Array 27 (2025) 100436

Contents lists available at ScienceDirect

Array

journal homepage: www.sciencedirect.com/journal/array

Improved K-means clustering and adaptive distance threshold for energy reduction in WSN-IoTs

Azamuddin Ab Rahman *, Sakib Iqram Hamim

Faculty of Computing, Universiti Malaysia Pahang Al-Sultan Abdullah, Malaysia

ARTICLE INFO

Keywords:
Energy
Improved k-means
Distance threshold
Clustering algorithm
Wireless sensor networks
Internet of things

ABSTRACT

The Internet of Things (IoTs) increasingly depends on Wireless Sensor Networks (WSNs) for real time data collection and communication. However, due to the limited battery capacity of sensor nodes, energy efficiency remains a critical challenge, especially since data transmission consumes the most energy. This study introduces an enhanced energy aware clustering approach that combines an improved K-Means algorithm with an adaptive distance threshold to optimize relay node selection and cluster formation. The method considers node proximity, residual energy, and overall network conditions to achieve balanced energy distribution across the network. The proposed approach was evaluated against established protocols including Hybrid Energy-Efficient Distributed Clustering (HEED), Threshold-Sensitive Energy-Efficient Sensor Network (TEEN), and previous versions of the Energy Efficient Cluster and Routing (EECR) protocol under three different deployment scenarios. Experimental results show that the enhanced EECR protocol reduces energy consumption by 5 % and significantly extends network lifetime, outperforming conventional techniques. The inclusion of adaptive distance thresholds proves effective in minimizing unnecessary energy drain and improving the reliability of data transmission. These results highlight the method's potential as a scalable and energy efficient solution for future IoT applications involving large scale sensor networks.

1. Introduction

The rapid expansion of IoTs applications has positioned WSNs as a fundamental technology for real-time data collection, processing, and transmission across interconnected systems. In such networks, energy efficiency remains a primary concern due to the limited battery life of sensor nodes, especially in remote or inaccessible environments. Clustering techniques have emerged as a key solution, where the selection of optimal Cluster Heads (CHs) is crucial to balancing network load and reducing energy consumption. However, conventional clustering methods often fall short in dynamic or dense network conditions, resulting in uneven energy distribution and reduced network lifetime. These challenges have led to increased research efforts to design smarter clustering algorithms that adapt to real-time conditions and maximize the operational lifespan of WSNs [1].

WSNs are often deployed in settings where replacing or recharging nodes is impractical, such as agricultural fields, industrial zones, or military applications [2]. To address the energy constraint, recent studies have explored adaptive CH selection strategies that consider residual energy, communication distance, and signal quality. Among the notable protocols, HEED and TEEN have been widely adopted for their energy-aware features. HEED utilizes residual energy and intra-cluster communication cost for CH election, while TEEN introduces threshold-based mechanisms to reduce transmission frequency. Despite their effectiveness, both approaches exhibit limitations in load balancing and long-term energy optimization, particularly in large-scale or heterogeneous networks [3].

Another key challenge in WSNs lies in the placement and number of relay nodes, which significantly impact both energy efficiency and communication reliability. Relay nodes serve as intermediaries to reduce direct transmissions to the base station, but their effectiveness depends heavily on optimal positioning and selection criteria. Poorly placed relays can lead to increased energy drain, particularly when nodes are far from the base station or when data paths are congested. Multi-hop communication helps mitigate energy loss, but if not carefully managed, it may also accelerate the energy depletion of intermediate nodes [4]. Thus, designing a mechanism that dynamically selects relay nodes based on energy levels, proximity, and centrality is essential for

* Corresponding author.
E-mail addresses: azamuddinrahman@umpsa.edu.my (A.A. Rahman), iqramhamim.cse.iiuc@gmail.com (S.I. Hamim).

https://doi.org/10.1016/j.array.2025.100436
Received 21 October 2024; Received in revised form 9 April 2025; Accepted 18 June 2025
Available online 18 June 2025

achieving balanced energy consumption and longer network life.

To address these gaps, this study introduces an improved K-Means clustering algorithm integrated with an adaptive distance threshold to enhance relay node selection and energy efficiency in WSNs. The proposed method takes into account node density, distance to the base station, and residual energy, resulting in more balanced cluster formations and smarter data routing. Unlike traditional static approaches, this adaptive model responds to changing network conditions, effectively extending the network lifetime and reducing node failures.

The key contributions in this study are as follows.

1. An improved K-Means algorithm optimizes centroid selection, ensuring balanced energy consumption across nodes. Optimal network lifetime is achieved with seven relay nodes in Tier-1 and twelve in Tier-2, reducing energy consumption by 5 %. Adjusting the distance threshold (15 m in Tier-1, 30 m in Tier-2) extends network lifetime by 27 % compared to EECR 2.
2. The EECR scheme prioritizes relay nodes based on centrality, proximity to the base station, and residual energy, thereby minimizing communication costs. The proposed path selection strategy also prevents node conflicts by selecting routes with the highest energy and shortest distance.
3. EECR demonstrates superior performance, extending network lifetime by 84 % over HEED and 63 % over TEEN. Evaluations based on First Node Dead (FND), Last Node Dead (LND), cluster count, and residual energy across multiple scenarios confirm its effectiveness under various network conditions.

The remainder of the paper is structured as follows. Section 2 discusses related works, while Section 3 presents the network model. Section 4 details the proposed improved k-means clustering technique, followed by an explanation of the adaptive distance threshold in Section 5. Section 6 discusses the energy consumption model, and Section 7 provides simulation results and analysis. Section 8 presents the experimental findings, while Section 9 concludes the study.

2. Related works

In WSNs, managing energy consumption remains one of the most critical challenges, as it directly influences how long a network can operate effectively. Most sensor nodes are powered by small batteries and are often deployed in environments where replacing or recharging them is not practical. Since different nodes perform different roles, from sensing to data forwarding, the amount of energy they consume can vary significantly. Among all activities, data transmission uses the most energy. The way data is routed through the network plays a key role here, especially since network conditions can change frequently due to node failures, mobility, or signal interference [5]. These changes demand frequent updates to routing paths, which increases energy use even further. As a result, factors like residual energy, signal strength, and communication range must be considered carefully to keep the network running smoothly.

The structure of the network itself also plays a major role in energy efficiency [6]. In most WSNs, sensor nodes send data either directly to a base station (BS) or through a series of other nodes such as cluster heads (CHs). When the BS is located far away from the sensors, the energy needed for communication increases sharply. Placing the BS more centrally can help reduce this burden [7]. Likewise, using a single hop to send data over long distances consumes more energy compared to multi hop communication, where data is relayed through nearby nodes [8]. However, using too many relay nodes or placing them inefficiently can result in duplicated transmissions and wasted energy [9]. This is why selecting the right number and position of relay nodes is so important for maintaining a balanced and energy aware network.

Cluster Based Routing (CBR) is a popular strategy that improves both energy efficiency and data handling in WSNs. This method involves two steps: selecting cluster heads and forming clusters. Sensor nodes are grouped into clusters, and each CH is responsible for collecting data from its group and forwarding it to the BS. Well known CBR protocols such as Low Energy Adaptive Clustering Hierarchy (LEACH) [10], HEED [11], Power Efficient Gathering in Sensor Information System (PEGASIS) [12], Clustering with Routing Protocol Design (CRPD) [13], and Energy Efficient Cluster Based Secure Routing (EECSR) [14] all aim to extend network lifetime by using energy more wisely. Studies have shown that using multi hop communication between sensor nodes and CHs is often more energy efficient than direct communication [8]. However, when CHs are located far from the BS, they need to transmit data with greater power, which can lead to faster energy depletion. To address this, threshold based relay selection can help reduce the load. Other improved protocols such as multi hop LEACH [15], TEEN [16], and Energy Efficient LEACH-C (EELEACH-C) [17] have demonstrated how effective relay strategies can further enhance energy performance.

In most deployments, sensor nodes are randomly scattered across the area, resulting in some being much closer to the BS than others. This uneven placement creates an energy hole problem, where nodes closer to the BS run out of energy faster because they handle more data forwarding [9]. CBR typically splits communication into three parts which from the node to the cluster head, from the cluster head to a relay node, and then from the relay to the BS. Breaking long transmissions into smaller steps helps conserve energy, since the power needed for transmission increases with distance. Still, many studies focus on the shortest physical path without accounting for the actual signal loss that occurs in wireless environments [18]. Since real world signal loss is not always linear, optimizing routing based solely on distance can be misleading. A better approach involves segmenting communication paths in a way that minimizes both distance and signal loss.

In practice, relay nodes are often placed in tough environments, and some end up too far from the BS. To keep energy usage efficient, these nodes should ideally be placed as close as possible to the BS, and only the minimum number necessary should be used. Some methods choose relay nodes based on how close they are to the BS using distance metrics. For example [19], used Euclidean distance to select optimal relay nodes, assuming the BS has accurate location data from each node, potentially using Global Positioning System (GPS) coordinates.

Recent studies have increasingly focused on addressing two longstanding challenges in WSNs which are uneven energy consumption and node failure [20]. introduced a clustering technique that aims to distribute the workload more evenly across the network. This helps to reduce the dependence on a small number of relay nodes and prevents premature depletion of any particular node. Their approach contributes to greater network stability and longer operational lifetime. In a similar effort [21], examined fault tolerance within CBR protocols. Acknowledging that sensor nodes are susceptible to environmental stress, battery depletion, and physical wear, their method involves a two-step process that monitors node status and reallocates failed nodes to neighbouring clusters. This avoids the need for complete reconfiguration of the network. Although the strategy improves overall resilience, it may occasionally lead to cluster heads being required to transmit over longer distances which could increase energy consumption in some areas of the network.

These challenges have inspired the development of more adaptive and intelligent clustering protocols. One of the most widely cited is HEED which selects cluster heads based on residual energy and communication cost [22]. Unlike earlier protocols such as LEACH that rely on random selection, HEED uses a more systematic approach that achieves better energy balancing particularly in heterogeneous networks. Despite these advantages, HEED still presents several limitations. The iterative process required for cluster head selection increases control messaging and introduces additional energy overhead [23]. Moreover, while HEED improves load distribution when compared to LEACH, it does not fully resolve the hotspot problem where cluster heads located near the base station experience faster energy drain due to increased data forwarding duties. Its static clustering model also makes it less flexible in environments with mobile nodes or variable network densities. These constraints suggest a continued need for clustering mechanisms that can dynamically adapt to changing network conditions.

In parallel to deterministic protocols like HEED, reactive methods such as TEEN offer another perspective for managing energy in WSNs especially in time sensitive applications. TEEN reduces unnecessary communication by allowing nodes to transmit only when a sensed value exceeds a preset threshold [16]. This significantly reduces energy consumption and prolongs network lifetime which is particularly useful for applications such as disaster response, industrial control, and military operations. However, the reliance on fixed thresholds means that some important data may be missed if changes remain below the reporting limit making TEEN unsuitable for applications requiring continuous monitoring such as environmental sensing. Furthermore, TEEN also employs static clustering which over time can lead to uneven energy depletion as certain cluster heads are tasked with more communication responsibilities. In response to these limitations, an enhanced version known as Adaptive TEEN (APTEEN) has been proposed. This protocol combines event driven reporting with periodic data transmission to strike a better balance between responsiveness and energy efficiency [24].

3. Network model

The communication of this network model is based on Time Division Multiple Access (TDMA) mode. In this network model, it consists of the BS, the sensor nodes, the relay nodes, and the CHs. The network model representation of EECR is based on as shown in Fig. 1.

The deployment of sensor nodes was determined based on the total number of nodes and the geographical dimensions of the network area. As an illustrative example, for a configuration involving 100 nodes, the distribution was allocated as 25 nodes in Tier-1 and 75 nodes in Tier-2. The communication model implemented within the network follows a multi-tier architecture. Tier-2 adopts a quad hop communication structure, while Tier-1 operates on a triple hop mechanism. In Tier-2, sensor nodes transmit data to their respective cluster heads. The data is then relayed through nodes within the same tier and subsequently forwarded to relay nodes located in Tier-1. This hierarchical forwarding continues until the data ultimately reaches the base station. In Tier-1, the process is relatively more direct, where sensor nodes send their data to both the cluster heads and nearby relay nodes, which then handle the final transmission to the base station. This tiered and hop based communication model is designed to optimize energy usage and ensure more efficient data delivery across the network.

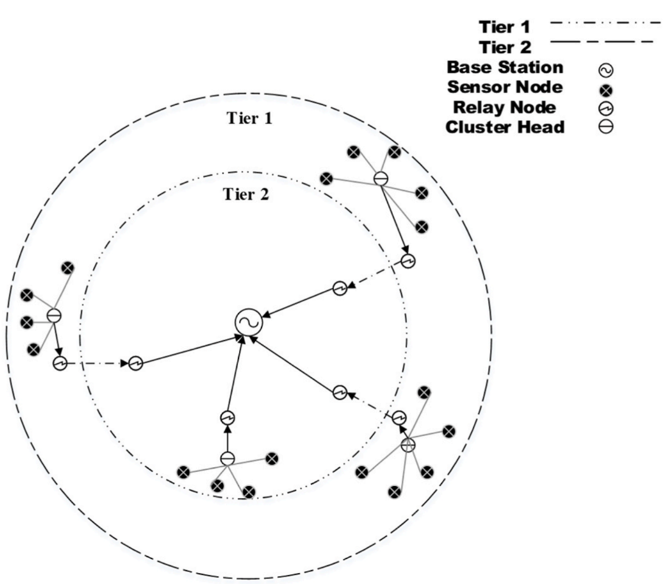


Fig. 1. The N-tier network Partition.

4. Improved K-means algorithm for clustering

The improved K-Mean is used for clustering where each cluster is defined and formed by its respective centre. In traditional K-Means clustering algorithm, some cluster could be overlapped to each other. Therefore, the improved K-Means can address such issues by assigning each cluster head to specific relay nodes in multi-tier environment. Moreover, the nodes are not forced to join the cluster but rather they will be assigned to join any clusters based on the probabilities between 0 and 1. The membership probabilities of the nodes at the edge maybe low so that it close to the centre of the cluster.

This gives more advantage to improved K-Means as opposed to traditional K-Means clustering algorithm. The improved K-Means is utilized for the CHs to form cluster with its child nodes. There are a few steps involved as follows.

1. Initial steps: The initial step is to identify and determine the point (i.e., count of relay nodes in Tier-1 and Tier-2).
2. Relay Node Selection: The selection to choose the relay nodes is build and based on K-Optimal approach as follows:

$$R = \left(\sqrt{\frac{I_i}{2\pi}}\right) * \sqrt{\left(\frac{\alpha_{fs}}{\beta_{mp}}\right)} * \frac{I}{d^2} \tag{1}$$

where I_i displays the number of the sensor nodes, I displays the number of nodes in Tier-1 and Tier-2, α_{fs} displays the amplification energy in the free space model; where the power loss is proportional to d^2, ε_{mp} is the amplification energy in the multi-path fading model where the power loss scale with d^4, and d is presents the mean distance among the nodes and BS. Therefore, it can determine the sets of $= \{r_{1,} r_2 \ldots.. r_n\}$. By considering the set of nodes $N = \{n_1, n_2 \ldots. n_k\}$, the cost function is measured as:

$$f(X, V, L) = \sum_{s=1}^{j} \sum_{t=1}^{s} m_{st} ||x_t - \partial_s||^2 \tag{2}$$

where $L = (\partial_s; s = 1, \ldots\ldots j)$ is the clusters' centroid and $X = (m_{st}; s = 1, \ldots\ldots j; t = 1, ..s)$ displays the probability value of X.

3. The m_{st} is the probability value of the tth node to the sth clusters. It m_{st} can be defined as:

$$m_{st} = \frac{\theta^{-\mu||x_t - \partial_s||^2}}{\sum_{k}^{j} \theta^{-\mu||x_t - \partial_s||^2}} \tag{3}$$

where θ represents the stiffness parameter, which effects the likelihood of node membership. An optimal clustering outcome is achieved by minimizing t. This approach deviates from traditional k-means by incorporating weighted squared errors into the cost function, as opposed to mere squared errors. The outcome of the improved k-means algorithm is contingent upon selected value of θ, which will be explored further in the discussion of simulation results.

4. To satisfy with the objective function in (2), the m_{st} need to meet with the following:

- Each node is assigned a probabilistic membership value, which ranges from 0 to 1, indicating its likelihood of association with a given cluster.

$$m_{st} \in [0, 1], s = 1, \ldots. j, k = 1, \ldots.. n \tag{4}$$

- The cumulative membership probabilities for a particular given node across all clusters adds to unity.

$$\sum_{s=1}^{j} m_{st} = 1, t = 1, \ldots..n. \tag{5}$$

- At least one node in each cluster should have a probability greater than "0" for being a part of that cluster.

$$\sum_{t}^{s} m_{st} > 0, \mathrm{s} = 1, \ldots, \mathrm{j} \tag{6}$$

- The coordinates of the cluster centres can be determined by optimizing the respective objective function, as:

$$\partial_s = \frac{\sum_{t=1}^{n} m_{st} x_t}{\sum_{t=1}^{n} m_{st}} \tag{7}$$

4.1. The influence of the stiffness parameter (θ) in the improved K-means algorithm

In the improved K-Means clustering algorithm, the stiffness parameter denoted as θ plays an important role in deciding how sensor nodes are grouped into clusters. This directly affects the accuracy of clustering and the energy efficiency of the network. Essentially, θ controls how "soft" or "strict" the cluster boundaries are. When tuned properly, it helps strike a balance: nodes are assigned in a way that avoids constant reassignments but still adapts well to the network structure. The ultimate goal is to reduce energy usage while keeping the network running longer.

The probability that a sensor node i belongs to a specific cluster j, written as P_{ij}, can be calculated using:

$$P_{ij} = \frac{\exp\left(-\theta \cdot d_{ij}^2\right)}{\sum_{k=1}^{C} \exp\left(-\theta \cdot d_{ik}^2\right)} \tag{8}$$

Here, d_{ij} is the distance between the node and the cluster center, and C is the total number of clusters. When θ is small, the probability values are more spread out. This means nodes can belong to more than one cluster with similar likelihood, which makes the clustering flexible especially for nodes near the border of two clusters. But if θ is too low, this flexibility can become a problem. Nodes may switch clusters too frequently, wasting energy.

On the other hand, when θ is large, the probability distribution becomes sharp. A node is very likely to belong to just one cluster. This reduces cluster switching and improves stability. However, this also means that nodes at the edge might be forced to join a distant cluster, which increases their communication energy.

This balance is shown in the clustering cost function:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{C} P_{ij}^{\theta} \cdot d_{ij}^2 \tag{9}$$

Where.

- J is the total clustering cost (a measure of how well the clustering configuration performs)
- N is the total number of sensor nodes in the network
- C is the total number of clusters
- P_{ij} is the probability that node i belongs to cluster j, based on the stiffness parameter
- θ id the stiffness parameter that controls the sharpness of the membership distribution
- d_{ij} is the Euclidean distance between node i and cluster centre j

To get the best results from the improved K-Means clustering, the stiffness parameter θ should not be fixed. Instead, it should be adjusted dynamically based on the current state of the network. Factors such as node density, remaining energy, and communication load can influence the ideal value of θ. For example, in dense areas where most nodes have high energy, a larger θ can improve clustering stability by reducing unnecessary reassignments. On the other hand, in sparse or low-energy zones, a smaller θ allows for more flexible clustering, helping to distribute the workload more evenly. This kind of adaptive tuning ensures that the clustering mechanism remains efficient under changing conditions, avoids excessive energy drain, and supports a longer, more stable network operation. The stiffness parameter θ may seem like a small setting, but it has a big impact. By adjusting it smartly, the improved K-Means algorithm can create cluster structures that are both efficient and adaptable, leading to better energy usage and a longer-lasting network. The improved K-Means clustering algorithm effectively handles outliers to ensure balanced energy consumption, but the stiffness parameter (θ) is critical in determining cluster assignments. An improper θ value may cause inefficient routing or frequent reassignments, leading to energy imbalances. To address this, the proposed method dynamically adjusts θ based on node density, residual energy, and transmission load, rather than using a fixed value. This adaptive tuning ensures optimal cluster formation, prevents energy drain, and maintains network stability.

5. Adaptive distance threshold

This section presents the adaptive distance threshold mechanism integrated into the (EECR protocol, which serves as a foundation for an enhanced energy management model in WSNs. Rather than relying on arbitrary relay node selection, as seen in conventional approaches, EECR introduces a more structured and data-driven method. The protocol determines threshold distances based on optimal node counts identified during cluster formation in both Tier-1 and Tier-2. Relay nodes are then selected through the K-Optimal process, but only if they fall within the specified threshold distance. This strategy ensures that node selection is not only proximity aware but also energy conscious, contributing to improved network balance and extended operational lifetime.

Empirical results suggest that the deployment of 7 relay nodes in Tier-1 and 12 in Tier-2 provides an optimal configuration. Introducing additional relay nodes beyond this configuration does not necessarily enhance performance and can, in fact, be counterproductive. An excessive number of relay nodes tends to increase redundant transmissions and leads to higher contention among nodes, ultimately resulting in greater energy consumption and reduced efficiency. Conversely, too few relay nodes can create communication bottlenecks, overburdening individual nodes and accelerating energy depletion. By applying the K-Optimal method dynamically, the network is able to adjust relay deployment in real time, promoting a balanced workload and preserving energy across the communication tiers.

To further enhance transmission efficiency, the relay nodes are deployed within tier specific distance thresholds that are adaptively determined. These thresholds are calculated based on the average distance between each node and its nearest neighbours, ensuring optimal coverage while minimizing communication range. Reducing the threshold below the recommended tier radius narrows each relay's coverage area and limits its effectiveness. Therefore, strategic placement within the computed threshold is essential to reduce the hop distance to the base station and avoid unnecessary energy expenditure. The decision making process for relay selection is detailed in Algorithm 1 for Tier-1 and Algorithm 2 for Tier-2, both of which illustrate the criteria used for selecting relay nodes based on proximity, centrality, and energy availability.

Algorithm 1: Relay Selection Proposed Algorithm in Tier-1

Input: i = tier, U_i = nodes, d = distance, d_1 = distance threshold
Initialize d_1 dynamically based on the centrality of the nodes
While ($i \leq 1$) **do**
  Update d_1 centrality of the nodes
  For (every nodes, U_i)
    If (U_i is alive) and ($d < d_1$)
      Selects the Relay Nodes using K-Optimum
      Connects Relay Nodes with CHs in Tier-1
    else
      Child nodes and CHs perform data transmission
    end if
end while

Algorithm 1 outlines the process for selecting relay nodes within Tier-1, excluding CHs. The K-Optimal formula serves as the foundation for this selection, ensuring an optimal number of relays. However, this selection is exclusively applied to nodes situated within a distance threshold. This threshold is established as the average distance separating each node from its nearest neighbours. Subsequently, the chosen relay nodes collaborate with CHs to form clusters, facilitating efficient data transmission. Centrality, a metric employed in the selection process, represents the cumulative distance between a node and its neighbours, normalized by the overall total number of nodes. The Euclidean distance formula, depicted as Equation (10), is utilized to calculate the distance between node i and node j, thereby contributing to the centrality value.

$$\mathrm{d}_{i,j} = \sqrt{\left(\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2\right)\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \tag{10}$$

Where $(x_i \text{ and } \mathrm{y}_i)$ represents the position of the node i and (x_j, y_j) represents the position of node j which is distributed in the network. Hence, based on Eq. (10), the centrality of node C_s can be defined as:

$$\text{Centrality}(C_s) = \frac{\left(\sum_{r=1}^{v} d_{i,j}\right)}{n} \tag{11}$$

Where k represents the count of nodes in the network, $d_{i,j}$ represents the distance between node i and node j. The $d_{i,j}$ is 0 when the node is a CH. Therefore, based on this value, it can be utilized to setup the far, along with the satisfactory and near centrality. Within this formula, v signifies the total number of nodes encompassed within the whole network. The term $d_{i,j}$ displays the distance separating node i from node j, as calculated using the Euclidean distance formula. Notably, the centrality value becomes 0 (Centrality(i) = 0) when the node under consideration is designated as a CH. This characteristic allows for the classification of nodes into three categories based on their centrality values: 'far,' 'satisfactory,' and 'near.'

Algorithm 2: Relay Selection Proposed Algorithm in Tier-2

Input: i = tier, U_i = nodes, d = distance, d_2 = distance threshold
Initialize d_2 dynamically based on the centrality of the nodes.
While ($i > 1$ && $i \leq 2$) **do**
  Update d_2 centrality of the nodes.
  For (every nodes, U_i)
    If (U_i is alive) and ($d < d_2$)
      Select (Relay Nodes) using K-Optimum
      Connects the Relay Nodes to the CHs- in Tier-2
    else
      Child nodes and CHs perform data transmission
    end if
end while

Similarly, Algorithm 2 shows the same method of relay node selection in Tier-1. The relay nodes in Tier-2 formed clusters with CHs in Tier-2 for data transmission to BS. The quad-hop and triple-hop transmission are utilized to obtain the shortest path to the BS. A routing table is regularly updated to maintain current paths with a small number of forwarding neighbour nodes. The total count of established nodes, the path of the nodes, and the average distance to the BS are taken into considerations to provide more accurate energy estimation. Let S_q is a square area. d_r is an average distance between nodes and relay nodes and d_{BS} is an average distance between nodes and BS. It can be calculated as:

$$d_r = \int_{s_q}^{d_{BS}} \sqrt{x^2 + y^2 \frac{1}{s_q} ds_q} \tag{12}$$

$$d_{BS} = \iint \left(\sqrt{x^2 + y^2}\right)\beta(x,y)dxdy \tag{13}$$

d_r and d_{BS} can be utilized to calculate the energy dissipated per round as:

$$\gamma_r = C^*\left(2T^*\varepsilon_e + T^*\varepsilon_d + v + \varepsilon_m{}^* d_r + T^*\varepsilon_f{}^* d_{BS}\right) \tag{14}$$

Where C is the number of clusters formed, v represents the number of bits, ε_d represents the data aggregation of the relay node, ε_m displays the amplifier energy of multi-path fading, and ε_f represents the amplifier energy of free space model. The energy of the nodes in Tier-1 (γ_1) and Tier-2 (γ_2) can be measured as:

$$\gamma_1 = \gamma_s(f) + \gamma_r(g) + \gamma_s(g)$$

$$= vE_e + v\varepsilon_m d(f,g)^2 + v\varepsilon_e + vE_e + v\varepsilon_m d(g,h)^2$$

$$= 3vE_e + v\varepsilon_m d\left[d(f,g)^2 + d(g,h)\right]^2 \tag{15}$$

$$\gamma_2 = \gamma_s(e) + \gamma_r(f) + \gamma_s(f) + \gamma_s(g) + \gamma_r(h)$$

$$= vE_e + v\varepsilon_m d(e,f)^2 + kE_{elect} + kE_{elect} + k\varepsilon_{mp} d(f,g)^2 + kE_{elect} + k\varepsilon_{mp} d(g,h)^2$$

$$\left[d(i,j)^2 + d(j,k)\right]^2 + d(k,l)\Big]^2 \tag{16}$$

The computation of the next hop for Tier-1 and Tier-2 is based on the weight function as shown in equation (17) and equation (18) as follows:

$$H(f,g) = \left[\frac{v(f).E}{v(f).max}\right] + \frac{\left[d(f,g)^2 + d(g,h)^2\right]}{d(f,h)^2} \tag{17}$$

$$W(e,h) = \left[\frac{v(e).E}{v(e).max}\right] + \frac{\left[d(e,f)^2 + d(f,g)^2 + d(g,h)^2\right]}{d(e,h)^2} \tag{18}$$

Here, *k(i)*.max depicts the initial energy of relay node The *v.(f).E* and *v.(e)E* are the remaining energy of the relay nodes in Tier-1 and of Tier-2 respectively. The weight function is used to calculate and determine the relay nodes with the closest distance with source and highest remaining energy.

6. Energy consumption model

In WSNs, most energy is used when sending data. Unlike wired networks, where only the sender uses energy, both the sending and receiving nodes use energy in WSNs. This is because radio communication, used for data transmission in WSNs, consumes energy on both sides, as shown in Fig. 2.

This model for energy consumption during data transmission in EECR comes from Ref. [25]. In their model, the energy used by transmitter nodes is heavily influenced by both the size (v bits) of the data packet being sent and also including the distance it travels. The formula to calculate this energy consumption as follows:

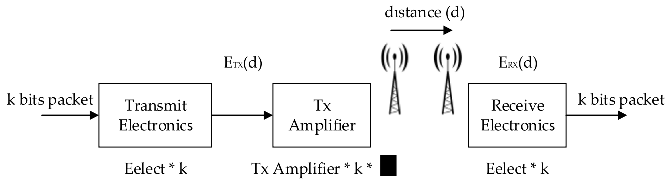


Fig. 2. Energy consumption model.

$$C_m(v.d) = \begin{cases} v^*E_e + v^*\varepsilon_e{}^*d^2, d \leq d_o \\ v^*E_e + v^*\varepsilon_m{}^*d^4, d \geq d_o \end{cases} \tag{19}$$

Where.

- C_{tx} is the total transmission energy cost,
- v is the data packet size in bits,
- E_e is the electronics energy per bit (fixed at 50 nJ/bit),
- ε_f is the free-space amplifier energy coefficient,
- ε_m is the multipath fading amplifier energy coefficient,
- d is the distance between transmitter and receiver,
- d_o is the threshold distance, calculated as $d_o = \sqrt{\frac{\varepsilon_f}{\varepsilon_m}}$

The formula includes several factors affecting energy use. C_m represents the total transmission cost. Distance d between transmitter and receiver is factored in. There's also an energy term E_e representing energy lost per data bit v-bits due to aspects like modulation and signal processing. To simplify calculations, a fixed value of 50 nJ/bit was used for E_e. The formula also considers amplifier energy consumption based on distance. When the distance falls lesser than a specific threshold, a fixed energy consumption per bit, ε_e applies. The formula then uses a different fixed energy consumption per bit, ε_m for distances exceeding the threshold. The value d_o is calculated as $\sqrt{\varepsilon_f/\varepsilon_m}$. Upon data reception, energy consumption was computed as:

$$C_X = v^*E_e \tag{20}$$

In WSN simulations, factors like distance between nodes, transmitter power, and number of packets affect total energy use. However, most simulations ignore the receiver's power consumption, which is needed for signal quality measurement. This can lead to inaccurate energy estimates. The text proposes using Received Signal Strength Indicator (RSSI) as an alternative for measuring signal strength, addressing this shortcoming. Additionally, it highlights that CHs in WSNs use more energy than regular nodes due to tasks like data aggregation. The energy of CH loses during transmission is calculated as:

$$C_{ch} = \left(\frac{K}{W}-1\right)^* v^* E_e + \frac{K}{W}{}^* k^* E_a + v^* E_e + v^* \varepsilon_f{}^* d^2 \tag{21}$$

In this equation.

- K is the total number of deployed sensor nodes,
- W is the number of clusters,
- d is the average distance between CHs and the BS,
- k is the number of signals being aggregated,
- E_a is the energy required for data aggregation, typically set to 5 nJ/bit/signal,
- ε_f is the free-space amplifier energy coefficient.

Where W represents the number of clusters, the d displays the average distance between CHs and the BS. The term E_a captures the energy used for the data aggregation, that is a constant value of 5nJ/bit/signal. Finally, total number of sensor nodes established/deployed in the network are presented as K.

7. Simulation results and analysis

This section validates the efficiency of our proposed EECR model against the existing methods (i.e., HEED and TEEN). Final results were examined and evaluated under three scenarios, including also the number of the nodes (i.e., 100, 200 and 800 nodes) which are shown in Fig. 3. The BS is positioned at the centre of the network to give fair transmission among the nodes.

7.1. Simulation settings

Within the simulated environment, sensor nodes were emulated as stationary wireless devices deployed in two network configurations which are Tier-1 and a Tier-2 architecture. An initial set of experiments employed 100 sensor nodes, a common practice for preliminary investigations *(X, Y)*. However, to evaluate the system's scalability and analyze energy consumption across varying network sizes, the count number of sensor nodes was subsequently raised to 800. Our simulation is based on the first-order radio model, which is widely used in WSN studies. Table 1 provides a detailed breakdown of the key parameters used in our simulation experiments.

Additionally, the simulation environment employs quad-hop communication for Tier-2 and triple-hop communication for Tier-1, allowing for efficient multi-hop data transmission towards the BS. The relay node selection process is optimized based on the K-Optimal approach, ensuring balanced energy distribution and efficient routing paths.

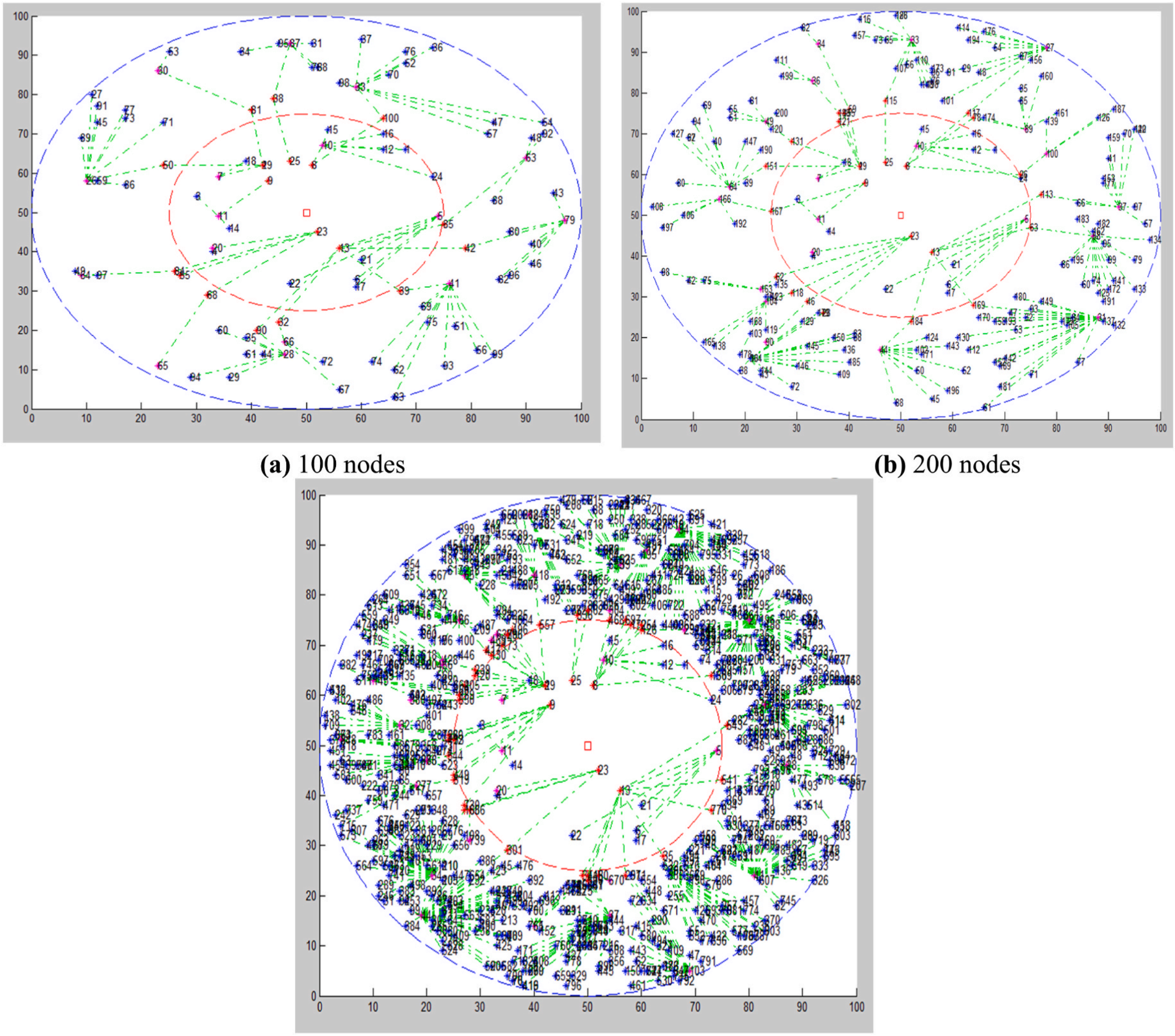
The effectiveness of the EECR protocol was evaluated through computer simulations, focusing on network performance. Its reliability was further assessed by comparing it to EECR 2 (which omits the distance threshold implementation) and the HEED and TEEN protocols. The evaluation considered three key performance metrics: FND, which represents the time elapsed before the first sensor node depletes its energy reserves, LND, which indicates when the final node becomes inoperable, and Standard Deviation of Residual Energy (SDRE), which measures the distribution of remaining energy per node across each simulation iteration. These metrics provide a comprehensive analysis of the protocol's efficiency and its impact on network longevity.

8. Experimental results

8.1. Improved K-means analysis

This section evaluates the energy balancing capabilities of the EECR 2 protocol by incorporating the K-Means clustering approach. The analysis considers two network sizes, comprising 100 and 200 sensor nodes, to represent typical and high-density deployment scenarios, respectively. To assess the uniformity of energy distribution across the network, the SDRE is employed as the primary evaluation metric. A lower SDRE value indicates a more balanced energy consumption pattern among the nodes, which is essential for extending the overall network lifetime and minimizing the risk of premature node failures. By analysing the SDRE under different network densities, this evaluation provides insight into how effectively EECR 2 manages energy balance when integrated with the K-Means clustering strategy.

Figs. 4–6 shows the SDRE of EECR 2 with and without K-Means approach with 100, 200 and 800 nodes. As shown in Fig. 4, the EECR 2 with K-Means outperforms EECR 2 without K-Means with low value of SDRE. At about the 3000th iteration, there was sharp decrease of SDRE for EECR 2 with K-Means of 100 nodes. The value of SDRE decreased from 0.0346 to 0.0089. The SDRE of EECR with K-Means is nearly 94 % more efficient in terms of energy balance than EECR without K-Means at the 5000th iteration. On the other hand, the SDRE for EECR 2 without K-Means of 100 nodes showed a steady increase until the 5000th iteration. Similarly, it showed a steady increase in SDRE for EECR 2 without K-Means when the nodes increased to 200 nodes. For EECR 2 with K-



(a) 100 nodes **(b)** 200 nodes

(c) 800 Nodes

Fig. 3. The Node density.

Table 1
The simulation parameters settings.

Parameter	Value
Number of sensor nodes	100, 200, 800
Network area	100m × 100m
Base station position	Centre (50,50)
Initial node energy	1 J, 2 J
Tier-1 radius	25m
Tier-2 radius	50m
Distance threshold (Tier-1)	15m
Distance threshold (Tier-2)	30m
*E*elect	5 nJ/bit/signal

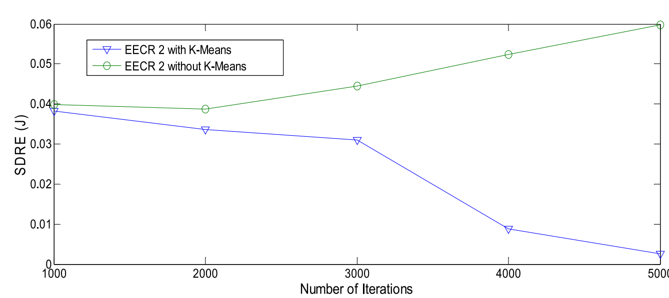


Fig. 4. The SDRE with 100 nodes.

Means, the same trends with 100 nodes were shown. There was a sharp decrease at the 4000th to 5000th iterations. It can be compared that the value of SDRE in EECR with K-Means with 200 nodes was lower than EECR without K-Means. In contrast, the SDRE of EECR without K-Means with 200 nodes was higher as compared to 100 nodes. In fact, the graphs showed opposite trends between them. For 800 nodes, the SDRE pattern follows a similar trend to what we saw with 100 and 200 nodes. The EECR 2 with K-Means consistently performs better, maintaining a lower SDRE value, which means it balances energy more efficiently. In contrast, EECR 2 without K-Means shows a steady increase in SDRE as the iterations go up, indicating a decline in energy efficiency. By the

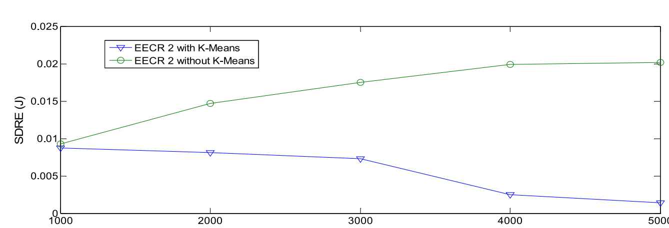


Fig. 5. The SDRE with 200 nodes.

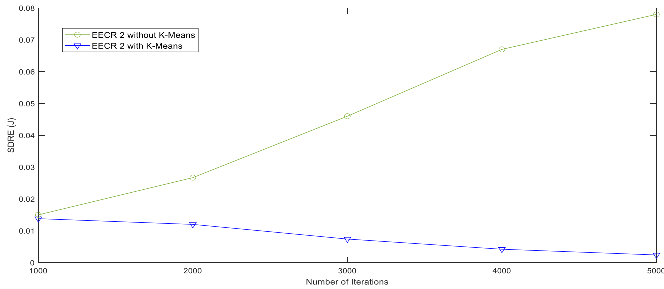


Fig. 6. The SDRE with 800 nodes.

time it reaches 5000 iterations, the version with K-Means is clearly outperforming the one without it. This reinforces the idea that using K-Means in EECR 2 significantly improves energy distribution, especially as the number of nodes grows.

For a network with 800 nodes, EECR 2 with K-Means maintains its advantage over the non-K-Means by ensuring better energy distribution and lower energy imbalance. However, as the network grows, challenges such as increased data traffic, relay node congestion, and higher energy consumption will become more pronounced. While EECR 2 with K-Means has already shown significant improvements in energy balance for 100 and 200 nodes, the efficiency gap may slightly narrow at larger scales due to the higher workload on relay nodes. Without K-Means, energy distribution is likely to become even more uneven, leading to rapid energy depletion in some nodes while others remain underutilized. To sustain performance at higher node densities, an adaptive relay node placement strategy will be crucial, ensuring that energy usage is evenly spread across the network. Additionally, multi-hop routing could further optimize energy consumption by preventing excessive strain on relay nodes and reducing long-distance transmissions. Overall, while EECR 2 with K-Means remains the more energy-efficient and stable approach, fine-tuning relay placement and clustering mechanisms will be essential to maximize network lifetime and maintain efficiency as the network scales up.

The results showed that EECR 2 with improved K-Means performed better than EECR 2 without K-Means. The K-Means approach in EECR was more energy balanced as compared to the other ones due to the following reasons. The optimal number of relay nodes acquired by utilizing the K-Optimal in EECR 2 gave the perfect value of θ in improved K-Means in order to produce similar probabilities among the cluster. The unsuitable value of k in K-Means resulted in high variance of the distance among relay nodes and CHs. Moreover, it can be noted that when the number of k was unsuitable (i.e. non-optimal number of relay nodes), high variation was observed, which meant the CHs might pass data to farther relay nodes. Finally, EECR 2 with K-Means was sensitive to outlier nodes. The relay node was centralised among CHs to balance energy consumption over relay paths to BS. In other words, high variance of residual energy among the nodes indicated instability of the network. This was due to unbalanced data transmission among the nodes. When unbalanced communication occurs, some nodes would have more burden in receiving and transmitting the data. This led to quick energy depletion and hence consumed high energy. This experiment corroborated that the K-Means approach in EECR provided energy balance in reducing energy consumption with small and high count of nodes in the network. The network lifetime of EECR 2 could be significantly improved using distance threshold for relay node placement.

8.2. Network lifetime of EECR 2

The current investigation focuses on the network lifetime and also on the energy balance of the EECR 2 protocol. This evaluation aims to elucidate the impact of varying count of relay nodes and the K-Means clustering approach employed within EECR 2 on these crucial network performance metrics. In this experiment, the number of relay nodes in (Tier 1 as well as in Tier 2) were varied as: (1,1), (1,3), (1,7), (1,15), (3,1), (3,3), (3,7), (3,12), (3,15), (7,1), (7,3), (7,7), (7,12), (7,15), (12,1), (12,3), (12,7), (12,12), (12,15), (15,1), (15,3), (15,7), (15,12), and (15,15) to investigate their network lifetime.

The overall lifespan of a WSNs is closely influenced by the number and arrangement of relay nodes within its architecture. As illustrated in Fig. 7, network lifetime varies significantly across different configurations of relay node deployment. The configuration involving seven relay nodes in Tier-1 and twelve in Tier-2 resulted in the highest recorded network lifetime of 6131 rounds, outperforming all other tested scenarios. In contrast, the shortest network lifetime of 5420 rounds was observed when only one relay node was deployed in each tier. Notably, across all configurations within Tier-1, the presence of seven relay nodes consistently yielded higher network longevity. For instance, when combined with varying relay node counts in Tier-2, the following lifespans were achieved: 5521 (7,1), 5625 (7,3), 5823 (7,7), 6491 (7,12), and 6002 (7,15). These results clearly indicate that the (7,12) setup offers optimal performance, with configurations such as (7,7) and (7,15) following closely behind. Interestingly, increasing the number of relay nodes in Tier-1 beyond seven—such as to 12 or 15 led to a gradual decline in network lifetime, regardless of the Tier-2 configuration.

This trend underscores the importance of strategic relay node deployment, which in this study is guided by the K-Optimal methodology. This method takes into account both the total number of deployed nodes and their average distances to the base station, enabling the identification of an optimal relay configuration for both tiers. The integration of the K-Optimal approach into the EECR 2 protocol provides a systematic way to balance energy consumption and minimize overhead. Complementing this, the use of the K-Means clustering algorithm plays a critical role in managing data transmission between cluster heads and relay nodes. Together, these two techniques which K-Optimal relay selection and K-Means based clustering form a cohesive strategy for enhancing the energy efficiency and sustainability of the network.

On the other hand, deploying an excessive number of relay nodes, as seen in configurations such as (15,7), (15,12), and (15,15), does not guarantee improved performance. Although a 2-Tier architecture inherently requires a considerable number of relay nodes, simply increasing their count can lead to diminishing returns. The energy consumption per node rises proportionally with the total number of active relays, resulting in inefficiencies and unnecessary overhead. In such cases, excessive relay deployment can lead to underutilization and energy wastage, while also accelerating battery depletion in nodes tasked with frequent communication. Conversely, configurations with too few relay nodes fail to distribute the workload evenly, placing disproportionate demands on individual nodes. Taken together, the findings confirm that the combination of seven relay nodes in Tier-1 and twelve in Tier-2, paired with the K-Means clustering strategy, represents the most energy efficient and balanced setup for prolonging network lifetime under the EECR 2 framework.

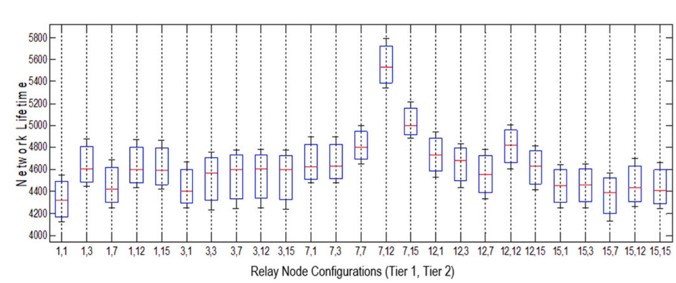


Fig. 7. Network lifetimes with Different Number of RNs.

8.3. Adaptive distance threshold (EECR 3)

This section investigates the effect of the relay placement using adaptive distance thresholds. The distance threshold for Tier-1 and Tier-2 were set as (5m,5m), (15m,30m), (20m,40m), (25m,45m) and (25m,50m) as used by Refs. [26,27] for distance evaluation. Fig. 8 shows the results of network lifetime for varied distance threshold.

It can be observed that a distance threshold between a sensor node and the BS has a significant impacts on network lifetime. The longest network life (7029) was achieved with a threshold of 15 m–30 m. As the threshold increased, the network lifetime decreased: 6576 for 20m–40m and 6718 for 25m–45m. Without a threshold (25m–50m), the lifetime dropped to 6106. The shortest lifetime (5432) occurred when nodes were closest to the BS (5m–5m). Interestingly, a wider threshold (25m–45m) performed better than a narrower one (20m–40m).

This study revealed that placing relay nodes very close (close vicinity) to the BS was not the most effective strategy for EECR 3. The closest threshold (5m–5m) resulted in the worst performance due to a phenomenon called the hotspot problem. The best network lifetime was achieved with a threshold of 15 m–30 m for a couple of reasons. First, keeping the BS and CHs close together significantly improved network lifetime in EECR 3. This is because the energy used for communication between two nodes increases with the square of the distance between them. So, as the threshold distance increased, the network lifetime decreased, except in the case of the closest threshold (5m–5m). Second, the threshold of 15m–30m allowed for an optimal number of relay nodes, which balanced energy consumption and workload distribution. Therefore, using a threshold of 15m–30m for relay node selection in EECR 3 was generally beneficial.

The threshold of 25m–45m achieved a better network lifetime than 20m–40m because it was closer to the ideal tier radius in EECR 3 (25m–50m). The 20m–40m threshold resulted in uneven load distribution within the transmission radius, resulting to a shorter network lifetime as compared to the 25m–45m threshold. Overall, a threshold of 15m–30m is recommended for EECR 3 to minimize the distance of communication between relay nodes and BS.

8.4. Performance evaluation of HEED, TEEN, EECR 2 and EECR 3

In this section, EECR 2, EECR 3, TEEN, and HEED were evaluated based on network lifetime, the number of dead nodes, and residual energy. The primary objective of this experiment was to compare the energy consumption of nodes among EECR 2, EECR 3, TEEN, and HEED.

Figs. 9 and 10 illustrate performance enhancements in terms of FND and LND for EECR 2, EECR 3, TEEN, and HEED protocols. Across all three scenarios (i.e., 1, 2, and 3), EECR 3 achieved the latest FND, with iterations reaching 4,098th, 4,123rd, and 4,734th, respectively. Furthermore, EECR 3 demonstrated a network lifetime improvement of 2 %, 3 %, and 1 % compared to EECR 2, TEEN, and HEED based on the FND metric in Scenarios 1, 2, and 3, respectively. Similar trends were observed for EECR 2, where FND occurred at the 4,005th, 4,018th, and 4,695th iterations. Meanwhile, TEEN achieved FND at the 3,711th, 3,854th, and 4,600th iterations, while HEED exhibited earlier FND at 3698th, 3814th, and 4554th iterations, respectively. These results suggest that EECR 3 outperformed all other protocols in terms of FND, demonstrating superior energy efficiency and network stability.

When evaluating LND as a network lifetime metric, EECR 3 once again emerged as the most efficient protocol, achieving the longest network lifetimes of 7,032nd, 7,283rd, and 12,542nd iterations in Fig. 11 as Scenario 1, Fig. 12 as Scenario 2, and Fig. 13 as Scenario 3, respectively. Compared to EECR 2, TEEN, and HEED, EECR 3 exhibited improvements of 21 %, 40 %, and 44 % in Scenario 1, 21 %, 34 %, and 44 % in Scenario 2, and 12 %, 63 %, and 84 % in Scenario 3, respectively. These findings confirm that EECR 3 provides a more energy-efficient clustering mechanism, significantly enhancing network lifetime and reducing premature node depletion compared to TEEN and

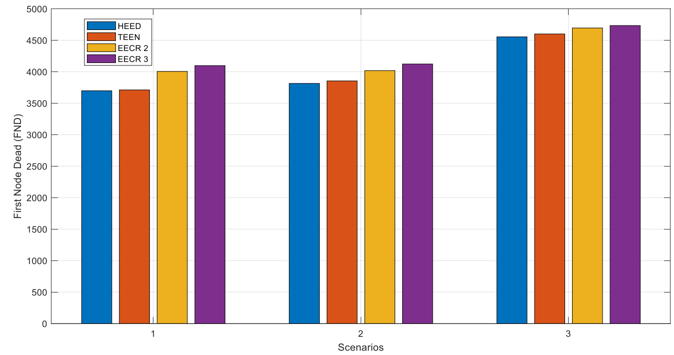


Fig. 9. Fnd of EECR 2, EECR 3, TEEN, and HEED.

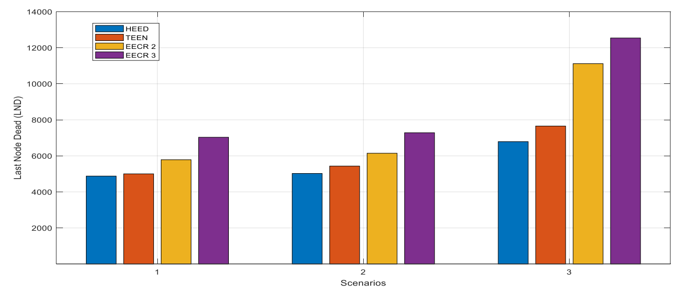


Fig. 10. Lnd of EECR 2, EECR 3, TEEN, and HEED.

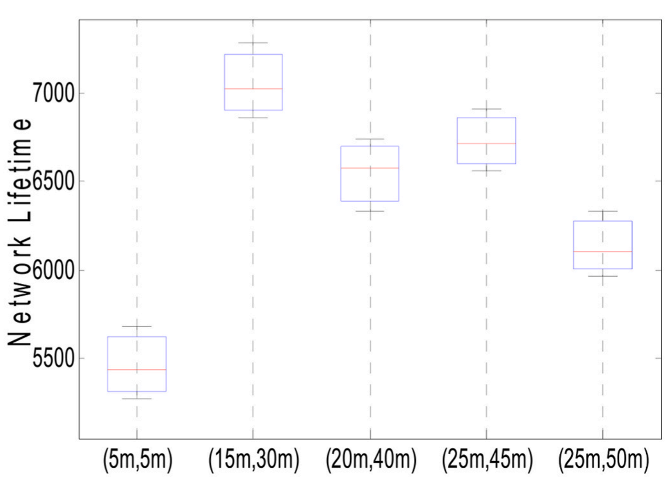


Fig. 8. Network lifetime of various distance threshold.

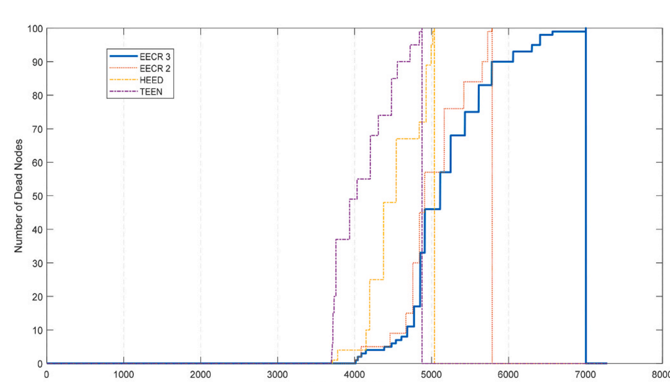


Fig. 11. Number of dead nodes for Scenario 1.

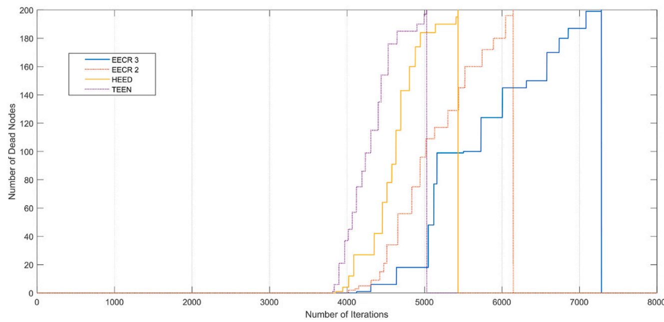


Fig. 12. Number of dead nodes for Scenario 2.

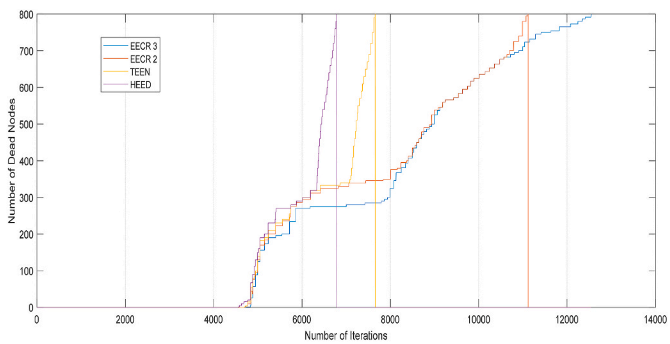


Fig. 13. Number of dead nodes for Scenario 3.

HEED.

The performance analysis of HEED, TEEN, EECR 2, and EECR 3 highlights the clear advantage of the EECR 3 protocol in enhancing network lifetime and energy efficiency. Across all three scenarios, EECR 3 consistently achieved the latest occurrence of FND and LND. For example, in Scenario 1, the FND for EECR 3 occurred at iteration 4098, outperforming EECR 2 at 4005, TEEN at 3711, and HEED at 3698. This upward trend continued in Scenario 3, where EECR 3 reached the LND at iteration 12,542, significantly higher than EECR 2 at 11,118, TEEN at 7654, and HEED at 6789. These findings suggest that EECR 3 is more effective in delaying node depletion and sustaining overall network activity.

A closer examination of the number of dead nodes over time further reinforces the superiority of EECR 3. The results demonstrate that HEED experienced the fastest increase in node deaths, indicating unbalanced energy usage and rapid depletion. TEEN showed a more gradual trend, yet still lagged behind in sustaining node activity. EECR 2 performed better by integrating enhanced clustering, but it was EECR 3 that showed the slowest rate of node failures throughout all three scenarios. This improvement can be attributed to the combination of adaptive relay node placement and intelligent clustering using improved K-Means. By optimizing the placement of nodes based on residual energy and proximity to the base station, EECR 3 successfully maintained energy balance and extended node survival.

Overall, the results indicate that EECR 3 offers a more reliable and energy-conscious solution for WSNs. Its ability to dynamically adapt to network conditions through distance threshold tuning and optimal relay node selection ensures efficient communication and prolonged network lifetime. While EECR 2 also demonstrates commendable performance, it lacks the adaptive thresholding feature that provides EECR 3 with its distinct advantage. Comparatively, TEEN and HEED are less effective in managing energy distribution over time. Therefore, EECR 3 stands out as a robust protocol for real-world sensor deployments where sustainability and energy efficiency are essential.

9. Conclusion

This research introduces two enhanced techniques under the EECR protocol, aimed at improving energy efficiency and network longevity in WSNs. The first enhancement involves an improved K-Means clustering algorithm, which employs probabilistic membership values to achieve balanced cluster formation and reduce the frequency of re-clustering. The second is an adaptive distance threshold mechanism that dynamically selects optimal relay nodes based on node proximity and residual energy, effectively lowering communication costs and enhancing energy distribution. Simulation results demonstrated that EECR significantly outperformed other protocols such as HEED and TEEN, particularly in terms of delaying node depletion and extending network lifetime. By maintaining a slower increase in the number of dead nodes and achieving higher values for both FND and LND metrics, EECR has proven to be a more sustainable and energy conscious protocol. Future research could explore incorporating real-time network dynamics, mobility patterns, and environmental factors into the relay node selection process to further optimize performance. These findings contribute meaningfully to the development of scalable, long-lasting IoT communication systems, where energy conservation and operational stability are critical for real-world deployment.

CRediT authorship contribution statement

Azamuddin Ab Rahman: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Sakib Iqram Hamim:** Writing – review & editing, Visualization, Resources, Investigation, Formal analysis, Conceptualization.

Disclosure of any funding to the study

This work was supported by the Fundamental Research Grant Scheme (FRGS) [grant numbers RDU230128].

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Azamuddin Bin Ab Rahman reports financial support was provided by Malaysia Ministry of Higher Education. Azamuddin Bin Ab Rahman reports a relationship with University of Malaysia Pahang Al-Sultan Abdullah that includes: employment. Sakib Iqram Hamim reports a relationship with University of Malaysia Pahang Al-Sultan Abdullah that includes: non-financial support. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) for providing the financial support under Fundamental Research Grant Scheme (FRGS Grant (RDU230128)). In addition, we would like to acknowledge and pay our gratitude to the Faculty of Computing (UMPSA) and specifically our Computer Network Research Group (CNRG) group for their support and for their facilities.

Data availability

Data will be made available on request.

References

[1] Begum BA, Nandury SV. Data aggregation protocols for WSN and IoT applications – a comprehensive survey. J King Saud Univ Comput Inf Sci Feb. 2023;35(2):651–81. https://doi.org/10.1016/j.jksuci.2023.01.008.
[2] Fahmy HMA. WSNs applications. 2023. p. 67–242. https://doi.org/10.1007/978-3-031-20709-9_3.
[3] Rahman A Ab. RSSI-guided cluster head selection for optimal optimization in IoT-enabled WSNs. In: 2023 IEEE 8th international conference on software engineering and computer systems (ICSECS). IEEE; Aug. 2023. p. 195–9. https://doi.org/10.1109/ICSECS58457.2023.10256366.
[4] Yang X, Yan J, Wang D, Xu Y, Hua G. THSI-RP: a two-tier hybrid swarm intelligence based node clustering and multi-hop routing protocol optimization for wireless sensor networks. Ad Hoc Netw Oct. 2023;149:103255. https://doi.org/10.1016/j.adhoc.2023.103255.
[5] Zou Y, et al. A survey of fault tolerant consensus in wireless networks. High-Confidence Computing Jun. 2024;4(2):100202. https://doi.org/10.1016/j.hcc.2024.100202.
[6] Hamim SI, Rahman ABA. Optimizing wireless sensor networks: a survey of clustering strategies and algorithms. International Journal of Computer Networks and Applications Oct. 2024;11(5):673–89. https://doi.org/10.22247/ijcna/2024/42.
[7] Hada RPS, Srivastava A. Dynamic cluster head selection in WSN. ACM Trans Embed Comput Syst Jul. 2024;23(4):1–27. https://doi.org/10.1145/3665867.
[8] Krishna KPR, Thirumuru R. Energy efficient and multi-hop routing for constrained wireless sensor networks. Sustainable Computing: Informatics and Systems Apr. 2023;38:100866. https://doi.org/10.1016/j.suscom.2023.100866.
[9] Ullah Khan S, Ulalh Khan Z, Alkhowaiter M, Khan J, Ullah S. Energy-efficient routing protocols for UWSNs: a comprehensive review of taxonomy, challenges, opportunities, future research directions, and machine learning perspectives. J King Saud Univ Comput Inf Sci Sep. 2024;36(7):102128. https://doi.org/10.1016/j.jksuci.2024.102128.
[10] Ghazy AS, Kaddoum G, Singh S. Low-latency low-energy adaptive clustering Hierarchy protocols for underwater acoustic networks. IEEE Access 2023;11:50578–94. https://doi.org/10.1109/ACCESS.2023.3277395.
[11] Shaban AM, et al. Performance evaluation of HEED protocol in underwater wireless sensor networks: throughput and delay analysis. IET Conference Proceedings Mar. 2025;2024(34):494–8. https://doi.org/10.1049/icp.2025.0132.
[12] V Anjali S, Paul B, Dhanya S, Pradeep A. Performance analysis of leach and pegasis routing protocols in a wireless sensor networks for smart water quality monitoring. In: 2024 15th international conference on computing communication and networking technologies (ICCCNT). IEEE; Jun. 2024. p. 1–6. https://doi.org/10.1109/ICCCNT61001.2024.10724635.
[13] Nasirian S, Pierleoni P, Belli A, Mercuri M, Palma L. Pizzza: a joint sector shape and minimum spanning tree-based clustering scheme for energy efficient routing in wireless sensor networks. IEEE Access 2023;11:68200–15. https://doi.org/10.1109/ACCESS.2023.3291915.
[14] Hilmani A, Koundi M, Sabri Y, Maizate A. Energy-efficient routing protocol for wireless sensor networks based on progressive and concentric clusters. Int J Reconfigurable & Embedded Syst ISSN 2023;2089(4864):4864.
[15] Chuhang W, Huangshui H, Tingting W. An improved multi-hop LEACH protocol based on chaotic genetic algorithm for wireless sensor networks. Wirel Pers Commun Feb. 2024;134(3):1843–62. https://doi.org/10.1007/s11277-024-10988-6.
[16] Maddi PA, Uppala P, Jitte SS, Raju MR. Energy efficient clustering in wireless sensor networks: a grey wolf-based TEEN protocol. In: 2024 international conference on integrated intelligence and communication systems (ICIICS). IEEE; Nov. 2024. p. 1–7. https://doi.org/10.1109/ICIICS63763.2024.10859939.
[17] Saxena M, Sardhara A, Raina S. An approach to increase the lifetime of traditional LEACH protocol using CHME-LEACH and CHP-LEACH. 2024. p. 133–45. https://doi.org/10.1007/978-981-99-9037-5_11.
[18] Jamaesha SS, Kumar RS, Gowtham MS. Cluster based hybrid optimization and kronecker gradient factored approximate optimum path curvature network for energy efficiency routing in WSN. Peer Peer Netw Appl May 2024;17(3):1588–609. https://doi.org/10.1007/s12083-024-01675-1.
[19] Tsai A-H. Relay network planning with geography information system and genetic algorithm in an actual multicell environment. IEEE Syst 2023;J:1–12. https://doi.org/10.1109/JSYST.2023.3307151.
[20] Abdulzahra AMK, Al-Qurabat AKM, Abdulzahra SA. Optimizing energy consumption in WSN-based IoT using unequal clustering and sleep scheduling methods. Internet of Things Jul. 2023;22:100765. https://doi.org/10.1016/j.iot.2023.100765.
[21] Rahul P, Kaarthick B. An optimal cluster head and gateway node selection with fault tolerance. Intelligent Automation & Soft Computing 2023;35(2):1595–609. https://doi.org/10.32604/iasc.2023.025762.
[22] Nasri M, Omari M, Kaddi M. "The application of multihop HEED protocol. Wirel Sens Netw 2022:557–65. https://doi.org/10.1007/978-3-030-92038-8_56.
[23] Rahman ABA, Hamim SI. Enhancing energy efficiency in multi-layer wireless sensor networks through cluster head fuzzy logic type 2 and multi-hop node strategies. International Journal of Computer Networks and Applications Oct. 2024;11(5):690–705. https://doi.org/10.22247/ijcna/2024/43.
[24] Anita, Asthana A. Routing protocols for wireless sensor network: a review and open research challenges. 2023. https://doi.org/10.3233/ATDE221313.
[25] Karthick GS. Energy-aware reliable medium access control protocol for energy-efficient and reliable data communication in wireless sensor networks. SN Comput Sci Jun. 2023;4(5):449. https://doi.org/10.1007/s42979-023-01869-z.
[26] Brahmam MV, Gopikrishnan S. Adaptive threshold based outlier detection on IoT sensor data: a node-level perspective. Alex Eng J Nov. 2024;106:675–90. https://doi.org/10.1016/j.aej.2024.08.063.
[27] Mote TS, Jagtap SK. Adaptive threshold residual energy-based efficient sensor network protocol (ATREEN). J Sens Jan. 2023;2023(1). https://doi.org/10.1155/2023/1629438.