Simplified clustering and improved intercluster cooperation approach for wireless sensor network energy balanced routing

Yanxin Yao[1], Wei Chen[2,3,4]* , Jie Guo[1], Xiaoyu He[1] and Ruixuan Li[1]

* Correspondence: chenwdavior@
163.com
[2]School of Computer Science and
Technology, China University of
Mining Technology, Xuzhou 221116,
Jiangsu, China
[3]Mine Digitization Engineering
Research Center of the Ministry of
Education, China University of
Mining and Technology, Xuzhou
221116, Jiangsu, China
Full list of author information is
available at the end of the article

Abstract

The equilibrium use of energy is very important for wireless sensor networks (WSN) with limited energy in order to avoid premature network collapse. The existing methods either need too complex calculations for precise clustering, or are too simple to overburden a few cluster heads. In order to solve these problems, we proposed energy balanced clustering routing (EBCR) in this paper. It could maximize the WSN life in energy non-harvesting scenario or improve energy utilization efficiency in energy harvesting scenario without increasing the amount of calculations. It gives a complete solution to the process of cluster head election, clustering, and intercluster routing algorithm. Firstly, a light weight cluster head election and a distributed cluttering method are proposed by introducing dynamic cluster radius and intersection region node division schemes with new principles. Thus, lightweight distributed clustering achieves the advantages of balancing the burden of cluster heads and alleviating hot zone problem. Then we optimized the cluster cooperative routing algorithm by analyzing cooperation and competition among cluster heads. The intercluster cooperative routing algorithm greatly improves the transmission efficiency between cluster heads. Moreover, this paper analyzes the reasons why the algorithm achieves more balanced energy usage, higher energy efficiency, and fewer calculations compared to the existing mainstream algorithms. At last, simulation results show that EBCR algorithm has advantages in terms of network energy consumption, number of surviving nodes in energy non-harvesting scenario compared with the delay-constrained energy-efficient cluster-based multi-hop routing (DCEM) method. Simulation also gives EBCR algorithm performance under various energy harvesting scenarios, which is quite satisfactory in energy utilization efficiency comparing with DCEM method. EBCR algorithm has superior performance in terms of balanced energy usage, low computation complexity, and high energy efficiency.

Keywords: WSN, Clustered routing, Hot zone, Clustering, Cluster head selection, Relay selection

1 Introduction

The WSN consists of a large number of inexpensive micro sensor nodes deployed in the monitoring area, and forms a self-organizing network system through wireless communication. The purpose of the WSN is to cooperatively sense, collect, and process the information of the sensing objects in the monitoring area, and send them to the observer. The biggest drawback of WSN is that the node energy is limited, and resources such as storage space and computing power are also very limited. With limited hardware resource, large-scale data collection can easily cause excessive energy consumption of nodes. If the node energy decays too quickly and the number of exhausted nodes is too high, network congestion will increase unreliable data transmission in the WSN [1]. Therefore, designing an energy-efficient network protocol to balance the nodes' energy usage and maximize network lifetime has always been the research focus of WSN. Because long distance transmission cost lots of energy, multi-hop transmission is a better solution under limited energy condition. According to the different topology structures, the routing protocols of WSN can be divided into planar routing protocols and hierarchical routing protocols [2, 3]. In planar routing protocol, the status of the network nodes is equal, but the application is limited due to the poor scalability and small network size. For a network of large size, each node directly transmitting information to the Sink node will cause too poor network scalability, too large energy consumption, and too heavy control load of the Sink node. In order to avoid these problems, WSN is often managed as a hierarchical network. Clustered routing protocols are able to manage a larger area without reducing the quality of service [4–7]. So it has been widely used due to their excellent performance in scalability, fault tolerance, and energy saving. In order to avoid the loss of network connectivity and premature network collapse due to the excessive exploiting of individual nodes, balanced energy clustering routing protocol needs to be studied under the premise that each node can transmit its own data to the Sink node [8]. When the energy of the nodes in the sensor network can be supplemented, an energy harvesting sensor network is formed. In order to make the energy efficiency of the network as high as possible [9, 10], simplify the routing maintenance cost and reduce the cluster update rate; it is also necessary to make the energy consumption of the sensor nodes as balanced as possible [11]. This paper focuses on energy-balanced clustering routing schemes. The delay-constrained energy-efficient cluster-based multi-hop routing (DCEM) algorithm [12] is analyzed in detail and the problems appearing are the motivation for this research. Considering design steps of cluster head election, clustering, and inter-cluster routing algorithm, a new complete solution is proposed in this paper. Without increasing the amount of calculations, it has achieved longer network life or higher energy efficiency than the existing algorithms. In this paper, a new method called energy balanced cluster and routing (EBCR) is proposed to solve the energy balance clustering routing problem. The major contributions of this research are as follows:

1   A complete energy balance clustering routing scheme is proposed, which can maximize the WSN life of non-energy collection scenarios or increase the

energy utilization of energy collection scenarios without increasing the amount of calculation.

2   By introducing a dynamic clustering radius with a new principle and a node division scheme for crossing regions, a lightweight clustering head election and distributed clustering method is proposed. The overall advantages of this distributed clustering method are lightweight distributed clustering, balancing the burden of cluster heads, and alleviating hot zone issues.

3   An optimized inter-cluster cooperative routing algorithm is proposed. It gives a complete consideration for the relay cooperation principle, and the cooperative competition relationship among cluster heads, which greatly improves the transmission efficiency between cluster heads.

4   This paper analyzes the reasons why the algorithm achieves more balanced energy usage, higher energy efficiency, and smaller calculations compared with the existing mainstream algorithms.

5   Simulation results show that EBCR algorithm performs better in terms of network energy consumption, has more nodes survived, and cost less energy compared to DCEM.

The paper is organized as follows: In Section 2, The WSN cluster and routing method is divided into two categories for review: uniform clustering and non-uniform clustering. It also summarizes the motives, schemes, and advantages of the proposed method. In Section 3, this part deals with the network and energy model. In Section 4, we describe the problems existing in DCEM algorithm in aspects of cluster head election, clustering and inter-cluster routing mechanism. In Section 5, we propose EBCR method and points out its merits compared with the existing methods. In Section 6, simulations are performed to compare EBCR method and DCEM method. In Section7, we conclude the paper.

2 Related work

In WSN, the main characteristics of clustering routing protocols lie in the aspects of cluster head election, clustering, and inter-cluster routing. The clustering routing protocol is mainly divided into uniform clustering algorithm and non-uniform clustering algorithm according to the clustering radius. It will be described hierarchically.

2.1 Uniform clustering algorithm

We discussed about cluster head election and clustering mode regarding uniform clustering algorithm.

2.1.1 Cluster head election

In the WSN, the cluster head election mechanism of the uniform clustering algorithm mainly includes the following types. Low-energy adaptive clustering hierarchy (LEACH) [13] and threshold-sensitive energy-efficient sensor network protocol (TEEN) [14] algorithm randomly select their cluster heads. The node randomly generates 0–1 random numbers. If it is greater than a threshold, it can become a

cluster head. The threshold is calculated by the cluster head election probability and the current running round number of the network. The disadvantage of this method is that the cluster head selection is completely random, and it cannot guarantee the selected nodes are suitable. The cluster head selection of LEACH-C [15] algorithm is not random, but entirely controlled by the Sink node. By collecting the global node location and energy information, the Sink node calculates the mean energy in the network, clusters, and selects cluster heads for each cluster using the simulated annealing algorithm. The disadvantage is that it is only suitable for small-scale systems. Large-scale network flooding will lead to a lot of energy consumption and even network congestion. The cluster head selection of hybrid energy-efficient distributed clustering (HEED) [16] algorithm is based on the residual energy, the density of the nodes, and the proximity to its neighbor nodes. The cluster head is determined through multiple iterations, consuming a lot of energy by cyclic iteration. The delay-constrained energy multi-hop (DCEM) [5] algorithm proposed in 2016 is a high energy efficiency clustering multi-hop routing algorithm based on time constraints. The cluster head is selected in a distributed way according to the relative energy and distance information of the node. It has advantages in energy consumption and the number of surviving nodes compared with the classical algorithms, such as LEACH, HEED, and other algorithms [17]. A particle swarm optimization based energy-efficient cluster head selection (PSO-ECHS) algorithm selects the cluster head among the nodes that exceed the average residual energy, based on particle swarm optimization (PSO) according to the location and energy information of all node s[18].

2.1.2 Uniform clustering mode

In the clustering way of HEED algorithm, a node decides which cluster to join according to the average reachable energy of average minimum reachability power (AMRP) carried in the received cluster head message. This solves the problem of which cluster a node eventually joins when covered by multiple cluster heads at the same time. Nodes that do not receive the cluster header message identify themselves as "isolated nodes" and declare themselves as cluster heads. The disadvantage is that isolated nodes will be generated, and the distance factor from the neighbor node to the cluster head node is not considered. Moreover, large clusters will easily cause some nodes to die too quickly, thus affecting the stability of the network. Although DCEM algorithm is superior to classical algorithms such as LEACH, HEED, etc., in terms of energy consumption and number of surviving nodes. The DCEM clustering method has the shortcomings: as long as the node is within the communication distance of the dominant node, the dominant node is used as its cluster head, causing the dominant node in a certain area to occupy the entire area. The dominant node will quickly fail, requiring frequent replacement of clusters. If the update time is slightly slower than the failure time, the cluster information will be lost. In the clustering method of PSO-ECHS, the non-cluster head node calculates the weight function of the cluster head, which takes into account the energy, distance, and node degree of the cluster head, and chooses the cluster with the maximum weight of the cluster head.

2.1.3 Non-uniform clustering algorithm

The above algorithms are all uniform clustering algorithms, which divide the network into multiple clusters of similar size, where the numbers of members in clusters is almost the same due to the uniform distribution of nodes. Because the cluster head nodes close to the Sink node need to forward data from other clusters, they will exhaust their energy too early and fail into the dormant state, causing network segmentation and reducing the network survival time. The researchers call this a "hot zone" problem. In order to solve the "hot zone" problem, non-uniform clustering methods came into being. The representative non-uniform clustering algorithms are energy-efficient non-uniform clustering (EEUC) algorithm [19, 20], unequal clustering algorithm for WSN based on fuzzy logic and improved ACO (UCFIA) algorithm [21], and uneven clustering dynamic routing based on energy efficiency and balanced (UDEB) algorithm [22]. EEUC algorithm is an energy-efficient non-uniform clustering algorithm. In each clustering period, the node first randomly selects the candidate cluster head according to probability T, and then selects the cluster head according to the residual energy in the candidate cluster head. In order to construct a non-uniform cluster, EEUC algorithm proposes a clustering competition radius related to the node position. The closer the node is to the Sink node, the smaller is the clustering radius. However, due to the difficulty of selecting the parameter of the EEUC algorithm for controlling the competition radius, the performance is restricted. During the election course, if a candidate cluster head Si announces its election victory, all the candidate cluster heads within 's competition radius cannot become the final cluster head and need to withdraw from the election process. When the candidate cluster head node competes, each candidate cluster head broadcasts a message to announce whether it is successful and at the same time receives such messages sent by the neighboring candidate cluster head nodes. This negotiation mechanism to determine the cluster head requires broadcasting a large number of messages, consuming a lot of energy. In the inter-cluster routing stage, each cluster head selects a cluster head with more residual energy as the next hop relay node, in each round of multi-hop link selection, and does not consider whether it is beneficial to save energy. Moreover, in the stable transmission phase of the EEUC algorithm, each cluster head will transmit data according to a fixed link. This will cause the nodes on the link to consume too much energy and a rapid death [19, 20]. UCFIA algorithm is a distributed self-organizing clustering algorithm. In the cluster head election, UCFIA algorithm uses local information such as the residual energy of the node, the distance from the Sink node, and density as the reference index for the cluster head election and cluster competition radius, and constructs the non-uniform cluster using the fuzzy theory model. In the data transmission phase, multi-hop routing is selected using an adaptive max-min ant colony optimization (ACO) algorithm. The UCFIA algorithm uses fuzzy theory and ACO to construct an energy-efficient WSN, which simplifies the clustering difficulty. However, many parameters in the UCFIA algorithm, such as the maximum local density and the maximum competition radius, are not optimal and can be further improved. These parameters can be further optimized to improve the performance of the network [21]. UDEB is a non-uniform clustering algorithm for dynamic routing. The algorithm takes the

Sink node as the center and divides the network from inside to outside into concentric circles with uniformly spacing. The interval between the rings is equal. Through the energy consumption analysis of the single-hop routing model and the multi-hop routing model, when the ring spacing is within the critical distance of the network, the nodes between the rings can perform multi-hop communication, and the nodes in a ring perform single-hop communication to reduce unnecessary energy consumption. The UDEB algorithm calculates the energy consumption of the whole network theoretically, determines the optimal number of cluster heads in each ring, calculates the cluster head's competition radius according to the optimal number of cluster heads, and guarantees a cluster head in each area in the network. The cluster heads will be elected. The UDEB algorithm considers the energy cost and the residual energy comprehensively when choosing the routing, and gives the routing probability in a weighted manner, which simplifies the complexity of the algorithm. UDEB algorithm solves the problem of energy hole by constructing a non-uniform clustering network of sub-rings.

However, the cluster head node uses two weights when choosing the routing, does not study these two parameters quantitatively, which restricts the improvement of the life cycle of WSNs [22]. In summary, the clustering methods such as distributed clustering of DCEM and EEUC are comparatively more advantageous, because of the small amount of computation in the energy-efficient clustering and multi-hop routing algorithms. Non-uniform clustering methods such as UCFIA and UDEB requiring globally centralized calculations are effective, but the amount of calculation is large [23–25]. The selection of inter-cluster multi-hop routing adopted by DCEM is based on minimizing energy consumption, which can lead to uneven energy consumption of nodes in the network [26–28].

2.2 A better solution

The paper mainly studies the distributed clustering method. A distributed clustering routing is proposed in this paper with a clustering method which combines a greatly simplified cluster head selection, dynamic cluster radius, cross-regional division, and an optimized routing scheme for inter-cluster cooperative communication. In the cluster head selection, nodes with high residual energy (supply energy) are directly selected as cluster head nodes, which could be performed distributedly, and the entire network coverage can be quickly realized without calculating and determining the optimal cluster head number. Because only the energy of the node is considered without considering information such as the location and density of the node, it is less affected by external changes and suitable for large-scale and dynamically changing networks. The clustering method in this paper contains a dynamic cluster radius and division of nodes in the intersection region scheme. The dynamic cluster radius scheme adopts a mechanism that a larger node energy and greater distance from the sink node corresponds to a larger cluster radius, so as to alleviate the hot zone problem. Based on the selected cluster head and cluster radius, the intended responsible area of the cluster head is formed. Due to simplified cluster head and cluster radius election methods, the number of cluster heads may be relatively large, and cluster heads may be relatively dense in local areas. To solve this problem, intersection region node division and intercluster cooperative routing algorithms are proposed. For the intersection regions covered by

multiple cluster heads, the cluster division principle is derived, which not only considers the energy factor but also considers the distance factor. The advantages are lightweight distributed clustering, balancing the burden of cluster heads and alleviating hot zone issues. In order to avoid the direct information transmission from each cluster head to the Sink node resulting in excessive energy consumption, a more effective intercluster communication collaboration scheme is adopted in this paper. Intercluster cooperation can only be used when it is beneficial to energy saving. The principle of cooperation among cluster heads is derived in this paper. It not only gives out when cooperation is beneficial but also considers the cooperative competition relationship among cluster heads. Thus, a complete intercluster routing multi-hop algorithm is designed, which is simpler, more energy efficient, and more balanced in energy consumption compared to DCEM's minimal energy path search method. Compared with the EEUC intercluster routing algorithm, the burden of the next hop candidate node is more reasonably analyzed and only the next hop node that is beneficial to energy saving is selected. No fixed routing is taken, so the energy of the cluster head is more balanced and energy efficient. The intercluster routing multi-hop algorithm, along with the simplified cluster head election and cluster intersecting area division methods together solves the problem of transmission efficiency when the cluster heads are densely distributed. As a whole, this method can achieve more balanced energy consumption compared to distributed clustering methods such as DCEM and EEUC, effectively reducing cluster head failure probability, greatly improving cluster update time, reducing clusters maintenance costs, and overcoming the disadvantages of the UCFIA and UDEB non-uniform clustering methods that require a lot of centralized calculations.

3 Network and energy model

3.1 Network model

A group of wireless sensor nodes randomly distributed in the area are considered. The hierarchical network model is shown in Fig. 1 set as follows:
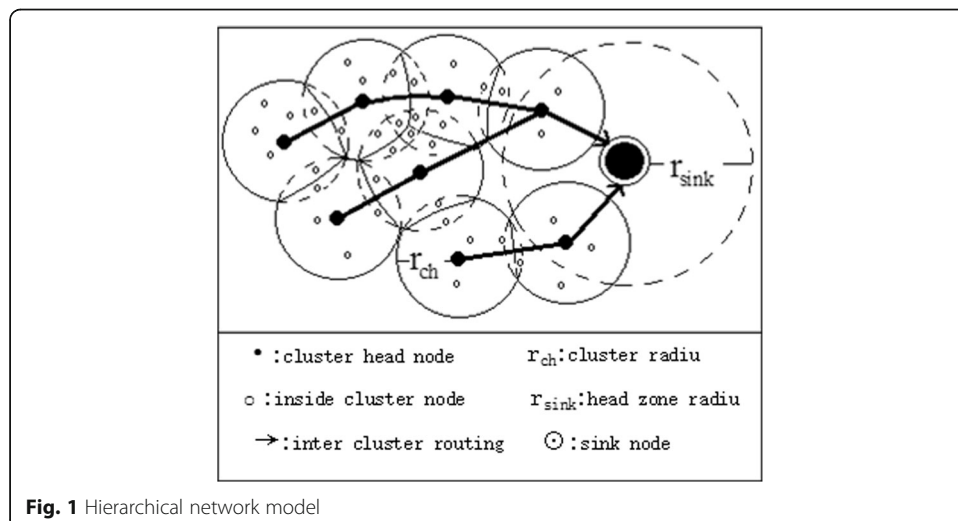
Fig. 1 Hierarchical network model

1 The WSN includes nodes acting as three kinds of nodes, namely Sink node, cluster head node, and cluster member node.

2 Node location is fixed during network operation.

3 The energy of the Sink node is unlimited and the communication range covers the entire network.

4 Other nodes could have energy supplement from environment. Each node can obtain its own location information and remaining energy information, and can adjust the transmission power according to the transmission distance.

5 The node can calculate the distance from the sending node based on the signal strength contained in the received information.

6 The cluster head nodes aggregate the data of the nodes in the cluster and transmit the integrated data to the Sink node through multiple hops.

7 The data sensed by the node is related, so the cluster head node can merge the collected information to reduce the total sent data.

3.2 Wireless channel model

In the wireless channel, the power of the electromagnetic wave weakens as the distance between the receiver and the transmitter increases. The energy consumed by a wireless sensor node to send and receive l bit information can be represented as E_{Rx} and E_{Tx}, respectively.

$$E_{Tx}(l,d) = \begin{cases} l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2, d < d_0 \\ l \cdot E_{elec} + l \cdot \varepsilon_{amp} \cdot d^4, d \geq d_0 \end{cases} \tag{1}$$

$$E_{Rx}(l,d) = l \cdot E_{elec} \tag{2}$$

where E_{elec} is the circuit energy loss coefficient for sending and receiving unit bit information. If the transmission distance is less than the threshold d_0, power amplifier losses use a free-space model. Otherwise, the attenuation model is used. ε_{fs} and ε_{amp} are the required energy for each bit of information sending by the amplifier in the two models, respectively. The smaller the d_0 is, the greater the probability of using the multipath attenuation model is, leading a greater the energy consumed and shorter the network lifetime. Commonly used constant values are as follows:

$$d_0 = 60\,\text{m}, E_{elec} = 50\text{nJ/bit}, \varepsilon_{fs} = 10\text{pJ}/(\text{bit} \cdot \text{m}^2), \varepsilon_{amp} = 0.0013\text{pJ}/(\text{bit} \cdot \text{m}^4).$$

3.3 Energy consumption calculation

Because the nodes in the cluster only need to send the collected data to the cluster head node, the energy consumption of the member nodes in each cluster is:

$$E_{mem}(j) = l \cdot E_{elec} + l \cdot \varepsilon_{fs} \cdot d^2(j) \tag{3}$$

where $d(j)$ is the distance between the member node j in the cluster and its cluster head node. Because the cluster head node needs to merge the data of all nodes in the cluster, and then transmits the integrated data. The energy consumption of each cluster head node is:

$$E_{CH}(i) = E_R(i) + E_F(i) + E_{Tx}(i) \tag{4}$$

$$E_R(i) = l \cdot E_{elec} \cdot Num_{CH}(i) \tag{5}$$

$$E_F(i) = l \cdot E_{fuse} \cdot Num_{CH}(i) \tag{6}$$

In the above formula, $E_R(i)$ is the energy consumed by the cluster head node i to collect the data of all nodes in the cluster. $E_F(i)$ is the energy consumed by the cluster head node i for data fusion of the collected node data in the cluster. $Num_{CH}(i)$ is the number of nodes in the cluster that belong to cluster head node i. $E_{Tx}(i)$ is the energy consumption of the cluster head node i to transmit l bit data to other cluster head nodes or sink node, calculated by (1). The constant $E_{fuse} = 5nJ/bit$.

3.4 Energy harvesting model

Assume that all wireless sensor nodes are equipped with energy harvesting devices, such as solar panels. The available energy may be momentarily different on a single node. At the same time, there may be variations in harvested energy for different nodes. We assume that each node has a separate energy harvesting rate. The harvested energy is stored in a storage device, $E_{S,\,i}$ represents the stored energy, and $E_{M,\,i}$ represents the maximum battery capacity. Considering the main energy cost, the energy model of EH-WSN for node i is

$$E_i(\tau) = \begin{cases} \min\left(E_i(\tau-1) + P_{EH,i}(\tau-1), E_{M,i}\right) - E_{mem}(i), & i \text{ is a member node in a cluster} \\ \min\left(E_i(\tau-1) + P_{EH,i}(\tau-1), E_{M,i}\right) - E_{CH}(i), & i \text{ is a cluster head node} \end{cases} \tag{7}$$

Here, a discrete-time system is considered in each sensor node. At the end of each time slot τ, $E_i(\tau)$ represents the remaining energy of node i. $E_i(\tau-1)$ represents the residual energy of node i at the end of the last time slot. At the beginning of each time interval τ, node i receives the energy supplement accumulated in the previous time slot, expressed as $P_{EH,\,i}(\tau-1)$. At any time, the maximum energy at node i is not allowed to exceed $E_{M,\,i}$. Each node should keep $E_i(\tau) > 0$; otherwise, the node fails until it has harvested enough energy to start again.

4 Previous method-DCEM algorithm

The DCEM algorithm proposed in 2016 is an energy-efficient clustering multi-hop routing algorithm with time constraint. Because the algorithm selects the cluster heads in a distributed way according to the relative energy and distance information of the nodes, it is superior to classical algorithms such as LEACH, HEED, etc. in terms of energy consumption, the number of surviving nodes, and so on. The DCEM clustering method has the following problems: as long as the node is within the communication distance of the dominant node, the dominant node is used as the cluster head, causing the dominant node in a certain area to occupy the entire area, and the dominant node will quickly fail, requiring frequent replacement of clusters. If the update time is slightly slower than the failure time, the cluster information will be lost. Intercluster routing in DCEM uses minimal energy consumption routing. The specific algorithm is described as follows.

4.1 Algorithm description

The clustering method in DCEM algorithm is as follows. (1) The sink node broadcasts ADV information with a certain power and starts the neighbor node discovery process. (2) Each node calculates the distance to the Sink node based on the strength of the received signal. (3) Each node has a timer and each node waits, and then starts sending ADV (ID, E) information to the neighbor nodes. ID is the node number, and E is the remaining energy of the node. Each node compares the energy in itself and the ADV information. If the energy of the node is low, the timer is turned off. This node is a member node that belongs to the cluster head with higher energy. In this way, a node with large energy first sends out information, and a node with smaller energy receives the information from a large energy node and automatically becomes a member node in the cluster. (4) The cluster head candidate node is either a node that has transmitted ADV information but does not receive ADV information of other nodes because distance is too far away, or a node which has sent ADV information, and has received ADV information of other nodes, and has higher energy. When two cluster head candidate nodes have the same energy and can communicate, the node nearer to the sink node is selected. Each candidate cluster head node wait time is:

$$w = \frac{1}{\text{TED}_i} \tag{8}$$

$$\text{TED}_i = \left(\frac{E_i}{E_{\text{total}}}\right)^{\alpha} + \left(\frac{1}{d_{(i,s)}}\right)^{\beta} \tag{9}$$

where TED_i is the energy and delay equalization factor of the candidate cluster head node i. E_i is the remaining energy of the candidate cluster head node i. E_{total} is the sum of energy of other candidate cluster head nodes in the received ADV information. $d_{(i,s)}$ is the distance from the cluster head candidate node's i to the Sink node. It can be seen that the node's waiting time is shorter for a node with larger energy and smaller distance from the Sink node. The node earlier sends out the information that it is cluster head nodes, and the candidate cluster head nodes that receive this information becomes member nodes in the cluster. The disadvantage of this method is that as long as a node is within the communication distance of the most dominant node, it becomes a member node. This makes the dominant node in a certain area occupy the entire area. That is, in the area where the clusters intersect, the intersection area is assigned to the cluster head node with the largest energy. This will lead to many problems, as indicated in Section 4.2. In terms of intercluster routing, DCEM selects the multi-hop routing with the least total energy consumption and does not consider energy balance.

4.2 Problem description

The clustering method of DCEM will cause many problems. When large energy nodes with small energy gap are near enough, the node with relatively smaller energy will not be selected as cluster head and the energy will be wasted until clusters are updated. Two examples are given to explain the problem. As shown in situation 1 in Fig. 2, it is assumed that the ADV information of node A is (ID = A, E = 10) and the ADV information of node B is (ID = B, E = 9). According to the DCEM, since the energy of node A is higher than the energy of node B; node A is a cluster head, and B is the node in

Fig. 2 Situation 1 for DCEM problem
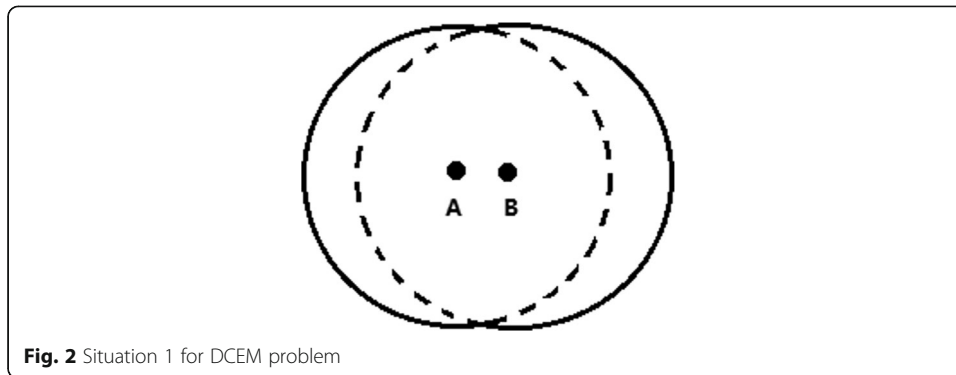
the cluster. The intersecting area is completely divided to node A. The energy difference between node A and node B is not significant, but it is obvious that the load of node A is much greater than that of node B. The energy of node A will quickly deplete, leading to a frequent replacement of cluster heads. If the update time is slightly slower than the failure time, it will lose this cluster information. As shown in situation 2 in Fig. 3, it is assumed that the ADV information of nodes A, B, and C is (ID = A, E = 10), (ID = B, E = 9), (ID = C, E = 5). Node A and node C are cluster heads of two clusters, and node B is in the intersecting area. According to the DCEM algorithm, since the energy of node A is higher than that of node C, the intersecting area is owned by node A. Although node B has a higher load capacity than node C because of its higher energy, node B cannot be a cluster head and its energy is wasted or only able to fully be utilized when clusters are updated. In order to solve the above problems, a new cluster routing scheme is proposed.

5 Proposed method-EBCR algorithm

In this paper, a new clustering routing algorithm EBCR is proposed, with low complexity cluster head election and distributed clustering scheme to cover the entire network more simply and quickly. An optimized intercluster routing scheme and the intersection region node division are proposed in order to solve the problem that cluster heads may be densely distributed in the method.

5.1 Cluster head selection

When the network is initialized, the Sink node collects the ID information, location information, residual energy information, and energy harvesting rate of the entire network nodes and stores them, and then updates the residual energy of the nodes. The

Fig. 3 Situation 2 for DCEM problem
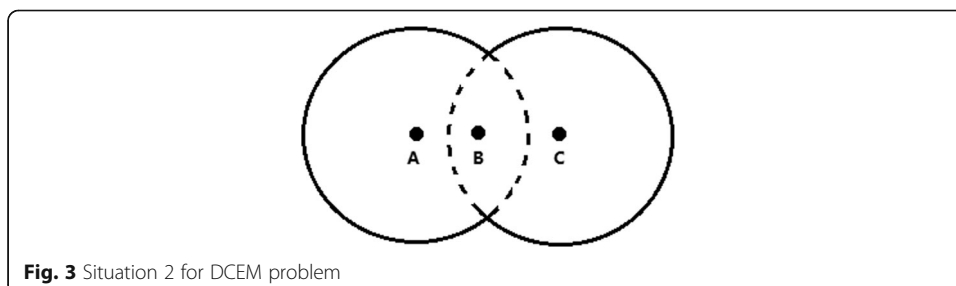
Sink node sorts all nodes in descending order of energy. Take the first 10% of the nodes as cluster heads and search for the nearby nodes with radius r_{ch} to form clusters. If the entire area cannot achieve full coverage, for the remaining unclustered nodes, take the nodes whose energy is the top α_{head} = 10% as the cluster heads, and search for the nearby nodes with radius r_{ch} to form clusters. This loop iterates until the entire area can be fully covered. In the cluster head selection, nodes with relatively high residual energy (supply energy) are directly selected as cluster head nodes. The advantage lies in that the entire network coverage can be quickly realized distributedly without calculating and determining the optimal cluster head number, along with the distributed clustering method. Because only the energy is considered, without considering information such as the location and density of the node, it is suitable for large-scale and dynamically changing networks. It is more rational compared to the random selection of cluster heads in EEUC method, and has lower computation compared to those requiring complicated calculations.

5.2 Dynamic cluster radius

The network is divided into multiple clusters of similar size. Due to the uniform distribution of nodes, the number of members in the cluster is also approximately the same. The cluster head nodes close to the Sink node need to forward data from other clusters, causing them in overload state, and leading to network segmentation and shorter network survival time. The researchers call this "hot zone" problem. In order to solve the "hot zone" problem, non-uniform clustering of dynamic cluster radius is considered in this paper. According to the principle that larger node energy and greater distance from the sink node corresponds to a larger search radiusr_{ch}, the search radius r_{ch} is

$$r_{ch} = \left\{ \alpha_1 \frac{d_{i\_sink}}{d_{max}} + \beta_1 \frac{\min\left[E_i(\tau-1) + P_{EH,i}(\tau-1), E_{M,i}\right]}{E_{M,i}} \right\} \cdot \alpha_3 \cdot d_0 \tag{10}$$

$$\alpha_1 + \beta_1 = 1 \tag{11}$$

where d_{i_sink} is the distance from any cluster head node i to the sink. $d_{max}$is the farthest distance from the selected cluster head nodes to the sink. $E_i(\tau - 1)$ is the residual energy of the cluster head node i at the end of the previous time interval τas shown in (7). At the beginning of the time interval τ, node i checks the energy harvested in the previous time slot denoted as $P_{EH,\ i}(\tau - 1)$. $E_{M,\ i}$ is the maximum battery capacity. At any time, the maximum energy at node i cannot exceed $E_{M,\ i}$. d_0 is the transmission threshold in the wireless channel model. α_1, β_1, and α_3 are parameters controlling cluster radius.

5.3 Node division in intersecting regions

The intended responsible area of the cluster head is formed according to the cluster head and cluster radius. Due to simplified cluster head and cluster radius election methods, the number of cluster heads may be relatively large, and cluster heads may be relatively dense in local areas. For the nodes located in the intersecting region of the clusters, distance from the cluster head node, the remaining energy of the cluster head, the energy absorption rate, and the distance of the cluster head from the sink should be considered when clustering. The cluster head node close to the sink node is

overloaded because it needs to forward data from other clusters. In Fig. 4, the distance from the cluster head node $i_1(i_2)$ and the sink node is smaller than certain threshold, and the distance from the cluster head node j_1 and the sink node is smaller than certain threshold. In order to avoid "hot region" problem, nodes in the intersecting regions of cluster head nodes i_1 and j_1 are assigned to cluster head j_1. The nodes in the intersecting regions of cluster head nodes of i_1 and i_2 are assigned to be the intra-cluster member nodes of the cluster head node with a large f which is calculated as follows:

$$f_{m\_j} = \alpha_2 \min\left[E_j(\tau-1) + P_{EH,j}(\tau-1), E_{M,j}\right] + \frac{\beta_2}{d_{m\_j}} \tag{12}$$

where f_{m_j} represents the fitness value of the node m in the intersection region to the cluster head j. $E_j(\tau-1)$ denotes the residual energy of the cluster head j at the end of the previous time interval τ as shown in (7). The supplementary energy accumulated by the node j in the previous time slot τ is $P_{EH,\,j}(\tau-1)$. $E_{M,\,j}$ is the battery capacity. The energy of node j cannot exceed $E_{M,\,j}$ at any time. d_{m_j} represents the distance from the node m in the intersection region to the cluster head j. α_2, and β_2 are weights that can be adjusted as needed. Another situation is shown in Fig. 5, when the distance from the cluster heads i_1, i_2, j_1 to the sink node are all larger than threshold, the nodes in the intersecting area of these nodes compute their $f$value respectively, and the nodes in the intersecting area become the intra-cluster members of the cluster head node with the largest $f$value.

Remark 1 *The EBCR algorithm reduces the burden on cluster heads and helps to avoid excessive consumption of cluster heads in hotspots, alleviating hot zone problems.*

The clustering method of DECR combining the methods of cluster head, dynamic cluster radius, and division of intersection regions solves the clustering problems of DCEM, in Figs. 2 and 3. In Fig. 2, energy of both A and B are relatively high and the energy difference is not large. The clustering method in DCEM makes the node with the larger energy in A and B become the cluster head, responsible for all the nodes within the cluster radius. The workload of cluster head nodes is large, leading to frequent replacement of cluster heads. The method proposed by us takes both nodes A and B as cluster heads. Because the energy difference between A and B is not large, the intersecting parts are mainly divided according to distance. The workload of cluster head nodes is obviously reduced, and cluster
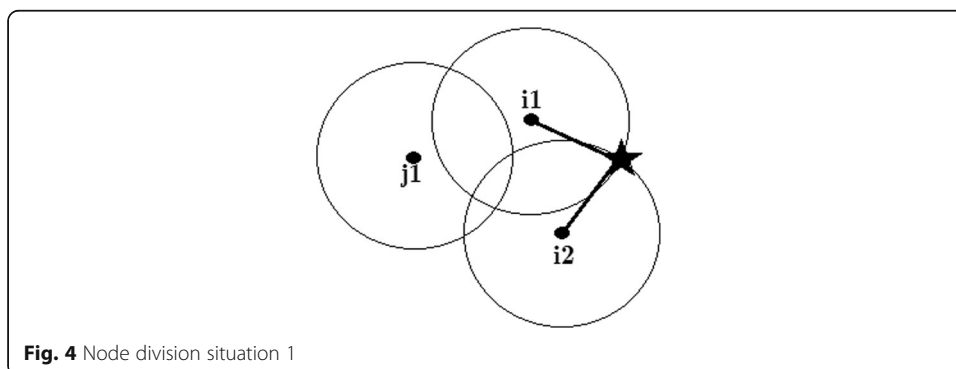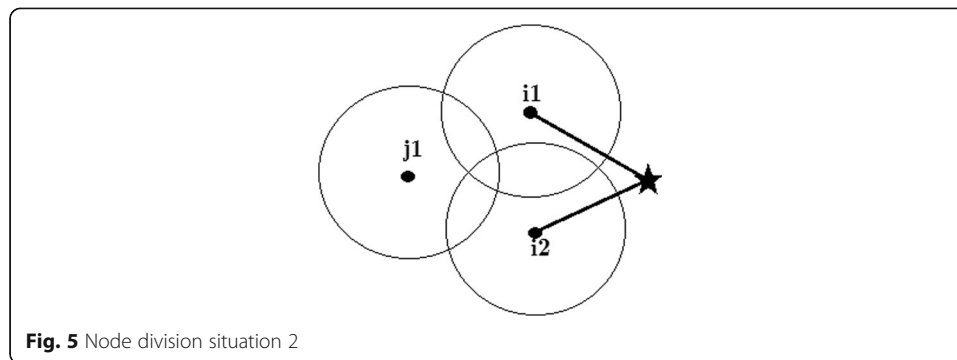

Fig. 4 Node division situation 1

Fig. 5 Node division situation 2

heads do not need to be replaced frequently. As shown in situation 2 in Fig. 3, DCEM clustering method wasted more energy in node B. According to our method, node B becomes a cluster head. This not only does not waste node B but also reduces the workload of nodes A and C. In summary, our proposed non-uniform clustering method better utilizes the nodes with relatively larger energy and allows the division of intersecting regions. It avoids a dominant node takes over all the nodes in its communication area. This could balance the burden on cluster heads and help to avoid excessive consumption of cluster heads in hotspots and thus alleviate "hot zone" problems. The above analysis shows that the EBCR clustering method is better than to DCEM. The clustering radius inherits considering the distance between the node and the Sink node from the EEUC. In addition, the energy of the cluster head is considered, which is more beneficial to energy balance. The clustering competition radius also has a clear selection basis. It is more beneficial to balance energy for its exploiting the relatively high energy nodes, instead of the EEUC clustering rule that does not allow cluster heads to appear within the competition radius of the selected cluster head. In addition, it avoids transmitting message to compete for cluster head node that EEUS needed and therefore saves a lot of energy. In the EBCR method, cluster head selection, clustering radius, and clustering methods are performed in a distributed, lightweight manner. The amount of calculation is much lower than the PSO-ECHS method based on particle swarm optimization. Because the consideration premise is that the wireless sensor nodes are distributed evenly, the factor of node density is not taken into consideration. So there is no complicated global calculation compared with UCFIA. Compared with UDEB, there is no theoretical calculation for energy consumption of the entire network to determine the optimal number of cluster heads in each ring, so the amount of calculation is low.

5.4 Relay communication analysis

Theorem 1. *When $d_1^2 + d_2^2 < d^2$, relay cooperation is beneficial for energy saving. d_1 is the distance from the source node to the cooperating node, d_2 is the distance from the cooperating node to the destination node, and d is the distance from the node to the destination node.*

Proof: With an ideal omnidirectional antenna, the free space loss is:

$$L = \frac{P_t}{P_r} = \frac{(4\pi d)^2}{\lambda^2} = \frac{(4\pi f d)^2}{c^2} \tag{13}$$

where P_t is the signal power of the transmitting antenna, P_r is the signal power of receiving antenna, λ is the carrier wavelength, d is the propagation distance between the antennas, and c is the speed of light. First, the favorable conditions for relaying will be analyzed under the simplified condition when the relay cluster head and the source data cluster head are responsible for the same amount of transmission data. As shown in Fig. 6, if the cluster heads A and B collect l bits data respectively and the total amount of data is $2l$ bits. Then

$$d_2^2 = d_1^2 + d^2 - 2d_1 d \, \cos a \tag{14}$$

$$L_{AC} = l \cdot (4\pi d/\lambda)^2 \tag{15}$$

$$L_{BC} = l \cdot (4\pi d_2/\lambda)^2 \tag{16}$$

$$L_0 = L_{AB} + L_{BC} \tag{17}$$

$$L_{AB} = l \cdot (4\pi d_1/\lambda)^2 \tag{18}$$

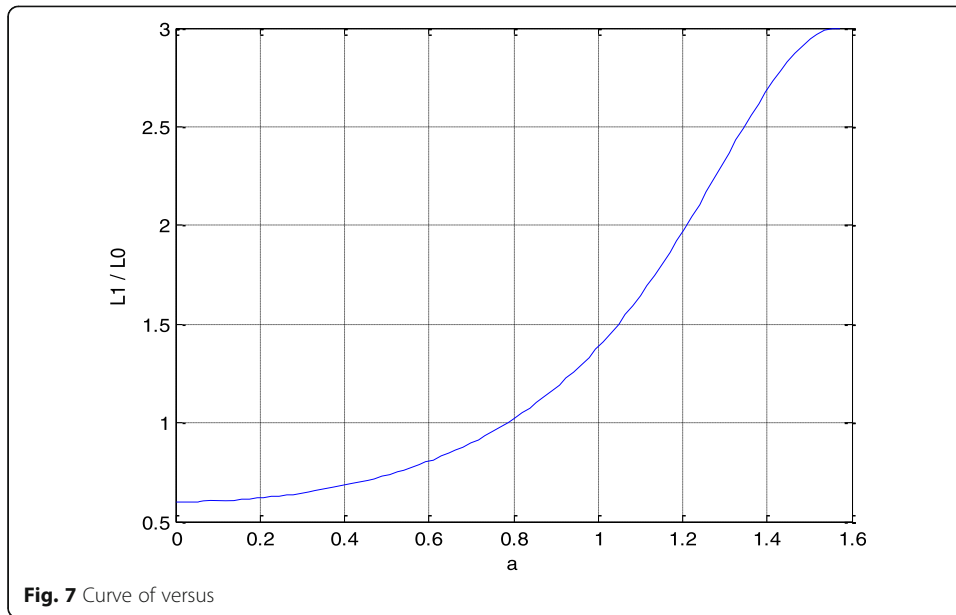$$L'_{BC} = 2l \cdot (4\pi d_2/\lambda)^2 \tag{19}$$

$$L_1 = L_{AB} + L'_{BC} \tag{20}$$

$$\frac{L}{L_0} = \frac{d_1^2 + 2d_2^2}{d^2 + d_2^2} = \begin{cases} \dfrac{3}{4\cos^2 a + 1}, d_1 = d_2 \\ 1 + \dfrac{2d_1^2 - 2d_1 d \, \cos a}{2d^2 + d_1^2 - 2d_1 d \, \cos a}, d_1 \neq d_2 \end{cases} \tag{21}$$

From (21), we can see that if $L_1/L_0 < 1$, then $d_1^2 + d_2^2 < d^2$. It is the cooperation beneficial condition. In particular, when $d_1 = d_2$, i.e., $a = \pi/4$, $L_1/L_0 = 1$. It can be seen from Fig. 7, when the angle $a < \pi/4$, the cooperative transmission is beneficial, and once again proves the conclusion that when $d_1^2 + d_2^2 < d^2$, cooperative transmission is beneficial.

Next, the favorable conditions for relaying will be analyzed under the condition when the amount of data taken charge of by relay cluster head and the source cluster head are not the same. In Fig. 8, the energy of cluster head A is greater than that of cluster head B. The intersection region needs to be divided. Assuming that the total data

Fig. 6 Cooperation derivation situation

Fig. 7 Curve of versus
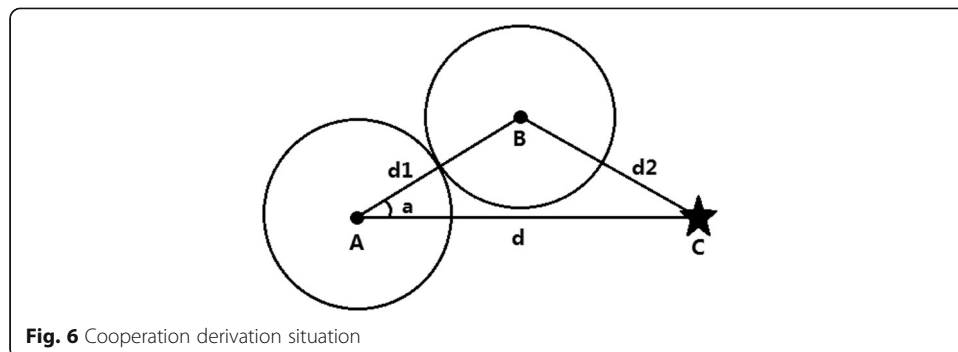
amount is $2l$ bits, if the amount of data collected by cluster head A is xl, the amount of data collected by cluster head B is $(2 - x)l$. We have

$$L_{AC} = xl \cdot (4\pi d/\lambda)^2 \tag{22}$$

$$L_{BC} = (2-x)l \cdot (4\pi d_2/\lambda)^2 \tag{23}$$

$$L_{AB} = xl \cdot (4\pi d_1/\lambda)^2 \tag{24}$$

$$L_{BC} = 2l \cdot (4\pi d_2/\lambda)^2 \tag{25}$$

$$\frac{L}{L_0} = \frac{L_{AB} + L_{BC}}{L_{AC} + L_{BC}} = \frac{xd_1^2 + 2d_2^2}{xd^2 + (2-x)d_2^2} \tag{26}$$

when $L_1/L_0 < 1$, then cooperation is beneficial, i.e., when $d_1^2 + d_2^2 < d^2$. Thus, Theorem 1 is validated.

5.5 Intercluster routing

The intercluster routing adopts the "single-hop" or "multi-hop" data transmission method and needs to be judged according to the distance between the cluster head and

Fig. 8 Cooperation derivation situation 2
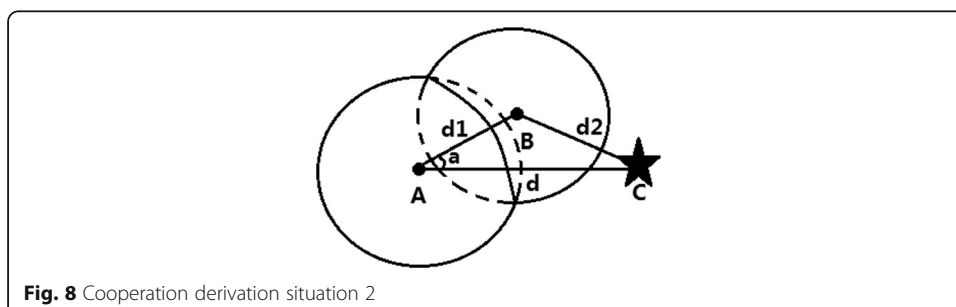
the Sink node. The distance d_{hop} is the critical distance. The distance from the cluster head to the base station is d_{i_sink}. When $d_{i\_\text{sink}} > d_{\text{hop}}$, if the cooperation beneficial condition is satisfied, from Section 5.4, then a "multi-hop" approach is used. Otherwise the "single-hop" approach is adopted, i.e., the cluster head node i sends information directly to the sink node. The "single-hop" approach is suitable for two situations. Either when $d_{i\_\text{sink}} < = d_{\text{hop}}$, or when $d_{i\_\text{sink}} > d_{\text{hop}}$, but the cooperation beneficial condition is not satisfied, the "single-hop" approach is adopted. The algorithm for inter-cluster "multi-hop" routing is as follows.

Algorithm 1 Inter-cluster "multi-hop" routing for EBCR

1: Sort all cluster heads in descending order of distance from sink.

2: $for(j = 0, j < n, j ++)$ % n is the number of cluster head nodes

3: { $while(d_{i\_\sin k} > d_{\text{hop}})$ % d_{hop} is the critical distance

4:      { $for(j = 0, j < n, j ++)$

5:          { $if(j ! = i \& \& d_{i\_j}^2 + d_{j\_\sin k}^2 < d_{i\_\sin k}^2)$

6:               { Calculate $G(j)$ as (27); }

7:          }

8: Take the cluster head node j with the largest $G(j)$ value as the next hop node of the cluster head node i；

9:      $i = j$；

10:     }

11: Output an inter-cluster route;

12: }

Sorting all the cluster head nodes in descending order of distance from the sink is to find the next hop node for the cluster head node farthest from the sink firstly until an intercluster route from the node to the sink is formed. Next, finding the next-hop node for the farthest distance among the remaining nodes outside the hot zone, until an intercluster route from the node to the sink is formed. Iterate through this loop until all cluster head nodes have explicit routes. From Section 5.4, it can be seen that cooperation are favorable when $d_1^2 + d_2^2 < d^2$. Therefore, in the fifth row of the algorithm, when finding the next hop node for the cluster head i, the node j that satisfies the $d_{i\_j}^2 + d_{j\_\sin k}^2 < d_{i\_\sin k}^2$ condition is the next hop candidate node. Calculate G value for each candidate node, and sorts them in descending order. The node j with the largest G value is selected as the next-hop node. Selecting a suitable node from the candidate next hop node j requires considering not only the current real-time residual energy of the candidate node but also considering the number of times that the candidate node helps other cluster head nodes to forward data. According to the principle that the next

hop node is selected as the candidate node with higher energy and fewer help forwarding times, the judgment value is

$$G(j) = \min\{E_j(\tau-1) + P_{EH,j}(\tau-1), E_{M,j}\} - E_R(j) - E_F(j) + \frac{\lambda}{N_j + 1} \tag{27}$$

where $G(j)$ is a judgment value for judging whether the candidate next hop node j is appropriate. Sort G for candidate nodes in descending order. The node with largest G is selected as the next hop node. $E_R(j)$ is the energy consumed by cluster head node j for collecting the data of member nodes in clusters as (5). $E_F(j)$ is the energy consumed by the cluster head node j for fusing the collected member node data as (6). N_j is the number of packets that cluster head node j helps other cluster head nodes to forward data, and λ is a parameter.

Remark 2 *The proposed intercluster "multi-hop" routing Algorithm 1 considers and solves the competing relationship between nodes.*

As shown in Fig. 9, nodes 2, 3, and 4 can be used as candidate next hop nodes for node 1 and node 5. If node 1 and node 5 both select node 2 as a cooperation node according to (27), will there be competition? According to the intercluster "multi-hop" routing algorithm, it is known that the next-hop cooperative node is first searched for node 1 farthest from the sink, until an intercluster route from the node 1 to the sink node is formed. Then look for the next hop collaboration node for node 5 until the intercluster route from the node 5 to the sink node is formed. Because of the sequential order, it involves the competition between source nodes. Also, the next-hop cooperative node is determined according to (27). Considering the number of times to help forward the data, the competition between next hop nodes is also considered. The situation that node 2 is simultaneously the cooperating node for node 1 and node 5 can only take place under the following condition: after node 2 has helped node 1, it is still better than nodes 3 and 4 for helping node 5, according to (27). Then node 2 acts as a cooperator for node 5. It is also reasonable. Advantage explanation: The intercluster routing algorithm proposed in this paper shows that when the nodes meet the cooperation beneficial condition, relay cooperation transmission is performed. The principle of cooperation among cluster heads is derived in this paper, which not only gives cooperation beneficial condition but also gives the cooperation order among cluster heads, the cooperative competition relationship among cluster heads. A complete intercluster routing multi-hop algorithm is designed, which is simple, efficient, and energy balanced compared to DCEM's minimal energy path search method. Compared to the EEUC intercluster routing algorithm, it also considers the number of determining to help other nodes forward data in addition to considering the current energy, and more reasonably analyzes the burden of the next candidate node, so as to make cluster
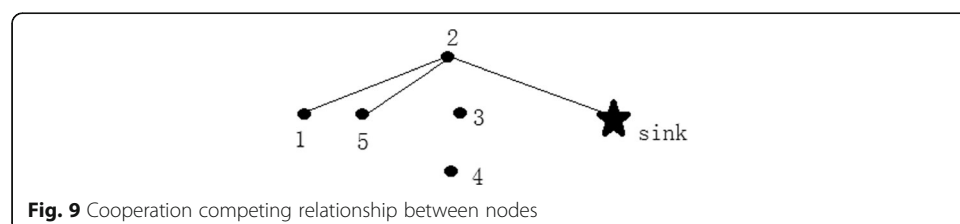
Fig. 9 Cooperation competing relationship between nodes

head energy more balanced. More important, collaboration is only done when the co-operation is favorable for energy saving, compared with EEUC and DCEM. The choice of routing is more flexible and judged on each transmission. The route selection is more flexible, and it is judged on each transmission, which avoids the node energy consumption on the path generated by the fixed routing method being too fast. The inter-cluster route multi-hop algorithm along with the simplified cluster head election and the cluster intersection region division method greatly improves the transmission efficiency between cluster heads, and solves the transmission efficiency problem when cluster heads are densely distributed.
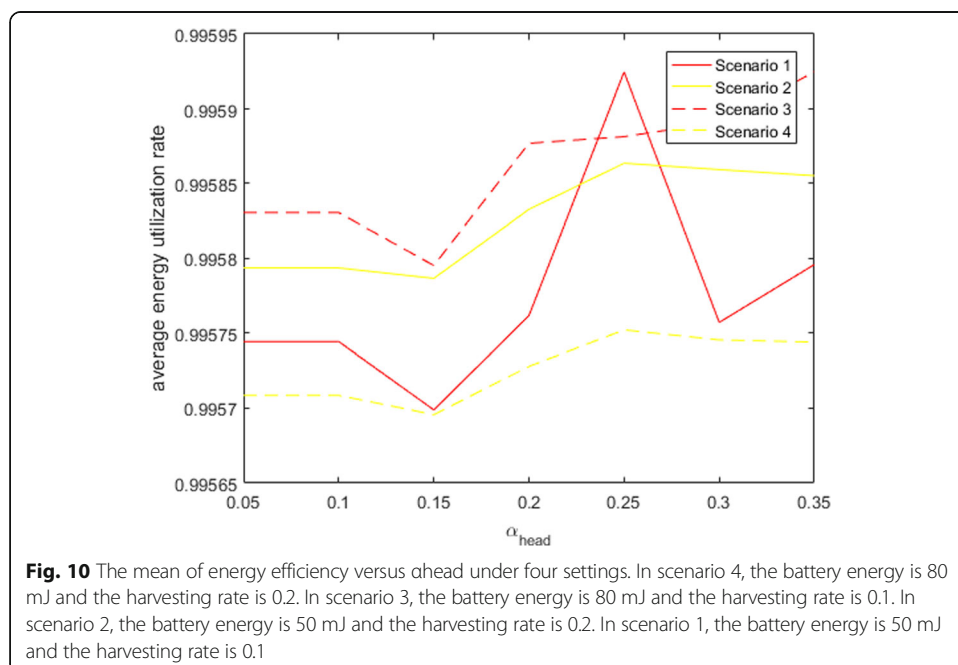
5.6 Parameter study

There are several important parameters with the method. In this subsection, we study how parameter α_3, α_{head}, and d_{hop} affect the performance of network, and also find out what values should be used for these three parameters through experiment. Here, we choose energy efficiency (equals to the rate of $[E_{\text{consume}}(k) + E_{\text{left}}(k)]/[E_{\text{harvest}}(k) + E_{\text{initial}}]$) as the metric to measure performance. The general simulation setting is discussed in Section 7.
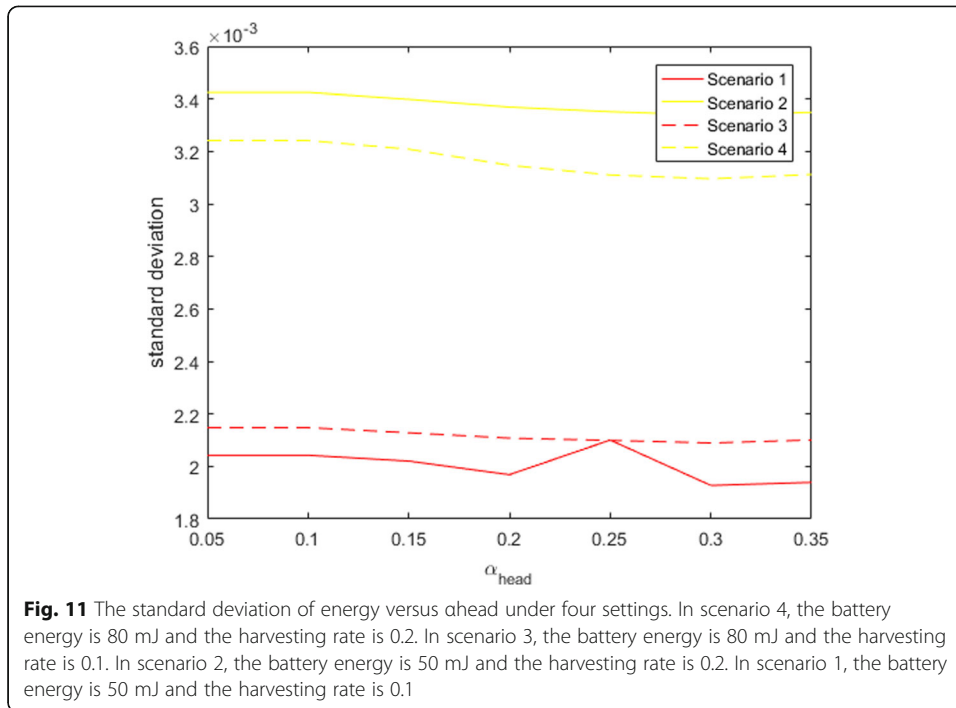
1   The parameters affecting the number of cluster heads

The radius r_{ch} is the search radius of node to form clusters. α_3 is the parameter controlling the cluster radius r_{ch} relative to the transmission distance threshold d_0. α_{head} is the rate to choose residual nodes as cluster head nodes. The larger α_3 is, the more nodes could be included in the selected cluster heads, and the fewer nodes will be remained unclustered, which leads to less probability that new cluster heads come into being. The larger α_{head} is, the more cluster heads will be selected generally, which leads to more cluster heads. So α_3 and α_{head} are important parameters to control the number of cluster heads. Too few cluster heads will cause the cluster heads to decay quickly due to the heavy load of forwarding data for the nodes in the cluster and intercluster communication with relatively long distance, which is not favorable to energy balance use. However, too many cluster heads will increase total energy consumption because the energy consumed on intercluster communication is too large due to long distance, and the energy consumed on intra-cluster communication is too little. In limitation case, too small radius will lead to each node being cluster head on their own, which makes the network a non-hierarchical network. So for certain network, the parameters of α_3 and α_{head} must be carefully chosen. We use the experiment to study the parameters. In our simulation, there are 100 nodes uniformly deployed in the 100 m × 100 m area. The Sink node is located at (100, 50) m which is assumed to have unlimited energy. The commonly used constant value for energy consumption calculations is $E_{\text{fuse}} = 5$ nJ/bit, $E_{elec} = 50$ nJ/bit, $\varepsilon_{amp} = 0.0013$ pJ/(bit · m^4), $\varepsilon_{fs} = 10$ pJ/(bit · m^2). The packet length is $l = 100$ bit. $d_0 = 87$ m. The control parameters are $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$, $\lambda = 1$. With $\alpha_3 = 1/2$, we vary the value of α_{head} from 5 to 25% at space 5%, under four settings of energy initial and harvesting conditions. The four settings are as follows: The battery capacity is the same for all nodes, $E_M = 50, 80$ mJ, respectively. The initial energy is distributed randomly in the range of [0, 25], [0, 50] mJ, respectively. The harvested energy

in each time slot is 10%, 20% of the initial energy of the node, respectively. The mean of energy efficiency with different value of α_{head} after 1000 rounds are plotted in Fig. 10 under four different settings.
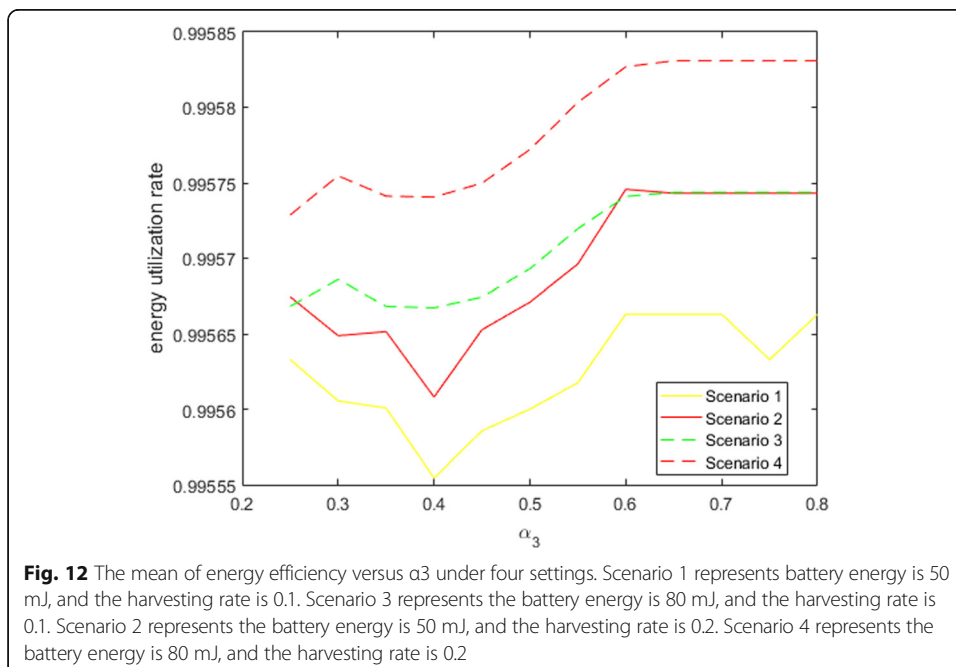
In Figs. 10 and 11, scenario 4, the battery energy is 80 mJ and the harvesting rate is 0.2. In scenario 3, the battery energy is 80 mJ and the harvesting rate is 0.1. In scenario 2, the battery energy is 50 mJ and the harvesting rate is 0.2. In scenario 1, the battery energy is 50 mJ and the harvesting rate is 0.1. It can be observed from the figures that the average energy efficiency is maximum and the standard deviation is minimum in each case when α_{head} is 0.25; that is, the energy can be best utilized and the energy used in each round is relatively stable. It can be found that α_{head} of 25% is enough. With α_{head} = 25%, changing the value of α_3 from 0.25 to 0.80 at space 0.05, under four settings of initial energy and harvesting conditions. The four settings are similar with the above four cases: The battery capacity is the same for all nodes, E_M = 50, 80 mJ, respectively. The initial energy is distributed randomly in the range of [0, 25], [0, 50] mJ, respectively. The harvested energy in each time slot is 10%, 20% of the initial energy of the node, respectively. The mean of energy efficiency and the packet drop rate with different value of α_{head} after 1000 rounds are plotted in Fig. 11, under four different settings. In Figs. 12 and 13, scenario 1 represents battery energy is 50 mJ, and the harvesting rate is 0.1. Scenario 3 represents the battery energy is 80 mJ, and the harvesting rate is 0.1. Scenario 2 represents the battery energy is 50 mJ, and the harvesting rate is 0.2. Scenario 4 represents the battery energy is 80 mJ, and the harvesting rate is 0.2. In four cases, when α_3 is 0.6, the drop rate is the lowest, the energy efficiency is the best, and the standard deviation is the smallest. Through comprehensive analysis, α_3 is selected as 0.6. And the range of node is 100 m × 100 m, relatively small for d_0 = 87 m. So whether the radius is small or large does not affect the node selection process much. When the node distribution range is larger, the effect of α_3 may be much larger.

Fig. 10 The mean of energy efficiency versus αhead under four settings. In scenario 4, the battery energy is 80 mJ and the harvesting rate is 0.2. In scenario 3, the battery energy is 80 mJ and the harvesting rate is 0.1. In scenario 2, the battery energy is 50 mJ and the harvesting rate is 0.2. In scenario 1, the battery energy is 50 mJ and the harvesting rate is 0.1

Fig. 11 The standard deviation of energy versus αhead under four settings. In scenario 4, the battery energy is 80 mJ and the harvesting rate is 0.2. In scenario 3, the battery energy is 80 mJ and the harvesting rate is 0.1. In scenario 2, the battery energy is 50 mJ and the harvesting rate is 0.2. In scenario 1, the battery energy is 50 mJ and the harvesting rate is 0.1
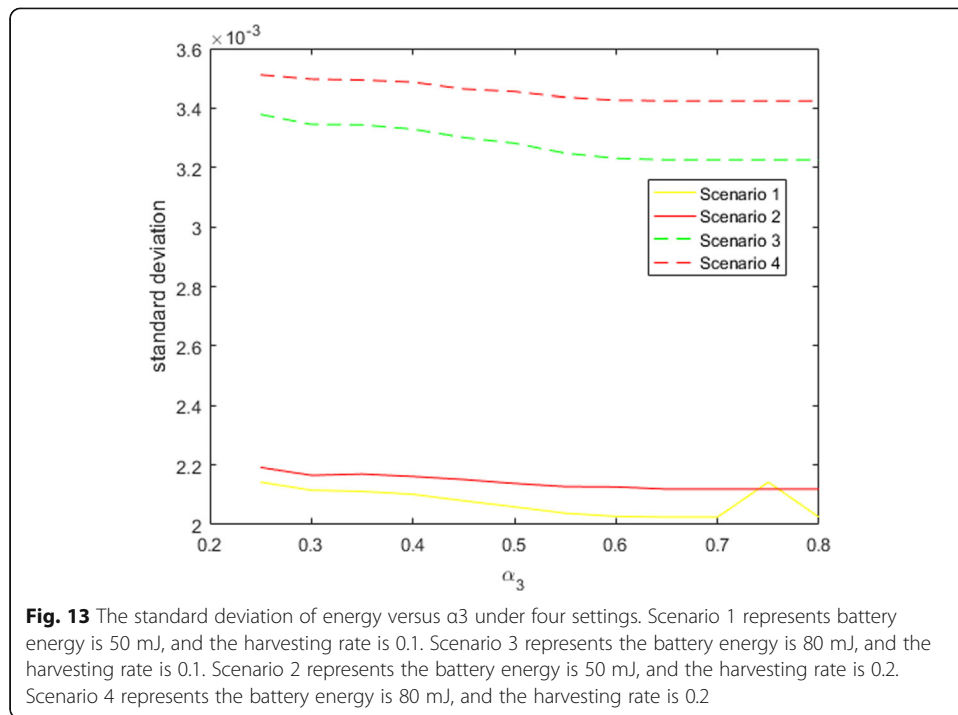
2 The parameters affecting the intercluster routing

The distance d_{hop} is the critical distance to control "hot zone" problem. Originally the nodes near the Sink node will exhaust their energy too quickly due to the heavy load that they help others to forward. If there is a control circle with radius d_{hop} round the Sink node, in which the nodes will directly communicate with the Sink node
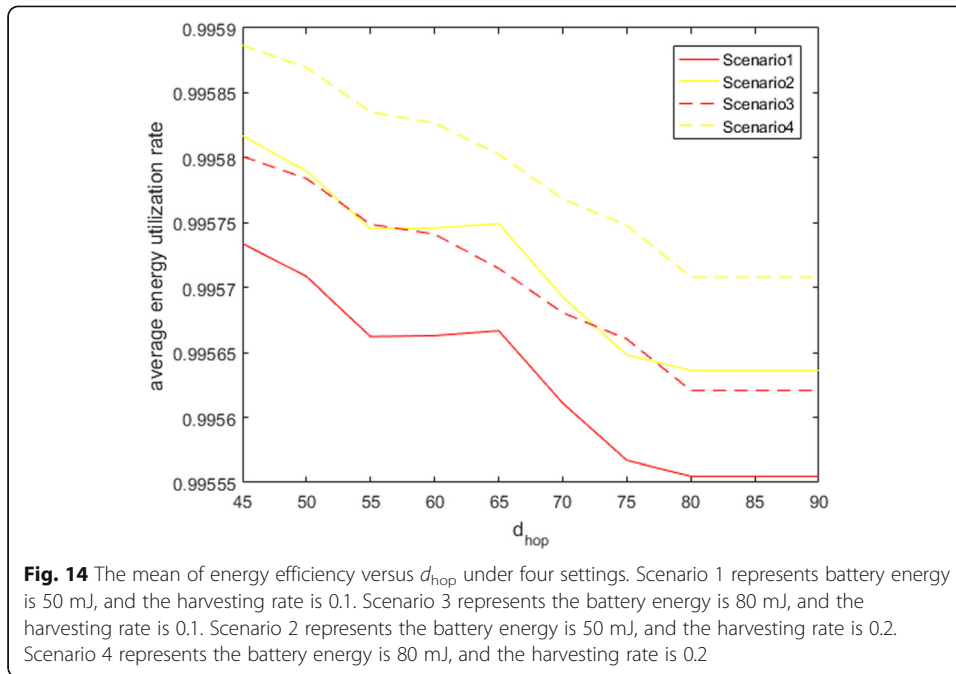

Fig. 12 The mean of energy efficiency versus α3 under four settings. Scenario 1 represents battery energy is 50 mJ, and the harvesting rate is 0.1. Scenario 3 represents the battery energy is 80 mJ, and the harvesting rate is 0.1. Scenario 2 represents the battery energy is 50 mJ, and the harvesting rate is 0.2. Scenario 4 represents the battery energy is 80 mJ, and the harvesting rate is 0.2

Fig. 13 The standard deviation of energy versus α3 under four settings. Scenario 1 represents battery energy is 50 mJ, and the harvesting rate is 0.1. Scenario 3 represents the battery energy is 80 mJ, and the harvesting rate is 0.1. Scenario 2 represents the battery energy is 50 mJ, and the harvesting rate is 0.2. Scenario 4 represents the battery energy is 80 mJ, and the harvesting rate is 0.2

without the help of relay nodes, the nodes in the inner circle round the Sink will be less burdened. According to Section 5.5, the larger the distance d_{hop} is, the more nodes in the control zone are, and the more balanced the energy will be used among the nodes in the control. But too large d_{hop} will cause too many nodes to directly communicate with the Sink node, which leads to too much energy consumption of the network. So there must be a proper value for d_{hop}. We take the network deployment as the above as an example. α_{head} is set as 0.25. α_3 is set as 0.6. Other parameters, such as α_1, β_1, l, are as same as the above settings. As shown in Fig. 13, the average energy efficiency and the standard deviation with different value of d_{hop} from 50 to 90, after 100 rounds are plotted . It can be observed from Figs. 14 and 15 that when d_{hop} is 45, the average energy efficiency is maximum and the standard deviation is minimum in each case; that is, when the value is 45, the energy can be best utilized, and each round of energy is relatively stable. The critical d_{hop} is determined as 45 m as indicated by the simulation results.

6 Results and discussion

The general simulation scenario for network node distribution is shown in Fig. 16. Ninety-nine nodes randomly and uniformly distributed in the 100 m × 100 m area. The Sink node is located at (100, 50) m which is assumed to have unlimited energy. Each node periodically generates packets, and packets are transmitted to sink nodes. The packet length is $l = 1$ bit. The commonly used constant value for energy consumption calculations is $E_{\mathrm{fuse}} = 5$ nJ/bit, $E_{elec} = 50$ nJ/bit, $\varepsilon_{amp} = 0.0013$ pJ/(bit · m^4), $\varepsilon_{fs} = 10$ pJ/(bit · m^2), $d_0 = 60$ m. For DCEM algorithm, the radius of clusters is 35 m. For EBCR algorithm, the other parameters of α_{head}, α_3, and d_{hop} could use the best studied values. Here, in order to show the general performance, we set the parameters as 0.2, 0.2, 87
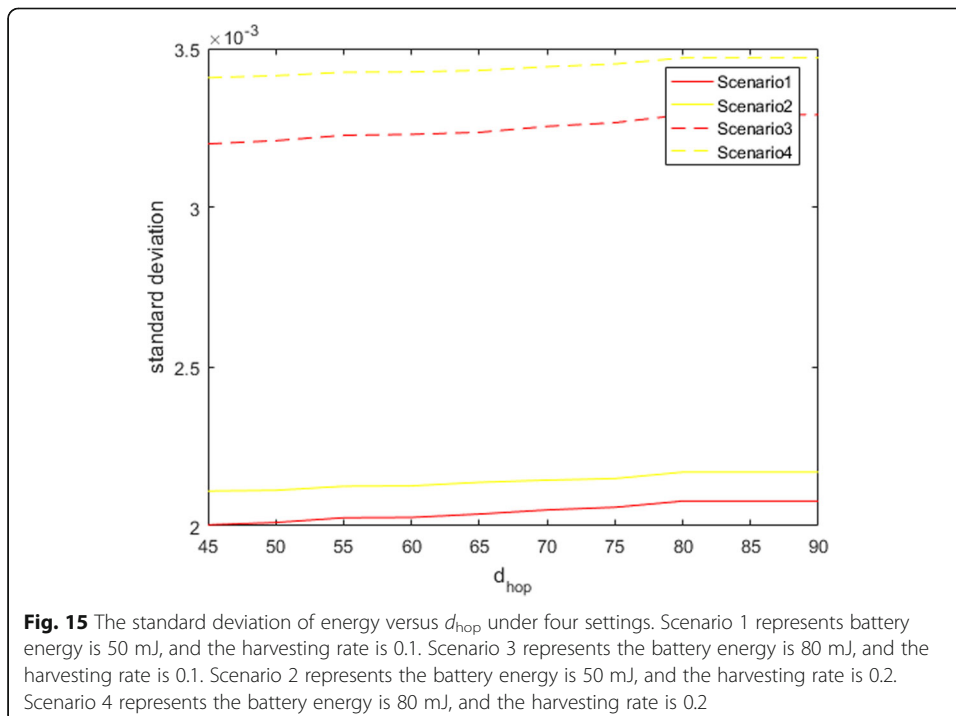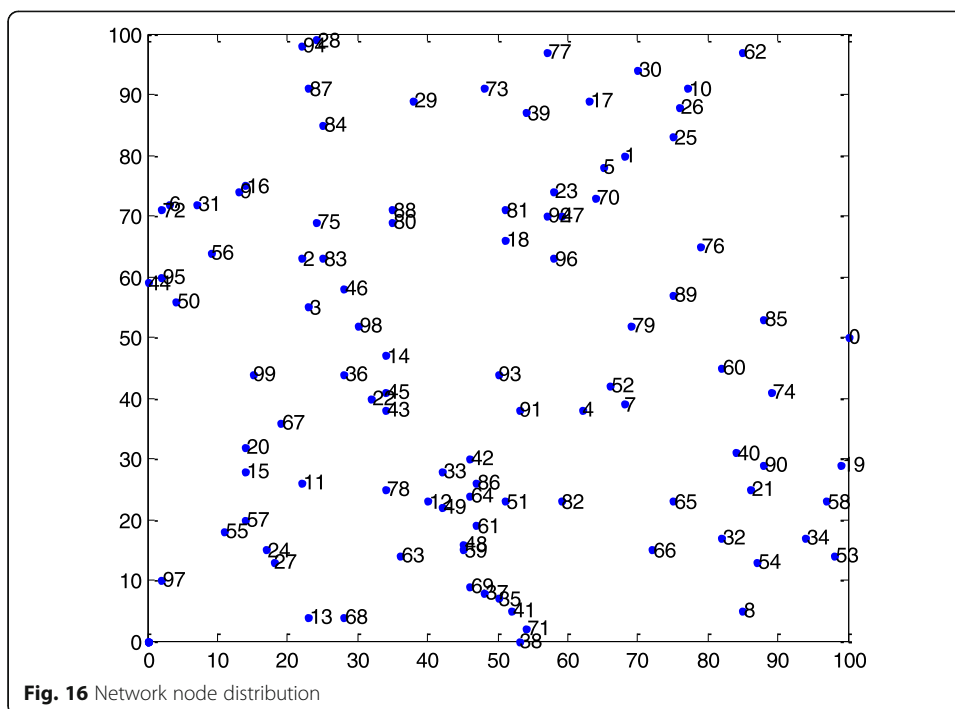
Fig. 14 The mean of energy efficiency versus d_{hop} under four settings. Scenario 1 represents battery energy is 50 mJ, and the harvesting rate is 0.1. Scenario 3 represents the battery energy is 80 mJ, and the harvesting rate is 0.1. Scenario 2 represents the battery energy is 50 mJ, and the harvesting rate is 0.2. Scenario 4 represents the battery energy is 80 mJ, and the harvesting rate is 0.2

m. The threshold for help in the hot zone is 3. The control parameters are $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$, $\lambda = 1$. For DCEM algorithm, its control parameters are $\alpha = \beta = 0.5$.

6.1 Scenario without energy harvesting

6.1.1 Simulation scenario 1

It is a scenario without energy harvesting. The hardware configuration of the 100 nodes is the same, i.e., the battery capacity is the same for all nodes $E_M = 0.12$ mJ. The initial

Fig. 15 The standard deviation of energy versus d_{hop} under four settings. Scenario 1 represents battery energy is 50 mJ, and the harvesting rate is 0.1. Scenario 3 represents the battery energy is 80 mJ, and the harvesting rate is 0.1. Scenario 2 represents the battery energy is 50 mJ, and the harvesting rate is 0.2. Scenario 4 represents the battery energy is 80 mJ, and the harvesting rate is 0.2

Fig. 16 Network node distribution

energy is distributed randomly in the range of [0.10, 0.12] mJ. For the WSN without energy harvesting, the biggest drawback is that the node energy is limited. If the node energy exhaustion is too fast, too many exhausted nodes will cause paralysis of the network and unreliable data transmission. Therefore, reducing the energy consumption as much as possible and extending the network life cycle are the primary goals. Therefore, the total energy consumption of the network and the number of surviving nodes are important indicators for evaluating the clustering algorithm.

a   Total network residual energy

The total network energy consumption of the EBCR and DCEM algorithms are shown in Fig. 17. The EBCR algorithm does not find an advantage node in the selection of cluster heads and to take over the nodes in the entire communication area in clustering phase, but makes good use of the nodes with larger energy as cluster heads and designs dynamic search radius according to their energy and distance from the Sink node. It also allows the division of the cross regions of the clusters and gives a reasonable division method. This effectively reduces the burden on the cluster head, helps to avoid excessive consumption of cluster heads in the hot zone, and alleviates the hot zone problem. In the intercluster routing: firstly, the candidate next hop nodes are selected according to the beneficial inequality relations among the cluster head nodes. During determining the next hop nodes, the cooperation order and coordination relationship between the cluster head nodes are considered, enabling more balanced consumption. As shown in Fig. 15, that the total energy consumption of nodes in the network is lower than that of the DCEM algorithm in each round of network operation of the EBCR algorithm, the network life cycle is significantly longer than the DCEM algorithm, which effectively extends the network life cycle.
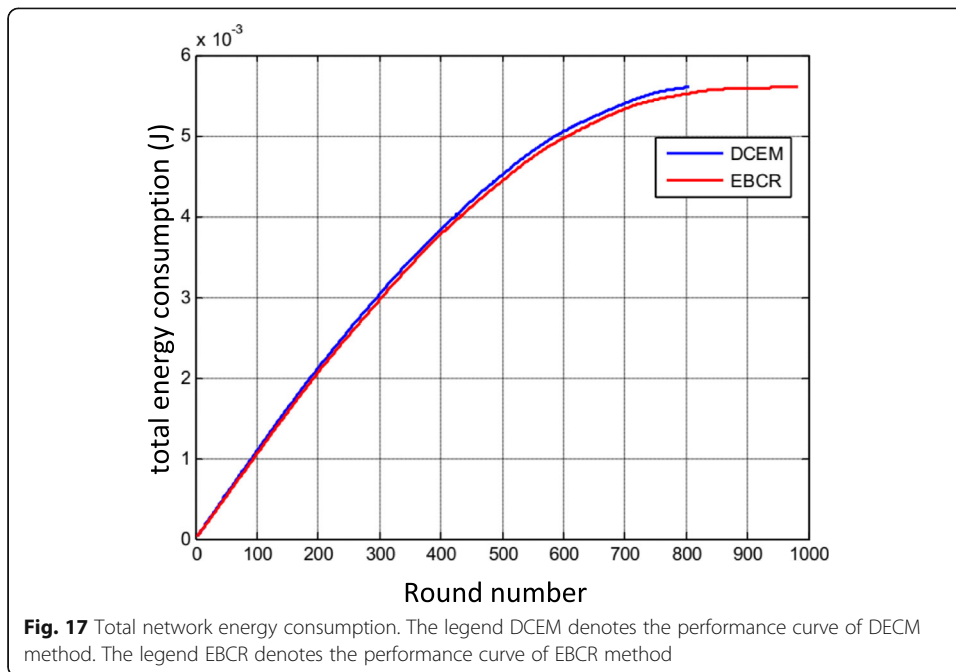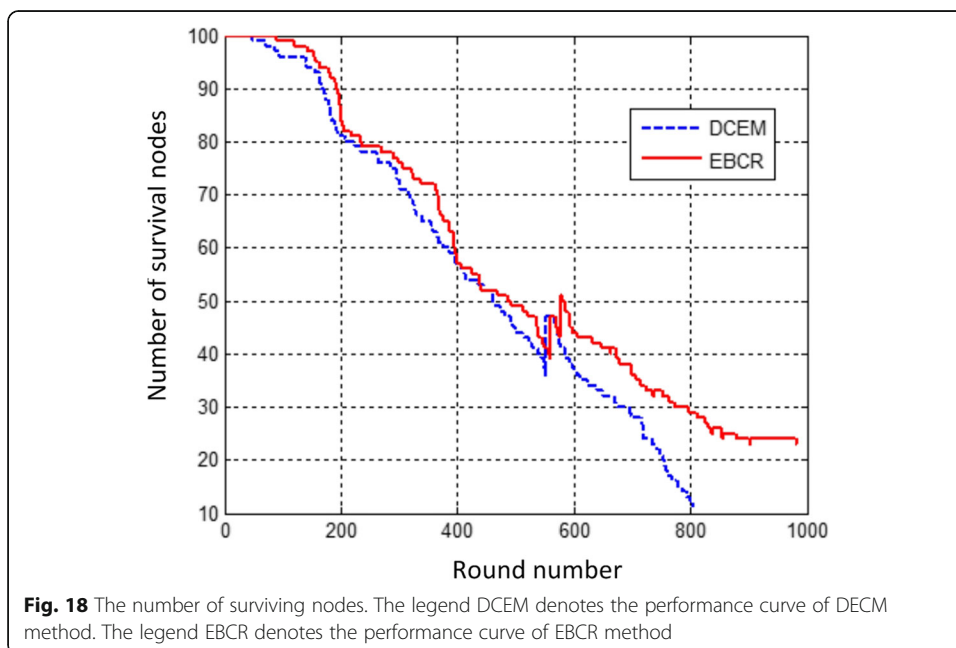
Fig. 17 Total network energy consumption. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

b   Number of surviving nodes

The numbers of surviving nodes of the EBCR algorithm and the DCEM algorithm are shown in Fig. 18. The comparison of the round numbers of the EBCR and the DCEM algorithms is shown in Table 1.

Performance elevation rate of EBCR compared to DCEM is:

Fig. 18 The number of surviving nodes. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Table 1 Comparison of the number of rounds

Algorithm	The round number of the first death node	The round number of the 50th death node	The total running round number
DCEM	48	466	805
EBCR	90	493	982
Promotion	87.5%	5.8%	22%

$$\text{promotion} = \frac{n_{\text{EBCR}} - n_{\text{DCEM}}}{n_{\text{DCEM}}} * 100\% \tag{28}$$

where n_{EBCR} is the number of dead rounds in the network using the EBCR algorithm and n_{DCEM} is the number of dead rounds in the network using the DCEM algorithm. It can be seen that the EBCR algorithm is optimized for cluster head selection, clustering, and intercluster routing, so that the network load are balanced among nodes and there are less fail nodes in the same round. The first death node failed in the 90th round and the 50th death node failed in the 493th round. The total number of running rounds is 982 rounds. The performance is significantly improved by 87.5%, 5.8%, and 22%, respectively in the corresponding aspects compared with the DCEM algorithm. The network life cycle is significantly longer.

6.2 Energy harvesting scenario with uniform nodes

For WSNs with energy harvesting, attention should be paid to energy efficiency as well as total consumption energy. The energy utilization efficiency is shown in (29).

$$\text{efficiency}(k) = \frac{E_{\text{consume}}(k) + E_{\text{left}}(k)}{E_{\text{harvest}}(k) + E_{\text{initial}}} \tag{29}$$
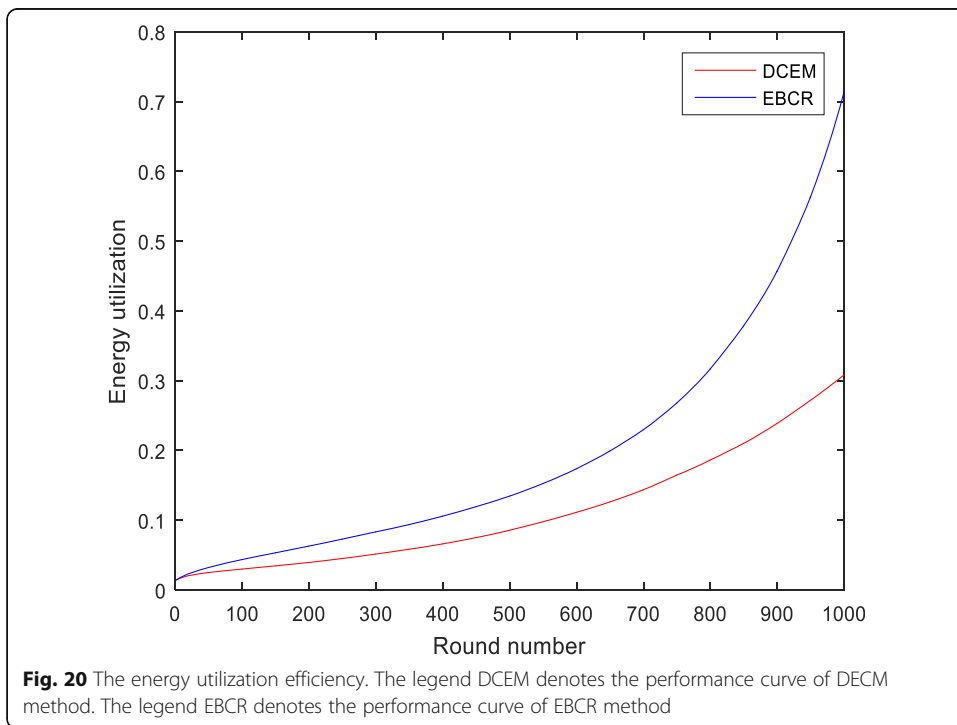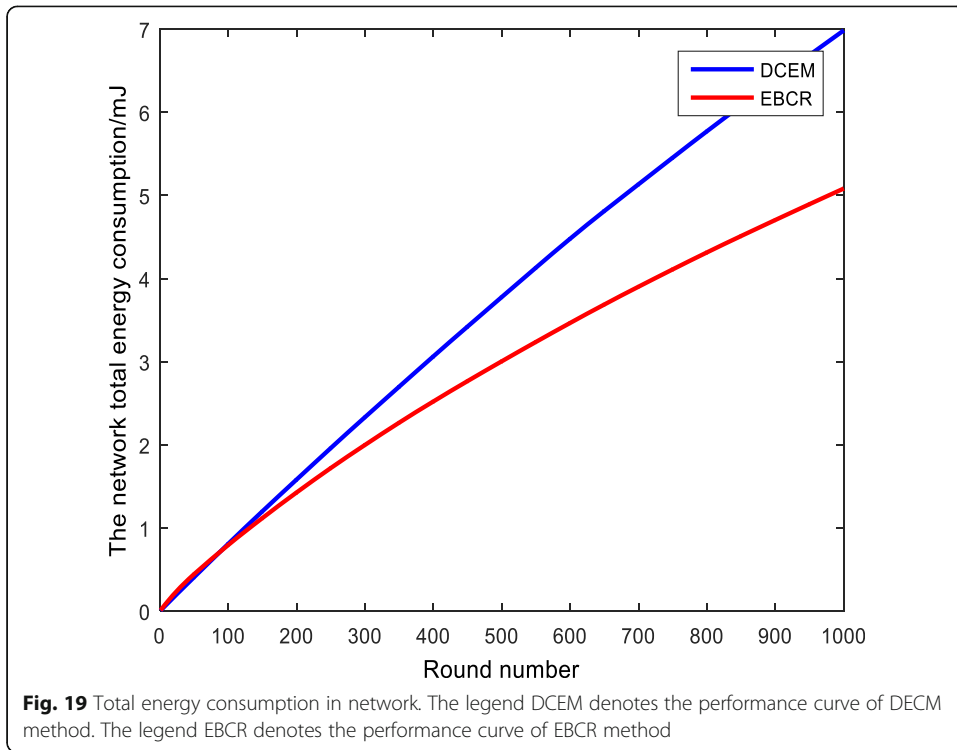
In (29), $E_{\text{consume}}(k)$ is the total data transmission consumption energy by time slot k, $E_{\text{left}}(k)$ is the total residual energy in the time network, and $E_{\text{harvest}}(k)$ is the total absorption energy in the network by time slot k. The energy harvesting rate is quite different among these scenarios, which indicates that the total energy available for use is not the same among these scenarios. Moreover, except energy consumption, there is overflow energy waste due to limited energy capacity. Because of these, the total energy consumption is not a useful measurement to method performance. The energy utilization efficiency is more useful to measure the method performance.

6.2.1 Simulation scenario 2

It is an energy harvesting scenario. The hardware configuration of the 100 nodes is the same. The battery capacity is the same for all nodes, E_M = 0.12 mJ. The initial energy is distributed randomly in the range of [0, 0.12] mJ. The harvested energy in each time slot is 10% of the initial energy of the node. After running for 1000 rounds, the total energy consumption in network and the energy utilization efficiency are shown in Figs. 19 and 20.

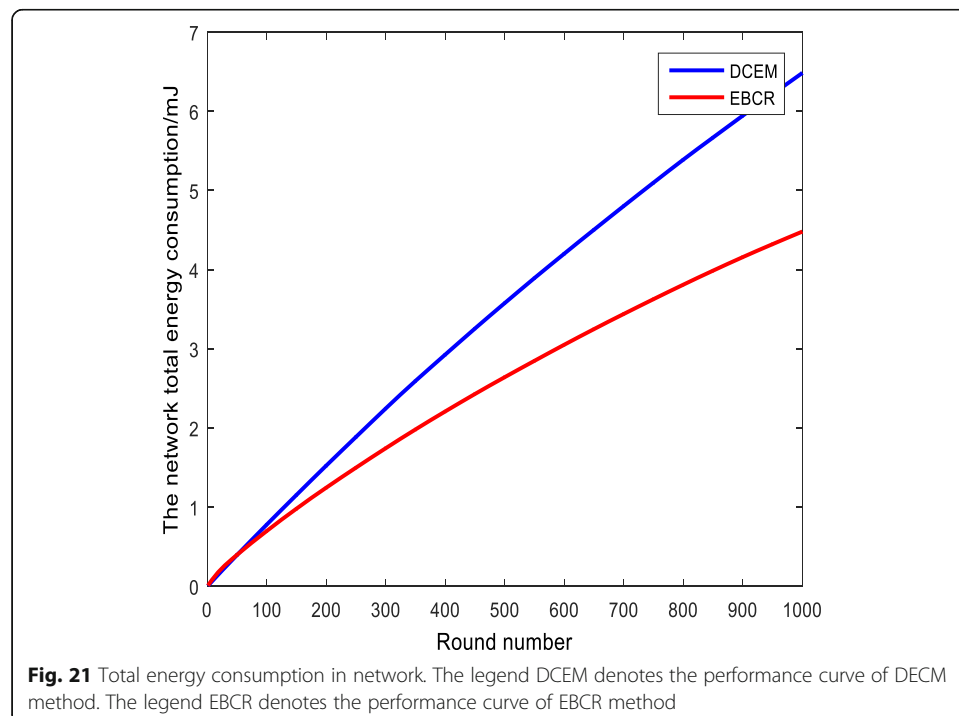### 6.2.2 Simulation scenario 3

The scenario is the same with simulation scenario 2 except that the harvested energy in each time slot for each node is randomly distributed. The harvested energy rate for all

Fig. 19 Total energy consumption in network. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Fig. 20 The energy utilization efficiency. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

nodes meets a normal distribution with mean of 10%, variance of 0.04. After running for 1000 rounds, the total energy consumption in network and the energy utilization efficiency are shown in Figs. 21 and 22.

6.3 Energy harvesting scenario with non-uniform nodes
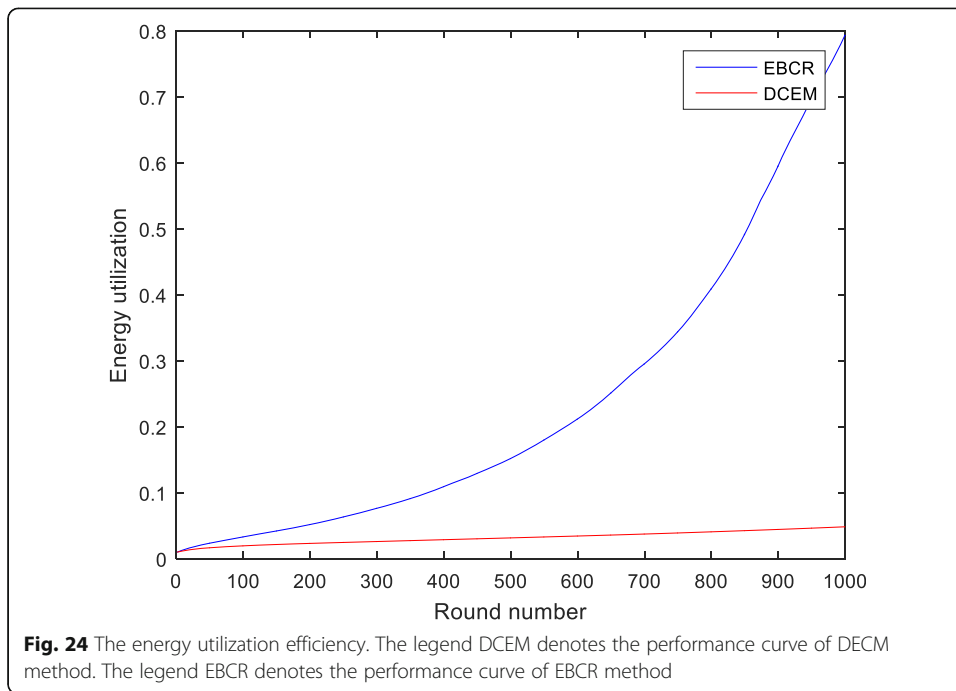
6.3.1 Simulation scenario 4

It is an energy harvesting scene. Assume that the hardware configuration of the 100 nodes is different. There are five special nodes with special large battery capacity E_M= 0.24 mJ. The energy harvested by the five special nodes in each slot is 20% of the initial energy of the node. The remaining 95 nodes are ordinary nodes, of which the battery capacity is 0.12 mJ. The initial energy of all nodes is random distributed in the range of [0, 0.12] mJ. The energy harvested by the 95 ordinary nodes in each slot is 10% of the initial energy of the node. After running for 1000 rounds, the total energy consumption in network and the energy utilization efficiency are shown in Figs. 23 and 24.
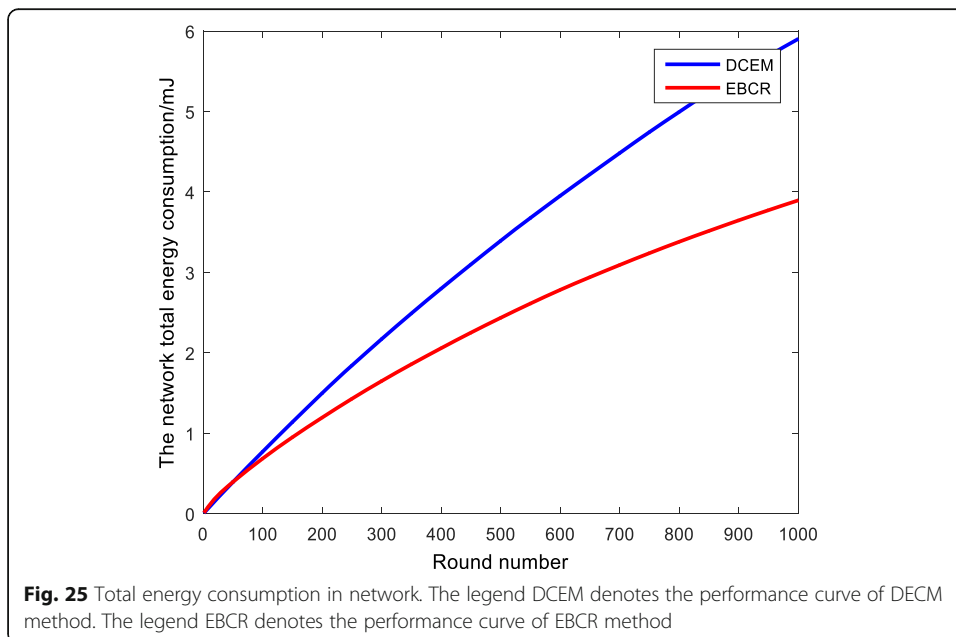
6.3.2 Simulation scenario 5

The scenario is the same with simulation scenario 4 except that the initial energy of the special five nodes is random distributed in the range of [0.06, 0.08] mJ. The initial energy of the ordinary 95 nodes is random distributed in the range of [0, 0.12] mJ. After running for 1000 rounds, the total energy consumption in network and the energy utilization efficiency are shown in Figs. 25 and 26. Figures 19, 21, 23 and 25 show the performance of DCEM and EBCR methods in scenario 2, 3, 4, and 5, respectively with the networks running 1000 rounds. It can also be seen that the total energy consumption of EBCR algorithm is less than that of DCEM algorithm after each run of the network. The slope of the total energy consumption curve of EBCR algorithm is smaller
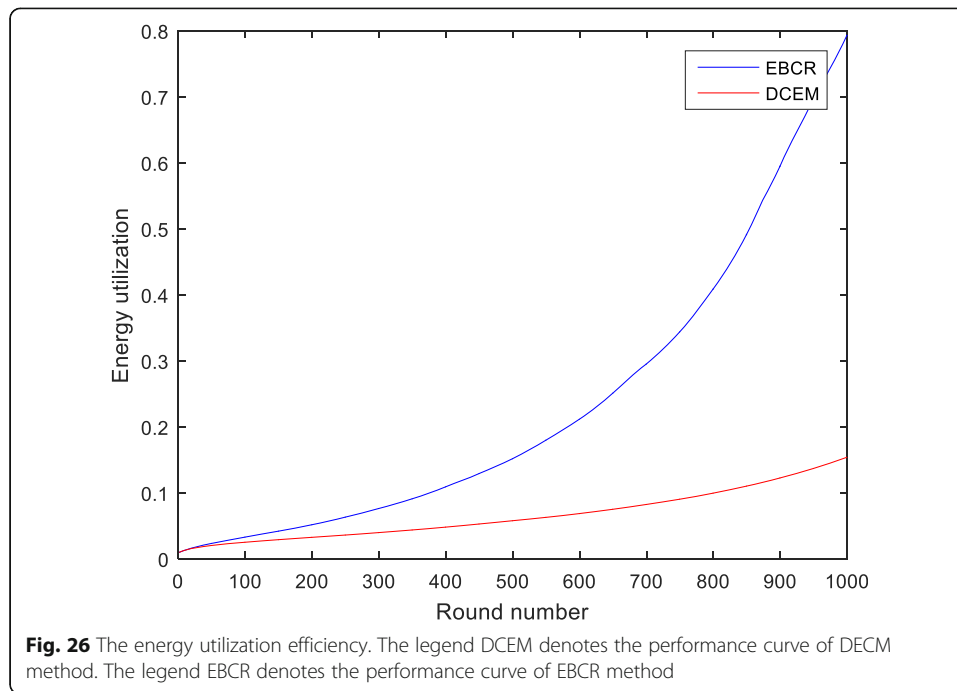

Fig. 21 Total energy consumption in network. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Fig. 22 The energy utilization efficiency. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Fig. 23 Total energy consumption in network. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Fig. 24 The energy utilization efficiency. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

than DCEM method in all scenarios and the energy efficiency of EBCR method is higher than DCEM method in all scenarios. This validates the better performance of EBCR method than DCEM method. In scenarios 2 and 3, the energy harvesting rate of nodes is more unbalanced in scenario 2, which makes the network condition more un-balanced than in scenario 3. Simulations show DCEM and EBCR methods both have better performance in scenario 3 than in scenario 2. The more balanced the network is, the better performance the methods are. Moreover, under scenario 2, the gap of energy utilization efficiency between the two methods is bigger. This shows that EBCR has

Fig. 25 Total energy consumption in network. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

Fig. 26 The energy utilization efficiency. The legend DCEM denotes the performance curve of DECM method. The legend EBCR denotes the performance curve of EBCR method

stronger adjustment ability of energy balanced utilization than DCEM, so that the energy consumption can be more efficient. Under the condition of more unbalanced energy harvesting, the energy equilibrium effect of EBCR is more obvious. In scenarios 4 and 5, the energy harvesting rate of nodes is more unbalanced in scenario 4, which makes the network condition more unbalanced than in scenario 5. Simulations show that DCEM method has better performance in scenario 5 than in scenario 4. The performance of EBCR is not obviously elevated as DCEM method in scenario 5 because the selection of cluster head is performed one by one by energy order and position relationship in DCEM method, while in EBCR the cluster heads are selected by energy amount of largest 10% and then larger 10% among the remaining unclustered nodes. The 10% generally contains all 5% special nodes. Moreover, in scenario 4, the energy of the special 5% nodes is generally larger than in scenario 5. The energy utilization not only relates to balanced property of energy but also relates to mean harvesting energy of the nodes. From scenario 2, scenario 3, scenario 4, and scenario 5, under various classic scenarios, BECR is always better than DCEM in energy utilization efficiency, and the total consumed energy in the network by EBCR is always smaller than DCEM in every round of network. This validates the superior performance in terms of balanced energy use and high energy efficiency of our proposed EBCR method.

7 Conclusions

In this paper, the method of distributed clustering routing is studied. A distributed clustering method combining greatly simplified cluster head election, dynamic cluster radius, and intersection region division mechanism is proposed, and an intercluster cooperation routing scheme is proposed. According to simulation results, EBCR performs better in balanced energy usage, leading to high network energy efficiency in energy

harvesting scenario and a longer network survival life in energy non-harvesting scenario. It also effectively reduces the cluster head failure probability, greatly improves the cluster update time, and reduces the maintenance cost of the clusters. On the whole, the method achieves a more balanced energy consumption performance than the distributed clustered DCEM and EEUC. It also requires less calculation than UCFIA and UDEB. The proposed EBCR method is generally better than the existing method in terms of energy balanced use and avoids complex computations. In the future, we will discuss the parameter selection of the method.

Abbreviations
EBCR: Energy balanced clustering routing; DCEM: Delay-constrained energy-efficient cluster-based multi-hop routing; WSN: Wireless sensor network; LEACH: Low-energy adaptive clustering hierarchy; TEEN: Threshold-sensitive energy-efficient sensor network protocol; HEED: Hybrid energy efficient distributed clustering algorithm; PSO-ECHS: Particle swarm optimization based energy-efficient cluster head selection; PSO: Particle swarm optimization; AMRP: Average minimum reachability power; ACO: Ant colony optimization; AMRP: Average minimum reachability power; EEUC: Energy-efficient uneven clustering; UCFIA: Unequal clustering algorithm for WSN based on fuzzy logic and improved ACO; UDEB: Uneven clustering dynamic routing based on energy efficiency and balanced

Authors' contributions
The five authors of the paper have extensively participated in all of the paper writing. YY and CW mainly worked on the idea and innovative research method. GJ wrote part of the paper. HX programmed and performed most of the simulations. LR revised this paper. All authors read and approved the final manuscript.

Author's information
Not applicable.

Availability of data and materials
Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Consent for publication
We agree to the publication of the paper.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Advanced Equipment Intelligent Perception and Control, Beijing International Cooperation Base for Science and Technology, Beijing Information Science and Technology University, Beijing, China. [2]School of Computer Science and Technology, China University of Mining Technology, Xuzhou 221116, Jiangsu, China. [3]Mine Digitization Engineering Research Center of the Ministry of Education, China University of Mining and Technology, Xuzhou 221116, Jiangsu, China. [4]School of Earth and Space Sciences, Peking University, Beijing 100871, China.

References
1. N. Liu, J. Pan, T. Nguyen, A bi-population quasi-affine transformation evolution algorithm for global optimization and its application to dynamic deployment in wireless sensor networks. J Wireless Com Network **2019**, 175 (2019)
2. Shanthi M, Ramadevi E. A cluster based routing protocol in wireless sensor network for energy consumption[J]. International Journal of Advanced Networking&Applications, 2017:1-10.
3. W. Guo, W. Zhu, Z. Yu, J. Wang, B. Guo, A survey of task allocation: contrastive perspectives from wireless sensor networks and mobile crowdsensing. IEEE Access **7**, 78406–78420 (2019)
4. S. Yadav, R.S. Yadav, A review on energy efficient protocols in wireless sensor networks[J]. Wirel. Netw **22**(1), 1–16 (2015)

5.   H. Cheng, Z. Xie, L. Wu, et al., Data prediction model in wireless sensor networks based on bidirectional LSTM. J Wireless Com Network **2019**, 203 (2019)

6.   Chi-Hua Chen, Feng-Jang Hwang, Hsu-Yang Kung, Travel time prediction system based on data clustering for waste collection vehicles, IEICE Trans. Inf. Syst., 2019, Volume E102.D, Issue 7, Pages 1374-1383.

7.   X. Zheng, W. Zheng, Y. Yang, et al., Clustering based interest prediction in social networks. Multimed. Tools Appl. **78**, 32755–32774 (2019)

8.   H. Cheng, Z. Xie, Y. Shi, N. Xiong, Multi-step data prediction in wireless sensor networks based on one-dimensional CNN and bidirectional LSTM. IEEE Access **7**, 117883–117896 (2019)

9.   G. Martinez, S. Li, C. Zhou, Wastage-aware routing in energy-harvesting wireless sensor networks[J]. Sensors Journal IEEE **14**(9), 2967–2974 (2014)

10. R. Lin, S. Wang, W. Guo, An overview of co-clustering via matrix factorization. IEEE Access **7**, 33481–33493 (2019)

11. A.S.M.Z. Kausar, A.W. Reza, M.U. Saleh, et al., Energizing wireless sensor networks by energy harvesting systems: Scopes, challenges and approaches[J]. Renew. Sust. Energ. Rev. **38**(5), 973–989 (2014)

12. T.T. Huynh, A.V. Dinh-Duc, C.H. Tran, Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks[J]. J. Commun. Netw. **18**(4), 580–588 (2016)

13. J. Kaur, V. Sahni, Survey on Hierarchical Cluster Routing Protocols of WSN[J]. Int. J. Comput. Appl. **130**(17), 18–22 (2015)

14. Manjeshwar A, Agrawal DP. TEEN: A Routing protocol for enhanced efficiency in wireless sensor networks[C]// Parallel and Distributed Processing Symposium. Proceedings, International. IEEE, 2002:400.

15. Heinzelman WB, Chandrakasan AP, Balakrishnan H. An application specific protocol architecture for wireless microsensor networks[C]// IEEE Transactions on Wireless Communication. 2002:660-670.

16. Krishnamoorthy S. Enhanced adaptive clustering mechanism for effective cluster formation in WSN[J]. Social Science Electronic Publishing, 2017.

17. L. Yang, Y.Z. Lu, Y.C. Zhong, et al., A hybrid, game theory based, and distributed clustering protocol for wireless sensor networks[J]. Wirel. Netw **22**(3), 1007–1021 (2016)

18. P.C.S. Rao, P.K. Jana, H. Banka, A particle swarm optimization based energy efficient cluster head selection algorithm for wireless sensor networks[J]. Wirel. Netw **23**(7), 2005–2020 (2017)

19. Lv Y, Miao Z, Zhang D, et al. A low energy uneven clustering topology control algorithm for wireless networks[C]. International Conference on Information Science and Control Engineering. IEEE, 2016:1203-1207.

20. Jiang C, Ren Y, Zhou Y, et al. Low-energy consumption uneven clustering routing protocol for wireless sensor networks[C]. International Conference on Intelligent Human-Machine Systems and Cybernetics. IEEE, 2016.

21. M. Song, Unequal clustering algorithm for WSN based on fuzzy logic and improved ACO[J]. The Journal of China Universities of Posts and Telecommunications **18**(6), 89–97 (2011)

22. Mao S, Zhao C, Zhou Z, et al. An improved fuzzy unequal clustering algorithm for wireless sensor network[C]// International ICST Conference on Communications and networking in China. IEEE, 2012.

23. T. Liu, T. Gu, N. Jin, et al., A mixed transmission strategy to achieve energy balancing in wireless sensor networks[J]. IEEE Trans. Wirel. Commun. **16**(4), 2111–2122 (2017)

24. Z. Xu, L. Chen, C. Chen, et al., Joint Clustering and routing design for reliable and efficient data collection in large-scale wireless sensor networks[J]. IEEE Internet Things J. **3**(4), 520–532 (2016)

25. D. Wei, Y. Jin, S. Vural, K. Moessner, R. Tafazolli, An energy-efficient clustering solution for wireless sensor networks. IEEE Trans. Wirel. Commun. **10**(11), 3973–3983 (2011)

26. S.A. Nikolidakis, D. Kandris, D.D. Vergados, C. Douligeris, Energy efficient routing in wireless sensor networks through balanced clustering. Algorithms **6**(1), 29–42 (2013)

27. O. Younis, S. Fahmy, HEED: A hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks[J]. IEEE Trans. Mob. Comput. **3**(4), 366–379 (2004)

28. Y. Wu, W. Liu, Routing protocol based on genetic algorithm for energy harvesting-wireless sensor networks [J]. Wireless Sensor Systems Iet **3**(2), 112–111 (2013)

Publisher's Note