Received: 4 April 2017 | Revised: 5 February 2018 | Accepted: 23 February 2018
DOI: 10.1002/dac.3577

RESEARCH ARTICLE

WILEY

An intelligent and knowledge-based overlapping clustering protocol for wireless sensor networks

Sohrab Khanmohammadi[1] | Mohammad Samadi Gharajeh[2]

[1] Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

[2] Young Researchers and Elite Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran

Correspondence
Mohammad Samadi Gharajeh, Young Researchers and Elite Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran.
Email: m.samadi@iaut.ac.ir

Summary

Overlapping is one of the topics in wireless sensor networks that is considered by researchers in the last decades. An appropriate overlapping management system can prolong network lifetime and decrease network recovery time. This paper proposes an intelligent and knowledge-based overlapping clustering protocol for wireless sensor networks, called IKOCP. This protocol uses some of the intelligent and knowledge-based systems to construct a robust overlapping strategy for sensor networks. The overall network is partitioned to several regions by a proposed multicriteria decision-making controller to monitor both small-scale and large-scale areas. Each region is managed by a sink, where the whole network is managed by a base station. The sensor nodes are categorized by various clusters using the low-energy adaptive clustering hierarchy (LEACH)-improved protocol in a way that the value of p is defined by a proposed support vector machine–based mechanism. A proposed fuzzy system determines that noncluster heads associate with several clusters in order to manage overlapping conditions over the network. Cluster heads are changed into clusters in a period by a suggested utility function. Since network lifetime should be prolonged and network traffic should be alleviated, a data aggregation mechanism is proposed to transmit only crucial data packets from cluster heads to sinks. Cluster heads apply a weighted criteria matrix to perform an inner-cluster routing for transmitting data packets to sinks. Simulation results demonstrate that the proposed protocol surpasses the existing methods in terms of the number of alive nodes, network lifetime, average time to recover, dead time of first node, and dead time of last node.

KEYWORDS

fuzzy logic, multicriteria decision making (MCDM), overlapping management, support vector machine (SVM), wireless sensor networks (WSNs)

1 | INTRODUCTION

Wireless sensor networks (WSNs)[1-3] are composed of low-energy, large-scale, and low-price wireless sensor nodes. Such sensors, called nodes, detect phenomena data (eg, temperature) to transmit sensed data to a center (eg, sink and base station). Wireless sensor networks have some applications in environmental monitoring,[4] traffic control,[5] fire systems,[6,7] etc. There are various challenges in these networks such as routing, fault tolerance, and overlapping.[8-14] An appropriate strategy presented for WSNs tries to solve one or some of the existing problems. A schematic of a WSN for fire systems is shown in Figure 1. The nodes sense environmental data (eg, smoke) and deliver sensed data to local sinks. Afterward, sinks transmit obtained data to a base station. Finally, the base station forwards received data packets to a center (eg, a fire department) via a gateway (eg, internet).

Overlapping happens in clustering sensor networks because there are common areas between some of the clusters. A proper overlapping management strategy can enhance the performance of many applications such as inter-cluster routing, topology discovery, and node localization.[8] Number of alive nodes,[15] network lifetime,[16-18] and average time to recover[19] are some of the important factors to evaluate the efficiency of any presented overlapping methods for WSNs. Figure 2 shows an illustrative example of overlapped clusters in a single-hop WSN. The clusters have the common areas covered by 2 or more cluster heads (CHs). Any sensor node placed at such areas is named as boundary node. A boundary node, in the most cases, transmits data packets to the nearest CH. This causes the efficiency of the network to be reduced considerably. The reason is that the selected CH will involve a high traffic load and will consume much more power energies. If a boundary node can associate with several CH nodes using an appropriate overlapping mechanism, the efficiency of the network would be enhanced.

An overlapping strategy should consider the mentioned factors to enhance the performance of a sensor network. Most of the existing methods do not utilize intelligent procedures to control overlapping conditions in WSNs. They cause the number of alive nodes to be decreased, network lifetime to be reduced, and average recovery time to be increased considerably. An intelligent and knowledge-based overlapping clustering protocol for WSNs is proposed in this paper to achieve 3 objectives: (1) increasing number of alive nodes, (2) prolonging network lifetime, and (3) decreasing average time to recover. This protocol uses some of the smart and knowledge-based systems to properly manage and control overlapping in clustering sensor networks. Multicriteria decision making (MCDM)[20] is used to split the network surface into several regions to reduce transmission distances among nodes. The nodes are grouped into various clusters in a way that CHs are selected by the low-energy adaptive clustering hierarchy (LEACH)-improved protocol[21] and the value of p is calculated by a proposed support vector machine (SVM).[22] A noncluster head (NCH) can associate with more than one CH to enhance the reliability and lifetime of CHs. All CHs are changed in a period, which are determined by a suggested utility function. Data packets are aggregated by CHs using a criteria-based mechanism to reduce the number of transmitted data packets to sinks and the base station. Finally, CHs apply an inner-cluster routing to transmit data packets to local sinks based on a suggested weighted criteria matrix.

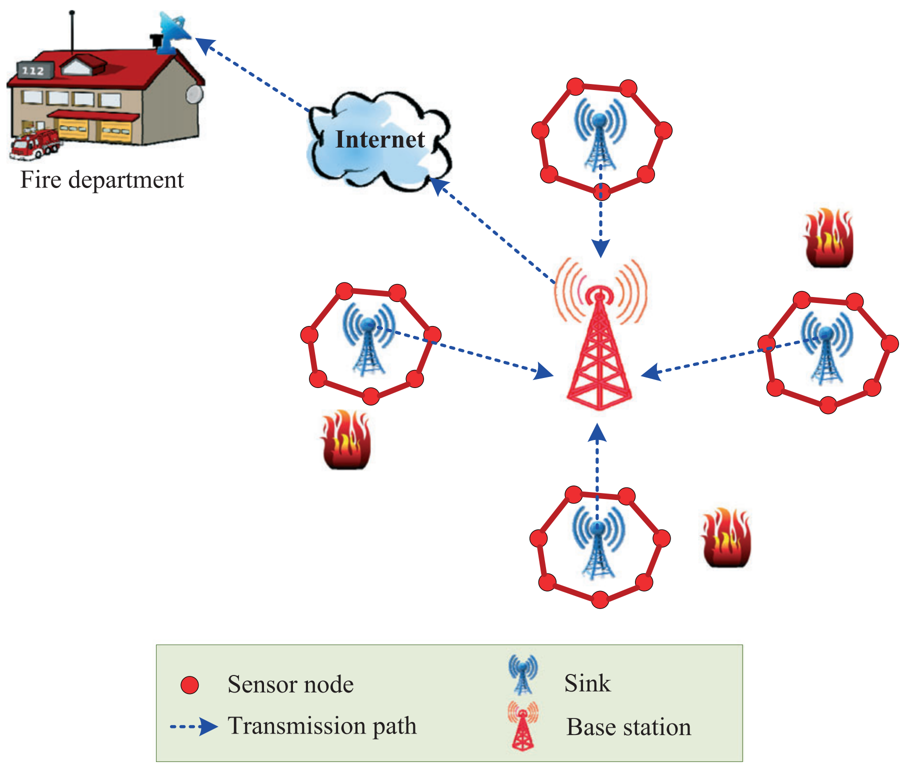

FIGURE 1 On overall view of a wireless sensor network

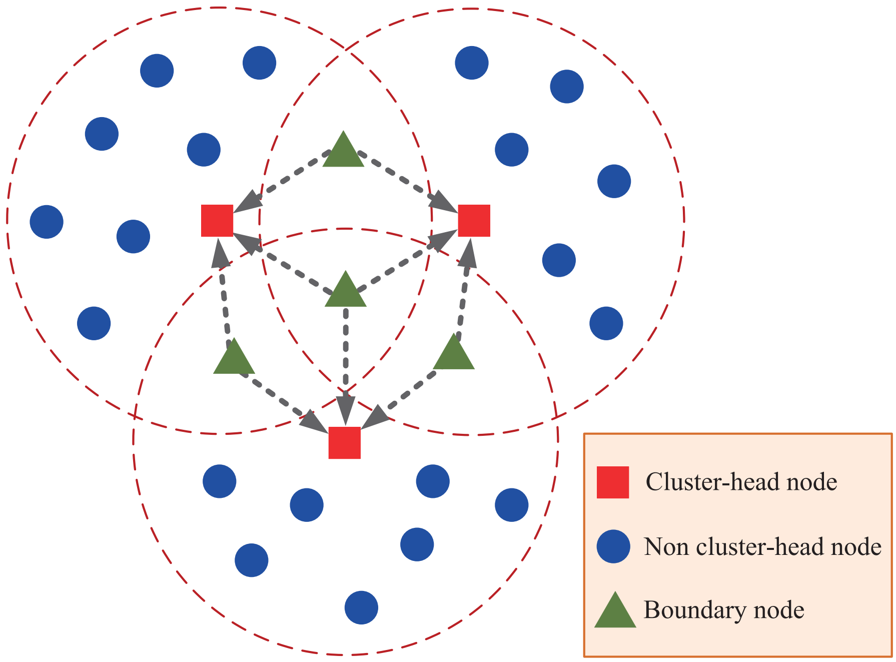

FIGURE 2 An example of overlapped clusters in a single-hop wireless sensor network

The remainder of this paper is organized as follows. Some of the existing related works are represented in Section 2. The proposed overlapping protocol is explained in Section 3 that includes the proposed MCDM controller, weighted criteria matrix, fuzzy system, utility function, criteria-based mechanism, and SVM-based mechanism. Simulation and comparison results are carried out in Section 4 to evaluate the efficiency of the proposed strategy compared with some of the existing methods. Finally, the paper is concluded in Section 5.

2 | RELATED WORKS

A randomized, distributed multihop overlapping clustering algorithm is presented in Youssef et al[8] to organize sensor nodes into overlapping clusters. It formulates the overlapping *k*-hop clustering problem to manage overlapping conditions in clusters. The multihop overlapping clustering algorithm meets both coverage and overlapping conditions with a high probability. In this algorithm, the requirements of sensor networks can be met by some heuristic procedures to decrease the processing and message complexity. It can make overlapping clusters with the average overlapping degree having high probability in a way that overlapped clusters boost the network robustness against clustering failures.

A randomized, distributed *k*-hop clustering algorithm (KOCA) is presented in Youssef et al[23] to solve the overlapping clustering problem. It formulates this problem carefully to generate the overlapping multihop clusters in WSNs. The KOCA generates the connected overlapping clusters in a way that the entire network is covered with a specific average overlapping degree. It generates approximately equal-sized clusters to enable distributing the traffic load over different clusters. Furthermore, the KOCA terminates in a constant number of iterations, independent of the network size.

A clustering algorithm is presented in Suharjono and Hendrantoro[24] that can control the overlapping conditions among clusters and maintain the balance of energy consumption. It is a modification version of the LEACH protocol called overlapping LEACH. The authors have used an appropriate mechanism to control the overlapped areas among clusters. Noncluster head nodes join the CH node that involves largest received signal strength. Moreover, they associate with another CH node involving a received signal strength of larger than *X*% from main CH node. Note that *X* limit value is specified based on the expected overlapping degree.

The problem of boundary discovery based on overlapping clustering in nonuniform WSNs is presented in Liao et al.[25] The paper presents a self-organization overlapping algorithm for clustering in nonuniform distribution (Nu-SOAC). This algorithm generates overlapping clusters and performs clusters boundary fusion to realize WSNs boundary. The main goals of the algorithm are to enhance network life cycle and generate network boundary. The Nu-SOAC forms the stable and reasonable clusters for improving the network life cycle and network robustness, efficiently.

A *k*-connected overlapping clustering method with energy awareness (K-OCHE) is presented in Ma et al[26] and Qin et al.[27] The basic goal of this method is to choose a CH node having energy availability status. The K-OCHE method adopts a sleep scheduling strategy to balance energy distributions, efficiently. In this strategy, neighboring nodes will remain awake to keep the method *k* connected. Cluster head nodes keep information about not only their own clusters but also adjacent clusters (eg, boundary nodes). Boundary nodes belong to multiple overlapping clusters for connecting different clusters to each other, during data transmission process. Evaluation results indicate that the method obtains a balanced load distribution, a longer network lifetime, and a quicker routing recovery time.

Moreover, some of the other routing protocols such as enhanced cluster based routing algorithm (ECBRP),[28] Bollinger band based energy efficient routing (BB),[29] coverage based energy efficient LEACH algorithm (CVLEACH),[30] and Bio-inspired based optimized algorithm[31] are presented in this area that discuss about energy efficiency in clustering sensor networks.

3 | THE PROPOSED PROTOCOL

This section describes the proposed intelligent and knowledge-based overlapping clustering protocol for WSNs, called IKOCP. This protocol not only manages overlapped clusters throughout the network but also uses intelligent mechanisms to provide an appropriate routing approach for increasing the efficiency of sensor networks. The network surface is split into several regions by a proposed MCDM controller to enable monitoring the small-scale and large-scale networks. Each region is controlled by a sink, and the whole network is controlled by a base station. The nodes are categorized into various clusters by the LEACH-improved protocol in a way that the value of *p* is determined by a proposed SVM-based mechanism. Note that the value of *p* indicates the percentage value of nodes to be selected as CH. Noncluster head nodes join several clusters to manage overlapping conditions over the network, by using a proposed fuzzy system. Since the energy consumption of CH nodes and the failure probability should be decreased in the service phase, all CH nodes are changed in a period. This process is carried out by a suggested utility function. A data aggregation mechanism is proposed to transmit only important data packets from CH nodes to sinks. Furthermore, a weighted criteria matrix is proposed to use an inner-cluster routing for transmitting data packets from CH nodes to sinks. The remainder of this section explains details of the proposed protocol. Note that the proposed MCDM controller and SVM-based mechanism are used in the deployment phase and the other mechanisms are used in the service phase.

3.1 | The network model

Figure 3 illustrates a schematic of the network model assumed by the proposed protocol. There are 9 regions in the network in a way that each one involves a sink and various clusters. The clusters are formed into each region using the LEACH-improved protocol. Each cluster has a CH node and some NCH nodes so that an NCH node transmits its sensing data to its CH node. Moreover, it can include the boundary nodes that are placed at overlapped areas. Ordinary nodes are not associated with any cluster so that their sensing data are not transmitted to the base station. Cluster head nodes transmit gathered data packets to the sink placed at the center of their region. Finally, sinks forward data packets toward a stationary base station. All sensor nodes, used in the protocol, are homogeneous so that they have the same initial energy, radio signal, and buffer size. Furthermore, the sinks and base station do not contain any energy constraints.

3.2 | Split the network by the proposed MCDM controller

The sensor network is partitioned into several regions by the proposed MCDM controller. This process is carried out to allow the proposed IKOCP protocol to monitor both small-scale and large-scale areas. This controller uses 2 input parameters including “surface area” and “number of nodes” to determine the output parameter “number of regions.” Parameter “surface area” uses 4 linguistic variables including “Small,” “Mediocre,” “Large,” and “Very Large.” Parameter “number of nodes” applies 4 linguistic variables including “Small,” “Medium,” “Large,” and “Very Large.” Finally, output parameter “number of regions” utilizes 5 linguistic variables including “Feeble,” “Few,” “Medium,” “Many,” and “So Many.” Since MCDM uses multicriteria rules to determine the best answer for any problem, Table 1 represents the possible rules of the proposed MCDM controller.

The proposed MCDM controller uses 3 stages to determine the number of regions based on surface area and number of nodes as (1) convert input quantity amounts to linguistic variables, (2) choose an appropriate output linguistic

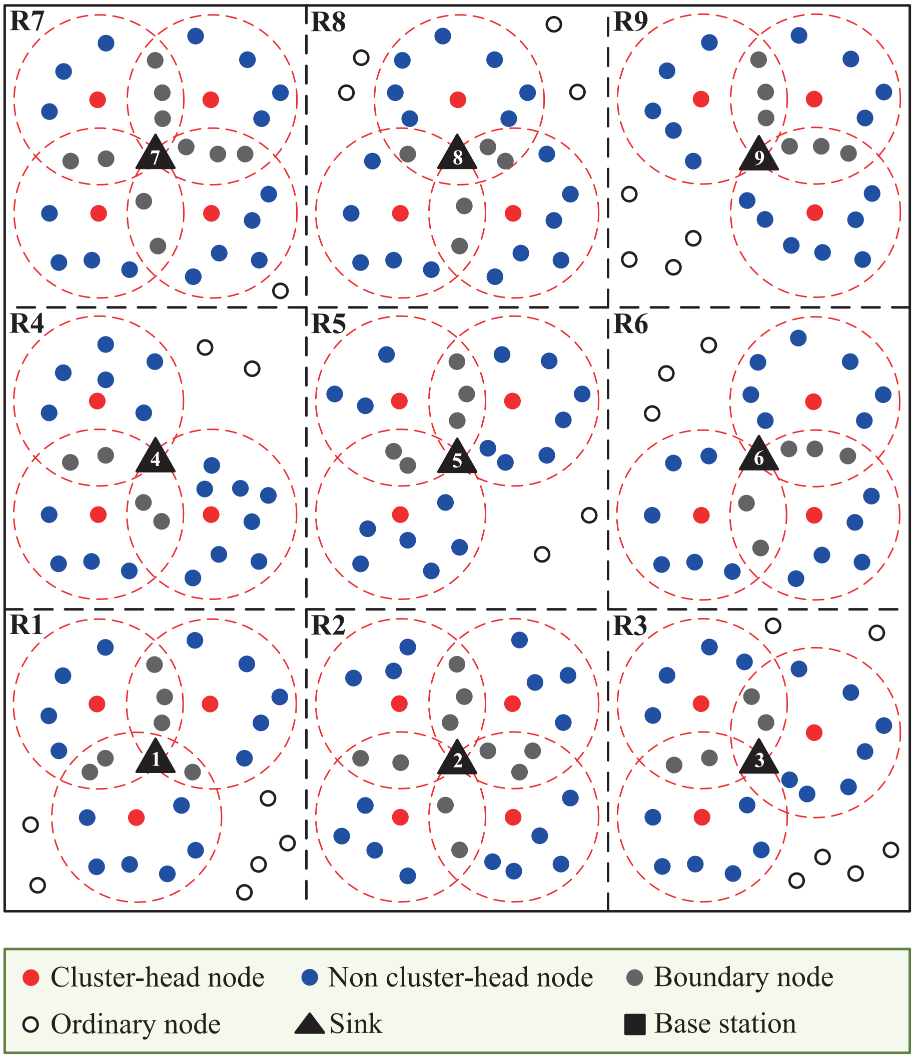

FIGURE 3 A schematic of the considered network model

TABLE 1 The multicriteria rules of the proposed MCDM controller

Number of nodes surface area	Small	Medium	Large	Very large
Small	Feeble	Feeble	Feeble	Few
Mediocre	Few	Few	Medium	Medium
Large	Medium	Medium	Many	Many
Very Large	Many	Many	So Many	So Many

variable, and (3) convert output linguistic variable to a quantity amount. Quantity amount of the surface area is calculated as

$$\mathrm{QA}_{\mathrm{SA}} = N_{\mathrm{L}} \times N_{\mathrm{W}}, \tag{1}$$

where N_{L} is the network length and N_{W} is the network width. Quantity amount of parameter "surface area" is converted to a linguistic variable as

$$\mathrm{SA}(x) = \begin{cases} \text{"Small,"} & 0 \le x < 0.25\ \mathrm{km}^2 \\ \text{"Mediocre,"} & 0.25\ \mathrm{km}^2 \le x < 1\ \mathrm{km}^2 \\ \text{"Large,"} & 1\ \mathrm{km}^2 \le x < 2.25\ \mathrm{km}^2 \\ \text{"Very Large,"} & \text{otherwise} \end{cases}, \tag{2}$$

where x is the quantitative amount of the network surface area. Quantity amount of parameter "number of nodes" is converted to a linguistic variable as

$$\mathrm{NN}(y) = \begin{cases} \text{"Small,"} & 0 \le y < 250 \\ \text{"Medium,"} & 250 \le y < 500 \\ \text{"Large,"} & 500 \le y < 750 \\ \text{"Very Large,"} & \text{otherwise} \end{cases}, \tag{3}$$

where y is the quantitative amount of the number of nodes. After determining the linguistic variables of the input parameters based on their amounts, an appropriate output linguistic variable is selected based on the approximate reasoning mentioned above. Afterward, the selected linguistic variable is converted to a crisp quantity to specify the number of regions as

$$\mathrm{NR}(z) = \begin{cases} 1, & z = \text{"Feeble"} \\ 4, & z = \text{"Few"} \\ 9, & z = \text{"Medium"} \\ 16, & z = \text{"Many"} \\ 25, & z = \text{"So Many"} \end{cases}, \tag{4}$$

where z is the selected linguistic variable of parameter "number of regions" calculated by the above equation. As indicated in the above equation, the minimum number of regions is considered as 1 and the maximum number of regions is considered as 25. For example, if surface area and number of nodes are large, the number of regions will be many. That is, they have positive impact in determining the number of regions. Suppose network length is 600 m, network width is 800 m, and number of nodes is 700; then, surface area will be 0.48 km^2, SA(0.48) will be "Mediocre," and NN(700) will be "Large" according to Equations 1 to 3. Afterward, the linguistic variable of the parameter "number of regions" will be "Medium" based on the approximate reasoning rules represented in Table 1. Finally, NR("Medium") will be equal to 9 according to Equation 4 that indicates the number of regions.

The main objectives to split the whole network into several regions and use multisinks throughout the network are (1) to decrease transmission distance between CH nodes and the base station, (2) to reduce energy consumption of CH nodes to transmit data packets toward the base station, and (3) to prolong network lifetime. Since the cost of sink is more than that of sensor node, we had to define a limited number of regions to reduce financial cost of the system while in the implementation phase. Therefore, if the maximum number of the regions was more than a certain number (eg, 25 regions), the number of sinks would be increased and thereby financial cost of the system would be enhanced noticeably.

Geographical coordinates of the sinks are calculated based on network length, network width, and number of regions. X geographical coordinates of the sinks are calculated as

$$\mathrm{Sink}_X(i) = \begin{cases} \frac{L(2\sqrt{n}-1)}{2\sqrt{n}}, & i \bmod \sqrt{n} = 0 \\ \frac{L}{2\sqrt{n}}, & i \bmod \sqrt{n} = 1 \\ \frac{3L}{2\sqrt{n}}, & i \bmod \sqrt{n} = 2 \\ \frac{5L}{2\sqrt{n}}, & i \bmod \sqrt{n} = 3 \\ \frac{7L}{2\sqrt{n}}, & i \bmod \sqrt{n} = 4 \end{cases}, \tag{5}$$

where i is the identifier number of each sink, n is the number of regions, and L is the network length. Moreover, Y geographical coordinates of the sinks are determined as

$$\mathrm{Sink}_Y(i) = \begin{cases} \frac{W}{2\sqrt{n}}, & \lceil i/\sqrt{n} \rceil = 1 \\ \frac{3W}{2\sqrt{n}}, & \lceil i/\sqrt{n} \rceil = 2 \\ \frac{5W}{2\sqrt{n}}, & \lceil i/\sqrt{n} \rceil = 3 \\ \frac{7W}{2\sqrt{n}}, & \lceil i/\sqrt{n} \rceil = 4 \\ \frac{W(2\sqrt{n}-1)}{2\sqrt{n}}, & \lceil i/\sqrt{n} \rceil = 5 \end{cases}, \tag{6}$$

where i indicates the identifier number of each sink, n indicates the number of regions, and W indicates the network width. For example, if the number of regions is 4, network length is 600 m, and network width is 800 m, then X and Y coordinates of the 4 sinks are considered as (150 m, 200 m), (450 m, 200 m), (150 m, 600 m), and (450 m, 600 m) according to Equations 5 and 6.

3.3 | Construct clusters based on the proposed SVM-based mechanism

The proposed IKOCP protocol uses LEACH-improved[21] to construct clusters of regions. Existing clustering routing methods, in the most cases, specify the value of p in a static manner so that this value is not changed throughout the network. If the value of p is determined based on network conditions, the efficiency of the sensor network can be enhanced noticeably. The IKOCP protocol proposes an SVM-based mechanism to define the value of p. Before describing the proposed mechanism, it is essential to explain how clusters are formed by the LEACH-improved protocol. LEACH-improved selects the CH nodes having high residual energy to prolong network lifetime. It constructs k clusters in a way that the algorithm defines a random number for each node; if the number is less than a given threshold value $T(n)$, the node becomes a CH node. This threshold value is calculated for each node based on initial energy, remaining energy, and the value of p as

$$T(n) = \frac{p \times E_C}{(1-p\times(r \mod (1/p)))\times E_I}; n \in G, \tag{7}$$

where p is the percentage value of nodes to be selected as CH, E_C is the remaining energy of the node, r is the currently circulating round number, E_I is the initial energy of all nodes, and G is a list of available sensor nodes on the network.[21]

The proposed SVM-based mechanism uses 2 input parameters including "surface area" and "number of nodes." The value of p is separately determined for each region so that the input parameters are specified based on the surface area and number of nodes for each region. The proposed SVM function is defined as

$$p = \sum (W \times P(i) + \mathrm{bias}) = \alpha_1 \frac{\mathrm{SA}_i}{\mathrm{SA}} + \beta_1 \frac{\mathrm{NN}_i}{\mathrm{NN}} + \mathrm{bias}, \tag{8}$$

where i is the identifier number of each region, SA_i is the surface area of region i, SA is the surface area of the network, NN_i is the number of nodes placed at region i, NN is the number of all sensor nodes, α_1 is the effect of parameter "surface area" on the function, β_1 is the effect of parameter "number of nodes" on the function, and bias is the tolerance amount of the function. Surface area of any region is calculated as

$$\mathrm{SA}_i = \frac{L \times W}{n}, \tag{9}$$

where i is the identifier number of each region, L is the network length, W is the network width, and n is the number of regions. The simulation results have demonstrated that if the value of p is less than 0.3, the number of clusters will be many. Therefore, it is indicated that $\alpha_1 + \beta_1 \leq 0.2$ and bias ≤ 0.1. Since both surface area and number of nodes have positive effects on the function, if they are so many, then the value of p will be very high. Table 2 represents the value of p and number of clusters under various changes on surface area and number of nodes. The results of each instance are carried out for one of the existing regions based on surface area of the network, surface area of the region, number of

TABLE 2 The effects of surface area and number of nodes on the value of *p* and number of clusters

No.	Surface area of the network, m^2	Surface area of a region, m^2	Number of all sensor nodes	Number of nodes at a certain region	The value of *p*	Number of clusters
1	400	400	10	10	0.3	8
2	2500	2500	50	50	0.3	21
3	10 000	10 000	50	50	0.3	28
4	40 000	40 000	100	100	0.3	40
5	90 000	90 000	100	100	0.3	46
6	250 000	27 778	500	87	0.12537	74
7	250 000	27 778	1000	168	0.12507	128
8	810 000	90 000	1800	357	0.12	245
9	1 000 000	62 500	2000	142	0.11293	266
10	4 000 000	160 000	2500	326	0.11252	332
11	6 250 000	250 000	2500	439	0.11478	340
12	9 000 000	360 000	3000	117	0.10795	393
13	16 000 000	640 000	3000	336	0.1116	366
14	25 000 000	1 000 000	4000	257	0.10921	479
15	25 000 000	1 000 000	4500	158	0.10776	518

all sensor nodes, and number of nodes at the region. It is evident that number of clusters will be many while the value of *p* is less than 0.3.

3.4 | Associate NCH nodes with CHs using the proposed fuzzy system

The main goal of this paper is to manage and control overlapped clusters in WSNs. It can be achieved by a system based on fuzzy logic[32,33] that uses prior knowledge through an inference engine to make appropriate decisions under various conditions. The proposed fuzzy system controls overlapping conditions in WSNs by using 3 essential parameters including distance, remaining energy, and number of members. Each NCH node associates with all covered CH nodes to transmit sensing data to them according to a given percentage. Figure 4 shows the main purpose of the proposed fuzzy

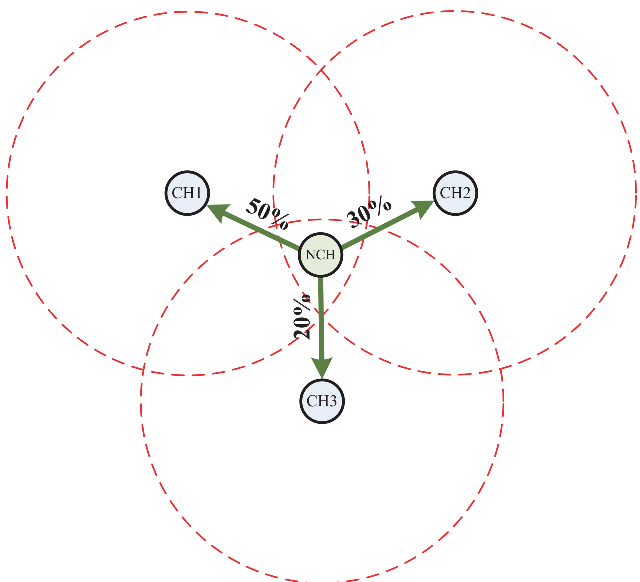

FIGURE 4 An example for associating a noncluster head (NCH) node to several cluster head (CH) nodes

system to manage overlapped clusters. It represents that each NCH node existed in an overlapped area can interact with all available CH nodes with a certain percentage. In this example, NCH node has 50% association with CH1, 30% with CH2, and 20% with CH3. That is, nearly 50% of sensing data will be transmitted to CH1, nearly 30% of sensing data will be transmitted to CH2, and nearly 20% of sensing data will be transmitted to CH3. The fuzzy system calculates the value of these percentages via a knowledge-based procedure.

The proposed fuzzy system consists of 3 input parameters including "distance," "remaining energy," and "number of members" and 1 output parameter "hit rate." Parameter "distance" represents geographical distance between NCH node and each CH node, parameter "remaining energy" represents residual energy of the CH node, parameter "number of members" represents number of NCH nodes at any cluster, and parameter "hit rate" represents rate of the CH node to interact with the NCH node. Note that the percentage of each CH node for associating to an NCH node will be determined based on the calculated value of parameter "hit rate." For this work, it is supposed that the linguistic terms of parameter "distance" are {"Near," "Medium," "Far"}, the linguistic terms of parameter "remaining energy" are {"Low," "Medium," "High"}, the linguistic terms of parameter "number of members" are {"Few," "Medium," "Large"}, and the linguistic terms of parameter "hit rate" are {"Low," "Medium," "High"}. Furthermore, the universe of discourse for parameter "distance" is {0, 20, 40, 60, 80} meters, the universe of discourse for parameter "remaining energy" is {1, 2, 3, 4, 5} joules, the universe of discourse for parameter "number of members" is {1, 3, 5, 7, 10}, and the universe of discourse for parameter "hit rate" is {0, 0.25, 0.5, 0.75, 1}. Membership functions of the input and output parameters used in the proposed fuzzy system are depicted by Figure 5.

Figure 6 illustrates the main components of the proposed fuzzy system that is composed of fuzzification, rule base, inference engine, and defuzzification components. The system uses bell-shaped function[34] to fuzzify parameters "distance," "remaining energy," and "hit rate" as well as uses triangular function[35] to fuzzify parameter "number of members." The rule base is built by fuzzy rule making[36] based on the fuzzy rules represented in Table 3. Note that these rules

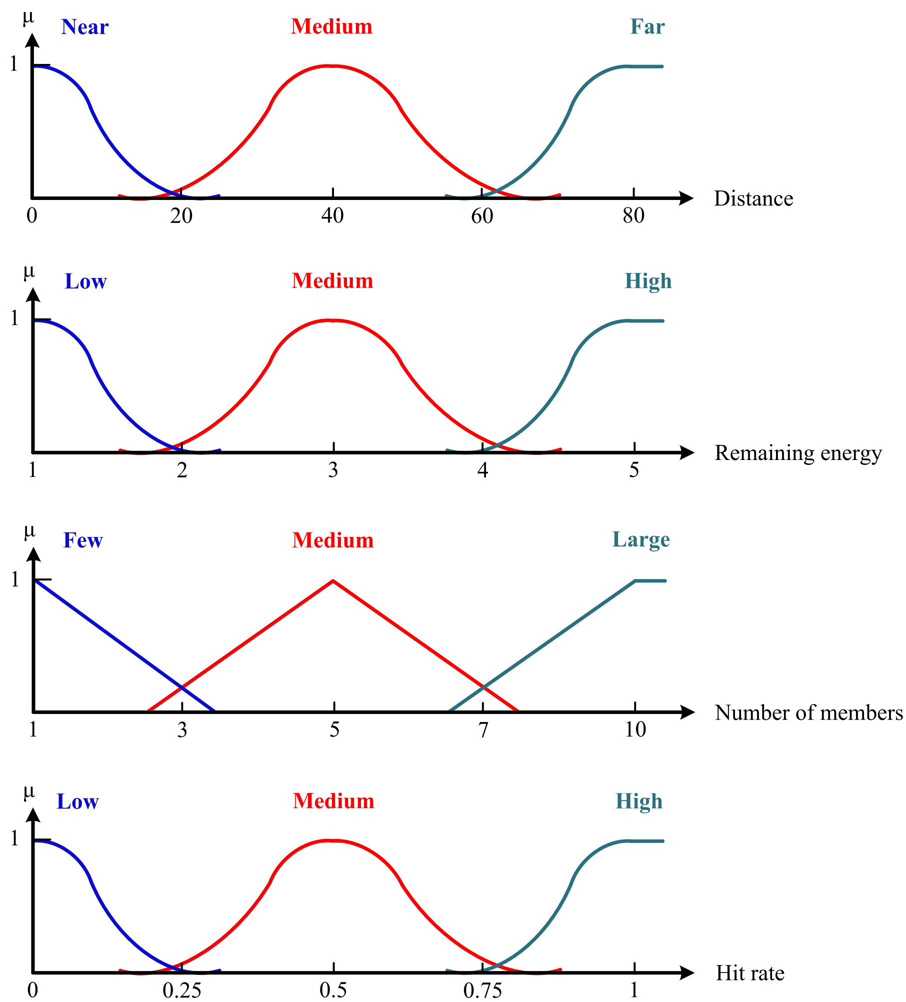

FIGURE 5 Membership functions of the proposed fuzzy system

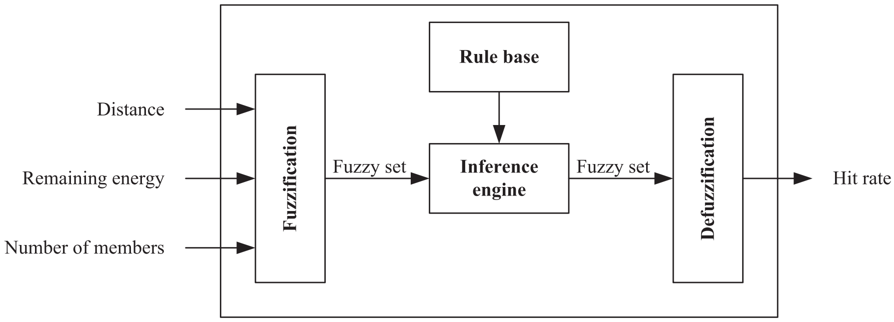

FIGURE 6 A schematic of the proposed fuzzy system

TABLE 3 The fuzzy rules used in the proposed fuzzy system

Rule no.	Antecedent			Consequent
	Distance	Remaining energy	Number of members	Hit rate
1	Near	Low	Few	Medium
2	Near	Low	Medium	Medium
3	Near	Low	Large	Low
4	Near	Medium	Few	High
5	Near	Medium	Medium	Medium
6	Near	Medium	Large	Medium
7	Near	High	Few	High
8	Near	High	Medium	High
9	Near	High	Large	Medium
10	Medium	Low	Few	Medium
11	Medium	Low	Medium	Low
12	Medium	Low	Large	Low
13	Medium	Medium	Few	High
14	Medium	Medium	Medium	Medium
15	Medium	Medium	Large	Low
16	Medium	High	Few	High
17	Medium	High	Medium	Medium
18	Medium	High	Large	Medium
19	Far	Low	Few	Low
20	Far	Low	Medium	Low
21	Far	Low	Large	Low
22	Far	Medium	Few	Medium
23	Far	Medium	Medium	Medium
24	Far	Medium	Large	Low
25	Far	High	Few	Medium
26	Far	High	Medium	Medium
27	Far	High	Large	Low

are aggregated together by the OR operation to build the total fuzzy rule. Fuzzy inference is one of the main components in a fuzzy system that determines appropriate decisions based on input fuzzy sets. This fuzzy system applies Mamdani type[37] in the inference engine to specify the output fuzzy set based on the input fuzzy sets. Afterward, output fuzzy set is converted to a crisp value by the center-of-gravity method[38] to specify hit rate of each CH node.

After determining the hit rate of CH nodes by the proposed fuzzy system, the percentage of each CH node for any desired NCH node will be calculated based on their rates. First, the sum of hit rates is determined for all CH nodes, and then the percentage of each CH node is calculated as

$$P_i = \frac{\mathrm{HR}_i}{S} \times 100; i = 1, 2, ..., n, \tag{10}$$

where n is the number of CH nodes, i is the identifier number of each CH node, HR_i is the hit rate of CH node i, and S is the sum of all hit rates. When an NCH node has a sensing data, it selects one of its CH nodes based on the percentage of each CH node. Note that NCH nodes can even associate with the CH nodes that are placed in other regions. Algorithm 1 represents the procedure that an NCH node uses to select an appropriate CH node to transmit sensing data. Lines 1 to 5 indicate the input and initial variables of the procedure. In line 4, a random number $R \in [1, 100]$ is selected by the algorithm; and in line 5, the percentage of the first CH node is initialized in variable S. Lines 6 to 13 describe how the algorithm selects the identifier number of the best CH node to transmit sensing data from the NCH node to the selected CH node.

Algorithm 1 Select an appropriate CH node by an NCH node

```
1:P[] ← Percentages of CH nodes
2:N ← Number of CH nodes
3:I ← 1
4:R ← A random number between 1 and 100
5:S ← P [1]
6:While (I ≤ N){
7:    If (R ≤ S)
8:        Return I
9:    Else {
10:        I ← I + 1
11:        S ← S + P[I]
12:    }
13:}
```

As an example, assume that there are 3 CH nodes that are covered by an NCH node as represented in Table 4. The hit rate of each CH node is specified using the proposed fuzzy system as described before. Afterward, the percentage of each CH node is calculated based on the hit rates and sum of all rates. Finally, the range of each CH node and the generated random number determine which CH node is the best target to receive sensed data from the NCH node. In this case, if the random number equals 55, CH2 will be selected as the best CH node.

3.5 | Change of CHS by the suggested utility function

Because the remaining energy of CH nodes will be reduced and their failure probability will be enhanced, it is essential that CH nodes switch between NCH nodes during a specified period. The IKOCP protocol uses a suggested utility function to select new CH nodes based on power energy, distance, and buffer status. First, this function is performed for all sensor nodes existing in clusters; then, the node with the highest utility will be selected as the best CH node into each cluster. The mathematical equation of this function is defined as

TABLE 4 An example to select an appropriate CH node by an NCH node

CH node no.	Hit rate	Percentage	$R \in [1, 100]$
CH1	0.9	$0.9/2.1 \times 100 = 43\%$	[1-43]
CH2	0.8	$0.8/2.1 \times 100 = 38\%$	[44-81]
CH3	0.4	$0.4/2.1 \times 100 = 19\%$	[82-100]
	$\sum = 2.1$		

$$\text{Utility}_i = \alpha_2 \frac{\text{Rem}_{\text{E}}^{i}}{\text{Initial}_{\text{E}}} + \beta_2 \frac{1}{\text{Avg}_{\text{D}}^{i}} + \gamma_2 \frac{\text{Free}_{\text{B}}^{i}}{\text{Initial}_{\text{B}}}; i = 1, 2, \ldots, n, \tag{11}$$

where n is the number of nodes of a cluster, i is the identifier number of each node, $\text{Rem}_{\text{E}}^{i}$ is the remaining energy of node i, $\text{Initial}_{\text{E}}$ is the initial energy of all sensor nodes, $\text{Avg}_{\text{D}}^{i}$ is the average distance from node i to other nodes of the cluster, $\text{Free}_{\text{B}}^{i}$ is the free buffer size of node i, $\text{Initial}_{\text{B}}$ is the initial buffer size of all sensor nodes, α_2 is the effect of remaining energy on the utility function, β_2 is the effect of distance, and γ_2 is the effect of node's buffer size. As indicated in the above function, if remaining energy of a node is high, average distance from the node to other nodes is low, and the free buffer size of the node is many, the success rate of the node to be selected as new CH node will be high. Average distance from a node to other nodes of a cluster is calculated as

$$\text{Avg}_{\text{D}}^{i} = \frac{\sum_{j=1}^{n} \sqrt{\left(X_j - X_i\right)^2 + \left(Y_j - Y_i\right)^2}}{n}; i = 1, 2, \ldots, n \wedge j = 1, 2, \ldots, n, \text{where } i \neq j, \tag{12}$$

where n is the number of nodes of the cluster, i is the identifier number of each node, j is the identifier number of the neighboring nodes of node i, (X_i, Y_i) is the geographical coordinates of node i, and (X_j, Y_j) is the geographical coordinates of node j. Also, the free buffer size of a node is calculated as

$$\text{Free}_{\text{B}}^{i} = \text{Initial}_{\text{B}} - \text{Occupied}_{\text{B}}; i = 1, 2, \ldots, n, \tag{13}$$

where n indicates the number of nodes of the cluster, i indicates the identifier number of each node, $\text{Initial}_{\text{B}}$ indicates the initial buffer size of all sensor nodes, and $\text{Occupied}_{\text{B}}$ is the number of data packets presented on the node's buffer.

3.6 | Data aggregation using the proposed criteria-based mechanism

In some cases where the number of sensing data by sensor nodes is so many, it is not essential to transmit all sensed data to the base station (eg, fire systems). Therefore, a data aggregation mechanism can be used to aggregate sensed data to transmit only crucial data packets to the base station. In the proposed protocol, CH nodes aggregate data packets by a criteria-based mechanism. This mechanism includes 3 stages: (1) grouping the sensed data into 5 categories including "Very Low," "Low," "Medium," "High," and "Very High"; (2) selecting the category having the largest number of sensed data; and (3) transmitting only the minimum, maximum, and average of the selected category. Figure 7 illustrates a schematic of the proposed criteria-based data aggregation.

For example, if temperature is the type of sensed data, then Algorithm 2 represents how the proposed mechanism aggregates gathered data of temperature, in a period. Lines 1 to 4 define some initial variables of the procedure. The algorithm categorizes sensed data into 5 arrays according to their data ranges, in lines 5 to 17. All data packets of the category having the most sensed data are copied to new array in line 18 in order to transmit only 3 data to the related sink. Lines 19 to 21 define some variables to select appropriate sensed data. In line 20, the initial values of the defined variables are set by the first element of the selected category. The minimum, maximum, and average data of the selected category are calculated from lines 22 to 29. Finally, they are returned to main program in line 30.

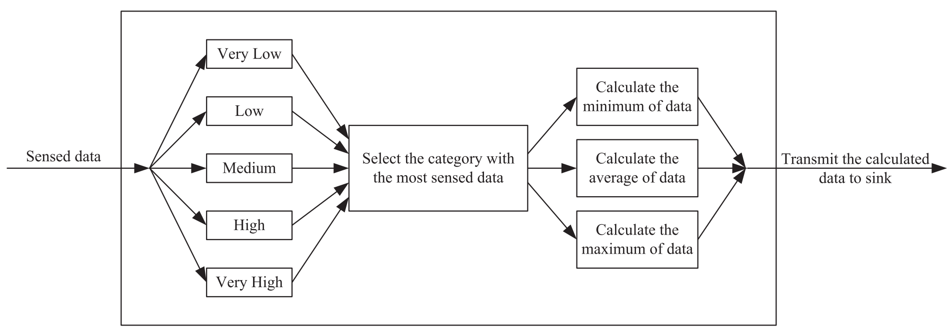

FIGURE 7 A schema of data aggregation by the criteria-based mechanism

Algorithm 2 Aggregate sensed data by the criteria-based mechanism

```
D[] ← Sensed data
N ← Number of sensed data
VL[] ← Empty; L[] ← Empty; M[] ← Empty; H[] ← Empty; VH[] ← Empty
I ← 1
While (I ≤ N){
    If (D[I]
        Append D[I] to VL
    Else If (D[I] ≥ −10 and D[I]
        Append D[I] to L
    Else If (D[I] ≥ 5 and D[I]
        Append D[I] to M
    Else If (D[I] ≥ 20 and D[I]
        Append D[I] to H
    Else
        Append D[I] to VH
    I ← I + 1
}
M[] ← Copy data of the category having the most sensed data
Len ← The length of M
Min ← M[1]; Max ← M[1]; S ← M[1]
J ← 2
While (J ≤ Len){
    If (M[J]
        Min ← M[J]
    If (M[J] > Max)
        Max ← M[J]
    S ← S + M[J]
}
Avg ← round(S / Len)
Return Min, Max, Avg
```

3.7 | Inner-cluster routing by the suggested weighted criteria matrix

Cluster head nodes use an inner-cluster routing to transmit data packets to their assigned sinks via one or more intermediate CH nodes. This routing is performed using a suggested weighted criteria matrix based on 4 parameters including "number of nodes," "remaining energy," "free buffer size," and "distance." These parameters are investigated for all neighboring CH nodes to select an appropriate CH to transmit data packets to the sink via hop-to-hop delivery. The "number of nodes" parameter indicates number of NCH nodes defined to a neighboring CH, the "remaining energy" parameter indicates residual energy of the neighboring CH, the "free buffer size" parameter indicates free buffer size of the neighboring CH, and the "distance" parameter indicates geographical distance from the neighboring CH to the sink of the related region. A schematic of the suggested procedure for inner-cluster routing is shown in Figure 8. When a CH node has data packets to transmit toward the sink, it selects one of its neighboring CH nodes based on the above parameters with the aid of a weighted criteria matrix. It transmits data packets to the sink via the neighboring CH node having the highest success rate. Afterward, the next CH node transmits data packets to its appropriate neighboring CH node and so on, until all data packets are delivered to the sink. For example, CH1 selects CH3 as its appropriate neighboring CH node, CH3 selects CH2 as an appropriate neighboring CH node, and finally data packets are delivered to the sink via CH2.

The suggested routing procedure uses a multicriteria matrix to select appropriate CH nodes. It consists of 2 stages: (1) determine success rate of neighboring CHs based on their characteristics and (2) select the CH node having the highest success rate. Table 5 represents the criteria matrix constructed for the proposed inner-cluster routing. This

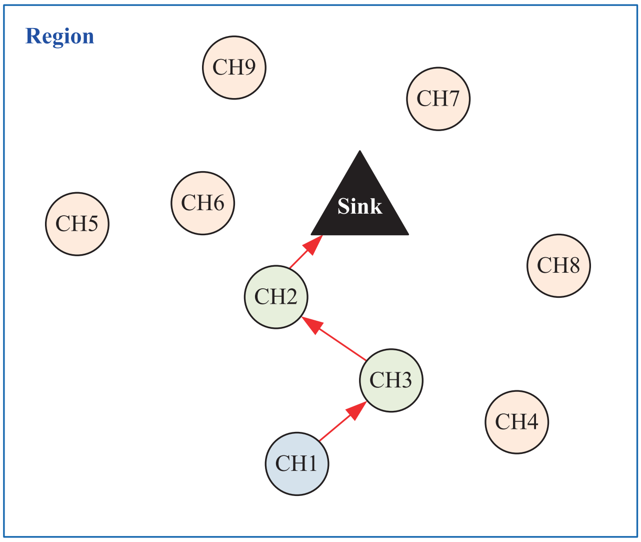

FIGURE 8 An overall view of the inner-cluster routing using the suggested weighted criteria matrix. CH, cluster head

TABLE 5 The elements of the criteria matrix used in the proposed inner-cluster routing mechanism

Neighbor CH no.	Number of nodes	Remaining energy	Free buffer size	Distance
1	m_1	e_1	b_1	d_1
2	m_2	e_2	b_2	d_2
...	...	...	...	...
n	m_n	e_n	b_n	d_n

matrix includes the input parameters of the suggested routing including "number of nodes," "remaining energy," "free buffer size," and "distance."

To select the best neighboring CH, the normalized weighted criteria matrix should be defined to calculate success rate of each CH node. Some of the above parameters have positive impact on the success rate, and some others have negative impact. Hence, the parameters should be normalized to participate in the weighted criteria matrix. The number of nodes determined for each neighboring CH is normalized as

$$m'_x = \frac{\min(m_1, m_2, ..., m_n)}{m_x}; x = 1, 2, ..., n, \tag{14}$$

where i is the identifier number of each neighboring CH node. The remaining energy of each neighboring CH is normalized as

$$e'_x = \frac{e_x}{\max(e_1, e_2, ..., e_n)}; x = 1, 2, ..., n. \tag{15}$$

The free buffer size of each neighboring CH is normalized as

$$b'_x = \frac{b_x}{\max(b_1, b_2, ..., b_n)}; x = 1, 2, ..., n. \tag{16}$$

Finally, the distance from each neighboring CH node to the sink is normalized as

$$d'_x = \frac{\min(d_1, d_2, ..., d_n)}{d_x}; x = 1, 2, ..., n. \tag{17}$$

After calculating the normalized values of all parameters, effect of each parameter should be defined separately. The vector [$\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{w}_3$, $\mathbf{w}_4$] represents the effects of the parameters where the sum of these weights equals 1. Table 6

TABLE 6 Details of the normalized weighted criteria matrix applied by the proposed inner-cluster routing

Neighbor CH no.	Number of nodes	Remaining energy	Free buffer size	Distance	Success rate
1	$\mathbf{w}_1\ m'_1$	$\mathbf{w}_2\ e'_1$	$\mathbf{w}_3\ b'_1$	$\mathbf{w}_4\ d'_1$	s'_1
2	$\mathbf{w}_1\ m'_2$	$\mathbf{w}_2\ e'_2$	$\mathbf{w}_3\ b'_2$	$\mathbf{w}_4\ d'_2$	s'_2
…	…	…	…	…	…
n	$\mathbf{w}_1\ m'_n$	$\mathbf{w}_2\ e'_n$	$\mathbf{w}_3\ b'_n$	$\mathbf{w}_4\ d'_n$	s'_n

represents the normalized weighted criteria matrix built for the proposed inner-cluster routing. The success rate of each neighboring CH node i is calculated as

$$s'_x = \mathbf{w}_1 m'_x + \mathbf{w}_2 e'_x + \mathbf{w}_3 b'_x + \mathbf{w}_4 d'_x; x = 1, 2, ..., n, \tag{18}$$

where x is the identifier number of each neighboring CH node. The neighboring CH having the highest success rate will be selected as an appropriate node to transmit data packets to the sink.

4 | PERFORMANCE EVALUATION

This section evaluates the performance of the proposed IKOCP protocol in terms of the number of alive nodes, network lifetime, average time to recover, dead time of first node, and dead time of last node. The simulation process is conducted in the MATLAB simulator considering any simulation requirements for WSNs (eg, data generation process, network traffic, and routing procedure).[39] Moreover, we have used the Prowler toolbox under MATLAB[40,41] to control some of the essential procedures in the simulation phase. The protocol is compared with some of the existing related methods including LEACH,[42] hybrid and energy-efficient approach (HEED),[43] weighted clustering algorithm (WCA),[44] Nu-SOAC,[25] greedy perimeter stateless routing (GPSR),[45] K-OCHE,[26,27] advanced LEACH routing protocol (ALEACH),[46] low energy adaptive clustering hierarchy with deterministic cluster-head selection (LDCHS),[47] and data dissemination based on ant swarms (TANT).[48] Furthermore, effects of the proportion of dead nodes, number of nodes, initial energy, and network size on the network lifetime and average time to recover are investigated. The simulation process is carried out during 10 hours in a topographical area of dimension 1000 m × 1000 m. The simulated network is considered as event-driven model so that several actuators generate events throughout the network. All actuators move over the network according to random waypoint mobility model.[49] The transmitting and receiving energies of sensor nodes are consumed according to the energy model presented in previous studies.[42,50] Based on this model, total energy consumption to transmit an l-bit packet over a distance d can be defined as

$$E_{\mathrm{tx}}(l,d) = \begin{cases} lE_{\mathrm{elec}} + l\varepsilon_{\mathrm{fx}} d^2, & d < d_0 \\ lE_{\mathrm{elec}} + l\varepsilon_{\mathrm{mp}} d^4, & d \geq d_0 \end{cases}, \tag{19}$$

where l is the length of the data packet, E_{elec} is the dissipated energy per bit for running the receiver circuit or transmitter, d is the distance between sender node and destination node, $\varepsilon_{\mathrm{fx}}$ or $\varepsilon_{\mathrm{mp}}$ is the amplifier energy associated with the transmitter amplifier model, and d_0 is the threshold distance. Threshold distance can be determined as

$$d_0 = \sqrt{\varepsilon_{\mathrm{fx}}/\varepsilon_{\mathrm{mp}}}. \tag{20}$$

Moreover, total energy consumption for receiving an l-bit packet can be calculated as

$$E_{\mathrm{rx}}(l) = lE_{\mathrm{elec}}, \tag{21}$$

where l is the length of the data packet and E_{elec} is the dissipated energy per bit for running the receiver circuit or transmitter. Simulation parameters are represented in Table 7. The number of regions in the simulation process was obtained equal to 16 by the proposed MCDM controller, based on the network characteristics. Note that the simulation process is repeated for each scenario at 10 times.

TABLE 7 Simulation parameters

Parameter	Default value
Simulation time, h	10
Topographical area, m	1000 × 1000
The location of the base station, m	(500, 1100)
Number of nodes	2000
Initial energy of sensors, J	5
Radio signal of sensor nodes, m	50
Initial buffer size of sensor nodes	200 packet
Number of actuators	50
Operating range of actuators, m	100
Packet data size, kbit	3
Data generation rate, data/min	1000
Number of RWP points	100
α_1	0.15
β_1	0.05
bias	0.1
Period time to change CHs, min	30
α_2	0.5
β_2	0.3
γ_2	0.2
Time to transmit data packets from CHs to sinks, min	5
E_{elec}, pJ/bit	50
ε_{fx}, pJ/bit/m^2	10
ε_{mp}, pJ/bit/m^4	0.0013

Abbreviations: CHs, cluster heads; RWP, random waypoint.

4.1 | Simulation results

The effect of the initial energy on the number of alive nodes is shown in Figure 9. The simulation results indicate that the number of alive nodes increases when the initial energy of sensor nodes enhances. The reason is that a high initial energy prolongs the dead times of sensors and increases the number of alive nodes. Note that the number of alive nodes

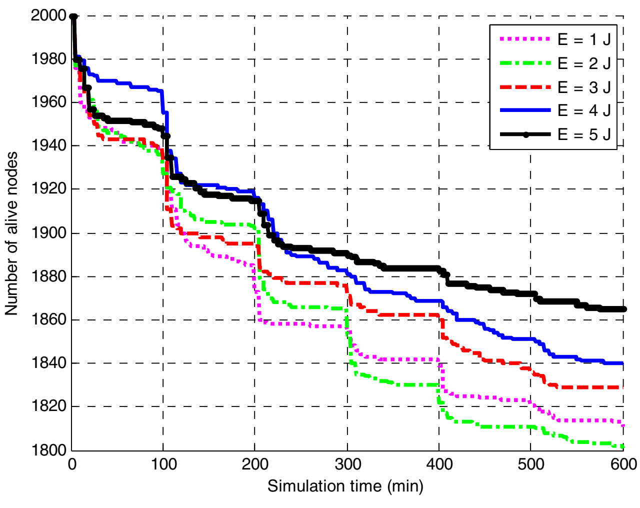

FIGURE 9 The effect of initial energy on number of alive nodes

are large at the initial times of the simulation process and they decrease when the simulation time increases because of consuming more power energies by sensors.

Figure 10 shows effect of the data generation rate on number of alive nodes. It illustrates that number of alive nodes decreases when data generation rate increases. The reason is that increasing the data generation rate causes the number of generated data packets to be increased, the power energy consumed by sensor nodes to be enhanced, and the number of alive nodes to be decreased considerably. Since the number of transmitted data packets increases along the simulation process, the number of alive nodes decreases during the simulation time.

Effect of the number of actuators on number of alive nodes is illustrated in Figure 11. As noted before, actuators are used in the simulation process to generate events at anywhere on the network. If there is a few number of actuators, then there will be a few number of events and generated data packets; therefore, the total energy consumption of sensor nodes will be low and number of alive nodes will be large and vice versa. Simulation results show that the number of alive nodes decreases when the number of actuators increases along the simulation process.

4.2 | Comparison of results

The proposed IKOCP protocol is compared with LEACH, HEED, WCA, and Nu-SOAC in terms of network lifetime. Figure 12 shows the effect of the proportion of dead nodes and number of nodes on network lifetime of the simulated methods. Figure 12A demonstrates that the network lifetime obtained by the proposed protocol is higher than that

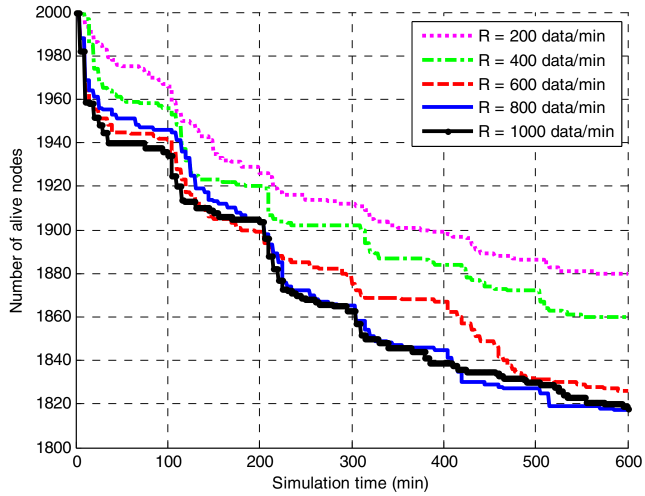

FIGURE 10 The effect of data generation rate on number of alive nodes

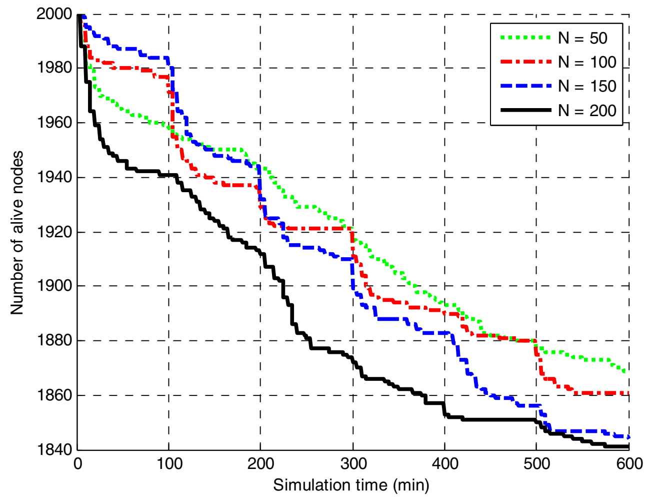

FIGURE 11 The effect of the number of actuators on number of alive nodes

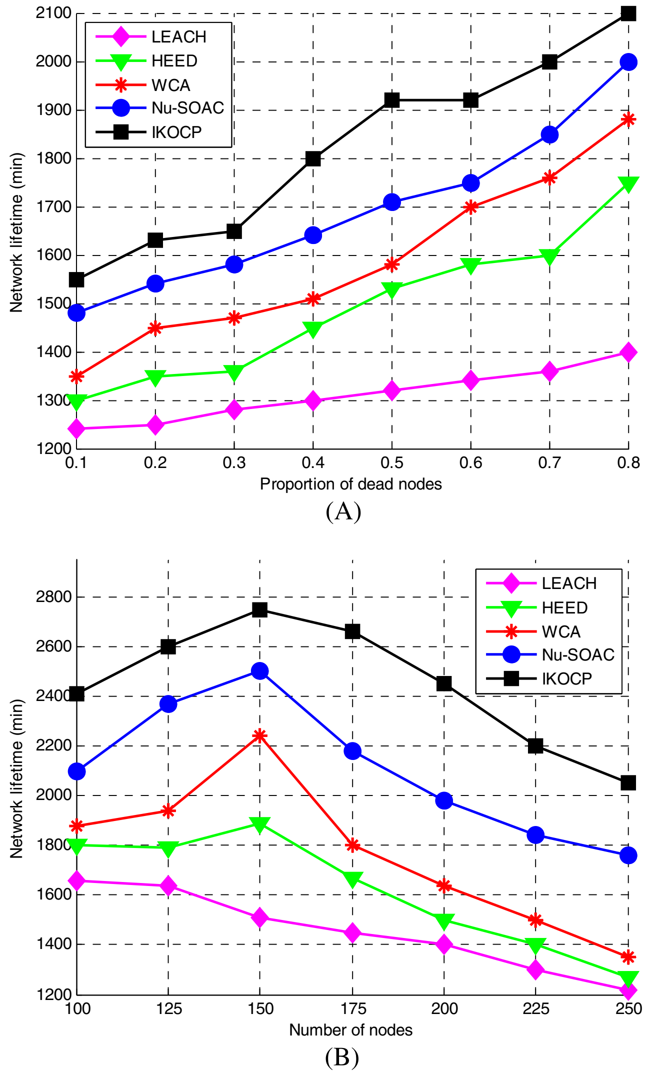

FIGURE 12 The effect of some input parameters on network lifetime. A, Proportion of dead nodes. B, Number of nodes

obtained by the other methods. The reason is that the proposed protocol uses some intelligent mechanisms to manage overlapped clusters. Figure 12B indicates that the network lifetime achieved by the IKOCP protocol is higher than that achieved by the other methods. Note that the network lifetimes obtained by all methods reduce when the number of nodes increases, because of increasing the number of transmitted data packets.

The proposed IKOCP protocol is compared with GPSR and K-OCHE in terms of the network lifetime and average time to recover. Figure 13 shows the effect of initial energy on network lifetime for GPSR, K-OCHE, and IKOCP. The comparison results indicate that the performance of IKOCP protocol is better than the GPSR and K-OCHE methods for different initial energies of sensors. The reason is that the proposed protocol considers the remaining energy of sensor nodes in most of the suggested mechanisms so that network lifetime obtained by this protocol is high compared with that obtained by the other methods. Because initial energy of nodes directly effects on network lifetime, the network lifetime of all simulated methods enhances when the initial energy increases.

Average time to recover is one of the main parameters to evaluate the efficiency of the presented methods in WSNs. Effect of the network size on average time to recover for GPSR, K-OCHE, and IKOCP is illustrated in Figure 14. The comparison results demonstrate the high efficiency of the proposed protocol compared with the other methods. It is shown that the average time to recover in the IKOCP protocol is low compared with that in the GPSR and K-OCHE methods for various network sizes. When the network size increases, geographical distances among sensor nodes will be increased, and hence, recovery time of all clusters will be grown significantly.

Dead time of the first and last nodes is one of the main factors to evaluate the efficiency of any routing protocol in WSNs. Table 8 represents the dead time of the first node and Table 9 indicates the dead time of the last node for the

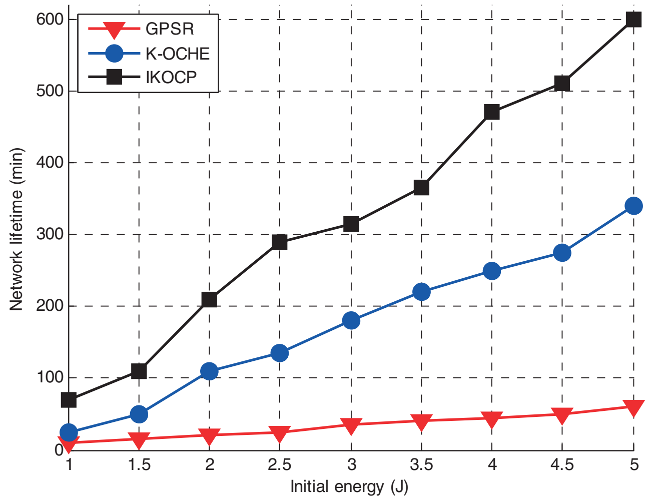

FIGURE 13 The effect of initial energy on network lifetime

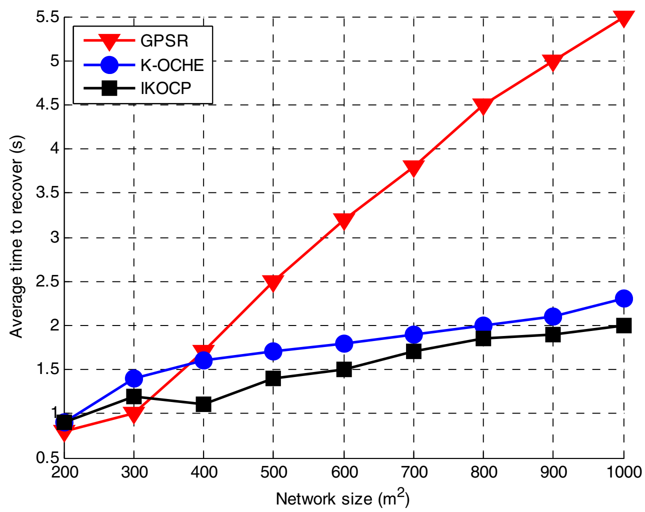

FIGURE 14 The effect of network size on average time to recover

TABLE 8 The dead time of the first node under various changes on number of nodes

Method		Number of nodes				
		100	200	300	400	500
Time, s	HEED	141	143	145	148	151
	WCA	142	145	147	149	154
	Nu-SOAC	201	204	207	211	216
	ALEACH	180	185	189	193	200
	LDCHS	140	141	144	146	150
	LEACH	139	141	143	146	149
	TANT	138	141	142	145	147
	IKOCP	231	235	240	243	249

proposed IKOCP protocol compared with the HEED, WCA, Nu-SOAC, ALEACH, LDCHS, LEACH, and TANT clustering routing methods. These results are achieved according to various changes on the number of nodes. As described in the comparison results, the dead times of proposed protocol for both the first node and the last node are more than those of the other routing methods. The reason is that IKOCP uses a multisink strategy via several intelligent and knowledge-based systems (eg, SVM and fuzzy system) to improve the most deployment and routing tasks of WSNs.

TABLE 9 The dead time of the last node under different number of nodes

Method		Number of nodes				
		100	200	300	400	500
Time, s	HEED	397	403	426	443	471
	WCA	401	409	433	451	479
	Nu-SOAC	853	864	876	885	896
	ALEACH	800	830	845	870	880
	LDCHS	400	420	445	470	490
	LEACH	390	395	420	435	450
	TANT	370	385	390	415	430
	IKOCP	918	926	934	936	946

5 | CONCLUSIONS

Overlapping cluster is one of the main challenges in WSNs. This paper proposed an intelligent and knowledge-based overlapping clustering protocol for WSNs, called IKOCP. It uses some of the intelligent and knowledge-based methods to manage and control overlapping conditions in WSNs. The IKOCP protocol applies several mechanisms to split the whole network, constitutes several clusters on the network, and associates noncluster heads with several CHs. Besides, it switches CHs of all clusters with NCH nodes, aggregates data packets, and selects appropriate transmission paths from CHs to sinks. A proposed MCDM controller splits the network surface into several regions using 2 parameters including surface area and number of nodes. An individual sink manages each region as well as a base station manages the whole network. The proposed SVM-based mechanism determines the value of p defined in the LEACH-improved clustering protocol into each region using some parameters such as surface area of each region, surface area of the network, and number of nodes. Noncluster heads associate with several CHs by a proposed fuzzy system to control the overlapping situations throughout the network. This process is carried out using 3 parameters including distance, remaining energy, and number of members. A suggested utility function changes CHs into all clusters in a period using some parameters such as remaining energy of each node, average distance from the node to other nodes, and free buffer size of each node. A data aggregation mechanism is proposed to aggregate data packets by CHs according to the ranges of sensed data. A suggested weighted criteria matrix selects an appropriate path from CHs to sinks via one or more intermediate CHs existed in the related cluster. This process is performed using 4 parameters including number of nodes, remaining energy, free buffer size, and distance.

The IKOCP protocol is simulated to demonstrate the efficiency of the proposed mechanisms. The simulation results indicate the effect of initial energy, data generation rate, and number of actuators on the number of alive nodes. Furthermore, the proposed protocol is compared with LEACH, HEED, WCA, Nu-SOAC, GPSR, K-OCHE, ALEACH, LDCHS, and TANT in terms of the network lifetime, average time to recover, dead time of first node, and dead time of last node. The comparison results show that the IKOCP protocol has high efficiency compared with the other methods depending on the proportion of dead nodes, number of nodes, initial energy, and network size.

ORCID

Mohammad Samadi Gharajeh http://orcid.org/0000-0002-2727-7742

REFERENCES

1. Yick J, Mukherjee B, Ghosal D. Wireless sensor network survey. *Comput Netw*. 2008;52(12):2292-2330.
2. Islam K, Shen W, Wang X. Wireless sensor network reliability and security in factory automation: a survey. *IEEE T Syst Man Cy C*. 2012;42(6):1243-1256.
3. Jadhav P, Satao R. A survey on opportunistic routing protocols for wireless sensor networks. *Procedia Comput Sci*. 2016;79:603-609.
4. Hodge VJ, O'Keefe S, Weeks M, Moulds A. Wireless sensor networks for condition monitoring in the railway industry: a survey. *IEEE T Intell Transport Sy*. 2015;16(3):1088-1106.

5. Kafi MA, Challal Y, Djenouri D, Doudou M, Bouabdallah A, Badache N. A study of wireless sensor networks for urban traffic monitoring: applications and architectures. *Procedia Comput Sci*. 2013;19:617-626.
6. Chen Y, Hwang JK. A power-line-based sensor network for proactive electrical fire precaution and early discovery. *IEEE T Power Deliver*. 2008;23(2):633-639.
7. Luis JA, Galán JAG, Espigado JA. Low power wireless smoke alarm system in home fires. *Sensors*. 2015;15(12):20717-20729.
8. Youssef A, Younis M, Youssef M, Agrawala A. WSN16-5: distributed formation of overlapping multi-hop clusters in wireless sensor networks. In: IEEE Global Telecommunications Conference (GLOBECOM'06); 27 November 2006–1 December 2006; San Francisco, CA: IEEE. pp. 1-6.
9. Cheraghlou MN, Babaie S, Samadi MLRC. Novel fault tolerant local re-clustering protocol for wireless sensor network. *J Comput*. 2012;4:99-104.
10. Mahapatroa A, Khilar PM. Transient fault tolerant wireless sensor networks. *Procedia Technol*. 2012;4:97-101.
11. Gharajeh MS, Khanmohammadi S. Static three-dimensional fuzzy routing based on the receiving probability in wireless sensor networks. *Comput Secur*. 2013;2:152-175.
12. Gharajeh MS. Determining the state of the sensor nodes based on fuzzy theory in WSNs. *Int J Comput Commun Contr*. 2014;9(4):419-429.
13. Bhuiyan MZA, Wang G, Cao J, Wu J. Deploying wireless sensor networks with fault-tolerance for structural health monitoring. *IEEE T Comput*. 2015;64(2):382-395.
14. Gharajeh MS, Khanmohammadi S. Dispatching rescue and support teams to events using ad hoc networks and fuzzy decision making in rescue applications. *J Control Syst Eng*. 2015;3(1):35-50.
15. Thakkar A, Kotecha K. Alive nodes based improved low energy adaptive clustering hierarchy for wireless sensor network. In: Kumar Kundu M, Mohapatra D, Konar A, Chakraborty A, eds. *Advanced Computing, Networking and Informatics—Volume 2*. Berlin: Springer; 2014:51-58.
16. Cardei M, MacCallum D, Cheng MX, et al. Wireless sensor networks with energy efficient organization. *J Interconnect Netw*. 2002;3:213-229. 03n04
17. Li J, AlRegib G. Network lifetime maximization for estimation in multihop wireless sensor networks. *IEEE T Signal Proces*. 2009;57:2456-2466.
18. Rahman MN, Matin MA. Efficient algorithm for prolonging network lifetime of wireless sensor networks. *Tsinghua Sci Technol*. 2011;16(6):561-568.
19. Dengchang Z, Zhulin A, Yongjun X. Time synchronization in wireless sensor networks using max and average consensus protocol. *Int J Distrib Sensor Networ*. 2013;2013:1-10.
20. Triantaphyllou E. *Multi-Criteria Decision Making Methods: A Comparative Study*. 44 Berlin: Springer Science & Business Media; 2013.
21. Guo D, Xu L. LEACH clustering routing protocol for WSN. In: International Conference on Information Engineering and Applications (IEA); 2013; vol. 219, pp. 153-160.
22. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.
23. Youssef MA, Youssef A, Younis MF. Overlapping multihop clustering for wireless sensor networks. *IEEE T Parall Distr*. 2009;20(12):1844-1856.
24. Suharjono A, Hendrantoro G. Dynamic overlapping clustering algorithm for wireless sensor networks. In: IEEE International Conference on Electrical Engineering and Informatics (ICEEI); 17–19 July 2011; Bandung: IEEE. pp. 1-6.
25. Liao Y, Wang X, Qi H, Li W. Overlapping clustering for boundary discovery in non-uniform wireless sensor networks. *J Comput Inform Syst*. 2011;7:3550-3559.
26. Ma C, Wang L, Xu J, Qin Z, Shu L, Wu D. An overlapping clustering approach for routing in wireless sensor networks. In: IEEE Wireless Communications and Networking Conference (WCNC); 7–10 April 2013; Shanghai, Shanghai, China: IEEE. pp. 4375-4380.
27. Qin Z, Ma C, Wang L, Xu J, Lu B. An overlapping clustering approach for routing in wireless sensor networks. *Int J Distrib Sensor Networ*. 2013;2013:1-11.
28. Han U-P, Park S-E, Kim S-N, Chung Y-J. An enhanced cluster based routing algorithm for wireless sensor networks. In: International Conference on Parallel and Distributed Processing Techniques and Applications; 2006; vol. 1.
29. Thakkar A, Kotecha K. A new Bollinger band based energy efficient routing for clustered wireless sensor network. *Appl Soft Comput*. 2015;32:144-153.
30. Thakkar A, Kotecha K. CVLEACH: coverage based energy efficient LEACH algorithm. *Int. J. Comput. Sci. Netw*. 2012;1.
31. Thakkar A, Kotecha K. Bio-inspired based optimized algorithm for cluster head election using RSSI and LQI. *Int J Comput Sci*. 2011;1:19-29.
32. Ross TJ. *Fuzzy Logic with Engineering Applications*. New York, NY, USA: John Wiley & Sons; 2009.
33. Godo L, Gottwald S. Fuzzy sets and formal logics. *Fuzzy Set Syst*. 2015;281:44-60.

34. Zhao J, Bose BK. Evaluation of membership functions for fuzzy logic controlled induction motor drive. In: IEEE 28th Annual Conference of the Industrial Electronics Society (IECON); 5–8 November 2002; vol. 1, pp. 229-234.
35. Sumathi S, Paneerselvam S. Computational Intelligence Paradigms: Theory & Applications using MATLAB. Boca Raton, FL, USA: CRC Press, 2010.
36. Mamdani EH, Assilian S. An experiment in linguistic synthesis with a fuzzy logic controller. *Int J Hum-comput St*. 1999;51(2):135-147.
37. Cococcioni M, Foschini L, Lazzerini B, Marcelloni F. Complexity reduction of Mamdani fuzzy systems through multi-valued logic minimization. In: IEEE International Conference on Systems, Man and Cybernetics (SMC); 12–15 October 2008; Singapore: IEEE. pp. 1782-1787.
38. Runkler TA. Selection of appropriate defuzzification methods using application specific properties. *IEEE T Fuzzy Syst*. 1997;5(1):72-79.
39. Ali QI, Abdulmaowjod A, Mohammed HM. Simulation & performance study of wireless sensor network (WSN) using MATLAB. In: 1st International Conference on Energy, Power and Control (EPC-IQ); 30 Nov.−2 Dec. 2010; pp. 307-314.
40. Nayyar A, Singh R. A comprehensive review of simulation tools for wireless sensor networks (WSNs). *J Wirel Netw Comm*. 2015;5:19-47.
41. Prowler: probabilistic wireless network simulator. 20 Oct. 2017. Available URL: http://www.isis.vanderbilt.edu/projects/nest/prowler/
42. Heinzelman WR, Chandrakasan A, Balakrishnan H. Energy-efficient communication protocol for wireless microsensor networks. In: IEEE 33rd Annual Hawaii International Conference on System Sciences; 4–7 January 2000; vol. 2, pp. 1-10.
43. Younis O, Fahmy S. Distributed clustering in ad-hoc sensor networks: a hybrid, energy-efficient approach. In: IEEE 23 AnnualJoint Conference of the IEEE Computer and Communications Societies (INFOCOM); 7–11 March 2004; vol. 1.
44. Chatterjee M, Das SK, Turgut DWCA. A weighted clustering algorithm for mobile ad hoc networks. *Cluster Comput*. 2002;5(2):193-204.
45. Karp B, Kung HT GPSR: greedy perimeter stateless routing for wireless networks. In: 6th Annual International Conference on Mobile Computing and Networking; 2000; pp. 243-254.
46. Ali MS, Dey T, Biswas R. ALEACH: advanced LEACH routing protocol for wireless microsensor networks. In: IEEE International Conference on Electrical and Computer Engineering (ICECE); 20−22 Dec. 2008; pp. 909-914.
47. Handy MJ, Haase M, Timmermann D. Low energy adaptive clustering hierarchy with deterministic cluster-head selection. In: 4th IEEE International Workshop on Mobile and Wireless Communications Network; 9−11 Sept. 2002; pp. 368-372.
48. Selvakennedy S, Sinnappan S, Shang Y. Data dissemination based on ant swarms for wireless sensor networks. In: IEEE Consumer Communications and Networking Conference (CCNC); 8−10 Jan. 2006; vol. 1, pp. 132-136.
49. Bettstetter C, Hartenstein H, Pérez-Costa X. Stochastic properties of the random waypoint mobility model. *Wirel Netw*. 2004;10(5):555-567.
50. Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application-specific protocol architecture for wireless microsensor networks. *IEEE T Wireless Commun*. 2002;1(4):660-670.

How to cite this article: Khanmohammadi S, Gharajeh MS. An intelligent and knowledge-based overlapping clustering protocol for wireless sensor networks. *Int J Commun Syst*. 2018;e3577. https://doi.org/10.1002/dac.3577