Uneven Clustering Routing Algorithm Based on Optimal Clustering for Wireless Sensor Networks

Jianpo Li and Junyuan Huo
School of Information Engineering, Northeast Dianli University, Jilin, 132012, China
Email: jianpoli@hotmail.com; huojunyuan@sina.com

Abstract—Hierarchical routing algorithm as an energy optimization strategy has been widely considered as one of the effective ways to save energy for wireless sensor networks. In this paper, we propose an uneven clustering routing algorithm based on optimal clustering. The algorithm considers the calculation of optimal cluster number, cluster head selection, cluster radius calculation, and isolated node management. Firstly, a new method to calculate the optimal cluster number is presented by considering data fusion rate and location adaptability. Secondly, we propose a cluster head selection algorithm by considering residual energy, initial energy, average energy consumption, and node degree. Thirdly, in order to solve "hot spot" problem, a self-adaptive uneven clustering algorithm is proposed, which takes node degree into consideration. Besides, we propose a solution to solve "isolated nodes problem". The isolated nodes join the nearest clusters according to the cost for joining near cluster and sending data to the base station through the nearest cluster head. The simulation results show the superiority in terms of network lifetime, the number of alive nodes, and the total energy consumption.

Index Terms—Wireless sensor networks, energy optimization strategy, optimal cluster number, cluster head selection, uneven clustering algorithms, isolated nodes

I. Introduction

Wireless Sensor Networks (WSN) typically consist of a large number of energy-constrained sensor nodes with limited onboard battery resources which form a dynamic multi-hop network. In a lot of applications supported by wireless sensor networks, node energy is difficult to renew [1]. Therefore, energy optimization is a critical issue in the design of wireless sensor networks [2]-[4].

At present many techniques have been proposed for the improvement of energy efficiency in energy-constrained, distributed wireless sensor networks. These techniques include energy optimization strategy based on node transmission power, such as common power (COMPOW) protocol [5]; energy optimization strategy based on routing protocol, such as low energy adaptive clustering hierarchy (LEACH) protocol [6]; energy optimization strategy based on medium access control (MAC) protocol, such as Sensor MAC (SMAC) protocol [7]; energy optimization strategy based on data fusion, such as maximum lifetime data gathering with aggregation (MLDA) algorithm [8]; energy optimization strategy based on node sleeping scheme, such as dynamic balanced-energy sleep scheduling scheme [9]. Among these techniques, energy efficiency routing protocol has been widely considered as one of the effective ways to save energy. Existing routing protocols can be generally divided into two categories: flat routing and hierarchical routing. Flat routing protocol is easy to implement, without additional cost of topology maintenance and packet routing. However, it has several shortcomings such as message implosion, overlay, and resource blindness [1]. Hierarchical routing protocol, also known as clustering routing protocol, such as LEACH protocol [6] and Hybrid Energy-Efficient Distributed clustering (HEED) protocol [10], has proposed the methods that using cluster heads to form the clusters. Researches show the hierarchical routing protocol is better than flat routing protocol in adaptability and energy efficiency.

LEACH protocol is one of the most popular hierarchical routing protocols for wireless sensor networks. In this protocol, the entire network is divided into several clusters. The cluster head node is used as a router to the base station. All members of cluster transmit their data to the cluster head. Then the cluster head aggregates and compresses all the received data and sends them to the base station. The operation of LEACH is divided into rounds. Each round includes a set-up phase and a steady-state phase. In the set-up phase, each node has the equal probability to become a cluster head randomly by using a distributed algorithm. Based on the received signal strength of the advertisement from each cluster head, each non-cluster head node determines its cluster in this round. It chooses the cluster head as minimum communication energy. Then the cluster head node sets up a Time Division Multiple Address (TDMA) schedule and transmits this schedule to the nodes in the cluster. This ensures that there are no collisions among data messages and allows the radio components of each non-cluster head node to be turned off at all times except during their transmission period. In the steady-state phase, the time is divided into frames. Nodes send their data to the cluster head at most once per frame during their allocated transmission slot. Non-cluster head nodes send the collected data to the cluster head node. Once the

Manuscript received August 22, 2015; revised February 17, 2016.
This work was supported by National Natural Science Foundation of China (No. 61501106), Science and Technology Foundation of Jilin Province (No. 20150520097JH), Science and Technology Foundation of Jilin City (No.201464056), and Innovation Foundation of Northeast Dianli University (No. Y2014021).
Corresponding author email: jianpoli@hotmail.com.
doi:10.12720/jcm.11.2.132-142

cluster head receives all the data, it performs data aggregation and sends them to the base station directly. Compared with the normal flat multi-hop routing algorithm and static hierarchical algorithm, the network lifetime of LEACH can be prolonged by 15%. However, there are also some shortcomings. For example, the residual energy of node is not taken into consideration during the cluster head selection; uneven distribution of cluster heads and cluster sizes, due to the random cluster head selection mechanism, may causes the decline in the balance of network load. In large-scale network, single-hop data transmission will lead to some cluster heads die in advance, which are far away from the base station. So the lifetime of the whole network will be affected. To avoid uneven distribution problem of cluster heads and cluster sizes in LEACH, references [11] and [12] propose LEACH-C (LEACH-Centralized) and LEACH-F (LEACH with Fixed clusters) algorithm. The LEACH-C utilizes the base station for cluster formation, unlike LEACH where nodes self-configure themselves into clusters. In the LEACH-C, firstly, the base station receives information regarding the location and energy level of each node in the network. Then it selects a predetermined number of cluster heads and configures the network into clusters. The cluster groupings are chosen to minimize the energy required for non-cluster head nodes to transmit their data to their respective cluster heads. In LEACH-F, clusters are created using the centralized cluster formation algorithm developed for LEACH-C. The base station uses simulated annealing algorithm to determine optimal clusters and broadcasts the cluster information to the nodes. The nodes can determine the TDMA schedule and the order to become cluster heads. The first node listed in the cluster list becomes cluster head in the first round, the second node listed in the cluster list becomes cluster head in the second round, and so forth. In LEACH-F, there is no set-up requirement for the different rounds. The nodes implicitly know when they are cluster heads and when they are non-cluster heads. The steady-state phase of LEACH-F is the same as that of LEACH. LEACH-F is not able to be applied into any sort of dynamic systems. The new nodes are not allowed to join the system in this fixed protocol schedule. And the system does not adjust its behavior when some nodes die or move. Therefore, LEACH-F is not a useful protocol architecture for real system. To avoid cluster heads premature death in LEACH, reference [13] establishes a new threshold which reflects the node energy, distance between node and base station, distance between cluster head and base station. Simulation results show that the algorithm is better in balancing the node energy and prolonging the network lifetime. But the protocol does not solve uneven distribution problem of cluster heads and cluster sizes. LEACH does not take residual energy and the distance that between node and base station into consideration while choosing the cluster head. The improved LEACH algorithm considers residual energy and distance between node and base station to select cluster head [14]. To save energy, the paper proposes to start the steady-state operation of a node only if the value sensed by a node is greater than the predetermined threshold value. The threshold value will be set by the terminal user at the application layer. LEATCH (Low Energy Adaptive Tier Clustering Hierarchy) algorithm offers a good compromise between delay and energy consumption. It proposes two level hierarchical approach. Each cluster is divided into some small clusters that are called Mini Clusters. For each Mini Cluster, the authors define a Mini Cluster Head (MCH). Simulation results show that LEATCH performs better in term of energy, delay, coverage, and scalability [15].

All improved protocols based on LEACH presented above can't solve the uneven distribution problem of cluster heads and cluster sizes. Reference [10] proposes HEED protocol. However, the clustering process requires a number of iterations. During each iteration, a node becomes a cluster head with a certain probability which considers the mixture of energy, communication cost, and average minimum reachability power (AMRP). All other nodes, which are not cluster heads, select the cluster head which has the lowest intra-cluster communication cost and directly communicate with cluster heads. Unlike LEACH, HEED creates well-balanced clusters. It has more balanced energy consumption and longer network lifetime. To achieve a longer network lifetime and cluster formation than HEED protocol, reference [16] presents a distributed dynamic clustering protocol based on HEED which exploits non-probabilistic approach and Fuzzy Logic (HEED-NPF). In this protocol, cluster head selection is finished by Fuzzy Logic which uses node degree and node centrality as input parameters. The output is the Fuzzy cost. Every node selects the cluster head with least cost and join it. Non-probabilistic cluster head selection is implemented through introducing delay, which is inversely proportional to residual energy for each node. Consequently node with more residual energy has more chance to become cluster head. The approach is more effective in prolonging the network lifetime than HEED and provides better cluster formation in the field.

To avoid "hot spot" problem, reference [17] proposes an Unequal Clustering Size (UCS) model for network organization, which can lead to more uniform energy consumption among the cluster head nodes and prolong network lifetime. At the same time the authors expand this approach to homogeneous sensor networks. The simulation results show that UCS model can lead to more uniform energy consumption in a homogeneous network as well. However, the assumptions in UCS are not in accordance with the actual situation. Reference [18] proposes and evaluates an Energy-Efficient Unequal Clustering (EEUC) mechanism for periodical data gathering applications in WSN. It wisely organizes the network via unequal clustering and multi-hop routing. EEUC is a distributed competitive algorithm. Unlike LEACH and HEED, the cluster heads are selected by localized competition without iteration. The node's competition range decreases as its distance to the base station decreasing. The result is that clusters closer to the base station are expected to have smaller cluster sizes. They will consume lower energy during the intra-cluster data processing and preserve more energy for the inter-cluster relay traffic. In the proposed multi-hop routing protocol for inter-cluster communication, a cluster head chooses a relay node from its adjacent cluster heads according to the node's residual energy and its distance to the base station. Simulation results show that EEUC successfully balances the energy consumption over the network and achieves a remarkable network lifetime improvement. The EAUCF (Fuzzy Energy-Aware Unequal Clustering) algorithm [19] aims to decrease the intra-cluster work of the cluster-heads. Fuzzy logic approach is adopted to handle uncertainties in cluster-head radius estimation. The algorithm is compared with some popular clustering algorithms. The experiment results show that EAUCF performs better in terms of first node dies, half of the nodes alive, and energy-efficiency metrics in all scenarios.

In this paper, based on the protocols presented above, we propose an uneven clustering routing algorithm based on optimal clustering. The main work consists of calculating optimal cluster number, choosing cluster heads, calculating the cluster radius, and dealing with isolated nodes. The rest of this paper is organized as follows. In Section II, the basic process of hierarchical routing is briefly described. In Section III, we propose a new method to calculate the optimal cluster number to replace the expression in LEACH which does not work when the base station is at (0, 0). The new algorithm considers fusion rate and suits for any situation for the base station. It converts energy consumption optimization problem to the optimal clusters calculation. In Section IV, we briefly summarize the shortcomings in random probability selection as LEACH and single factor iteration selection as HEED. Furthermore we propose a new cluster selection algorithm by considering more factors such as residual energy, initial energy, average energy consumption, node degree(the number of neighbor nodes within the cluster range) et al. In section V, in order to solve "hot spot" problem, a self-adaptive uneven clustering algorithm is proposed, which takes node degree into consideration. In traditional algorithm, the isolated nodes in WSN usually form clusters separately and send data to the base station directly which will cost a lot of energy. So, in Section VI, we propose a solution in which isolated nodes join the nearest clusters according to the cost for joining near cluster and send data to the base station through the nearest cluster head. Section VII shows the simulation and numerical analysis. Final conclusion remark are made in Section VIII.

II. The Basic Process of Hierarchical Routing

The operation of hierarchical routing for WSN can be divided into set-up phase and steady-state phase. The main work of set-up phase consists of calculating the optimal cluster number, choosing cluster heads, calculating cluster radius, broadcasting cluster head message, joining cluster, dealing with isolated nodes and creating TDMA schedules. The main work of steady-state phase consists of constructing route path, gathering data to cluster head, data fusion and sending data to the base station.

In the set-up phase, we need the optimal number of clusters which can be obtained by using energy model. A radio model proposed in LEACH [6] is shown in Fig. 1.

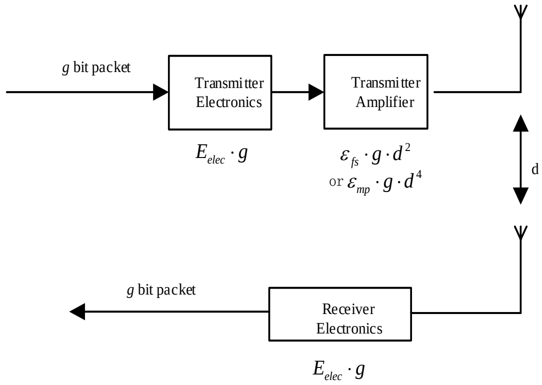

Fig. 1. The radio energy consumption model

where E_{elec} is the transmitter energy consumption per bit, g is the number of bits, ε_{fs} is proportional constant of the energy consumption for the transmit amplifier in free space channel model ($\varepsilon_{fs} \cdot g \cdot d^2$ power loss), ε_{mp} is proportional constant of the energy consumption for the transmitter amplifier in multipath fading channel model ($\varepsilon_{mp} \cdot g \cdot d^4$ power loss), the distance between transmitter and receiver is d, the transmitter energy consumption to run the transmitter or receiver circuitry is $E_{elec} \cdot g$, the energy consumption in transmitter amplifier is $\varepsilon_{fs} \cdot g \cdot d^2$ or $\varepsilon_{mp} \cdot g \cdot d^4$.

In this model, the free space channel model and multipath fading channel models are used, which depend on the distance between transmitter and receiver. If the distance is less than a threshold, the free space model is used; otherwise, the multipath fading channel model is used.

Each cluster head dissipates energy to receive data from the nodes, aggregate the data and transmit the aggregated data to the base station. Since the base station is far from the nodes, assume that the energy dissipation follows the multipath fading channel model (d^4 power loss). Each non-cluster head node only need to transmit its data to the cluster head once during a frame. The energy cost follows the free space channel model (d^2 power loss) under the assumption that the distance from the non-cluster head node to the cluster head is small. The optimal cluster number k [11], [12] can be obtained as:

$$k = \sqrt{\frac{\varepsilon_{fs} M^2 N}{2\pi \varepsilon_{mp} {d_{toBS}}^4}} \quad (1)$$

where M^2 stands for $M \times M$ random distributed area for nodes, N is the number of nodes, d_{toBS} is the distance between node and base station [6].

In LEACH, the simulation results show that the optimal proportion of cluster head is 5% [6]. In HEED [10] and many other improved protocols [19], [20], the above proportion is also adopted. In these protocols, the optimal cluster number is obtained as (1). When the base station is at coordinate (75, 175), $k = 1 \sim 6$, but when the base station is at coordinate (0, 0), $k = 1 \sim \infty$. So the equation does not suit the situation when the base station is at (0, 0). In addition, fusion rate is not suitable because LEACH assumes it has perfect data fusion.

After obtaining the optimal cluster number, the next step is to choose appropriate nodes as cluster heads which can gather data from intra-cluster nodes, compress data and send them to the base station. There are two kinds of algorithms to choose cluster heads, random probability selection as LEACH and probability iteration selection as HEED. The selection probability for the i^{th} node to be a cluster head node in LEACH [6] is given by

$$P(i) = \begin{cases} \dfrac{k/N}{1 - \dfrac{k}{N} \cdot \left[R \bmod N/k \right]} & i \in G \\ 0 & other \end{cases} \quad (2)$$

where k is the optimal cluster number, G is the set of nodes that have not been selected as cluster heads in the last $R \bmod N/k$ rounds. The distribution of cluster heads and cluster sizes are uneven because the cluster heads are selected randomly. As a result, some nodes may die earlier. In order to solve this problem, LEACH-C and LEACH-F are proposed in [11] and [12], respectively. However, both of them are centralized based approaches, in which the cluster heads are selected by the base station in limited area, and are not suitable for large-scaled networks. HEED protocol [10] takes residual energy and AMRP into consideration. The probability for the i^{th} node to become a cluster head is given by

$$P'(i) = \max\left(\frac{k}{N} \cdot \frac{E_{residual}}{E_{\max}}, p_{\min}\right) \quad (3)$$

where k/N is the rate of the optimal cluster number to the node number, $E_{residual}$ is the estimated node residual energy and $E_{\max}$ is a reference maximum energy (corresponding to a fully charged battery), which is typically identical for all nodes. The $P'(i)$, however, is not allowed to fall below a certain threshold p_{min} (e.g. 10^{-4}). p_{min} is inversely proportional to $E_{\max}$.

HEED obtains well-balanced clusters and energy consumption, longer network lifetime. At the same time it has preferable scalability in data fusion. But it only take the residual energy and AMRP into consideration while choosing the cluster heads. More factors such as initial energy, distance between node and base station, node degree, and average energy consumption et al may be considered.

After the cluster heads have been selected, the cluster heads will broadcast an advertisement message (ADV) to let all the other nodes know the cluster information for the current round. In LEACH, the area occupied by each cluster is approximately M^2/k, the radius for each cluster [11], [12] is

$$R_{LEACH} \approx \sqrt{\frac{M^2}{\pi k}} \quad (4)$$

To avoid "hot spot" problem in equal radius cluster, reference [17] and [18] propose UCS and EEUC model, respectively. In EEUC, the uneven cluster radius is

$$R_{EEUC} = \left[1 - c \cdot \frac{d_{toBS\_MAX} - d_{toBS}}{d_{toBS\_MAX} - d_{toBS\_MIN}} \right] \cdot R_c^0 \quad (5)$$

where d_{toBS_MAX} and d_{toBS_MIN} denote the maximum and minimum distance between the sensor nodes and the base station, respectively, d_{toBS} is the distance between the cluster head node and the base station, c is a constant coefficient between 0 and 1, R_c^0 is the maximum competition radius which is predefined.

EEUC solves "hot spot" problem, but c is determined by experience which has an influence on the cluster radius. In addition, node degree can be taken into consideration to increase its adaptation.

Once receiving ADV, each non-cluster head node determines its cluster for this round by choosing the cluster head that requires the minimum communication energy, based on the received ADV signal strength from each cluster head in LEACH, or minimum AMRP in HEED. Those nodes which can't receive message from the cluster head will become isolated nodes.

After each node having selected the cluster it belongs to, it must inform the cluster head node that it will be a member of the cluster. Each node transmits a join-request message (Join-REQ) to the chosen cluster head. The cluster head node sets up a TDMA schedule and transmits this schedule to the nodes in its cluster. Since the isolated nodes have not received the TDMA schedule, their data will be sent to the base station directly. After the TDMA schedule has been known by all nodes (except the isolated nodes) in the cluster, the set-up phase is completed and the steady-state operation will begin. Once the cluster head receives all the data, it performs data aggregation to enhance the common signal and reduces the uncorrelated noise among the signals. The resultant data are sent to the base station by routing path or directly.

In order to solve the problems mentioned in set-up and steady-state phase, we propose some new algorithms. Firstly, we propose a new method to obtain the optimal cluster number to replace that used in LEACH which does not work when the base station is at (0, 0). The new algorithm takes fusion rate into consideration and is suitable for any situation of the base station. Secondly, we propose a new cluster selection algorithm by taking more factors into consideration, such as residual energy, initial energy, average energy consumption, node degree *et al*. besides, we propose a self-adaptive uneven clustering algorithm which takes node degree into consideration to increase its adaptation. Finally, we give some ideas to deal with isolated nodes to reduce energy consumption.

III. Calculation of the Optimal Cluster Number

Clustering protocols, such as LEACH and HEED, all assume that they have perfect data fusion, which is not practical. Furthermore, (1) used to obtain the optimal cluster number does not work properly when the base station is located at or close to coordinate (0, 0).

Here we use the radio model shown in Fig.1. Assume there are N nodes distributed uniformly in an $M \times M$ region. There are k clusters, each has N/k nodes (one cluster head node and $N/k-1$ non-cluster head nodes). Each cluster head dissipates energy to receive data from the non-cluster head nodes, aggregate the data and transmit them to the base station. Since the base station is far from the nodes, we can assume that the energy consumption follows the multipath fading channel model (d^4 power loss). Therefore, the energy consumption of a cluster head in a round can be obtained as

$$E_{CH} = g\left[\left(\frac{N}{k}-1\right)E_{elec} + \frac{N}{k}E_{DA} + f_{agg}\left[\varepsilon_{mp}E\left(d_{toBS}^4\right)+E_{elec}\right]\right] \tag{6}$$

where E_{DA} is the energy consumption for data fusion per bit, d_{toBS} is the distance between node and the base station, $E\left(d_{toBS}{}^4\right)$ is the expectation of $d_{toBS}{}^4$, f_{agg} is fusion rate. In order to compare the performance of LEACH and the proposed algorithm in same condition, we can define $f_{agg}=1$. The value of fusion rate can vary from 1 to N in different conditions.

Each non-cluster head node only needs to transmit its data to the cluster head once during a frame. Assume the distance to the cluster head is small, the energy consumption follows the free space channel model. The radio expends

$$E_{NCH} = g\left[E_{elec} + \varepsilon_{fs}E\left(d_{toCH}^2\right)\right] \tag{7}$$

where d_{toCH} is the distance between the node and the cluster head, $E(d_{toCH}{}^2)$ is the expectation of $d_{toCH}{}^2$. In each frame, all the nodes expend:

$$\begin{aligned} E_{total} &= k\left[E_{CH} + \left(\frac{N}{k}-1\right)E_{NCH}\right] \\ &= gk\left\{\left(\frac{N}{k}-1\right)E_{elec} + \frac{N}{k}E_{DA} + \right. \\ &\quad f_{agg}\left[\varepsilon_{mp}E\left(d_{toBS}^4\right)+E_{elec}\right] + \\ &\quad \left.\left(\frac{N}{k}-1\right)[E_{elec}+\varepsilon_{fs}d_{toCH}{}^2]\right\} \\ &\approx g\left\{NE_{elec} + NE_{DA} + kf_{agg}\left[\varepsilon_{mp}E\left(d_{toBS}^4\right)+E_{elec}\right] + \right. \\ &\quad \left. N\left[E_{elec}+\varepsilon_{fs}\frac{M^2}{2\pi k}\right]\right\} \end{aligned} \tag{8}$$

By making the derivative of the function E_{total} equal to 0, the optimal number of k can be obtained as

$$k = \left\lceil \sqrt{\frac{\varepsilon_{fs}\cdot M^2\cdot N}{2\pi f_{agg}\cdot\left[\varepsilon_{mp}\cdot E\left(d_{toBS}{}^4\right)+E_{elec}\right]}} \right\rceil \tag{9}$$

where $\lceil a \rceil$ denotes the smallest integer which is greater than or equal to the argument a.

IV. Cluster Head Selection

HEED only considers residual energy and AMRP during the cluster head selection. Some other factors should be considered, such as energy consumption speed, node degree, distance between the node and the base station.

The node with fast energy consumption should not be considered as the cluster head. The node with larger node degree or short distance to the base station should have higher probability to be the cluster head. Considering all the factors above, we introduce energy consumption factor $E_{avecons}/E_{consume}$ [20], factor of node degree N_{nbr}/N_{alive} [10], factor of distance between the node and the base station $(d_{toBS\_MAX}-d_{toBS})/(d_{toBS\_MAX}-d_{toBS\_MIN})$ [18] as the factors determining the probability of cluster head selection

$$P_{NEW} = \max\left(\frac{k}{N}\cdot\left(\frac{E_{residual}}{E_{\max}}\cdot\frac{E_{avecons}}{E_{consume}}\right)\cdot(\alpha\cdot\frac{N_{nbr}}{N_{alive}} + (1-\alpha)\cdot\frac{d_{toBS\_MAX}-d_{toBS}}{d_{toBS\_MAX}-d_{toBS\_MIN}}), p_{\min}\right) \tag{10}$$

where k/N is the rate of the optimal number of cluster head nodes to the number of nodes, α is a constant coefficient between 0 and 1. Considering the influence of the two factors, the factor of node degree has more effect than the factor of distance between the node and the base station, so we can assume $\alpha = 0.6$ at first. In part VII we will discuss the value and look for the best value. $E_{residual}$ is the estimated node residual energy, $E_{\max}$ is a reference maximum energy, $E_{consume}$ is the energy consumption of the i^{th} node, $E_{avecons}$ is the average energy consumption of the whole network during the last round of data transmission, N_{nbr} is the number of neighbor nodes, $N_{a\ l\ i}$ is the number of alive nodes, p_{min} is a certain threshold (e.g., 10^{-4}).

In round R, the larger energy consumption is, the smaller $E_{avecons}/E_{consume}$ is and the lower probability to be a cluster head in the next round. On the contrary, the probability to be a cluster head in next round will be larger. At the same time, the larger N_{nbr}/N_{alive} is, the larger probability to be a cluster head.

In addition, the node should have larger probability to be cluster head when its distance between the node and the base station is short. We use $(d_{toBS\_MAX}-d_{toBS})/(d_{toBS\_MAX}-d_{toBS\_MIN})$ to solve the problem. In (10), α is used to make P_{prob} more reasonable.

V. Uneven Clustering Algorithm

Based on (5), node degree is introduced to improve the adaptation. The new uneven cluster radius is

$$R_{NEW}=\left(1+\frac{d_{toBS}-E\left(d_{toBS}\right)}{\beta\cdot(d_{toBS\_MAX}-d_{toBS\_MIN})}\right)\cdot\left(1-\frac{N_{nbr}}{\gamma\cdot N_{alive}\cdot R_{LEACH}}\right)\cdot R_{LEACH} \quad (11)$$

where $\beta\in(0,1)$, $\gamma\in(1,5)$,they are constant coefficients considering the influence of the factors, we can assume $\gamma=2$ and $\beta=0.7$ at first. we will discuss the value and find the best value in Section VII, R_{LEACH} can be calculated by (4) and will be demonstrated in Section VII.

In (11), cluster radius can adjust according to the distance between the node and the base station. When d_{toBS} is larger, the cluster radius is larger too. On the contrary, the radius will be smaller. In addition, the node degree has influence on the radius. The larger the node degree is, the smaller the cluster radius is.

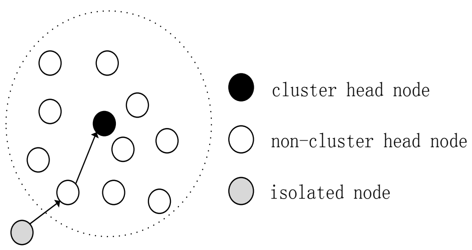

Fig. 2. Isolated node join near cluster

VI. Isolated Node Dealing Algorithm

If the nodes have not received advertisement message from cluster heads, they will become isolated nodes. In LEACH and HEED, the isolated nodes will become cluster heads and send data to the base station directly which will cost a lot of energy. Some algorithms that let isolated nodes join near cluster according to certain rules is shown in Fig. 2.

In order to find suitable cluster, we introduce the cost for joining near cluster which is directly proportional to energy consumption for sending data to the base station by multi-hop.

It will consume energy when the data are sent to next hop node, then the node receives the data and transmits them to the cluster head.

$$\begin{aligned}E_{v-u-CH}&=g\left[E_{elec}+\varepsilon_{fs}d_{vtou}^{2}\right]+2gE_{elec}+g\left[E_{elec}+\varepsilon_{fs}d_{utoCH}^{2}\right]\\&=g\varepsilon_{fs}\left[d_{vtou}^{2}+d_{utoCH}^{2}\right]+4gE_{elec}\end{aligned} \quad (12)$$

where d_{vtou} is the distance between the isolated node v and the next hop node u, d_{utoCH} is the distance between node u and the cluster head.

Cluster head consumes energy in receiving data from node u and sending them to the base station.

$$E_{CH-BS}=g\left[E_{elec}+f_{agg}\varepsilon_{mp}d_{toBS}^{4}+E_{DA}\right] \quad (13)$$

where E_{DA} is the energy consumption for data fusion per bit. Then the total energy consumption is

$$\begin{aligned}E_{v-u-CH-BS}&=g\varepsilon_{fs}\left[d_{vtou}^{2}+d_{utoCH}^{2}\right]+4gE_{elec}+\\&\quad g\left[E_{elec}+E_{DA}+f_{agg}\varepsilon_{mp}d_{toBS}^{4}\right]\\&=g\left[\varepsilon_{fs}\left(d_{vtou}^{2}+d_{utoCH}^{2}\right)+5E_{elec}+\right.\\&\quad\left.E_{DA}+f_{agg}\varepsilon_{mp}d_{toBS}^{4}\right]\\&=g\left[\varepsilon_{fs}\left(d_{vtou}^{2}+d_{utoCH}^{2}\right)+f_{agg}\varepsilon_{mp}d_{toBS}^{4}\right]+\\&\quad g\left[5E_{elec}+E_{DA}\right]\end{aligned} \quad (14)$$

No matter which cluster head is selected, $g\cdot\left[5E_{elec}+E_{DA}\right]$ is always the same and can be removed. So

$$E_{v-u-CH-BS}\propto\left[\varepsilon_{fs}\left(d_{vtou}^{2}+d_{utoCH}^{2}\right)+f_{agg}\varepsilon_{mp}d_{toBS}^{4}\right] \quad (15)$$

Assume the cost to join near cluster for an isolated node is

$$Cost_{v-u-CH-BS}=\left[\varepsilon_{fs}\left(d_{vtou}^{2}+d_{utoCH}^{2}\right)+f_{agg}\varepsilon_{mp}d_{toBS}^{4}\right] \quad (16)$$

Now the isolated node can determine whether to join near cluster according to the cost of joining near cluster $Cost_{v-u-CH-BS}$.

VII. Simulation and Numerical Analysis

In NS2, we distribute randomly 100 nodes (in equations (1)-(3), (6), (8)-(10), $N=100$) in the area of $100\times100m^{2}$ (in (1), (4), (8) and (9), $M=100m$). The initial energy of all the sensor nodes are equal (in (3) and (10), $E_{\max}=4J$). In (5), $c=0.5$ and $R_{c}^{0}=50$. In (10), the optimal cluster number $k=4$, In (1), (7)-(9), (12)-(16), $f_{agg}=1$, $\varepsilon_{fs}=10\text{ pJ/bit/m}^{2}$, $\varepsilon_{mp}=0.0013\text{ pJ/bit/m}^{4}$, $E_{elec}=50\text{ nJ/bit}$, $E_{DA}=5\text{ nJ/bit}$. In (10), $\alpha=0.6$. In (3) and (10), according to the reference [10] and R_c^0, we set $p_{\min} = 0.0005$, $R_{LEACH} = 50$ [6,10-12,18,19] .

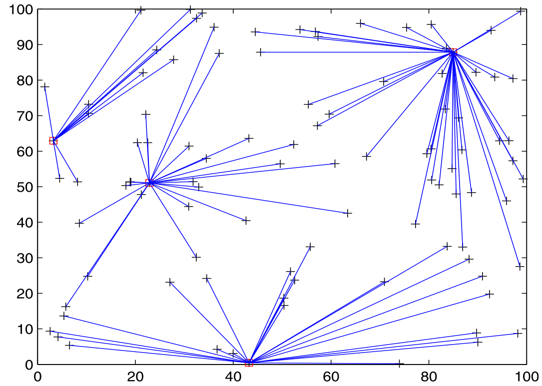

Fig. 3. The network topology of LEACH

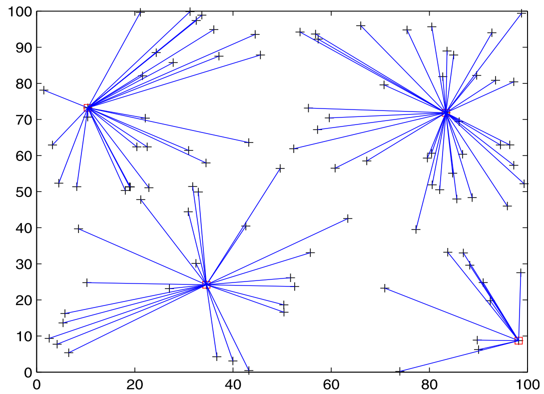

Fig. 4. The network topology of HEED

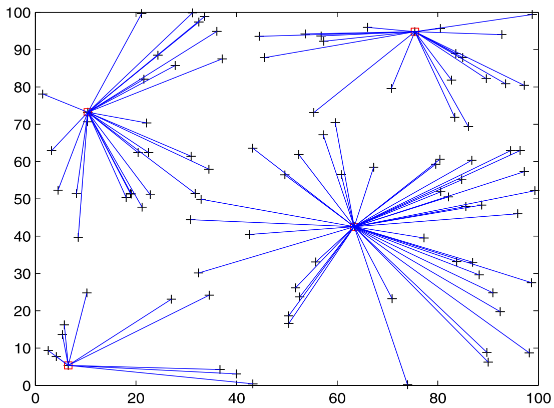

Fig. 5. The network topology of EEUC

The network topology of LEACH, HEED, EEUC and the proposed new algorithm are shown in Fig. 3, Fig. 4 Fig. 5 and Fig. 6, respectively. "□" represents cluster head nodes, "+" represents non-cluster head nodes, "Δ" represents base station. The distribution of the cluster heads and cluster sizes is uneven in Fig. 3 because the cluster heads are selected randomly in LEACH. The distribution of cluster heads and cluster sizes is uniform in Fig. 4 and Fig. 5. To avoid "hot spot" problem in even cluster routing algorithm, cluster sizes are smaller nearing the base station in Fig. 6. Non-cluster head nodes keep dormant until it is time to transmit data to the cluster heads continually using a combination TDMA/CDMA schedule.

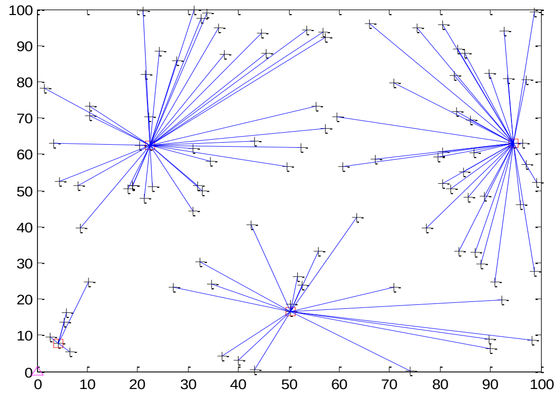

Fig. 6. The network topology of proposed new uneven clustering routing algorithm

Fig. 7 shows the influence of the cluster number on the number of alive nodes in LEACH. The simulation result shows the time that the first node dies is about 250th round when $k = 4$ which is prolonged by about 31.4% comparing with that when $k = 5$ (the optimal proportion of cluster head is 5% in LEACH). The time that the network no longer provides acceptable quality results is about 347th round when $k = 4$ which is prolonged by about 17.7% comparing with that when $k = 5$. Therefore, we obtain the maximum network lifetime when the cluster number $k = 4$. The simulation result is consistent with that obtained by (9).

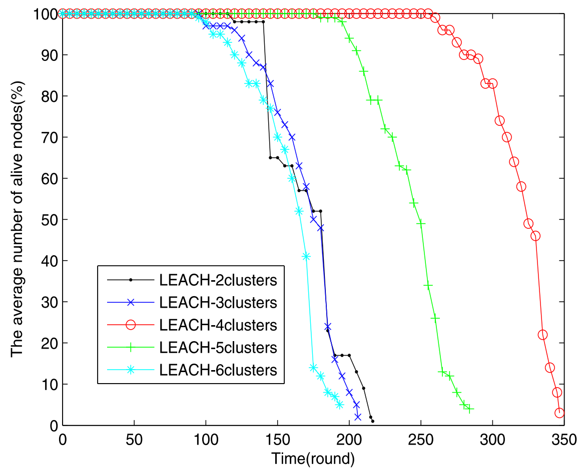

Fig. 7. The average number of alive nodes by changing the cluster number in LEACH

Fig. 8 shows the influence of the cluster number on the total energy consumption in LEACH. The simulation result shows that, when $k = 2$, the total energy consumption is less than that in other situations in the front 50 rounds. But, when $k = 4$, the energy consumption grows more slowly than that in other situations after 50th round. therefore, when the cluster number $k = 4$, the maximum network lifetime is obtained. The simulation result is consistent with that obtained by (9).

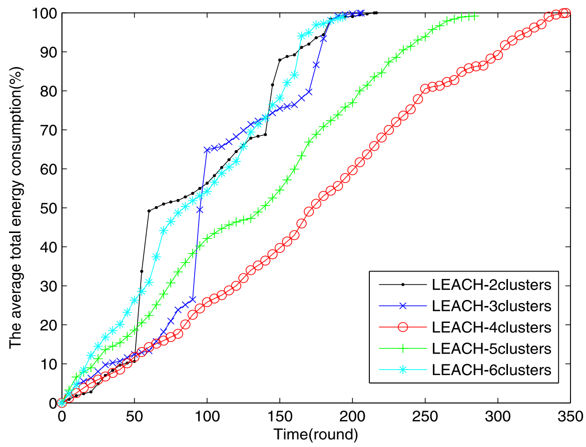

Fig. 8. The average total energy consumption by changing the cluster number in LEACH

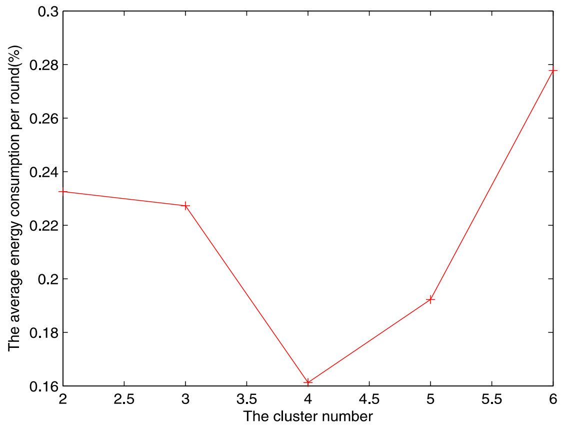

Fig. 9. The average energy consumption per round in LEACH

Fig. 9 shows the average energy consumption per round in LEACH as the cluster number varies from 2 to 6. It shows that, when the cluster number $k=4$, the average energy consumption is minimal, which is consistent with our algorithm.

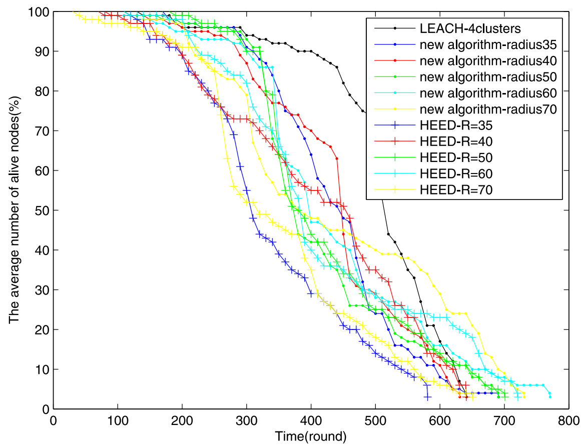

Fig. 10. The average number of alive nodes by changing cluster radius in HEED and the new cluster head selection algorithm

Fig. 10 shows the influence of the cluster radius on the number of alive nodes in HEED and new cluster head selection algorithm. The horizontal axis represents time (round). The vertical axis represents the number of alive nodes ($\alpha=0.6$ in the new cluster head selection algorithm). After 400th round, the number of alive nodes in the new algorithm is larger than that in other situations when $R_{LEACH}=50$ or $R_{LEACH}=60$. When $R_{LEACH}=60$, the time that the network no longer provides acceptable quality results is about 770th round in the new algorithm, which is prolonged about 20.1% than that in LEACH ($k=4$). Therefore, the maximum network lifetime is obtained in HEED and the new algorithm when $R_{LEACH}=60$.

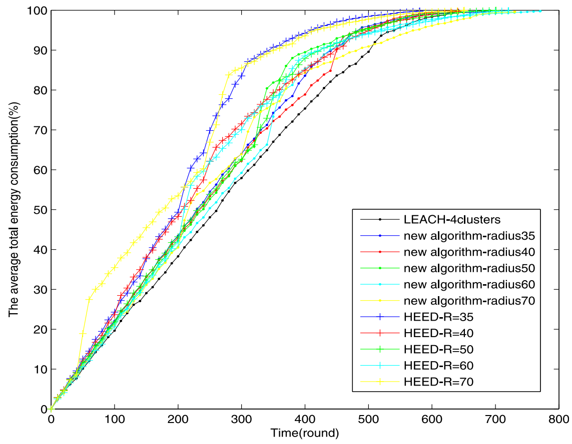

Fig. 11. The average total energy consumption by changing cluster radius in HEED and the new cluster head selection algorithm

Fig. 11 shows the influence of cluster radius on the total energy consumption in HEED and the new cluster head selection algorithm. The horizontal axis represents time (round). The vertical axis represents the average total energy consumption ($\alpha=0.6$ in the new cluster head selection algorithm). The simulation result shows that maximum network lifetime is obtained when $R_{LEACH}=60$.

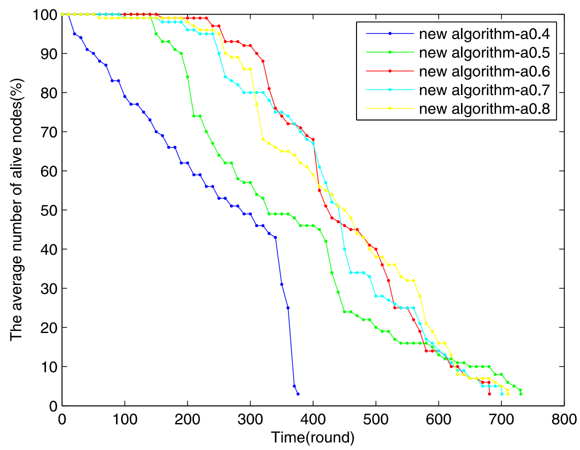

Fig. 12. The average number of alive nodes by changing the α value in the new cluster head selection algorithm

Fig. 12 shows the influence of α value on the number of alive nodes in the new cluster head selection algorithm. The simulation result shows that, when $\alpha=0.5$, maximum network lifetime is obtained. The time that the network no longer provides acceptable quality results is about 730th round when $\alpha=0.5$, which is prolonged by about 7.4% than that when $\alpha=0.6$.

Fig. 13 shows the influence of α value on the total energy consumption in the new cluster head selection algorithm. The simulation result shows that maximum network lifetime is obtained when $\alpha = 0.5$.

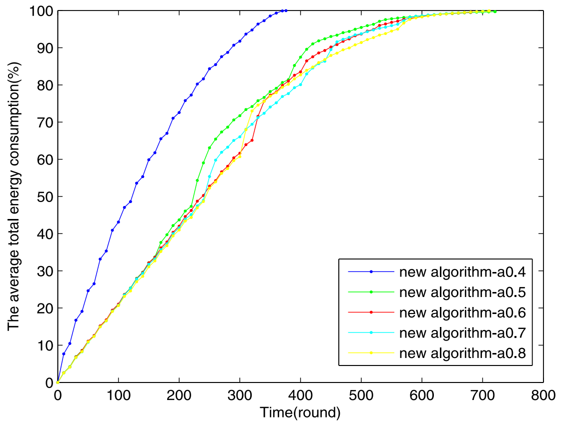

Fig. 13. The average total energy consumption by changing the α value in the new cluster head selection algorithm

Fig. 14 shows the influence on the number of alive nodes when setting $c = 0.5$ and $R_c^0 = 50$ in EEUC and changing the value of β and γ in the new uneven clustering algorithm. The horizontal axis represents time (round). The vertical axis represents the number of alive nodes ($\gamma = 2$ in the new uneven clustering algorithm). The time that the network no longer provides acceptable quality results is about 780th round in the new cluster head selection algorithm, which is prolonged by 13% than that in EEUC. The simulation result shows that maximum network lifetime is obtained when $\beta = 0.7$.

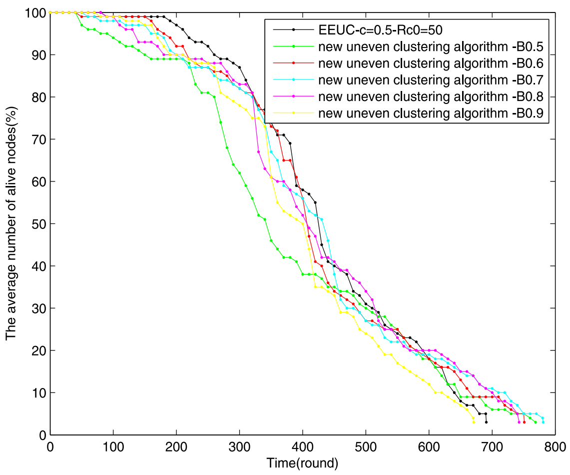

Fig. 14. The average number of alive nodes by setting $c = 0.5$ and $R_c^0 = 50$ in EEUC and changing the value of β and γ in the new uneven clustering algorithm

Fig. 15 shows the influence on the total energy consumption when setting $c = 0.5$ and $R_c^0 = 50$ in EEUC and changing the value of β and γ in the new uneven clustering algorithm. The horizontal axis represents time (round). The vertical axis represents the total energy consumption ($\gamma = 2$ in the new uneven clustering algorithm). The simulation result shows that maximum network lifetime is obtained when $\beta = 0.7$.

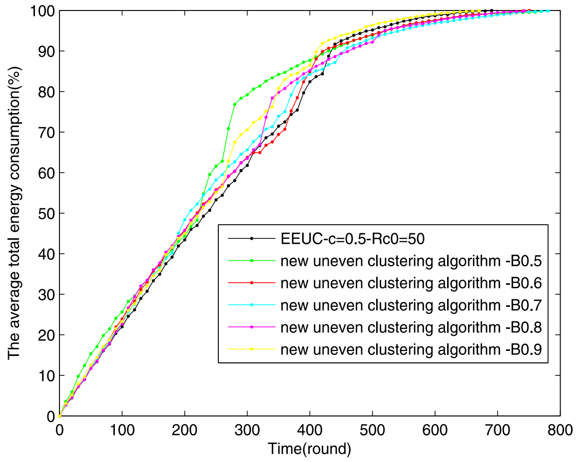

Fig. 15. The average total energy consumption by setting $c = 0.5$ and $R_c^0 = 50$ in EEUC and changing the value of β and γ in the new uneven clustering algorithm

Fig. 16 shows the influence on the number of alive nodes when setting $c = 0.5$ and $R_c^0 = 50$ in EEUC and changing the γ value in the new uneven clustering algorithm. The horizontal axis represents time (round). The vertical axis represents the number of alive nodes ($\beta = 0.7$ in the new uneven clustering algorithm). The time that the network no longer provides acceptable quality results is about 800th round when $\gamma = 2$,which is prolonged by about 15.9% than that in EEUC. Therefore, the maximum network lifetime is obtained when $\gamma = 2$.

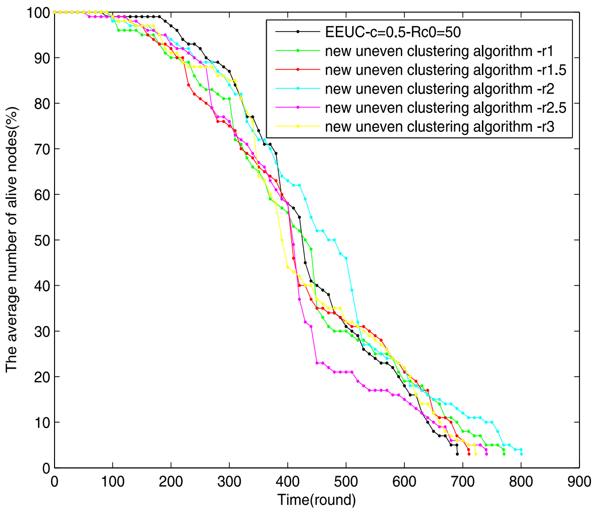

Fig. 16. The average number of alive nodes by changing the γ value in the new uneven clustering algorithm

Fig. 17 shows the influence of the γ value on the total energy consumption in the new algorithm. The horizontal axis represents time (round). The vertical axis represents the total energy consumption ($\beta = 0.7$ in the new uneven clustering algorithm). The simulation result shows that maximum network lifetime is obtained when $\gamma = 2$.

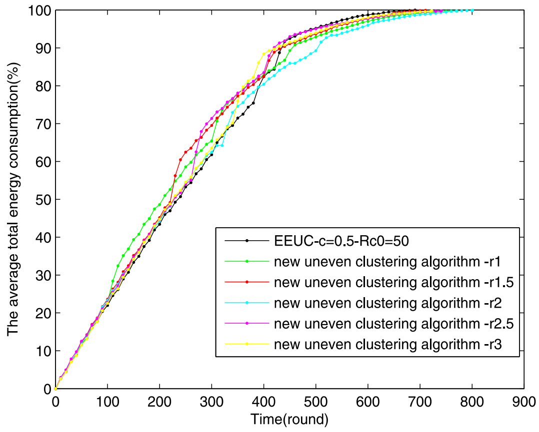

Fig. 17. The average total energy consumption by changing the γ value in the new algorithm

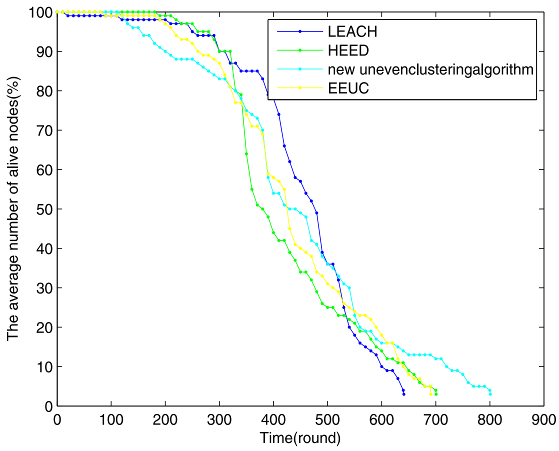

Fig. 18. The average number of alive nodes for different algorithms

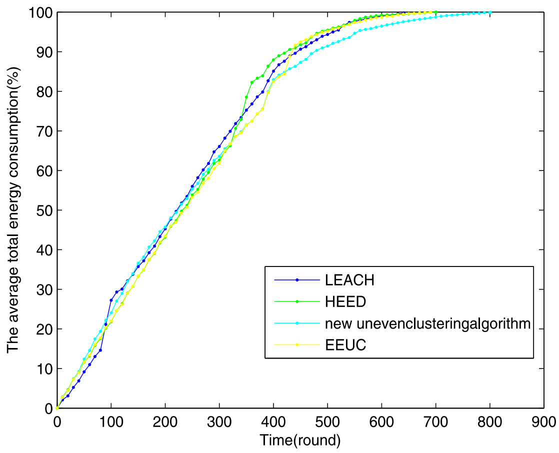

Fig. 19. The average total energy consumption for different algorithms

Fig. 18 shows the average number of alive nodes for LEACH ($k=4$), HEED ($R_c^0=50$), EEUC, and the new uneven clustering routing algorithm ($\alpha=0.5$, $\beta=0.7$ and $\gamma=2$). The simulation result shows the first node dies is about 110th round in the new uneven clustering algorithm , which is prolonged by about 60% than that in LEACH ($k=4$). The time that the network no longer provides acceptable quality results is about 800th round in the new uneven clustering algorithm, which is prolonged by about 24.8% than that in LEACH ($k=4$). So the new algorithm has the superiority in terms of network lifetime and the number of alive nodes.

Fig. 19 shows the average total energy consumption for LEACH ($k=4$), HEED ($R_c^0=50$), EEUC and the new uneven clustering algorithm ($\alpha=0.5$). The simulation result shows that the total energy consumption in the new uneven clustering algorithm grows more slowly than that in other situations. It demonstrates that the new algorithm has the superiority in terms of network lifetime and the total energy consumption over other approaches.

VIII. Conclusion

In this paper, we propose an uneven clustering routing algorithm based on optimal clustering for wireless sensor networks. Some new methods are proposed to obtain the optimal cluster number, select cluster heads, calculate the cluster radius, and deal with isolated nodes. Firstly, to obtain the optimal cluster number, we introduce data fusion rate and location adaptability to the base station. The experiment results show the optimal cluster number is consistent with that obtained by the new calculation method. Secondly, we propose a new cluster head selection algorithm by considering residual energy, initial energy, average energy consumption, and node degree. The algorithm obtains well-balanced clusters and energy consumption, longer network lifetime than traditional algorithms. Thirdly, in order to solve “hot spot” problem and “isolated nodes problem”, we propose a self-adaptive uneven clustering algorithm and a solution to deal with isolated nodes, which can reduce and balance energy consumption significantly. The experiment results show the time that the first node dies is delayed by about 60% than that in LEACH. The time that the network no longer provides acceptable quality results in the new algorithm is prolonged by about 24.8% than that in LEACH. The number of data received at the base station in the new algorithm is less than that in LEACH and almost same as that in HEED. The total energy consumption in the new algorithm grows more slowly than that in other situations. It is effective in reducing the energy consumption and prolonging the network lifetime.

References

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “Wireless sensor networks: A survey,” *Computer Networks*, vol. 38, no. 4, pp. 393-422, March 2002.

[2] T. Rault, A. Bouabdallah, and Y. Challal, “Energy efficiency in wireless sensor networks: A top-down survey,” *Computer Networks*, vol. 67, pp. 104-122, July 2014.

[3] S. Lonare and G. Wahane, “A survey on energy efficient routing protocols in wireless sensor network,” in *Proc. 4th International Conference on Computing, Communications and Networking Technologies*, Tiruchengode, India, July 2013, pp. 1-5.

[4] A. Boonsongsrikul, S. Kocijancic, and S. Suppharangsan, "Effective energy consumption on wireless sensor networks survey and challenges," in *Proc. 36th International Convention on Information & Communication Technology Electronics & Microelectronics*, Opatija, Croatia, May 2013, pp. 469-473.

[5] M. Magno, S. Marinkovic, B. Srbinovski, and E. M. Popovici, "Wake-Up radio receiver based power minimization techniques for wireless sensor networks: A review," *Microelectronics Journal*, vol. 45, pp. 1627-1633, December 2014.

[6] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient communication protocol for wireless microsensor networks," in *Proc. 33rd Annual Hawaii International Conference on System Sciences*, Hawaii, USA, January 2000, pp. 1-10.

[7] S. Toklu and O. A. Erdem, "BSC-MAC: Energy efficiency in wireless sensor networks with base station control," *Computer Networks*, vol. 59, pp. 91-100, February 2014.

[8] H. Li, K. Li, W. Qu, and I. Stojmenovic, "Secure and energy-efficient data aggregation with malicious aggregator identification in wireless sensor networks," *Future Generation Computer Systems*, vol. 37, pp. 108-116, July 2014.

[9] A. More and V. Raisinghani, "Discharge curve backoff sleep protocol for energy efficient coverage in wireless sensor networks," *Procedia Computer Science*, vol. 57, pp. 1131-1139, July 2015.

[10] O. Younis and S. Fahmy, "HEED: A hybrid energy-efficient distributed clustering approach for ad-hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 3, no. 4, pp. 366-379, October 2004.

[11] W. B. Heinzelman, "Application-Specific protocol architectures for wireless networks," Ph.D. dissertation, Dept. Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, 2000.

[12] W. B. Heinzelman, A. P. Chandrakasan, and H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks," *IEEE Transactions on Wireless Communications*, vol. 1, no. 4, pp. 660-670, October 2002.

[13] R. Sharma, N. Mishra, and S. Srivastava, "A proposed energy efficient Distance Based Cluster Head (DBCH) algorithm: An improvement over LEACH," *Procedia Computer Science*, vol. 57, pp. 807-814, July 2015.

[14] S. H. Gajjar, K. S. Dasgupta, S. N. Pradhan, and K. M. Vala, "Lifetime Improvement of LEACH protocol for wireless sensor network," in *Proc. 3rd Nirma University International Conference on Engineering*, Ahmedabad India, December 2012, pp. 1-6.

[15] W. Akkari, B. Bouhdid, and A. Belghith, "LEATCH: Low energy adaptive tier clustering hierarchy," *Procedia Computer Science*, vol. 52, pp. 365-372, May 2015.

[16] H. Taheri, P. Neamatollahi, M. Naghibzadeh, and M. H. Yaghmaee, "Improving on HEED protocol of wireless sensor networks using non probabilistic approach and fuzzy logic (HEED-NPF)," in *Proc. 5th International Symposium on Telecommunications*, Tehran, Iran, December 2010, pp. 193-198.

[17] S. Soro and W. B. Heinzelman, "Prolonging the lifetime of wireless sensor networks via unequal clustering," in *Proc. 19th International Parallel & Distributed Processing Symposium*, Denver, USA, 2005, pp. 236-244.

[18] G. H. Chen, C. F. Li, M. Ye, and J. Wu, "An unequal cluster-based routing protocol in wireless sensor networks," *Wireless Networks*, vol. 15, no. 2, pp. 193-207, February 2009.

[19] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Applied Soft Computing*, vol. 13, pp. 1741-1749, April 2013

[20] J. Li, X. Jiang, and I. T. Lu, "Energy balance routing algorithm based on virtual MIMO scheme for wireless sensor networks," *Journal of Sensors*, 2014.

Jianpo Li was born in China in 1980. He received his B.S., M.S., and Ph.D. from the Department of Communication Engineering, Jilin University, China, in 2002, 2005, and 2008, respectively. In 2008, he joined the Department of Information Engineering, Northeast Dianli University, where he is currently an associate professor. His research interests are wireless sensor networks and intelligent signal processing.

Junyuan Huo was born in China in 1989. He received his B.S. from the Department of Information Engineering, Shenyang Institute of Engineering. He is currently pursuing his MS in Northeast Dianli University. His main research interest is wireless sensor network.