

An energy-efficient overlapping clustering protocol in WSNs

Yuan Hu¹ · Yugang Niu¹ 

Published online: 29 December 2016
© Springer Science+Business Media New York 2016

Abstract The limited battery power supply system makes energy efficiency a major concern in WSNs. An effective method is to organize the sensors into clusters to avoid redundancy and long-distance data transmission in the network. In traditional clustering methods, the cluster heads not only serve as leaders to collect the coming data from their cluster members but also play the roles of relay nodes to transmit the aggregated data to the sink node simultaneously, such that CHs consume much more energy than ordinary nodes. From the perspective of energy balancing, it is better to select the different nodes as CHs and relay nodes. In this paper, an energy-efficient overlapping clustering protocol is proposed, which assigns the boundary nodes in the overlapping area to relay the aggregated data to the sink node. Thereby the relay nodes are uniformly distributed near the CHs. Comparisons with LEACH and SEECH protocols show that the proposed protocol achieves better performance in terms of lifetime and load-balancing.

Keywords Clustering · Overlapping region · Energy efficiency · Load balancing

1 Introduction

With the development of micro-electro-mechanical systems (MEMS), data processing and wireless communication technologies, WSNs become more and more popular in academic research and practical applications, such as health-care monitoring, enemy tracking, environmental monitoring, and industrial detection [1, 2].

In WSNs, a large number of sensor nodes are deployed in a large area. In most cases, these sensors are equipped with limited energy supply units that cannot be replaced or recharged. Therefore, energy efficiency is a major concern in WSNs [3, 4]. Reducing energy consumption from the node level and balancing energy consumption from the network level are the two aspects to improve the energy efficiency for WSNs [5]. Energy consumption in a sensor includes: transmitting/receiving data, processing, and idle listening to the media. Transmitting and idle listening contribute the majority of energy consumption [1, 3, 5]. Hence, reducing data transmission and sleep scheduling are two appropriate methods to prolong the network lifetime [6].

Clustering is a key routing technique to extend the lifetime of WSNs as it can reduce the data transmission in the network [7]. In a cluster-based network, most sensors work as cluster members (CMs) to perform the sensing tasks, while a few sensors work as cluster heads (CHs) to aggregate the coming data and forward it to the sink node directly or via relay nodes. Single-hop communication consumes huge amount of energy due to long distance transmitting, and causes energy unbalancing as sensors far away from the sink node consume more energy [7, 8]. Meanwhile, multi-hop method also causes energy unbalancing since sensors near the sink node burden more relay traffic [9]. Hence, it is of great significance to minimize the

✉ Yugang Niu
acniuyg@ecust.edu.cn
Yuan Hu
huyuan-ecust@qq.com

¹ Key Laboratory of Advanced Control and Optimization for Chemical Process of Ministry of Education, East China University of Science and Technology, Shanghai 200237, China

energy expenditure of the sensors near the sink node as well as minimizing the total energy expenditure of the whole network [10, 11].

The remainder of this paper is organized as follows. In Sect. 2, some existing clustering protocols are briefly introduced. In Sect. 3, the network model and overlapping clustering problem is formulated. Section 4 provides the details of EEOC protocol. The protocol analysis is given in Sect. 5. The simulation results on network lifetime and distribution of energy consumption is given in Sect. 6, which is followed by the conclusions in Sect. 7.

2 Related work

In past decades, many clustering protocols have been proposed to improve energy efficiency and balance the energy consumption of the network. As CHs consume more energy than ordinary nodes, a fundamental way to balance the energy depletion is CH rotation during the data gathering rounds, which not only can balance the intra-cluster energy consumption, but also can balance energy consumption among CHs in the inter-cluster communication [12]. For example, in LEACH [13], the CH roles are periodically rotated after a predefined number of data gathering rounds. To avoid nodes serving as CHs continuously while avert high frequency of CH updating, DEECIC [14] introduces an energy threshold E_{th} for each CH in the current round to decide whether to continue as a CH in the coming round. The CH updating process is operated only when the residual energy of current CHs falls below E_{th} .

Considering the distributions of node's energy consumption, some existing clustering protocols introduce energy heterogeneity to balance the energy consumption and prolong the network lifetime. In heterogeneous WSNs, some advanced nodes with more initial energy are deployed into the network. During the CH selection phase, the nodes with more initial energy have higher chances to serve as CHs [15–18]. Some clustering protocols assign relay nodes to serve as inter-cluster routers without sensing the phenomenon in the monitoring area [19–22].

Besides, some researchers design unequal clustering protocols by decreasing the cluster radius or number of CMs with heavier relay traffic near the sink node [23–27]. As a result, CHs near the sink node burden lighter local traffic and remain more energy to forward the relay traffic. The distances from the sink node [23–25, 27] or hop counts [26] are used to form clusters. In a non-uniform distributed network, the node density impacts intra-cluster energy consumption which should be taken into consideration to limit the cluster size [23].

However, CH rotations will result in extra energy consumption and make the protocol more complex [14]. Meanwhile, both of the unequal clustering and CH rotation methods can only balance energy consumption from the network level. A node selected as a CH will burden much more traffic than other ordinary nodes such that it will die early and the network lifetime is shortened.

In view that CHs take more roles in the data gathering phase, a reasonable contemplation for balancing energy consumption of the network is to assign the data forwarding task to other ordinary nodes, avoiding sensors serving as CHs and relay nodes simultaneously. To authors' best knowledge, most existing works explore relay node placement issue in cluster-based networks without considering cluster formation. Although the SEECH clustering protocol [28] investigates how to select relay nodes during the cluster set-up phase, it selects CHs and relay nodes separately. In SEECH, the nodes with high node degree and residual energy are prioritized as CHs, and other ordinary nodes located closer to the sink node and remaining more energy have larger chances of being relay nodes. Each CH selects the nearest relay node as the next step and forward data to the sink node via the next relay node. Based on the idea of allocating aggregation task and forwarding task separately, two aspects can be considered to further improve the energy efficiency of the network:

- To keep a good balance of energy consumption among CHs, the relay nodes should be distributed uniformly in the area between CHs and the sink node, and the distance between relay nodes and CHs and the distance between relay nodes and sink node should be traded to minimize the relay energy consumption.
- From the perspective of balancing energy consumption in different regions of the monitoring, the cluster radius should be adjusted according to the node density.

Based on the above discussion, an energy-efficient overlapping clustering protocol called EEOC is proposed in this paper to balance energy consumption among nodes in different region of the monitoring area. Overlapping clustering means that a sensor node may belong to more than one cluster simultaneously. Overlapping clustering can be used in many applications, such as providing robustness and connectivity in the inter-cluster routing phase [29], guaranteeing global time synchronization [30], CH failure recovery [31] and node localization [32]. To the best of our knowledge, only one clustering method produces overlapping regions in the clusters formation phase called KOCA [33]. KOCA generates connected multi-hop overlapping clusters with the desired overlapping degree to cover the entire network. However, KOCA is a static clustering method that there are no cluster heads rotations in the working lifetime. As a result, load unbalancing is

inevitable in KOCA. Different from KOCA, the overlapping regions produced in our protocol aims at choosing proper relay nodes near the CHs to balance the energy consumption of the whole network. The main feature of the proposed protocol includes: (1) generating overlapped clusters with designed overlapping area; (2) and selecting proper high residual energy nodes in the overlapping areas to serve as relay nodes. Different from many existing clustering protocols, the proposed protocol in this work selects different nodes to undertake the roles of CHs and relay nodes, avoiding a node to serve as the CH and relay node simultaneously. Besides, the CHs and the relay nodes are selected jointly to ensure that the relay nodes are distributed uniformly near CHs to reduce energy consumption among CHs. An overlapping cluster formation mechanism is implemented such that the candidate relay nodes for a CH can be determined in its overlapping area. Moreover, the node density is also taken into consideration during the cluster-setup phase.

3 Network model and problem formulation

Consider a homogenous wireless sensor network with the sensors randomly distributed over a $M \times M$ square field. Some assumptions are made as follows:

- The sink node is located outside of the monitoring area and all nodes are stationary after deployment.
- Each sensor node is equipped with identical initial energy E_0 , and the sink node has no energy constraint.
- Sensor nodes send the gathered data to the sink node periodically.
- Sensor nodes are location-unaware. Each sensor is assigned with an unique ID.
- The links are symmetric. A pair of node can communicate with each other using the same power.
- Power control is allowed for sensors to adjust the transmission power according to the Euclidean distance.

3.1 Energy model

The energy consumption model as in [13] is utilized in this work. According to the distance between the receiver and the transmitter, the free space and the multi-path channel models are employed. The energy dissipation for sending l bits over a distance d is

$$E_T(l, d) = \begin{cases} l \times E_{elec} + l \times \varepsilon_{fs} \times d^2 & \text{if } d \leq d_0 \\ l \times E_{elec} + l \times \varepsilon_{mp} \times d^4 & \text{if } d > d_0 \end{cases} \quad (1)$$

where E_{elec} is the energy dissipation per bit per unit to run

the circuitry of nodes, ε_{fs} and ε_{mp} are the energy required by amplifier in the energy consumption model, $d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$ is the threshold distance for the model. The energy consumption to receive l bits message is

$$E_R = l \times E_{elec} \quad (2)$$

The energy consumption to aggregate l bits message is

$$E_{AG} = l \times E_{DA} \quad (3)$$

where E_{DA} is the energy dissipation to aggregate one bit message.

3.2 Problem formulation

A sensor network can be modeled as a undirected graph $G = (V, E)$, where the vertices V represent nodes and the edges E represent the connectivity between each pair of nodes in the network. Suppose that the number of vertices in the graph $(|V|)$ is N .

Some notations are defined as follows:

- **Closed set of cluster members of CH_i (S_{CH_i}):** is the set which contains the CMs in cluster CH_i and its CH.
- **inter-overlapping degree between CH_i and CH_j (O_{ij}):** is the number of common nodes in the overlapping area between them, which is defined as

$$O_{ij} = |S_{CH_i} \cap S_{CH_j}| \quad (4)$$

When $O_{ij} = 0$, it means that Cluster CH_i and Cluster CH_j are completely separated from each other.

- **intra-overlapping degree of cluster CH_i (OD_i):** is the total number of clusters which share overlapping area with CH_i .

$$OD_i = \sum_{k \in K_{CH}} L_{ik} \quad (5)$$

where K_{CH} is the set of CHs in the network, and L_{ik} is defined as

$$L_{ik} = \begin{cases} 1 & \text{if } O_{ik} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

- **direction fitness between cluster CH_i and CH_j (DF_{ij}):** is the ratio of the distance from CH_i to the sink node to the sum of the distance from CH_i to CH_j and the distance from CH_j to the sink node.

$$DF_{ij} = \frac{d_{CH_i}}{d_{ij} + d_{CH_j}} \quad (7)$$

where d_{ij} is the distance between two CHs of CH_i and CH_j , and d_{CH_i} (d_{CH_j}) represent the distance from the CH of Cluster CH_i (CH_j) to the sink node. As the sum

of any two sides of a triangle is greater than the third side, $0 < DF_{ij} < 1$.

- **node degree**: is the number of neighbor nodes for a node.
- **boundary nodes**: are the nodes in the overlapping area between two clusters.

The objective of this work is to find the set of CHs K_{CH} to form overlapping clusters satisfying:

- the intra-overlapping degree of any cluster CH_i satisfies $OD_i > 0$. It ensures that each CH can find at least one candidate relay nodes in its communication range.
- the inter-overlapping degree between cluster CH_i and CH_j satisfies $O_{ij} \in (\tau_1, \tau_2)$, where τ_1 is a threshold to guarantee the redundancy of relay nodes, τ_2 is a threshold to control the overlapping degree between cluster CH_i and CH_j .

After the clusters are organized, the relay nodes are selected according to the following rules:

- the residual energy $E_{re} > E_{th}$, where E_{th} is a threshold to evaluate whether the node can continue to be a relay node.
- select the nodes in the overlapping area with high inter-overlapping degree and high direction fitness DF_{ij} to serve as relay nodes. It is beneficial to reduce the total energy consumption for sending the aggregated data to the sink node.
- the intra-overlapping degree $OD_i < \varphi$ such that the relay nodes will not be distributed too concentrated.

As a result, the energy consumption is balanced and the network lifetime is prolonged.

4 The EEOC protocol

The EEOC protocol in this work is operated after distributing sensors in the monitoring area, which is divided into rounds and each round contains cluster-setup phase and steady-state phase. Network topology is established in the cluster-setup phase, which contains four steps: information collection, cluster selection, cluster formation and relay nodes selection. In the steady-state phase, network data is collected and sent to the sink node periodically. To reduce energy consumption of the cluster-setup phase, the data transmission phase is much longer than the cluster-setup phase. Different from traditional clustering protocols, each CH selects one of its boundary nodes with higher residual energy and proper direction as relay node. The general scheme of EEOC is shown in Fig. 1.

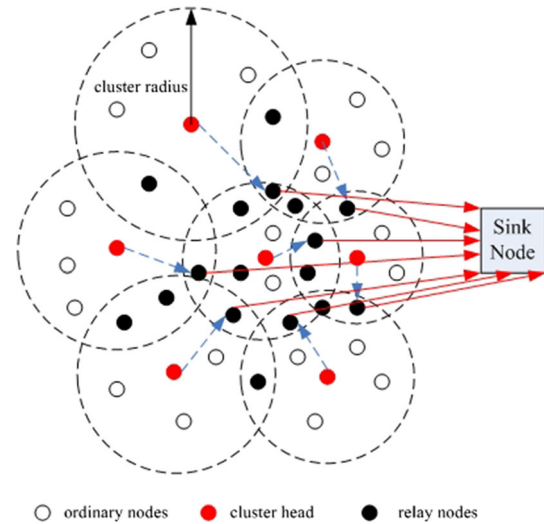


Fig. 1 General scheme of EEOC protocol

4.1 Cluster radius

The lower bound of the cluster radius in EEOC protocol is set to ensure connectivity of the nodes with a high probability and reduce the interference of cluster from one another. According to the result in [10], the nodes are connected with a probability of at least $1 - \epsilon$ if the communication radius satisfies

$$R_s \geq \sqrt{\frac{M^2}{N\pi} \log\left(\frac{N}{\epsilon}\right)} \tag{8}$$

The minimum cluster radius is set as

$$R_{c_min} = \frac{1}{2}R_{s_min} = \frac{1}{2}\sqrt{\frac{M^2}{N\pi} \log\left(\frac{N}{\epsilon}\right)} \tag{9}$$

In a cluster-based network topology, the connectivity of network can be guaranteed if the communicate radius and the cluster radius satisfies $R_s \geq 2R_c$. We set $R_{c_min} = \frac{1}{2}R_{s_min}$ such that all nodes can be connected with a high probability in all cases of sub-network division.

In order to save energy, we set the maximum distance between any pair of nodes in a cluster to be shorter than the threshold d_0 :

$$R_{c_max} = \frac{d_0}{2} \tag{10}$$

that is, the energy depletion of intra-cluster is based on the free-space energy model, which is more energy-efficient than the long-distance transmission.

4.2 Information collection of the network

At the beginning of data gathering rounds, the sink node broadcasts a ‘Hello message’ to the whole network. Each sensor hears the message and estimates the distance from the sink node based on the RSSI. Afterwards, all sensors will broadcast the ‘hand-shaking message’ within its communication radius R_s , which contains the information of node ID, residual energy and distance to the sink node. Therefore, each node establishes a neighboring table based on the ‘hand-shaking message’. Each node will calculate the node degree and the average residual energy based on its neighboring table. To obtain reliable infrastructure of the network, the ‘hand-shaking message’ should be broadcast at the beginning of each round.

4.3 Cluster head selection

In this step, CHs are selected in a distributed fashion. Generally, nodes with more residual energy and larger node degree are more appropriate to be CHs. To balance energy consumption, CHs should be distributed uniformly in the network. Previous clustering protocols for WSNs with non-uniform distribution of nodes arrange more CHs in dense area to form smaller clusters. However, forming clusters too small will leading to high inter-cluster traffic in the dense area.

In this protocol, the nodes with higher node degree and residual energy have higher chances to be CHs. Each node i will calculate the parameter p_i to compete for serving as a CH:

$$p_i = \begin{cases} \frac{E_{res_i} \cdot degree_i}{E_{aver_i} \cdot deg_{aver_i}} & \text{if } E_{res_i} > \varepsilon \cdot E_{aver_i} \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where deg_{aver_i} and E_{aver_i} are the average node degree and the average residual energy for the neighbors of node i , respectively. ε ($0 < \varepsilon < 1$) is a constant to avoid nodes with little residual energy but high node degree serving as CHs. The value of ε is set artificially and determined by the energy consumption rate of nodes in the network. Higher the energy consumption rate is, higher value ε should be set.

At the beginning of the proposed protocol, each node remains most initial energy so that E_{res_i} varies in a small range and the node degree is different due to the random distribution. According to Eq. (11), the node degree metric will play a major role in prioritizing nodes for being CHs. After some nodes are dead during the data gathering

rounds, E_{res_i} and deg_i will be completely different from each other due to the unbalancing energy consumption between CHs and CMs and the death of some low energy nodes. As a result, both of them make significant contribution for CH competing.

As mentioned above, CHs should be distributed uniformly in the network. A competing broadcast delay is set associated with a competing radius for each node to ensure that there is only one CH in the competing area and all nodes are connected in the cluster topology. A node i will wait for the expiration of its delay time T_{wait}^i before deciding whether it should declare as a CH:

$$T_{wait}^i = e^{\frac{1}{p_i + \alpha}} \quad (12)$$

where α is a constant to ensure that T_{wait}^i is a finite value if $p_i = 0$.

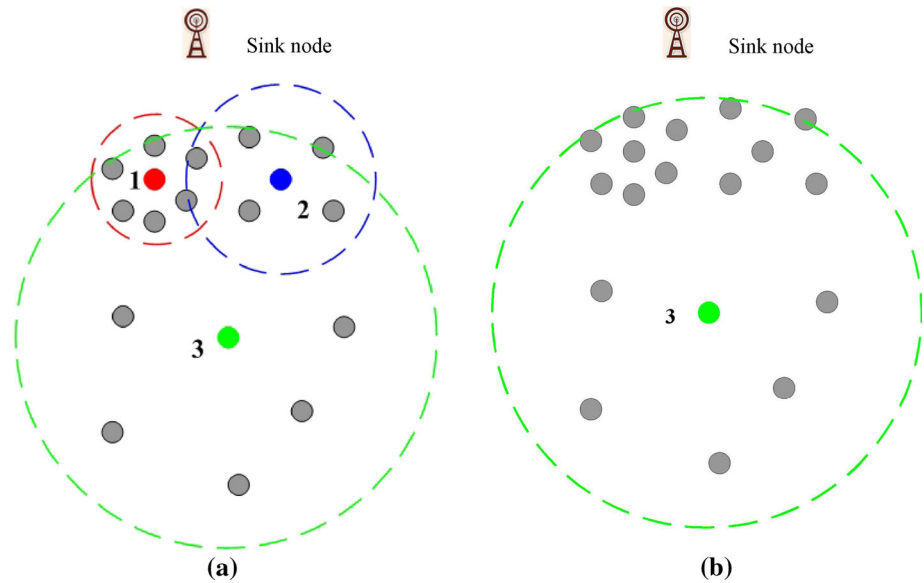
A node with more residual energy and higher node degree has shorter waiting time so that it has a higher chance to be CH. If a node does not receive any ‘CH_Competing’ message during the waiting time, it will broadcast the ‘CH_Competing’ message in its competing radius r_c^i to declare itself as a CH:

$$r_c^i = r_0 \cdot \left(1 - \frac{\varphi(i) - deg_{aver_i}}{\varphi_{max}(i) - \varphi_{min}(i)} \right) \quad (13)$$

where r_0 is a constant depending on the characteristic of sensors (initial energy, communication ratio) in the network, $\varphi(i) = \frac{degree_i}{deg_{aver_i}}$, $\varphi_{min}(i)$ and $\varphi_{max}(i)$ are the minimum and maximum node degree for the neighbors of node i .

All nodes in the competing radius hearing the message will give up the CH competing operation. The competing radius is related to the distance to the sink node and the node density. According to expression (11)–(13), the CH selection operation is started in the dense area with a smaller competing radius, which can avoid unbalance CH distribution in the network. For instance, as shown in Fig. 2(a), the red node in dense area near the sink node has the top priority to broadcast its ‘CH_Competing’ message within the red radius. And then, the blue node broadcasts its ‘CH_Competing’ message within the blue radius. The green nodes far away from the sink node broadcasts its ‘CH_Competing’ message in the green radius followed by the blue one. Finally, CHs will be distributed uniformly in the network. However, if the green node broadcasts its ‘CH_Competing’ message firstly, all nodes in its competing area will give up the competing for being CHs and there will be only one node selected in the network, as shown in Fig. 2(b). The cluster head selection algorithm is given in Algorithm 1.

Fig. 2 Cluster head competing started from different nodes. **a** started from node 1, **b** started from node 3



Algorithm 1 Cluster head Selection

```

procedure CH SELECTION( $p(i), T_{wait}^i, r_c^i$ )
  set the CH competing timer as  $T_{max}$ ;
  for  $i = 1 : 1 : n$  do
    calculate  $p(i), T_{wait}^i, r_{compet}^i$ ;
    set the counter value as  $T_{wait}^i$ ;
    while  $T < T_{max}$  do
      if  $T = T_{wait}^i$  then
        if receive ‘CH_Competing’ messages then
          give up the cluster head competing;
        else
          being a CH;
          broadcast a ‘competing message’ in  $r_c^i$ ;
        end if
      end if
    end while
  end for
end procedure

```

4.4 Cluster formation

At the beginning, each CH broadcasts a ‘CH_Announce’ message in its communication radius R_s . The ‘CH_Announce’ message contains the information of node ID, distance to the sink node and the residual energy. Like

the overlapping clustering protocol in [33], each ordinary nodes maintain a CH-table containing three elements $[CHID, d_{toCH}, d_{toBS}]$, where CHID is the sender node ID, d_{toCH} is the distance from the sender node, d_{toBS} is the distance from the source node to the sink node. And then each ordinary node will send a ‘master-head’ message to

the nearest CH to form un-overlapping clusters, which contains the information of node ID and *CH-table*. The CHs hearing the ‘master-head’ message will hold a *CM-table* with $[d_{ioCH}, CMID, NEI_CHID, d_{CHtoCH}, d_{NEItoBS}]$, where *CMID* is the ID of sender node, *NEI_CHID* is the neighbor CH ID connected through the sender node, d_{CHtoCH} is the distance to the neighbor CH, $d_{NEItoBS}$ is the distance from the neighbor CH to the sink node. Each ordinary node will compare its residual energy with the residual energy of other CHs in its *CH-table*. If the ordinary node remains more energy than any of the neighbor CHs, it will declare itself as a ‘Hide-high-energy node’ (HHE node) by sending a ‘high_energy’ message to the CHs with lower energy and the CH in its master cluster. The ‘high_energy’ message contains the CH IDs (the CHs with lower energy and the CH in its master cluster) and the residual energy of its own.

The overlapping cluster formation starts from the HHN nodes. A pair of CHs (e.g. CH_i and CH_j) containing the same HHE nodes will calculate the direction fitness DF_{ij} . As shown in Fig. 3, there are three CHs (A, B, C) in the network, we consider A as a better relay station compared with C because the sum distance from A to B and B to S is

shorter compared with the distance through $A - C - S$ ($DF_{AB} = \frac{d_{CH_A}}{d_{AB}+d_{CH_B}} > \frac{d_{CH_A}}{d_{AC}+d_{CH_C}} = DF_{AC}$). The parameter *DF* is a metric to evaluate which clusters are proper to overlap. A cluster CH_i will select the cluster with highest *directionfitness* (CH_h) to form overlapping clusters by broadcasting a ‘Overlap’ message.

If the HHE nodes between them $n_{HHE} > \tau_1$, the inter-overlapping degree $O_{ih} > \tau_1$, the redundancy condition of relay nodes is satisfied. Otherwise, CH_i will select another cluster with the second highest *directionfitness* to form overlapping clusters and inform the *HHE* nodes between them to form overlapping area. If no HHN nodes exists between the pair of CHs, τ_1 number of boundary nodes remaining most energy between CH_i and CH_j are selected as boundary nodes by broadcasting a ‘Overlap’ message. To avoid too many relay nodes concentrated in one area, the intra-overlapping degree of each cluster is limited to φ and the inter-overlapping degree between any pair of CHs is limited to τ_2 . If there are more than φ neighbor CHs connected by a CH, it will send a ‘refuse_CH’ message to the redundant CHs with lower *directionfitness*. If there are more than τ_2 boundary nodes between any pair of CHs, the CH with larger ID will send a ‘refuse_CM’ message to the redundant boundary nodes with lower residual energy. The sketch map of cluster formation is shown in Fig. 4, and the cluster formation algorithm is given in Algorithm 2.

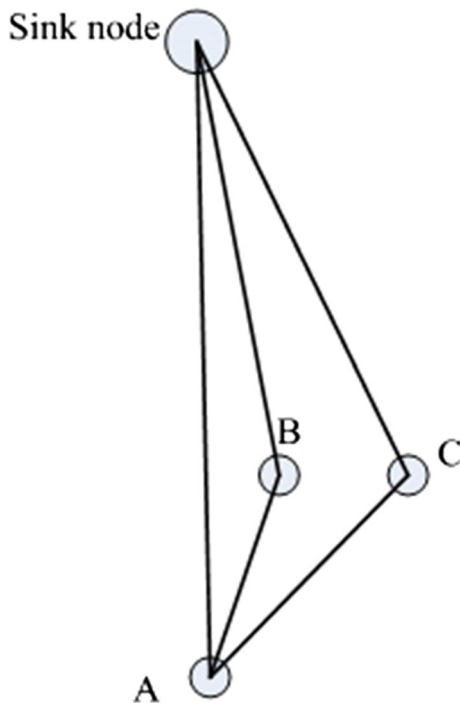


Fig. 3 Direction fitness

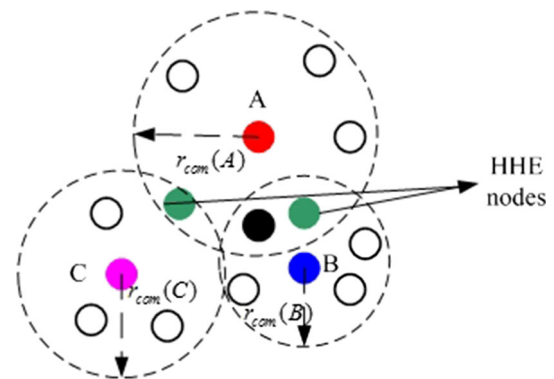


Fig. 4 Cluster formation

Algorithm 2 Cluster Formation

```

for  $i=1:1:n$  do
  if node_type( $i$ )='C' then ▷ for CHs
    establish a 'CM-table';
    calculate  $OD_i$ ;
    for  $j=1:1:n$  do
      if node_type( $j$ )='C' &  $j \in S_{CH}^{nei}(i)$  then
        calculate  $O_{ij}, DF_{ij}$ ;
        if  $O_{ij} < \tau_1$  then
          send 'overlap' messages to BN( $i,j$ );
        end if
        if  $OD_i = 0 \& DF_{ij} > \theta$  then
          send 'overlap' messages to BN( $i,j$ );
        end if
        if  $O_{ij} > \tau_2$  then
          send 'refuse.CH' messages to  $CH_j$ ;
        end if
        if  $OD_i > \varphi \& DF_{ij} > \theta$  then
          send 'refuse.CM' messages to BN( $i,j$ );
        end if
      end if
    end for
  else ▷ for ordinary nodes
    establish a 'CH-table';
    send 'master-head' message to nearest CH;
    if node  $i$  is a HHE node then
      send 'HHE messages' to corresponding CHs;
    end if
    if receive 'overlap' message then
      connect to the source CH;
    end if
  end if
end for

```

4.5 Relay nodes selection

Each CH selects a boundary node in its own overlapping area as the relay node. First, a CH choose the overlapping area with highest DF as candidate relay area according to the direction fitness metric. All nodes in the candidate relay area form the relay nodes set. And then the CH sends a 'relay-message' to the boundary node with highest residual energy in the relay node set to inform it serving as a relay node. As shown in Fig. 5, as $DF_{AB} > DF_{AC}$, the overlapping area between Cluster A and Cluster B are the

candidate relay area, and all nodes within it (the blue nodes) are selected as candidate relay nodes. The node remaining most energy is arranged to forward the aggregated data to the sink node. In the proposed clustering protocol, we assign the data collecting task and the data relaying task to different nodes, avoiding a node serving as CH and relay node simultaneously. Each CH collects the coming data from its CMs and forwards the aggregated data to the sink node via its corresponding relay node. As a result, the maximum number of hops from any node to the sink node is three. For large-scale WSNs the protocol needs

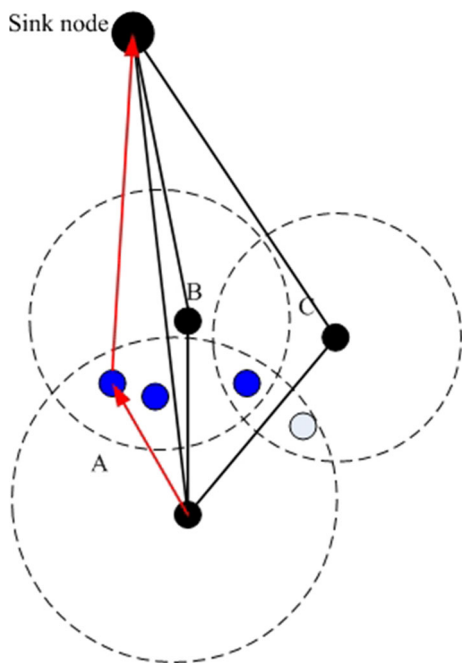


Fig. 5 Relay node selection

to be extended by selecting multi-hop relay nodes. Due to the overlapping cluster formation scheme and the metric setting of direction fitness, the proper next-hop relay nodes can be easily determined by choosing boundary nodes in the next overlapping regions based on the ‘direction fitness’ metric.

In some special cases shown in Fig. 6, two CHs B and C are located on the connection line between A and the sink node, $DF_{AB} = DF_{AC} = 1$. In this case, we prefer to choose the boundary nodes between Cluster A and Cluster B as relay nodes of Cluster A rather than the the boundary nodes between Cluster A and Cluster C. Assigning relay nodes close to CHs is one of the major difference between EEOC and SEECH. In SEECH, the nodes near the sink node are prioritized as relay nodes rather than the nodes near CHs. It should be pointed out that the EEOC protocol is more effectively to balance the energy of the whole network since it can reduce the energy consumption of CHs by requesting relay nodes to play the role of long-distance transmission. The relay node selection is given in Algorithm 3.

Algorithm 3 Relay Node Selection

```

for i=1:1:n do
    if node_type(i)='C' then
        for j=1:1:n do
            if node_type(j)='C' & j ∈ SneiCH(i) then
                calculate DFij;
            end if
        end for
        select k, DFik = max(Dij);
        for m ∈ BN(i, k) do
            if Em0re = max Emre then
                select m0 as the relay node;
            end if
        end for
    end if
end for
    
```



Fig. 6 Special case

4.6 Data transmission and cluster rotation

The CHs create a TDMA schedule to decide which CM can transmit the sensed data within the allocated time, other members will turn off to save energy. Relay nodes communicate with CHs and the sink node within the inter-cluster communication time slot. DSSS (Direct-Sequence Spread Spectrum) is applied to reduce inter-cluster interference. Base station assigns different spreading code to different CHs, and sensors in the same cluster using the same spreading code during the intra-cluster communication period. In data transmission phase, all boundary nodes only need to send data once to its master-CHs. To balance energy consumption among nodes, the CH and relay node roles are rotated periodically after predefined data gathering rounds, and the rotation of CHs and relay nodes are separately.

5 Protocol analysis

5.1 Control message complexity

At the beginning, each node broadcasts a ‘Hand-shaking’ to its neighbors, and the sink node broadcasts a ‘Hello message’ to all nodes. There are $n + 1$ messages in the information collection step. In the CH selection step, suppose that there are n_0 ($n_0 < n$) nodes selected as CHs, n_1 ($n_1 < (n - n_0)$) nodes as HHE nodes, n_2 ($n_2 < n - n_0 - n_1$) ordinary nodes as boundary nodes, and n_3 ($n_3 < n_0$) clusters contain more than φ overlapping regions. As a result, n_0 ‘CH_competing’ messages are broadcast in the network. Hence, there are n_0 ‘CH-Announce’ messages, $(n - n_0)$ ‘master-head’ messages, n_1 ‘high_energy’ messages, n_2 ‘overlap’ messages and n_3 ‘refuse’ messages produced in the cluster formation step. Subsequently, there are n_0 ‘relay-messages’ produced in the relay node selection step. Therefore, the control message complexity in the whole cluster-setup phase is

$$\begin{aligned} T &= n + 1 + n_0 + n_1 + n - n_0 + n_2 + n_3 + n_0 \\ &= 2n + n_1 + n_2 + n_3 + n_0 \end{aligned} \quad (14)$$

Thus, $2n < T < 6n$, the overall communication overhead complexity for cluster-setup phase in the network is $O(n)$.

5.2 Cluster heads and relay nodes distribution

The CHs are properly distributed in the network. In the CH selection step, the broadcast delay related to nodes residual energy and node density is introduced. The node with most residual energy in the dense area will send the competing message within the competing radius. All the nodes in its competing radius receive the message and give up the competing. As a result, the node with most residual energy is the only CH in its competing radius.

The relay nodes are selected based on the ‘direction fitness’ metric within the overlapping area. As a result, the relay nodes are distributed properly near the CHs within the area between CHs and the sink node.

5.3 Energy consumption

Appropriate distribution of CHs can reduce the energy consumption of members in a cluster, and balance the energy consumption among CHs in the different region of the network. There are three types of nodes in the cluster: ordinary nodes, relay nodes and CHs. The required energy for ordinary nodes i is calculated as

$$E_{nonCH}(i) = l_0 \left(E_{elec} + \varepsilon_{fs} \times d_{ioCH}(i)^2 \right) \quad (15)$$

where $d_{ioCH}(i)$ is the distance from node i to its CH. The expectation of square distance from any node to its CH is

$$E[d_{ioCH}^2(i)] = \frac{(1 - \mu)^2 R_i^2}{2} \quad (16)$$

where R_i is the cluster radius defined as the maximum distance from any node to its CH in the cluster, μ is a adjustment coefficient to amend the area of cluster caused by overlapping. Combining expression (15) and (16), we obtain

$$E_{nonCH}(i) = l_0 \left(E_e + \varepsilon_{fs} \times \frac{(1 - \mu)^2 R_i^2}{2} \right) \quad (17)$$

The total energy consumption for all ordinary nodes in a round is

$$E_{nonCH}^{total} = (n - n_0) l_0 E_e + \varepsilon_{fs} \times \frac{\sum_{i=1}^{n_0} n_i (1 - \mu)^2 R_i^2}{2} \quad (18)$$

The node density in cluster CH_i is

$$\rho_i = \frac{n_i}{\pi R_i^2}. \quad (19)$$

Substituting expression (19) into (18), we obtain

$$E_{nonCH}^{total} = (n - n_0)l_0E_e + \epsilon_{fs} \times \frac{\pi \sum_{i=1}^{n_0} \rho_i (1 - \mu)^2 R_i^4}{2}. \tag{20}$$

The energy consumption for CH i to receive and send the data to its corresponding relay node is

$$E_{CH}(i) = l_0(2E_e + E_{DA} + \epsilon_{fs} \times (1 - \mu)^2 R_i^2). \tag{21}$$

where the distance from CH i to its relay node is approximated as the cluster radius $(1 - \mu)R_i$. The total energy consumption for all CHs in a round is

$$E_{CH}^{total} = n_0l_0(2E_e + E_{DA}) + (1 - \mu)^2 \epsilon_{fs} \sum_{i=1}^{n_0} R_i^2. \tag{22}$$

We assume that all relay nodes adopting the multi-path channel energy consumption model in transmission. According to the expression (56) in [34], the expected fourth power distance from the sink node to the sensing field is

$$E[d_{toBS}^4] = \frac{193M^4}{720} \tag{23}$$

The energy consumption for relay node i to receive and send the aggregated data to the sink node is

$$E_{relay}(i) = l_0(E_e + \epsilon_{mp}d_{toBS}^4(i)) \tag{24}$$

The total energy consumption for all relay nodes of the network in a round is

$$E_{relay}^{total} = n_0l_0 \left(E_e + \epsilon_{mp} \frac{193M^4}{720} \right) \tag{25}$$

The total energy consumption of the whole network in a round is

$$E_{total} = E_{nonCH}^{total} + E_{CH}^{total} + E_{relay}^{total} \tag{26}$$

According to expressions (20), (22) and (25), we can see that after the number of CHs n_0 is determined, the total energy consumption of the sensor network in a data gathering round primarily depends on the distribution of CHs and node density in the sensing field and overlapping degree of the formatted clusters. Uniform distribution of CHs yields proper cluster radius and sensors in dense area should be grouped into small clusters to reduce energy consumption.

6 Simulations

In this section, we evaluate the EEOC protocol from energy efficiency perspective to demonstrate the superiority of our protocol in comparison with LEACH [13] and SEECH [28] via MATLAB. For simplicity, we assume an

ideal PHY layer with no losses, and error-free links with no collisions are guaranteed by the TDMA schedule. Moreover, we assume that all the coming packets from CMs can be aggregated into a constant length by their CH in each cluster. The average residual energy and the number of nodes alive are two major metrics used to explore the energy efficiency. To investigate the impact of node density and the distance from monitoring area to the sink node on the proposed protocol, we evaluate the performance from three aspects under two different network size (Case 1: 100 m × 100 m with 100 nodes, $W=75$; Case 2: 100 m × 100 m with 200 nodes, $W=100$). First, the energy efficiency of the whole network is evaluated. Second, we further calculate the energy consumption among nodes with different roles (CHs, relay nodes and CMs) to investigate the influence of the distributions of node’s energy consumptions on the network lifetime. At last, the energy consumption from different region of the network is evaluated to prove the advantage of the protocol in balancing energy consumption at different levels of the network topology. The parameters of simulation scenario is described in Table 1.

6.1 Energy efficiency of the proposed protocol

In the simulations, when a node runs out of the limited energy, it is taken to be dead. There are different definitions on the lifetime of network, for example, the time when the first node dies (FND), the time when all nodes die (AND) and the time when a specified percentage (10 %) of nodes die (TND) [9]. All of them are used in this paper to evaluate the energy efficiency performance of the proposed protocol.

The number of living nodes and their average residual energy are shown in Figs. 7 and 8. Result in Fig. 7 indicates that EEOC achieves longer stable work stage (from start to the first node dead, TND) and survival time (from start to ten percent of nodes dead, TND) compared with LEACH and SEECH, while gets shorter AND time compared with LEACH in both cases. The comparison of FND, TND and AND is shown in Fig. 9. The result in Fig. 8 indicate that EEOC remains more residual energy in the whole survival time due to appropriate CH distribution and relay nodes distribution. Meanwhile, a shorter period between FND time and AND time means that the energy consumption is well balanced among nodes in the network. A long FND time and the rapid decrease in the number of dead nodes for our protocol shown in Fig. 7 demonstrates that it has obvious advantages in saving total energy and balancing energy consumption. Compared the results in Figs. 7(a), 8(a) and 9(c) with the results in Figs. 7(b), 8(b) and 9(c), it can be seen that our protocol and the SEECH protocol both achieve high energy efficiency in the

Table 1 Experimental parameters

Parameters	Value	Parameters	Value
Area	100 m × 100 m	ε	0.2
n	100 (Case 1) 200 (Case 2)	W	75 (Case 1); 100 (Case 2)
τ_1	2	Packet size	4000 bits
φ	4	τ_2	4
E_0	0.5 J	r_0	30
ε_{fs}	10 PJ/bit · m ²	E_{elec}	50 nJ/bit
		ε_{mp}	0.0013 PJ/bit · m ⁴

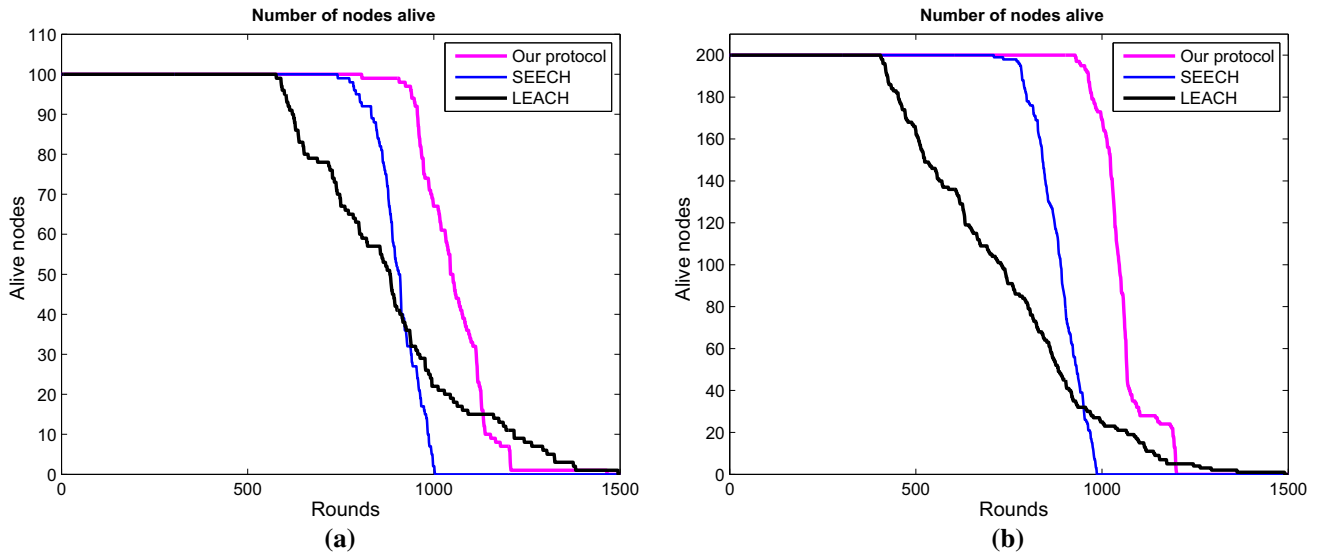


Fig. 7 Number of nodes alive. **a** Case 1, **b** Case 2

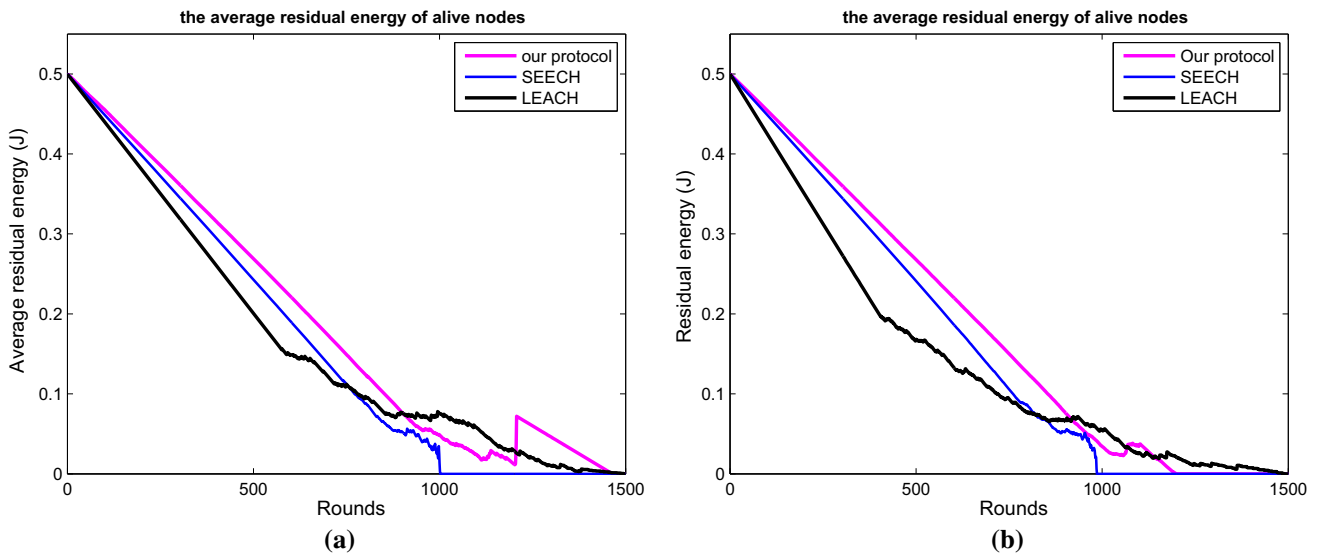


Fig. 8 Average residual energy of living nodes. **a** Case 1, **b** Case 2

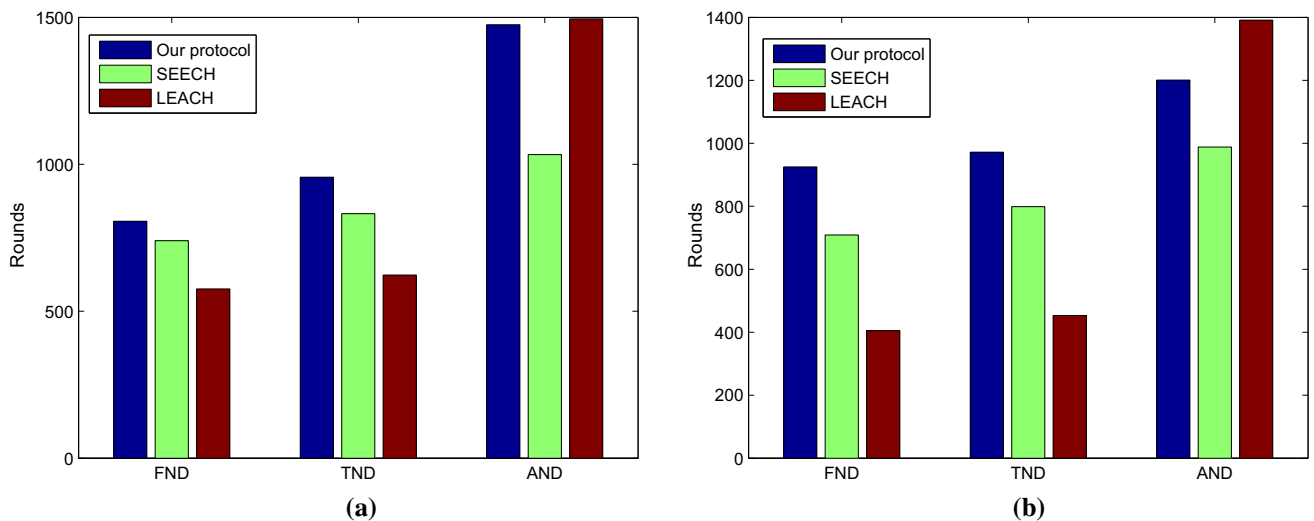


Fig. 9 Lifetime compared among three protocols. a Case 1, b Case 2

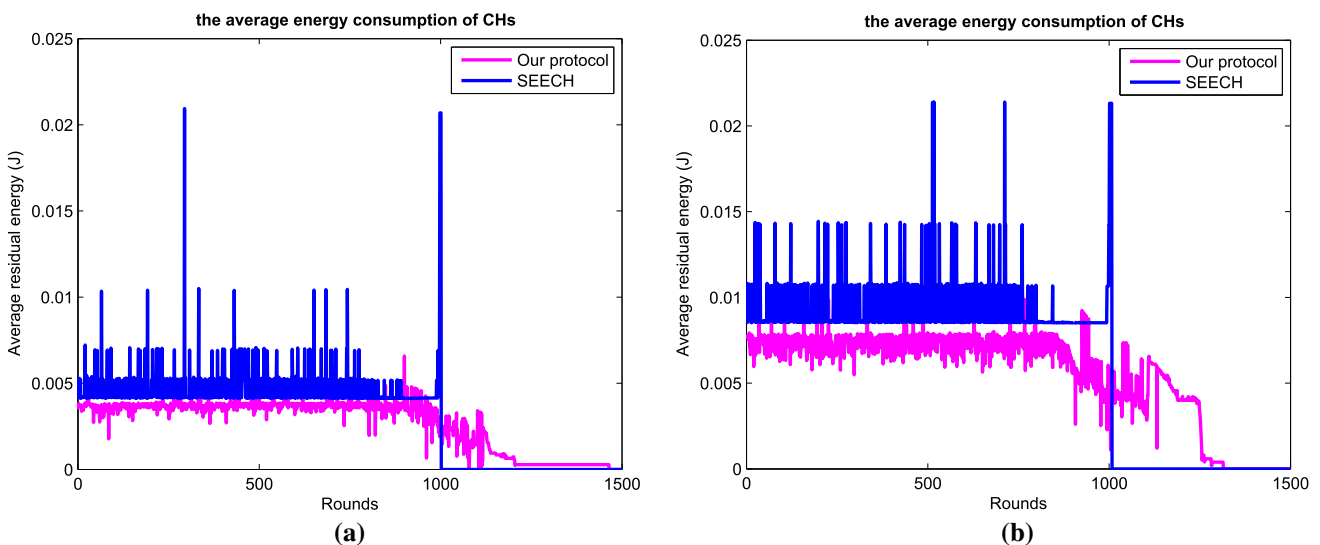


Fig. 10 Average energy consumption of CHs. a Case 1, b Case 2

two simulation cases. This observation indicates the robustness our proposed protocol and the SEECH protocol.

6.2 Energy consumption among different roles of nodes

By avoiding a node serving as a CH and a relay node simultaneously, the EEOC and SEECH protocol both get excellent energy efficiency. Furthermore, the EEOC protocol gets even better results than SEECH. To investigate the difference between EEOC and SEECH, the energy consumption among different roles of nodes is compared. We form the same number of clusters ($K_{CH}=5$) as SEECH

formed in its first simulation scene (the same to our simulation setting) to obtain the reasonable comparison result.

The average energy consumption of CHs, relay nodes and cluster members in each rounds are shown in Figs. 10, 11 and 12. As we prefer to choose the relay nodes near the CHs rather than the ones near the sink node, the energy depletion for CHs caused by remote communication is avoid, whereas the selected relay nodes should consume more energy to forward data to the sink node due to a longer transmission distance. Compared the results in Figs. 10(a), 11(a) and 12(a) with the results in Figs. 10(b), 11(b) and 12(b), it can be seen that CHs and relay nodes consume more energy during each data gathering round in

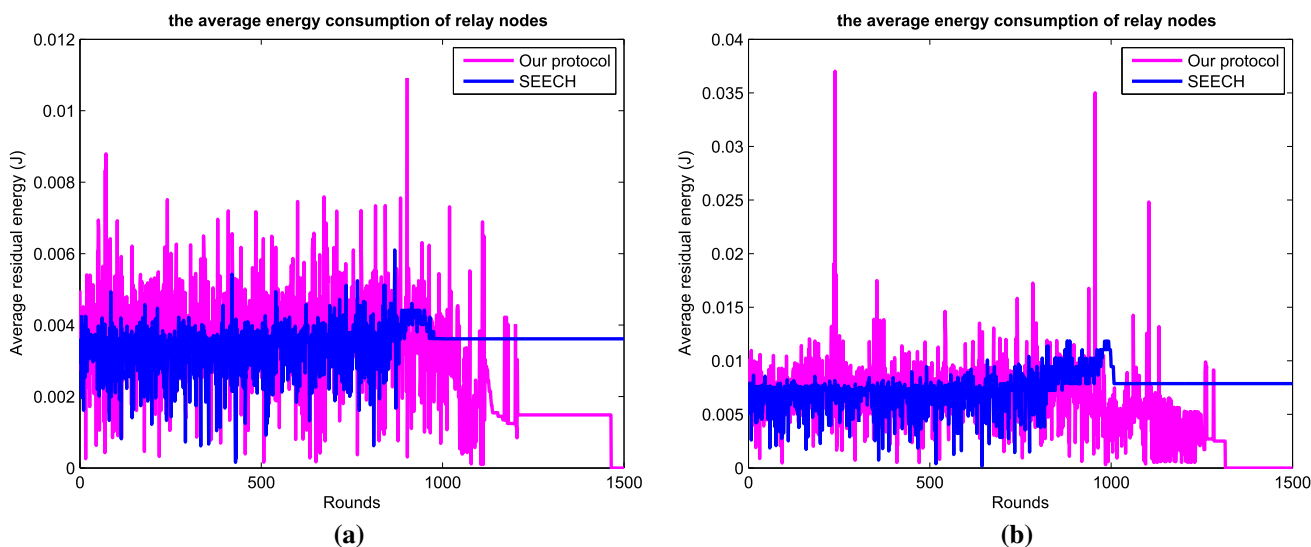


Fig. 11 Average energy consumption of relay nodes. **a** Case 1, **b** Case 2

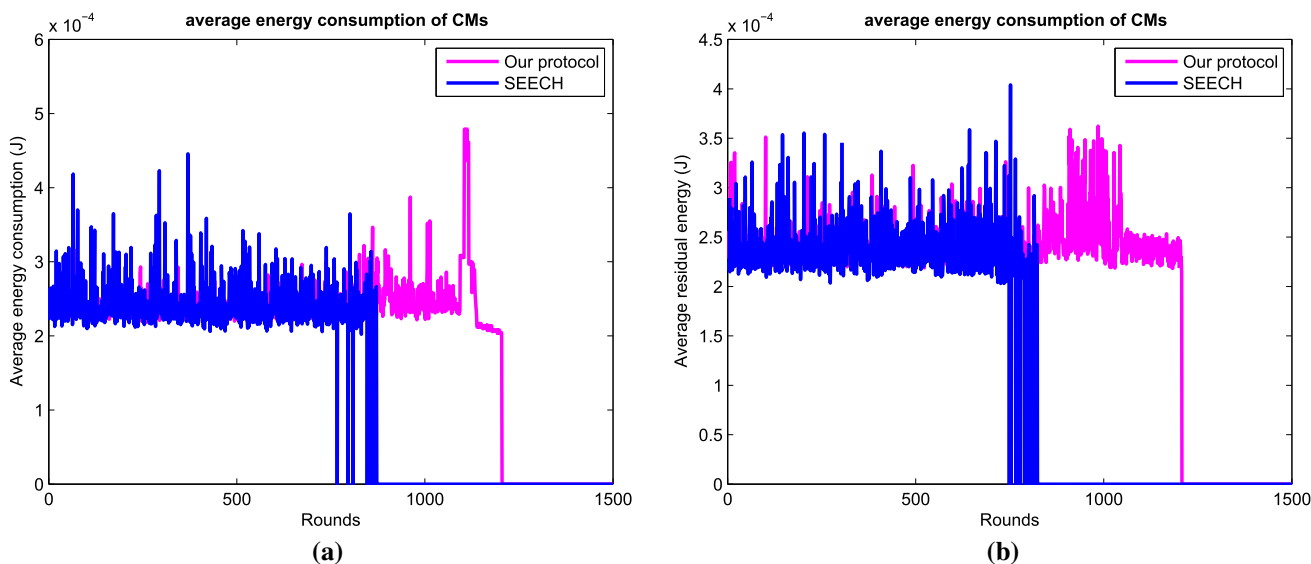


Fig. 12 Average energy consumption of CMs. **a** Case 1, **b** Case 2

a denser nodes deployment with a farer base station (Case 2). The results in Figs. 10 and 11 indicate that the average energy consumption of EEOC is much lower at the cost of higher energy consumption of relay nodes compared with SEECH. Since CHs consume huge energy to collect data from its CMs, it is beneficial to assign relay nodes near CHs to reduce inter-cluster energy consumption, avoiding untimely death of CHs. From Fig. 12, we can see that the average energy consumption of CMs in EEOC is approximately the same during the survival time compared with the energy consumption of CMs in SEECH. Combining the results shown in Figs. 7, 8, 9, 10, 11 and 12, it can be seen that energy balancing among CHs makes a great contribution for prolonging the network lifetime.

6.3 Energy consumption in different regions of network

The EEOC protocol aims at balancing energy consumption among all nodes in the network. The energy consumption rate of different regions in the network should be as synchronous as possible. We equally divide the network area into three regions (a_1 region, a_2 region and a_3 region) as shown in Fig. 13, and investigate the average energy consumption in the three regions of the network with different protocols.

Simulation results in Figs. 14, 15 and 16 indicate that the EEOC protocol achieves best energy balancing performance in different region of the network under both

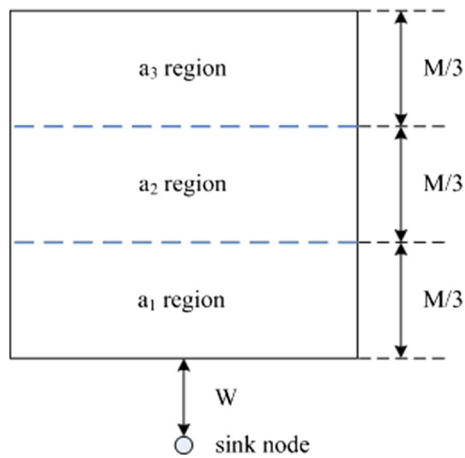


Fig. 13 Different regions of the network

simulation cases. Figure 14 shows the result of LEACH. Since the distance metric is not considered during the cluster-setup phase and CHs communicate with the sink directly, the average residual energy of each region reduces with the distance to the sink node. As shown in Fig. 15, SEECH achieves better performance than LEACH. It should be pointed out that the average residual energy in a_1 region is the lowest, even though a_1 region is closest to the sink node. The reason is that most of the relay nodes are selected near the sink node (in the a_1 region) in SEECH. The energy consumption are well balanced in three regions by using the EEOC protocol, shown in Fig. 16. The excellent performance is achieved because of the proper distribution of relay nodes near CHs in the region between CHs and the sink node. Compared the results in

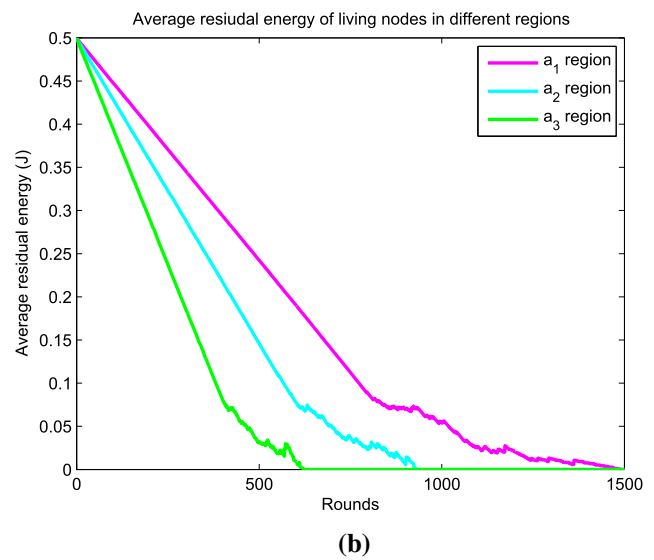
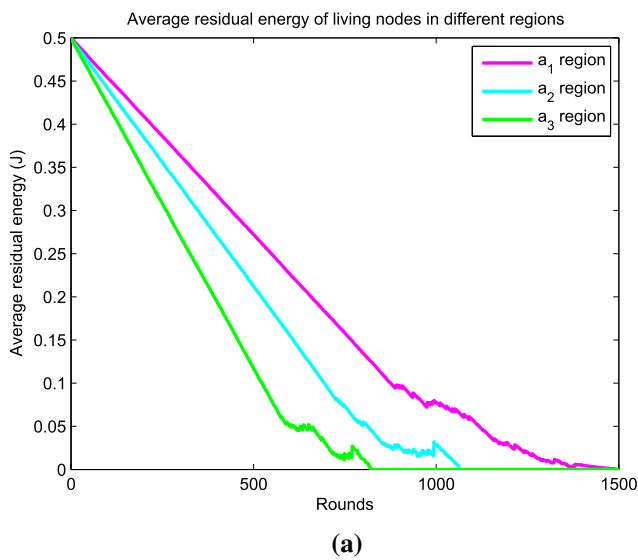


Fig. 14 Average residual energy in different areas in LEACH. **a** Case 1, **b** Case 2

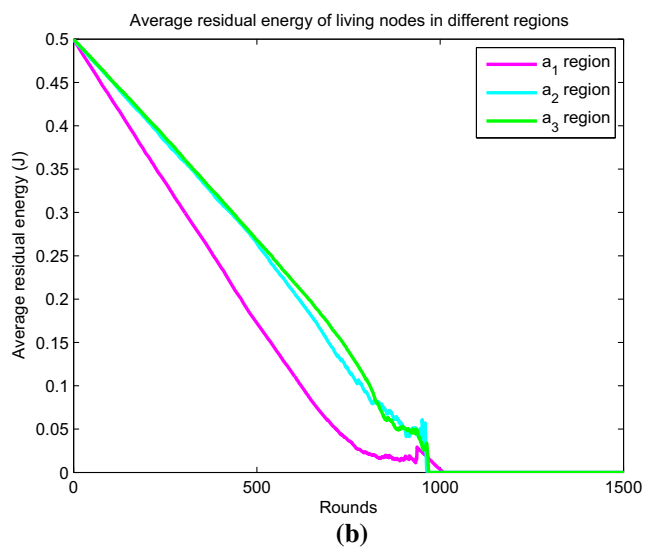
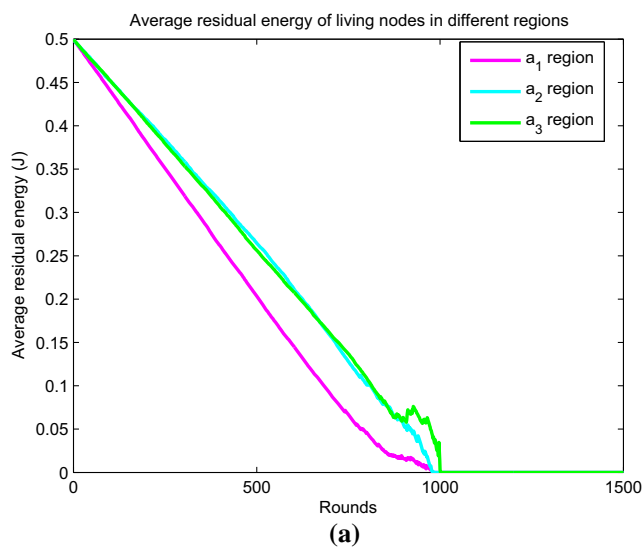


Fig. 15 Average residual energy in different areas in SEECH. **a** Case 1, **b** Case 2

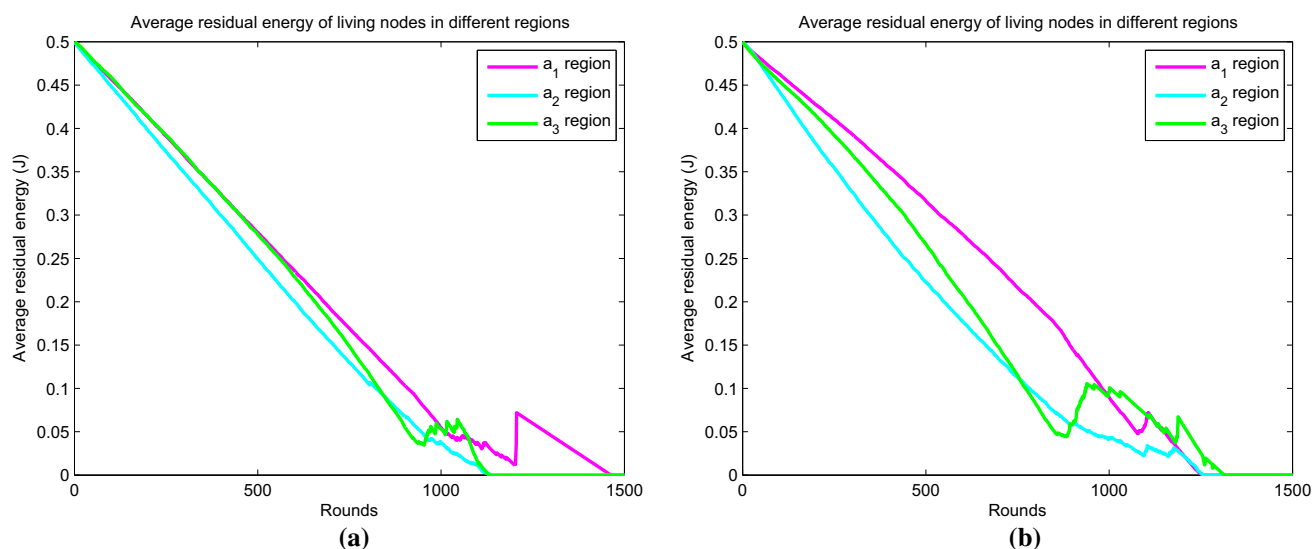


Fig. 16 Average residual energy in different areas in our protocol. **a** Case 1, **b** Case 2

Figs. 14(a), 15(a) and 16(a) with the results in Figs. 14(b), 15(b) and 16(b), it can be observed that the energy consumption can be more balanced when the sink node is closer to the monitoring area (Case 1).

7 Conclusion

Reducing and balancing energy consumption both can prolong the network lifetime, which should be considered simultaneously. An energy-efficient overlapping clustering protocol has been proposed for WSNs in this paper. Based on the overlapping cluster formation mechanism, the selected CHs are uniformly distributed, leading to low energy consumption during the intra-cluster communication phase. The proper high residual energy nodes in the overlapping areas are selected as relay nodes according to the distance metric, which can reduce energy consumption caused by long distance transmission between CHs and relay nodes, and hence balance energy consumption of the whole network. Simulation results show that the EEOC protocol can prolong the network lifetime better than LEACH protocol and SEECH protocol. Furthermore, it has been shown from simulations that EEOC achieves better energy-balancing quality than SEECH and LEACH.

Acknowledgements This work is supported by National Natural Science Foundation (NNSF) from China (61273073, 61374107).

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. *Computer Networks*, 38, 393–422.
2. Yick, J., Mukherjee, B., & Ghosal, D. (2008). Wireless sensor network survey. *Computer Networks*, 52, 2292–2330.
3. Abdelaal, M., & Theel, O. (2014). Recent energy-preservation endeavours for longlife wireless sensor networks: A concise survey. In *IEEE 7th conference on wireless and optical communications networks (WOCN)* (pp. 1–7).
4. Rault, T., Bouabdallah, A., & Challal, Y. (2014). Energy efficiency in wireless sensor networks: A top-down survey. *Computer Networks*, 67, 104–122.
5. Keskin, M. E., Altanel, I. K., Aras, N., & Ersoy, C. (2014). Wireless sensor network lifetime maximization by optimal sensor deployment, activity scheduling, data routing and sink mobility. *Ad Hoc Networks*, 17(6), 18–36.
6. Ishmanov, F., Malik, A. S., & Kim, S. W. (2011). Energy consumption balancing (ECB) issues and mechanisms in wireless sensor networks (WSNs): A comprehensive overview. *European Transactions on Telecommunications*, 22(4), 151–167.
7. Afsar, M. M., & Tayarani-N, M. H. (2014). Clustering in sensor networks: A literature survey. *Journal of Network and Computer Applications*, 46, 198–226.
8. Abbasi, A. A., & Younis, M. (2007). A survey on clustering algorithms for wireless sensor networks. *Computer Communications*, 30, 2826–2841.
9. Liu, A. F., Zhang, P. H., & Chen, Z. G. (2011). Theoretical analysis of the lifetime and energy hole in cluster based wireless sensor networks. *Journal of Parallel and Distributed Computing*, 71(10), 1327–1355.
10. Mhatre, V., & Rosenberg, C. (2004). Design guidelines for wireless sensor networks: Communication, clustering and aggregation. *Ad Hoc Networks*, 2, 45–63.
11. Lai, W. K., Fan, C. S., & Lin, L. Y. (2012). Arranging cluster sizes and transmission ranges for wireless sensor networks. *Information Sciences*, 183, 117–131.
12. Chen, G. H., Li, C. F., Ye, M., & Wu, J. (2009). An unequal cluster-head routing protocols in wireless sensor networks. *Wireless Networks*, 15, 193–207.
13. Heinzelman, W. B., Chandrakasan, A. P., & Balakrishnan, H. (2002). An application-specific protocol architecture for wireless microsensor networks. *IEEE Transactions on Wireless Communications*, 1(4), 660–670.
14. Liu, Z. X., Zheng, Q. C., X, L., & Guan, X. P. (2012). A distributed energy-efficient clustering algorithm with improved

- coverage in wireless sensor networks. *Future Generation Computer Systems*, 28(5), 780–790.
15. Smaragdakis, G., Matta, I., & Bestavros, A. (2004). SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. In *Second international workshop on sensor and actor network protocols and applications (SANPA 2004)*.
 16. Qing, L., Zhu, Q., & Wang, M. (2006). Design of a distributed energy-efficient clustering algorithm for heterogeneous wireless sensor networks. *Computer Communications*, 29(12), 2230–2237.
 17. Zhou, H., Wu, Y., Hu, Y., & Xie, G. Z. (2010). A novel stable selection and reliable transmission protocol for clustered heterogeneous wireless sensor networks. *Computer Communications*, 33(15), 1843–1849.
 18. Zhen, H., Li, Y., & Zhang, G. J. (2013). Efficient and dynamic clustering scheme for heterogeneous multi-level wireless sensor networks. *Acta Automatica Sinica*, 39(4), 454–460.
 19. Xu, K. N., Hassanein, H., Takahara, G., & Wang, Q. H. (2010). Relay node deployment strategies in heterogeneous wireless sensor networks. *IEEE Transactions on Mobile Computing*, 9(2), 145–159.
 20. Wang, F., Wang, D., & Liu, J. C. (2011). Traffic-aware relay node deployment: Maximizing lifetime for data collection wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 22(8), 1415–1423.
 21. Cui, Q., Yang, X. J., Tao, X. F., & Zhang, P. (2014). Optimal energy-efficient relay deployment for the bidirectional relay transmission schemes. *IEEE Transactions on Vehicular Technology*, 63(6), 2625–2641.
 22. Chang, J. Y., & Lin, Y. S. (2014). A clustering deployment scheme for base stations and relay stations in multi-hop relay networks. *Computers and Electrical Engineering*, 40, 407–420.
 23. Liao, Y., Qi, H., & Li, W. Q. (2013). Load-balanced clustering algorithm with distributed self-organization for wireless sensor networks. *IEEE Sensor Networks*, 13(5), 1056–1498.
 24. Kuila, P., Gupta, S. K., & Jana, P. K. (2013). A novel evolutionary approach for load balanced clustering problem for wireless sensor networks. *Swarm and Evolutionary Computation*, 12, 48–56.
 25. Chanak, P., Banerjee, I., & Rahaman, H. (2015). Load management scheme for energy holes reduction in wireless sensor networks. *Computers and Electrical Engineering*, 48, 1–15.
 26. Liu, T., Li, Q. R., & Liang, P. (2012). An energy-balancing clustering approach for gradient-based routing in wireless sensor networks. *Computer Communications*, 35, 2150–2161.
 27. Baranidharan, B., & Santhi, B. (2016). DUCF: Distributed load balancing unequal clustering in wireless sensor networks using Fuzzy approach. *Applied Soft Computing*, 40, 495–506.
 28. Tarhani, M., Kaviani, Y., & Siavoshi, S. (2014). SEECH: Scalable energy efficient clustering hierarchy protocol in wireless sensor networks. *IEEE Sensor Journal*, 14(11), 3944–3954.
 29. Liu X. F., et al. (2011). Energy efficient clustering for WSN-based structural health monitoring. In *IEEE INFOCOM* (pp. 2768–2776).
 30. Ammar, I., Miskeen, G., & Awan, I. (2013). Overlapped schedules with centralized clustering for wireless sensor networks. In *IEEE 27th international conference on advanced information networking and applications* (pp. 33–40).
 31. Kalyanasundaram, B., & Younis, M. (2013). Using mobile data collectors to federate clusters of disjoint sensor network segments. In *IEEE ICC-Ad-hoc and sensor networking symposium* (pp. 1496–1500).
 32. Dai, G. Y., et al. (2015). A novel distributed clustering-based MDS algorithm for nodes localization in WSNs. *International Journal of Grid Distribution Computing*, 8(2), 79–90.
 33. Youssef, M. A., Youssef, A., & Younis, F. (2009). Overlapping multihop clustering for wireless sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 20(12), 1844–1856.
 34. Amini, A., Vahdatpour, A., Xu, W. Y., Gerla, M., & Sarrafzadeh, M. (2012). Cluster size optimization in sensor networks with decentralized cluster-based protocols. *Computer Communications*, 35, 207–220.



Yuan Hu received his B.S. degrees from the Wuhan University of Science & Technology in 2011. Currently, he is pursuing his Ph.D. degree in the Department of Automation at East China University of Science & Technology. His current research interests include wireless sensor networks and congestion control.



Yugang Niu received his M.Sc. and Ph.D. degrees from the Nanjing University of Science & Technology in 1992 and 2001, respectively. In 2003, he joined the School of Information Science & Engineering, East China University of Science & Technology where he is now a Professor of Control Science and Engineering. Professor Niu is Associate Editor of the International Journal of System Sciences, Information Science, Neurocomputing, and Conference Editorial Board of IEEE Control Systems Society, and a member of Editorial board of IET Control Theory & Applications. His research interests include stochastic systems, sliding mode control, Markovian jump systems, networked control systems, wireless sensor network, smart grid.