Journal of Information Hiding and Multimedia Signal Processing ©2019 ISSN 2073-4212
Ubiquitous International Volume **10**, Number **1**, January **2019**

An Uneven Clustering Routing Protocol based on Improved K-means Algorithm for Wireless Sensor Network in Coal-mine

Xixi Dong
School of Electronic Science and Engineering
Nanjing University Of Posts And Telecommunications
9 Wenyuan Road, Qixia District, Nanjing 210000, P.R.China
15396758786@163.com

Yun Zhang and Shujuan Yu
School of Electronic Science and Engineering
Nanjing University Of Posts And Telecommunications
9 Wenyuan Road, Qixia District, Nanjing 210000, P.R.China
y021001@njupt.edu.cn;yusj@njupt.edu.cn

Received January 2018; revised August 2018

Abstract. *Aiming at the long band-type topology of Wireless Sensor Network in coal-mine and the uneven energy consumption between the nodes, this paper proposes an uneven clustering routing protocol based on improved K-means algorithm. According to the distance of the node to the base station, the initial centroid selection procedure in K-means algorithm is improved to construct clusters of different size, and the clusters close to the base station have smaller size than those far away from the base station. The reselection of cluster heads is based on residual energy and geographical location, so that the nodes of low energy will not be selected as cluster heads. In the data transmission phase, according to the approximate linear distribution of cluster heads, a multi-hop routing algorithm between cluster heads is designed, and the energy balance between forwarding nodes is considered. The simulation results show that the routing protocol can effectively balance network node energy consumption and achieve an obvious improvement on the network lifetime.*
Keywords: Wireless Sensor Network; coal-mine; K-means algorithm; uneven clustering; initial centroid

1. **Introduction.** In recent years, the frequent occurrence of major accidents in coal-mine has caused huge losses to state property and citizen's life. Mine safety production and emergency rescue has become the focus of the society. Due to the advantages of convenient deployment, low cost, flexible structure and strong survivability, Wireless Sensor Network (WSN) is particularly suitable for comprehensive monitoring of coal-mine [1-3]. Because the topology of WSN in coal-mine is long band-type, the direction of node information flow is mainly unidirectional from the excavation to the exit of the roadway. The routing information and the amount of data transmitted by the nodes in different areas are greatly different [4-6], resulting in a serious problem of uneven energy consumption, data redundancy and data latency.

At present, many scholars have studied the clustering routing protocol. Heinzelman proposed LEACH [7], a classical low energy adaptive clustering protocol, which allocates the extra energy consumption of cluster heads to nodes in the network by periodically changing nodes into cluster heads. There is a serious "hot spot" effect [8] due to the single-hop mode and uneven energy consumption between clusters. The PEGASIS [9] protocol proposed by Lindsey periodically organize nodes into a chain, and the data is transmitted to the base station after fused on this chain. The essence of PEGASIS protocol is to allocate the energy consumption of cluster heads by reconstructing clusters. Hybrid energy-efficient distributed clustering (HEED) [10] proposed by Younis et al thought that in order to achieve inter-cluster load balancing in a local area, a node should select clusters with the lowest power consumption to join. But the head of the election campaign using iterative high-energy way. Anundita Ray et al [11] studied an energy-efficient clustering protocol EECPK-means based on the improved K-means algorithm, which uses a midpoint algorithm to improve the initial centroid selection process and finally balances the cluster head load and prolongs the network lifetime. Li [12] proposed an Energy-Efficient Uneven Clustering algorithm (EEUC), a distributed uneven clustering algorithm. The distance between the candidate cluster heads and the base station is taken as the parameter to calculate the uneven competition radius. Thus the cluster heads close to the base station are smaller in size. The above three protocols of clustering routing protocols are designed for the general scene without considering the topology of linear networks. In[13], the authors consider that the sensors equipped with solar-powered equipment no longer have the limited battery life problem, but the renewable energy do not apply to the environment of coal-mine.

This work proposes an uneven clustering routing protocol based on improved K-means algorithm (UCRPK-means). Based on the analysis of the long band-type network topology in coal-mine, the initial centroid selection procedure in K-means algorithm is modified to generate uneven clusters. The initial cluster head (CH) nodes are which closest to the centroid, and then dynamically replaced according to the remaining energy and geographic location. The multi-hop routing algorithm among clusters is designed according to the approximate linear distribution of cluster heads in the data transmission phase. The simulation results show that the routing protocol can effectively balance network node energy consumption and achieve an obvious improvement on the network lifetime.

2. Network and Radio Energy Consumption Model.

2.1. Network model.

For the roadway characteristics in coal-mine, the evaluation of the proposed protocol is performed on a WSN with N nodes deployed randomly on a given geographical area ($M1 \times M2$, $M1 >> M2$). This network model is assumed[14]:

(1) All nodes and the base station (BS) are assumed stationary when they are deployed in the environment.

(2) The BS is located outside the area of sensor nodes (exit of roadway), its computational power and energy is not limited.

(3) All nodes are homogeneous, the nodes can be aware of their remaining energy and have capability of data fusion.

(4) According to the distance from the receiver, the node can freely adjust its transmission power to save energy consumption.

(5) Knowing each other's transmission power, the node can calculate the sender's proximity distance to himself based on the strength of the received signal (RSSI).

2.2. Radio energy consumption model.

In this paper, we use the same radio energy consumption model as in [7]. Therefore, the energy consumption for transmitting a l-bit message to a distance is given by Eq.(1).

$$E_{Tx}(l,d) = \begin{cases} lE_{elec} + l\varepsilon_{fs}d^2 & if \quad d < d_0 \\ lE_{elec} + l\varepsilon_{mp}d^4 & if \quad d \geq d_0 \end{cases} \tag{1}$$

Here E_{elec} represents the energy consumption in the electronic system for sending or receiving a bit, ε_{fs} and ε_{mp} is the energy consumption of the free-space model amplifier and the energy consumption of the multi-channel attenuating amplifier, respectively. The distance threshold d_0 is calculated as

$$d_0 = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}} \tag{2}$$

The energy consumption of receiving a l-bit message is calculated by Eq.(3)

$$E_{Rx}(l) = lE_{elec} \tag{3}$$

Data fusion also consumes a certain amount of energy, E_{DA} represents the energy consumed by the fusion 1-bit data. We assume that the data collected by neighboring nodes have high redundancy, and the CHs can fuse the data of its members into a fixed-length data packet.

3. **Proposed UCRPK-means Protocol.** In the network deployment phase, the BS broadcasts a signal to the network with a given transmitted power. After receiving this signal, each sensor node calculates its approximate distance to the base station according to the intensity of the receiving signal. This distance is not only helpful for sensor nodes to select the appropriate transmission power when transmitting data to base stations, but also one of the necessary information for constructing clusters of different size. Fig.1 shows the architectural organization of the protocol proposed in this paper. The circle of unequal size indicates uneven cluster size, and the closer to the base station, the smaller it is. The bold lines with arrows indicate the multi-hop data transfer between CHs. As shown in Fig.2, the UCRPK-means protocol collects data by the cycle, and once the uneven clusters formed, the unevenly clusters will no longer be regrouped by the round. Only dynamic reselection of cluster head and data transmission is performed in each round.

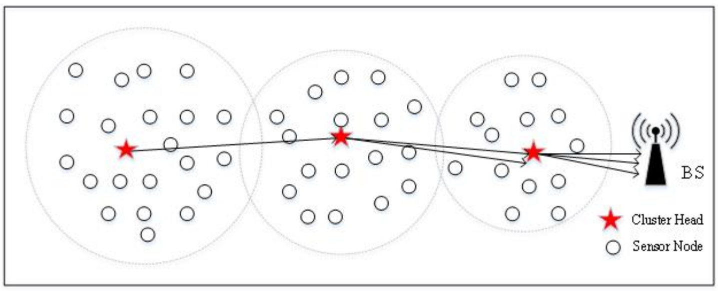


FIGURE 1. Architectural organization of proposed protocol.

3.1. **Cluster formation and initial cluster head selection.** In the long band-type roadway of coal-mine, the cluster is adjacent and linearly distributed. As the large amount of data transmitted by the cluster head near the base station, cluster heads will exhaust energy prematurely and fail. Therefore, uneven clusters are constructed so that the size of the cluster near the BS is less than the cluster of far away from the BS, and the energy consumption of the CHs can be equalized. The value of clustering number in K-means

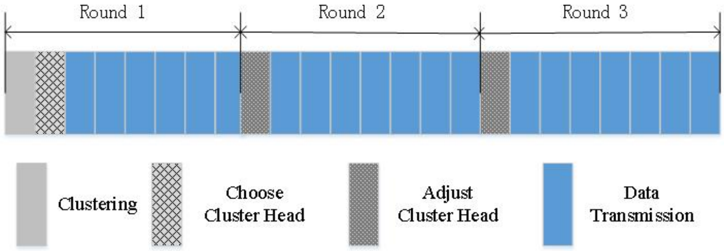


FIGURE 2. Round schematic of network operation.

algorithm is quite important. In this paper, the optimal number of clusters[11] is given by:

$$K_{opt} = \frac{\sqrt{N}}{\sqrt{2\pi}}\sqrt{\frac{\varepsilon_{fs}}{\varepsilon_{mp}}}\frac{\sqrt{S}}{d_{toBS}^2} \tag{4}$$

Where d_{toBS} is the average distance from all sensor nodes to the BS, and S is the area of the long band-type area under the coal-mine environment. Table 1 describes the steps of cluster formation and initial cluster head selection.

3.2. **Dynamic reselection of cluster head.** In order to ensure that the network works properly, the remaining energy of the current CH needs to be checked in each round of operation. If the residual energy of the CH is lower than the threshold energy, the node with the minimum ID and larger than the ID of the current CH node is selected as the new CH. The new CH sends a broadcast message to the member node in the cluster, informing the member node in the cluster of the node ID selected as the CH in the next round. Fig.3 shows the sorting of the sensor nodes with the ID number. The ID number of the node indicates the order when it is selected as CH. Therefore, ID number plays an important role in the reselection of CHs. This phase is described clearly in Table 2.

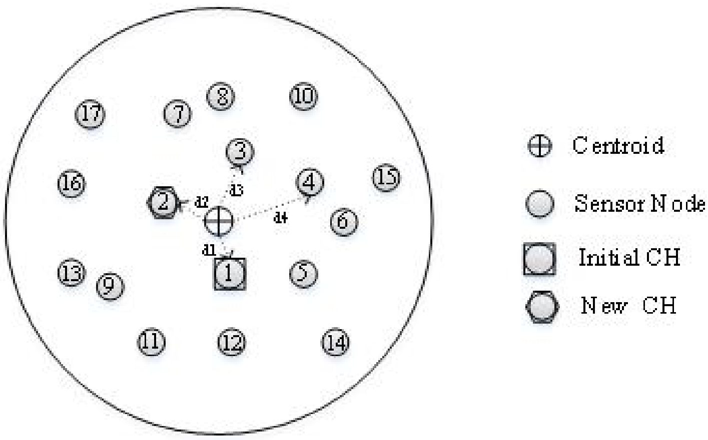


FIGURE 3. Reselection of cluster head according to the nodes with ID number.

The node ID is sorted according to the distance of the centroid in the cluster, and the closer to the centroid is selected as the new cluster head. So the energy consumption of each node in the next round will be as balanced as possible. The reselection principle of CHs is according to synthesize the residual energy and geographical position, which is more in line with the actual situation. The energy threshold is given according to the

TABLE 1. Cluster formation and initial cluster head selection.

Input: a sample set of sensor nodes $\rightarrow X = \{x_1, x_2, \cdots, x_N\}$;

optimum number of clusters $\rightarrow K_{opt}$;

Output: a set of K_{opt} clusters $\rightarrow C = \{C_1, C_2, \cdots, C_{K_{opt}}\}$;

K_{opt} number of initial CH $\rightarrow CH = \{CH_1, CH_2, \cdots, CH_{K_{opt}}\}$;

Steps:

1:Calculate the distance from each sensor node to the BS and sort them in ascending order;

2:Add up the total number of nodes whose distance to the base station is less than $d_0 \rightarrow count$;

3:$r = \dfrac{N - K_{opt} \times count}{1 + 2 + \cdots + (K_{opt} - 1)}$;

4: For $i = 1 : K_{opt}$

$size(i) = count + r \times (i - 1)$;

End

5: Divide the sorted nodes in step 1 into K_{opt} uneven sets $\rightarrow size(i)$ represents the size of each uneven set;

6: In each set, take the intermediate node as the initial centroid μ_k;

7: Each sensor node is assigned to the nearest centroid based on the Euclidean distance;

8: Calculate the new centroid for each cluster $\rightarrow Centroid(X, Y) = (\frac{1}{S}\sum_{i=1}^{S} x_i, \frac{1}{S}\sum_{i=1}^{S} y_i)$;

9: Repeat steps 7 and 8 until the error function $J = \sum_{n=1}^{N}\sum_{k=1}^{K_{opt}} r_{nk} \left\| x_n - \mu_k \right\|^2$ takes the minimum value and the iteration terminates;

where $r_{nk} = \begin{cases} 1, & if \quad n \in C_k \\ 0, & else \end{cases}$;

10: After the cluster is formed, an ID number is allotted to each node of a cluster according to the distance to the centroid.Assign smaller number to the closer one [15] ;

11: The node closest to the centroid of each cluster is selected as the initial cluster head [16], and a broadcast message is sent to notify nodes in the cluster.

information of the alive nodes:

$$E_{threshold} = lE_{elec}(\frac{countlive}{K_{opt}} - 1) + lE_{DA}\frac{countlive}{K_{opt}} + lE_{elec} + l\varepsilon_{mp}d_{toBS}^4 \quad (5)$$

TABLE 2. Dynamic reselection of cluster head.

Input: the nodes with ID number;
Output: the ID number of cluster head;
Steps:
1: For all selected CHs;
2: If ($E_{residual} \geq E_{threshold}$);
3: The node will remain as cluster head in the next round;
4 else
5: Check ID number of all sensor nodes in that cluster;
6: The node in the next order of ID number is selected as a new CH;
7: End if
8:End for

Where *countlive* is the total number of alive nodes.

3.3. **The phase of data transmission.** When the cluster head node transmits data, it needs to measure the distance from the BS firstly. If the distance between the CH node and the BS is less than d_0, the CH directly establishes communication with the BS by one-hop. Otherwise, along the direction of the CH node to the BS, looking for the nearest other CH node as an intermediate node to transmit data between BS and CH in a multi-hop manner. Before the data transmission, the network transmission routing table in in this environment is established according to the above rules, and each CH node can query its next hop node.

By means of the intermediate node to send data information to the BS, the energy consumption of CH is given by

$$E_{ICH} = \begin{cases} lE_{elec}[(n_c-1)+n_i] + lE_{DA}n_c + n_i \times (lE_{elec} + l\varepsilon_{fs}d_{toICH}^2) & if\ d_{toICH} < d_0 \\ lE_{elec}[(n_c-1)+n_i] + lE_{DA}n_c + n_i \times (lE_{elec} + l\varepsilon_{mp}d_{toICH}^4) & if\ d_{toICH} >= d_0 \end{cases} \tag{6}$$

The energy consumption of the CH nodes directly connected with the BS by single hop is as given by

$$E_{CH} = lE_{elec}[(n_c-1) + \frac{K_c}{K_{opt}-K_c}] + lE_{DA}(n_c + \frac{K_c}{K_{opt}-K_c}) + lE_{elec} + l\varepsilon_{fs}d_{CHtoBS}^2 \tag{7}$$

In formula (6) and (7), n_c is the number of members in the cluster; n_i represents the total number of hops for the CH as intermediate node; d_{toICH} is the distance between CH and intermediate nodes; K_c represents the number of CH that transmit data in a multi-hop routing to the BS; d_{CHtoBS} is the distance from the CH to the BS.

The energy consumed by each cluster member nodes per round is:

$$E_{non-CH} = \begin{cases} lE_{elec} + l\varepsilon_{fs}d_{toCH}^2 & if\ d_{toCH} < d_0 \\ lE_{elec} + l\varepsilon_{fs}d_{toCH}^4 & if\ d_{toCH} >= d_0 \end{cases} \tag{8}$$

Where d_{toCH} is the distance between member nodes and corresponding cluster head.

The total energy consumption of each round by the whole network during the data transmission phase is as follows:

$$E_{round} = \sum_{K_c} E_{ICH} + \sum_{K_{opt}-K_c} E_{CH} + (countlive - K_{opt})E_{non-CH} \tag{9}$$

4. **Simulations and Analysis of Results.** In order to verify the performance of UCRPK-means protocol in the environment of coal-mine, MATLAB have been used to evaluate and analyze UCRPK-means protocol, compared with EECPK- means and EEUC protocol. Simulation parameters are summarized in Table 3.

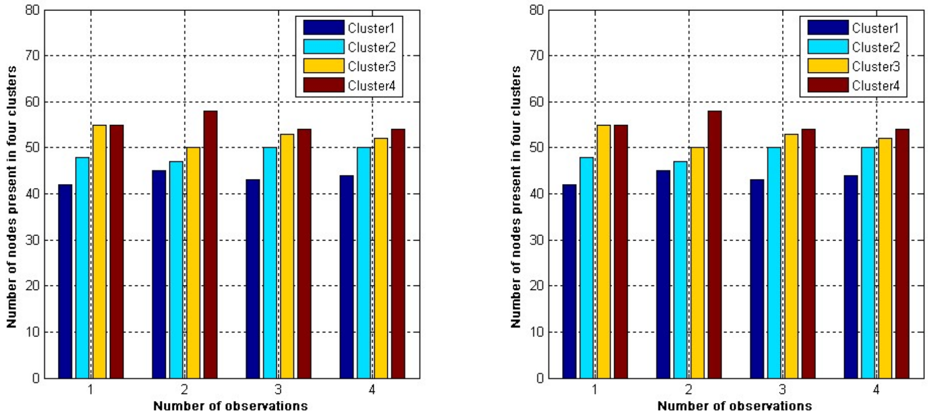
Table 3. Simulation parameters.

Parameters	Value
Simulation area	$500 * 50m^2$
BS location	(500,25)
Number of sensor nodes	200
Initial energy per nodes	1J
E_{elec}	50 nJ/bit
ε_{fs}	$10pJ/bit/m^2$
ε_{mp}	$0.0013pJ/bit/m^4$
d_0	87m
Data package size	4000 bit
Control package size	200 bit

4.1. **Comparison of our proposed UCRPK-means with EEUC protocol.** In [12], the EEUC protocol proposed by Li constructs clusters of different sizes by using uneven competition ranges of candidate clusters, so that the clusters close to the BS are smaller than the clusters far away from the BS. In order to verify the energy balance and effectiveness of UCRPK-means in coal-mine, we compare UCRPK-means and uneven clustering protocol EEUC in terms of cluster size and energy consumption of CH firstly. As can be seen from Fig.4, when the network is divided into four clusters, the number of cluster members in both protocols increases with the distance from the BS, but the cluster size of the UCRPK-means protocol is more uneven.

Fig.5 shows the variance of the energy consumption by the CH in each of the eight randomly selected rounds. It is clear from the figure that UCRPK-means has the lowest variance compared to the EEUC protocol, thus best balanced the energy consumption of the CH, demonstrating that the non-uniform design of UCRPK-means protocol is better than EEUC protocol.

4.2. **Comparison of the three protocols.** When the number of alive nodes in the network is less than 20% of the total number of initial nodes, the system cannot continue to work. In Fig.6, it shows the number of nodes which are alive using EECPK-means, EEUC and our proposed UCRPK-means protocol. It can be seen from the figure that UCRPK-means significantly prolongs the death time of the first node. The span between the death of the first node and the end of the network can reflect the balance of energy



(a) Using EEUC Protocol. (b) Proposed UCRPK-means Protocol.

FIGURE 4. Number of nodes in four clusters.

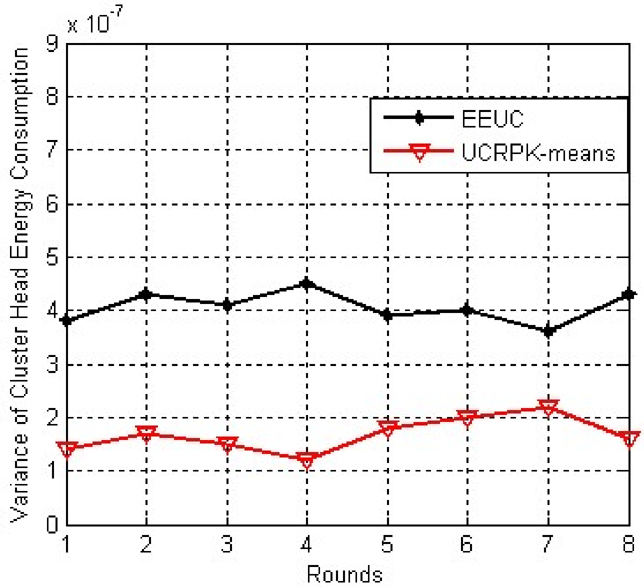


FIGURE 5. The variance of the energy consumed by the CH.

consumption. The span of EECPK-means and EEUC is 1400 and 1300 rounds respectively, while the UCRPK-means protocol is 1600 which are 14.2% and 23.1% higher than the EECPK-means and EEUC protocols respectively. The UCRPK-means protocol enables the network to operate 2300 rounds, which is 1.2 times that of the EECPK- means and EEUC protocols. This shows that UCRPK-means protocol can better balance the network energy consumption and extend the network lifetime.

In Fig.7, it shows the comparison of total energy consumption in the network of EECPK-means, EEUC and UCRPK-means. It can be seen that the total energy consumption using UCRPK-means protocol is always lower than EECPK- means and EEUC. EECPK-means and EEUC protocol cannot continue to work after 1900 rounds, and energy consumption reaches 95% of total energy. With the same energy consumption, the UCRPK-means protocol enables the network to operate around 2300 rounds. As can be seen, the UCRPK-means protocol can effectively reduce energy consumption compared to above mentioned protocols.

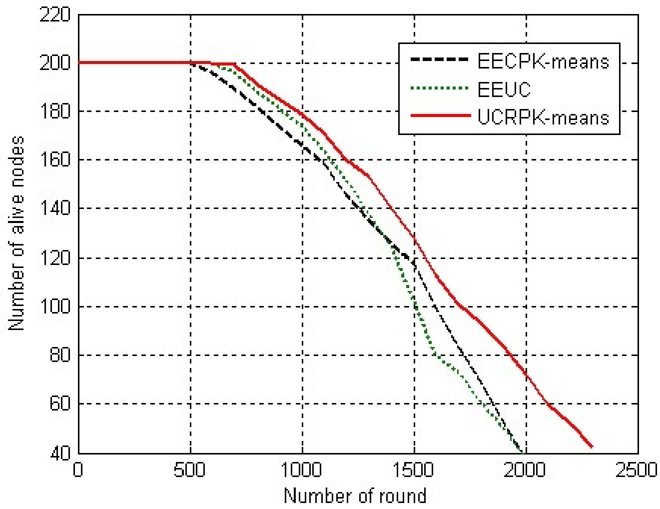


Figure 6. Number of alive nodes with respect to number of rounds.

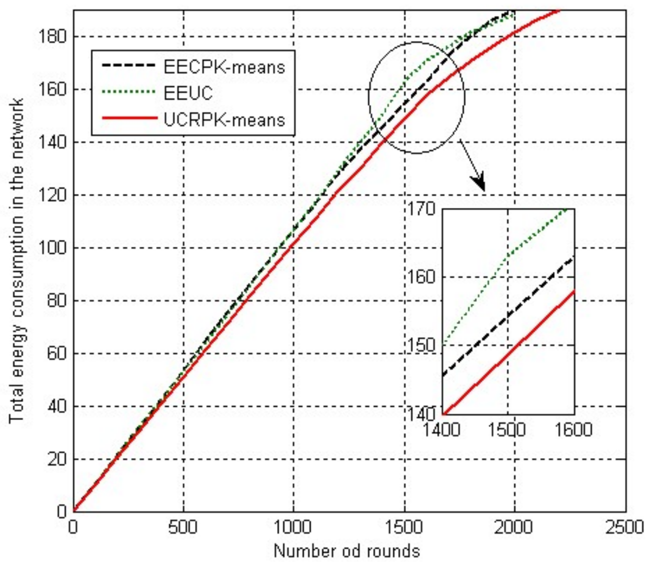


Figure 7. Energy consumption with respect to number of rounds.

5. **Conclusion.** In this paper, based on the analysis of long band-type topology of WSNs in the coal-mine, we propose an uneven clustering routing protocol based on improved K-means to solve the serious energy consumption imbalance in this kind of network. This protocol constructs clusters with different size through the selection of initial centroid in K-means algorithm to balance the energy consumption among cluster heads. The selection and reselection of cluster heads take numbering mechanism, which is based on the remaining energy and geographic location. Multi-hop routing in the data transmission phase is designed according to the special linear topology of the coal-mine, which not only

focuses on the link energy consumption but also emphasizes node load balancing. Simulation results show that this protocol can balance the network load more effectively than the existing several typical clustering protocols, so as to optimize the network performance and prolong the network lifetime.

Acknowledgement. The authors would like to thank the support of this work by Grants from the National Natural Science Foundation of China (NSFC)(No. 61302155, No. 61274080) and Nanjing University of Posts and the introduction of talent Project (No. NY2l4052).

REFERENCES

[1] W. Yang, Y. Huang, Wireless sensor network based coal mine wireless and integrated secure monitoring information system, *International Conference on Networking IEEE*, pp. 13–15, 2007.
[2] X. G. Niu, X. Huang, X. Zhao, et al. The design and evaluation of a wireless sensor network for mine safety monitoring, *IEEE GLOBECOM 2007-IEEE Global Telecommunications Conference IEEE* , pp.1291-1295, 2007.
[3] D Meana-Llorin, CG Garca, V Garca-Daz, et al, SenseQ: Replying questions of Social Networks users by using a Wireless Sensor Network based on sensor relationships, *Data Science and Pattern Recognition*, vol. 1, no. 1, pp.1–12, 2017.
[4] H. L. Wang, S. Zhang, J. J. Wu, Improvement of LEACH routing protocol based on long strip coal mine roadway, *Safety in Coal Mines*, vol. 42, no. 10, pp.63–65, 2011.
[5] B. Shen, S. Y. Zhang, S.Zhong, Clustering routing protocol for wireless sensor network, *Journal of Software*,vol. 17, no. 7, pp.1588–1600, 2006.
[6] Krishnamachari. B, Estrin. D, Wicker. S, The impact of data aggregation in wireless sensor networks, *International Conference on Distributed Computing Systems Workshops, 2002. Proceedings IEEE*, vol. 5, pp.575-578, 2002.
[7] Heinzelman. W, Chandrakasan. A, Balakrish. H, Energy-efficient communication protocol for wireless micro sensor networks, *Hawaii International Conference on System Sciences IEEE* , vol. 18, pp.3005-3014, 2000.
[8] J. S. Pan, T. T. Nguyen, T. K. Dao, et al, Clustering Formation in Wireless Sensor Networks: A Survey, *Journal of Network Intelligence*, vol. 2, no. 4, pp. 287-309, 2017.
[9] Lindsey. S, Raghavendra. C, Slvalingam. K. M, Data gathering algorithms in sensor networks using energy metrics, *IEEE Transactions on Parallel and Distributed Systems*, vol. 13, no. 9, pp.924-935, 2002.
[10] Younis O, Fahmy S, HEEDa hybridenergy-efficientdistributed clustering approach for ad hoc sensor networks, *IEEE Transactions on Mobile Computing* , vol. 3, no. 4, pp.366-379, 2004.
[11] A. Ray, D. De, Energy efficient clustering protocol based on K-means (EECPK-means)-midpoint algorithm for enhanced network lifetime in wireless sensor network, *IET Wireless Sensor Systems*, vol. 6, no. 6, pp.181-191, 2016.
[12] C. F. Li, G. H, Chen, M. Ye, An uneven cluster-based routing protocol for wireless sensor networks, *Chinese Journal of Computers*, vol. 30, no. 1, pp.27-36, 2007.
[13] H. Gwo-Jiun, C. Tun-Yu, W. Hsin-Te, The Adaptive Node-Selection Mechanism Scheme in Solar-Powered Wireless Sensor Networks, *Journal of Network Intelligence*, vol. 3, no. 1, pp.58-73, 2018.
[14] C. Li, M. Ye, G. Chen, et al, An energy-efficient unequal clustering mechanism for wireless sensor networks, *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference IEEE*, vol. 2005, pp.8 pp.-604, 2005.
[15] G. Y. Park, H. Kim, H. W. Jeong, et al, A novel cluster head selection method based on K-means algorithm for energy efficient wireless sensor network, *International Conference on Advanced Information NETWORKING and Applications Workshops IEEE*, pp. 910-915, 2013.
[16] P. Sasikumar, S. Khara, K-means clustering in wireless sensor networks , *Fourth International Conference on Computational Intelligence and Communication Networks IEEE Computer Society*, vol. 59, pp.140-144, 2012.