Data Clustering Method in Wireless Sensor Networks Based on Residual Energy Perception

Xudong Yang
Chongqing Vocational Institute of Safety & Technology,Wanzhou ,China
13609433383@163.com

Abstract—To prolong the survival time of wireless sensor network, an iterative scheme was proposed. First of all, spectrum clustering algorithm iteratively segmented the network into clusters, and cluster head nodes in each sub cluster were determined depending on the size of residual energy of sensor nodes. Then, a data forwarding balance tree was constructed in each sub cluster. Data forwarding path of each non-cluster head node was defined, and the moving path of a mobile data collector was determined, which used the residual energy as the basis for the network optimization. Finally, this scheme was simulated, and two traditional data gathering algorithms were compared. The results showed that the algorithm designed in this experiment could effectively balance energy consumption among all WSN nodes and had great performance improvement compared with the traditional data collection algorithm. To sum up, this algorithm can significantly reduce the energy consumption of the network and improve the lifetime of the network.

Keywords—Mobile data collection, residual energy perception, data forwarding tree, spectral clustering

1 Introduction

In the current era, people generally believe that information can be seen everywhere, but the difficulty is how to get effective information quickly and conveniently. The way of information acquisition has become the bottleneck of development. At present, there are many ways to obtain information. One of the fastest developing branches is wireless sensor network. In the existing applications, wireless sensor networks have a good performance in all fields. In military terms, the aircraft is used for putting sensors. The microsensors with small volume and strong concealment are scattered in the enemy area for monitoring; in the industry, sensor nodes are deployed in the pipeline to monitor the operation of each link, so as to avoid the accident; in the environmental protection, the sensor nodes are deployed in the contaminated area and high radiation area to capture the environmental pollution degree and these places are not close to the staff; in life, sensor nodes are also used in medical treatment and home furnishing scene to monitor ill health of seriously diseased patients, and timely

alarm; in the monitoring family safety, it prevents the elderly and children event accident.

Therefore, the problem of data collection in wireless sensor network (WSN) has been widely studied in the last ten years. Considering the characteristics of sensor nodes, when people study WSN, the most important factor is energy saving, and the sending and receiving process of data is the most serious part in energy consumption of sensor nodes. As is known to all, one of the most important features of sensor nodes is that their energy cannot be supplied continuously, and only limited energy is provided by the battery. Since the sensor nodes are deployed to remain stationary, we want the WSN to survive as long as possible. For the entire sensor network, its lifetime is limited to the first single sensor node that consumes energy. Therefore, in order to extend its lifetime, we must study the energy consumption of a single sensor node.

2   Literature review

In terms of wireless sensor network multi-parameter data acquisition, Bayram and Şeker as stated in [1] used MICAz nodes to collect low-frequency vibration signals for transformer fans and compressors in substation, and the sensor bandwidth is only 50Hz. Although these nodes have realized multi-parameter data acquisition, the performance is generally low and cannot meet the testing of the mechanical parameters with high frequency vibration signals. Liu and others as stated in [2] already introduced the high frequency vibration signal acquisition system used for sanding monitoring in oil well. Ruela and others as stated in [3] designed memetic and evolutionary design of wireless sensor networks based on complex network characteristics. To improve the node data acquisition frequency accuracy of wireless sensor network, Ahmed proposed a kind of norm HMF and suppressed sampling interval vibration in the process of data collection. According to the sampling interval, Niaki and others as stated in [4] designed a variable sampling interval. HMF statistics sampling interval is used and suppressed vibration, but it does not consider the frequency drift in the process of sampling. Naderi as stated in [5] used a scalable and energy aware protocol to increase network lifetime in wireless sensor networks. Rao and others as stated in [6] found residual energy aware mobile data gathering in wireless sensor networks.

For suppressing data acquisition frequency drift, Liu and others as stated in [7] proposed wireless sensor network node based on cross layer design. Nodes obtain accurate SFD to measure and make up for node crystal oscillator drift through cross layer. It also inhibits data acquisition frequency drift and increases the sampling interval time vibration. The accuracy of the data acquisition frequency is improved by the software algorithm, but the effect is not good. In addition, Logambigai and Kannan as stated in [8] proposed the unequal clustering based on wireless sensor networks. At the same time, Azad and Sharma as stated in [9] discussed the maximum residual energy based clustering scheme for wireless sensor networks. Later, Heed as stated in [10] studied the data gathering method based on a mobile sink for minimizing the data loss in wireless sensor networks. Aiming at the problem of efficient transmission of

massive data caused by high frequency acquisition of mechanical testing, Bhuiyan put forward event sensitive adaptive sampling to reduce data amount and energy consumption in WSN. These methods are applicable only to specific data, and the complexity of algorithm is high.

To sum up, the existing research on data clustering methods is not enough, such as the effectiveness of clustering method in sensor networks. To solve the problem of wireless sensor networks, the calculation method based on residual energy perception is proposed to prolong the survival time of wireless sensor network. This method can significantly reduce the energy consumption of the network and improve the lifetime of the network.

3 Mobile data collection method in Wireless Sensor Networks based on residual energy

3.1 Energy consumption model

The dissipative model of the antenna is used to describe the energy consumption, and the model is used to measure and analyze the energy consumption of each sensor node.
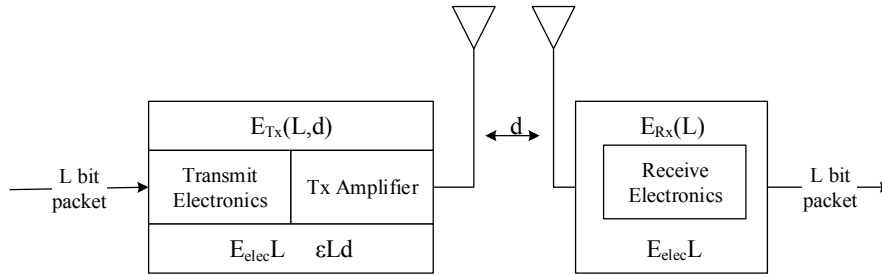

Fig. 1. Energy Consumption Model

d represents the distance between the sender and the receiver. *L* indicates the length of the data to be sent, and the E_{elec} refers to the energy consumed by sending or receiving every bit data. The energy consumption formula of the transmitting antenna is as follows:

$$E_{tx}(\text{L},d) = \begin{cases} L * E_{elec} + L * E_{fs} * d^2, d < d_0 \\ L * E_{elec} + L * E_{mp} * d^4, d > d_0 \end{cases} \tag{1}$$

$E_{tx}(L, d)$ indicates the energy consumed for sending L−bit data when two node shave the distance of *d*. The parameter E_{fs} represents the dissipative energy in the free space model (d^2 energy loss rate), and the parameter E_{mp} suggests the dissipative ener-

gy in multichannel attenuation channel model (d^4 energy loss rate). The threshold d_0 is defined as:

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \qquad (2)$$

Similarly, the energy consumption formula of the data sender is as follows:

$$E_{rx}(\mathrm{L}, \mathrm{d}) = E_{elec} * L \qquad (3)$$

Based on this energy consumption model, it is easy to find that the transmission distance has a great impact on the energy consumption. Therefore, the distance factor is the first factor to be considered in the clustering process.

3.2 Algorithm description

The major objective for wireless sensor networks data collection protocol of residual energy perception is, by dividing a large connected network into small independent networks, selecting a "central node" in every small independent network. The so-called "central node" is abstract, and it can be both virtual node and entity node. It can be used as a confluence node of data as well as a data forwarding task alone. The mobile data collector stays in each "central node" for a period of time in accordance with a certain route. During this period, the mobile data collector broadcasts the token signal received by the WSN nodes in the surrounding communication area, and all the sensor nodes that receive the token signal will flow to the location where the mobile data collector is. At the beginning of the algorithm, based on the optimal energy consumption, we divide the entire wireless sensor network into two sub clusters, further find sub clusters satisfying subdivision conditions, and subdivide the sub cluster equally. Since each subdivision increases the stopping point of the mobile data collector, we repeat the process until the mobile data collector runs longer than the system set threshold. When all the sub clusters are formed, we construct a data forwarding tree based on the residual energy for each cluster to improve the energy utilization. Each member of the cluster works along the data forwarding tree.

3.3 Clustering stage

First of all, spectral clustering is used for the network clustering. Assuming that $A \square R^{N \square N}$ is the similarity matrix of graph G, each element value in A is associated with the distance between any two WSN nodes in the network. This is the Gauss similar type, which is defined as follows:

$$A = [a_{ij}] = \begin{cases} \exp(-\dfrac{d^2(i,j)}{2\sigma^2}) , i \neq j \\ 0, otherwise \end{cases} \qquad (4)$$

In the iteration of the spectral clustering, we can get k sub clusters that may be subdivided again. In order to not lose the generality, we supposed k=2. That is to say, we started from the original network and generate two sub clusters after the iteration. After n round iterations, we formed n+1 small enough sub clusters and mobile data collector drive on path point formed in these head clusters of sub cluster.

Spectral clustering is a class of algorithms. In spectral clustering mapping methods with different criteria and different bases, the steps of the algorithm are different. Here, we have a little change in the general spectral clustering algorithm, and use the following steps to calculate:

A sensor is mapped to a point and a complete graph is constructed. The distance between any pair of sensor nodes is mapped to the weight of the edges between points. The complete graph is expressed in the form of the adjacency matrix, recorded as W and used as the similarity matrix in the spectral clustering.

The list elements of W are summed and the total number of N is obtained, consisting of a diagonal matrix (the data except for the diagonal are 0 data matrix), resulting in a matrix of N*N, denoted as D. Let L=D - W.

The first k feature values and feature vector of L are calculated and the feature vector space is constructed.

The first k feature vectors are combined to form a matrix of N*k. In the N*k matrix, each row vector is the vector in the k dimension space. The typical clustering method such as k-means algorithm is used to cluster the N vectors and the N vectors represent the N nodes in the original graph. Then, in the clustering results, the category that each row vector belongs to indicates the class of nodes in the original graph. In this way, the classification of original data point is obtained.

In general, the task of the cluster head node is to collect all the data generated by its member nodes and wait for the mobile data collector being close for the data upload. It is clear that the cluster head nodes are loaded in this way, causing them to run out of energy quickly.

However, because of the introduction of a mobile data collector, we no longer define traditional cluster head nodes. In fact, we use "pull node" to replace the cluster head. A simple exposition is made: because the mobile data collector can move at any location in two-dimensional plane, its candidate stopping point is infinite theoretically. In order to simplify the degree of discretization of two-dimensional plane, and try to make as many sensor nodes as possible to get closer to mobile data collector, we define a pulping node as a common WSN node. For the mobile data collector, a pull node is only a location point used to identify its location. When the mobile data collector reaches a pull node, it takes the pull node as the center; all the neighbor nodes in the same cluster will receive the signal and begin to upload their sensing data to the mobile data collector. Because these neighbor nodes are in the range of pull node, they can directly transmit the data packet to the mobile data collector rather than the

pull node. The traditional data collection method is that the data of the neighbor nodes are reported by the cluster head nodes, as shown in Figure 2.

Figure 2 shows a schematic diagram of the traditional cluster head node collection data. Among them, the line frame represents the subdivision of the sub cluster, and the node A and node B are the cluster head nodes of two clusters, respectively.

Figure 3 shows the flow chart of data that is collected at the pulling node when the mobile data collector is involved in the work. Among them, the car represents the mobile data collector. Nodes A and B are still cluster heads in Figure 7. The difference is that they no longer collect data in the sub cluster, but only identify a stop position of the mobile data collector, and all the data are uploaded to the car.
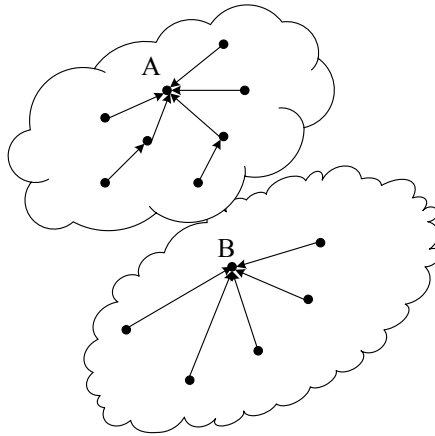

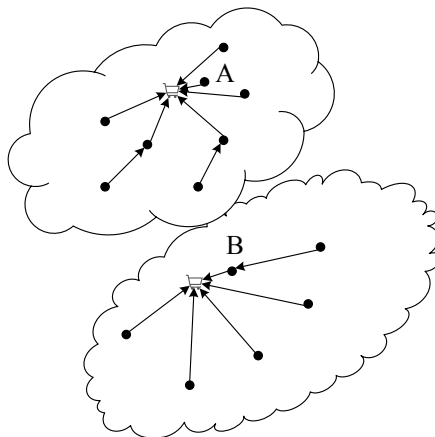**Fig. 2.** Traditional Cluster Head Date Gathering Example

Fig. 3. Polling Point Date Gathering Example

The main reason for this consideration is the balance of energy consumption. As mentioned above, the lifetime of a single bottleneck node affects the lifetime of the whole network. We arrange the nodes around the traditional cluster heads reasonably to slow down the pressure of cluster heads, and the load of cluster heads is much relieved.

3.4 Determination of cluster subdivision

Once the pull node is determined, the mobile data collector can move to these special WSN node locations to collect data. If the time for mobile data collector traversing all the pull nodes is less than the threshold of the system, we can further use the mobile data collector to save the energy consumption of sensors. In order to find out which sub cluster can be subdivided into small clusters, a strategy is designed to complete. The main reason for this algorithm is the trade-off between the reduction of the energy consumption of sensor nodes and the increase of the path of the mobile data collector. In every stage of the algorithm, two indexes are considered: one is ΔP and the other is ΔL. Assuming that, in the t round iteration, the cluster C_i is subdivided into sub clusters C_i' and C_i'', and these two sub clusters are used in the t+1 round iteration. As the sub cluster C_i is subdivided, the inevitable result is the increase in the number of pull nodes within the sub cluster C_i, resulting in an increase in the total travel time of the mobile data collector. ΔL is the increase in the amount of time, which can be defined as:

$$\Delta L = TSP(\text{t}+1) - TSP(\text{t}) \tag{5}$$

TSP(t+1) indicates that the total driving time for mobile data collector traversing all the pull nodes in t-th round data collection.

Similar to the cause of path growth, the total energy consumption in the original C_i sub cluster is reduced due to the subdivision of the cluster. This is because the sensor nodes at the edge of the two fine clusters no longer transmit the data to each other, but are completely separated. ΔP is defined as the reduction of energy value. The number of data packets without forwarding depends on two aspects. One is the number of edge passed through by the dotted line and the other is the number of vertex segmented and independent after the edge passed through by the dotted line is removed. Therefore, ΔP is defined as:

$$\Delta P = C * - N \tag{6}$$

In order to measure the importance of two parameters to the selection results, the definition is:

$$\beta = \frac{\Delta P}{\Delta L} \tag{7}$$

This parameter is used to determine which sub clusters should be subdivided in some rounds iteration. In general, we always want a subdivision that saves the maxi-

mum energy and increases the path length to the minimum. As a result, the larger the β value of a cluster, the more subdivided the sub cluster is.

3.5 Determination of the data forwarding tree

When the mobile data collector stays at a pull node, how should all the sub cluster members forward the packet to the mobile data collector? The solution is to design a data forwarding tree. It is still based on the concept of residual energy to balance the energy consumption on a data forwarding tree.

In order to make better use of the energy of the whole network, we redefine the load balancing: the more the remaining energy of the sensor node is, the more data it will transmit. This assumption is reasonable. To solve this problem, we present an algorithm to solve the load balancing problem. It is described as follows: when all the pull nodes are determined in the previous steps, all the sensor nodes in the sub cluster are accessed in the order of BFS. Suppose U_i is the collection of sensor nodes in i-th layer in the BFS process, then the remaining energy set of members in U_i is P_i. Suppose BFS has a total of n layers throughout the process, then the iteration starts from the farthest layer. Assuming that U_i^k is k-th element in U_i, k=1, 2,... size (U_i). N (U_i^k) represents the set of one hop neighbor nodes of U_i^k. For the sake of simple description, another variable is set: $C_i(U_i^k)=U_i \cap N(U_i^k)$.

It represents one hop neighbor node set. That is, $C_i(U_i^k)$ indicates some brother nodes of $C_i(U_i^k)$, while $C_{i-1}(U_i^k)$ suggests some parent nodes of U_i^k. U_i^k is added to the data forwarding tree with two possible ways: U_i^k directly sends data packet to U_{i-1}^k, where U_{i-1}^k is a node in $C_{i-1}(U_i^k)$; or U_i^k firstly sends data packet to U_i^j where U_i^j is a node in $C_i(U_i^k)$, and then U_i^j forwards data to U_{i-1}^p, which is a node in $C_{i-1}(U_i^p)$.

These two different processes represent directly forwarding data from the upper layer node and transferring to the upper node after the layer node. The two processes have different energy consumption. If U_i^k sends data through the first way, then we need to reduce *count*$\square\Delta E$ for P_i^j. Among them, *count* is the number of all data packets to be sent and ΔE is the energy consumed for sending a data packet. On the contrary, if the maximum residual energy of every node in $C_i(U_i^k)$ is less than $C_{i-1}(U_i^j)$, then U_i^k should send the data to his neighbor node U_i^j, then U_i^j transmits all data to its parent node U_{i-1}^p. In this way, we can reduce the load of the nodes with less energy left, and the current high energy node load increases. It is worth mentioning that each sensor node should keep one of his parent nodes id.

4 Experimental results

Some parameters results of the algorithm studied here (REAMDC-WSN algorithm), traditional LEACH algorithm and mobile data collection algorithm SCRC-WSN proposed by Brahim are given, which proves the performance advantage of the algorithm in improving the network lifetime.

4.1 Relationship between path length and network lifetime

The main goal of the REAMDC-WSN algorithm is to improve the lifetime of the network without exceeding the default time of the system. Figure 4 shows a diagram of the relationship between the path length and the total network lifetime of the mobile data collector.
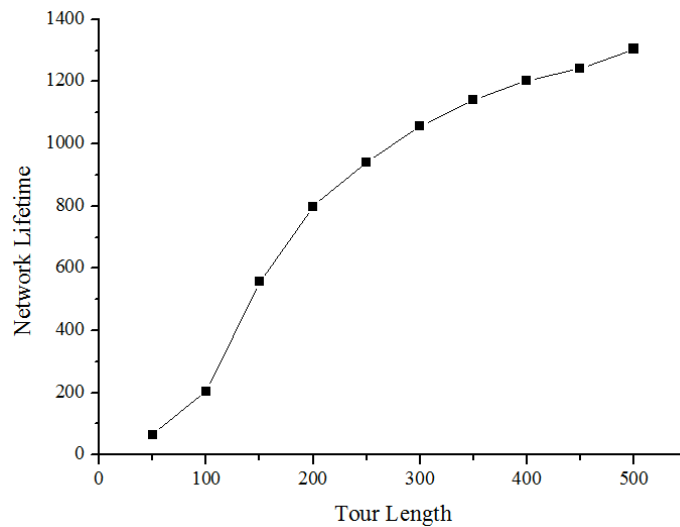

Fig. 4. Relationship Between Path Length And Network Life Time

It can see from the figure that, when the path length is increased, the lifetime of the sensor network is not increasing at constant rate. On the contrary, the slope increases rapidly at the beginning, then becomes slower and slower, and eventually maintains a stable state. This phenomenon can be interpreted as: when the path length is limited to a smaller value, such as L=100, which suggest that the cluster is relatively large, data packets needed forwarding within the cluster are relatively more. It will make more energy consumed, and thus the network overall survival time becomes shorter. Once the length of the path is increased, the size of the cluster becomes smaller and smaller, so the lifetime of the network increases rapidly. The graph indicates that once L is increased to 400, the lifetime of the network will be significantly improved, that is, the whole network is subdivided into many small clusters, and almost any node can easily access the node. The mobile data collector has made a great contribution to improving the network lifetime. However, to a certain extent, more system threshold cannot significantly improve the network survival time, and its growth rate is slow. If we continue to increase the path length, the lifetime of the network will not continue to improve. This is because the mobile data collector can access every sensor node.

4.2 Average distance between sensor nodes

Figure 5 shows the relationship between the network lifetime and the number of sub cluster with the average distance between the sensor nodes for a given path L.
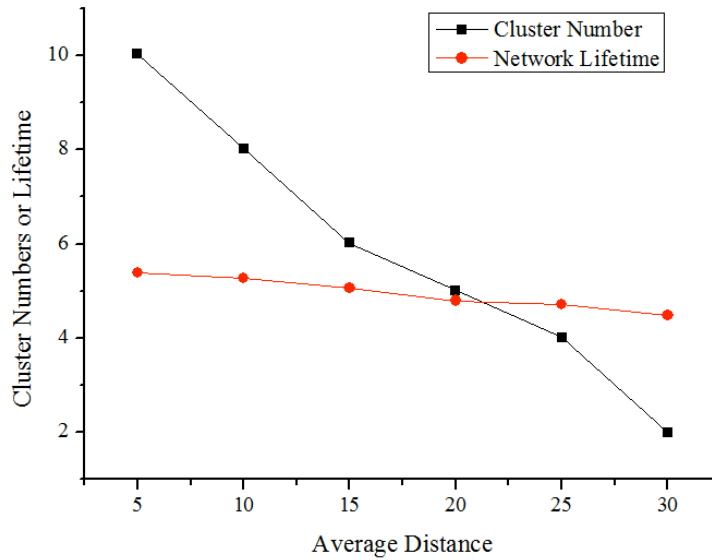

Fig. 5. Relationship Between Average Distance And Network Life Time

Since the WSN nodes are randomly distributed in the detection range, the average distance between them is different. The lines in the graph indicate that once the average distance becomes larger, the number of sub clusters and the lifetime of the network will be smaller. This is because the larger the average distance of the cluster is, the sparer the distribution of sensor nodes is. Since the path length of a mobile data collector is fixed, it can't get to too far the cluster to collect data. In this case, remote sensor nodes can only transmit data through other sensor nodes forwarding, which will lead to additional energy consumption. In addition, in the energy consumption model we use, the consumption of transmitting and receiving energy between a pair of sending sensors and receiving sensors is a function of their distance d. Therefore, the average distance has great influence on the formation and energy consumption of sub clusters.

4.3 Relationship between the residual energy and the network lifetime in the network

The node energy is one of the important indexes to be considered by WSN, and the residual energy is an important parameter that can prolong the lifetime of WSN. Therefore, we compare the energy of the three algorithms. Figure 6 shows the changes in the residual energy in the entire network in each of the three algorithms.
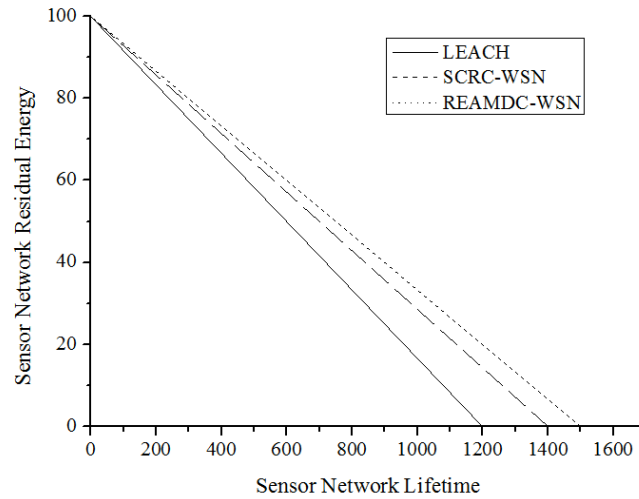
Fig. 6. Relationship Between Residual Energy And Network Life Time

It can be seen that the residual energy changes of the whole network in the LEACH and SCRC-WSN algorithms are faster than that of the REAMDC-WSN algorithm. The slope of the LEACH algorithm is 0.094, the slope of SCRC-WSN is 0.072, and the slope of REAMDC-WSN is only 0.067. This is because we save a lot of energy in the stage of fine molecular cluster. We always choose the sub cluster with the smallest energy consumption to subdivide, so in the three algorithms, the residual energy index is the most excellent. It is also fully demonstrated that our REAMDC-WSN algorithm has a reasonable planning and utilization for the residual energy of the WSN node.

4.4 Average data collection energy consumption comparison diagram

Because the three algorithms are divided into rounds to collect data, we compare each round of data collection energy consumption. In that neither of the three algorithms has data fusion, for the sake of convenience, we measure the energy consumption of a round of data collection based on the amount of data packets sent, as shown in Figure 7.

In the LEACH algorithm, due to the need to broadcast the cluster head node selection process, and there are more data in the cluster forwarding, LEACH algorithm has the maximum energy consumption in average each round data collection; SCRC-WSN algorithm uses spectral clustering for clustering division, which to some extent reduces the transmission amount of broadcast data. However, because there is no mobile data collector involved, the average consumption of each round data of the mobile phone is still relatively high; in the REAMDC-WSN algorithm, the participation of mobile data collector significantly improves the average data collection energy consumption.
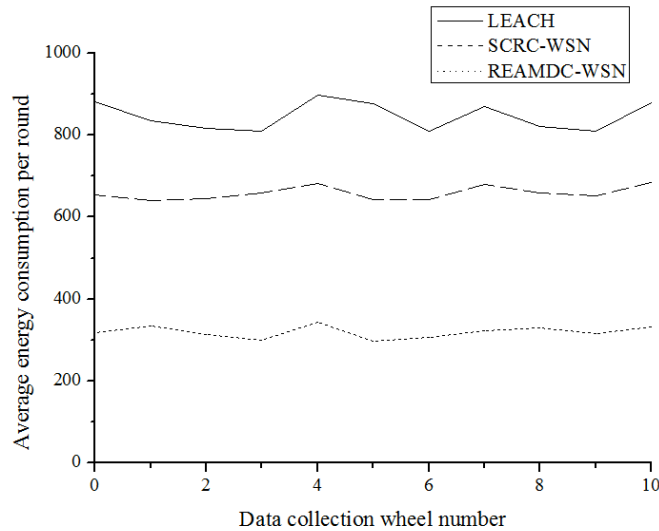
Fig. 7. Relationship Between Residual Energy And Network Life Time

5 Conclusions

In this paper, a spectrum clustering algorithm is used to segment the network into clusters. Then, a data forwarding balance tree was constructed in each sub cluster. The results showed that the algorithm designed in this experiment could effectively balance energy consumption among all WSN nodes. as a result, the following conclusions can be summarized:

First, combining the mobility and the clustering, balance is made between delay and energy consumption. It guarantees delay parameters that the system can tolerate and improves the survival time of the network as much as possible.

Second, a centralized data collection scheme by REAMDC-WSN is provided. By spectral clustering, dynamic selection of pull nodes is conducted, dynamic subdivision of cluster heads is performed and the data forwarding tree reporting data is completed. As a result, the WSN data collection scheme with good performance is obtained.

Finally, compared with the classical LEACH algorithm, the SCRC-WSN algorithm in recent years, and the REAMDC-WSN algorithm, in terms of mobile data collector path length, node average distance, network lifetime and residual energy, this scheme can have effective energy balance between each WSN node. After comparison with the traditional data collection algorithm, the performance can be significantly improved.

6   References

[1] Bayram, D., & Şeker, S. Wavelet based neuro-detector for low frequencies of vibration signals in electric motors. Applied Soft Computing, 2013, vol. 13.5, pp. 2683-2691. https://doi.org/10.1016/j.asoc.2012.11.019

[2] Liu, G., Liu, P., & Han, J. High frequency vibration signal acquisition system used for sanding monitoring in oil well. Petroleum Geology & Recovery Efficiency, 2013, vol. 20.6, pp. 108-110.

[3] Ruela, A. S., Cabral, R. D. S., Aquino, A. L. L., & Guimarães, F. G. Memetic and evolutionary design of wireless sensor networks based on complex network characteristics, 2017, vol. 1(2), pp. 33-53.

[4] Niaki, S. T., FazlollahMasoumiGazaneh, & MoslemToosheghanian. Economic design of variable sampling interval x-bar control charts for monitoring correlated non-normal samples. Communications in Statistics, 2013, vol. 42(18), pp. 3339-3358. https://doi.org/10.1080/03610926.2011.628769

[5] Naderi, H. ScEP: A Scalable and Energy Aware Protocol to Increase Network Lifetime in Wireless Sensor Networks. Wireless Personal Communications, 2015, vol. 82(1), pp. 611-623. https://doi.org/10.1007/s11277-014-2243-8

[6] Rao, X., Huang, H., Tang, J., & Zhao, H. Residual energy aware mobile data gathering in wireless sensor networks. Telecommunication Systems, 2016, vol. 62(1), pp. 31-41. https://doi.org/10.1007/s11235-015-9980-1

[7] Liu, Q., & Liu, M. Energy-efficient clustering algorithm based on game theory for wireless sensor networks. International Journal of Distributed Sensor Networks, 2017, vol. 13(11), pp. 155014771774370. https://doi.org/10.1177/1550147717743701

[8] Logambigai, R., & Kannan, A. Fuzzy logic based unequal clustering for wireless sensor networks. Wireless Networks, 2016, vol. 22(3), pp. 945-957. https://doi.org/10.1007/s11276-015-1013-1

[9] Azad, P., & Sharma, V. Maximum residual energy based clustering scheme for wireless sensor networks. Advanced Science Focus, 2013, vol. 1(2), pp. 111-119. https://doi.org/10.1166/asfo.2013.1021

[10] Heed, A. A Data Gathering Method Based on a Mobile Sink for Minimizing the Data Loss in Wireless Sensor Networks. International Journal of Distributed Sensor Networks, 2015, vol. 5, pp. 242.

7   Author

Xudong Yang is with Chongqing Vocational Institute of Safety & Technology,Wanzhou, 404020, China.