

Received October 24, 2020, accepted October 30, 2020, date of publication November 3, 2020, date of current version November 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3035624

# An Improved Energy-Efficient Clustering Protocol to Prolong the Lifetime of the WSN-Based IoT

ALI ABDUL-HUSSIAN HASSAN<sup>1,2</sup>, WAHIDAH MD SHAH<sup>2</sup>, ABDUL-HUSSEIN HASSAN HABEB<sup>3</sup>,  
MOHD FAIRUZ ISKANDAR OTHMAN<sup>2</sup>, (Member, IEEE),  
AND MOHAMMED NASSER AL-MHIQANI<sup>2</sup>, (Member, IEEE)

<sup>1</sup>College of Education for Pure Sciences, University of Kerbala, Karbala 56001, Iraq

<sup>2</sup>Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya 76100, Malaysia

<sup>3</sup>Mathematics Department, Al-Zahraa University for Women, Karbala 56001, Iraq

Corresponding author: Ali Abdul-Hussian Hassan (altaeali801@gmail.com)

**ABSTRACT** A wireless sensor network (WSN) is an important part of the Internet of Things (IoT). However, sensor nodes of a WSN-based IoT network are constraining with the energy resources. A clustering protocol provides an efficient solution to ensure energy saving of nodes and prolong the network lifetime by organizing nodes into clusters to reduce the transmission distance between the sensor nodes and base station (BS). However, existing clustering protocols suffer from issues concerning the clustering structure that adversely affects the performance of these protocols. In this study, we propose an improved energy-efficient clustering protocol (IEECP) to prolong the lifetime of the WSN-based IoT. The proposed IEECP consists of three sequential parts. First, an optimal number of clusters is determined for the overlapping balanced clusters. Then, the balanced-static clusters are formed on the basis of a modified fuzzy C-means algorithm by combining this algorithm with a mechanism to reduce and balance the energy consumption of the sensor nodes. Lastly, cluster heads (CHs) are selected in optimal locations with rotation of the CH function among members of the cluster based on a new CH selection-rotation algorithm by integrating a back-off timer mechanism for CH selection and rotation mechanism for CH rotation. In particular, the proposed protocol reduces and balances the energy consumption of nodes by improving the clustering structure, where IEECP is suitable for networks that require a long lifetime. The evaluation results prove that the IEECP performs better than existing protocols.

**INDEX TERMS** Wireless sensor network, Internet of Things, clustering protocol, energy consumption, network lifetime.

## I. INTRODUCTION

Internet of Things (IoT) is a significant source of technological solutions in several applications. The IoT is pillars by a wireless sensor network (WSN) which decreases the cost of the new technology. Literature verifies that this technology integration will reduce costs and ensure convenience in daily life through smart sensor node networks whereby the nodes have access to internet [1], [2].

WSN, an inexpensive legacy system, has been applied in several fields, such as industrial control, environmental monitoring, military surveillance, and intelligent transportation systems [1], providing large-scale physical data that can

be further utilized. Thus, by integrating the IoT and WSN applications, no massive paradigm shift is needed [2].

WSN-based IoT is advantageous for its convenient deployment and low cost. Furthermore, it can function independently in harsh or high-risk places where human presence is not possible. However, WSNs have defects that need to be addressed [1]. The network lifetime problem is the main challenge in WSN [3].

The sensor's lifetime is only related to its batteries, which are difficult or impossible to replace or recharge due to the rugged environments where they are operating [4]. This problem undermines the integration of the WSN into IoT, elevating the costs of new technology. Accordingly, prolonged network lifetime is considered as a major challenge in the WSN-based IoT. Consequently, to prolong the network's lifetime and improve energy consumption, a clustering approach

The associate editor coordinating the review of this manuscript and approving it for publication was Gongbo Zhou.

is used in the WSN. The clustering protocol, where the sensor nodes are divided into small clusters, is an effective technique to reduce energy consumption and prolong network lifetime by avoiding long-distance communication [5], [6]. Each cluster employs one node as a cluster head (CH) that has duties more than member nodes (MNs). Practically, each MN in the cluster transmits its sensing data to its CH, and then the CH transmits these data to BS via a single-hop or multi-hop manner.

Although the clustering protocol is considered as an effective way to conserve energy for the nodes in WSNs, clustering structure remains a major issue, which adversely affects the network lifetime through the inefficient energy consumption of nodes [4], [7]–[10]. Furthermore, the WSN poor clustering structure frequently affects the subsequent procedures of the network, such as data aggregation and routing discovery, where it prepares the network for operation [11]. Consequently, the clustering structure efficiency has a considerable effect on the WSN lifetime.

The first of those issues is when determining a sub-optimal number of clusters (less or more than the optimal number), leading to the increase in the energy consumption of nodes [12]. Most of the clustering protocols that create balanced overlapping clusters suffer from the inaccurate determination of the optimal number of clusters when using current mathematical models because the distance to the CH has not been estimated correctly. The second issue is related to cluster formation, which can drastically affect the lifetime of WSN [13]. On certain occasions, an FCM algorithm (that is widely used in the WSN domain for cluster formation) produces unbalanced clusters (large and small) because of the random nodes deployment in the area, hence, resulting in unbalanced energy consumption for nodes. In large clusters, the selected CHs are burdened by data more than the CHs of the other clusters, thus, consuming more energy for the transmission of data [10], [14]. The third issue is improper CH selection, where most of the distributed methods do not take into account the routing information as a parameter in the CH selection. Consequently, an irregular distribution of the CHs occurs, where the transmission distances among the CHs in the network are uneven. Hence, some CHs are obliged to increase their signal strength in order to transmit data to the next hop, leading to unbalanced energy consumption for CHs in the network [15], [16]. The final issue of the clustering structure is the rotation of the CH function among members of the cluster. As a fixed value of energy is used as a threshold for CH rotation, the nodes demonstrate a dysfunctionality in terms of the CH and MN functions in the cluster. This dysfunctionality leads to unbalanced energy consumption for nodes that have been sequentially selected as CHs in the cluster, which subsequently accelerates the first node death (FND). As a result, two problems occur in relation to the CH that generate unbalanced energy consumption. The problems include the unbalanced transmission distances among CHs in the network, and the use of a static value of the threshold to rotate the CH function among members of the cluster.

Therefore, this research is very significant as it addresses the main research question: how to prolong the network lifetime for the WSN-based IoT? Several sub-questions are identified as follows:

How to determine the optimal number of clusters in case of the formation of the overlapping-balancing clusters?, How to form balanced clusters with little cost of the intra-distance of clusters in the random nodes distribution?, How to achieve balanced energy consumption among the CHs of clusters?, How to achieve balanced energy consumption of the successive CHs in the cluster?

To address these issues pertaining to the clustering structure that adversely affects the network lifetime through the inefficient energy consumption of nodes, and to answer the posed questions, this study proposes an improved **energy-efficient clustering protocol (IEECP)** to prolong the lifetime of the WSN-based IoT which consists of three parts:

Firstly, a modified mathematical model is proposed based on the analysis of the energy consumption model for multi-hop communications and overlapping clusters in order to determine the optimal number of clusters. Secondly, a modified fuzzy C-means algorithm (M-FCM) is proposed in order to produce balanced cluster. Thirdly, a new algorithm is proposed known as CH selection and rotation algorithm (CHSRA) that integrates the back-off timer mechanism for CH selection, with a new rotation mechanism for CH rotation among members of the cluster.

The main contribution by the proposed protocol is the prolonging of the WSN-based IoT lifetime that depends on the node's battery, which extensively increases the applications' range of the WSN-based IoT. This major contribution can be achieved through the following tasks:

- 1) **Selecting the optimal number of clusters based on the modified mathematical model by considering the overlapping case among clusters and multi-hop communications,**
- 2) **Forming balanced clusters that reduce the cost in the intra-distance based on modified fuzzy C-means algorithm (M-FCM) that result from a combination of the FCM algorithm with a centralized mechanism,**
- 3) **Reducing the energy overhead that results from the CH selection process in each round by a new integration of the back-off timer mechanism for CH selection with rotation mechanism in one algorithm known as CH selection and rotation model (CHSRA),**
- 4) **Balancing the communication distance among the CHs in the network based on a new objective function for the back-off mechanism, and**
- 5) **Balancing the life of the selected CHs in the cluster based on a new dynamic threshold.**

For straightforward ease in reading, most of the abbreviations used in this study are illustrated in Table 1. The rest of this paper is organized as follows: Section 2 provides a brief survey of the clustering algorithms and their advantages and disadvantages in literature. In section 3, the radio energy consumption model is introduced. Section 4 details the proposed

**TABLE 1. Abbreviations.**

Notation	Description
$L$	The size of messages (bit)
$E_{ini}$	The initial energy for node
$d_0$	The distance threshold between the transmitter and receiver
$E[d_{CH}]$	mathematical expectation of the distance from the member nodes to the CH
$N$	The total number of nodes in the network
$n$	The number of member nodes in the cluster
$R$	The radius of the transmission area for CH
$M^2$	The sensing area
$j$	The current cluster number, $j = 1, 2, \dots, k$
$i$	The current nodes, $i = 1, 2, \dots, n$
$E_{non-CH}$	The energy consumption for MN
$\rho$	node's density
$K$	The number of clusters
$C_{over}$	The overlapping area between clusters
$R_{over}$	The radius for the overlapping clusters
$E_{total}$	The total energy consumption for clusters
$\mu$	The membership of node (i)
$C_j$	The cluster centroid
$d(x_i, c_j)$	The distance between a node (i) and centroid (cj)
$Th_{cluster}$	The cluster threshold
$Pe$	The permittivity value
$\varepsilon$	The termination condition for the FCM operation
$m$	The value of the fuzzifier
$T_b$	The node's timer
$ADV$	The advertisement message for the CH declaration
$F$	The objective function for CH selection
$E_r$	The residual energy for node
$E_{com}$	The consumption energy of the node
$d_{FCH}$	the distance between a specific node to the forward CH
$d_{BCH}$	the distance between a specific node to the backward CH
$ACD$	The adjustment coefficient for distances
$T$	The fixed ratio from the initial energy
$E_{TH}$	The energy threshold for CH rotation
$R_{CHs}$	Rounds of all CHs in the cluster at the $E_{TH}$
$R_n$	Rounds of the member nodes in the cluster at the $E_{TH}$
$E_{CH-per-round}$	The energy consumption for the CH per round
$E_{n-rnd}$	The energy consumption for the nodes per round
$STD$	The standard deviation
$\bar{x}$	The mean of the cluster size
$S_j$	cluster size (j)
$FND$	First Node Dies
$LND$	Last Node Dies
$WSND$	Weighted First Node Dies
$HND$	Half Node Die

protocol. Then, the results of the discussion are explicated in Section 5. Finally, section 6 concludes the study.

## II. RELATED WORK

Among the principal goals of the cluster-based protocol is an effective clustering structure of the network, which enables the decrease in the energy being consumed, and offers balanced energy consumption [5]. The first proposed clustering protocol is the LEACH protocol [17]. The primary idea is to select the CH in a distributed manner at each round and let the nodes join the nearest CH to form a dynamic cluster. This network topology is formulated based on the selected CHs, which is basically not efficient due to its disregard for the residual energy of nodes [18]. If the CHs are not optimally selected for the frequent rounds, the network will suffer from a poor clustering setup even when it has the ability to adapt, thus, weakening the protocol performance. Therefore,

the major challenge for dynamic clustering is in the CH selection [16]. Furthermore, the priority for CH selection leads to the formation of dynamic clusters at each round, thereby increasing the energy overhead as a result of the cluster formation after each re-selection process for CHs [19], [20]. Another version of the protocol is a LEACH-centralized protocol (LEACH-C) [21], where the optimal number of clusters  $K$  is determined based on a mathematical model. In contrast to LEACH concerning the CH selection and cluster formation, Base Station BS is responsible for these processes through the utilization of the simulated annealing optimization method, where at every round, the nodes that have more than the average energy will transmit their information to the BS. In addition, nodes with less than average energy do not stand a chance of being a candidate for the CH function. However, energy overhead remains when the information is transmitted to BS and the round trip is time-consuming at the CH selection process [1], [22].

To address the energy consumption and delay issues, the energy delay index for the trade-off (EDIT) protocol is proposed [23]. This protocol uses a back-off timer mechanism to select the CHs using the objective function that depends on the residual energy and distance to the BS with the number of neighbor nodes. Then, the nodes will join the nearest CH to form the clusters. For data transmission, this protocol uses a multi-hop method to send the sensing data to the BS based on the energy-delay function. This protocol enables the reduction of the overhead in CH selection by using the back-off timer mechanism. However, it does not guarantee that the selected CHs are evenly distributed in the monitoring area [20], leading to unbalanced energy consumption and longer time-consumption due to the re-selection of all CHs simultaneously [1]. Furthermore, it suffers from the formation of unbalanced clusters by letting nodes join the nearest CH. In addition, dynamic clustering overhead is also formed.

Another proposed protocol used to minimize the overhead of the CH selection and delay is delay-constrained energy multi-hop (DCEM) [24]. In this protocol, the CHs are selected based on the back-off timer mechanism according to the residual energy and distance to BS for the nodes. Then, it lets the nodes join the nearest CH. This proposed protocol is a fully distributed approach, which is similar to the EDIT protocol. Furthermore, the multi-hop communication method is used by transmitting data to BS based on the cost-function of energy and end-to-end delay. However, this protocol has the same problems as the EDIT protocol.

Ray and De [25] proposed an energy-efficient clustering protocol based on K-means (EECPK-means). At the initial stage, this protocol determines the optimal number of clusters based on the mathematical model. Next, it overcomes the problems caused by the formation of the dynamic clusters by generating the static and balanced clusters through the improved KM algorithm that is executed at the BS. Subsequently, the CHs are selected in the distribution by a given ID number for each node based on the distance of the node from

the centroid of the cluster. To reduce the overhead of nodes, the CH re-selection process is not executed in every round but through a rotation mechanism that is based on a fixed value of energy as a rotation threshold. If the residual energy of the CH node is less than the threshold value, then the CH changes. For data transmission to the BS, this protocol uses an energy-cost function for the next-hop selection to save the energy for the CHs, hence, prolonging the network lifetime.

Jain [14] proposed a Traffic-Aware Channel Access Algorithm for Cluster-Based Wireless Sensor Networks to form balanced and static clusters. In the later sections, this protocol is called (TACAA) for simplicity. This protocol presents a modified-FCM by re-arranging the degrees of belonging for the nodes to produce balanced clusters to overcome the random node deployment issue in the sensing area. In terms of the CH selection, after the current CH works for a set of rounds as a threshold for CH rotation, the current CH selects the next CH in the cluster according to residual energy and degrees of belonging of the node. However, the reliance on the degrees of belonging is inefficient because of a normalization condition in the membership function [26], where the actual distance from the centroids is not indicated. Consequently, this condition impacts the energy consumption of nodes by increasing the intra-distance for each cluster. In addition, the reliance on the set of rounds as the threshold value is inefficient [27]; if the selected CH does not have enough energy to perform its job over this number of rounds, then the cluster becomes an island that is isolated from the network. OCM-FCM [10] also improves the structure of the cluster by presenting a new mathematical model to determine the optimal number of clusters based on the energy consumption model analysis for nodes. Furthermore, static clusters are formed based on the improved FCM algorithm. This protocol uses the distributed approach for the CH selection that reduces the overhead on nodes and the round-trip time by transmitting their information to the BS, where the current CH selects the next CH of the cluster based on residual energy for nodes. However, the CH re-selection process is conducted at every round, increasing energy consumption through the exchange of the control message among members of the cluster. Moreover, this process suffers from the same problem in terms of the CH distribution in the monitoring area because it relies only on the residual energy of the node in the CH selection.

Based on literature, the clustering protocol should consider four aspects: 1) the optimal number of clusters, 2) the formation of balanced and static clusters, 3) the evenly distribution of the selected CHs in the monitoring area with low overhead in the selection process, and finally, 4) the CH rotation process that relies on a threshold value. However, these factors have not been addressed in depth by the existing studies, hence, affecting the clustering protocol performance.

### III. RADIO ENERGY CONSUMPTION MODEL

The energy consumption for nodes is measured by using the radio energy consumption model that depends on the

distance between the transmitter and receiver [28]. Based on the distance between them, the free space or multi-path models can be utilized. Therefore, as a message ( $L$  bit) is transmitted through a distance ( $d$ ), the energy consumption for a transmitter node can be formulated as

$$E_{TX}(l, d) = \begin{cases} E_{elec} * l + E_{AD} * l + \varepsilon_{fs} * L * d^2, & d \leq d_0 \\ E_{elec} * l + E_{AD} * l + \varepsilon_{amp} * L * d^4, & d > d_0 \end{cases} \quad (1)$$

where  $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{amp}}$  is the distance threshold between the transmitter and receiver, which equals to (78.7);  $E_{elec}$  represents the energy consumption for the electronic system, whether sending or receiving one bit;  $E_{AD}$  represents the energy consumption for data aggregation; and  $\varepsilon_{fs}$  and  $\varepsilon_{amp}$  are the energy consumption of the free space propagation and power consumption of multipath propagation, respectively:

$$E_{RX}(L) = E_{elec} * L \quad (2)$$

where  $E_{RX}(L)$  is the energy consumption at the node as one bit of the message is received.

### IV. IECCP

This section clarifies the proposed protocol which consists of three parts: determination of the optimal number of clusters based on a modified mathematical model, formation of balanced clusters based on a modified fuzzy C-means (M-FCM) and selection and rotation of the CH for clusters based on the CH selection-rotation algorithm (CHSRA).

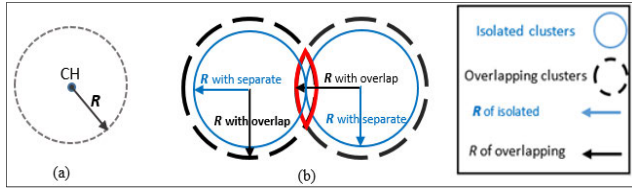
#### A. DETERMINATION OF THE OPTIMAL NUMBER FOR CLUSTERS

The mathematical model is popularly used in the domain to ascertain the number of clusters. This method is less time-consuming in finding the number of clusters, where the number of clusters is defined prior to the execution of the deployment process for nodes. Accordingly, it is suitable for all types of applications, especially for real-time applications, hence, drawing many studies to utilize this method to determine the optimal number of clusters. This method is often executed by the BS.

The mathematical model relies on a disk model to represent the distance to the CH. The disk model [7] is often utilized for studying the WSN communication, taking into account the coverage area for the transmission and entailing a disk of the plane with radius  $R$ , as shown in Figure 1-a. This value represents the distance to the CH in the mathematical model. The estimated value of radius ( $R$ ) has a significant effect on the final result of the number of clusters, as proven later in this paper. Although other studies consider the distance to CH is the same whether the clusters are overlapping or isolated, in reality, the value of the radius is greater in the overlapping clusters, as shown in Figure 1-b.

We assume that  $N$  nodes are deployed randomly in a square sensing area ( $M2$ ). If  $K$  clusters exist, then the means number of the total nodes for each cluster is  $N/K$  (one CH and





**FIGURE 1. (a) disk model, and (b) difference in radius between overlapping clusters and separate clusters.**

$N/K - 1$  of MNs). Every cluster head consumes an amount of energy when receiving data from MNs, aggregating them from MNs, and transmitting the aggregate data to the BS. As the BS is located outside the sensing area, the multi-hop communication is used to transmit the sensing data to BS. Therefore, the energy consumption follows the free space model ( $d < d_0$ ) shown in Eq. 1.

$$E_{CH} = LE_{elec} \left( \frac{N}{K} - 1 \right) + LE_{AD} \frac{N}{K} + LE_{elec} + L\epsilon_{fs} d_{BS}^2, \quad (3)$$

where  $d_{BS}^2$  refers to the distance from the CH to the next hop in a multi-hop communication with the assumption of perfect aggregation for the data.

The MN of the cluster merely needs to send its data to the CH. As the distance inside the cluster between MNs and their CH is not big, the energy consumption also observes the free-space model. Consequently, the energy consumption utilized in each MN is as follows

$$E_{non-CH} = LE_{elec} + L\epsilon_{fs} d_{CH}^2, \quad (4)$$

where  $d_{CH}^2$  refers to the distance from MN to the CH. The area occupied by each cluster is around  $M^2/K$  [29]. Generally, this is an arbitrary-shaped area with a node deployment  $\rho(x, y)$ .

The predicted distance from MNs to their CH that is supposed to be at the center of the cluster is given by

$$E[d_{CH}] = \iint (x^2 + y^2) \rho(x, y) dx dy, \quad (5)$$

$$E[d_{CH}] = \iint r^2 \rho(r, \theta) r dr d\theta. \quad (6)$$

In the separated clusters, the area is a circle with radius  $R = M/\sqrt{\pi K}$  and  $\rho(r, \theta)$  is constant for  $r$  and  $\theta$  [10].

Nonetheless, the radius  $R_{over}$  of the overlapping clusters is more than the radius  $R_{spt}$  of the separated clusters for the same distributed area. Therefore, the area of the overlapping clusters is a circle by radius ( $R_{over}$ ); thus, the radius of the overlapping clusters becomes

$$R_{over} = R_{spt} + C_{over}, \quad (7)$$

To estimate the appropriate value of  $C_{over}$ , we execute intensive simulations at various values of  $C_{over}$ . Based on these simulation results, the range of the appropriate  $C_{over}$  value is from 0.04 to 0.09 of  $R$ . The determination of  $C_{over}$  value [30] used the sum of the weighted values of the range, so the value

of  $C_{over} \cong 0.06$ , as illustrated in the Results and Discussion section.

$$R_{over} = R + 0.06R = 1.06R, \quad (8)$$

Therefore, Eq. 6 is as follows:

$$E[d_{CH}] = \rho \int_{\theta=0}^{2\pi} \int_{r=0}^{1.06M/\sqrt{\pi K}} r^3 dr d\theta = \frac{1.262\rho M^4}{2\pi K^2}. \quad (9)$$

If the density of nodes is uniform throughout the cluster area, then

$$\rho = \frac{K}{M^2}, \quad (10)$$

$$E[d_{CH}] = \frac{1.31 M^2}{2\pi K}, \quad (11)$$

$$E_{non-CH} = LE_{elec} + L\epsilon_{fs} \left( \frac{1.262 M^2}{2\pi K} \right). \quad (12)$$

The energy consumption in the cluster is as follows:

$$\begin{aligned} E_{cluster} &= E_{CH} + \left( \frac{N}{K} - 1 \right) E_{non-CH} \\ &\cong E_{CH} + \left( \frac{N}{K} \right) E_{non-CH}. \end{aligned} \quad (13)$$

For  $K$  clusters, the total energy consumption is as follows:

$$E_{total} = K * E_{cluster} \quad (14)$$

$$\begin{aligned} E_{total} &= K \left[ \frac{N}{K} (LE_{elec} + L\epsilon_{fs} d_{CH}^2) + LE_{elec} \left( \frac{N}{K} - 1 \right) \right. \\ &\quad \left. + LE_{AD} \frac{N}{K} + LE_{elec} L\epsilon_{fs} d_{BS}^2 \right] \end{aligned} \quad (15)$$

$$\begin{aligned} &= K \left[ \frac{N}{K} \left( LE_{elec} + L\epsilon_{free} \left( \frac{1.262 M^2}{2\pi K} \right) \right) \right. \\ &\quad \left. + LE_{elec} \left( \frac{N}{K} - 1 \right) \right. \\ &\quad \left. + LE_{AD} \frac{N}{K} + LE_{elec} + L\epsilon_{fs} d_{BS}^2 \right] \end{aligned} \quad (16)$$

$$\begin{aligned} E_{total} &= NLE_{elec} + NL\epsilon_{fs} \left( \frac{1.262 M^2}{2\pi K} \right) + NLE_{elec} \\ &\quad + NLE_{AD} + KL\epsilon_{fs} d_{BS}^2 \end{aligned} \quad (17)$$

By equating the derivative of the total energy of the network with respect to  $K$  to zero,

$$0 = -NL\epsilon_{free} \left( \frac{1.262 M^2}{2\pi K} \right) + KL\epsilon_{fs} d_{BS}^2. \quad (18)$$

The optimum number of clusters can be obtained and given by

$$K = \sqrt{\frac{1.262N}{2\pi} \frac{M}{d_{BS}}}. \quad (19)$$

## B. FORMATION OF BALANCED CLUSTERS

For the formation of balanced clusters, a modified fuzzy C-means algorithm (M-FCM) is proposed in this study by combining the FCM with a centralized mechanism.

Before discussing the proposed algorithm to form balanced clusters, the conventional FCM algorithm is illustrated in the next section to provide a general idea.

### 1) FCM ALGORITHM OVERVIEW

The FCM algorithm has been widely used in the clustering processes for WSN cluster formation. This algorithm was originally presented by Dunn [31]. The goal of FCM is to form better clusters by reducing the summation of distances between the objects (N) and the cluster centers (C) by using the objective function. In WSN, the objects refer to nodes that are already distributed in the sensing area. The FCM objective function for organizing nodes into clusters in the WSN can be formulated as follows:

$$J_{FCM} = \sum_{i=1}^n \sum_{j=1}^K \mu_{ij}^m d(x_i, c_j)^2, \quad i = 1, 2, \dots, N, j = 1, 2, \dots, K \quad (20)$$

The normalization condition is

$$\sum_{j=1}^K \mu_{ij} = 1, \quad \mu \in [0, 1] \quad (21)$$

$$\mu_{ij} = \frac{1}{\sum_{j=1}^K \left( \frac{d(x_i, c_j)}{d(x_i, c_k)} \right)^{\frac{2}{m-1}}}, \quad (22)$$

$$C_j = \frac{\sum_{i=1}^n (\mu_{ij})^m * d(x_i, c)}{\sum_{i=1}^n (\mu_{ij})^m}, \quad (23)$$

where K refers to the number of clusters, N refers to the number of nodes,  $\mu$  refers to the membership of node (i) to cluster (j),  $C_j$  refers to cluster centroid; d refers to the distance between a node (i) and centroid ( $c_j$ ), commonly described by Euclidean distance; and  $m$  is the value of the fuzzifier that is chosen as a real number greater than 1 ( $m \in [1, \infty)$ ).  $m$  approaches to 1 clustering tend to become crisp (same as K-means algorithm) but when it reaches to the infinity, clustering becomes fuzzified (unreliable) [32]. Therefore, the value of fuzzifier is usually chosen as 2 in most of the applications [33], [34]. To terminate the algorithm, we use the condition  $U_{ij}(t) - U_{ij}(t-1) < \varepsilon$ , where  $t$  is the current iteration, and  $\varepsilon$  is a very small number close to zero (e.g., 0.001). On certain occasions, FCM produces unbalanced clusters because of the nature of the random deployment of sensor nodes in the monitoring area [35], as shown in Figure 2. This situation leads to unbalanced energy consumption for nodes, which adversely affects the network lifetime [10], [14], [25]. Some of the studies sought to overcome this problem by rearranging the degrees of belonging for nodes to produce balanced clusters, as shown in [14]. However, relying on the degrees of belonging is inefficient because of a normalization condition in the membership function, leading to an increase in the intra-distance for the clusters. Consequently, this condition increases the energy consumption of nodes [26].

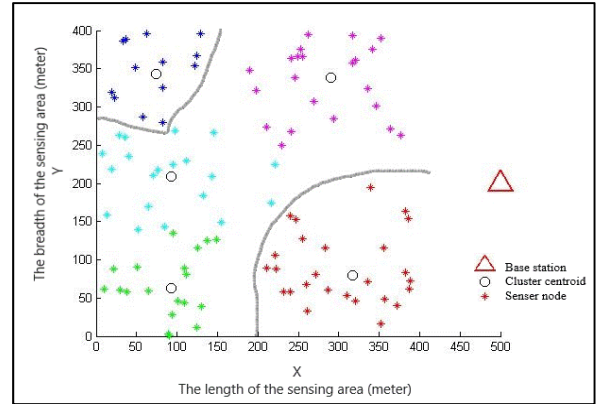


FIGURE 2. Formation of unbalanced cluster size using FCM.

To address this issue, a modified clustering algorithm has been proposed in this study to form balanced clusters with minimal intra-cluster distance by relying on the actual distance from centroids rather than the degrees of belonging for nodes.

### 2) MODIFIED FCM (M-FCM)

The proposed clustering algorithm is executed at the BS and consists of two phases: 1) initial cluster formation, which is based on the FCM, and 2) balanced cluster formation, which is based on the CM. In the initial cluster formation, the FCM is applied to form the clusters as shown in the algorithm, and then the process shifts to the second phase. The balanced cluster formation phase consists of two subphases. The first subphase consists of the following steps:

- 1) The cluster threshold ( $Th_{cluster}$ ) is determined based on Eq. 24.
- 2) Clusters are sorted based on size.

Minimum cluster size is compared with that of the  $Th_{cluster}$ . If the size is greater than the  $Th_{cluster}$ , then the FCM creates balanced clusters. Otherwise, the process shifts to the second subphase.

$$Th_{cluster} = \frac{N * Pe}{K}, \quad (24)$$

where  $Pe$  is the permittivity value equals to 0.85 [25], and K signifies the number of clusters.

In the second subphase, CM considers the final centroids of the clusters that were produced from the previous phase (FCM phase) as initial points to form balanced clusters. Steps of the CM are as follows:

1. The distance between the initial points and nodes is determined.
2. Nodes are arranged based on their distance from the initial points.
3. The initial points select the nearest number of nodes that are equal to the threshold of the cluster value to join it.
4. The remaining nodes that are still non-joined join the nearest initial point to construct the final clusters.

**Algorithm 1** M-FCM**Input**

Number of sensor nodes =  $N$   
 Number of clusters =  $K$   
 maximum iteration = 100  
 Improvement value =  $\epsilon$   
 Permittivity value  $Pe = 0.85$

**Output**

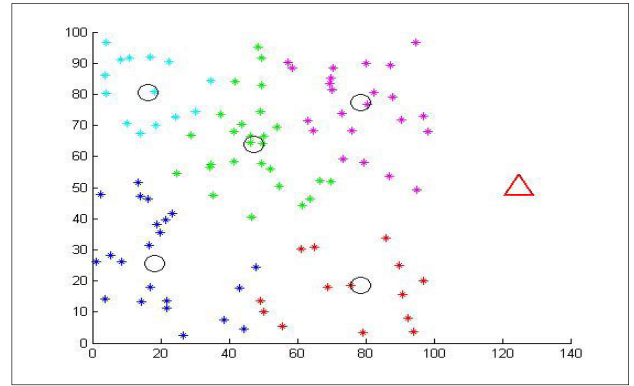
Balanced clusters

**Process**

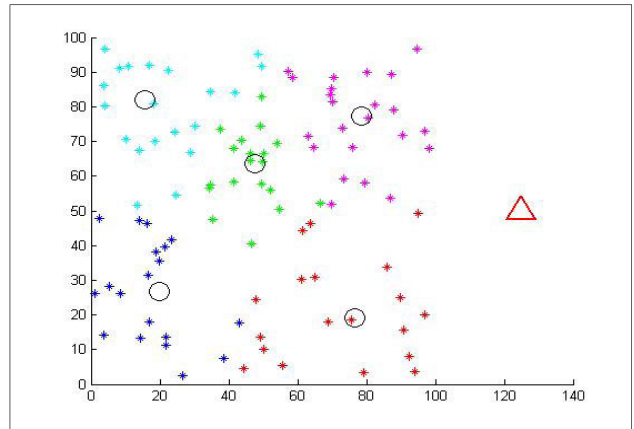
1. Select the random  $K$  point as an initial centroid  $C^{(0)}$ .
2. For  $i = 1$  to maximum iteration, **do**
3.   update the membership  $\mu$  matrix using Eq. 22.
4.   Calculate the new  $C$  centroids using Eq. 23.
5.   Calculate the new objective function  $J$  using Eq. 20.
6.   If  $|J^{(i)} - J^{(i-1)}| < \epsilon$ , **then**
7.     break;
8.   else
9.      $J^{(i)} = J^{(i-1)}$
10.   end if
11. end for
- 12.
13. Calculate the cluster threshold ( $Th_{cluster}$ ) using Eq. 24
14. for  $x = 1$  to  $C$  **do**
15.   sort cluster based on cluster size
16.   find the minimum cluster size ( $min\_cluster$ )
17. end for
18. if  $mincluster > Th_{cluster}$ , **then**
19.   break;
20. else
21.   for  $x = 1$  to  $C$  **do**
22.     for  $j = 1$  to  $N$  **do**
23.       calculate the distance between the final centroid (initial point) and nodes
24.       sort nodes based on distance from the initial point
25.       end for
26.       end for
27.     for  $x = 1$  to  $C$  **do**
28.       initial point selects a  $Th_{cluster}$  number of nearest nodes
29.       end for
30.       the remaining nodes are joined to the nearest initial point based on distance
31.       for  $x = 1$  to  $C$  **do**
32.          $C_{final}(x, y) = \left( \left( \frac{1}{n} \sum_1^n x_i \right), \left( \frac{1}{n} \sum_1^n y_i \right) \right)$
33.       end for
34.     end if
35. end

5. Each cluster determines the new centroid based on the means for node locations.

This procedure ensures that the minimum cluster size is equal to or greater than the threshold cluster range with lower intra-cluster distance.



a) Formation of clusters based on conventional FCM



b) Formation of clusters based on M-FCM

**FIGURE 3.** Formation of clusters according to FCM and M-FCM.

Figure 3 shows an example of the cluster formation based on the FCM and M-FCM algorithms in the same node distribution, where the number of clusters ( $K$ ) are 5, and the total number of nodes ( $N$ ) are 100. In an optimal situation, we assume that each of the cluster embraces 20 as the mean of these nodes. As shown in Figure 3-a, when applied to the conventional FCM, the cluster with the sky nodes has only 14 nodes, but the cluster with the green nodes has 28 nodes. The size of these clusters varies considerably from the mean value of the cluster as aforementioned. When applied to M-FCM, the cluster with the sky nodes has only 20 nodes, and the cluster with the green nodes has 19 nodes, as shown in Figure 3-b.

Accordingly, the size of these clusters matches the mean value of the cluster as aforementioned, which is 20 nodes.

**C. CHSRA**

The CH selection and rotation issues have gained a great interest in researchers. Furthermore, in this study, a new algorithm has been proposed by integrating the back-off timer mechanism for CH selection with a rotation mechanism called CHSRA. In this algorithm, the CH is selected accurately by using a new objective function. Furthermore, the CH function is rotated among the members of the cluster based on

a new rotation mechanism, where it is executed without any contribution to the BS.

The goal of CHSRA is to reduce the overhead by selecting the CH within members of the cluster only. Furthermore, it balances the distance among CHs in adjacent clusters by adopting the routing information in the CH selection process that leads to balanced energy consumption for CHs. Besides, the CHSRA ensures the balance in energy consumption for the successive CHs of the cluster. The CHSRA comprises two phases:

- 1) CH selection phase implemented by the back-off timer mechanism, and
- 2) CH rotation phase implemented by the dynamic threshold mechanism.

### 1) CH SELECTION PHASE

The back-off timer mechanism is used to select the CH, which is a distributed mechanism. This mechanism is widely used in the literature because it reduces the overhead for nodes and has the least delay in the selection process [7], [36]. In this mechanism, each node in the cluster sets its timer. The node is set as either CH or CM according to its timer (Tb) and the advertisement (ADV) message is received before the timer terminates. If the node received the ADV message from another node in the cluster, then it will cancel its timer and become CM. However, if the timer expires and the node does not receive any message, it broadcasts the ADV message and becomes a CH [37]. The timer value is set based on an objective function (F) of the node, where the timer value is the converse of the objective function as follows:

$$Tb = 1/F. \quad (25)$$

This is presumably, the first time that the back-off timer mechanism is applied to select the CH within members of the cluster. In the current study, this mechanism is applied to the CH selection in all network nodes, thus, increasing time and energy consumption. Another significant contribution concerning the CH selection is to propose a new objective function for this mechanism that provides efficient distribution for the selected CHs in the network through selecting them in the optimal location. In this new objective function, the distance between a specific node to the forward CH (FCH) and the backward CH (BCH) is adopted along with the adjustment of coefficient for distances (ACD), in order to show the balance of distance between FCH and BCH and the residual energy of the node as the selection parameters for the CH selection process. This procedure ensures that the selected CH is in an optimal location according to the adjacent CHs of the other clusters. The proposed objective function relies on the aforementioned parameters rather than the residual energy of the node only [3], [36] or the residual energy of the node and distance to the BS [24], as they do not guarantee the efficient distribution of the CHs in the network. Figure 4 shows the effect of distances on the CH selection.

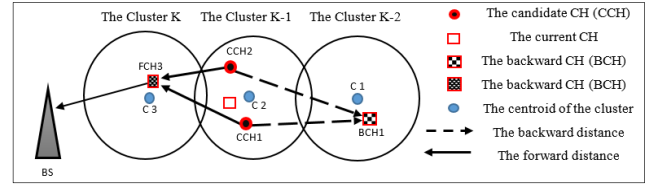


FIGURE 4. CH selection mechanism using backward and forward distance.

Consequently, each node in certain clusters computes the following parameters to define the objective function F, which are:

- residual energy  $E_r$  to prevent selecting CH with low energy

$$E_r = E_{ini} - E_{con}, \quad (26)$$

where  $E_{ini}$  refers to the initial energy and  $E_{con}$  refers to the consumption energy of the node.

- 1) Euclidean distance from the nearest forward CH ( $d_{FCH}$ ) to reduce the energy consumption for the candidate CH (j)

$$d_{FCH} = \begin{cases} \min(\|x_i - CH_j\|) d_{BS} > d_F \\ < d_0, \text{otherwise} \\ \|x_i - BS\| \end{cases} \quad j = 1, \dots, K \quad \text{and } i = 1, 2, \dots, n \quad (27)$$

- 2) Euclidean distance from the nearest backward CH ( $d_{BCH}$ ) to reduce energy consumption for backward CH (j-1)

$$d_{BCH} = \begin{cases} \min(\|x_i - CH_{j-1}\|) j > 1; \text{Otherwise} \\ 0 \end{cases} \quad (28)$$

- 3) ACD for the node; this coefficient is responsible for showing the balance of distance between FCH and BCH.

$$ACD = \frac{\min(d_{FCH}, d_{BCH})}{\max(d_{FCH}, d_{BCH})} \quad (29)$$

According to these parameters, the objective function F for CH selection is

$$F_i = \frac{E_r}{d_{FCH} + d_{BCH}} + ACD. \quad (30)$$

The selected CH based on this proposed algorithm overcomes the following two issues:

- 1) The energy overhead (additional energy cost) in the CH selection process is minimized by using the back-off time mechanism with members of the cluster rather than using all nodes in the network as in the current studies.
- 2) The CH is selected optimally because the required criteria for a balanced energy consumption in the selection are considered.



## 2) CH ROTATION PHASE

To solve the problem of unbalanced energy consumption for the successive CHs in the cluster, we set a dynamic threshold value for the CH rotation mechanism rather than the fixed value as in the other studies, where this value gradually increases with each process of the CH reselection. In this proposed mechanism, the energy consumed and the ratio from the initial energy (T) are used to estimate the threshold value. The first action taken by the selected CH directly after selection is calculating the value of its threshold for rotation ( $E_{TH}$ ) based on Eq. 31.

$$E_{TH} = \begin{cases} E_{con} + (E_{ini} - T) * E_{ini} & \text{if } (E_{con} + T * E_{ini}) \leq E_{ini} \\ E_{con} + E_r & \text{if } (E_{con} + T * E_{ini}) > E_{ini}, \end{cases} \quad (31)$$

where  $E_{con}$  is the consumption energy of the node,  $E_{ini}$  is the initial energy of the node,  $E_r$  is the residual energy of the node, and T is a constant value of initial energy but may differ from one cluster to another subjects to the number of members in the cluster. The T value is estimated only once for the cluster throughout the network lifetime. The T value can be calculated as follows:

$$R_{CHs} = (n - 1) * (E_{TH} / E_{CH-per-rnd}), \quad (32)$$

$$E_{rth} = E_{ini} - E_{TH}, \quad (33)$$

$$R_n = E_r / E_{n-rnd}, \quad (34)$$

where  $R_{CHs}$  refers to the rounds of all CHs in the cluster at the  $E_{TH}$ , with  $E_{rth}$  as the residual energy of the node at the  $E_{TH}$  value,  $R_n$  represents the rounds of the member nodes in the cluster at the  $E_{TH}$ ,  $E_{CH-per-rnd}$  is the energy consumption per round for the CH,  $E_{n-rnd}$  is the energy consumption per round for the nodes; and  $E_{TH}$  is the threshold value within the range from 0.1 to 0.9 of the initial energy values for the node.

$$T = R_{CHs} \cap R_n, \quad (35)$$

The appropriate value of T represents the intersection point of the curve of all CH rounds with the curve of members' rounds.

At the end of each round, the CH in the cluster compares its energy with the threshold value that is computed based on Eq. 32. If the residual energy of the current CH is equal to or less than the threshold value, then the current CH changes. Otherwise, the CH continues its function.

Finally, the preceding algorithms are combined in the IEECP framework that represents the main objective of this study, as shown in Figure 5.

For data transmission, members of the cluster directly transmit the sensing data to their selected CH. Then, the CH transmits this data to the BS by using a multi-hop manner. This manner is considered an advantageous option used to reduce energy consumption in case of relatively long-distance transmission [28]. In this manner, we adopt the same mechanism as portrayed in the [25]; the chosen CH checks if its transmission distance to BS is less than  $d_0$ ; it transmits the

## Algorithm 2 CHSRA

### Input:

Number of sensor nodes = N

Number of clusters = K

Cluster size = [ ]

maximum iteration = X

### Output:

Cluster heads with dynamic threshold for rotation

### Process:

1. for i = 1 to K **do**
2.   for j = 1 to cluster size (i) **do**
3.     each node in the cluster computes the  $E_r$ ,  $d_{FCH}$ ,  $d_{BCH}$ , and ACD according to Eqs. 26, 27, 28, and 29.
4.     setup the timer for nodes based on objective function (F) in Eq. 30
5.   end for
6.   max (F) = CH (i)  $\rightarrow$  broadcasts join message for members of the cluster
7.   members that received the join message cancel their timer and join CH
8.   Calculate the energy threshold  $E_{th}$  for CH rotation using Eq. 31
9.   end for
10. for x = 1 to max iteration **do**
11.   for i = 1 to K **do**
12.     for j = 1 to cluster size (i) **do**
13.       members send sensing data to CH
14.       update the residual energy  $E_r$  of nodes based on the energy model
15.     end for
16.     If  $E_{com}(\text{CH}) > E_{th}$  then
17.       re-select the CH (return to step 3)
18.     end if
19.   end for
20. end for
21. end

data directly to BS. Otherwise, it selects the nearest forward CH (towards BS) to reduce the energy consumption for CHs.

Although this mechanism will lead to some delay in data arrival at BS, it can bypass the long transmission distance for the selected CH, leading to energy saving for CHs.

## D. COMPLEXITY ANALYSIS FOR IEECP

As mentioned earlier, the execution of the IEECP protocol processes occurs in two different places. The first place is the BS, where the number of clusters is computed initially based on the modified mathematical model, and then the balanced clusters are formed based on the M-FCM. The second place is the node, where the CH selection and rotation are processed based on the CHSRA algorithm.

The determination of the number of clusters does not contribute to any time complexity, hence, it is deemed suitable for real-time applications.

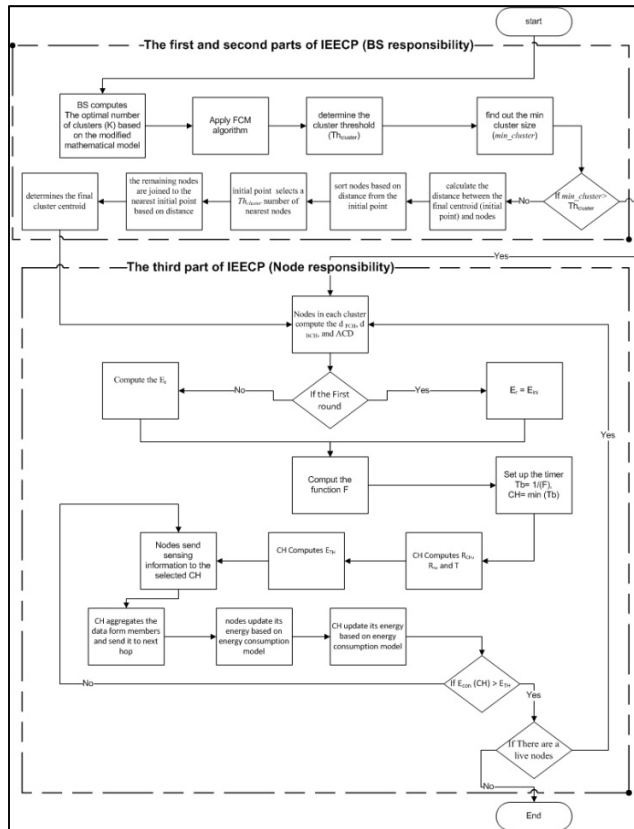


FIGURE 5. IEECP flowchart.

For the M-FCM, the time complexity is  $O(NK^2 \times I_{FCM}) + NK$ . The time complexity for FCM is  $O(NK^2 \times I_{FCM})$  as reported by [20], where  $N$  is the number of sensor nodes of the network,  $K$  is the required number of clusters, and  $I_{FCM}$  is the number of FCM iterations. In M-FCM, the time complexity has been increased by  $NK$  rather than in FCM. However, due to the one-time execution of this procedure through the BS prior to the network operation, there is no contribution to the time complexity related to M-FCM at the network operation. Moreover, since the BS does not have constraints related to the memory as found in the sensors, the space complexity of the M-FCM algorithm does not constitute any obstacle at the formation of clusters. Therefore, the parts of the IEECP protocol executed at the BS do not perform any time and space complexity with regard to the network operation.

Since CHSRA is a distributed algorithm which is being applied within the cluster (it is not applied for the whole network), the member of the cluster updates its information at each round. The CH re-selection phase occurs when the energy consumption of the current CH is more than the threshold; members of the cluster ( $n$ ) need to update their information to select the next CH for the cluster by relying on CHs for other clusters ( $K-1$ ). Thus, the analysis of the time complexity is based on the equations given in (30). Therefore, the time complexity is  $O(n \times I_{round} + K-1)$ , where  $I_{round}$  is

the number of rounds for the node until it dies. Consequently, the time complexity for the CHSRA is identical for the linear function, which is a small contribution for time complexity in terms of CH selection and rotation processes [38]. Furthermore, the space complexity of the CHSRA is  $O(K^2 + 50)$ , where it is an acceptable contribution of the space complexity for processes of the CH selection and rotation.

As for the overload complexity, the nodes do not suffer from overhead during the formation of clusters due to the fact that this process occurs only once in the network, executed by the BS, getting the benefit of forming static clusters through a centralized approach. Likewise, for CHSRA, the overload for the CH selection is reduced to the maximum possible extent, as during the selection process only the node that will be the CH broadcasts ADV message to the rest of the cluster members for joining it. Similarly, for the re-selection process, only the CH broadcasts the message to the rest of the cluster members for the re-election. Therefore, the overhead complexity of the CHSRA is dependent upon the number of CH in the network, which equal to the number of clusters  $K$ . The overhead complexity of the CHSRA is constant and identical to an  $O(1)$  because  $K$  is a predefined fixed value. In other words, the overhead complexity is independent of the network size.

## V. RESULTS AND DISCUSSION

This section presents the significant results of the proposed protocol implemented in Matlab. In this study, we adopted two pertinent phases of the evaluation to show the significance and reliability of this study, where the first phase involved the evaluation of the proposed algorithms separately. The second phase detailed the IEECP performance evaluation by comparing it with related works.

For the network scenario, the proposed protocol was performed on a WSN with nodes that were randomly deployed on a two-dimensional square area of length ( $M \times M$ ), where only one base station was situated outside the network. Other assumptions of the network are observed as follows:

- Nodes, as well as the base station, are stationary after deployment.
- Nodes have equal initial energy and are not physically accessible. Thus, these nodes could not be rechargeable.
- The base station is not limited in terms of energy, memory, and computational power.
- Base station knows the identifier of all node.
- Nodes are subject to the radio energy consumption model.
- Nodes know their geographical position.

### A. EVALUATION OF THE PROPOSED ALGORITHMS

This section reports the ability of each algorithm to overcome the problems that drive the proposal. The proposed algorithms are separately evaluated according to some evaluation parameters, as stated in the following:

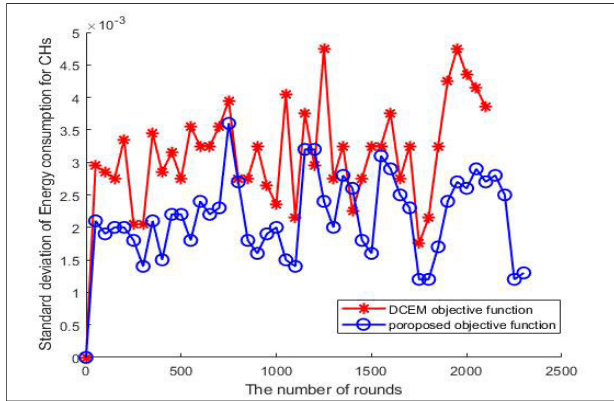


FIGURE 6. STD of energy consumption for CHs.

### 1) DISPERSION MEASUREMENT OF ENERGY CONSUMPTION AMONG CHS

The performance of the proposed objective function for CH selection of the back-off timer mechanism was compared with another objective function of the back-off timer mechanism that used the energy and distance to the BS for the CH selection as contained in the DCEM [24]. In this evaluation, we used the standard deviation (STD) to measure the variance of the energy consumption in the CHs according to the objective function of DCEM and proposed objective function, respectively. Figure 6 illustrates the (STD) of the energy consumption for the CHs over the network lifetime, where the STD of energy consumption based on the proposed objective function is less than the STD of the energy consumption based on DCEM objective function. Consequently, the energy consumption of the selected CHs based on proposed objective function was much more balanced than the energy consumption of the CHs that were selected based on the DCEM objective function. While the distance to the forward CH ( $d_{FCH}$ ) and the backward CH ( $d_{BCH}$ ) along with the adjustment coefficient for distances (ACD) was used in the proposed objective function, the distance to BS was used in the DCEM objective function.

### 2) ENERGY CONSUMPTION EVALUATION FOR THE SUCCESSIVE CHS OF THE CLUSTER

The performance of proposed rotation mechanism is evaluated through a comparison with the rotation mechanisms for the EECPK-means protocol [25] based on the energy consumption for the selected CH nodes.

The threshold energy for the EECPK-means protocol was equal to the sum of energy consumption for the CH when receiving, aggregating, and transmitting data to the BS for the mean number of nodes [25]; thus, it was a static value for all CHs of the cluster. As shown in Figure 7, in the EECPK-means algorithm, the first node that became a CH for the cluster consumed more than 98% of energy in its function as a CH until the threshold value was reached, and then it was transformed into an ordinary node. Its remaining energy

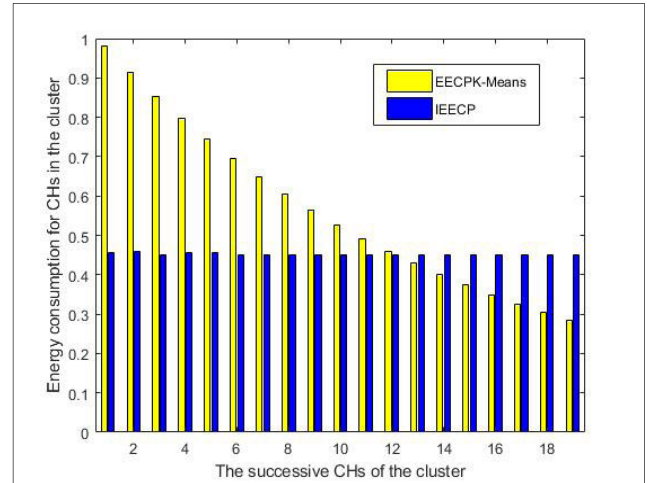


FIGURE 7. Energy consumption of selected CHs based on threshold.

was less than 2%, which was deemed impossible for the node to survive longer in the network. On the other hand, the last node that became a CH for the cluster consumed most of its energy as an ordinary node before becoming a CH for the cluster, enabling it to remain for a long duration of time in the network. Besides, it consumed 24.8% of energy until the threshold value was reached. This problem is common to all protocols that use a static threshold value. In contrast, due to a dynamic energy threshold reliance for the CH rotation that was gradually increasing according to the energy consumption of the node with each CH re-selection, the energy consumption was almost balanced for all the successive CHs in the cluster of the proposed protocol, as shown in Figure 7.

### 3) EVALUATION OF BALANCED CLUSTER FORMATION

The formation of balanced clusters in the proposed protocol was evaluated through a comparison made with the clustering algorithms in the EECPK-means [25], TACAA [14], and OCM-FCM [10] protocols based on measurement parameters used in the literature namely balanced cluster size and cost of intra-cluster distance. This is due to the clusters become more balanced at the cost of the intra-cluster distance [39]. Seven observations similar to those presented in [25] were compared, as shown in Table 2.

#### a: VARIATION IN SIZE OF CLUSTERS (VSC)

This parameter measured the dissimilarities of the sizes among the clusters, where the smaller the factor, the better. This condition signified that a balance existed in the cluster size.

$$VSC = \frac{\sum |S_j - \bar{x}|^2}{k}, \quad (36)$$

$$\bar{x} = \frac{\sum_{j=1}^k S_j}{k}, \quad (37)$$

where  $S_j$  refers to cluster size ( $j$ ) and  $\bar{x}$  refers to the mean of the cluster size. As the network has 100 sensor nodes with

**TABLE 2.** Cluster size in various observations.

Observe	EECPK-Means					OCM-FCM					TACAA					IEECP				
	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Clust 1	Clust 2	Clust 3	Clust 4	Clust 5	Clust 1	Clust2	Clust 3	Clust 4	Clust 5
1	19	21	24	18	18	15	18	21	22	24	18	24	19	21	18	18	18	20	20	24
2	21	18	22	21	18	24	17	20	19	20	23	18	18	18	23	24	18	19	18	21
3	19	20	21	20	20	16	27	16	18	23	20	17	22	19	22	18	23	18	19	22
4	21	19	19	20	21	25	23	14	21	17	23	21	17	21	18	23	21	17	21	18
5	20	24	23	17	16	15	23	21	25	16	19	23	20	22	16	18	23	21	21	17
6	19	21	16	22	22	23	23	17	15	22	23	22	18	17	20	22	22	18	17	21
7	17	23	18	22	20	20	15	20	22	23	18	17	22	21	22	20	17	19	22	22

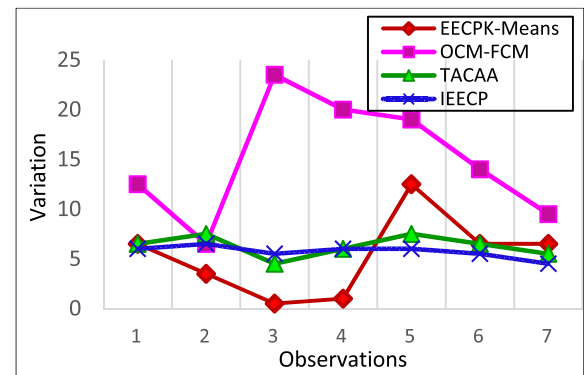
**TABLE 3.** Comparison of results.

Observe	EECPK-Means		OCM-FCM		TACAA		IEECP	
	VSC	Total distance DT (m)	VSC	Total distance DT (m)	VSC	Total distance DT (m)	VSC	Total distance DT (m)
1	6.5	2400.8	12.5	1573.20	6.5	1589.14	6	1574.23
2	3.5	2692.7	6.5	1799.50	7.5	1815.76	6.5	1804.81
3	0.5	2618.1	23.5	1481.49	4.5	1529.34	5.5	1491.05
4	1	2613.1	20	1633.37	6	1651.84	6	1651.84
5	12.5	2731.6	19	1633.76	7.5	1662.9	6	1658.56
6	6.5	2639.9	14	1619.91	6.5	1644.17	5.5	1642.79
7	6.5	2545.6	9.5	1537.32	5.5	1559.79	4.5	1542.51

five clusters, the value of  $\bar{x} = 20$ . The results are illustrated in Table 3. As shown in Figure 8, the variation in cluster size that was formed based on the EECPK-means was less than the variation in cluster size for OCM-FCM (which was desirable in the cluster formation) and was occasionally less than the variation in cluster size for the TACAA protocol and proposed algorithm M-FCM. Likewise, the variation in cluster size for the proposed protocol was less than that of the variation of the clusters for OCM and TACAA. It was also occasionally less than the variation in cluster size for the EECPK-means protocol. However, M-FCM was deemed more superior than other protocols based on the stability in the results of the variation, signifying that the performance of the M-FCM algorithm was stable in the formation of balanced clusters for all observations.

#### b: COST OF INTRA-CLUSTER DISTANCE ( $D_T$ )

This evaluation parameter was crucial because it showed the significance of producing balanced clusters on the total energy consumption in the network. Therefore, when forming the balanced clusters, the increase in the intra-distance should be in acceptable range and not significant. The total

**FIGURE 8.** Variation in size of clusters.

intra-distance DT could be computed based on Eq. 38.

$$D_T = \sum_{j=1}^K \sum_{i=1}^n d(x_i, c_j)^2, \quad (38)$$

where  $d(x_i, c_j)$  is the distance from a node  $x_i$  to the cluster centroid  $c_j$ ,  $n$  is the number of cluster members, and  $K$  is the number of clusters. The results are illustrated in Table 3.



**TABLE 4.** Simulation parameters.

Parameter	Value
$E_{elec}$	50 nJ/bit
$E_{free}$	10 pJ/bit/m <sup>2</sup>
$\epsilon_{amp}$	0.0013 pJ/bit/m <sup>4</sup>
$L$	3200 bit
$E_{ini}$	1 J
$E_{AD}$	5nJ/bit
$N$	100; 1000
$M$	100*100; 1000*1000
$BS$ location	(50,125); (500,1250)

Based on Table 2, the OCM–FCM protocol had the lowest intra-distance of clusters because it improved only at the initial selection of the cluster’s centroid. On the other hand, the intra-distance of clusters for IIECP was slightly more than that for OCM–FCM and less than that of the other protocols, where the cost of IIECP hit the highest level by the intra-distance (24.793 m) in the fifth observation. This cost was equal to 1.5% of the total intra-distance for OCM–FCM for the same observation. This cost of the intra-distance was acceptable when compared with the cost of the protocols, hence, indicating that IIECP was better than others in reducing the cost of producing balanced clusters. Overall, M-FCM of our proposed protocol was significantly superior in comparison to other clustering algorithms of the existing protocols in terms of the formation of balanced clusters.

### B. EVALUATION OF THE PROPOSED PROTOCOL PERFORMANCE

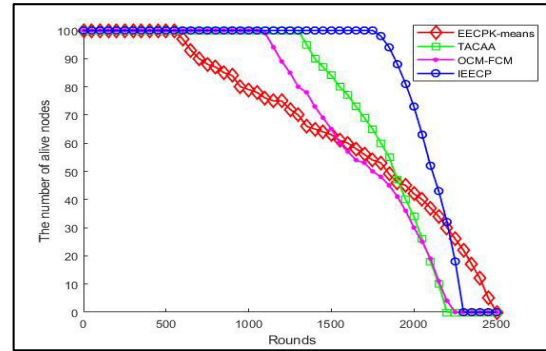
The effectiveness of IIECP over the existing protocols was evaluated by means of comparison with certain selected studies based on the clustering structure factors, where these protocols were known as the EECPK-means [25], OCM–FCM [10], and TACAA [14]. In this phase of the evaluation, two different scenarios were used to indicate the scalability of the proposed protocol, as follows:

- 100 nodes were deployed randomly in the  $100 \times 100\text{m}^2$  WSN sensing area size, as the literature widely utilizes this scenario.
- 1000 nodes were deployed randomly in the  $1000 \times 1000\text{m}^2$  WSN sensing area size, where in current times, the WSN includes a huge number of nodes that are deployed in large-scale areas. Other details are illustrated in Table 4.

Some of the measurement parameters used in this evaluation were as follows:

#### 1) NETWORK LIFETIME

Two measuring parameters based on the first node dies (FND), or the last node dies (LND) were used for the network lifetime. Numerous protocols rely on the FND to measure the network lifetime [18], [14], where it is occasionally called the stable period. Furthermore, there are certain protocols that

**FIGURE 9.** Number of live nodes of the first scenario.**TABLE 5.** Results based on the first scenario.

Protocol	FND	HND	LND	WFND
EECPK-Means	549	1840	2498	0.28
TACAA	1274	1865	2189	1.39
OCM-FCM	1062	1740	2250	0.893
IIECP	1744	2113	2292	3.18

are dependent upon the LND to measure the network lifetime as shown in [40].

There is a stable and also an unstable period of the network lifetime that is defined from the FND to LND. Therefore, we used the FND and LND in this evaluation, in addition to a new factor that showed the relationship between the stable and unstable periods, namely weighted first node dies (WFND). Whenever the WFND is high, the network has a good stable period. Consequently, for improved network performance, the highest value for FND and WFND should be achieved along with the high value for LND. The WFND was computed based on Eq. 39.

$$WFND = \frac{FND}{LND - FND} \quad (39)$$

The results were based on the first scenario; the number of alive nodes over time for different comparison protocols are illustrated in Figure 9 and Table 5.

The nodes of the EECPK-means protocol possessed the lowest lifetime according to FND, and the highest lifetime according to LND among the protocols, owing to a small value of the threshold of the CH rotation for this protocol. In the same context, TACAA and OCM–FCM protocols demonstrated better FND and WFND than the EECPK-means protocol, but with LND less than IIECP and EECPK-means protocols. The IIECP demonstrated great stability and maintains the lives of all nodes as long as possible compared with other protocols. Furthermore, the proposed protocol possessed the highest WFND among other protocols as shown in Table 5, where all the nodes in it died in a narrow period of time, thereby, reducing the unstable period.

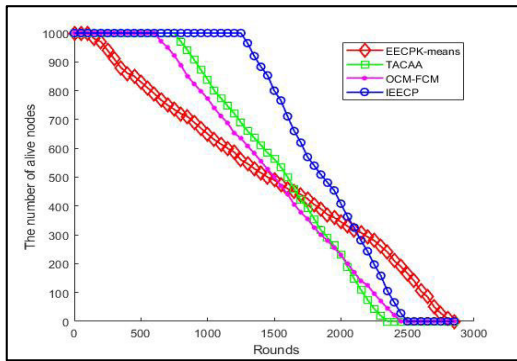


FIGURE 10. Number of live nodes of the second scenario.

TABLE 6. Results based on the second scenario.

Protocol	FND	HND	LND	WFND
EECPK-Means	123	1460	2854	0.045038
TACAA	786	1600	2347	0.503523
OCM-FCM	634	1492	2489	0.341779
IEECP	1282	1843	2509	1.044825

For the second scenario, the number of alive nodes over time for comparison with different protocols are illustrated in Figure 10 and Table 6. The results of this scenario are similar to the first one in terms of arranging the protocols for the FND and LND, where the effectiveness of IEECP over other protocols according to FND and WFND is demonstrated. Additionally, it is revealed that the EECPK-Means protocol has achieved a better LND.

The results of the second scenario indicated that the FND was reduced for all protocols even with the increase in the number of nodes in the network. This is due to the increase in the sensing area as compared with the first scenario where the number of relay points increased, signifying an increase in the overhead for the CHs, and an acceleration of their deaths. Moreover, there was an increase in the average cluster size compared with the first scenario, where the size of the message transmitted by the CH increased, signifying an increase in the energy consumption for CHs. In contrast, the LND for all protocols increased, as this is due to the increase in the total number of nodes that only impacted the extension of the network lifetime in terms of the LND metric.

Based on this experiment, there are significant differences in the lifetime among the baseline protocols according to the FND in both scenarios, as illustrated in Tables 5 and 6. The IEECP has the highest lifetime duration according to the FND and WFND, which extends the stability period of the network more than other protocols, and enhances the network performance. In contrast, there are no considerable differences in the lifetime span among the baseline protocols according to the LND.

TABLE 7. Result of paired t-test for the lifetime.

SCEN	Comparison of IEECP	t value	P value (one tailed)	Significance of Null Hypothesis	Confidence interval	
					lower	upper
1 <sup>st</sup>	EECPK-Means	4.00	0.003	< 0.05	51.0	83.6
	TAACA	2.20	0.027	< 0.05	50.8	94.5
	OCM-FCM	3.8	0.002	< 0.05	50.9	84.8
2 <sup>nd</sup>	EECPK-Means	2.80	0.005	< 0.05	338.6	703.4
	TAACA	1.86	0.037	< 0.05	353.9	737.9
	OCM-FCM	2.09	0.023	< 0.05	338.2	712.9

## 2) STATISTICAL INFERENCE FOR NETWORK LIFETIME

In this section, the statistical inference of the network lifetime for the baseline protocols in the above scenarios was illustrated by using the paired t-test, as shown in [41].

The network lifetime followed a bivariate normal distribution when pairwise comparison for the network lifetime was performed between IEECP with EECPK-Means, TACAA, and OCM-FCM, where the lifetime of the IEECP was coupled with each lifetime for the existing protocols as pair of (xi, yi) that xi was the lifetime for IEECP and yi was the lifetime for the baseline protocol.

The lifetime for the pairwise comparison of IEECP with EECPK-Means, TACAA, and OCM-FCM was supposed to be equal based on the Null Hypothesis (H0). In contrast, the Alternative Hypothesis (H1) showed that the lifetime of the IEECP was greater than the existing protocol.

The t-test was defined with n-1 degrees of freedom as follows:

$$t = \bar{x} / (\text{STD} / \sqrt{(n-1)}) \quad (40)$$

where  $\bar{x}$  and STD refers to the mean and standard deviation of the difference of the lifetime for two correlated sized of the Comparison-protocols.

In T-test, the p refers to the probability of the calculated value with n-1 degrees of freedom. If this value is less than 0.05, the Null Hypothesis is rejected at 5% significance level. Thus, the Alternative Hypothesis was accepted at 95% confidence level. Table 7 illustrated the results obtained by using t-test of the IEECP with EECPK-Means, TACAA, and OCM-FCM, respectively. In each case, the P value was less than 0.05 and the t value was less than t-critical, so the Null Hypothesis was rejected at 5 % significance level and Alternative Hypothesis was accepted at 95% confidence level, assuming that the lifetime of the IEECP was greater than the existing protocols.

## 3) ENERGY DISSIPATION IN INEFFICIENT NETWORK

We used the measures of the energy dissipation for the network to demonstrate the superiority of the proposed protocol

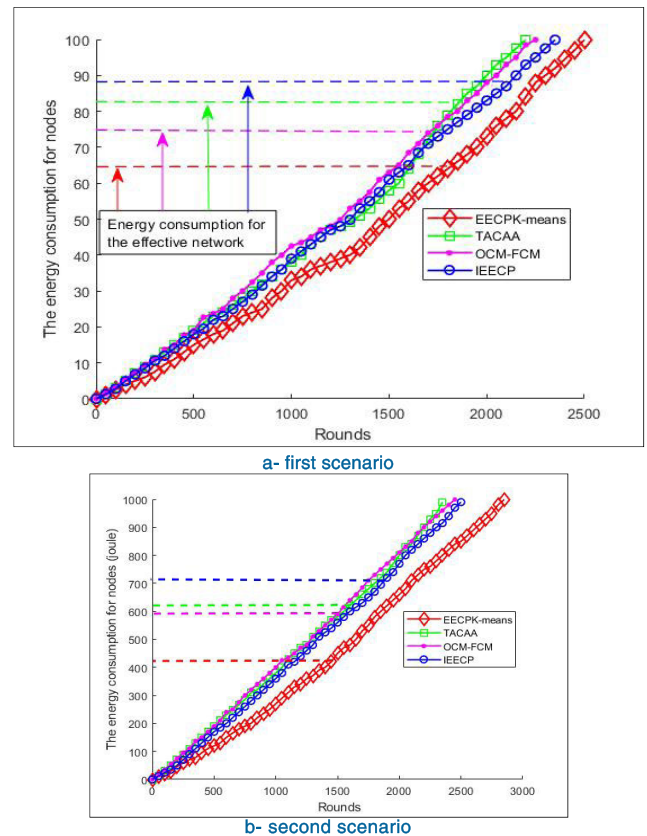
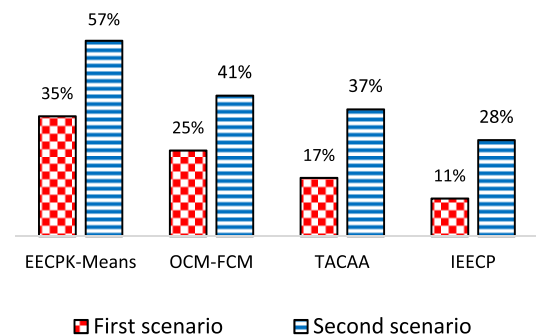
**TABLE 8.** Energy management over lifetime.

Protocol	HND occurred		Energy consumption till HND		Energy dissipation for ineffective-network	
	Scenario		Scenario		Scenario	
	1 <sup>st</sup>	2 <sup>nd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>	1 <sup>st</sup>	2 <sup>nd</sup>
EECPK-Means	1840	1460	65%	43%	35%	57%
TACAA	1865	1600	83%	63%	17%	37%
OCM-FCM	1740	1492	75%	59%	25%	41%
IEECP	2113	1843	89%	72%	11%	28%

in managing energy consumption over the related works. After half of the nodes die (HND), a major change in the network topology occurs [42], and the current topology of the network becomes ineffective relative to the energy consumption and greatly weakens the network performance in the next rounds. Accordingly, after the death of half of the network nodes, the energy consumed in this ineffective network is considered as energy dissipation. In other words, the sum of the original or initial energy for the nodes is not fully used in an effective network. However, part of this energy is wasted in a network that is unable to sense the entire monitoring area. Therefore, the energy of nodes needs to be managed properly to make the network as efficient as possible.

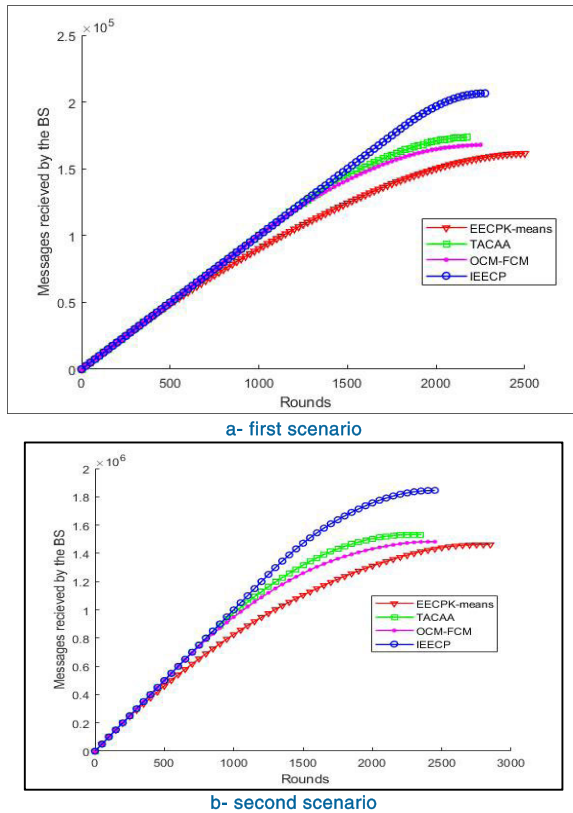
Based on Table 5 and 6, the HND for the EECPK-means protocol occurred in round 1840 for the first scenario and in round 1460 for the second scenario. Therefore, the energy consumption for the effective network (from the start of the network operation until the occurrence of HND) was 65% of the initial energy for nodes for the first scenario and 43% for the second scenario, as shown in Figure 11 and Table 8. Accordingly, the remaining energy for nodes, which were 35% and 57% of the initial energy had been consumed in an inefficient network for the first and second scenario, respectively, as shown in Figure 12. This ratio of the remaining energy was considered as the energy dissipation for the EECPK-means protocol, which showed the highest energy dissipation among the protocols i.e. it had the lowest energy management for nodes among protocols.

Similarly, the occurrence of the HND occurred in round 1740, and 1492 for the OCM-FCM protocol, as well as 1865 and 1600 for the TACAA protocol, within the first and second scenario respectively, as illustrated in Tables 5 and 6. For the OCM-FCM of these rounds, the energy consumptions for the network were 75% and 59% of the initial energy for nodes in both scenarios. Therefore, the energy dissipation were 25% and 41% of the initial energy for the first and second scenario, respectively, as shown in Figure 12. In terms of TACAA protocol, the energy consumptions for the network were 83% and 63% of the initial

**FIGURE 11.** Energy consumption for effective network for the first and second scenario.**FIGURE 12.** Energy dissipation in ineffective network at the first and second scenario.

energy for nodes in both scenarios. Therefore, the energy dissipation were 17% and 37% of the initial energy for the first and second scenario, respectively, as shown in Figure 12.

For IEECP, the HND occurred in rounds 2113 and 1843 for the first and second scenario, respectively. At these rounds, the energy consumption of the nodes was 89% of the initial energy for nodes in the first scenario and 72% in the second scenario, as shown in Figure 11 and Table 7. Accordingly, the remaining energy for nodes were 11% and 28% of the initial energy for nodes had been consumed in an inefficient network for the first and second scenario, respectively, as shown in Figure 12 and Table 7.



**FIGURE 13.** Number of messages received by BS.

Consequently, the energy of the nodes was efficiently consumed in the proposed protocol more than in other protocols prior to the network topology changes and became an ineffective network, where the proposed protocol had the lowest energy dissipation among the protocols. The reason behind this is that the proposed protocol has the highest lifetime according to FND and WFND, which results in the HND occurring at the later rounds compared with other protocols. As a result, the proposed protocol not only prolong the network lifetime duration but also efficiently utilizes and manages the energy of nodes to ensure that the network retains an efficient operation for as long as possible.

#### 4) NUMBER OF MESSAGES ARRIVED IN BS

The main goal of WSN was to sense the environment and then transmitted the sensed data to the BS. Therefore, when the BS received more sensing data, the performance of the network improved. Accordingly, the parameter of the number of messages received by the BS is used in this evaluation, where this parameter is important in clustering protocols such as the previous evaluation parameters [1].

As shown in Figure 13, the numbers of messages that arrived at the BS by utilizing the proposed protocol were similar to those utilizing EECPK-means, TACAA, and OCM-FCM until the 600<sup>th</sup> round for the first scenario, and 350<sup>th</sup> round for the second scenario.

Nevertheless, after the 600<sup>th</sup> and 350<sup>th</sup> rounds for the first and second scenario, respectively, the difference in the messages that arrived at the BS was shown among the protocols.

The highest number of messages that arrived at the BS were found in the proposed protocol, whereas the lowest number of messages were found in the EECPK-means protocol for both scenarios, although it had the highest lifetime duration according to LND. This is due to the prolonged stable period of the network which IIECP possesses, that is deemed advantageous more than other protocols.

As discovered from this evaluation, it can be seen that the EECPK-Means protocol performs better with regard to the LND. However, it has the lowest energy management for nodes and the number of messages received by the BS in both scenarios. Therefore, it is revealed that the LND parameter is not a sufficient indicator of good performance for the protocol in prolonging the network lifetime.

Accordingly, the proposed protocol shows the best performance among all the other protocols in all the evaluation parameters. This outcome is expected because the proposed protocol seeks to overcome the problems caused by inefficient and unbalanced energy consumption of nodes in the clustering protocol.

## VI. CONCLUSION

In this significant work, we propose an improved energy-efficient clustering protocol (IIECP) to prolong the lifetime of WSN-based IoT network through overcoming the problems of the clustering structure that adversely affect the protocol performance. Evidently, the proposed protocol reduces and balances the energy consumption of nodes by improving the clustering structure. Hence, the IIECP is deemed suitable for networks that require a longer lifetime. In general, the results yield that the IIECP performs better than the existing protocols. Our proposed protocol will be a beneficial contribution to the field that will enhance the daily operations in many areas of life, which utilize WSN in the IoT world. The energy consumption of the network is analyzed to compute the optimal number of clusters based on the distance to the CH in the case of the overlapping clusters. Then, the modified FCM algorithm (M-FCM) is proposed by combining it with a centralized mechanism to form static and balanced clusters. Finally, a new CH selection-rotation algorithm (CHSRA) is presented by integrating the back-off timer mechanism for the CH selection with the rotation mechanism for CH rotation. The CHSRA has relied on a new objective function for selecting CHs in optimal locations to balance the energy consumption among CHs for the clusters. Furthermore, it has relied on a new dynamic threshold for CH rotation within members of clusters to balance the energy consumption for the successive CHs in the cluster. In future work, we aim to enhance the protocol by improving the FCM algorithm concerning the random initial selection. Moreover, we believe that improving the objective function of CH selection through the reliance on weighted energy-based distance for adjacent CHs is also crucially significant. We anticipate that the future



clustering protocol can perform excellently when these limitations are taken into consideration.

## ACKNOWLEDGMENT

The authors would like to thank UTeM Zamalah Scheme for contributing in this work.

## REFERENCES

- [1] J. Shen, A. Wang, C. Wang, P. C. K. Hung, and C.-F. Lai, "An efficient centroid-based routing protocol for energy management in WSN-assisted IoT," *IEEE Access*, vol. 5, pp. 18469–18479, 2017.
- [2] V. Reddy and P. Gayathri, "Integration of Internet of Things with wireless sensor network," *Int. J. Electr. Comput. Eng.*, vol. 9, no. 1, pp. 439–444, 2019.
- [3] H. P. Gupta, S. V. Rao, A. K. Yadav, and T. Dutta, "Geographic routing in clustered wireless sensor networks among obstacles," *IEEE Sensors J.*, vol. 15, no. 5, pp. 2984–2992, May 2015.
- [4] Q. Wang, S. Guo, J. Hu, and Y. Yang, "Spectral partitioning and fuzzy C-means based clustering algorithm for big data wireless sensor networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–11, Dec. 2018.
- [5] S. Dehghani, B. Barekatin, and M. Pourzaferani, "An enhanced energy-aware cluster-based routing algorithm in wireless sensor networks," *Wireless Pers. Commun.*, vol. 98, no. 1, pp. 1605–1635, Jan. 2018.
- [6] A. Sheta and B. Solaiman, "Evolving a hybrid K-means clustering algorithm for wireless sensor network using PSO and GAs," *Int. J. Comput. Sci. Issues*, vol. 12, no. 1, pp. 23–32, 2015.
- [7] L. Kong, Q. Xiang, X. Liu, X.-Y. Liu, X. Gao, G. Chen, and M.-Y. Wu, "ICP: Instantaneous clustering protocol for wireless sensor networks," *Comput. Netw.*, vol. 101, pp. 144–157, Jun. 2016.
- [8] M. S. Tinker and S. Chinara, "Energy conservation clustering in wireless sensor networks for increased life time," in *Proc. 2nd Int. Conf. Adv. Comput. Commun. Eng.*, May 2015, pp. 7–10.
- [9] A. Bouyer, A. Hatamlou, and M. Masdari, "A new approach for decreasing energy in wireless sensor networks with hybrid LEACH protocol and fuzzy C-means algorithm," *Int. J. Commun. Netw. Distrib. Syst.*, vol. 14, no. 4, pp. 400–412, 2015.
- [10] S. Su and S. Zhao, "An optimal clustering mechanism based on fuzzy-C means for wireless sensor networks," *Sustain. Comput., Inf. Syst.*, vol. 18, pp. 127–134, Jun. 2018.
- [11] S. B. Shah, Z. Chen, F. Yin, I. U. Khan, and N. Ahmad, "Energy and interoperable aware routing for throughput optimization in clustered IoT-wireless sensor networks," *Future Gener. Comput. Syst.*, vol. 81, pp. 372–381, Apr. 2018.
- [12] 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. Bioinf.*, vol. 14, no. 1, pp. 76–84, Jan. 2017.
- [13] H. El Alami and A. Najid, "CFFL: Cluster formation using fuzzy logic for wireless sensor networks," in *Proc. IEEE/ACS 12th Int. Conf. Comput. Syst. Appl. (AICCSA)*, Nov. 2015, pp. 1–6.
- [14] A. Jain, "Traffic aware channel access algorithm for cluster based wireless sensor networks," *Wireless Pers. Commun.*, vol. 96, no. 1, pp. 1595–1612, Sep. 2017.
- [15] C. Gherbi, Z. Aliouat, and M. Benmohammed, "A novel load balancing scheduling algorithm for wireless sensor networks," *J. Netw. Syst. Manage.*, vol. 27, no. 2, pp. 430–462, Apr. 2019.
- [16] Z. Xu, L. Chen, C. Chen, and X. Guan, "Joint clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks," *IEEE Internet Things J.*, vol. 3, no. 4, pp. 520–532, Aug. 2016.
- [17] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annu. Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, p. 10.
- [18] S. A. Sert, A. Alchihabi, and A. Yazici, "A two-tier distributed fuzzy logic based protocol for efficient data aggregation in multihop wireless sensor networks," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 6, pp. 3615–3629, Dec. 2018.
- [19] Z. M. Zahedi, R. Akbari, M. Shokouhifar, F. Safaei, and A. Jalali, "Swarm intelligence based fuzzy routing protocol for clustered wireless sensor networks," *Expert Syst. Appl.*, vol. 55, pp. 313–328, Aug. 2016.
- [20] M. Shokouhifar and A. Jalali, "Optimized Sugeno fuzzy clustering algorithm for wireless sensor networks," *Eng. Appl. Artif. Intell.*, vol. 60, pp. 16–25, Apr. 2017.
- [21] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 660–670, Oct. 2002.
- [22] M. T. Rahama, M. Hossen, and M. M. Rahman, "A routing protocol for improving energy efficiency in wireless sensor networks," in *Proc. 3rd Int. Conf. Electr. Eng. Inf. Commun. Technol. (ICEEICT)*, Sep. 2016, pp. 1–6.
- [23] A. Thakkar and K. Kotecha, "Cluster head election for energy and delay constraint applications of wireless sensor network," *IEEE Sensors J.*, vol. 14, no. 8, pp. 2658–2664, Aug. 2014.
- [24] T.-T. Huynh, A.-V. Dinh-Duc, and C.-H. Tran, "Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks," *J. Commun. Netw.*, vol. 18, no. 4, pp. 580–588, Aug. 2016.
- [25] A. Ray and D. De, "Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network," *IET Wireless Sensor Syst.*, vol. 6, no. 6, pp. 181–191, Dec. 2016.
- [26] Y. Zhang, L. Li, Y. Bi, and Y. Chen, "Distributed data fusion based on fuzzy C-means for wireless sensor network," *J. Comput. Inf. Syst.*, vol. 8, no. 22, pp. 9167–9174, 2012.
- [27] R. Dutta, S. Gupta, and M. K. Das, "Low-energy adaptive unequal clustering protocol using fuzzy C-means in wireless sensor networks," *Wireless Pers. Commun.*, vol. 79, no. 2, pp. 1187–1209, Nov. 2014.
- [28] H. Mostafaei, "Energy-efficient algorithm for reliable routing of wireless sensor networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5567–5575, Jul. 2019.
- [29] H. El Alami and A. Najid, "ECH: An enhanced clustering hierarchy approach to maximize lifetime of wireless sensor networks," *IEEE Access*, vol. 7, pp. 107142–107153, 2019.
- [30] M. Elshrkawey, S. M. Elsherif, and M. Elsayed Wahed, "An enhancement approach for reducing the energy consumption in wireless sensor networks," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 30, no. 2, pp. 259–267, Apr. 2018.
- [31] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973.
- [32] A. Shokrollahi and B. M.-N. Maybodi, "An energy-efficient clustering algorithm using fuzzy C-means and genetic fuzzy system for wireless sensor network," *J. Circuits, Syst. Comput.*, vol. 26, no. 1, pp. 1–22, 2017.
- [33] J. Qin, W. Fu, H. Gao, and W. X. Zheng, "Distributed k-means algorithm and fuzzy C-means algorithm for sensor networks based on multiagent consensus theory," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 772–783, Mar. 2017.
- [34] N. Shivappa and S. S. Manvi, "Fuzzy-based cluster head selection and cluster formation in wireless sensor networks," *IET Netw.*, vol. 8, no. 6, pp. 390–397, Nov. 2019.
- [35] A. A. Hassan, W. Shah, M. Fairuz, and I. Othman, "Evaluate the performance of K-means and the fuzzy C-means algorithms to formation balanced clusters in wireless sensor networks," *Int. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 1515–1523, 2020.
- [36] Z. Xu, L. Chen, L. Cao, T. Liu, D. Yang, and C. Chen, "DARC?: A distributed and adaptive routing protocol in cluster-based wireless sensor networks," *Int. J. Distrib. Sens. Netw.*, vol. 11, no. 12, pp. 1–14, 2015.
- [37] S. Fang, S. M. Berber, and A. K. Swain, "An overhead free clustering algorithm for wireless sensor networks," in *Proc. IEEE GLOBECOM-IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 1144–1148.
- [38] M. Baniata and J. Hong, "Energy-efficient unequal chain length clustering for wireless sensor networks in smart cities," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–12, Sep. 2017.
- [39] V. Pal, G. Singh, and R. P. Yadav, "Balanced cluster size solution to extend lifetime of wireless sensor networks," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 399–401, Oct. 2015.
- [40] S. A. Sert, A. Yazici, and T. Dokeroglu, "Fuzzy processing in surveillance wireless sensor networks," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2016, pp. 1509–1515.
- [41] S. Ghosh, S. Mondal, and U. Biswas, "Fuzzy C means based hierarchical routing protocol in WSN with ant colony optimization," in *Proc. 2nd Int. Conf. Appl. Theor. Comput. Commun. Technol. (iCATccT)*, Jul. 2016, pp. 348–354.

- [42] 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, Mar. 2016.



lar ad hoc networks, clustering methods, and wireless sensor networks.

**ALI ABDUL-HUSSIAN HASSAN** received the B.Sc. degree in communication technical engineering, and the M.Sc. degree in information and communication technology from Universiti Utara Malaysia (UUM), in 2004 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Universiti Teknikal Malaysia Melaka (UTeM). He focuses on energy-efficient clustering protocols in the wireless sensor network. His research interests include mobile ad-hoc networks, vehicular



published in statistics, operations research, clustering analysis, mathematical statistics, statistical inference, and survival systems. He held many administrative positions since 1979. He is currently the Vice President of Al-Zahraa University for Women. He is currently a Professor in the statistics field.

**ABDUL-HUSSIEN HASSAN HABEB** received the B.Sc. degree in statistics from the University of Baghdad, Iraq, in 1976, the PGDip. degree from the University of Technology, Iraq, in 1980, the master's degree in statistics from the University of Baghdad, Iraq, in 1989, and the Ph.D. degree in statistics from the University of Baghdad, Iraq, in 2000. He is involved in several research projects. He has supervised numerous postgraduate students. He has many books and research papers



in Information Security, IT governance and management, and other related topics in computer networks and computer security.

**MOHD FAIRUZ ISKANDAR OTHMAN** (Member, IEEE) received the Ph.D. degree in information technology from the Queensland University of Technology and the master's degree in internetworking from the University of Technology, Sydney. He is currently a Senior Lecturer with the Department of Computer Systems and Communication, Faculty of Information and Communication Technology, FTMK (UTeM). His research interests include human behavioral issues



research interests include system and networking, network security, and IoT-related technology.

**WAHIDAH MD SHAH** received the B.Tech. degree from Universiti Utara Malaysia, the master's degree in computer science from Universiti Teknologi Malaysia, and the Ph.D. degree in computer science from Lancaster University, U.K. She is currently a Senior Lecturer with the Department of Computer Systems and Communication, Universiti Teknikal Malaysia Melaka. She is a member of the Information Security, Digital Forensic, and Computer Networking Research Group. Her



machine learning, deep learning, and image processing.

**MOHAMMED NASSER AL-MHIQANI** (Member, IEEE) received the B.Sc. degree in computer science (computer networking) and the M.Sc. degree in computer science (internetworking technology) from Universiti Teknikal Malaysia Melaka (UTeM), in 2014 and 2015, respectively, where he is currently pursuing the Ph.D. degree. His research interests include cybersecurity, cyber-physical system security, and insider threats in cybersecurity, wireless sensor network,

...