

## Research Article

# CPEH: A Clustering Protocol for the Energy Harvesting Wireless Sensor Networks

Yu Han <sup>1</sup>, Jian Su <sup>2</sup>, Guangjun Wen <sup>1</sup>, Yiran He <sup>1</sup> and Jian Li <sup>1</sup>

<sup>1</sup>University of Electronic Science and Technology of China, 611731, China

<sup>2</sup>Nanjing University of Information Science & Technology, 210044, China

Correspondence should be addressed to Jian Li; [lj001@uestc.edu.cn](mailto:lj001@uestc.edu.cn)

Received 10 February 2021; Revised 4 March 2021; Accepted 26 March 2021; Published 12 April 2021

Academic Editor: Yuanpeng Zhang

Copyright © 2021 Yu Han et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the last decade, energy harvesting wireless sensor network (EHWSN) has been well developed. By harvesting energy from the surrounding environment, sensors in EHWSN remove the energy constraint and have an unlimited lifetime in theory. The long-lasting character makes EHWSN suitable for Industry 4.0 applications that usually need sensors to monitor the machine state and detect errors continuously. Most wireless sensor network protocols have become inefficient in EHWSN due to neglecting the energy harvesting property. In this paper, we propose CPEH, which is a clustering protocol specially designed for the EHWSN. CPEH considers the diversity of the energy harvesting ability among sensors in both cluster formation and intercluster communication. It takes the node's information such as local energy state, local density, and remote degree into account and uses fuzzy logic to conduct the cluster head selection and cluster size allocation. Meanwhile, the Ant Colony Optimization (ACO) as a reinforcement learning strategy is utilized by CPEH to discover a highly efficient intercluster routing between cluster heads and the base station. Furthermore, to avoid cluster dormancy, CPEH introduces the Cluster Head Relay (CHR) strategy to allow the proper cluster member to undertake the cluster head that is energy depletion. We make a detailed simulation of CPEH with some famous clustering protocols under different network scenarios. The result shows that CPEH can effectively improve the network throughput and delivery ratio than others as well as successfully solve the cluster dormancy problem.

## 1. Introduction

The automatic and unmanned operation is one of the most prominent characters in the Industry 4.0 era [1, 2]. To monitor the industrial system's state and detect errors, the Industry 4.0 applications need a large number of sensors attached to machines to report their sensing data to the control center periodically [3]. Traditional wireless sensor networks (WSN) face the weakness of energy supply [4]. When the sensors run out of their energy, the network will suffer the performance wreck. Replacing the sensor's battery will bring extra cost and even be impossible in some harsh scenarios. There are currently many protocols designed for WSN trying to extend the practical network lifetime [5–8] by saving and balancing the energy consumption among sensors. However, the innate defect of WSN cannot be covered up. Since the sense is a long-term task, traditional WSN will no longer be suitable in the Industry 4.0 era. In the last decade, the rapid develop-

ment of energy harvesting technology [9] makes tiny commercial sensors with energy harvesting ability available. Researchers proposed the energy harvesting wireless sensor network (EHWSN) to realize the long-term operation [10, 11]. Sensors in EHWSN harvest energy from their surrounding environments and utilize the harvested energy to power themselves. EHWSN eliminates the battery requirement and has an unlimited lifetime, in theory, making it a promising technology in Industry 4.0 era.

EHWSN typically consists of a base station (BS) and a larger number of sensors. Sensors generate their sensing data shortly and periodically to guarantee the monitoring accuracy and quickly respond to the error. EHWSN shall handle the heavy generated data timely and transmit all the information to the BS, which is high energy costing. However, in the majority of scenarios, the energy harvesting rates of sensors are limited and individual [12]. To improve the network performance, effective routing protocols are necessary for

EHWSN to cut down the energy consumption while better utilizing the harvested energy [13]. As a hierarchy approach, clustering has been proved to have the inherited advantages on energy efficiency and scalability compared with flat routing protocols [14, 15], making it suitable for EHWSN.

Clustering protocols work in rounds. Every round, the cluster head (CH) will collect the data from its cluster members (CMs) and upload the handled data to the BS. Since CMs are usually close to each other, the data they generate may have high correlations. The CH will take local data fusion to remove the redundancy, thus reducing the data quantity. The protocol can balance the workload over different sensors and utilize the network harvested energy properly by reconstructing the clusters, whereas despite the advantages, at the present stage, most clustering protocols are designed for traditional WSN. There will be some challenges for them to work in the EHWSN condition. First, the primary design purpose of current clustering protocols is to extend the lifetime of WSN. Nevertheless, the leading aspiration of EHWSN has turned to maximize the network performance under the energy harvesting restrictions. Second, both the cluster formation and the intercluster routing of current clustering protocols do not consider the different energy harvesting rates between sensors, resulting in a performance decline. Last, in EHWSN, the sensor may deplete the energy and return to the sleep state when serving as a CH. Current clustering protocols cannot respond to this condition, causing the cluster dormancy phenomenon.

In this paper, we propose CPEH, which is a clustering protocol designed explicitly for EHWSN. CPEH cares about the diversity of the energy harvesting ability among sensors in both cluster formation and intercluster communication. In the cluster formation process, CPEH considers the sensor's local energy state, local node density, and remote degree. It then uses fuzzy logic to conduct the cluster head competition distributedly. In the intercluster routing discovery, the Ant Colony Optimization (ACO) is utilized to help the BS discover the proper paths for different CHs. To avoid cluster dormancy, CPEH introduces the Cluster Head Relay (CHR) strategy, which will choose the appropriate CM to undertake the sleep CH. Compared with other clustering protocols, the simulation result proves that CPEH will effectively improve the network delivery ratio and elevate the network throughput. Meanwhile, the CHR strategy can successfully maintain the cluster working normally when the initial CH runs out of its energy.

The rest of this paper is organized as follows. Section 2 gives a brief review of current clustering protocols. In Section 3, both the network model and the energy model used in this paper are described. In Section 4, we introduce our proposed protocol CPEH in detail. The simulation result is discussed in Section 5. Finally, Section 6 concludes the paper and introduces the further work.

## 2. Related Works

Clustering protocols have been well studied in the last decade. The Low Energy Adaptive Clustering Hierarchy (LEACH) [16] protocol is the most famous clustering proto-

col for periodical sensing applications. LEACH works in rounds and uses a random rotation to select the CH in order to balance the energy dissipation between sensors. The rotation guarantees each sensor to become a CH once and only once in every  $1/P$  round, where  $P$  is the expected cluster head ratio. Every CH adopts a local data aggregation to decrease the data amount sending to the BS. LEACH achieves a better performance than traditional flat routing protocols. However, researchers in [17] point out that the selection of CHs in LEACH does not consider the sensor's energy state. The low energy node will die rapidly, once it becomes a CH, limiting the network lifetime. To solve that, they propose an Energy-Efficient Clustering Scheme (EECS), which uses the residual energy as the concerned parameter of the CH selection. The candidate CHs with the most energy within the competitive radius will become the final CH. To balance the energy consumption, CMs in EECS choose their CHs based on a cost function that gives the CHs near the BS a higher priority. Hence, the energy consumption between CHs is balanced. Nevertheless, both LEACH and EECS are based on single-hop intercluster communication, causing them to be inefficient in large-scale networks due to long-range transmission. Hybrid Energy-Efficient Distributed Clustering (HEED) [18] employs an iteration strategy to construct the cluster and carry out a rarely uniform CH distribution. The communication between CHs and the BS is multihop to reduce the transmission cost. HEED achieves a good performance. However, the iteration process will increase the energy overhead caused by the exchanged control packets. Researchers in [19] propose the Unequal Cluster-based Routing (UCR) protocol which groups the nodes into clusters of unequal sizes to strike the energy dissipation imbalance between different CHs. The clusters near the BS will have smaller sizes in UCR so that the CHs can achieve the energy trade-off between the intracluster and intercluster.

It has been proved that the CH selection is typically an NP hard problem. Recently, with the fast development in computational intelligence, many researchers try to use the metaheuristic approach or fuzzy logic technology to resolve the CH selection. In [20], the BS uses particle swarm optimization (PSO) to optimize the set of CHs. The fitness function of PSO tries to decrease the maximum average intracluster distance in the meanwhile keeping the CHs with more energy. The CH in the Clustering Protocol based on the Metaheuristic Approach (CMPA) [21] is selected based on the Harmony Search (HS) algorithm. The selection focuses on reducing energy consumption in the meanwhile keeping a balanced energy distribution. CPMA also creatively introduces the Artificial Bee Colony (ABC) algorithm to optimize the protocol parameters, making it robust to various application scenarios. Clustering protocols using the metaheuristic approach are centralized and have a relatively high computational power requirement for the BS. For the large-scale network with strict scalability demand, it is a better choice to use fuzzy logic to handle the cluster head competition distributedly. In [22], the Fuzzy Energy-Aware Unequal Clustering (EAUCF) algorithm uses fuzzy logic to guide the clustering process. The fuzzy system considers both the residual energy and the distance to the BS. It is aimed at balancing the energy

consumption between CHs by assigning the CHs with different competitive radius. In [23], the researcher proposes a Multiobjective Fuzzy Clustering Protocol (MOFCA). MOFCA takes the sensors' nonuniform distribution and movement into account, making it suitable for both stationary and evolving networks. MOFCA has the same fuzzy system output as EAUCF. On the contrary, MOFCA additionally adds the sensor's local density to the fuzzy system input. The consideration of local density can effectively decrease the intracluster communication cost and maintain robustness over various network distribution and movement conditions, whereas like most clustering protocols, MOFCA does not care about the intercluster routing. The improper routing between CHs may downgrade the performance, especially in large-scale networks.

All the clustering protocols introduced above are designed for the traditional WSN. Currently, there are only several clustering protocols for EHWSN. In [24], the researchers propose the Energy Neutral Clustering (ENC) protocol to provide perpetual network operation. ENC employs a novel Cluster Head Group (CHG) mechanism that allows a cluster to use multiple cluster heads to share heavy traffic load. In every round, the sensors in the CHG will take the role of CH in turns to make sure the data sent by CMs can always be received. ENC uses convex optimization to choose the best combination of cluster number and the CHG number. The simulation shows ENC can improve the throughput than others. However, owing to the CHG mechanism, the areas in which CHG nodes locate may not be monitored correctly. A novel centralized clustering protocol for EHWSN is proposed in [25]. The BS in [25] runs a modified discrete particle swarm optimization (DPSO) algorithm to select CHs for the networks. The optimization considers both the transmission cost and the sensor's residual energy but ignores the difference of the energy harvesting ability between sensors, leading to a performance discount. Researchers in [26] pointed out that the energy harvesting process usually does not match the real energy demand, and sensors suffer from occasional energy shortages, mostly when they serve as the CH. To address this issue, they propose a Distance-and-Energy-Aware Routing with Energy Reservation (DEARER). DEARER encourages the sensor with more residual energy and a short distance to the sink more likely to become the CH. It also allows sensors to save their harvested energy for future use. The simulation results show that DEARER can achieve good performance. However, the complexity of DEARER is much higher than others, which causes it hard to implement.

### 3. System Models

**3.1. Network Model.** As shown in Figure 1, the EHWSN in this paper utilizes the CPEH, a multihop intercluster routing-based clustering protocol, to coordinate its network communication. The network assumptions are summarized as follows:

- (1) The BS is energy unlimited and always has sufficient knowledge of the entire network. Sensors are ran-

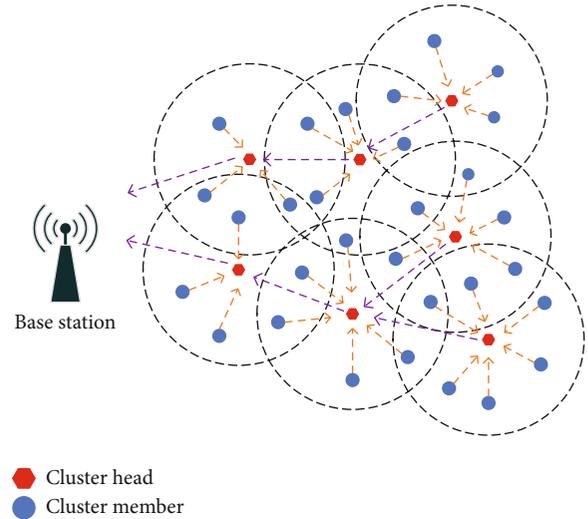


FIGURE 1: The primary network structure of the CPEH. CPEH is a multihop intercluster communication-based clustering protocol.

domly placed throughout the entire network. All the sensors have the same initial energy and maximum energy capacity. However, the energy harvesting rates among sensors are different. Both the BS and the sensor are motionless once deployed

- (2) All the sensors can communicate with the BS directly and have the ability to adjust their transmitting power according to the distance
- (3) All the network links are symmetric. Sensors can estimate the communication distance based on the received signal power
- (4) The network is homogeneous. The data generated by sensors in the same cluster have a relatively high correlation. The cluster head can make the local data aggregation to reduce the redundancy. The aggregation ratio is the same in different clusters

**3.2. Energy Model.** For energy harvesting, we assume that sensors can continuously harvest energy from their surrounding environments at a fixed speed. However, due to the diversity of location and hardware, the energy harvesting rates differ between sensors. *Since sensors usually harvest their energy from the heat or the vibration of their associated machines, we believe that this assumption is reasonable for the EHWSN used in the Industry 4.0 applications.* We use the *harvest-use* model to describe the energy availability of sensors. The harvested energy can be used immediately, which reduces the complexity of the energy management and easy to be implemented.

As shown in Figure 2, we employ the classical radio energy dissipation model used in [16]. The transmitter will consume energy to run the radio electronics and the power amplifier, whereas the receiver only dissipates energy to run the radio electronics. Represent the packet size as  $l$  bits and the communication distance as  $d$ ; the energy cost of

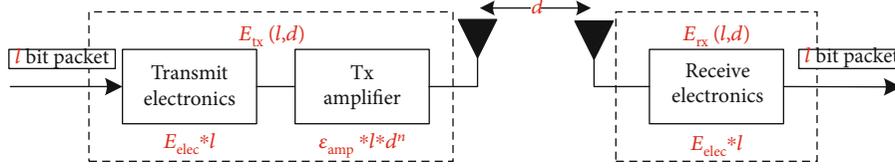


FIGURE 2: The radio energy dissipation model used in this paper.

transmitting ( $E_{tx}$ ) and receiving ( $E_{rx}$ ) a packet can be expressed as follows:

$$E_{tx}(l, d) = \begin{cases} E_{elec} \cdot l + E_{fs} \cdot d^2 \cdot l, & d \leq d_0, \\ E_{elec} \cdot l + E_{mp} \cdot d^4 \cdot l, & d > d_0, \end{cases} \quad (1)$$

$$E_{rx}(l, d) = E_{elec} \cdot l, \quad (2)$$

where  $E_{elec}$  is the energy cost per bit by the radio electronics of both transmitter and receiver and  $E_{fs} \cdot d^2$  and  $E_{mp} \cdot d^4$  represent the energy needed per bit for the power amplifier under the free space channel ( $d^2$ ) and the multipath channel ( $d^4$ ).  $d_0$  is the threshold distance defined as

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}}. \quad (3)$$

Notice that  $E_{elec}$ ,  $E_{fs}$ , and  $E_{mp}$  depend on the transceiver characters and the acceptable bit-error rate [19]. They should be set carefully. For data aggregation, assuming the CH will consume  $E_{da}$  energy to handle one-bit data and the cluster has  $W$  sensors (including the CH), the energy dissipation for the CH in the aggregation process is

$$E_{agg} = l \cdot W \cdot E_{da}. \quad (4)$$

## 4. CPEH Introduction

**4.1. CPEH Overall.** Figure 3 exhibits the basic communication timeline of CPEH, which is operated in rounds. Each round is comprised of a setup phase and a working phase. To better utilize the harvested energy and avoid some particular sensors' overuse, the CHs should be selected reasonably. Hence, awake sensors shall have different chances to be the CH based on the considerations of their energy states, locations, and surrounding conditions. The same considerations should also be taken to assign the CHs with different allowed sizes. At the beginning of the setup phase, every awake sensor will proceed with the cluster head election distributedly based on fuzzy logic, which we will give a detailed description in the following paper. After that, regular sensors will choose the CH to join based on the distance and size of the CH. Then, the CH will generate a TDMA schedule and broadcast it to all its members. At the end of the setup phase, the BS will use the ACO algorithm to discover the best inter-cluster route for each CH and broadcast the result to all the CHs.

Owing to the exchanging of control messages, the setup phase is relatively energy costing. To reduce the communication overhead, CPEH combines several frames into one working phase. In each frame, based on the TDMA schedule, the awake CM will transmit the sensing data to the CH in its associated slot and keep silent in others to save energy and avoid the collision. Once the CH detects the current energy lower than the sleep threshold, it will execute the CHR strategy to find the replacer and broadcast the result to all the awake CMs in the particular slot (Slot CHR). The chosen CM will become the successive CH in the following frames. Notice that if the network burden is heavy, the CHR strategy may be executed several times to maintain the cluster usually. After receiving the data from all the awake CMs, the CH will first aggregate the data and then upload the fusion data to the BS through the multihop routing.

The direct-sequence spread spectrum (DSSS) is utilized to avoid intercluster collision. CPEH assigns a universal spreading code and a unique spreading code to every sensor. The universal spreading code is used to exchange control messages in the setup phase and route packets between CHs in the working phase. On the contrary, once a sensor becomes a CH, it will confirm all the members using its unique spreading code for the intracluster communication. Hence, the adjacent clusters will have different spreading codes, eliminating the collisions among clusters.

Since the harvested energy may not support the continuous work, sensors under CPEH will operate in a sleep-wake mode. Define  $E_{capacity}$  as the sensor's energy storage capacity,  $E_{sleep}$  as the sensor's sleep threshold, and  $E_{awake}$  as the sensor's awake threshold. Then, we have  $E_{sleep} = 0.1 \cdot E_{capacity}$  and  $E_{awake} = 0.2 \cdot E_{capacity}$ . If the sensor detects that its residual energy has dropped off  $E_{sleep}$ , it turns to the sleep state to maintain the basic procedure and recharge. Once the energy reaches  $E_{awake}$ , the dormant sensor will wake up again and work normally.

### 4.2. CPEH Detailed

**4.2.1. Cluster Construction.** The cluster construction of CPEH is based on fuzzy logic. At the beginning of the setup phase, each awake sensor will broadcast a *State\_Message*, which contains its current residual energy and the energy harvesting rate, in the radius  $R$ . Based on the *State\_Message*, awake sensors will calculate the average residual and the average harvesting rate within  $R$ , which will be further used. Then, awake sensors will use the fuzzy logic system to estimate their chances to become the CHs and the allocated cluster sizes. The complicated fuzzy logic system working process is described as follows.

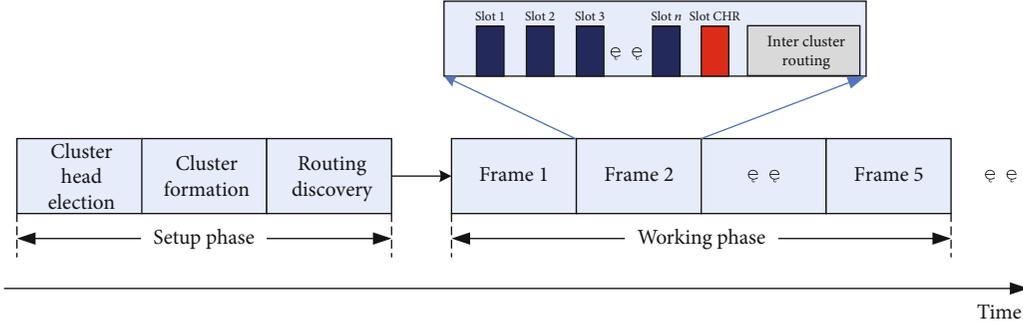


FIGURE 3: The basic communication timeline of CPEH. CPEH works in rounds. Every round consists of a setup phase and a working phase.

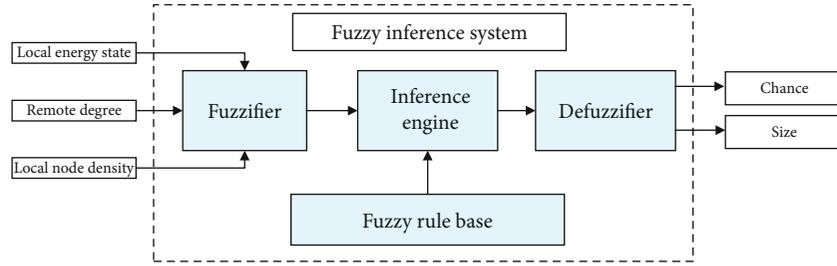


FIGURE 4: The structure of the fuzzy logic system. Sensor inputs the local energy state, remote degree, and the local node density. The system outputs the chance and the size.

The fuzzy logic system of CPEH is shown in Figure 4. It considers three different inputs: the local energy state, the local node density, and the distance to the BS.

(1) *Local Energy State*. The local energy state of sensor  $i$  is defined as

$$S_{i_{es}} = \left( \frac{S_{i_{re}}}{S_{i_{energy\_ave}}} \right)^\alpha \cdot \left( \frac{S_{i_{har}}}{S_{i_{har\_ave}}} \right)^\beta, \quad (5)$$

where  $S_{i_{re}}$ ,  $S_{i_{har}}$  means the residual energy and the energy harvesting rate of the sensor  $i$  and  $S_{i_{energy\_ave}}$ ,  $S_{i_{har\_ave}}$  represents the average residual energy and the average energy harvesting rate of all the awake sensors in the radius  $R$  of  $i$ .  $\alpha$  and  $\beta$  are the impact factors; we choose  $\alpha = 2$  and  $\beta = 1$  in this paper. The larger the local energy state is, the larger the chance and the size should be so that the sensor can fully exploit the harvested energy to undertake more works. The corresponding fuzzy linguistic variables are Good, Moderate, and Bad. Figure 5 shows the membership function of those linguistic variables.

(2) *Local Node Density*. The local node density of sensor  $i$  is defined as

$$S_{i_{den}} = \frac{\text{network\_area} \cdot N_{i_{local}}}{\pi \cdot R^2 \cdot N_{total}}, \quad (6)$$

where  $\text{network\_area}$  means the area of the entire network,  $N_{total}$  means network sensor number, and  $N_{i_{local}}$  means the

number of total awake sensors in the radius of sensor  $i$ . We believe that the fuzzy logic system should assign the sensor, which has a large density, with a higher chance and a bigger size so that the entire intracluster communication cost can be reduced. The corresponding fuzzy linguistic variables are Sparse, Normal, and Dense. Figure 6 shows the membership function of those linguistic variables.

(3) *Remote Degree*. Let  $d_{i\_to\_BS}$  means the distance between sensor  $i$  and the BS. Let  $d_{max}$  mean the max  $d_{i\_to\_BS}$  among all the sensors. We define

$$d_{i\_remote} = \frac{d_{i\_to\_BS}}{d_{max}}, \quad 0 < d_{i\_remote} \leq 1 \quad (7)$$

as the parameter to show the sensor's relative remote degree. The fuzzy logic system should give the CH nearing the BS a smaller cluster size to make the trade-off between intercluster and intracluster energy consumption. The corresponding fuzzy linguistic variables are Far, Medium, and Near. Figure 7 shows the membership function of those linguistic variables.

The fuzzy logic system has two outputs: chance and size. We give the chance nine fuzzy linguistic variables: Very Low, Low, Rather Low, Low Medium, Medium, High Medium, Rather High, High, and Very High. The corresponding membership functions are shown in Figure 8. The fuzzy linguistic variables of size are Very Small, Small, Rather Small, Low Medium, Medium, High Medium, Rather Large, Large, and

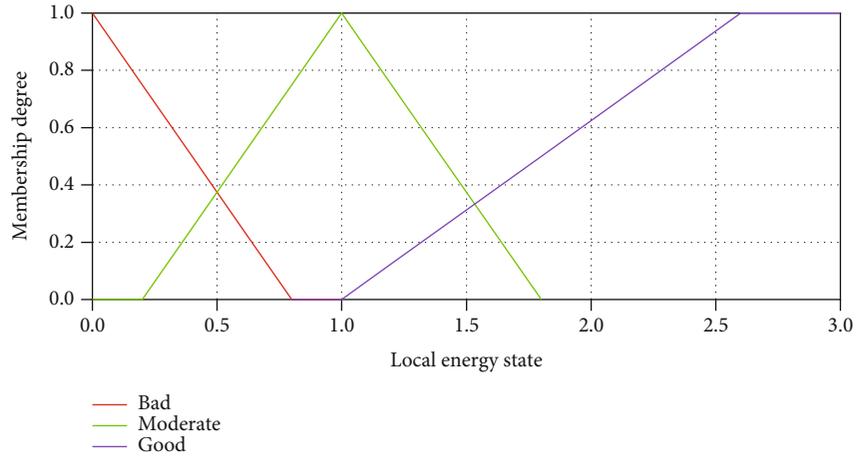


FIGURE 5: Membership function for local energy state.

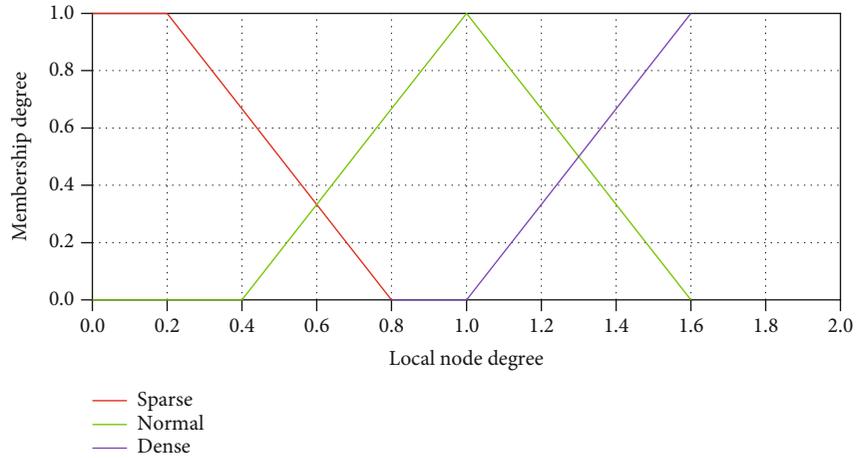


FIGURE 6: Membership function for local node degree.

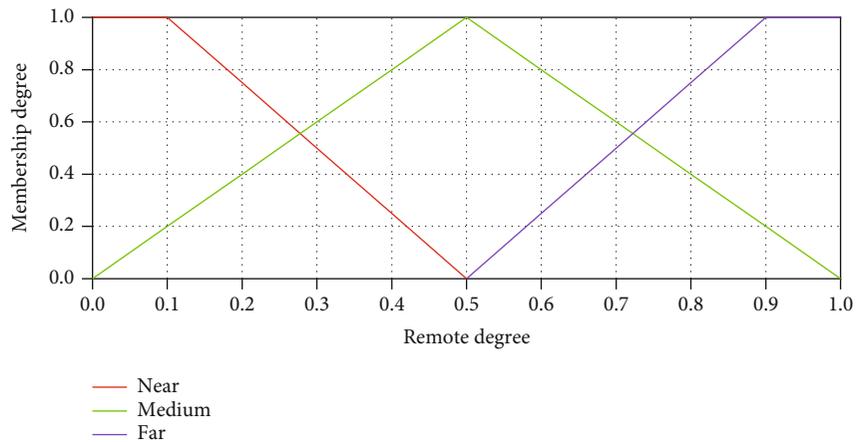


FIGURE 7: Membership function for remote degree.

Very Large. Figure 9 shows the membership function of those linguistic variables.

Once the sensor inputs the crisp values into the fuzzy logic system, the fuzzifier will transform those values to the

associated linguistic variables based on the membership functions. Then, the Mamdani [27] fuzzy inference engine will make the inference according to the input linguistic variables based on the fuzzy rules shown in Table 1. After that,

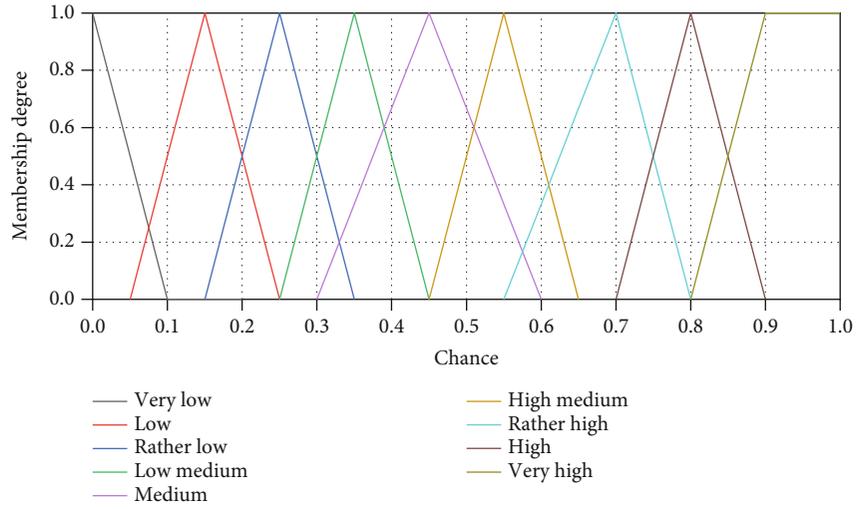


FIGURE 8: Membership function for remote degree.

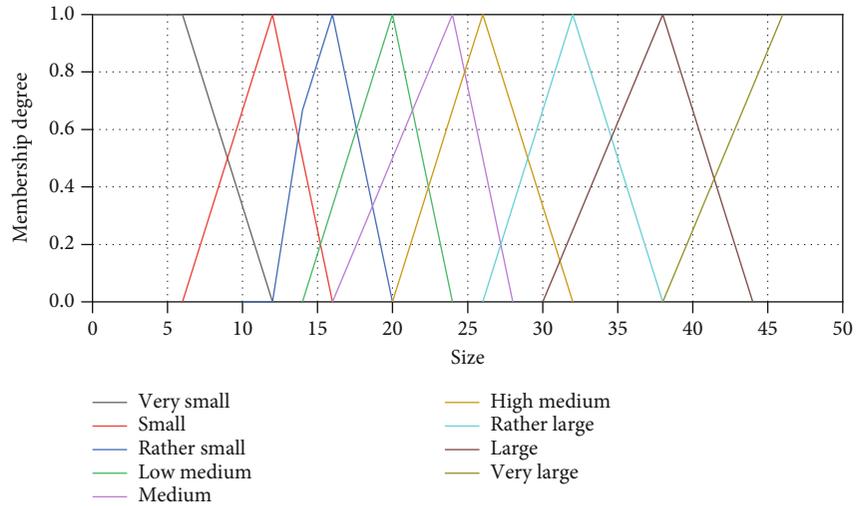


FIGURE 9: Membership function for size.

TABLE 1: The detailed fuzzy rules. The fuzzy rule base has 27 different rules.

No.	Input variables			Output variables	
	Local energy state	Remote degree	Local node density	Chance	Size
1	Bad	Near	Sparse	Very Low	Very Small
2	Bad	Near	Normal	Low	Very Small
3	Bad	Near	Dense	Rather Low	Small
13	Moderate	Medium	Sparse	Low Medium	Low Medium
14	Moderate	Medium	Normal	Medium	Medium
15	Moderate	Medium	Dense	High Medium	High Medium
25	Good	Far	Sparse	Rather High	Rather Large
26	Good	Far	Normal	High	Large
27	Good	Far	Dense	Very High	Very Large

the inferred linguistic variables will be defuzzified based on the center of area method. The fuzzy logic system finally outputs the crisp value of the chance and size.

After the fuzzy inference, the awake sensor will then broadcast a *Competition\_Message* containing its chance and ID, and it will record the corresponding information of its

```

Cluster Construction Process for sensor  $i$ 
1: At the beginning of the Round
2: If  $Si.state = awake$  then
3:   Broadcast a State_Message within  $R$ 
4: End if
5: On receiving the State_Message from its neighbor  $j$ 
6:   add  $j$  to its neighbor set  $Si.neighbor$ 
7:   calculate its local energy state and local node density
8:   infer its chance and size base on the fuzzy logic system
9: Broadcast a Competition_Message ( $Si.chance, Si.ID$ )
10: While the cluster head election time has not expired, do
11:   if  $Si.chance > Sj.chance, \forall Sj \in Si.neighbor$  then
12:     broadcast a Winner_Message ( $Si.ID$ ) and then EXIT
13:   end if
14:   on receiving a Winner_Message from sensor  $j$ 
15:   if  $Sj \in Si.neighbor$  then
16:     broadcast a Quit_Message ( $Si.ID$ ) and then EXIT
17:   end if
18:   on receiving a Quit_Message from sensor  $j$ 
19:   if  $Sj \in Si.neighbor$  then
20:     remove sensor  $j$  from  $Si.neighbor$ 
21:   end if
22: end while
23: If  $Si.role = CH$  then
24:   On receiving the Join_Message from sensor  $j$ 
25:   if the current cluster size less than  $Si.size$ 
26:     Accept the sensor  $j$ , sending back the Accept_Message
27:   else Sending back the Reject_Message
28:   end if
29: Else
30:   Sending the Join_Message to the nearest CH
31:   On receiving the Accept_Message, finish
32:   On receiving the Reject_Message, delete the CH and repeat 30
33:   if all the neighbor CHs refuse to accept sensor  $i$ 
34:      $Si.role = CH$ 
35:   end if
36: End if

```

PSEUDOCODE 1: Pseudocode of the cluster construction process in CPEH.

neighbors based on the received *Competition\_Message*. After that, the awake sensor will broadcast a *Winner\_Message* and become the final CH if it has the largest chance among its neighbors. If an awake sensor receives a *Winner\_Message* from its neighbors, it will quit the election and broadcast a *Quit\_Message*. The awake sensor will remove the sensor from its neighbors if it receives the corresponding *Quit\_Message*.

The non-CH sensors may receive more than one *Winner\_Message*. It will then choose the nearest CH to join and send a *Join\_Message* to the CH. Once the CH receives a *Join\_Message*, it will accept the sensor by sending back an *Accept\_Message* if the cluster is not full. Otherwise, a *Reject\_Message* will be sent to the sensor. The sensor will send another *Join\_Message* to the next nearest CH on receiving the *Reject\_Message* and may repeat this process until successfully joining a cluster. There may be the condition that all the clusters have no room for a sensor to join. In this case, the sensor becomes a CH by itself. Our simulation shows this condition is quite rare. We summarize the pseudocode of the cluster construction in Pseudocode 1.

After the cluster construction, the CH will broadcast a TDMA schedule that assigns the slots for its members. In the working phase, sensors will transmit their data in their associated slot and keep silent in others to save energy and avoid the collision. CPEH combines several frames into one round. Hence, awake sensors may deplete their energy and turn to sleep in the middle of the current round. If the depleted sensor is the CM, it will just give up its slot, whereas if the depleted sensor is the CH, the CHR strategy will be executed, which we will discuss later. Notice the sleep sensors at the setup phase may wake up halfway in the working round. However, those sensors will keep silent until the round ending since there is no slot designed.

Since the cluster construction is conducted fully distributedly by sensors, we analyze the control message complexity, reflecting the overhead caused by the message exchanging. Assuming there are  $N$  awake sensors at the setup phase, every awake sensor will broadcast a *State\_Message* and a *Competition\_Message*. If  $M$  sensors become the CH, there will be  $M$  *Winner\_Message* and  $N - M$  *Quit\_Message*. Then,

the number of control messages in the cluster head election process is  $N + N + M + (N - M) = 3N$ . Let  $k_i$  represent the number of attempts of  $CM_i$  before it joins a cluster. The number of control messages in the cluster formation process can be represented as  $\sum_1^{N-M} 2k_i = 2(N - M)K_{\text{average}}$ , where  $K_{\text{average}}$  means the average attempt number and is tested to be less than 3 in all the simulations. Hence, the total control messages in the cluster construction are  $(3 + 2K_{\text{average}})N - 2MK_{\text{average}}$ . The corresponding complexity is  $O(N)$ . Note that the control message number of CPEH is slightly larger than some cluster protocols such as LEACH, EAUCF, and MOFCA. However, we believe that this sacrifice will construct more proper clusters and enhance the final network performance.

**4.2.2. Intercluster Routing Discovery.** An effective intercluster routing protocol can significantly decrease network energy consumption and improve network performance. However, most clustering protocols only pay attention to the cluster construction but neglect the impact of intercluster routing on the final system performance. The CPEH considers both the energy cost and the energy state in the routing discovery and decides the final result in the view of the entire network. As the basic principle of ACO is the positive feedback mechanism [28], ACO can be treated as a reinforcement learning approach. A larger number of researches have proved that ACO has a unique advantage in solving the routing problem over other metaheuristic algorithms.

The routing discovery is based on iterations, which is both energy costing and complex. We assign this work to the BS, which can reduce the overhead of CHs and get the result quickly. After the cluster construction, each CH will transmit a control packet to the BS, containing its current residual energy and energy harvesting rate. The BS will calculate the energy states of each CH  $i$ :  $CH_i.es = (CH_i.re/E_{\text{cap}}) \cdot (CH_i.har/E_{\text{harmax}})$  and the possible next-hop set of each CH  $i$ :  $Set_i = \{CH_0, CH_j\}$ ,  $d_{CH_i \rightarrow BS} > d_{CH_j \rightarrow BS}$ , where  $E_{\text{cap}}$  means the energy capacity and  $E_{\text{harmax}}$  means the max harvesting rate. We let  $CH_0$  represent the BS. Then, the BS will discover the intercluster routes based on the following steps. We take the route between  $CH_i$  and the BS as an example.

**Step 1 (route searching).** The BS place  $n$  ants in  $CH_i$  at a fixed time interval to find the path to the BS. Each ant will choose the next-hop  $CH_j$  according to the following equation:

$$P_{ij} = \frac{[\tau_{ij}(t)]^\theta [\eta_{ij}]^\nu}{\sum_{k \in Set_i} [\tau_{ik}(t)]^\theta [\eta_{ik}]^\nu}, \quad (8)$$

where  $P_{ij}$  represents the probability of the ant in  $CH_i$  choosing  $CH_j$  as its next hop.  $\tau_{ij}(t)$  is the pheromone trail density on the edge of  $(CH_i, CH_j)$  at the time  $t$ . The routing design should consider both the total energy consumption and the energy states of sensors on the path. Hence, we define the heuristic information  $\eta_{ij}$  as

$$\eta_{ij} = \frac{CH_j.es}{[\text{dis}(CH_i, CH_j)]^2}. \quad (9)$$

$\theta$  and  $\nu$  are impact factors. We give  $\theta = 1$  and  $\nu = 3$  to let the algorithm own a good ability in both local and global search. The ant will continue the searching until it reaches the BS.

**Step 2 (route evaluating).** When the ant reaches the BS, the BS will get the route information that can be represented by  $X = (x_1, x_2, x_3, \dots, x_m, x_{m+1})$ , where  $m$  is the number of CHs in the route,  $x_1$  means the source CH and  $x_{m+1}$  means the BS. Then, the BS will evaluate the quality of the route based on the objective function  $F$ :

$$F = Q \cdot \frac{E_{\text{ave}} \cdot E_{\text{min}}}{E(X)}, \quad (10)$$

where  $Q$  is the constant factor.  $E_{\text{ave}}$  and  $E_{\text{min}}$  are the average energy state and minimum energy state of the sensors on the route.  $E(X)$  is the total energy consumption of transmitting a data packet over the route and can be summarized as follows:

$$\begin{aligned} E(X) &= l \cdot \left( \sum E_{\text{receive}} + \sum E_{\text{transmit}} \right) \\ &= \left( E_{\text{elec}} \cdot (m - 1) + E_{\text{elec}} \cdot m \right. \\ &\quad \left. + \sum_{i=1}^m \begin{cases} E_{\text{fs}} \cdot d_{x_i, x_{i+1}}^2, & d_{x_i, x_{i+1}} \leq d_0 \\ E_{\text{mp}} \cdot d_{x_i, x_{i+1}}^4, & d_{x_i, x_{i+1}} > d_0 \end{cases} \right) \\ &= \left( E_{\text{elec}} \cdot (2m - 1) + \sum_{i=1}^m \begin{cases} E_{\text{fs}} \cdot d_{x_i, x_{i+1}}^2, & d_{x_i, x_{i+1}} \leq d_0 \\ E_{\text{mp}} \cdot d_{x_i, x_{i+1}}^4, & d_{x_i, x_{i+1}} > d_0 \end{cases} \right). \end{aligned} \quad (11)$$

$$\tau_{ij}(t+n) = (1 - \rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}, \quad \Delta\tau_{ij} = F_{\text{best}}. \quad (12)$$

**Step 3 (pheromone updating).** Based on the objective function, the BS can find the best route in one iteration. CPEH then uses the max-min ant system model [29] to update the pheromone. The max-min ant system model will only update the pheromone on the path of the best route. Hence, the convergence rate will be improved.

The pheromone is updated according to (12).  $\rho$  is the pheromone decay coefficient used to escape the local optima.  $F_{\text{best}}$  is the best objective function value in this iteration.

The BS will repeat those steps until the iteration time is reached. The best route information will be broadcasted to all the CHs at the end of the setup phase.

**4.2.3. The CHR Strategy.** The fuzzy logic system of CPEH is aimed at selecting the sensor with a good local energy state to become the CH. However, the heavy workload may still exhaust the CH's energy halfway in the working round, especially when the frame number is large. Most clustering protocols cannot handle this condition, causing the whole cluster

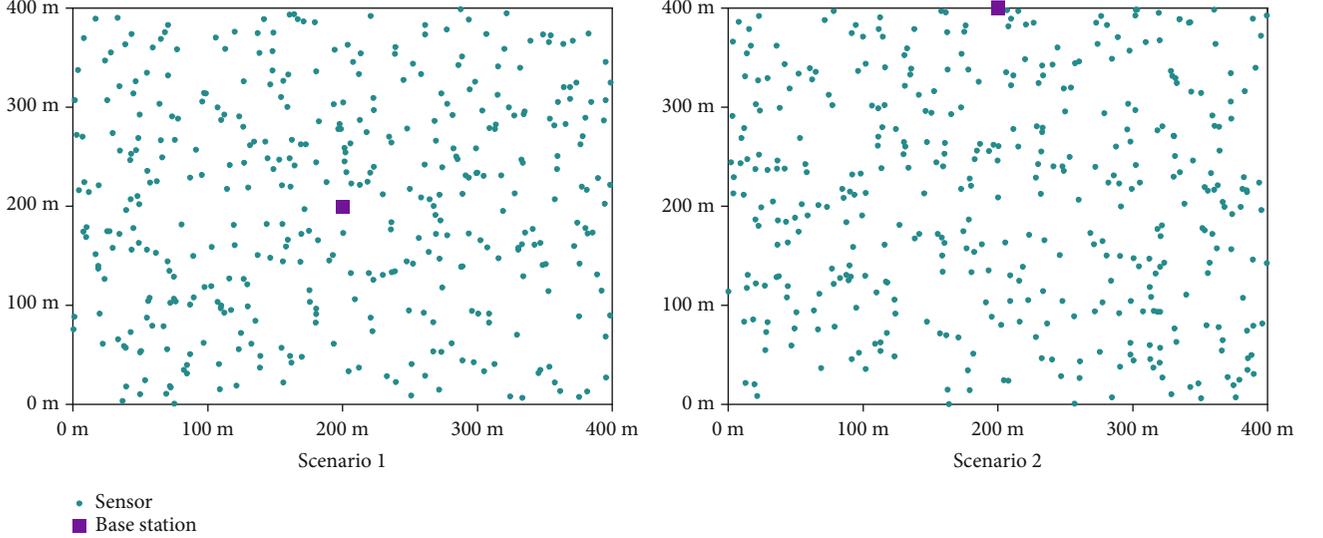


FIGURE 10: The sensors and BS deployment case in different scenarios.

to be silent until the end of the working round and wrecking the sensing accuracy. To solve this, CPEH utilizes the CHR strategy, a simple greedy approach, to help the depleted CH find the proper CM to undertake its job. Once the current CH finds that its energy has fallen below  $E_{\text{sleep}}$ , it will first choose  $k$ , which is set to 3 in this paper, the nearest awake CMs  $[CM_1, CM_2, \dots, CM_k]$ , and calculate the corresponding energy states for each chosen CM. Then, the CM with the best energy state will be selected as the new CH. The result will be broadcast to all the CMs in the *Slot CHR*.

We believe that the CHR strategy can maintain the cluster's stability and keep the intercluster routing effectively since only the CMs near the current CH will be the candidates. On the other hand, by choosing the CM with the best energy state among the candidates, the CHR strategy can also better utilize the network energy. The CHR strategy may be executed several times in one working round. It can visibly improve the network performance.

## 5. Performance Evaluation

**5.1. Simulation Setting.** This section compares the performance of CPEH with three popular clustering protocols: LEACH, EAUCF, and MOFCA. We focus on a  $400\text{m} \times 400\text{m}$  sensing area with 400 energy harvesting sensors randomly deployed throughout the network. The network is assumed to be homogeneous, with a data aggregation ratio of 0.1.

To have an intuitive insight into the advantages of CPEH, we consider two different scenarios. Precisely, Scenario 1: the BS is located in the middle of the network, and the harvest rates of sensors range from  $25\ \mu\text{W}$  to  $250\ \mu\text{W}$  randomly. Scenario 2: the BS is located on the edge of the network, and the harvesting rates range from  $50\ \mu\text{W}$  to  $500\ \mu\text{W}$ . We show a typical deployment case in Figure 10.

We define each round as 0.1 h, and the network total simulation time is 72 h. To test the network performance under different network loads, we change the number of the frames in each round from 50 to 120. Ten different sensor deploy-

TABLE 2: The corresponding simulation parameters used in this paper.

Parameters	Values
Network area	[400 m, 400 m]
Number of sensors	400
Base station location	(200, 200), (200, 400)
Sensor initial energy	1 J
Sensor energy capacity	2 J
Number of frames	50-120
Sensor energy harvesting rate	( $25\ \mu\text{W}$ - $250\ \mu\text{W}$ ), ( $50\ \mu\text{W}$ , $500\ \mu\text{W}$ )
Data packet size	500 byte
Control packet size	25 byte
$E_{\text{elec}}$	50 nJ/bit
$E_{\text{fs}}$	10 pJ/bit/m <sup>2</sup>
$E_{\text{mp}}$	0.0013 pJ/bit/m <sup>4</sup>
$E_{\text{DA}}$	5 nJ/bit/signal
$d_0$	87 m
Competition radius	$R = 60\text{ m}$
Impact factor in fuzzy logic	$\alpha = 2, \beta = 1$
Number of ants	10
Number of iterations in ACO	20
Initial pheromone trail density	$\tau_{i,j}(0) = 10$
Impact factor in ACO	$\theta = 1, \nu = 3$
Constant factor $Q$	$Q = 0.05$
Pheromone decay coefficient	$\rho = 0.25$
Number of CHR candidates	$k = 3$

ments are randomly generated in each scenario to get the average result to eliminate the contingency. The relevant parameters are summarized in Table 2. The simulation is

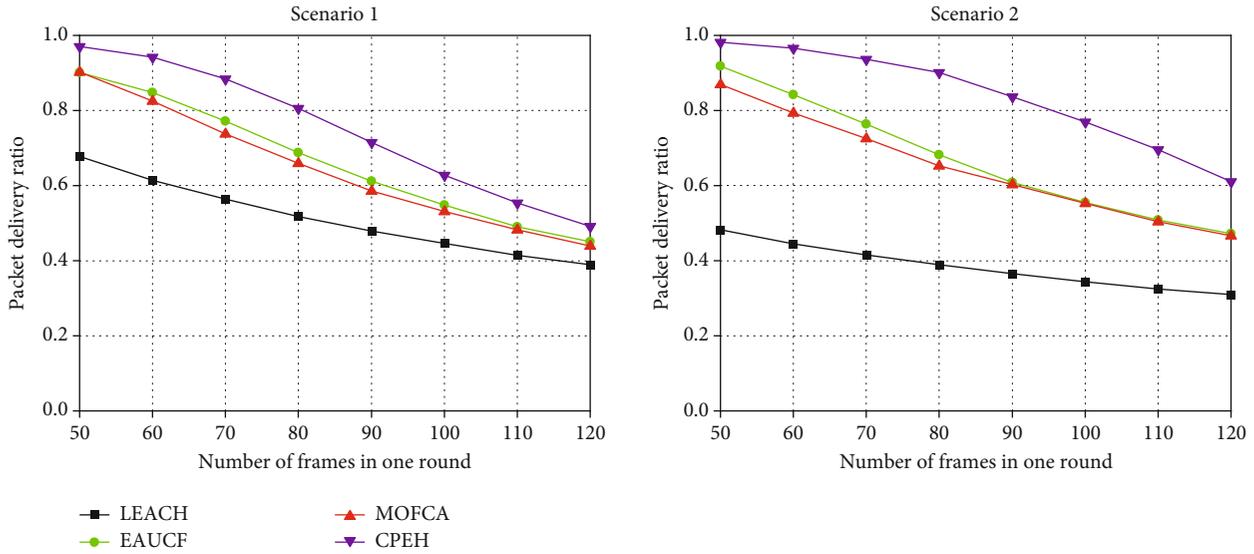


FIGURE 11: Network packer delivery ratio in different scenarios.

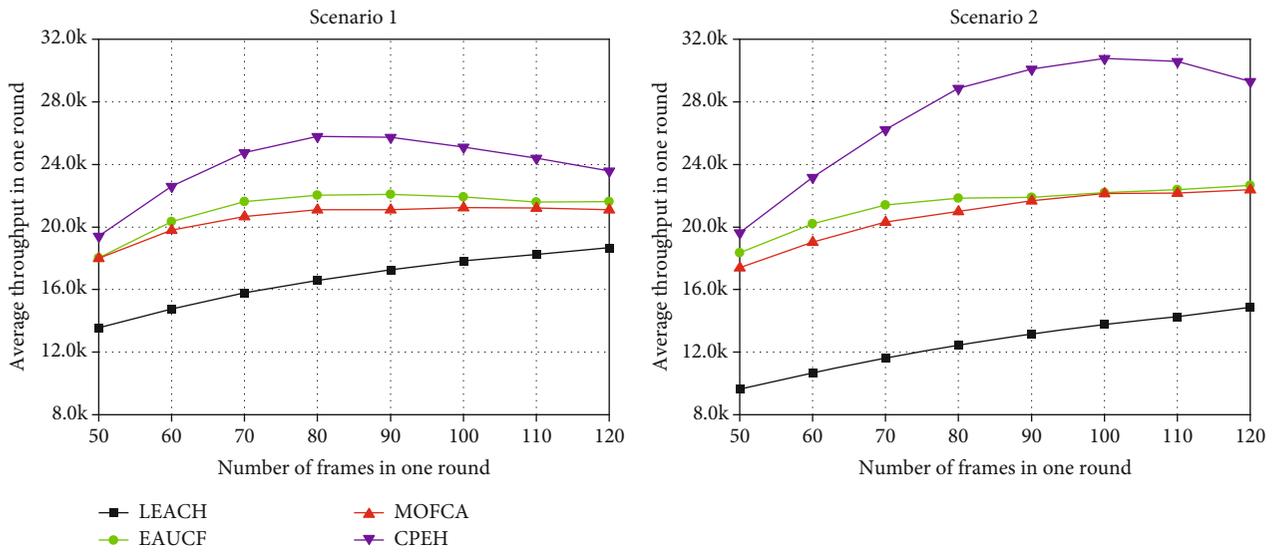


FIGURE 12: Average network throughput in one round in different scenarios.

executed by Matlab. All the energy dissipation of control packets and data packets is considered to guarantee veracity. The result is summarized in Section 5.2.

5.2. Simulation Result

5.2.1. Packet Delivery Ratio. Figure 11 shows the network packet delivery ratio under different frame numbers in two scenarios. With the increase of the frame number, the delivery ratio will gradually decrease under all the protocols, which can be easily explained. The more frame number, the more energy shall be needed to handle the data packets, leading to more packet losses owing to the energy shortage. However, it is noteworthy that CPEH outperforms others in all the conditions. For example, when the frame number is 80, in Scenario 1, CPEH can achieve a 55.6%, 17.1%, and 22.2%

increment over LEACH, EAUCF, and MOFCA separately. The improvement comes to 131.7%, 32%, and 38.2% in Scenario 2. We believe that the superiority of CPEH is owing to the more reasonable cluster and route construction.

5.2.2. Network Throughput. We illustrate the average network throughput in one round under different clustering protocols in Figure 12. When the frame number is low, increasing the frame number will improve the throughput under all the protocols. However, the throughput improvement becomes weak or even negative if the frame number is large. This phenomenon shows that we cannot improve the network throughput by naively assigning more frames in the working round. CPEH can achieve a 38.1%, 16.8%, and 21.5% and 107%, 35.9%, and 37.4% higher max

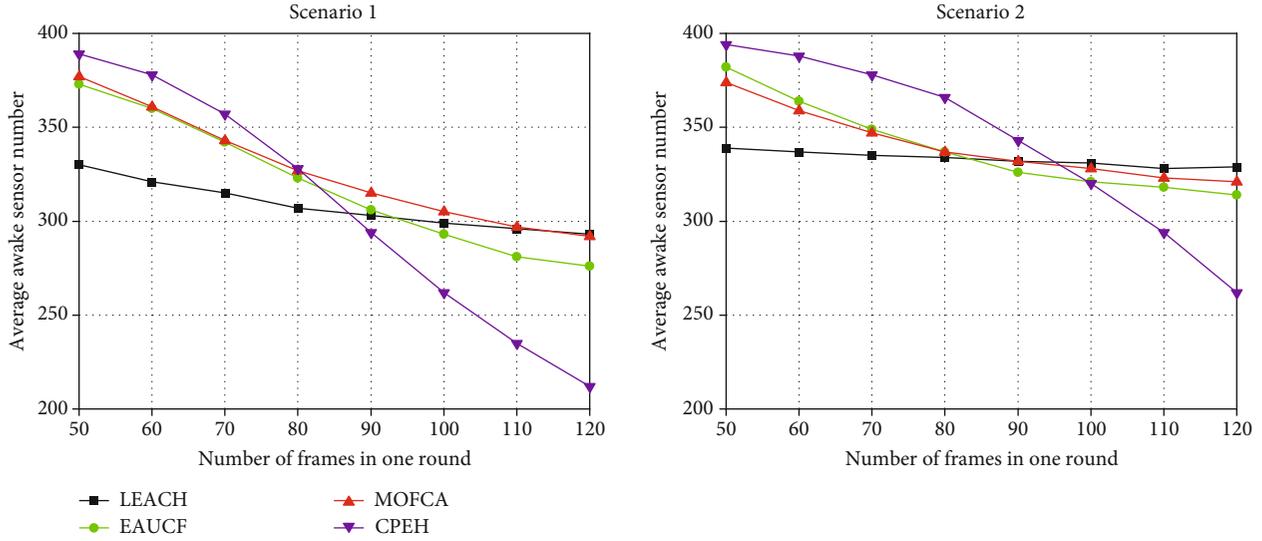


FIGURE 13: Average awake sensor number.

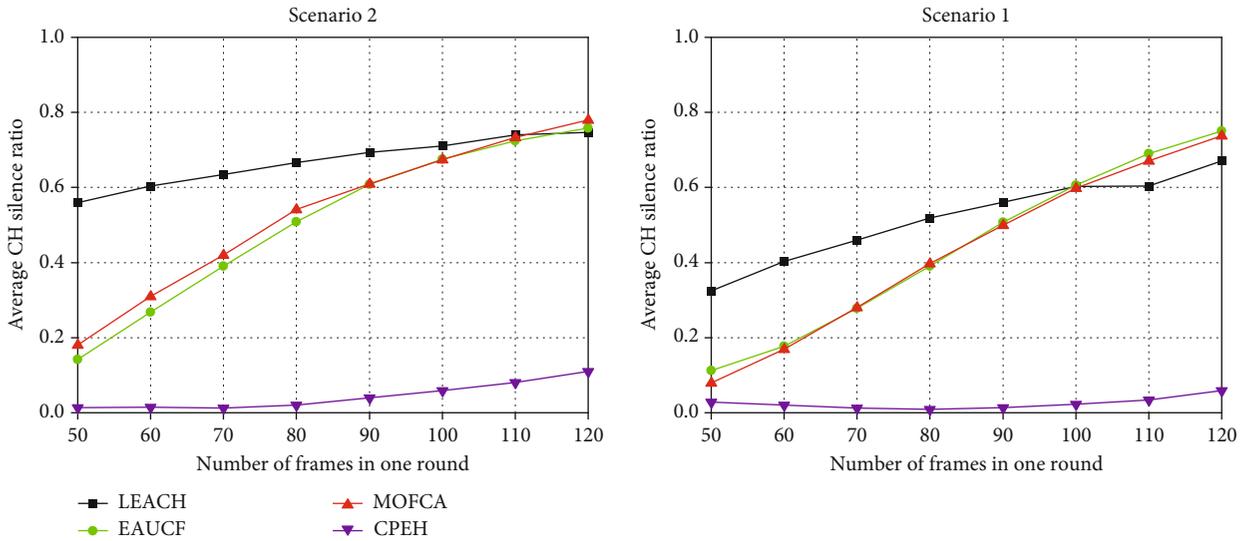


FIGURE 14: Average cluster dormancy ratio.

throughput in Scenario 1 and Scenario 2 than LEACH, EAUCF, and MOFCA, respectively, whereas the throughput and packet delivery ratio in CPEH has a trade-off relation, which shall be carefully considered.

**5.2.3. Awake Sensor Number and the CH Dormancy Ratio.** Since the sense is a cooperative work in the homogeneous network, the number of awake nodes will affect the sensing accuracy. We record the average number of awake nodes at every round beginning and summarize the result in Figure 13. The result shows that the average awake sensor number under CPEH will suffer a sharper decrease with the frame increase. Therefore, even though the CPEH can keep the network with more awake sensors when the frame number is relatively small, this index of CPEH becomes the worst under a large frame number. We think this phenomenon is

owing to the CHR strategy of CPEH. LEACH, EAUCF, and MOFCA do not consider the situation where the CH turns to sleep in the middle of the working round. In that case, even though CMs have enough energy to continue their work, they have to keep silent until the round ending, which will cause the next round to have more awake sensors. However, the CHR strategy can effectively solve this problem and let the awake CMs with enough energy continually work. We have to point out that even though CPEH may have fewer awake sensors in some conditions, it can still achieve a better delivery ratio and throughput.

Figure 14 shows the average CH dormancy ratio under different frame numbers. We can observe that the CHR strategy of CPEH can successfully avoid the cluster being silent in the working round. On the contrary, clusters under three other protocols may lose their function and become silent

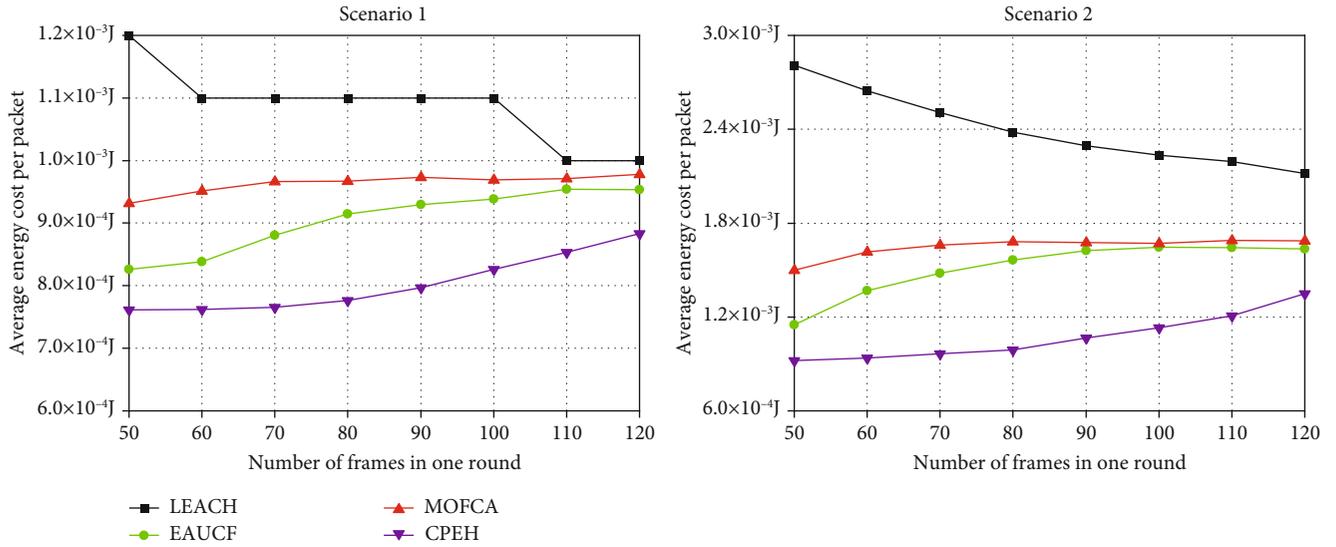


FIGURE 15: Average cluster silence ratio.

halfway, especially when the frame number is large. This conclusion can also prove that it is worthy of sacrificing few awake sensor numbers by taking the CHR strategy.

**5.2.4. Average Energy Cost per Packet.** We summarize the average energy cost of transmitting a data packet under different protocols and show the result in Figure 15. It is no doubt that CPEH is the most economical protocol in all network conditions, which proves its efficiency. We believe that, except for the more appropriate cluster construction and intercluster routing, the CHR strategy also contributes.

For EAUCF and MOFCA, when the CH falls into sleep during the working round, the routing topology will be changed. The last-hop and next-hop CHs of the sleeping CH will have to communicate directly, causing the long-range transmission and destroying the energy efficiency. Conversely, the CHR strategy can keep the intercluster routing topology robust and stable, leading to continuous, efficient intercluster communication.

## 6. Conclusion and Future Work

In this paper, we propose CPEH, which is a clustering protocol designed explicitly for EHWSN. CPEH mainly consists of two parts. The first part focuses on cluster construction. We adopt the fuzzy logic system to handle the uncertain nature of EHWSN and construct clusters more appropriately. The second part utilizes the ACO algorithm to optimize the intercluster routing, which can inherently achieve a better path than greedy algorithms used in most clustering protocols. We execute a comprehensive simulation of CPEH with some representative clustering protocols under different network conditions. The result proves that CPEH can always achieve the best performance in network delivery ratio and throughput. Furthermore, the CHR strategy of CPEH can effectively solve the cluster dormancy problem, ensuring the cluster works normally. The advantages of CPEH make it a suitable

protocol for Industry 4.0 era applications. For future work, we will extend CPEH to handle the multi-BS condition and consider sensors' movement.

## Data Availability

The authors do not list the raw simulation data in this paper owing to the spatial confined. Readers can get the detailed data by connecting to the corresponding author.

## Conflicts of Interest

The authors have declared that no competing interests exist concerning this study.

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under project contract nos. 61701082, 61701116, 61601093, 61971113, and 61901095, in part by National Key R&D Program under project contract nos. 2018YFB1802102 and 2018AAA0103203, in part by Guangdong Provincial Research and Development Plan in Key Areas under project contract nos. 2019B010141001 and 2019B010142001, in part by Sichuan Provincial Science and Technology Planning Program under project contract nos. 2018HH0034, 2019YFG0418, 2019YFG0120, 2020YFG0039, and 2018JY0246, in part by Ministry of Education China Mobile Fund Program under project contract no. MCM20180104, in part by Yibin Science and Technology Program—Key Projects under project contract nos. 2018ZSF001 and 2019GY001, and in part by Central University Business Fee Program under project contract no. A03019023801224 and Fundamental Research Funds for the Central Universities under Grant ZYGX2019Z022.

## References

- [1] S. Wang, J. Wan, D. Zhang, D. Li, and C. Zhang, "Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination," *Computer Networks*, vol. 101, pp. 158–168, 2016.
- [2] J. Wan, S. Tang, Z. Shu et al., "Software-defined industrial Internet of things in the context of industry 4.0," *IEEE Sensors Journal*, vol. 16, p. 1, 2016.
- [3] X. Li, D. Li, J. Wan, A. V. Vasilakos, C. F. Lai, and S. Wang, "A review of industrial wireless networks in the context of Industry 4.0," *Wireless Networks*, vol. 23, no. 1, pp. 23–41, 2017.
- [4] M. M. Krunz and T. Shu, "Coverage-time optimization for clustered wireless sensor networks," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 1, pp. 202–215, 2010.
- [5] M. Youssef, A. Youssef, and M. Younis, "Overlapping multi-hop clustering for wireless sensor networks," *IEEE Transactions on Parallel & Distributed Systems*, vol. 20, no. 12, pp. 1844–1856, 2009.
- [6] P. Ding, J. A. Holliday, and A. Celik, "Distributed energy-efficient hierarchical clustering for wireless sensor networks," in *IEEE International Conference on Distributed Computing in Sensor Systems*, Berlin, Heidelberg, 2005.
- [7] N. Sabor, M. Abo-Zahhad, S. Sasaki, and S. M. Ahmed, "An unequal multi-hop balanced immune clustering protocol for wireless sensor networks," *Applied Soft Computing*, vol. 43, pp. 372–389, 2016.
- [8] Z. Hanzalek and P. Jurcik, "Energy efficient scheduling for cluster-tree wireless sensor networks with time-bounded data flows: application to IEEE 802.15.4/ZigBee," *IEEE transactions on industrial informatics*, vol. 6, no. 3, pp. 438–450, 2010.
- [9] M. L. Ku, W. Li, Y. Chen, and K. R. Liu, "Advances in energy harvesting communications: past, present, and future challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1384–1412, 2017.
- [10] S. Ulukus, A. Yener, E. Erkip et al., "Energy harvesting wireless communications: a review of recent advances," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 3, pp. 360–381, 2015.
- [11] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 3, pp. 443–461, 2011.
- [12] S. Peng and C. P. Low, "Prediction free energy neutral power management for energy harvesting wireless sensor nodes," *Ad Hoc Networks*, vol. 13, pp. 351–367, 2014.
- [13] D. Hasenfratz, A. Meier, C. Moser, J. J. Chen, and L. Thiele, "Analysis, comparison, and optimization of routing protocols for energy harvesting wireless sensor networks," in *IEEE International Conference on Sensor Networks*, Newport Beach, CA, USA, 2010.
- [14] B. Baranidharan and B. Santhi, "DUCF: distributed load balancing unequal clustering in wireless sensor networks using fuzzy approach," *Applied Soft Computing*, vol. 40, pp. 495–506, 2016.
- [15] J. M. Kim, S. H. Park, Y. J. Han, and T. M. Chung, "CHEF: cluster head election mechanism using fuzzy logic in wireless sensor networks," in *2008 10th International Conference on Advanced Communication Technology*, vol. 1, pp. 654–659, Gangwon, Korea (South), 2008.
- [16] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient protocol for wireless microsensor networks," in *Proceedings of the 33rd annual Hawaii international conference on system sciences*, Maui, HI, USA, 2000.
- [17] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks, conference IEEE international performance," in *PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference*, Phoenix, AZ, USA, 2005.
- [18] O. Younis and S. Fahmy, "HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366–379, 2004.
- [19] G. Chen, C. Li, M. Ye, and J. Wu, "An unequal cluster-based routing protocol in wireless sensor networks," *Wireless Networks*, vol. 15, no. 2, pp. 193–207, 2009.
- [20] N. M. A. Latiff, C. C. Tsimenidis, and B. S. Sharif, "Energy-aware clustering for wireless sensor networks using particle swarm optimization," in *2007 IEEE 18th international symposium on personal, indoor and mobile radio communications*, pp. 1–5, Athens, Greece, 2007.
- [21] Y. Han, G. Li, R. Xu, J. Su, J. Li, and G. Wen, "Clustering the wireless sensor networks: a meta-heuristic approach," *IEEE Access*, vol. 8, pp. 214551–214564, 2020.
- [22] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Applied Soft Computing*, vol. 13, no. 4, pp. 1741–1749, 2013.
- [23] S. A. Sert, H. Bagci, and A. Yazici, "MOFCA: multi-objective fuzzy clustering algorithm for wireless sensor networks," *Applied Soft Computing Journal*, vol. 30, pp. 151–165, 2015.
- [24] S. Peng, T. Wang, and C. P. Low, "Energy neutral clustering for energy harvesting wireless sensors networks," *Ad Hoc Networks*, vol. 28, pp. 1–16, 2015.
- [25] J. Li and D. Liu, "DPSO-based clustering routing algorithm for energy harvesting wireless sensor networks," in *International Conference on Wireless Communications & Signal Processing*, pp. 1–5, Nanjing, China, 2015.
- [26] Y. Dong, J. Wang, B. Shim, and D. I. Kim, "DEARER: a distance and-energy-aware routing with energy reservation for energy harvesting wireless sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3798–3813, 2016.
- [27] T. Ross, *Fuzzy-Logik. Fuzzy Logic with Engineering Applications*, McGraw-Hill, Inc, 1995.
- [28] M. Dorigo and C. Blum, "Ant colony optimization theory: a survey," *Theoretical Computer Science*, vol. 344, no. 2–3, pp. 243–278, 2005.
- [29] J. Bai, G. K. Yang, Y. W. Chen, L. S. Hu, and C. C. Pan, "A model induced max-min ant colony optimization for asymmetric traveling salesman problem," *Applied Soft Computing*, vol. 13, no. 3, pp. 1365–1375, 2013.