Intelligent and Converged Networks
2023, 4(3): 206−224

ISSN 2708-6240
DOI: 10.23919/ICN.2023.0018

Relay-assisted wireless energy harvesting for multihop clustered IoT network

B. Pavani*, K. Venkata Subbareddy, and L. Nirmala Devi

Abstract: In large-scale networks such as the Internet of Things (IoT), devices seek multihop communication for long-distance communications, which considerably impacts their power exhaustion. Hence, this study proposes an energy harvesting-enabled, relay-based communication in multihop clustered IoT networks in a bid to conserve the battery power in multihop IoT networks. Initially, this study proposes an efficient, hierarchical clustering mechanism in which entire IoT devices are clustered into two types: the closest cluster (CC) and remote clusters (RCs). Additionally, Euclidean distance is employed for the CC and fuzzy c-means for the RCs. Next, for cluster head (CH) selection, this study models a fitness function based on two metrics, namely residual energy and distance (device-to-device distance and device-to-sink distance). After CH selection, the entire clustered network is partitioned into several layers, after which a relay selection mechanism is applied. For every CH of the upper layer, we assign a few lower-layer CHs to function as relays. The relay selection mechanism is applied only for the devices in the RCs, while for devices in the CC, the CH functions as a relay. Finally, several simulation experiments are conducted to validate the proposed method's performance. The results show the method's superiority in terms of energy efficiency and optimal number of relays in comparison with the state-of-the-art methods.

Key words: Internet of Things (IoT); energy management; clustering; cluster head selection; relay selection

1 Introduction

The Internet of Things (IoT) is reducing the gap between the cyber and physical worlds, and has become one of the emerging network standards. It has many distinctive applications, such as industrial monitoring and control, smart healthcare, military surveillance, smart transportation, smart cities, and smart agriculture[1, 2].

These applications continuously sense, compute, and forward acquired data to a server or sink node. Heterogeneous sensor devices are run on inadequate energy resources and are mostly deployed in adverse environments. Therefore, energy management becomes one of the major challenges in IoT. Recently, industry-wide research was conducted in the field of IoT to establish energy-efficient, reliable communication across devices with a view to improving the network lifetime and achieving enhanced resource allocation[3]. The literature shows that using sensor nodes (SNs) in IoT networks in an energy-efficient way is still an unexplored field that needs to be investigated[4, 5].

Many researchers suggested that clustering is one of the prominent solutions to reduce the energy consumption of the network[6–9]. Clustering reduces the interchange of redundant messages among the SNs by conserving the communication bandwidth[6]. The idea of data accumulation at cluster heads (CHs) assists SNs in saving energy. A hierarchical strategy called clustering increases the scalability of IoT networks[7]. In this strategy, the CH accumulates the data from SNs and sends them to the sink node. However, CHs near the sink node may experience considerable energy

• B. Pavani, K. Venkata Subbareddy, and L. Nirmala Devi are with the Department of Electronics and Communication Engineering, University College of Engineering, Osmania University, Hyderabad 500007, India. E-mail: beti.pavani@gmail.com; subvishk03@gmail.com; nirmaladevi@osmania.ac.in.
∗ To whom correspondence should be addressed.
Manuscript received: 2023-01-23; revised: 2023-05-12; accepted: 2023-05-30

© All articles included in the journal are copyrighted to the ITU and TUP. This work is available under the CC BY-NC-ND 3.0 IGO license: https://creativecommons.org/licenses/by-nc-nd/3.0/igo/.

consumption and rapid energy depletion, which can lead to hotspots in the network[8, 9]. One of the potential strategies to prevent the hot spot problem is uneven clustering, which results in varying cluster sizes with respect to the distance between the sink node and CHs; i.e., as distance increases, the cluster size also increases. Hence, we must concentrate on the reduction in intracluster energy consumption for the clusters nearer to the sink node to maintain adequate energy between the clusters.

Clustering is a successful approach to effectively manage energy. However, owing to weak connectivity between the nodes, when the SNs are moving away from the sink node, they are unable to transmit the sensed data to the sink node. This results in poor network performance in terms of network lifetime and energy consumption in multihop IoT networks. Accordingly, few researchers suggested that relaying is the best solution to improve the network lifetime. A benefit of a relay network is that it enables range extension while using less energy. Relay nodes farther from the sink node typically use additional energy for data transfer. Consequently, selecting the right relay is essential for sensor node energy saving because it considerably influences both the network and sensor lifetimes[10]. Another difficulty with any wireless network, such as IoT networks, is determining the number of ideal relay nodes. While selecting more nodes as relays may boost diversity, it also results in more energy consumption, which interferes with SN synchronization[11]. Additionally, most methods of selecting relay nodes suppose that the relay nodes are randomly distributed throughout the network[12]. When relay nodes are deployed in this manner, the resulting heterogeneous IoT network is fault tolerant. Although a lot of work has been put into using the relay selection technique to enhance the reliability of IoT networks, it still remains unclear how to prevent these scenarios.

This study proposes a relay-based hybrid and adaptive clustering mechanism to fill this research gap. In this mechanism, initially, the network is partitioned into a few layers, and the layer nearer to the sink node is considered the lowest layer where the closest cluster (CC) is located. Next, the layers other than the CC layer are considered upper layers, where the remote clusters (RCs) are located. Furthermore, the CH selection mechanism and relay selection mechanism are introduced to efficiently balance the entire network's energy. Therefore, the following are the major contributions of this paper:

(1) This study proposes a hybrid and adaptive clustering mechanism to efficiently use each node's energy. In this clustering, all SNs are grouped into two clusters, namely the CC and RCs.

(2) In a clustered IoT network, all the CHs consume more energy than the cluster members (CMs) owing to their continuous communication. This study proposes a new CH selection mechanism to reduce the energy consumption by CH. This mechanism employs one fitness function, which consists of the metrics, namely distance and residual energy.

(3) This study proposes a hierarchical structuring model and an energy-efficient, optimal relay selection mechanism to provide effective connectivity between the nodes for long-distance communication.

(4) This study proposes an effective hierarchical packet routing (HPR) mechanism to reduce the overall network energy consumption.

The rest of this paper is organized as follows. Section 2 describes the related past work. Section 3 explains the proposed work. Section 4 explores the simulation results and their analysis, and finally, Section 5 concludes this paper.

2 Related work

In this section, detailed literature related to relay selection mechanisms for clustered IoT networks is discussed. To maintain connectivity from each SN to the base station (BS), the literature contains numerous relay node placement techniques for wireless sensor networks (WSNs) or IoT networks. The main difficulty areas in employing the relays in the deployment area are low connectivity, scalability, and prolonged lifetime[13–15]. Shukla and Tripathi[16] proposed a scalable and energy-efficient routing protocol (SEEP) to prolong the network lifetime of WSNs in IoT. SEEP was employed in a multitier-based clustering framework, and a subarea division algorithm was used to partition the entire network into several zones with varying lengths and widths. The number of zones varied with the size of the network, and each zone was divided into a few clusters. The number of clusters near the base station was larger compared with distant zone clusters. The CH selection depended on the metrics, namely distance and energy. After selecting appropriate CH, optimum number of relays were selected in each cluster. The nodes in the zone near the base station could only perform direct communication with the base station, and there was no direct communication for distant zones. Hence, energy depletion increased as distance increased, and energy wastage was also high owing to more clusters nearer to the base station.

Pius Agbulu et al.[17] proposed a relaying algorithm for multihop, clustered WSNs. The authors used a hybrid K-means clustering algorithm for effective clustering and a gradient descent algorithm for relay selection. The metrics, namely residual energy and distance, were used for CH selection. Among the non-CH nodes, the node with the highest residual energy, minimal path losses between the CHs, shortest transmission distances, and adequate coverage was selected as the relay node. The energy consumption increased due to the selection of a non-CH as a relay node. Suman Prakash et al.[18] proposed a novel clustering algorithm to balance the energy by selecting optimal CHs and relay nodes on the basis of the metrics, namely delay, distance, link lifetime, and energy. The hybrid heuristic data aggregation protocol was used to select the optimal number of CHs and relay nodes. The authors computed the distance based on the received signal strength indicator (RSSI) to select the optimal number of relay nodes, but they did not consider the effective distance between the SNs and the sink node.

Luo et al.[19] concentrated on low-cost, fixed-clustering problems and proposed a random relay-based, fixed-clustering protocol called random relay selection clustering protocol for energy harvesting (RRCEH). This protocol split an entire network into k number of ring-shaped regions with equal width, where each region consisted of an equal number of CHs. Based on the transmitting area, the authors formulated a random relay matrix to select a few CHs as relays. Furthermore, the authors computed energy consumption in three states, namely data receiving state, data processing state, and intercluster communication state. They measured the consumed energy in two ways for intercluster communication, namely intercluster data transmission and reception. They measured the energy consumption but did not consider it for the relay selection mechanism. Darabkh et al.[20] suggested a routing scheme called low-power energy-aware and layering-based clustering and routing protocol (EA-CRP). The scheme was supported by a layering and clustering framework. Although EA-CRP effectively operated in scalable networks, fewer layers were formed as the network size grew, which caused long-distance communication between the layers. Consequently, the network lifetime was shorter for larger network areas.

Shukla and Tripathi[21] suggested an energy-efficient routing protocol and implemented it in three phases, namely hierarchical cluster formation, effective relay node selection, and efficient routing. They divided the entire network into a few clusters, and hierarchical communication was framed between the SNs and base station through relay nodes, cluster coordinators, and CHs. They performed the selection of relay nodes by considering node density and distance between the nodes and the sink node. Hence, the network lifetime was reduced when the node density increased. Jaiswal and Anand[22] proposed a CH selection mechanism based on the gray wolf optimization algorithm. They formulated a fitness function through which a CH was selected among the CM nodes which has maximum energy level, minimum neighboring distance. It also considers intra-cluster distance, minimum sink distance, and maximum node degree. Here, maximum node degree includes more neighboring nodes and maximum priority factor of the node. Furthermore, a distance-based relay selection strategy was introduced for effective and reliable intercluster routing. They did not consider the energy of distant nodes, which were incapable of transferring data toward the base station. Lin et al.[23] proposed a fan-shaped clustering mechanism for large-scale sensor networks. They divided the entire network into a few fan-shaped clusters and introduced various energy-saving schemes through CH selection, relay selection, and reclustering methods. Initially, they measured the central point in each layer and used it as a reference for distance measurement for all nodes in that layer. Relay selection was based on the distance only, and a non-CH node was selected as a relay node. The energy consumption increased, and the network lifetime decreased owing to the selection of a non-CH node as a relay node.

Barik et al.[24] proposed a device-to-device (D2D) relay node (DRN) selection and uplink transmission power allocation (DSPA) algorithm to reduce the energy consumption at CHs. They considered energy consumption during transmission and reception to evaluate the energy consumption. CH selection was performed based on the channel quality index. Two types of CHs were considered, namely active and inactive, and the inactive CH was regarded as a DRN. Furthermore, the base station constantly identified the CHs with low battery levels so that the DRNs could take over the transmission in place of the standard multihop communication. The energy consumption was effectively computed, but the distance between the SN and the base station was not considered. Zhang et al.[25] proposed a new mechanism called hybrid tree-based and cluster-based routing protocol for raw data collection to enhance the network lifetime. Using this mechanism, they collected raw data through multihop communication without any redundant nodes. They considered uniform and nonuniform distances between the nodes for the selection of relay nodes to save energy. They did not compute the energy consumed at each CH in transferring the data toward the destination.

Furthermore, although relays increase the network lifetime, their deployment needs an extra battery for power supply. In such cases, energy harvesting assists the network or node in sustaining for a longer time. Simultaneous wireless information and power transfer (SWIPT) is one of the energy harvesting techniques and is useful for cooperative relaying for next-generation wireless networks[26]. Mao et al.[27] evaluated the performance of multihop relaying employing energy harvesting. They used the amount of harvested energy for the next-hop relay selection. They also framed a relay selection mechanism for the time-switching (TS) and power-splitting (PS) protocols. They computed the largest number of hops from the given source node's initial energy. They concentrated on the amount of harvested energy but not on the number of relay nodes selected. Chen et al.[28] suggested a novel, multihop, cooperative relaying transmission technology in which the source and relays both harvested energy from cochannel interferences (CCIs) and subsequently used it for data transmission. Their proposed method identified the largest number of hops that could be supported on the basis of the amount of harvested energy.

Asiedu et al.[29] proposed a multihop, decode-and-forward SWIPT system, which transferred sensed data toward the destination via multihop relays. The authors used the PS protocol to harvest the energy from past-hop nodes and considered the metrics, namely distance and residual energy, for relay selection. Wu et al.[30] proposed a cooperative forwarding power (CFP) scheme for multihop wireless cooperative networks. In this scheme, multiple tasks were performed by the relays and receiver. Initially, the receiver received information from the past-hop node or transmitter and harvested the energy from the nearest relays. Explicitly, the relays nearer to the transmitter harvested the energy from the transmitter and sent it (not the data, only power) to the receiver. For next-hop relay selection, hop distance and available energy were considered, and the distance between the SN and sink node was not taken into account.

Wang et al.[31] designed an energy-efficient transmission (e-Trans) SWIPT for clustered WSNs. They introduced two types of transmissions, namely direct transmission (DT) and relay transmission (RT). The CH node was selected as a relay node on the basis of the received signal strength and relay selection factor. The SWIPT-PS architecture was considered for energy harvesting and data transmission. However, the relay selection became inefficient when the received signal strength was low. Han et al.[32] formulated a problem in multihop clustered networks for energy-efficient routing and proposed the energy-efficient cooperative SWIPT routing (EECSR) algorithm. They formulated two transmission links, namely information transmission and SWIPT, which were used to forward the information toward the destination either directly or through a relay. They selected one member node in between two CH nodes as an energy harvesting node, which was also considered a relay node. They measured energy efficiency for DT and RT links. Each CH consumed more energy and transmitted a weak radio frequency (RF) signal toward the member node to harvest the energy, and the number of relays required was more when the network density increased.

Babruvhan and Thippeswamy[33] proposed SWIPT-based energy-efficient packet transmission between the clusters in WSNs. In this transmission, each CH harvested energy from its member nodes, and this energy was spent on data collection and transmission. Next, a priority-based relay selection mechanism was used, which comprised two metrics, namely residual energy and distance. The highest priority node, i.e., the one with least distance and highest residual energy, was selected as a relay node. A non-CH node was selected as a relay node, following which the energy consumption increased.

Tables 1 and 2 present the existing state-of-the-art relay selection strategies for clustered, multihop IoT networks with and without energy harvesting, respectively. Most of the abovementioned works considered fixed-clustering mechanisms, leading to reduced network performance. Moreover, when the nodes are moved away from the sink node, then their residual energy starts to decrease due to uninterrupted communication with the respective destination. Most authors proposed CH and relay selection mechanisms based on their residual energy and distance, but they did not consider node density and coverage range. In addition, only the past-hop node's RF signal was considered for energy harvesting, and the remaining nodes' RF signals were not considered as a valid source for energy harvesting.

Problem outline. Upon reviewing all of the abovementioned methods, we observed that a uniform clustering mechanism was used for clustering the entire nodes and that non-CH nodes were selected as relay nodes. This increased the energy consumption and reduced the network lifetime. Moreover, most existing methods exhibit poor network performance when node density and network area increase. However, this study concentrates on the selection of an optimal number of

Table 1 List of the state-of-the-art methods cluster-based relay selection mechanisms without energy harvesting.

Reference	Methodology	Remark
[16]	1. A SEEP was proposed. 2. The metrics, namely distance and energy, were used for CH selection. 3. Relay node was selected from the CHs.	The number of clusters nearer to the base station was more, and energy depletion increased when the node was moved away from the base station.
[17]	1. A relaying algorithm for multihop, clustered networks was proposed. 2. A hybrid *K*-means algorithm was used for clustering, and a gradient descent algorithm for relay selection. 3. The metrics, namely residual energy and distance, were used for CH selection.	A non-CH node was selected as a relay node, and this increased the energy consumption in the network.
[18]	1. An HHDA protocol for optimal CHs and relay selection was proposed. 2. The metrics, namely distance, energy, link lifetime, and delay, were used for CH and relay selections.	The distance was computed based on the RSSI to select the optimal relay node, but the effective distance between the SNs and sink node was not considered.
[19]	1. An RRCEH technique for CH and relay selections was proposed. 2. A random relay matrix was formulated based on the transmitting area. 3. Energy consumption was evaluated in three phases.	Only distance was considered for relay selection, and available energy or consumed energy was not considered.
[24]	1. The DSPA algorithm was used to reduce energy consumption. 2. CH selection was performed based on the channel quality index. 3. Active and inactive types of CHs were considered, and the inactive CH was selected as a DRN.	Energy consumption was effectively computed, but the distance between the SN and base station was not considered.

Table 2 List of the state-of-the-art cluster-based relay selection mechanisms with energy harvesting.

Reference	Methodology	Remark
[30]	1. A CFP scheme for multihop wireless cooperative networks was proposed. 2. The relays nearer to the transmitter harvested energy from the transmitter and sent it (only power, not the data) to the receiver.	Limited performance was observed for large-scale networks because selected relays (SRs) were nearer to the sink node.
[31]	1. An e-Trans SWIPT for clustered WSNs was designed. 2. A CH node was selected as a relay node based on received signal strength and a relay selection factor. 3. DT and RT links were used.	Relay selection became inefficient when the received signal strength was low.
[32]	1. The EECSR algorithm was proposed. 2. They selected one member node in between two CH nodes as an energy harvesting node, which was also considered a relay node. 3. DT and RT links were used.	The number of relays required was high when node density increased.
[33]	1. SWIPT-based, energy-efficient packet transmission between the clusters in WSNs was proposed. 2. Each CH harvested energy from its member nodes. 3. A priority-based relay selection mechanism was used, which comprised two metrics, namely residual energy and distance.	A non-CH node was selected as a relay node, which resulted in fast energy depletion.

relays when node density increases, and it thus enhances the energy efficiency of the network.

3 Proposed work

We develop a novel clustering and relay selection mechanism with an aim to improving the network lifetime. We apply hierarchical clustering, in which all nodes are grouped into two types of clusters, CC and RCs. The CC is located within the surroundings of the sink node, while the RCs are located far away from the sink node. The CHs of the RCs are routed through the CH of CC (CCH) and few relays. The relay selection mechanism is applied only for the nodes in the RCs, while for nodes in the CC, the CH functions as a relay-based on the coverage capacity and set theorem at every layer. Our proposed mechanism determines the optimal number of relays that can cover the entire CHs in their upper layer.

3.1 Network model

The following assumptions are made to formulate the network for the proposed mechanism:

(1) The network comprises one sink node and N number of SNs. All these nodes are static nodes, and an equal amount of initial energy is provided to all SNs.

(2) All SNs are location-aware, heterogeneous, and battery operated.

(3) The sink node is placed at the center of the network, and it handles all signaling and routing issues.

(4) Each SN is equipped with a single antenna, and an energy harvesting capability is provided to each SN through a multisource-based hybrid SWIPT (H-SWIPT) technique[34]. Each SN in its communication range (CR) communicates with other SNs. The range of an SN is represented by $R_{\rm SN}$. Each node assigns a binary value to all other nodes in the network on the basis of the distance between them.

$$s=\begin{cases}1, & d \leqslant R_{\rm SN};\\ 0, & d > R_{\rm SN}\end{cases} \tag{1}$$

where d represents the distance between the target node and SN, and s the binary indicator. Figure 1 shows the network model comprising l number of layers, and the sink node is placed at the center of the network. In each layer, all SNs are clustered into two types based on a hybrid clustering mechanism, as discussed in Section 3.2. The CC is located in the layer nearer to the sink node, whereas RCs are located in the next layers. The energy model required for this research work is appended in Appendix A.

3.2 Hybrid clustering

Hybrid clustering mainly aims to improve the network lifetime and groups all the SNs into two types of clusters, namely CC and RCs. The CC is the cluster nearer to the sink node. The maximum number of nodes, $n_{\rm m}$, to be included in the CC is decided as

$$n_{\rm m} = p_n \times N \tag{2}$$

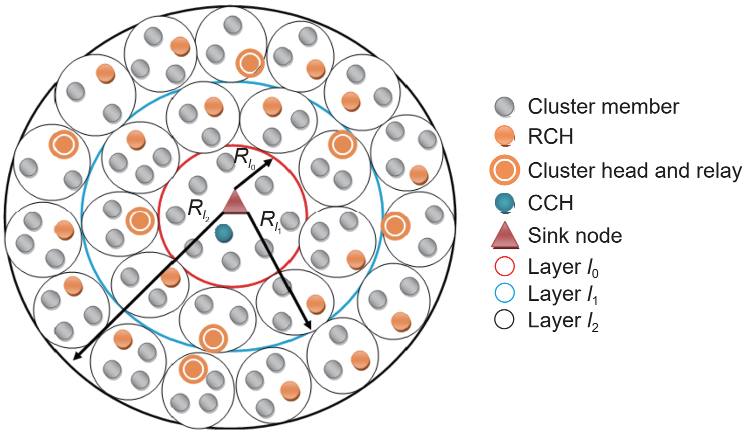

Fig. 1 Layered network model.

where N represents the total number of SNs and p_n represents the percentage of nodes in the CC. After clustering a few nodes into the CC by using Eq. (2), the remaining SNs are clustered into the RCs. The number of RCs[35] is calculated as

$$\mathrm{RC_{opt}} = \sqrt{\frac{N_{\mathrm{A}}^*}{2\pi} \times d_{\mathrm{Th}} \times \frac{M}{d_{\mathrm{CCH}}^2}} \tag{3}$$

where $N_{\mathrm{A}}^* = N_{\mathrm{A}} - N_{\mathrm{CC}}$ denotes the number of alive nodes except CC nodes (N_{A} represents the total number of alive nodes in the network and N_{CC} represents the total number of alive member nodes in the CC). d_{Th} represents the threshold distance. $M \times M$ denotes the transmitting region and d_{CCH} represents the average Euclidean distance between the CCH and the remaining nodes in the network. After calculating the number of clusters using Eq. (3), the fuzzy c-means (FCM) algorithm is applied to cluster the remaining nodes into an optimum number of RCs. FCM is one of the efficient clustering algorithms that optimally groups all the nodes into different clusters[36–39]. Here, it uses location information to cluster the nodes on the basis of membership values and centroids. The complexity incurred at clusters formation is demonstrated in Lemma A1 in Appendex A.

3.3 CH selection

After clustering all SNs into two types of clusters, CH selection is performed for effective data communication. The selection of CHs in the CC and RCs is discussed in the following sub-sections.

3.3.1 CCH selection

A fitness function is formulated for CCH selection, and it is the combination of the metrics, namely residual energy and relative distance. The fitness function for CCH selection is expressed as

$$F_{\mathrm{CCH}} = \beta \times X + (1-\beta) \times Y \tag{4}$$

where $X = E_{\mathrm{re}}/E_{\mathrm{i}}$ represents the energy (E_{re} denotes residual energy[40, 41], and E_{i} denotes the initial energy), $Y = \left(d_{\mathrm{Th}} - d_{\mathrm{CCM}}^{\mathrm{SinkNode}}\right)/d_{\mathrm{Th}}$ represents the relative Euclidean distance (d_{Th} denotes the threshold distance[42] and $d_{\mathrm{CCM}}^{\mathrm{SinkNode}}$ denotes the distance between the CM of CC and sink node), and β and $(1-\beta)$ the weights of the functions X and Y, respectively. When communication is initiated, the residual energy of each node decreases because of intermittent operations among the nodes. For each round of operations, the value of β[43] is updated based on the residual energy and threshold energy E_0. Noteworthily, E_0 is assessed based on γ and E_{i}; i.e., $E_0 = \gamma \times E_{\mathrm{i}}$, where γ assumes various such values as 0.20, 0.40, 0.60, and 0.80 to update β. Therefore, we have

$$\beta = \begin{cases} 0.5-0.8, \text{ if } E_{\mathrm{min}}^{\mathrm{C}} \geqslant \gamma_{\mathrm{i}} \times E_{\mathrm{i}}; \\ 0.9, \text{ otherwise} \end{cases} \tag{5}$$

where $\gamma_i = \{0.80, 0.60, 0.40, 0.20\}$, and $E_{\mathrm{min}}^{\mathrm{C}}$ represents the minimum residual energy among the CM nodes of the selected cluster and is given by

$$E_{\mathrm{min}}^{\mathrm{C}} = \begin{cases} \min\limits_i \left\{E_{\mathrm{re}}^i\right\}, \ i = 1,2,\ldots,N_{\mathrm{CC}}, \text{ for CC}; \\ \min\limits_i \left\{E_{\mathrm{re}}^i\right\}, \ i = 1,2,\ldots,N_{\mathrm{RC}}, \text{ for RC} \end{cases} \tag{6}$$

where N_{RC} represents the total number of CM nodes in the RC. The value of β is updated using Eqs. (4) and (5) and is used to calculate the fitness value of each CM, and based on the values obtained, the node with the optimal fitness value is selected as a CH.

3.3.2 RCH selection

In the RC, member nodes nearer to the centroid have maximum residual energy. A fitness function is used for CH selection, (i.e., RCH) based on the metrics such as residual energy and distance. If the remote CMs are far away from the centroid $\left(d_{\mathrm{RCM}}^{\mathrm{Centroid}} > d_{\mathrm{mean}}^{\mathrm{RC}}\right)$, then that function includes both metrics, namely distance and energy; otherwise, it includes only the metric of energy. Therefore, the fitness function of RCH is given as

$$F_{\mathrm{RCH}} = \begin{cases} \beta \times X + (1-\beta) \times Z, \text{ if } d_{\mathrm{RCM}}^{\mathrm{Centroid}} > d_{\mathrm{mean}}^{\mathrm{RC}}; \\ X, \text{ otherwise} \end{cases} \tag{7}$$

where β is updated in each round using Eq. (5), and $Z = \left(d_{\max}^{\mathrm{RC}} - d_{\mathrm{RCM}}^{\mathrm{Centroid}}\right) / d_{\max}^{\mathrm{RC}}$ represents the relative Euclidean distance between CM and centroid of a particular RC ($d_{\max}^{\mathrm{RC}}$ denotes the maximum distance between the CM and centroid is given by $d_{\max}^{\mathrm{RC}} = 1 + \max\left(d_{\mathrm{RCM}_i}^{\mathrm{Centroid}}\right), i = 1, 2, \cdots, n_{\mathrm{RC}}$), and

$$d_{\mathrm{mean}}^{\mathrm{RC}} = \mathrm{mean} \sum_{i=1}^{n_{\mathrm{RC}}} d_{\mathrm{RCM}_i}^{\mathrm{Centroid}} \tag{8}$$

Therefore, CH selection for the CC and RCs is performed after all the nodes are clustered into several clusters using the FCM algorithm. Equations (4) and (7) are used to calculate the fitness value of each CM, and the node with the maximum fitness value is selected as a CH.

Algorithm 1 describes the step-by-step process of CH selection for CC and RCs. As communication progresses, each node's residual energy decreases. After several rounds, the energy at each node in the network is not sufficient to forward the sensed or collected data to the respective CH or sink node, and this implies the death of a node. Therefore, a large data packet loss occurs and also leads to reduction in network lifetime. To avoid this loss, this study provides each node the ability to harvest energy by using a multisource-based H-SWIPT technique[34].

H-SWIPT is a combination of the TS and PS protocols. Here, multiple sources are considered for energy harvesting such as sink node's RF energy, CCI, and neighbor nodes' RF signal, in addition to the past-hop node's RF signal. In this technique, each node harvests energy using the RF signals transmitted in the network from multiple sources and simultaneously transfers the energy and data to the CH or sink node. For this purpose, the expression for residual energy is modified and is provided in Eq. (9). Accordingly, when the node's residual energy is less than the threshold energy, the corresponding node starts to harvest energy using the H-SWIPT technique; otherwise, it continues the process with the available energy. The modified residual energy[34] is expressed as

$$\widehat{E}_{\mathrm{re}} = \begin{cases} E_{\mathrm{re}} + E_{\mathrm{T}}^{\mathrm{eh}}, \text{ if } E_{\mathrm{re}} < E_0; \\ E_{\mathrm{re}}, \text{ otherwise} \end{cases} \tag{9}$$

where $E_{\mathrm{T}}^{\mathrm{eh}}$ represents total energy harvested from multiple sources using the H-SWIPT architecture at each sensor node in the network. Once the CH selection is completed for both types of clusters, the network is formulated into hierarchical layers.

Here, the network is divided into a few layers, and relays are introduced to reduce the burden on CHs in a bid to improve the network lifetime and reduce energy consumption. The complexity incurred at CH selection is demonstrated in Lemma A2 in Appendix A.

Algorithm 1 CH Selection

```
for CCH selection do
  Update β using Eqs. (5) and (6)
  for N_CC do
    CM obtains its F_CCH value using Eq. (4)
  end
  The CM with the highest F_CCH is called CCH
end
for RCH selection do
  Update β using Eqs. (5) and (6)
  for N_RC do
    CM obtains its F_RCH value using Eq. (7)
  end
  The CM with the highest F_RCH is called RCH
end
```

3.4 Hierarchical structuring

In hierarchical structuring, the entire clustered network is divided into several layers. For division into layers, we considered as a reference the maximum coverage (R) of the sink node. Figure 1 shows the layered network model of the proposed work. Initially, all clusters are divided into l number of layers, such as $l_0, l_1, \ldots$. The layer nearer to the sink node is considered l_0, where the CC is located. The distant layers are denoted as $l_1, l_2, \ldots$, at which only the RCs are located. In each layer, for each cluster, one node is elected as a CH, and all remaining nodes are considered CMs. In the CC, the CH acts as a relay, whereas in the RCs, a few CHs are selected as relays.

Furthermore, the CR of the sink node, R, is defined as the maximum Euclidean distance between the sink node and SN.

$$R = \max\left(d_{\text{SinkNode}}^{\text{SN}}\right) \tag{10}$$

Initially, the coverage range for layer l_0 is defined as the maximum distance between the sink node and the most distant member node of CC. Therefore, the coverage range for layerl_0 is calculated as

$$R_{l_0} = \max\left(d_{\text{SinkNode}}^{\text{CCM}}\right) \tag{11}$$

Similarly, the coverage range for layer l_1 is the cumulative range of coverage of layer l_0, R_{l_0}, and some percentage of the CR of coverage of the sink node, R. Thus, the coverage range for layer l_1 is expressed as

$$R_{l_1} = R_{l_0} + (X \times R/100) \tag{12}$$

where R_{l_1} denotes the range of coverage for layer l_1, R_{l_0} denotes the range of coverage for layer l_0, X denotes a constant, and R denotes the communication coverage range of the sink node. In a generalized form, the coverage range of the l_i-th layer is expressed as

$$R_{l_i} = R_{l_{i-1}} + (X \times R/100) \tag{13}$$

where R_{l_i} denotes the coverage range of layer l_i, where $i = 1, 2, \ldots$, and $R_{l_{i-1}}$ denotes the coverage range of layer l_{i-1}. After partitioning the entire network into a few layers, a few CHs in the l_i-th layer are selected as relays for the CHs of the l_{i+1}-th layer except layer l_0. This optimal relay selection reduces the energy consumption and effectively forwards the data packets toward the sink node.

3.5 Relay selection

Generally, relays are used for long-distance communication in multihop IoT networks. The relay selection mechanism is applied only for nodes in the RCs, whereas for nodes in CC, the CH functions as a relay. Few RCHs of the l_i-th layer are considered relays for the RCHs of the l_{i+1}-th layer. Let us consider that the RCHs of the l_i-th layer are represented by CH_i^p, where $p = 1, 2, \ldots, s$, and that those of the l_{i+1}-th layer are CH_{i+1}^q, where $q = 1, 2, \ldots, t$ and $i = 1, 2, \ldots$. Initially, the l_i-th layer's RCHs are considered candidate relays for the l_{i+1}-th layer's RCHs, which need support for data transmission. Furthermore, for optimal relay selection, initially, the Euclidean distance between the CH_i^p-th candidate relay and $\text{CH}_{(i+1)}^q$-th RCHs is computed. If the distance obtained is within the CR of the CH_i^p-th candidate relay ($\text{CR}(\text{CH}_i^p)$), then the $\text{CH}_{(i+1)}^q$-th RCH comes under the respective candidate relay's proximity. Therefore, the node proximity for the CH_i^p-th candidate relay is given as

$$P_i^p = \begin{cases} \text{CH}_{i+1}^q, \text{ if } d_{\text{CH}_i^p}^{\text{CH}_{i+1}^q} \leqslant \text{CR}\left(\text{CH}_i^p\right); \\ 0, \text{ otherwise} \end{cases} \tag{14}$$

Optimal relay selection is performed after obtaining the CH_i^p-th candidate relay's proximity. Thereafter, we check for a single relay in the l_i-th layer that covers the maximum number of RCHs of the l_{i+1}-th layer. Hence, the optimal relay for the l_i-th layer is selected as

$$\text{OR}_{l_i} = \max_p P_i^p \tag{15}$$

where OR_{l_i} represents a single optimal relay for the l_i-th layer and P_i^p represents the candidate relay set for the l_i-th layer. If the single relay is not sufficient to cover the entire CHs of the next layer, then we check for the combination of two or more relays. For this purpose, we check for the optimal combinations of candidate relays. Among possible combinations, the best combination, i.e., one with maximum node coverage capacity, is selected. As there exist varying numbers of nodes and CHs in the network, we adopt multirelay combinations (MRCs). In this, the number of combinational relays varies depending upon the node count and CH count. Optimal MRC selection is performed as

$$\text{OR}_{l_i} = \max_p\left(P_i^p \cup P_i^{p+1}\right) \tag{16}$$

where P_i^p and P_i^{p+1} represent the candidate relay sets for the l_i-th layer. The abovementioned equation checks for two relay combinations that have the maximum node coverage capacity in the l_i-th layer.

If the entire RCHs in the l_{i+1}-th layer are not covered with the two relay combination candidates, then we will check for more MRCs. Hence, the proposed relay selection mechanism selects an optimal number of relays from the set of candidate relays in the l_i-th layer until it covers the entire RCHs of the next layer, and it is given by

$$\mathrm{OR}_{l_i} = \max_{p}\left(P_i^p \cup P_i^{p+1} \cup \cdots \cup P_i^s\right) \tag{17}$$

After selecting optimal relays for effective data transmission in the l_i-th layer, the remaining CHs send their aggregated data packets to their respective destinations. Once all the CHs and relays have been selected, they propagate an advertisement message to their CMs and to all the CHs for packet routing. Few CMs and CHs of RCs that are ready for communication send their join request messages to appropriate CHs and relays. Furthermore, all the CHs and SRs follow a time division multiple access schedule to transfer the data from CMs to CHs. Similarly, all the CHs or relays send their aggregated data packets to the sink node using the carrier sense multiple access/collision avoidance scheduling algorithm based on the proposed routing mechanism. Algorithm 2 shows the relay selection mechanism for RCs.

Algorithm 2 Relay selection for RCs

for upper layers except l_0 **do**
  for $\mathrm{CH}_i^p, p = 1,2,\ldots,s$
    for $\mathrm{CH}_{i+1}^q, q = 1,2,\ldots,t$
      $P_i^p \leftarrow \varnothing$
      Compute the distance between CH_i^p and CH_{i+1}^q
      Compute the CR of CH_i^p
      if $d_{\mathrm{CH}_i^p}^{\mathrm{CH}_{i+1}^q} \leqslant \mathrm{CR}\left(\mathrm{CH}_i^p\right)$ **then**
        $P_i^p \leftarrow P_i^p \cup \left\{\mathrm{CH}_{i+1}^q\right\}$
    end
  end
  for relay selection **do**
    for $p = 1,2,\ldots,s$
      if $\mathrm{coverage}\left(\mathrm{CH}_i^p\right)$ = total number of CHs in l_{i+1}-th layer
        $\mathrm{OR}_{l_i} = \mathrm{CH}_i^p$
      else if $\mathrm{coverage}\left(\mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1}\right)$ = total number of CHs in l_{i+1}-th layer
        $\mathrm{OR}_{l_i} = \mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1}$
      else if $\mathrm{coverage}\left(\mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1} \cup \mathrm{CH}_i^{p+2}\right)$ = total number of CHs in l_{i+1}-th layer
        $\mathrm{OR}_{l_i} = \mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1} \cup \mathrm{CH}_i^{p+2}$
      else $\mathrm{coverage}\left(\mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1} \cup \cdots \cup \mathrm{CH}_i^s\right)$ = total number of CHs in l_{i+1}-th layer
        $\mathrm{OR}_{l_i} = \mathrm{CH}_i^p \cup \mathrm{CH}_i^{p+1} \cup \cdots \cup \mathrm{CH}_i^s$
    end
  end
end

3.6 Hierarchical packet routing

The proposed mechanism includes the HPR strategy for energy-efficient data transmission. Initially, all the CMs and CCH verify the distance between them and the sink node in layer l_0. In layer l_0, if the distance between the CM of CC and the sink node $\left(d_{\mathrm{CCM}}^{\mathrm{SinkNode}}\right)$ is less than or equal to the distance between the CCM and CCH $\left(d_{\mathrm{CCM}}^{\mathrm{CCH}}\right)$, then the CCM directly sends its data packets to the sink node; otherwise, it will send them to CCH. Hence, the destination of the CCM of layer l_0 can be expressed as

$$\mathrm{Dest}_{\mathrm{CCM}}^{l_0} = \begin{cases} \mathrm{SinkNode}, \text{if } d_{\mathrm{CCM}}^{\mathrm{SinkNode}} \leqslant d_{\mathrm{CCM}}^{\mathrm{CCH}}; \\ \mathrm{CCH}, \text{if } d_{\mathrm{CCM}}^{\mathrm{SinkNode}} > d_{\mathrm{CCM}}^{\mathrm{CCH}} \end{cases} \tag{18}$$

Similarly, the CMs of the RCs in layer l_1 send their data packets to respective RCHs. Next, the RCHs forward the collected data packets to either the sink node or CCH on the basis of the distance. Simultaneously, the SRs presented in layer l_1 collect and aggregate the data packets from next layer's RCHs and forward them to either the sink node or CCH. In layer l_1, if the distance between the RCH or SR and sink node $\left(d_{\text{RCH or SR}}^{\mathrm{SinkNode}}\right)$ is less than or equal to the distance between the RCH and CCH $\left(d_{\text{RCH or SR}}^{\mathrm{CCH}}\right)$, then the cumulated data packets are directly forwarded to the sink node; otherwise, they are sent to CCH. Therefore, the destination of CHs or SRs of layer l_1 can be expressed as

$$\mathrm{Dest}_{\text{RCH or SR}}^{l_1} = \begin{cases} \text{Sink Node, if } d_{\text{RCH or SR}}^{\mathrm{SinkNode}} \leqslant d_{\text{RCH or SR}}^{\mathrm{CCH}}; \\ \text{CCH, if } d_{\text{RCH or SR}}^{\mathrm{SinkNode}} > d_{\text{RCH or SR}}^{\mathrm{CCH}} \end{cases} \tag{19}$$

From layer l_2 layer onward, the SNs will not have direct communication with the sink node and will rely on the previous layer's relays to transfer data packets. The destination of the RCHs of layer l_2 is decided based on the distance between its RCH and SRs of layer l_1. Let us suppose that two relays of layer l_1 $\left(\mathrm{CH}_1^{\mathrm{r}1} \text{ and } \mathrm{CH}_1^{\mathrm{r}2}\right)$ cover entire RCHs of layerl_2, then the destination of the RCHs of layerl_2 can be expressed as

$$\mathrm{Dest}_{\mathrm{RCH}}^{l_2} = \begin{cases} \mathrm{CH}_1^{\mathrm{r}1}, \text{ if } d_{\mathrm{CH}_2^{\mathrm{RCH}}}^{\mathrm{CH}_1^{\mathrm{r}1}} \leqslant d_{\mathrm{CH}_2^{\mathrm{RCH}}}^{\mathrm{CH}_1^{\mathrm{r}2}}; \\ \mathrm{CH}_1^{\mathrm{r}2}, \text{ if } d_{\mathrm{CH}_2^{\mathrm{RCH}}}^{\mathrm{CH}_1^{\mathrm{r}1}} > d_{\mathrm{CH}_2^{\mathrm{RCH}}}^{\mathrm{CH}_1^{\mathrm{r}2}} \end{cases} \tag{20}$$

where $\mathrm{CH}_1^{\mathrm{r}1}$ and $\mathrm{CH}_1^{\mathrm{r}2}$represent the two SRs of layer l_1. The abovementioned equation describes the destination of the RCHs of layer l_2. If the distance between the RCH of layer l_2 $\left(\mathrm{CH}_2^{\mathrm{RCH}}\right)$ and the SR of layer l_1 $\left(\mathrm{CH}_1^{\mathrm{r}1}\right)$ is less than or equal to the distance between $\mathrm{CH}_2^{\mathrm{RCH}}$ and other SR of layer l_1 $\left(\mathrm{CH}_1^{\mathrm{r}2}\right)$, then the collected and aggregated data packets of respective RCHs are forwarded through the relay $\mathrm{CH}_1^{\mathrm{r}1}$;otherwise, they are forwarded through the relay $\mathrm{CH}_1^{\mathrm{r}2}$. In this manner, the upper layers maintain communication with CHs of the lower layers to transfer the data packets to the sink node.

4 Simulation

The performance of the proposed method is evaluated in this section via various simulation experiments. Initially, this section explores the particulars of the simulation setup and then the results.

4.1 Simulation setup

MATLAB is used to simulate the proposed method. We consider a network of 100 nodes randomly located in an area of 1000 m × 1000 m with the sink node located at the center. We assum a small-scale relay fading communication channel and that all the receiving nodes have equal priority, i.e., $\delta_{ij} = 1$, $\forall i$, $j \in d_{ij}$. Furthermore, we consider similar energy harvesting parameters to those used in Ref. [30]. We set the maximum transmitting power $P_{\max}$ to be 100 MW, antenna noise variance ${\sigma_{ij}}^2$ = −50, −40, and −30 dBm, and minimum harvesting energy to be 10% of $P_{\max}$. Table 3 presents the parameters required to simulate the proposed network.

Table 3 Simulation setup.

Parameter	Value
Node count (N)	100
Network area	1000 m × 1000 m
Size of the control packet	200 bit
Size of the data packet	4000 bit
Node placement	Random
Transmission range of each node	250 m
Initial energy (E_{i})	1 J
p_n	0, 0.2, 0.4, 0.6, 0.8, 1.0
Energy consumed in free space (E_{fs})	10 pJ/bit/m^2
Energy consumed in multipath propagation environment (E_{mp})	0.0013 pJ/bit/m^4
d_{Th}	87.7058 m
Energy consumed to transmit or receive 1-bit information (E_{Elec})	50 nJ/bit
${\sigma_{ij}}^2$	−50, −40, −30 dBm
Number of simulation rounds	1500

4.2 Simulation results and analysis

The performance of the proposed method is evaluated through the optimum value of p_n and energy efficiency metric under varying numbers of relays, noise variances, transmitting powers, number of rounds, and nodes under the RT and DT modes. The DT mode uses a DT link to transmit data to the sink node, whereas the RT mode uses relays for the same. Furthermore, the proposed method's performance is compared with that of the e-Trans SWIPT[31] and EECSR[32] methods.

4.2.1 Optimum value of p_n

The maximum number of nodes that can be included in the CC is decided based on p_n (see Fig. 2). We vary p_n from 0 to 1.0 in extensive simulation experiments to determine its optimum value. The number of rounds is

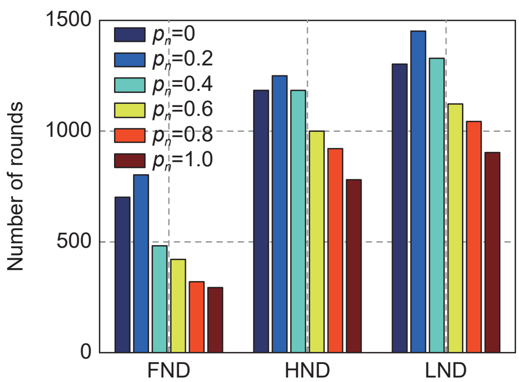

Fig. 2 Optimum value of p_n for closest clustering.

measured for each simulation for three performance metrics, namely first node death (FND), half node death (HND), and last node death (LND). The optimum value of p_n is determined when these three metrics acquire largest number of rounds. From Fig. 2, it can be seen that the optimum value of p_n is 0.20, as it covers the largest number of rounds for all the abovementioned metrics. This optimum value is fixed for the rest of the simulation experiments.

4.2.2 Energy efficiency

The energy efficiency of wireless communication systems quantifies the number of bits of information consistently conveyed from a transmitter to a receiver per unit of energy used at the transmitter. In most cases, the energy is consumed by the data processing circuitry, which is considered static power consumption, and to send and receive signals from various SNs through a wireless fading medium. The performance of the proposed method was evaluated using the metric of energy efficiency. Therefore, we considered system energy efficiency for our proposed method as

$$\text{Energy efficiency} = \frac{\text{throughput (bit)}}{\text{Total energy consumed (J)}} \tag{21}$$

Figure 3 shows the average energy efficiency (kbit/J) increment versus the increased number of relays under the two transmission modes with varying noise variance. From Fig. 3, we can observe that at the same noise level, the RT mode's energy efficiency is higher than that of the DT mode. Compared with RT, DT has larger fading effects, more path loss, and a smaller coverage area, leading to increased energy consumption and decreased energy efficiency.

Furthermore, for the two modes, the energy efficiency at −50 dBm noise variance is higher than that at −40 dBm. From Fig. 3, the average energy efficiency is observed at ${\sigma_{ij}}^2 = -50$ dBm to be approximately 3544 and 5394 kbit/J for the DT and RT modes, respectively. Similarly, the average energy efficiency at ${\sigma_{ij}}^2 = -40$ dBm for the DT and RT modes is approximately 2466 and 3857 kbit/J, respectively. Hence, at various noise variances, the RT mode of transmission is more energy efficient than the DT mode of data transmission.

Figure 4 shows the graph of energy efficiency (kbit/J) versus maximum transmit power (dBm) under the two transmission modes. An increase in maximum transmit power immediately results in a gain in the system's energy efficiency and is independent of the mode of transmission because a higher value of transmit power increases the signal-to-interference-plus-noise ratio (SINR) at the receiver.

Furthermore, when the maximum transmit power is greater than or equal to 20 dBm, the system's energy efficiency increases, reaches an optimal value, and remains constant. A further increase in transmission power would reduce the system's efficiency once the efficiency has reached its optimum level owing to the energy-limited sensors' excessive energy usage. From Fig. 4, we can observe that the RT mode's energy efficiency is higher than that of the DT mode. On average, the energy efficiency under the RT mode is

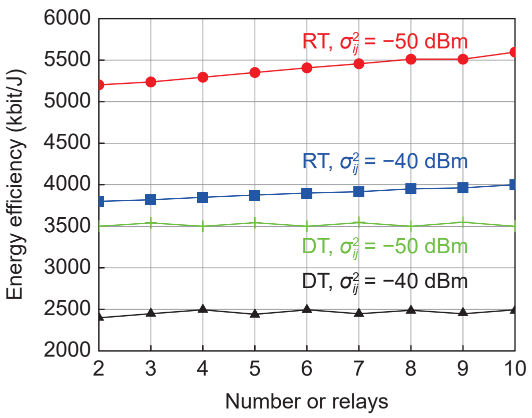

Fig. 3 Energy efficiency (kbit/J) with varying numbers of relays.

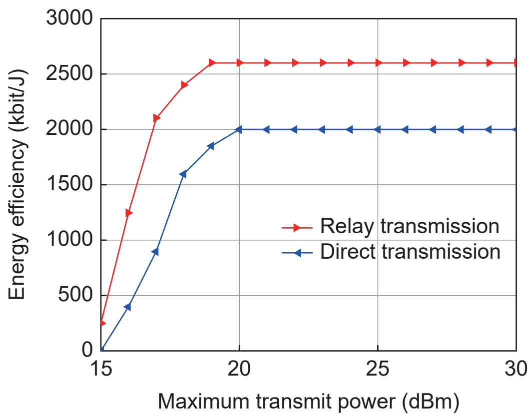

Fig. 4 Energy efficiency (kbit/J) with varying maximum transmit power (dBm).

700 kbit/J higher than that under the DT mode. Hence, in terms of average energy efficiency, the RT mode is 35% better than the DT mode.

Figure 5 shows the graph for energy efficiency versus the number of rounds under different noise variances and the number of relays. From Fig. 5, we can observe that the system's energy efficiency increases with decreased noise variance and an increased number of relays. Moreover, a decrease in noise power improves the SINR, and an increase in the number of relays can maximize the probability of selecting a more energy-efficient transmission path. At −50 dBm noise variance, the approximated average energy efficiency is 3533 kbit/J when the number of relays (NR) is 6 and 3266 kbit/J when NR is 3. Similarly, at −40 dBm noise variance, the approximated average energy efficiency is 2550 kbit/J when NR = 6 and 2266 kbit/J when NR = 3. Furthermore, at −30 dBm noise variance, the approximated average energy efficiency is 1700 kbit/J when NR = 6 and 1566 kbit/J when NR = 3. Hence, the percentage of improvements in the average energy efficiency for ${\sigma_{ij}}^2$ = −50 dBm over ${\sigma_{ij}}^2$= −40 and −30 dBm are 29% and 51%, respectively. Therefore, the energy efficiency is high when the noise variance is low and NRs are more.

Figure 6 shows the number of relays versus varying node density for the proposed method and the existing methods, namely e-Trans SWIPT[31] and EECSR[32]. From Fig. 6, we can observe that more relays are required to reach the destination as the number of nodes increases. The proposed method selects an optimal number of relays compared with e-Trans SWIPT and EECSR because the latter two methods select the relay node irrespective of the number of nodes when the distance between the two CHs is high and do not consider the distance between the CH and sink node.

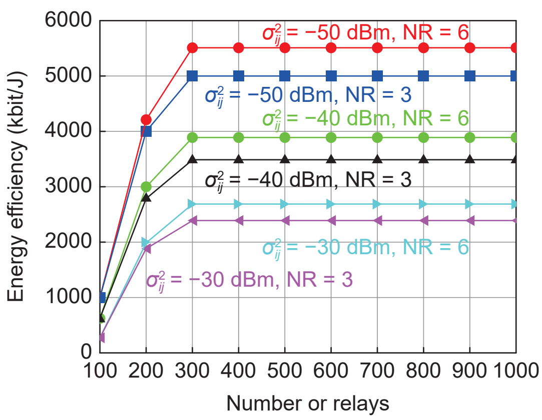

Fig. 5 Energy efficiency (kbit/J) with varying numbers of rounds.

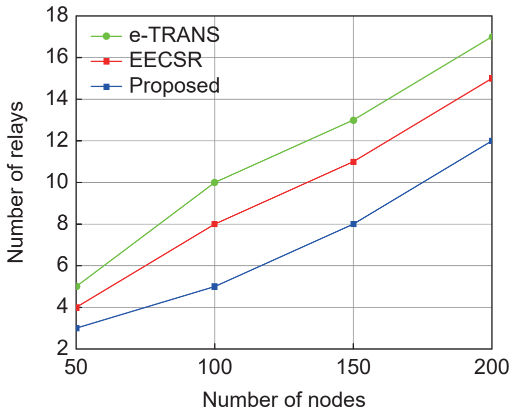

Fig. 6 Number of relays required versus varying numbers of nodes.

Here, we consider both distances and select an optimal number of relay nodes depending on the node density. From Fig. 6, we can observe that, on average, the NR required for the proposed method is 6, whereas it is 10 and 12 for the EECSR and e-Trans SWIPT methods, respectively. Hence, the proposed method requires a lower number of relay nodes when the node density is low, and as node density increases, an optimal NR is required when compared with the existing methods.

Figure 7 shows the graph for energy efficiency (kbit/J) versus the number of nodes under the RT mode. Here, the proposed method's energy efficiency is compared with that of the EECSR method at various noise variances under the RT mode of transmission with varying number of nodes. From the figure, we can observe that for the proposed method, when the number of nodes increases, the energy efficiency also increases at various noise variances. The percentage improvement in terms of energy efficiency at ${\sigma_{ij}}^2$ = −50 dBm of the proposed method compared with the EECSR method is 18.2% when the number of nodes is 50, 15.4% when the number of nodes is 100, and 11% when the number of nodes is 150.

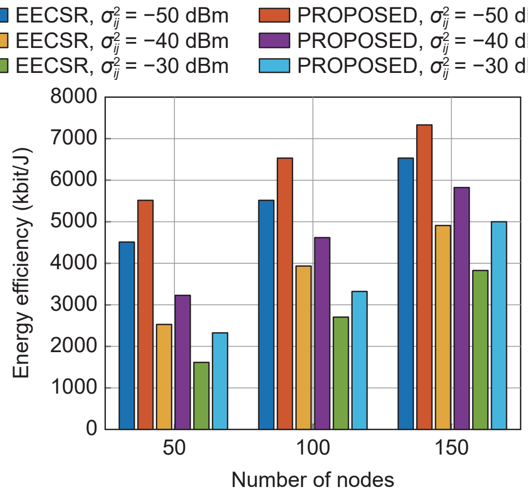

Fig. 7 Energy efficiency (kbit/J) with varying numbers of nodes under the RT mode.

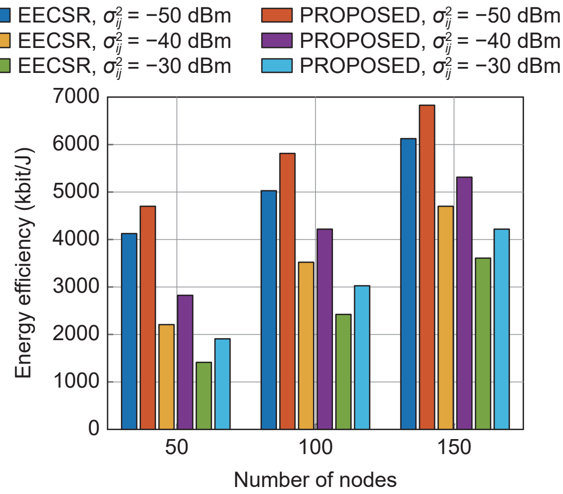

Fig. 8 Energy efficiency (kbit/J) with varying numbers of nodes under the DT mode.

Similarly, the percentage improvement in terms of energy efficiency at ${\sigma_{ij}}^2$ = −40 dBm of the proposed method compared with the EECSR method is 25% when the number of nodes is 50, 17% when the number of nodes is 100, and 15.5% when the number of nodes is 150. Furthermore, the percentage of improvement in terms of energy efficiency at ${\sigma_{ij}}^2$ = −30 dBm of the proposed method over the EECSR method is 36% when the number of nodes is fifty, 12.5% when the number of nodes is 100, and 24% when the number of nodes is 150. Therefore, under the RT mode, when the number of nodes increases, the proposed method's energy efficiency at fewer noise variances is higher compared with the EECSR method.

Figure 8 shows the graph of energy efficiency (kbit/J) versus the number of nodes under the DT mode. Here, the proposed method's energy efficiency is compared with that of the EECSR method at various noise variances under the DT mode with varying numbers of nodes. From Fig. 8, we can observe that for the proposed method, when the number of nodes increases, the energy efficiency also increases at various noise variances. The percentage average energy efficiency improvements of the proposed method compared with the EECSR method at ${\sigma_{ij}}^2$ = −50, −40, and −30 dBm are 12.5%, 21.4%, and 31.6%, respectively, when the number of nodes is 50.

Similarly, the percentage of average energy efficiency improvements of the proposed method over the EECSR method at ${\sigma_{ij}}^2$ =−50, −40, and −30 dBm are 13.8%, 16.6%, and 20%, respectively, when the number of nodes is 100. Furthermore, the percentage of average energy efficiency improvements of the proposed method over the EECSR method at ${\sigma_{ij}}^2$ = −50, −40, and −30 dBm are 10.3%, 9.4%, and 16.6%, respectively, when the number of nodes is 150.

From Figs. 7 and 8, we can observe that the RT mode's energy efficiency is higher than that of the DT mode because higher node density results in higher energy consumption owing to increased path loss and fading effects between the nodes and CHs. Hence, higher energy consumption results in lower energy efficiency.

Figures 9 and 10 show the required NR for varying CR at a node count of 100 and 200, respectively. The

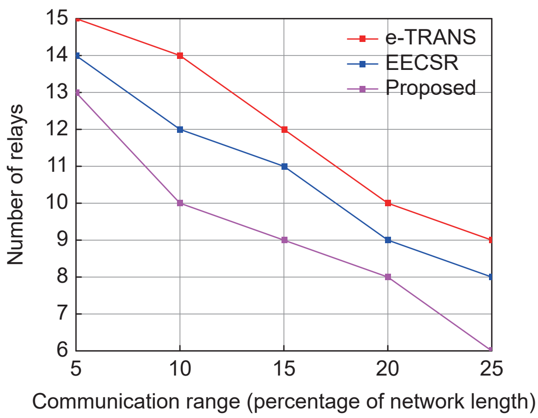

Fig. 9 Number of relays for varying CRs when the number of nodes is 100.

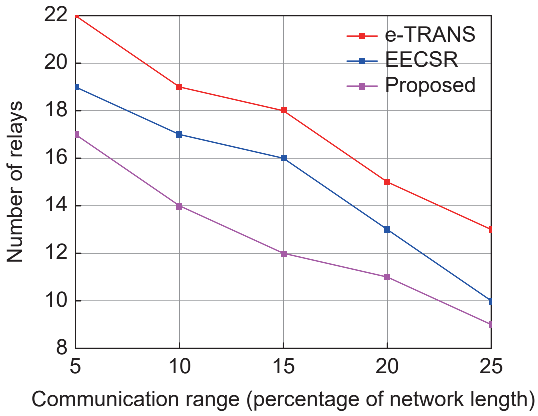

Fig. 10 Number of relays for varying CR when the number of nodes is 200.

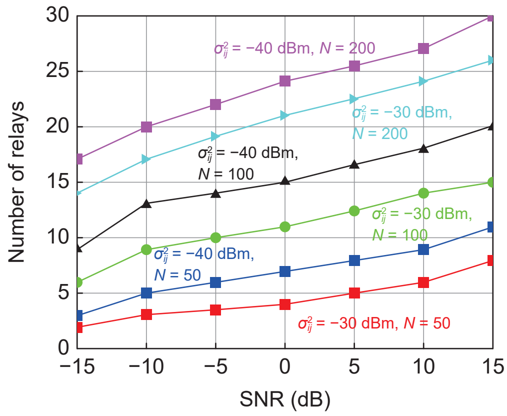

Fig. 11 Number of relays for varying SNR (dB) for different node counts and noise variances.

required NR decreases when the CR of each node increases, as shown in the figures. From Fig. 9, we can observe that when a node's CR is less, then it requires more NR because each CH covers only a few nodes.

However, when the CR of a node increases, then each CH covers a maximum number of nodes, and there are fewer requirements for higher number of relays. If the number of nodes increases from 100 to 200, then each CH cannot cover a maximum number of nodes, as a result of which it requires more NR when the CR is less, as shown in Figs. 9 and 10. On average, the NR required for the proposed, EECSR, and e-Trans SWIPT methods is 9, 1, and 12, respectively, for a node count of 100, as shown in Fig. 9. On average, the NRs required for the proposed EECSR and e-Trans SWIPT methods are 13, 16, and 18, respectively, for a node count of 200, as shown in Fig. 10. From the results, we can observe that the performance of the proposed method is superior to that of the existing methods; i.e., the optimum NR required for the proposed method is less than those required by the existing methods for various node counts.

Figure 11 shows the graph for the number of relays versus varying signal-to-noise ratios (SNRs (dB)) under varying node densities and noise variances. From Fig. 11, we can observe that as the SNR increases, the number of relays also increases for different node densities and noise variances. As the SNR increases, noise variance decreases, the CR of each node decreases, and more relays are required to transmit information toward the sink node. For the low-SNR region, the NR required for a node count of 200 is more than those required for node counts of 100 and 50 at the same noise level. At −40 dBm, the average NR required for node counts of 200, 100, and 50 is 24, 15, and 7, respectively. At −30 dBm, the average NRs required for node counts of 200, 100, and 50 are 21, 11, and 5, respectively.

5 Conclusion

Energy management poses one of the major challenges in IoT networks. Hence, this study proposed an energy-efficient relay selection mechanism for clustered IoT networks to efficiently use the network's energy and provide reliable communication between the nodes. Initially, this study proposes hybrid clustering in which the entire network is divided into two types of clusters, namely the CC and RCs. Furthermore, CH selection is introduced by formulating a fitness function based on the metrics, namely distance and residual energy. After CH selection, the entire network is divided into a few layers, and some CH nodes are selected as relay nodes in each layer. The relay selection is applied only for the upper layers, where the RCs are located, and the CCH function as a relay. The proposed method's performance is examined through several simulation experiments, and it exhibits better results compared with the existing methods under varying noise variance and node density.

Appendix

A Energy model

This study uses the free space energy model or the multipath propagation model, both of which are employed based on the distance between the sender and receiver nodes (d)[40, 41]. The threshold distance (d_{Th}) is used to decide which propagation model (d^2 or d^4) will be used for the energy consumption. Generally, energy consumption at each node is evaluated based on four states, namely transmitting, receiving, idle, and sleep. Among these four states, the states' idle and sleep consume a negligible amount of energy. Therefore, we consider energy consumption during the transmitting and receiving states only. The amount of energy consumed to transmit a packet of k-bit length over a distance d is given by the following:

$$E_{\mathrm{tx}}(k,d)=\begin{cases}k\times E_{\mathrm{el}}+k\times E_{\mathrm{fs}}\times d^2, \text{ if } d\leqslant d_{\mathrm{Th}};\\ k\times E_{\mathrm{el}}+k\times E_{\mathrm{mp}}\times d^4, \text{ if } d>d_{\mathrm{Th}}\end{cases} \tag{A1}$$

where E_{el} signifies the energy consumed to transmit or receive 1-bit information and E_{fs} and E_{mp} denote the energies consumed in free space and multipath propagation environments, respectively. Additionally, $d_{\mathrm{Th}}=\sqrt{\dfrac{E_{\mathrm{fs}}}{E_{\mathrm{mp}}}}$ denotes the threshold distance[42]. The amount of energy consumed to receive a packet of k-bit length is given by

$$E_{\mathrm{rx}}(k)=k\times E_{\mathrm{el}} \tag{A2}$$

The residual energy E_{re} is then evaluated based on $E_{\mathrm{tx}}(k,d)$ and $E_{\mathrm{rx}}(k)$ as

$$E_{\mathrm{re}}=E_{\mathrm{Current}}-E_{\mathrm{Cons}} \tag{A3}$$

where E_{Cons} represents the total energy consumed during the transmission and reception of a k-bit data packet and is equal to the summation of the transmitting and receiving energies, i.e.,

$$E_{\mathrm{Cons}}=E_{\mathrm{tx}}(k,d)+E_{\mathrm{rx}}(k) \tag{A4}$$

Equations (A1) and (A2) are restricted to only the SN. If the SN acts as a CH, then it collects the information from all the member nodes and transmits it to the sink node. Therefore, the amount of energy consumed to transmit and receive the information by the CH is given by

$$E_{\mathrm{tx}}(k\times n_{\mathrm{cm}},d)=\\ \begin{cases}(k\times n_{\mathrm{cm}}(E_{\mathrm{el}}+E_{\mathrm{CD}})+\left(k\times E_{\mathrm{fs}}\times d_{\mathrm{ch}}^2\right), \text{ if } d_{\mathrm{ch}}\leqslant d_{\mathrm{Th}};\\ (k\times n_{\mathrm{cm}}(E_{\mathrm{el}}+E_{\mathrm{CD}})+\left(k\times E_{\mathrm{mp}}\times d_{\mathrm{ch}}^4\right), \text{ if } d_{\mathrm{ch}}>d_{\mathrm{Th}}\end{cases} \tag{A5}$$

Additionally, we have

$$E_{\mathrm{rx}}(k\times n_{\mathrm{cm}})=k\times n_{\mathrm{cm}}(E_{\mathrm{el}}+E_{\mathrm{CD}}) \tag{A6}$$

where n_{cm} denotes the total number of CM nodes, E_{CD} denotes the total energy consumed during data collection from all the CMs by the CH in a cluster, and d_{ch} denotes the Euclidean distance between the CH and sink node.

Lemma A1 The time complexity incurred at cluster formation in the worst case is given by $O(N)$, where N denotes the total number of IoT nodes in the network. Next, the worst-case complexity for exchanging messages during cluster formation is given by $O(1)$.

Proof In the proposed approach, nodes are clustered into several clusters only at the beginning, i.e., before data transmission. The cluster formation is performed with the assistance of the CR of the nodes and the radius (length or width) of the network. Clustering all nodes into hierarchical clusters is a one-time process. The information regarding the clustering is broadcasted by the sink node. Because the nodes are aware of their locations, they can determine their cluster on the basis of their CR. Hence, the message exchanging time is constant for all nodes, i.e., $O(1)$, and the worst-case time complexity of the entire network is given by $O(N)$. At the time of cluster formation, the nodes won exchange any information, and, hence, the worst-case message exchanging complexity is given by$O(1)$.

Lemma A2 The time complexity and message exchanging complexity incurred during the CH selection in the worst case are $O(n)$, where $n\ll N$ denotes the total number of IoT nodes in the layer.

Proof Let us consider that a cluster has n number of nodes. To become the CH, each node in that cluster must check the probability of being selected as the CH for $n-1$ nodes. Hence, the time complexity of the proposed CH selection approach is given by $O(n)$. After the computation of probabilities of each node in the cluster, the node with the maximum probability is selected as the CH. After being selected as the CH, the node broadcasts a message to all the other nodes in the cluster, conveying that it is the CH. Hence, two messages must be broadcast, one before the CH selection and another after it. Thus, the message broadcasting complexity is given by $O(1)$. Thus, the overall complexity due to the exchange of messages in each cluster is given by $O(n)$.

References

[1] T. Torfs, T. Sterken, S. Brebels, J. Santana, R. van den Hoven, V. Spiering, N. Bertsch, D. Trapani, and D. Zonta, Low power wireless sensor network for building monitoring, *IEEE Sens. J.*, vol. 13, no. 3, pp. 909–915, 2013.

[2] J. E. Siegel, S. Kumar, and S. E. Sarma, The future Internet of Things: Secure, efficient, and model-based, *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2386–2398, 2017.

[3] J. N. Al-Karaki and A. E. Kamal, Routing techniques in wireless sensor networks: a survey, *IEEE Wirel. Commun.*, vol. 11, no. 6, pp. 6–28, 2004.

[4] R. Yarinezhad, Reducing delay and prolonging the lifetime of wireless sensor network using efficient routing protocol based on mobile sink and virtual infrastructure, *Ad Hoc Netw.*, vol. 84, pp. 42–55, 2019.

[5] M. Wang, S. Wang, and B. Zhang, APTEEN routing protocol optimization in wireless sensor networks based on combination of genetic algorithms and fruit fly optimization algorithm, *Ad Hoc Netw.*, vol. 102, p. 102138, 2020.

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

[7] K. Guleria and A. K. Verma, Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks, *Wirel. Netw.*, vol. 25, no. 3, pp. 1159–1183, 2019.

[8] H. M. Hassan and R. Priyadarshini, Optimization techniques for IoT using adaptive clustering, in *Proc. 2021 Int. Conf. Computing, Communication, and Intelligent Systems* (*ICCCIS*), Greater Noida, India, 2021, pp. 766–771.

[9] K. Guleria and A. K. Verma, Meta-heuristic ant colony optimization based unequal clustering for wireless sensor network, *Wirel. Pers. Commun. Int. J.*, vol. 105, no. 3, pp. 891–911, 2019.

[10] U. Uyoata, J. Mwangama, and R. Adeogun, Relaying in the Internet of Things (IoT): A survey, *IEEE Access*, vol. 9, pp. 132675–132704, 2021.

[11] A. Chugh and S. Panda, Strengthening clustering through relay nodes in sensor networks, *Procedia Comput. Sci.*, vol. 132, pp. 689–695, 2018.

[12] M. Bukhsh, S. Abdullah, A. Rahman, M. N. Asghar, H. Arshad, and A. Alabdulatif, An energy-aware, highly available, and fault-tolerant method for reliable IoT systems, *IEEE Access*, vol. 9, pp. 145363–145381, 2021.

[13] J. Tang, B. Hao, and A. Sen, Relay node placement in large scale wireless sensor networks, *Comput. Commun.*, vol. 29, no. 4, pp. 490–501, 2006.

[14] F. Senel and M. Younis, Relay node placement in structurally damaged wireless sensor networks via triangular steiner tree approximation, *Comput. Commun.*, vol. 34, no. 16, pp. 1932–1941, 2011.

[15] R. Magán-Carrión, J. Camacho, P. García-Teodoro, E. F. Flushing, and G. A. Di Caro, A Dynamical Relay node placement solution for MANETs, *Comput. Commun.*, vol. 114, pp. 36–50, 2017.

[16] A. Shukla and S. Tripathi, A multi-tier based clustering framework for scalable and energy efficient WSN-assisted IoT network, *Wirel. Netw.*, vol. 26, no. 5, pp. 3471–3493, 2020.

[17] G. Pius Agbulu, G. Joselin Retna Kumar, and A. Vimala Juliet, A lifetime-enhancing cooperative data gathering and relaying algorithm for cluster-based wireless sensor networks, *Int. J. Distrib. Sens. Netw.*, vol. 16, no. 2, p. 155014771990011, 2020.

[18] P. Suman Prakash, D. Kavitha, and P. Chenna Reddy, Delay-aware relay node selection for cluster-based wireless sensor networks, *Meas. Sens.*, vol. 24, p. 100403, 2022.

[19] X. Luo, C. Zhang, and L. Bai, A fixed clustering protocol based on random relay strategy for EHWSN, *Digit. Commun. Netw.*, vol. 9, no. 1, pp. 90–100, 2023.

[20] K. A. Darabkh, N. J. Al-Maaitah, I. F. Jafar, and A. F. Khalifeh, EA-CRP: A novel energy-aware clustering and routing protocol in wireless sensor networks, *Comput. Electr. Eng.*, vol. 72, pp. 702–718, 2018.

[21] A. Shukla and S. Tripathi, An effective relay node selection technique for energy efficient WSN-assisted IoT, *Wirel. Pers. Commun.*, vol. 112, no. 4, pp. 2611–2641, 2020.

[22] K. Jaiswal and V. Anand, A Grey-Wolf based Optimized Clustering approach to improve QoS in wireless sensor networks for IoT applications, *Peer-to-Peer Netw. Appl.*, vol. 14, no. 4, pp. 1943–1962, 2021.

[23] H. Lin, L. Wang, and R. Kong, Energy efficient clustering protocol for large-scale sensor networks, *IEEE Sens. J.*, vol. 15, no. 12, pp. 7150–7160, 2015.

[24] P. K. Barik, C. Singhal, and R. Datta, An efficient data transmission scheme through 5G D2D-enabled relays in wireless sensor networks, *Comput. Commun.*, vol. 168, pp. 102–113, 2021.

[25] Y. Zhang, L. Liu, M. Wang, J. Wu, and H. Huang, An improved routing protocol for raw data collection in multihop wireless sensor networks, *Comput. Commun.*, vol. 188, no. C, pp. 66–80, 2022.

[26] N. Ashraf, S. A. Sheikh, S. Ahmad Khan, I. Shayea, and M. Jalal, Simultaneous wireless information and power transfer with cooperative relaying for next-generation wireless networks: A review, *IEEE Access*, vol. 9, pp. 71482–71504, 2021.

[27] M. Mao, N. Cao, Y. Chen, and Y. Zhou, Multi-hop relaying using energy harvesting, *IEEE Wirel. Commun. Lett.*, vol. 4, no. 5, pp. 565–568, 2015.

[28] E. Chen, M. Xia, D. B. da Costa, and S. Aïssa, Multi-hop cooperative relaying with energy harvesting from cochannel interferences, *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1199–1202, 2017.

[29] D. K. P. Asiedu, H. Lee, and K. J. Lee, Simultaneous wireless information and power transfer for decode-and-forward multihop relay systems in energy-constrained IoT networks, *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9413–9426, 2019.

[30] Q. Wu, X. Zhou, Q. Cao, and H. Fang, Multihop capability analysis in wireless information and power transfer multirelay cooperative networks, *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–12, 2018.

[31] S. Wang, L. Ma, and W. Wu, An optimal energy-efficient transmission design for SWIPT systems in WSNs, in *Proc. 2021 Int. Wireless Communications and Mobile Computing* (*IWCMC*), Harbin, China, 2021, pp. 664–668.

[32] S. Han, X. M. Liu, H. Y. Huang, F. Wang, and Y. H. Zhong, Research on energy-efficient routing algorithm based on SWIPT in multi-hop clustered WSN for 5G system, *EURASIP J. Wirel. Commun. Netw.*, vol. 2021, no. 1, pp. 1–26, 2021.

[33] S. B. Babruvhan and B. M. Thippeswamy, Simultaneous wireless information and power transfer with a cooperative relay in wireless sensor network, *Int. J. Emerg. Technol. Comput. Sci. Electron.*, vol. 23, No. 5, pp. 108–113, 2016.

[34] B. Pavani, L. N. Devi, and K. V. Subbareddy, Energy enhancement and efficient route selection mechanism using H-SWIPT for multi-hop IoT networks, *Intell. Converged Netw.*, vol. 3, no. 2, pp. 173–189, 2022.

[35] A. S. Raghuvanshi, S. Tiwari, R. Tripathi, and N. Kishor, Optimal number of clusters in wireless sensor networks: An FCM approach, in *Proc. 2010 Int. Conf. Computer and Communication Technology* (*ICCCT*), Allahabad, India, 2010, pp. 817–823.

[36] D. E. Gustafson and W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in *Proc. 1978 IEEE Conf. Decision and Control*, *17th Symp. Adaptive Processes*, San Diego, CA, USA, 2007, pp. 761–766.

[37] M. S. Yang, A survey of fuzzy clustering, *Math. Comput. Model. Int. J.*, vol. 18, no. 11, pp. 1–16, 1993.

[38] A. Panchal and R. K. Singh, EHCR-FCM: Energy efficient hierarchical clustering and routing using fuzzy C-means for wireless sensor networks, *Telecommun. Syst.*, vol. 76, no. 2, pp. 251–263, 2021.

[39] Q. Ni, Q. Pan, H. Du, C. Cao, and Y. Zhai, A novel cluster head selection algorithm based on fuzzy clustering and particle swarm optimization, *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 14, no. 1, pp. 76–84, 2017.

[40] S. M. M. H. Daneshvar, P. Alikhah Ahari Mohajer, and S. M. Mazinani, Energy-efficient routing in WSN: A centralized cluster-based approach via grey wolf optimizer, *IEEE Access*, vol. 7, pp. 170019–170031, 2019.

[41] D. C. Hoang, R. Kumar, and S. K. Panda, Fuzzy C-Means clustering protocol for Wireless Sensor Networks, in *Proc. 2010 IEEE Int. Symp. on Industrial Electronics*, Bari, Italy, 2010, pp. 3477–3482.

[42] N. Wang and H. Zhu, An energy efficient algrithm based on LEACH protocol, in *Proc. 2012 Int. Conf. Computer Science and Electronics Engineering*, Hangzhou, China, 2012, pp. 339–342.

[43] N. Kumar and D. P. Vidyarthi, A green routing algorithm for IoT-enabled software defined wireless sensor network, *IEEE Sens. J.*, vol. 18, no. 22, pp. 9449–9460, 2018.

B. Pavani received the BTech degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2009 and the MTech degree in embedded systems from Jawaharlal Nehru Technological University, Hyderabad, India in 2012. She obtained the PhD degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 2023. She worked as an assistant professor in various engineering colleges in Telangana from 2012 to 2017. She is doing a post-doctoral program in Technology Innovation Hub (TIH) for Internet of Things (IoT) and Internet of Everything (IoE) at Indian Institute of Technology (IIT), Bombay, India. Her current research interests include the Internet of Things, wireless sensor networks, embedded systems, and next-generation wireless communications.

K. Venkata Subbareddy received the BTech degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Anantapur, India in 2012 and the MTech degree in electronics and communication engineering from Jawaharlal Nehru Technological University, Hyderabad, India in 2015. He is currently pursuing the PhD degree in electronics and communication engineering at the University College of Engineering, Osmania University, Hyderabad, India. His current research interests include the Internet of Things, computer vision, and machine learning.

L. Nirmala Devi received the BEng, MEng, and PhD degrees in electronics and communication engineering from Osmania University, Hyderabad, India in 1997, 2005, and 2014, respectively. She is currently working as a professor and head of department (HoD) in the Department of Electronics and Communication Engineering, Osmania University. She has more than 20 years of teaching experience in subjects like signals and systems, digital signal processing, adaptive signal processing, neural networks, AI, ML, IoT and wireless communications, wireless sensor networks, data and computer communications, analog communication, digital communication, and research methodology for engineering. Her research interests include Ad-hoc networks, wireless communication, wireless sensor networks, IoT, and signal processing, machine learning, and AI. Currently, she is working on various research projects sponsored by the Ministry of Electronics and Information Technology (MeiTY), Government of India, New Delhi, Department of Science & Technology (DST) and UGC. Recently she received a grant of 1 crore to work on smart city projects from RUSA 2.0 sponsored by MHRD, Govt of INDIA.