is assigned a specific data transmission slot to transmit the sensed data in that round. In the steady state phase, the CH nodes collect the data sent by all the MNs of that particular cluster, aggregate it, and forward the aggregated data to the sink node via direct or multi-hop communication.

3.2 Network model

The following are the assumptions that are made regarding the network model.

- Uniform deployment of sensor nodes is done throughout the sensing region and are immovable after the deployment.
- Sensor nodes are not rechargeable after deployment.
- The resource-rich base station collects sensor node data from outside the sensing region.
- Sensors operate in active or low-power sleep mode.
- Sensor nodes can die due to energy depletion only.
- Sensor nodes have heterogeneous capabilities regarding the initial amount of energy.

3.3 Radio model

As used in Baranidharan and Santhi (2016), this work uses the first-order radio model for data transmission across the simulated network. The energy consumption of a node depends on the distance and the amount of data that needs to be communicated. If the value of the communication distance is less than the threshold (i.e., d_0), energy consumption in a node happens according to the free space model that is proportional to d^2. Otherwise, the energy consumption in a node happens according to the multipath model, which is proportional to d^4.

For a node to transmit a K_1-bit data packet in the network, it has an energy consumption of:

$$Ene_{Tx}(K_1, d) = \begin{cases} E_{elec} \times K_1 + \epsilon_{fs} \times d^2 & if \ d < d_0 \\ E_{elec} \times K_1 + \epsilon_{mp} \times d^4 & if \ d > = d_0 \end{cases} \tag{1}$$

where E_{elec} is the transmitter unit energy dissipation per bit, ϵ_{fs} and ϵ_{mp} are the amplification energies required for free space and multipath models, and d_0 is the distance threshold.

Similarly, for a node to receive K bits of data, the energy consumption involved is:

$$Ene_{Rx}(K_1) = E_{elec} \times K_1 \tag{2}$$

where E_{elec} is the receiver unit energy dissipation per bit.

3.4 Notations

All notations used in this work are shown in Table 1.

4 Proposed work

In the current work, the proposed HCM framework has been implemented over HLBC approach even though the HCM framework can be implemented over any existing basic clustering paradigm. The working of the proposed HCM protocol happens in rounds and is shown in the Fig. 1. The proposed protocol operates in the following phases: Node deployment and Cluster set-up phase, CH selection phase, routing path establishment phase, Data collection and forwarding phase. During the node deployment and cluster set-up phase, sensor nodes are uniformly deployed throughout the network area. Following deployment, the network's clusters are set up in accordance with the basis clustering architecture scheme. In the current work, HLBC has been considered as basis clustering approach. The selection of CHs for every cluster is performed in the CH selection phase. In the routing paths establishment phase, intra-cluster and inter-cluster data communication paths are set up in the entire network, followed by TDMA schedule assignment. In the final phase, data collection and forwarding take place. Figure 2 shows the proposed protocol's clustering architecture after the network successfully passes through all these phases in a round. Algorithm 1 shows the procedural implementation of proposwed HCM protocol. Table 1 represents the various notations used in this paper.

Algorithm 1 Proposed HCM Protocol

procedure HCM(SN, E)
    Uniform distribution of n sensor nodes $SN[i]$ across the network area.
    Partitioning network area into m layers and cluster set-up happens acoording to HLBC Architecture.
    while (operating nodes) >0 **do** ▷ Nodes with positive residual energy; SN(i).E>0
        A node $[CH_i]_X$ is selected as CH in a cluster X using $Algorithm$ 2.
        Intra-cluster Communication paths are established among nodes $SN(i)$ and $SN(j)$ of a cluster X using $Algorithm$ 3.
        Inter-cluster communication paths are established between $[CH_i]_X$ and $[CH_j]_Y$ in the direction of the sink node.
        $[CH_i]_X$ transmits TDMA schedules to all corresponding member nodes. Data aggregation and forwarding is performed.
    end while
end procedure

4.1 Node deployment and cluster set-up phase

In this phase, the uniform deployment of sensor nodes in the network area is done. After deployment, the clusters are formed in the network according to the HLBC clustering architecture scheme. The clustering architecture of HLBC splits the entire network area into numerous layers of equal

Table 1 Notations used in this paper

Notation	Definition
RE_i	Residual energy of node i
dtR_i	Distance to relay node from node i
NC_i	Node centrality of node i
PS_j	Performance score of alternative j
L_j	Load on node j
E_j	Energy of node j
d_{ij}	Distance between nodes i and j

sizes and an equal number of sensor nodes across each layer. Each layer further splits into several clusters. The layer's distance from the sink determines its cluster size. Large-sized clusters are formed in the layer near the sink, and small-sized clusters are formed in a layer far from the sink. Let us consider a scenario where the sink partitions the network consisting of n nodes into m layers (i.e., $1, 2, 3, \ldots, m$). Let there be k clusters at layer-1 of sizes $\frac{n}{m \times k}$. Then the number of clusters formed at the subsequent layers varies as $2^{(m-1)} \times k$. At the end of this phase, the network gets partitioned into clusters, and the large-sized clusters are formed near the base station to avoid the energy hole problem.

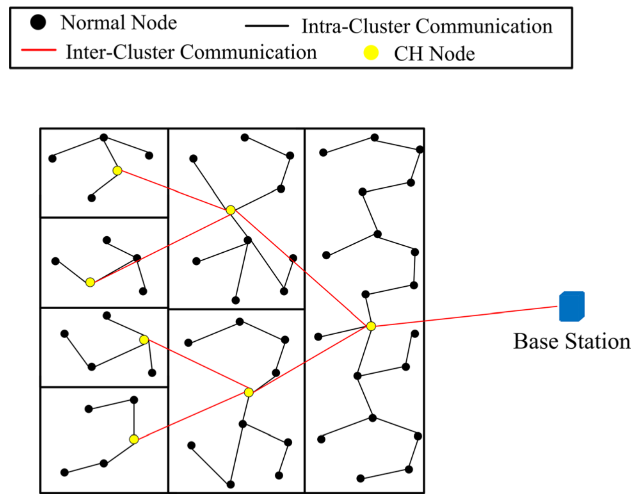

Fig. 2 Proposed HCM clustering architecture

4.2 CH selection phase

In this phase, CHs are elected for every cluster formed as per the basis approach and in the current work as according to the HLBC clustering architecture. The CHs play a crucial role in the network because they perform essential tasks like intra-cluster data gathering and inter-cluster data communication. Hence the best-suited nodes must be elected for this role among the available ones. The vital characteristics for a node to become an ideal CH in a cluster are:

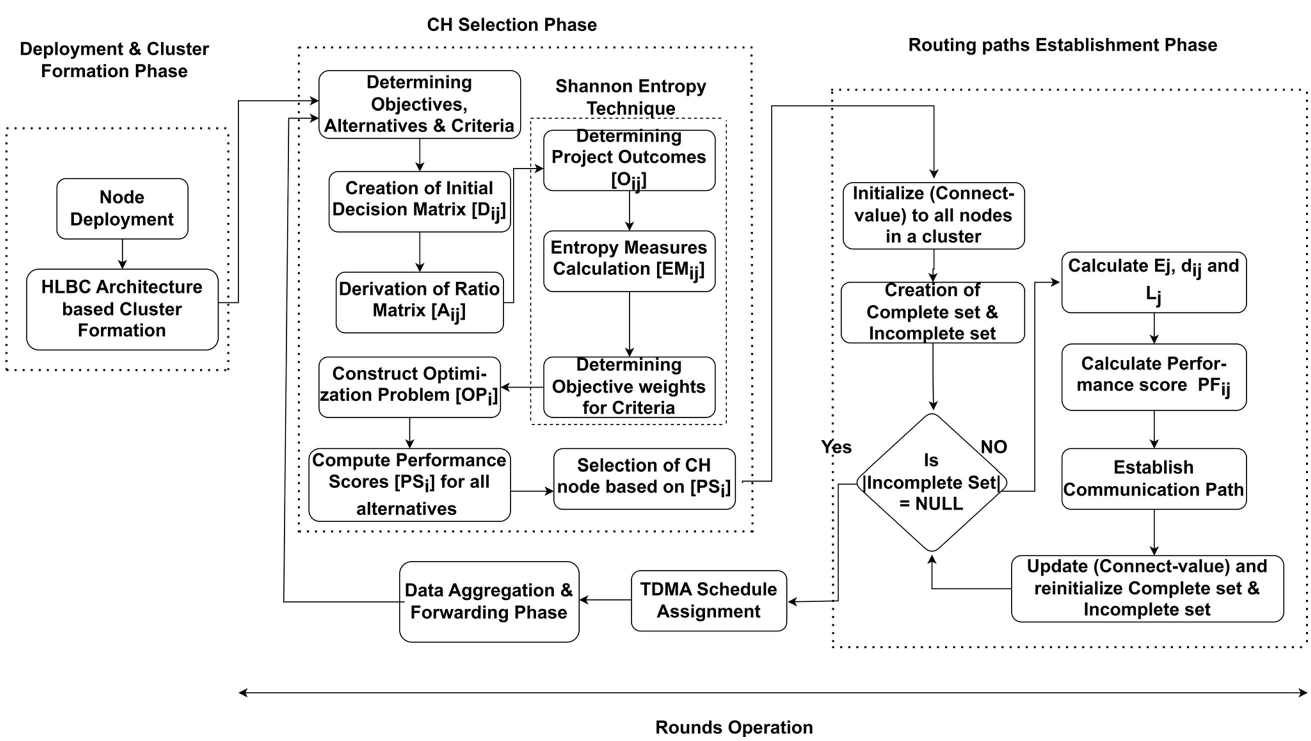

Fig. 1 Working of the proposed HCM protocol

- It should have higher residual energy when compared to other nodes, as CHs consume relatively more energy to perform the assigned tasks than non-CH nodes.
- It should be closer to the relay or sink node to avoid higher energy dissipation while performing inter-cluster communications.
- It must be relatively closer to its member nodes (i.e., overall intra-cluster distances in a cluster should be minimal).

Hence, it is essential to elect a node for the role of CH that achieves a balance among all these vital characteristics. In this work, the CH selection is made by employing a Multi-Objective Optimization based on Ratio Analysis (MOORA) method (Hwang and Yoon 1981; Chakraborty 2011). MOORA is a multi-criteria decision-making method where all possible alternatives are evaluated based on some relevant criterion, and each alternative's performance score is generated as an outcome. Next, the best alternatives can be selected based on performance scores.

Algorithm 2 : Proposed CH Selection Phase

Input: Cluster X, Alternatives $[N]$, Criteria $[C]$.
Important Variables: A_{ij}, D_{ij}, O_{ij}, EM_{ij}.
Output: CH node selected in a cluster X.

1: Determine Objectives, Alternatives [N] and the corresponding Criteria [C] corresponding to cluster X.
2: Initialize Decision matrix D_{ij} with alternatives i corresponding to criteria j as shown in $Eqn.(3)$.
3: Generate Ratio matrix A_{ij} from decision matrix D_{ij} using

$$[A_{ij}] = \frac{D_{ij}}{\sqrt{\sum_{k=1}^{N} D_{ij}}}$$

.
4: Determine beneficial and non-beneficial criteria corresponding to objectives.
5: Derive Project outcomes O_{ij} from Decision matrix D_{ij} using

$$[O_{ij}] = \frac{D_{ij}}{\sum_{k=1}^{N} D_{ij}}$$

.
6: Compute Entropy measures EM_{ij} using

$$[EM_{ij}] = -1 \times \frac{1}{ln(u)} \times \sum_{i=1}^{u} O_{ij} \times \ln(O_{ij})$$

.
7: Determine objective weights for all criteria C_j such that

$$[W_j] = \frac{(1 - EM_j)}{\sum_{j=1}^{v} (1 - EM_j)}$$

.
8: Construct the Optimization problem OP_i such that

$$[OP_i] = \sum_{j=1}^{y} W_j \times A_{ij} - \sum_{j=(y+1)}^{C} W_j \times A_{ij}$$

9: Compute Performance score PS_i for all alternatives using $Eqn.(9)$.
10: Sort alternatives in ascending order on the basis of performance score PS_i.
11: The node i having highest performance score PS_i is selected for CH role in cluster X.

The operation of the proposed MOORA-based CH selection method is done according to Algorithm 2 and has been explained in the following steps:

- **Step-I:** In this step, the objectives, alternatives, and criteria are to be determined. The objective here is to select the best possible CH for each cluster. The alternatives are the sensor nodes with positive residual energies competing for the role of CH. These competing alternatives are evaluated against three essential criteria: $\{$Residual-Energy (RE_i), distance-to-Relay (dtR_i), and Node-Centrality $(NC_i)\}$.
- **Step-II:** In this step, an initial decision table or matrix is created based on the available information about the alternatives and the criteria. A row in the decision matrix represents the information about a particular alternative (i.e., sensor node), and a column represents the value of a specific criterion. The decision matrix D_{ij} (Chakraborty 2011) is formed as shown below:

$$[D_{ij}]_{u\times v} = \begin{bmatrix} D_{11} & D_{12} & \dots & D_{1v} \\ D_{21} & D_{22} & \dots & D_{2v} \\ \vdots & \vdots & \dots & \vdots \\ D_{u1} & D_{u2} & \dots & D_{uv} \end{bmatrix} = \begin{bmatrix} RE_1 & dtR_1 & NC_1 \\ RE_2 & dtR_2 & NC_2 \\ \vdots & \vdots & \vdots \\ RE_u & dtR_u & NC_u \end{bmatrix} \tag{3}$$

  where, D_{ij} shows the performance measure of alternative i in correspondence to criteria j, u represents the competing alternatives (i.e., Sensor nodes of a cluster having positive residual energy), and v represents the criteria on which the competing alternatives are evaluated.
- **Step-III:** In this step, each alternative's performance on a criterion is compared to a representative denominator for all alternatives on that criterion. The square root of each alternative's sum of squares is used as the denominator. The ratio matrix A_{ij} (Chakraborty 2011) is derived as follows:

$$[A_{ij}]_{u\times v}=\begin{bmatrix}\frac{RE_1}{\sqrt{\sum_{j=1}^{u}RE_j^2}} & \frac{dtR_1}{\sqrt{\sum_{j=1}^{u}dtR_j^2}} & \frac{NC_1}{\sqrt{\sum_{j=1}^{u}NC_j^2}}\\ \frac{RE_2}{\sqrt{\sum_{j=1}^{u}RE_j^2}} & \frac{dtR_2}{\sqrt{\sum_{j=1}^{u}dtR_j^2}} & \frac{NC_2}{\sqrt{\sum_{j=1}^{u}NC_j^2}}\\ \vdots & \vdots & \vdots\\ \frac{RE_u}{\sqrt{\sum_{j=1}^{u}RE_j^2}} & \frac{dtR_u}{\sqrt{\sum_{j=1}^{u}dtR_j^2}} & \frac{NC_u}{\sqrt{\sum_{j=1}^{u}NC_j^2}}\end{bmatrix} \tag{4}$$

where RE_j, NC_j represents the residual energy and node centrality of node j, dtR_j shows the distance of node j to the relay.

- **Step-IV:** In this step, the beneficial and non-beneficial criteria are determined. The beneficial criteria have a positive influence, and the non-beneficial criteria have a negative impact while optimizing the objective function. In this work, the positive criteria are residual energy (RE_i) because its higher value is desired, whereas the negative criteria are distance to relay (dtR_i) and node centrality (NC_i) because the lesser values of these particular criteria are desirable. Next, the weights are derived for each criterion using the Shannon-entropy technique. Entropy is a concept in information theory that can be thought of as a measure of how uncertain a discrete probability distribution is. Entropy can be used efficiently in decision-making since it measures data contrasts and clarifies average intrinsic information. Here Shannon entropy determines the weights of the criterion dynamically on the basis of input measures. Steps (a), (b), and (c) represent the working procedure of Shannon-entropy technique.

  Step-(a): Project outcomes (O_{ij}) (Hwang and Yoon 1981) are derived by normalizing the decision matrix (D_{ij}). Normalizing raw data eliminates measurement unit and scale anomalies. This process standardises scales and units across criteria to enable comparisons.

$$[O_{ij}]=\frac{D_{ij}}{\sum_{j=1}^{u}D_{ij}}=\begin{bmatrix}\frac{RE_1}{\sum_{j=1}^{u}RE_j} & \frac{dtR_1}{\sum_{j=1}^{u}dtR_j} & \frac{NC_1}{\sum_{j=1}^{u}NC_j}\\ \frac{RE_2}{\sum_{j=1}^{u}RE_j} & \frac{dtR_2}{\sum_{j=1}^{u}dtR_j} & \frac{NC_2}{\sum_{j=1}^{u}NC_j}\\ \vdots & \vdots & \vdots\\ \frac{RE_u}{\sum_{j=1}^{u}RE_j} & \frac{dtR_u}{\sum_{j=1}^{u}dtR_j} & \frac{NC_u}{\sum_{j=1}^{u}NC_j}\end{bmatrix} \tag{5}$$

  Step-(b): Entropy measures (EM_{ij}) (Hwang and Yoon 1981) are computed on the basis of project outcomes (O_{ij}). It is done to calculate the contribution of each possible outcome concerning total entropy. To do this, multiply the outcome's probability by the outcome's probability's logarithm, narrowing down the range of potential values.

$$\begin{aligned}[EM_{ij}] &= -1\times\frac{1}{ln(u)}\times\sum_{i=1}^{u}O_{ij}\times\ln(O_{ij})\\ &=\left[\frac{-1}{ln(u)}\times Z_{RE}\quad \frac{-1}{ln(u)}\times Z_{dtR}\quad \frac{-1}{ln(u)}\times Z_{NC}\right]\end{aligned} \tag{6}$$

  where the values of $Z_{RE}, Z_{dtR}, and Z_{NC}$ are:

$$Z_{RE}=\sum_{i=1}^{u}\frac{RE_i}{\sum_{j=1}^{u}RE_j}\times\ln\left(\frac{RE_i}{\sum_{j=1}^{u}RE_j}\right)$$
$$Z_{dtR}=\sum_{i=1}^{u}\frac{dtR_i}{\sum_{j=1}^{u}dtR_j}\times\ln\left(\frac{dtR_i}{\sum_{j=1}^{u}dtR_j}\right)$$
$$Z_{NC}=\sum_{i=1}^{u}\frac{NC_i}{\sum_{j=1}^{u}NC_j}\times\ln\left(\frac{NC_i}{\sum_{j=1}^{u}NC_j}\right)$$

  Step-(c): Based on Entropy measures (EM_{ij}), the objective weights are determined for each criterion (Hwang and Yoon 1981).

$$[W_j]=\frac{(1-EM_j)}{\sum_{j=1}^{v}(1-EM_j)} \tag{7}$$

  The weights of the criteria W_{RE}, W_{dtR}, and W_{NC} are computed as follows:

$$W_{RE}=\frac{(1-Z_{RE})}{(1-Z_{RE})+(1-Z_{dtR})+(1-Z_{NC})}$$
$$W_{dtR}=\frac{(1-Z_{dtR})}{(1-Z_{RE})+(1-Z_{dtR})+(1-Z_{NC})}$$
$$W_{NC}=\frac{(1-Z_{NC})}{(1-Z_{RE})+(1-Z_{dtR})+(1-Z_{NC})}$$

- **Step-V:** In this step, the optimization problem (OP_i) is constructed by adding all the beneficial criteria attributes and subtracting all the non-beneficial criteria attributes (Hwang and Yoon 1981).

$$[OP_i]=\sum_{j=1}^{y}W_j\times A_{ij}-\sum_{j=(y+1)}^{v}W_j\times A_{ij} \tag{8}$$

  where, the criteria $\{1,2,\ldots,y\}$ are the beneficial criteria and $\{(y+1),\ldots,v\}$ are the non-beneficial criteria, W_j represents the criteria weights, and A_{ij} represents the ratio matrix values.

  In this work, the beneficial criteria are residual energy where as the non-beneficial criteria are node centrality and distance to relay. Next, the performance scores for each alternative are computed from the optimization problem. The performance score of an alternative i is computed as shown below:

$$PS_i = W_{RE} \times A_{(i,RE)} - W_{dtR} \times A_{(i,dtR)} - W_{NC} \times A_{(i,NC)} \quad (9)$$

where W_{RE}, W_{NC} and W_{dtR} represent the weights of criteria residual energy, distance to relay and the node centrality.

The performance score matrix $[PS]_i$ is generated as shown below:

$$[PS]_i = \begin{bmatrix} W_{RE} \times A_{(1,1)} - W_{dtR} \times A_{(1,2)} - W_{NC} \times A_{(1,3)} \\ W_{RE} \times A_{(2,1)} - W_{dtR} \times A_{(2,2)} - W_{NC} \times A_{(2,3)} \\ \vdots \\ W_{RE} \times A_{(u,1)} - W_{dtR} \times A_{(u,2)} - W_{NC} \times A_{(u,3)} \end{bmatrix}$$

- **Step-VI:** The performance scores generated in Step-V can be either negative or positive, depending on the number of beneficial and non-beneficial criteria in the system. Then $[PS]_i$ matrix is sorted in increasing order based on the performance scores of every alternative. The alternative with the highest performance score is designated the best alternative, and the one with the lowest performance score is designated the worst alternative. The best alternative is picked, and its corresponding node is marked as CH for that particular cluster in that round.

4.3 Routing paths establishment phase

In Sect. 4.1, clusters are formed as per the HLBC clustering architecture, and Sect. 4.2 deals with CH selection in each cluster. In this phase, routing paths get established among the sensor nodes. Direct or multi-hop communication can be employed to disseminate the sensed data from the sensor nodes to the base station. In most cases, direct communication is used for intra-cluster data communication, and multi-hop communication is used for inter-cluster communication. In the HLBC scheme, the member nodes send the sensed data directly to the CH node. The CH nodes employ multi-hop communication for inter-cluster data communication. The CH nodes communicate the aggregated data to the assigned relay node in the subsequent layer toward the base station. The nodes near the base station (i.e., layer 1) directly communicate with the base station. The shortcomings in the HLBC architecture are that direct communication is employed at the intra-cluster level. At the intra-cluster level, multi-hop data communication can be used instead of direct communication. This aids in further load balancing the network and improving the network lifetime. So, in this proposed work, the routing paths among the member nodes in a cluster are established by forming a minimum spanning tree (MST) with the help of the proposed modified Dijkstra based MST formation technique.

In literature, most of the works employing multi-hop communication for data forwarding consider distance as the only influential parameter while forming routing paths to optimize energy consumption. But it is critical to assess the load and energy level of the relay nodes along with the distance parameter to achieve balanced energy consumption among all nodes in the network. The load on a relay node specifies the number of relays the node will be performing in the already constructed network. In this work, while constructing an MST in a cluster, all the influential parameters- energy level, distance, and load on the relay node are taken into account. After the proposed MST technique is implemented in all the clusters, intra-cluster routing paths are formed among all non-CH nodes. The same routing paths established among the CH nodes as per the HLBC clustering architecture will be utilized for inter-cluster data communication. Algorithm 3 shows the proposed modified dijkstra based MST formation technique.

4.3.1 Proposed modified dijkstra based MST formation technique

- **Step-1:** Initially, complete and incomplete sets are created. The CH node will join the complete set, and the member nodes having positive residual energy will join the incomplete set. The *connect-value* for nodes in the complete set will be 1, and for nodes in the incomplete set will be 0.
- **Step-2:** Update complete and incomplete sets on the basis of *connect-value*. Every node from the incomplete set will calculate its performance score relative to all nodes from the complete set.

  Let us assume that node i is from the complete set and node j is from the incomplete set. The performance score of node i relative to node j (i.e., PF_{ij}) is calculated as follows:

$$(PF_{ij}) = W_E \times (E_j) + W_d \times \frac{1}{d_{(i,j)}} + W_L \times (L_j) \quad (10)$$

  where, $\{W_E, W_d, W_L\}$ represents the weightages corresponding to criteria remaining energy, distance and load respectively. After conducting several simulations the best performing weights for the crieria are found out to be $W_E = 0.25$, $W_d = 0.50$, and $W_L = 0.25$ Here, E_j is the residual energy of node j, $d_{(i,j)}$ is the euclidean distance from node i to node j, and L_j represents the load on node j. The load on node j (i.e., L_j) is computed as shown in (9).

$$L_j = 1 - \frac{|(Connected - nodes)_j|}{Size(cluster) - 1} \tag{11}$$

where, $(Connected - nodes)_j$ represents the number of nodes from the complete set that have established routing paths to node j and $Size(cluster)$ denotes the size of that particular cluster in which MST formation in being done.

- **Step-3:** Sort the values of PF_{ij} in the increasing order. The nodes corresponding to the highest performance score are selected, and a routing path is established between them. The *connect-value* of node i gets updated to 1 and joins the complete set and gets removed from the incomplete set.
- **Step-4:** Steps (2–4) are repeated until the incomplete set becomes empty.

After establishing intra-cluster and inter-cluster routing paths, synchronized TDMA schedules are constituted by all CHs and sent to their corresponding member nodes in the same way as done in the HLBC clustering architecture.

Algorithm 3 : Proposed Modified Dijkstra-based MST formation technique

Input: MN's of a cluster X, CH of cluster X.
Important Variables: $(Connect - value)$, complete set[], incomplete set[].
Output: Intra-cluster communication paths established in a cluster X.

```
Set (Connect − value) for CH node as 1 and for MNs as 0. ▷ helps in determining complete and in-complete sets
Initialize complete set and incomplete set based on (Connect − value) of nodes.
while Count(incomplete set) ≠ NULL do
    Initialize the value of K_2 with the count of nodes in the complete set.
    Select a node i from incomplete set.
    for j=1; j≤K_2; j++ do
        Calculate E_j and d_ij corresponding to node j from complete set.
        Compute L_j using Eqn.(11) corresponding to node j from complete set.
        Compute performance score PF_ij using Eqn.(10).
    end for
    Sort PF_ij in ascending order and select the node j from complete set having highest PF_ij value.
    Establish an Intra-cluster communication path between node i from the incomplete set and node j from the complete set.
    Update the (Connect − value) of node i to 1 and reinitialize complete and incomplete sets.
end while
```

4.4 Data aggregation and forwarding phase

In this phase, every node senses the network region and waits for its TDMA time slot for data forwarding. Every sensor node communicates its sensed data to its corresponding CH or relay via established routing paths. The CH nodes aggregate the received information and forward it to the next-level relay or the base station. The relay nodes also aggregate their relay data and the sensed data before forwarding it via the routing paths.

4.5 Algorithm analysis

The computational complexity of the proposed HCM protocol (i.e., Algorithm 1) combines the specific implementation of the deployment and cluster formation phases, the CH selection phase, and the routing path establishment phases. Deployment and cluster formation phases occur only once in the network based on the HLBC clustering architecture. In the deployment phase, the sensor nodes are uniformly deployed over the network area and have a worst-case time complexity of $O(N)$, where N is the number of sensor nodes in the network. In the cluster formation phase, the network is organized into layers and then into clusters according to HLBC architecture (i.e., shown in Fig. 2). This phase occurs in constant time only once and involves a time complexity of $O(m)$, where m is constant. The CH selection phase (i.e., shown in Algorithm 2) happens according to the MOORA-based technique, where every cluster runs Algorithm 2 individually to determine its CH. The weights determination for the considered criteria (i.e., C) is done using the Shannon entropy technique. The worst-case time complexity for this phase involves $O(K \times \frac{N}{K} \times C)$, where K is the number of clusters and C is the number of criteria used for weight estimation. For the routing path establishment phase (i.e., as shown in Algorithm 3), the modified Dijkstra-based MST formation technique is employed, where nodes in clusters get partitioned into complete and incomplete sets, with K CH nodes are present in the complete set and $(N - K)$ nodes are present in the incomplete set. The worst-case time complexity involved here is $O(N^2)$. The overall computational complexity in the worst case is the sum of all the phases: $O(N^2)$. The communication overhead involved for deployment and the cluster formation phase is $O(2N)$, where the sensor nodes send a control packet to the sink that includes their location information and receive another control packet with their corresponding cluster information. The communication overhead involved in the CH selection phase is $O(\frac{N}{K} \times K)$, where MNs receive CH information. In the routing path establishment phase, $O(N \times 1)$ communication overhead is involved where a control packet is transferred by the CH node, including the parent node information. Some potential drawbacks of the presented HCM protocol algorithms, compared to other protocols, can be: (i) In the CH selection algorithm, in every round, the alternatives and the weightages of the criteria are altered based on the input measures. (ii) Extra

Table 2 Simulation parameters considered in this work

Simulation areas	(120 m × 120 m) and (240 m × 240 m)
Location of sink node	(60 m, 200 m) and (120 m, 320 m)
Sensor node types	Homogeneous and heterogeneous
Initial energy (homogeneous case)	0.5 J
Initial energy (heterogeneous case)	0.5 and 2 J
Energy advancement factor (a)	3
Efs	10 pj/bit/m^2
Eelec	50 nj/bit
Emp	0.0013 nj/bit/message
Size of data packet	4000 bits
Size of control packet	200 bits
ETX	50×10^{-9} J/bit
ERX	50×10^{-9} J/bit
EDA	5×10^{-9} J/bit/message

Table 3 Network lifetime comparison of EMUC, HLBC and proposed HCM in homogeneous network

	EMUC	HLBC	HCM
Scenario-1 120 m × 120 m			
FND	201	341	417
HND	346	595	809
LND	412	1121	1674
Scenario-2 240 m × 240 m			
FND	254	309	382
HND	317	567	731
LND	340	960	1498

computations are to be performed while determining the CH node and the routes for data transmission. (iii) In very few cases, the intra-cluster communication paths set up among clustered nodes might not be minimal in terms of distance but happen in a load-balanced way with respect to other methods.

5 Performance evaluation

The proposed scheme's performance is evaluated in both homogeneous and heterogeneous scenarios. Its performance in homogeneous cases is compared with the EMUC (Assari et al. 2020) and HLBC (Prasad et al. 2021) protocols, and in heterogeneous cases is compared with the exisisting DEEC [19], SEP (Smaragdakis et al. 2004) and het-HLBC (Prasad et al. 2021) protocols. To do this, MATLAB 2021 software is used to simulate all the schemes mentioned above and can be found out at Prasad (2023). For simulation purposes, we have considered two different network scenarios, which helps us give a clear statement about the scalability and practical performance evaluation of the proposed scheme. The following two network scenarios are considered for performing simulations:

Scenario 1: A 120 × 120 m network is considered in which 120 nodes are distributed uniformly. The sink is placed exterior at location (60 m, 200 m). As per the HLBC architecture, the network is partitioned into three equal-sized layers (i.e., $Layer_1$, $Layer_2$ and $Layer_3$) of 40 nodes each per layer. The layer set up close to the sink is termed as $Layer_1$ having two clusters of sizes 20 nodes each. The layer set up distant from the sink is termed as $Layer_3$ having eight clusters of sizes five nodes each. The remaining is the intermediate layer (i.e., $Layer_2$) which constitutes four clusters with 10 nodes per cluster.

Scenario 2: A 240 × 240 m network is considered in which 240 nodes are distributed uniformly. The sink is placed exterior at location (120 m, 320 m). As per the HLBC architecture, the network is partitioned into three equal-sized layers (i.e., $Layer_1$, $Layer_2$ and $Layer_3$) of 80 nodes each per layer. The layer set up close to the sink is termed as $Layer_1$ having two clusters of sizes 40 nodes each. The layer set up distant from the sink is termed as $Layer_3$ having 8 clusters of sizes 10 nodes each. The remaining is the intermediate layer (i.e., $Layer_2$) which constitutes four clusters with 20 nodes per cluster.

Table 2 shows the simulation parameters considered in this work. The deployment in heterogeneous cases is done in such a way that each cluster will have at-least one heterogeneous node. All of the results depicted are an average of 50 simulation runs. The proposed Scheme's performance is evaluated using the following metrics.

1. **Network lifetime** represents the working period of a network. The rounds operated until the first node, 50% nodes and 99% of the total nodes have exhausted their energies in the network are represented as first node death (FND), half node death (HND) and last node death (LND) respectively.
2. **Node death rate** represents the total nodes that have drained out their energy across the rounds operated.
3. **Energy consumption** is the overall amount of energy consumed in the network throughout the course of the rounds.

Table 3 represents the network lifetime of all the considered approaches in homogeneous case. The first node death (FND) of EMUC, HLBC and the proposed HCM approach are at rounds 201, 341 and 417 for scenario-1 and 254, 309 and 382 for scenario-2, respectively. The half node death (HND) for the presented approaches are at rounds 346, 595 and 809 for scenario-1 and 317, 567 and 731 for scenario-2, respectively. The last node death (LND) are at rounds 412, 1121 and 1674 for scenario-1 and 340, 960 and 1498 rounds respectively for scenario-2.

Table 4 shows the network lifetime of all the simulated approaches in heterogeneous case. The FND of SEP, DEEC, het-HLBC and the proposed approach are at rounds 170, 196, 494 and 609 for scenario-1 and 32, 47, 405 and 513 for scenario-2, respectively. The HND for the presented approaches are at rounds 283, 397, 766 and 921 for scenario-1 and 181, 212, 696 and 765 for scenario-2, respectively. The LND are at rounds 1289, 1761, 2085 and 3245 for scenario-1 and 1154, 1209, 2042 and 2857 rounds respectively for scenario-2. It is evident that the proposed approach has performed better compared to existing approaches in both scenarios 1 and 2. This is because of the proficient clustering architecture and the CH selection technique employed in the proposed scheme. Our proposed HCM scheme successfully incorporates the benefits of the HLBC clustering architecture, which achieves balanced energy consumption by avoiding the hot-spot problem and enabling energy-efficient operation. Moreover, the proposed MOORA-based CH selection technique employed at each cluster by considering all vital parameters helps in selecting the best-suited nodes for the role of CH, and the MST formed in each cluster based on the proposed modified Dijkstra algorithm further load balances the network and makes it operate in energy efficient manner.

Table 4 Network lifetime comparison of SEP, DEEC Het-HLBC and proposed Het-HCM in heterogeneous network

	SEP	DEEC	Het-HLBC	Het-HCM
Scenario-1 120 m × 120 m				
FND	170	196	494	609
HND	283	397	766	921
LND	1289	1761	2085	3245
Scenario-2 240 m × 240 m				
FND	32	47	405	513
HND	181	212	696	765
LND	1154	1209	2042	2857

Figures 3 and 4 represent the node death rate in scenarios 1 and 2, respectively, for homogeneous cases. Figure 5 and 6 illustrate the node death rate in scenarios 1 and 2, respectively, for heterogeneous cases. It is evident that the proposed approach has attained a better stable region in comparison with all the approaches presented for both homogeneous and heterogeneous cases. The stable region represents the number of rounds operated in the network until the first node dies. The proposed approach has attained a better stable region and has performed better in terms of node death rate because of the balanced energy consumption happening in the network. The residual energy, load on the relay node, and distance parameter considered while constructing MST in all clusters help attain better-balanced energy consumption among all nodes within a cluster. The proposed technique also helps achieve lower intra-cluster and inter-cluster communication distances, enabling the network to operate energy-efficiently.

Figures 7 and 8 represent the energy consumption rate in scenarios 1 and 2, respectively, for homogeneous cases. Figures 9 and 10 illustrate the energy rate in scenarios 1 and 2, respectively, for heterogeneous cases. It is evident that the proposed approach has obtained better energy consumption when compared to all other approaches presented for both homogeneous and heterogeneous cases. This is because the proposed approach has achieved lower intra-cluster and inter-cluster communication distances compared to other presented approaches. The lower intra-cluster communication distances are achieved in the network because of the modified Dijkstra-based minimum spanning tree constructed in each cluster that considers all other vital parameters, like residual energy and load on the relay node, in addition to the distance parameter. Moreover, multi-hop communications employed among the CH nodes to next-level relays help attain further energy efficiency.

Table 5 shows the percentage improvement in terms of network lifetime between HLBC and the proposed HCM scheme. In scenario 1, for the homogeneous case, the proposed HCM approach has achieved a percentage improvement of 22.28%, 35.96% and 49.93% in FND, HND and LND, respectively. For the heterogeneous case, the proposed approach has achieved a percentage improvement of 23.27%, 20.23% and 55.65% in FND, HND and LND, respectively. In scenario 2, for the

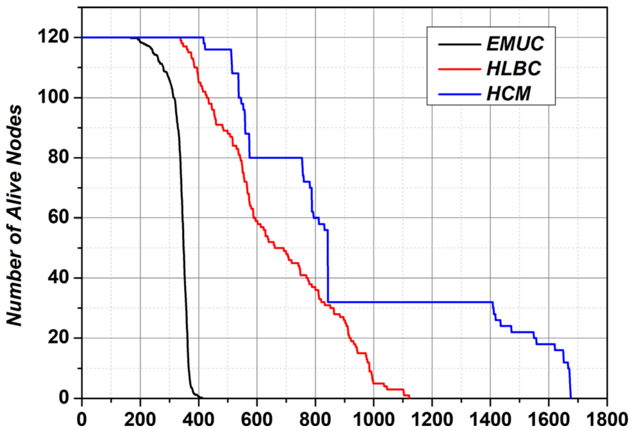

Fig. 3 Node death rate in scenario-1 (homogeneous case)

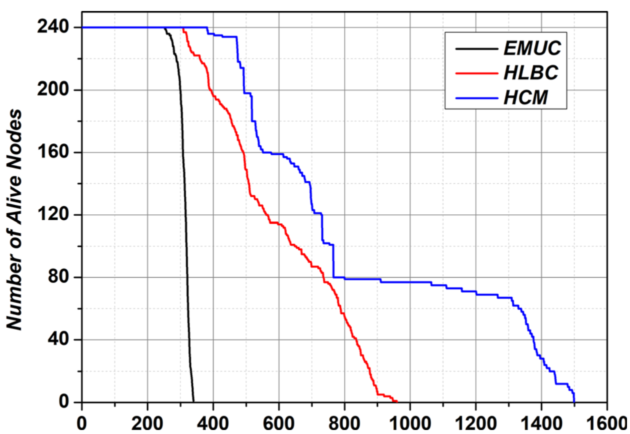

Fig. 4 Node death rate in scenario-2 (homogeneous case)

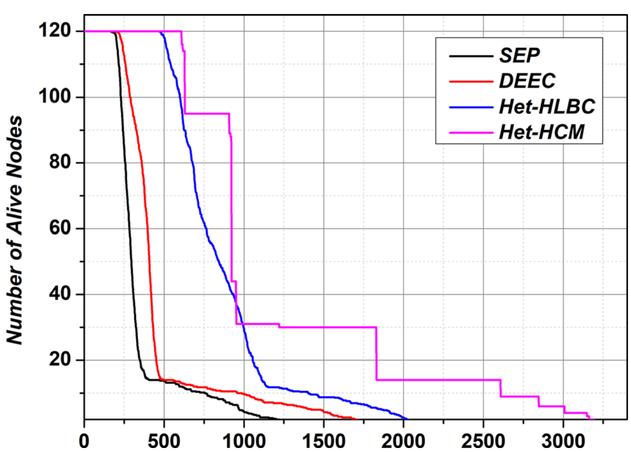

Fig. 5 Node death rate in scenario-1 (heterogeneous case)

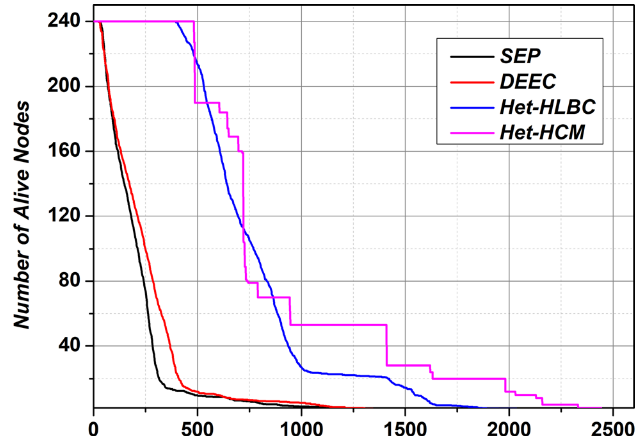

Fig. 6 Node death rate in scenario-2 (heterogeneous case)

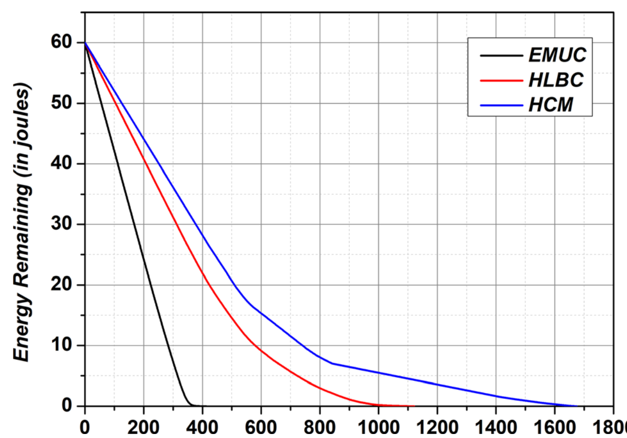

Fig. 7 Energy consumption rate in scenario-1 (homogeneous case)

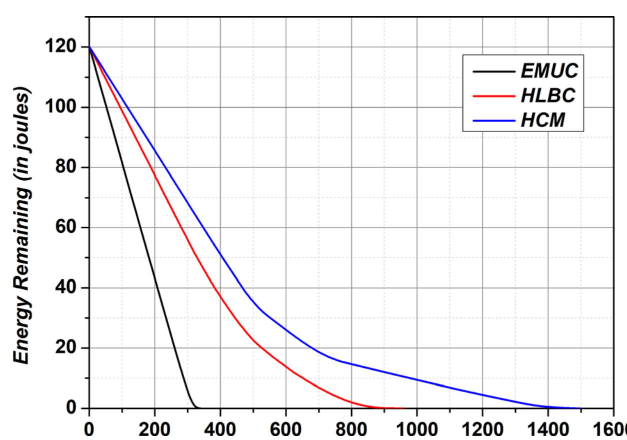

Fig. 8 Energy consumption rate in scenario-2 (homogeneous case)

homogeneous case, the proposed approach has achieved a percentage improvement of 23.62%, 28.92% and 56.04% in FND, HND and LND, respectively. For the heterogeneous case, the proposed approach has achie-ved a percentage improvement of 26.66%, 9.91% and 39.91% in FND, HND and LND, respectively. Although our proposed approach incorporates the basic clustering architecture from the HLBC scheme, it has been proven to operate better because of the proposed CH selection technique and modified Dijkstra-based minimum spanning tree for intra-cluster data communications.

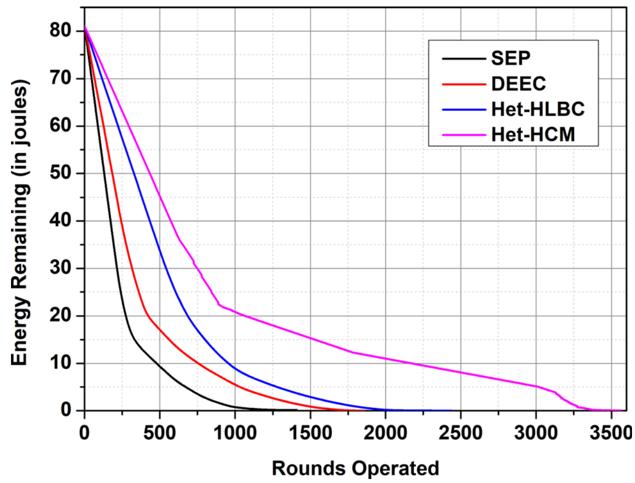

Fig. 9 Energy consumption rate in scenario-1 (heterogeneous case)

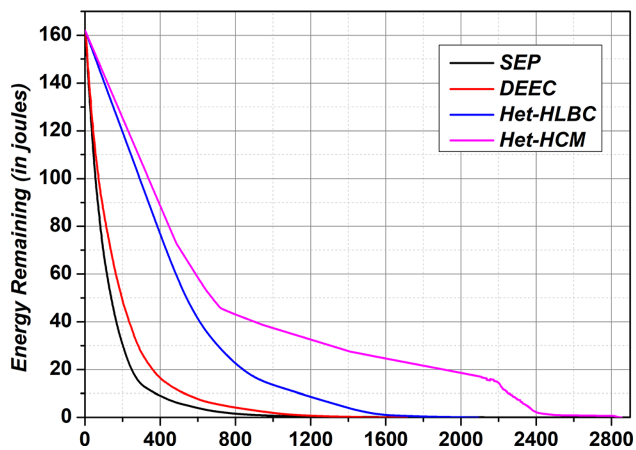

Fig. 10 Energy consumption rate in scenario-2 (heterogeneous case)

Table 6 shows the percentage of energy consumed in the network in scenarios 1 and 2 at the time of the FND, HND and LND in all the simulated protocols for both homogeneous and heterogeneous cases. For homogeneous cases, in scenario 1, the FND in EMUC, HLBC and the proposed HCM schemes happen when 59.89%, 54.53%, and 55.22% of total network energies are consumed, respectively. For heterogeneous cases, in scenario 1, the FND in SEP, DEEC, Het-HLBC and the proposed Het-HCM schemes happen when 49.90%, 40.76%, 60.13%, and 53.79% of total network energies are consumed, respectively. The HND for homogeneous cases in scenario 1 occurs when 98.09%, 84.42%, and 86.96% of total network energies are consumed for EMUC, HLBC and the proposed HCM schemes, respectively. The HND for heterogeneous cases in scenario 1 occurs when 76.41%, 73.08%, 79.82% and 73.0% of total network energies are consumed for SEP, DEEC, Het-HLBC and the proposed Het-HCM schemes, respectively. It can be observed that the proposed approach has operated for a greater number of rounds in an energy-efficient manner.

5.1 Discussion

The proposed MOORA-based CH selection technique employed at each cluster by considering all vital parameters (residual energy, node centrality, and distance to relay) helps in selecting the best-suited nodes for the role of CH. The weightages of all the vital parameters are estimated using the Shannon entropy technique and calculated for every round based on the input measures. The MST constructed in each cluster based on the proposed modified Dijkstra algorithm by considering energy level, distance, and load on the relay node helps further load balance the network and makes it operate energy-efficiently. The presented protocol is analyzed based on network lifetime, energy consumption, and node death rates. All the presented results revealed the effectiveness, load-balancing, and energy-efficient operation capabilities of the proposed HCM protocol. The considered scenarios (i.e., scenario 1, scenario 2) and cases (i.e., homogeneous, heterogeneous) proved the scalability and adaptability of the presented protocol. It is evident from the results that the proposed approach has attained a better stable region, network lifetime and energy consumption in comparison with all compared approaches presented for both homogeneous and heterogeneous cases. The proposed HCM protocol can be further analyzed in the aspects of (i) its performance if the location of the sink node is kept inside or at edges near the network area and (ii) by adding some more influential criteria while decision-making.

6 Conclusion

This paper presented a hierarchical clustering framework with a combined MOORA technique, Shannon entropy measures for CH selection, and a modified Dijkstra-based MST formation technique. The performance of the proposed HCM scheme in terms of network lifetime, node death rate, and energy consumption has been found to be better than EMUC and HLBC protocols for homogeneous cases and in comparison to SEP, DEEC, and Het-HLBC protocols for heterogeneous scenarios. Overall, in terms of performance, HCM is followed by HLBC. The percentage increase in the FND, HND, and LND of the HCM scheme in comparison to HLBC varies from [22–24]%,

Table 5 Percentage improvement w.r.t network lifetime for HLBC and the proposed HCM approach

	Homogeneous cases			Heterogeneous cases		
	HLBC	HCM	% Increase	Het-HLBC	Het-HCM	% Increase
Scenario-1 120 m × 120 m						
FND	341	417	22.28%	494	609	23.27%
HND	595	809	35.96%	765	921	20.23%
LND	1121	1674	49.33%	2085	3245	55.65%
Scenario-2 240 m × 240 m						
FND	309	382	23.62%	405	513	26.66%
HND	567	731	28.92%	696	765	9.91%
LND	960	1498	56.04%	2042	2857	39.91%

Table 6 Percentage energy consumption of various simulated approaches in homogeneous and heterogeneous cases

	Homogeneous cases			Heterogeneous cases			
	EMUC	HLBC	HCM	SEP	DEEC	Het-HLBC	Het-HCM
Scenario-1 120 m × 120 m							
FND	59.89%	54.53%	55.22%	49.90%	40.76%	60.13%	53.79%
HND	98.09%	84.42%	86.96%	76.41%	73.08%	79.82%	73.01%
LND	99.99%	99.98%	99.99%	99.81%	99.99%	99.98%	99.99%
Scenario-2 240 m × 240 m							
FND	81.25%	54.88%	54.79%	22.58%	27.19%	53.22%	57.11%
HND	98.65%	86.25%	85.82%	78.06%	71.77%	80.86%	72.68%
LND	99.99%	99.99%	99.99%	99.76%	99.56%	99.98%	99.99%

[28–36]%, and [49–57]%, respectively, for homogeneous network scenarios. Whereas the FND, HND, and LND of the HCM scheme vary from [23–27]%, [9–21]%, and [39–56]%, respectively, in comparison to the HLBC for heterogeneous network scenarios.

References

Alia OM (2018) A dynamic harmony search-based fuzzy clustering protocol for energy-efficient wireless sensor networks. Ann Telecommun 73(5):353–365

Amini M, Karimi A (2020) Two-level distributed clustering routing algorithm based on unequal clusters for large-scale internet of things networks. J Supercomput. https://doi.org/10.1007/s11227-019-03067-2

Assari Y, Al Fallah S, Elasari J, Arioua M, El Oualkadi A (2020) Energy-efficient multi-hop routing with unequal clustering approach for wireless sensor networks. Int J Comp Netw Commun 12:57–75. https://doi.org/10.5121/ijcnc.2020.12304

Baranidharan B, Santhi B (2016) DUCF: distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach. Appl Soft Comput 40:495–506. https://doi.org/10.1016/j.asoc.2015.11.044

Benelhouri A, Idrissi-Saba H, Antari J (2022) An improved gateway-based energy-aware multi-hop routing protocol for enhancing lifetime and throughput in heterogeneous wsns. Simul Model Pract Theory 116:102471

Bhattacharjya K, Alam S, De D (2021) TTCBT: two tier complete binary tree based wireless sensor network for FSR and LANMAR routing protocols. Microsyst Technol 27:1–11. https://doi.org/10.1007/s00542-018-3982-3

Chakraborty S (2011) Applications of the moora method for decision making in manufacturing environment. Int J Adv Manuf Technol 54:1155–1166. https://doi.org/10.1007/s00170-010-2972-0

Dutt S, Agrawal S, Vig R (2018) Cluster-head restricted energy efficient protocol (creep) for routing in heterogeneous wireless sensor networks. Wireless Pers Commun 100(4):1477–1497

El Alami H, Najid A (2019) ECH: an enhanced clustering hierarchy approach to maximize lifetime of wireless sensor networks. IEEE Access 7:107142–107153

Elkamel R, Messouadi A, Cherif A (2019) Extending the lifetime of wireless sensor networks through mitigating the hot spot problem. J Parallel Distrib Comput 133:159–169. https://doi.org/10.1016/j.jpdc.2019.06.007

Elshrkawey M, Elsherif SM, Elsayed Wahed M (2018) An enhancement approach for reducing the energy consumption in wireless sensor networks. J King Saud Univ Comput Inf Sci 30(2):259–267. https://doi.org/10.1016/j.jksuci.2017.04.002

Esmaeili H, Hakami V, Bidgoli BM, Shokouhifar M (2022) Application-specific clustering in wireless sensor networks using combined fuzzy firefly algorithm and random forest. Expert Syst Appl 210:118365

Fanian F, Rafsanjani MK, Saeid AB (2021) Fuzzy multi-hop clustering protocol: selection fuzzy input parameters and rule tuning for WSNS. Appl Soft Comput 99:106923

Gangwar S, Prasad IB, Yadav SS, Pal V, Kumar Patra S (2022) HSCR: hierarchical structured cluster routing protocol for load balanced wireless sensor networks. Software: Practice and Experience

Gangwar S, Prasad IB, Pal V, Yogita, Yadav SS (2023) Fuzzy clustering based cluster head selection for iot enabled wsns. In:

2023 10th International conference on computing for sustainable global development (INDIACom), pp 761–766
Hamzah A, Shurman M, Al-Jarrah O, Taqieddin E (2019) Energy-efficient fuzzy-logic-based clustering technique for hierarchical routing protocols in wireless sensor networks. Sensors 19(3):561
Heinzelman WR, 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
Hoang DC, Yadav P, Kumar R, Panda SK (2013) Real-time implementation of a harmony search algorithm-based clustering protocol for energy-efficient wireless sensor networks. IEEE Trans Ind Inf 10(1):774–783
Hung L-L, Leu F-Y, Tsai K-L, Ko C-Y (2020) Energy-efficient cooperative routing scheme for heterogeneous wireless sensor networks. IEEE Access 8:56321–56332
Hwang C-L, Yoon K (1981) Methods for multiple attribute decision making. Springer, Berlin, Heidelberg, pp 58–191
Jain D, Shukla P, Varma S (2022) Energy efficient architecture for mitigating the hot-spot problem in wireless sensor networks. J Ambient Intell Human Comput. https://doi.org/10.1007/s12652-022-03711-5
Le-Ngoc KK, Tho QT, Bui TH, Rahmani AM, Hosseinzadeh M (2022) Optimized fuzzy clustering in wireless sensor networks using improved squirrel search algorithm. Fuzzy Sets Syst 438:121–147
Mittal N, Singh U, Sohi BS (2019) An energy-aware cluster-based stable protocol for wireless sensor networks. Neural Comput Appl 31(11):7269–7286
Mondal S, Khatua S, Biswas U, Das R (2022) Energy efficient algorithms for enhancing lifetime in wireless sensor networks. Microsyst Technol. https://doi.org/10.1007/s00542-022-05256-1
Pal V, Singh G, Yadav RP (2015) Balanced cluster size solution to extend lifetime of wireless sensor networks. IEEE Internet Things J 2(5):399–401. https://doi.org/10.1109/JIOT.2015.2408115
Pal V, Singh G, Yadav R et al (2015) Effect of heterogeneous nodes location on the performance of clustering algorithms for wireless sensor networks. Proc Comp Sci 57:1042–1048
Pathak A (2020) A proficient bee colony-clustering protocol to prolong lifetime of wireless sensor networks. J Comput Netw Commun
Prasad IB, Pal V (2023) Matlab code for BCM protocol. https://github.com/bhanu3214/main-code. Accessed 15 Nov 2022
Prasad IB, Rout B, Pal V, Yogita (2020) M-way balanced tree data aggregation approach for clustered wireless sensor networks. In: Janyani V, Singh G, Tiwari M, Ismail T (eds) Opt Wireless Technol. Springer Singapore, Singapore, pp 607–613
Prasad IB, Yogita SS, Yadav V. Pal (2021) HLBC: a hierarchical layer-balanced clustering scheme for energy efficient wireless sensor networks. IEEE Sens J 21(22):26149–26160. https://doi.org/10.1109/JSEN.2021.3114506
Prasad IB, Yogita, Yadav SS, Kumar N, Pal V (2023) Bat: Barabasi-albert with topsis scale free topology evolution for load balanced WSNS. IEEE Sens J. https://doi.org/10.1109/JSEN.2023.3277461
Priyadarshi R, Soni SK, Nath V (2018) Energy efficient cluster head formation in wireless sensor network. Microsyst Technol 24(12):4775–4784. https://doi.org/10.1007/s00542-018-3873-7
Priyadarshi R, Rawat P, Nath V, Acharya B, Shylashree N (2020) Three level heterogeneous clustering protocol for wireless sensor network. Microsyst Technol 26(12):3855–3864. https://doi.org/10.1007/s00542-020-04874-x
Qing L, Zhu Q, Wang M (2006) Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. Comput Commun 29(12):2230–2237. https://doi.org/10.1016/j.comcom.2006.02.017
Sharma R, Vashisht V, Singh U (2019) EEFCM-DE: energy-efficient clustering based on fuzzy c means and differential evolution algorithm in WSNS. IET Commun 13(8):996–1007
Shyama M, Pillai AS, Anpalagan A (2022) Self-healing and optimal fault tolerant routing in wireless sensor networks using genetical swarm optimization. Comput Netw 217:109359
Singh J, Yadav SS, Kanungo V, Yogita, Pal V (2021) A node overhaul scheme for energy efficient clustering in wireless sensor networks. IEEE Sens Lett 5(4):1–4. https://doi.org/10.1109/LSENS.2021.3068184
Smaragdakis G, Matta I, Bestavros A et al. (2004) Sep: a stable election protocol for clustered heterogeneous wireless sensor networks. In: Second international workshop on sensor and actor network protocols and applications (SANPA 2004), vol 3, Boston, MA
Soni V, Mallick D (2017) FTGAF-HEX: fuzzy logic based two-level geographic routing protocol in wireless sensor networks. Microsyst Technol. https://doi.org/10.1007/s00542-017-3272-5
Thiagarajan R et al (2020) Energy consumption and network connectivity based on novel-leach-POS protocol networks. Comput Commun 149:90–98
Vahabi S, Mojab SP, Eslaminejad M, Dashti SE (2022) EAM: energy aware method for chain-based routing in wireless sensor network. J Ambient Intell Human Comput 13(9):4265–4277
Wang Z, Ding H, Li B, Bao L, Yang Z, Liu Q (2022) Energy efficient cluster based routing protocol for WSP using firefly algorithm and ant colony optimization. Wirel Pers Commun Int J 125(3):2167–2200
Yogita Y, Pal V, Yadav A (2022) DDC protocol to protract network lifetime of wireless sensor networks. Int J Comput Appl 44(4):316–323

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.