An Energy-Efficient Clustering Protocol for the Lifetime Elongation of Wireless Sensors in IoT Networks

Ali Mohammed Kadhim Abdulzahra and Ali Kadhum M. Al-Qurabat

Abstract The efficient use of energy is a crucial difficulty that must be addressed when constructing a wireless sensor network (WSN)-based Internet of Things. Thousands of battery-powered tiny devices known as sensors make up these networks. Sensors are limited-resource devices with a finite amount of energy. The lifetime of the entire network may be extended significantly by reducing the energy consumption of these nodes. WSN-based IoT clustering is a rapidly growing field of study. The main problems in clustered WSN-based IoT are determining the right number of clusters and then picking a cluster head (CH) node inside every created cluster. In this paper, we introduce a unique clustering approach for WSN-based IoT systems based on fuzzy c-means (FCM). The approach employs a FCM methodology to form the clusters and a decrease in the total energy used on each cluster to identify the optimal cluster head (CH). Instead of changing CHs with dynamic clustering at each period, we aim to apply an energy threshold to postulate CH dynamicity based on current energy levels, therefore extending the sensor network life span.

Keywords IoT · WSN · Clustering · Energy consumption · Fuzzy c-means

1 Introduction

The considerable advancement in the field of the Internet of Things (IoT) in the current modern era has made human existence more evolved (Al-Qurabat and Abdulzahra 2020). It has improved the standard of living for all people. Along with the development of WSNs, the IoT idea was broadened. A WSN is made up of N sensor nodes (SNs) that are scattered at random throughout a geographic region

A. M. K. Abdulzahra · A. K. M. Al-Qurabat (✉)
Department of Computer Science, College of Science for Women, University of Babylon, Babylon, Iraq
e-mail: alik.m.alqurabat@uobabylon.edu.iq

A. M. K. Abdulzahra
e-mail: ali.abdulzahra.gsci11@student.uobabylon.edu.iq

(Abdulzahra et al. 2021). These sensors collect information about the surroundings, including humidity, acoustics, light, vibration, and temperature. The data collected is sent to a sink node, which is a base station (BS). A WSN-based IoT application is made up of a collection of self-contained sensors that may collect data from their surroundings in order to create a general overview of the controlled region (Idan Saeedi and Al-Qurabat 2021).

The transmission of data is the most energy-intensive function for SNs in WSNs. As a result, fewer data transfers or lower transmission power are necessary to decrease energy consumption. Furthermore, a number of requirements in the network's architecture and operation must be satisfied in order for WSNs to be employed. Because SNs have a fixed quantity of energy, energy conservation is typically seen as the most important difficulty in maintaining network connectivity and prolonging the SN's life span, especially if the deployment region is harsh or hostile and the battery is not replaceable (Idrees and Al-Qurabat 2021; Saeedi and Al-Qurabat 2022).

Clustering is one viable option for addressing these issues and making the most of available energy. Clustering, which splits the network into clusters and forces SNs in each cluster to transmit data to a cluster head (CH), is responsible for this (Panchal and Singh 2021). Because the sensors are so near to the CHs, they may lower their transmission powers, reducing energy usage and increasing the network's life span. CHs are selected from among the SNs to oversee the collection of data from sensors within their clusters, aggregate it, and transfer it to the BS (Gantassi et al. 2020).

We offer a unique clustering protocol based on fuzzy c-means with distance- and energy-limited termed (FCMDE) for clustering, CH selection, and data transfer to increase the life span of WSN-based IoT. Instead of picking the node closest to the fuzzy c-means centroid as CH, as in earlier research (Panchal and Singh 2021; Qin et al. 2017), FCMDE picks the node closest to the majority of nodes in the network. The proximity criterion ensures that nodes in each cluster remain close to their CH at all times, allowing them to broadcast at much lower power levels. We aim to apply an energy threshold to hypothesis the dynamicity of CH based on current energy levels rather than substituting CHs for dynamic clustering at each period in this investigation. The change in how the CH is chosen has a significant impact on the network's energy consumption.

The remaining portions of the paper are described below. Section 2 includes the related works. Section 3 briefly introduces the network concept and energy consumption model. Section 4 contains a full overview of the suggested protocol. The simulation findings and discussions are presented in Sect. 5. Section 6 outlines the paper's conclusion.

2　Related Works

The sensor nodes in the network may be configured to operate as CH either centrally or distributedly. The first uses a BS to control CH selection, but the second is completely self-contained. Distributed protocols include low-energy adaptive clustering hierarchy (LEACH) (Heinzelman et al. 2000), hybrid energy-efficient distributed clustering (HEED) (Younis and Fahmy 2004), and others. Machine learning is increasingly being utilized to divide the network into clusters, from which CHs are selected depending on predefined criteria. This may be performed using algorithms like k-means (Gantassi et al. 2020; Qin et al. 2017; Cai et al. 2019) and fuzzy c-means (Panchal and Singh 2021; Qin et al. 2017), which are becoming more popular in WSNs, IoT, and crowd-sensing applications.

To deal with the uncertainty in WSNs, (Bhajantri and Sutagundar 2017; Ahamad and Kumar 2017) used clustering methods based on fuzzy logic. The authors in Bhajantri and Sutagundar (2017) introduced data processing and clustering for WSNs based on fuzzy logic. This method takes into account each node's energy level, bandwidth, and connection efficiency. The suggested work aims to increase network performance regarding the lifetime of the network, the number of live nodes, CH selection time, throughput, and energy usage. The authors in Ahamad and Kumar (2017) presented an energy-efficient clustering technique based on the fuzzy logic system to extend the WSN life span in a probabilistic approach model. With the support of an efficient CH selection approach, this effectively tackles the issue of low sensor node residual energy usage. The authors of Usha Kumari and Padma (2019) recommended three protocols: enhanced DEEC (EDEEC), developed DEEC (DDEEC), and DEEC. The rates of CH energy minimization and network longevity were investigated for each clustering strategy. In terms of sensor network longevity, the EDEEC protocol surpasses both the DECC and DDEEC protocols, according to the data.

In Rahimi and Chrysostomou (2019), the authors introduced a load balancing method that works intelligently dependent on a controller of fuzzy logic and the queue of priority to reduce and disseminate energy usage, resulting in an improvement in network life span. In Wang et al. (2018), the authors firstly offer an analytical approach for determining the ideal number of clusters in a WSN. Next, they present a centralized clustering approach dependent on the spectral division method. Following that, they offer a decentralized solution to the clustering technique based on the fuzzy c-means approach. Eventually, they ran extensive simulations, and the findings revealed that the suggested approaches beat the HEED clustering method in regards to energy cost and network longevity. The paper's authors Tarhani et al. (2014) suggest a Scalable Energy-Efficient Clustering Hierarchy (SEECH) for selecting CHs. High-degree SNs are categorized as CHs in this approach, whereas low-degree SNs are used as relays. It employs a distance-based method to assess the homogeneity of CHs for balancing clusters. As compared to the LEACH and TCAC methods, the suggested algorithm shows improved SEECH protocol performance in terms of sensor network life span.

3   Preliminaries

The energy consumption model and the network model are presented in this section.

3.1   Network Model

We provide a standard IoT monitoring environment for WSN-based IoT applications in this section. To assure the system's energy efficiency, we adopt a cluster-based architecture. The BS is at the center of a square sensing field with N randomly dispersed sensor nodes. The nodes continuously monitor the environment and report their results to the CH, which periodically transfers the data gathered to the BS (also known as the gateway). We make the following assumptions for our network model:

- The topology of the network remains static throughout the network operation.
- Sensor nodes based on the IoT are deployed in a uniform pattern but at random.
- The sensor nodes are all homogeneous.
- All sensor nodes are energy-restricted and start with the same amount of energy.
- The BS is supposed to be free of energy, computation, and network coverage limitations.
- Radio interference, as well as any obstruction or signal attenuation caused by the existence of physical objects, are not taken into account.
- We believe that the suggested protocol is extremely secure. This work's security considerations are outside the scope of this paper.

3.2   Energy Model

Sensor nodes need energy for remaining awake, network maintenance, data processing, packet receiving, packet transmission, and sensing, among other things. The amount of energy required to send a packet is proportional to the size of the packet and the distance traveled (Heinzelman et al. 2000; Al-Qurabat et al. 2021; Al-Qurabat 2022). The transmitter demands a quantity of energy to send an $w - $ bit packet across a distance of d, as given:

$$E_{\text{TX}}(w, d) = \begin{cases} E_{\text{elec}} \times w + \epsilon_{\text{fs}} \times w \times d^2 & \text{if } d < d_0 \\ E_{\text{elec}} \times w + \epsilon_{\text{mp}} \times w \times d^4 & \text{if } d \geq d_0 \end{cases} \qquad (1)$$

Receiving a $w - $ bit packet consumes the following amount of energy:

$$E_{\text{RX}}(w) = E_{\text{elec}} \times w \qquad (2)$$

The energy wasted per bit by the receiver or transmitter circuits is denoted by E_{elec} in (1) and (2). In a free space model and a multi-path fading channel model, we utilize ϵ_{fs} and ϵ_{mp}, respectively, to describe the energy usage of the amplifier per bit. The distance between the receiver and transmitter is indicated by the letter d. The d_0 threshold is formulated as having

$$d_0 = \sqrt{\frac{\epsilon_{\text{fs}}}{\epsilon_{\text{mp}}}} \tag{3}$$

The data aggregation power consumption, which is denoted as E_{da}, is another factor that is considered. We suppose that each cluster member delivers $w -$ bit packet to its CH during each period of data collection, and that the energy spent by a CH during one period of collecting data may be expressed as

$$E_{\text{CH}} = \frac{N}{c} \times E_{\text{elec}} \times w + \frac{N}{c} \times E_{\text{da}} \times w + \epsilon_{\text{mp}} \times w \times d_{\text{BS}}^4 \tag{4}$$

The CH wastes energy by collecting packets from nodes, aggregating them, and transferring the resulting packets to the BS. The number of clusters is given by c, while the average distance between a CH and a BS is given by d_{BS}.

4 The FCMDE Protocol

There are four phases to implementing the FCMDE protocol. The first phase is to decide on the best number of clusters to use. The fuzzy c-means method is used in the second phase to propose a centralized clustering technique. The third phase is CH selection. This phase considers the nodes' remaining energy as well as their cluster's position (with regard to other nodes). Instead of using the sensor node closest to the centroid, FCMDE uses a novel measure in which the sensor node closest to all other nodes is chosen as the CH. Transmitting data between nodes and CHs in the clusters is the last phase.

4.1 The Optimal Number of Clusters

Since the quantity of inter-cluster communication rises with c, determining the ideal number c of clusters is crucial. However, when c is lower, the number of intra-cluster communications increases considerably. Using the silhouette coefficient (SC) or silhouette score approach Younis and Fahmy (2004), we will determine the ideal number of clusters as in the following:

$$\mathrm{SC}(n_i) = \frac{b(n_i) - a(n_i)}{\max\{a(n_i), b(n_i)\}}, \tag{5}$$

where $SC(n_i)$ is the silhouette coefficient of the sensor node n_i; $a(n_i)$ denotes the average intra-cluster distance, that is, the average distance between sensor node n_i and all other sensor nodes in the cluster to which n_i belongs. The minimal average inter-cluster distance between sensor node n_i and all clusters to which n_i does not belong is denoted by $b(n_i)$.

The SC's value ranges from $[-1, 1]$. A score of 1 indicates that the sensor node is highly compact inside the cluster to which it belongs and is far distant from the other clusters. The poorest possible value is -1. Near-zero values indicate overlapping clusters.

4.2 FCM Clustering

To split the network into a fixed optimal number of clusters, we suggest a centralized clustering algorithm based on the fuzzy c-means (Idrees and Al-Qurabat 2021) approach in this section. We presume that the sink node is fully aware of the network architecture. The sink node connects all CHs and separates the sensor nodes into c clusters: $C_1, C_2, \ldots, C_c$. The goal of this protocol's cluster creation is to decrease the following objective function:

$$J_{\min} = \sum_{i=1}^{c} \sum_{j=1}^{N} u_{ij}^m d_{ij}^2, \tag{6}$$

where u_{ij} is the degree of membership to cluster i of sensor node n_j, d_{ij} denotes the distance between sensor node n_j and the cluster i's center point. With the actual parameter $m > 1$, the degree u_{ij} of sensor node n_j with regard to the cluster is determined and fuzzyfied as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{m-1}}} \tag{7}$$

In addition, the cluster center is being upgraded utilizing:

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m n_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{8}$$

The FCM-based clustering algorithm's behavior is determined by the clusters' number c in addition to the sensor nodes' number. Because the number of functioning

sensor nodes fluctuates from period to period, clustering occurs at the start of each period. During the clustering phase, the following activities are taken:

1. Set the clusters' number to c.
2. Assign c initial cluster centers at random.
3. Use (7) to compute the matrix of membership.
4. Use (8) to compute the center of the cluster.
5. Steps 3 and 4 should be repeated until equilibrium is achieved (fixed centers).

The FCM algorithm associates the cluster center coordinates with their sensor members; only the sensor node's membership is taken into account by our protocol.

4.3 CH Selection

Clustering is conducted prior to CH selection in this study to decrease the energy consumed in the process of cluster creation. Two factors must be met by the CH selection policy.

4.3.1 *Position Inside the Cluster*

Rather than the node closest to the cluster's center, the CH is chosen based on its proximity to the most other nodes. Because the aim of the proposed protocol is to minimize the energy required by sensors for sending to the CH, rather than to choose the node at the cluster's center, this requirement, which we call the proximity criterion, is more beneficial than closeness of the possible CH to the cluster's center. We develop a cost function, λ, that calculates the Euclidean distance between the selected node and all other in-cluster nodes to discover the sensor that is closest to the most other nodes and costs the least amount of energy to broadcast to inside its cluster.

$$\lambda = \sum_{j=1}^{c} \sum_{n_i \in C_j} d(n_i, X_j), \tag{9}$$

where n_i denotes the ith node in the network, X_j indicates the centroid of the sensor nodes in a specific cluster, and C_j comprises N_j nodes and the Euclidean distance $d(n_i, X_j)$ is given by

$$d(n_i, X_j) = \|n_i - X_j\|^2. \tag{10}$$

4.3.2 The Level of Residual Energy

A node's remaining energy should be over a certain threshold E_{TH} in order for it to be considered for CH selection. This criterion is required to prevent the CH from dying too soon, resulting in the network being disconnected. For every CH selection, we create an energy-related cost function that describes the toenergy usage for all sensor cluster members. When sensor node X_i is designated as CH, the energy-related cost functions $E_{Co}(CH_i)$ for each cluster of n_c sensonodes are expressed as the total consumed energy of all sensor nodes.

$$E_{Co}(CH_i) = \sum_{j=1, j \neq i}^{n_c} E_{TX}(x_j \rightarrow X_i) + (n_c - 1) \times E_{RX}(X_i) + E_{TX}(X_i \rightarrow BS),$$

$$(11)$$

where

- $E_{TX}(x_j \rightarrow X_i)$: The amount of energy expended by sensor node x_j to send a data packet to CH X_i.
- $E_{RX}(X_i)$: The amount of energy utilized by the CH X_i when it receives a data packet from a sensor node.
- $E_{TX}(X_i \rightarrow BS)$: The amount of energy expended by CH node X_i during the transmission of the aggregated data packet to the BS.

The BS calculates the cost functions relating to energy and closeness for each cluster and chooses the node with the lowest $E_{Co}(CH)$ and λ as the CH. The BS transmits a packet of information to every node in the network once the CH has elected, which includes the CH ID, and the cluster ID. Following the CH election, the BS changes the state of the nodes in its system (energy levels of nodes).

Following the completion of the initial configuration of the network, the CHs in the next period will compare their energy levels ($E_{CH-Th}(i)$) to the energy threshold function E_{TH}. The CH can maintain intra-cluster communication with member nodes of a cluster if the present CH remaining energy levels ($E_{CH-Th}(i)$) are equivalent or higher than the energy threshold level E_{TH}; otherwise, the CH must discontinue and demand the creation of a new cluster. Therefore, the CH could remain without change for consecutive periods until its residual energy falls below the threshold.

5   Transmission of Data

The sensor nodes begin transmitting data to the CHs when the CHs are identified. Due to the obvious shortest geographic distance to the CHs attained by the FCM algorithm, the transmitting power of cluster member nodes is enhanced. The CHs perform data aggregation, reducing the quantity of data and then transmit the aggregated data to the BS.

6 Simulation and Performance Evaluation

The simulation findings used to assess our suggested protocol (FCMDE) are presented in this section. The FCMDE supposes that the sensor nodes are deployed in a 1000×1000 m region for building the network model. The sink node is positioned in the center of the network (500, 500). Table 1 lists the factors that were utilized in the simulations. The FCMDE protocol is compared to the DDEEC (Qin et al. 2017) and SEECH (Tarhani et al. 2014) protocols in four aspects: network lifetime, throughput, and energy consumption.

It is critical that all sensor nodes remain operational as long as feasible since the performance of the network suffers when a node dies. As a result, knowing the death time of the first node is critical. The period of time during which the network's first node dies is described as the network's lifetime. First-SN, Half-SN, and Last-SN (periods' number during which the network's first, half, and last node die, respectively) are all being included in the study. Figure 1 shows the comparison of FCMDE with SEECH and DDEEC simulation results.

Table 1 Simulation factors

Factors	Value
WSN size	1000×1000 m
SNs no	100
Initial energy	100 J
E_{da}	5 nJ/bit/signal
ϵ_{fs}	10 pJ/bit/m^2
ϵ_{mp}	0.0013 pJ/bit/m^4
E_{elec}	50 nJ/bit
Packet size	512 bytes

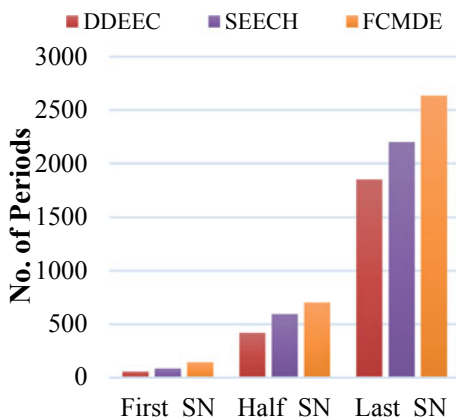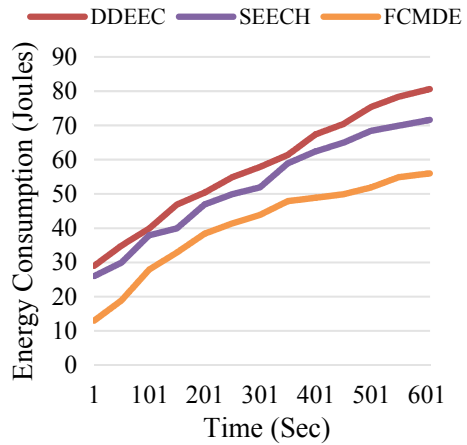**Fig. 1** First-SN, Half-SN, and Last-SN stages of the network

Fig. 2 Energy consumption
of the network

It is shown by the results obtained in Fig. 1; the suggested protocol (FCMDE) has a first-SN enhancement of roughly 250% and 168% when compared to DDEEC and SEECH, respectively. The Half-SN and Last-SN are also superior in comparison.

In the next experiment, the FCMDE protocol investigates how much energy is lost on average inside the network. Consumption of energy is among the ultimate important factors to consider when determining WSN's effectiveness. Figure 2 compares the suggested FCMDE protocol to the DDEEC and SEECH strategies in terms of energy usage. Experiment results demonstrate that the energy consumption during every period was already lowered. The FCMDE protocol uses roughly 10 and 18 J throughout the transmission of data, respectively, which is less than the SEECH and DDEEC protocols. The findings demonstrate that the FCMDE protocol performs best and saves more energy in comparison with the other two protocols.

Another simulation experiment was conducted to assess the network's throughput. Throughput is defined as the ratio of the packets that the CH acknowledges to the delay of the communication of packets in the process of transmitting, which defined as

$$\text{Throughput} = \frac{\text{total No. of packets received by CH}}{\text{delay in process of communication}} \qquad (12)$$

The analysis of the throughput of the suggested FCMDE protocol compared to SEECH and DDEEC protocols is shown in Fig. 3. When compared to the DDEEC and SEECH protocols, the quantity of packets sent to the CH in the suggested FCMDE protocol is 22% and 13% faster, respectively. As a result, as compared to previous techniques, throughput measuring has grown over time.

Fig. 3 Throughput of the network
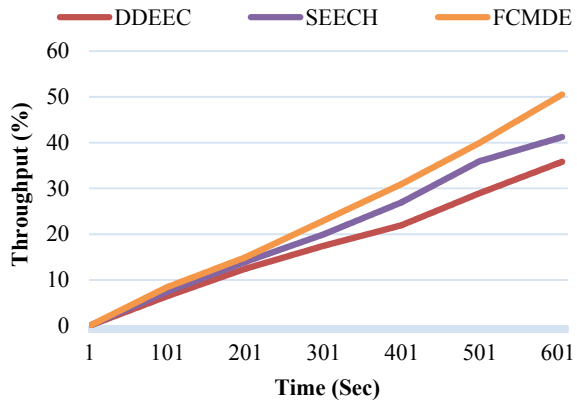
7 Conclusions

In this study, FCMDE was introduced as a clustering protocol for WSN-based IoT. The suggested FCMDE reduces energy drain and increases longevity while minimizing overhead costs. FCDME chooses a CH during clustering by combining fuzzy c-means, node location, and residual power. In an attempt to reduce transmission overhead costs and unnecessary CH changes for every transmission period, FCMDE uses functions of thresholds, namely the threshold of energy and the proximity criterion. The effectiveness of the suggested FCMDE protocol has been demonstrated through thorough simulation using a variety of possible assessment performance indicators. Average energy usage, network longevity, and throughput are all examples of these indicators. A comparison analysis of SEECH and DDEEC procedures was also conducted, demonstrating the superiority of the FCMDE approach.

References

S.A. Abdulzahra, A.K.M. Al-Qurabat, A.K. Idrees, Compression-based data reduction technique for IoT sensor networks. Baghdad Sci. J. **18**(1), 184–198 (2021)

F. Ahamad, R. Kumar, Energy efficient region based clustering algorithm for WSN using fuzzy logic, in *2016 IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2016—Proceedings* (2017)

A.K.M. Al-Qurabat, A lightweight huffman-based differential encoding lossless compression technique in IoT for smart agriculture. Int. J. Comput. Digit Syst. **11**(1), 117–127 (2022)

A.K.M. Al-Qurabat, S.A. Abdulzahra, An overview of periodic wireless sensor networks to the internet of things, in *IOP Conference Series: Materials Science and Engineering*, vol. 928, issue 3 (2020)

A.K.M. Al-Qurabat, Z.A. Mohammed, Z.J. Hussein, Data traffic management based on compression and MDL techniques for smart agriculture in IoT. Wirel. Pers. Commun. **120**(3) (2021)

L.B. Bhajantri, A.V. Sutagundar, Fuzzy logic based cluster head selection and data processing in distributed sensor networks, in *International Conference on Computing, Analytics and Security Trends, CAST 2016* (2017)

S. Cai, Y. Zhu, T. Wang, G. Xu, A. Liu, X. Liu, Data collection in underwater sensor networks based on mobile edge computing. IEEE Access **7** (2019)

R. Gantassi, B. Ben Gouissem, J. Ben Othmen, Routing protocol LEACH-K using K-Means algorithm in wireless sensor network. Adv. Intell. Syst., 1150 AISC (2020)

W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *Proceedings of the Hawaii International Conference on System Sciences* (2000)

I.D. Idan Saeedi, A.K.M. Al-Qurabat, A systematic review of data aggregation techniques in wireless sensor networks. J. Phys. Conf. Ser. **1818**(1) (2021)

A.K. Idrees, A.K.M. Al-Qurabat, Energy-efficient data transmission and aggregation protocol in periodic sensor networks based fog computing. J. Netw. Syst. Manag. **29**(1) (2021)

A. Panchal, R.K. Singh, EHCR-FCM: energy efficient hierarchical clustering and routing using fuzzy C-means for wireless sensor networks. Telecommun. Syst. **76**(2) (2021)

J. Qin, W. Fu, H. Gao, W.X. Zheng, Distributed k-means algorithm and fuzzy c-means algorithm for sensor networks based on multiagent consensus theory. IEEE Trans. Cybern. **47**(3) (2017)

P. Rahimi, C. Chrysostomou, Improving the network lifetime and performance of wireless sensor networks for IoT applications based on fuzzy logic, in *Proceedings—15th Annual International Conference on Distributed Computing in Sensor Systems, DCOSS 2019* (2019)

I.D.I. Saeedi, A.K.M. Al-Qurabat, Perceptually important points-based data aggregation method for wireless sensor networks. Baghdad Sci. J. **19**(4), 875–886 (2022)

M. Tarhani, Y.S. Kavian, S. Siavoshi, SEECH: scalable energy efficient clustering hierarchy protocol in wireless sensor networks. IEEE Sens. J. **14**(11) (2014)

C. Usha Kumari, T. Padma, Energy-efficient routing protocols for wireless sensor networks. Adv. Intell. Syst. Comput. **898** (2019)

Q. Wang, S. Guo, J. Hu, Y. Yang, Spectral partitioning and fuzzy C-means based clustering algorithm for big data wireless sensor networks. Eurasip. J. Wirel. Commun. Netw. **2018**(1) (2018)

O. Younis, S. Fahmy, HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans. Mob. Comput. **3**(4) (2004)