A Clustering Algorithm for Wireless Sensor Networks Using Geographic Distribution Information and Genetic Algorithms

Yu Song[1,2(✉)], Zhigui Liu[1], and He Xiao[1,3]

[1] School of Information Engineering, South West University of Science and Technology, Mianyang 621010, Sichuan, China
3958613@qq.com

[2] Department of Network Information Management Center, Sichuan University of Science and Engineering, Zigong 643000, China

[3] School of Computer Science, Sichuan University of Science and Engineering, Zigong 643000, China

Abstract. WSN consists of a large number of micro sensor nodes with limited resources. The limited battery resources of these nodes have become an important bottleneck to the development of WSN. In order to improve the energy efficiency and prolong the network life cycle, we propose a clustering method GDGA based on improved genetic algorithm. In this method, we divide the sensor area into two parts: the first part is that the distance from the node to BS is less than the transmission threshold of the node. For the nodes in this area, we do not cluster but directly transmit the data to BS. The part beyond the threshold of BS is divided into the second region. The nodes in this region will be clustered using the improved genetic algorithm according to the characteristics of node distribution. Simulation results show that compared with other four protocols, GDGA has the highest energy efficiency, lower average energy consumption of cluster head and longer life cycle of the whole network.

Keywords: Genetic algorithm · Clustering · CH selection · Clustering beyond threshold

1 Introduction

WSN is composed of many miniature and cheap sensor nodes. These sensor nodes are usually deployed in the monitoring area and form a self-organizing network through wireless communication. Sensor nodes can sense, collect and process data and transmit the data to the BS [1]. In recent years, WSN has been widely used in military, environmental monitoring and disaster management [2]. The existing WSN mainly relies on a limited battery supply and its life cycle depends entirely on its battery power. Because WSNs are often deployed in remote or harsh environments, these batteries are almost impossible to replace. Therefore, the energy optimization of WSN has been a hot topic of research.

X. Sun et al. (Eds.): ICAIS 2020, LNCS 12240, pp. 156–165, 2020.
https://doi.org/10.1007/978-3-030-57881-7_14

The most popular solution among WSN's many energy-saving solutions is: clustering technology. This technique divides sensor nodes into multiple clusters. Each cluster in the network has a unique cluster head node. Ordinary nodes will send the perceived data to the cluster head node [3]. Then, the cluster head node sends the data to the BS after collecting and aggregating data. Clustering technology has many benefits, including scalability, energy efficiency, and reduced routing latency. In the past research, many clustering methods have been proposed. The common goal of these clustering technologies is to save energy. The operations in a clustering protocol are usually divided into three phases: CH selection, cluster formation, and data transmission. The main part of each method is the CH selection algorithm that defines the energy efficiency of the network [4].

1.1 Clustering

LEACH protocol-full name is "Low Energy Adaptive Clustering Hierarchy". It was proposed by Heinzelman et al. In 2000. This protocol is WSN's first dynamic clustering protocol and is considered the basis of other advanced clustering protocols in WSN. Generally, it is a layered, random, and distributed single-hop protocol. LEACH is executed continuously in rounds. Each round can be divided into two phases: the establishment phase of the cluster and the stable phase of transmitting data. The cluster establishment process can be divided into 4 stages: selecting the cluster head node, broadcasting the cluster head node, establishing the cluster head node, and generating a scheduling mechanism [5].

TEEN (Threshold sensitive Energy Efficient Sensor Network protocol) is different from all nodes in LEACH that always have data to transmit. TEEN is specifically designed for applications that should send data to the BS when certain events occur. TEEN is the first hierarchical WSN routing protocol for responsive networks. TEEN works basically the same way as LEACH, except that after each re-clustering, the cluster head node needs to broadcast the following three parameters to the members in the cluster:

1) Characteristic value: The physical parameter of the data that the user cares about.
2) Hard Threshold (HT): The absolute threshold of the characteristic value of the monitoring data [6]. When the feature value monitored by the node exceeds this threshold, the transmitter is started to report this value to the cluster head node.
3) Soft Threshold (ST): Monitors a small range of characteristic values to change the threshold to trigger the node to start the transmitter to report data to the cluster head.

When the characteristic value of the node's sensing data exceeds its hard threshold for the first time, the node starts the transmitter to send the sensing data, and this characteristic value is also stored in the node's memory.

The node will start the next data transfer if and only if the following two conditions are met:

1) The feature value currently sensed is greater than the hard threshold.
2) The difference between the current feature value and the previous feature value is greater than or equal to the soft threshold.

However, there are some problems in TEEN. First, users will not get feedback from the area of interest until the threshold is reached. As a result, some nodes may die without the user being aware of their deaths because it has not received feedback.

1.2 Genetic Algorithm

Genetic algorithm is a kind of randomized search method that evolved from the evolutionary laws of the biological world (the genetic mechanism of survival of the fittest). Compared with traditional search algorithms, genetic algorithms have the following characteristics:

1) Genetic does not have too many mathematical requirements for the optimization problem to be solved. Genetic algorithms can handle any form of objective function and constraints, whether linear or non-linear, discrete or continuous, or even mixed search space.
2) The ergodicity of evolutionary operators makes genetic algorithms very efficient for global searches in the sense of probability, while traditional optimization methods transfer to better points by comparing neighboring points, thereby achieving a local search process that converges;
3) Using probabilistic optimization method, it can automatically obtain and guide the optimized search space, and adaptively adjust the search direction.
4) Genetic algorithms can provide great flexibility to mix domain-independent heuristics for various special problems, thereby ensuring the effectiveness of the algorithm [7].

2 System Model

2.1 Energy Model

The energy consumption of a classic wireless sensor network is mainly composed of two parts: information reception energy consumption and information transmission energy consumption.

$$E_{(\mathrm{Si})}(L,\ d) = \begin{cases} L \times E_{elec} + L \times E_{fs} \times d^2, (d < d_0) \\ L \times E_{elec} + L \times E_{mp} \times d^4, (d \geq d_0) \end{cases} \tag{1}$$

$$E_{(Ri)} = L \times E_{elec} \tag{2}$$

$$d_0 = \sqrt{E_{fs}/E_{mp}} \tag{3}$$

Where L is the total amount of data to be transmit or receive. E_{elec} is the energy consumed to send one bit of data; d is the distance between two sensor nodes. The signal energy consumption model is divided into two categories according to distance: free space model and multi-path attenuation model [8]. When the transmission distance is less than the threshold d_0, the free space energy consumption model is adopted in the communication mode, while the multi-path attenuation model is adopted in the other way. d_0 is a constant, and the value depends on the network environment.

2.2 System Model

We assume that there are N sensor nodes randomly distributed in the area of M × M. BS has infinite energy and is located in the center of the area.

1) Each node is equipped with a GPS device and knows its own coordinates.
2) After node deployment is completed, all nodes become static nodes. Each node can belong to only one cluster.
3) Each node has local information, including its own unique ID, CH node ID, information collection and transmission rounds, remaining energy level, and distance to its neighbors.
4) The nodes have the same structure, the same initial energy, the same computing and communication capabilities, and all have the ability to collect, calculate, store, and fuse data.

The whole operation of WSN is divided into several rounds according to time. In each round, the ordinary sensor node sends the sensed data to its cluster head [9].

3 Design and Analysis of Algorithms

3.1 Design of Clustering Algorithm

After the node arrangement is completed, the WSN forms C concentric circles with BS as the center. As shown in Fig. 1, the radius R of the innermost first circle of the concentric circles is equal to the threshold d_0. We call this circular area 0. In region 0, we believe that according to formula 1, nodes within the BS threshold (the threshold of BS in this article is equal to the threshold of the node) directly transmit data to the BS. The nodes in this area do not need to be clustered. Clustering instead wastes extra energy [10].

The radius of the 2nd to Cth concentric circles in the inner layer is 1/2 d_0. This is because during the data transmission phase, data will be transmitted from the nodes in the Cth concentric circle in the outermost layer to the nodes of the inner concentric circle in order. During data transmission, the transmission distance does not exceed the threshold d_0 [11].

Next, we determine the extent of each cluster within the area of each concentric circle. As shown in Fig. 1, in the first concentric circle (herein referred to as Zone 1, the concentric circles in the following and so on), the nodes in the area of points A and B form a new cluster. Point A is on the inner perimeter of the first concentric circle, and point B is on the outer perimeter of the first concentric circle. And the Euclidean distance between two points is d_0. Similarly, another cluster can be divided by points C and D, and the Euclidean distance between the two points is also d_0. In the same way, new clusters can be divided according to points E and F within the area of Zone 2. If the area is not enough after the division is completed, a small cluster is formed separately. Such as Cluster D in Zone 1 [12].

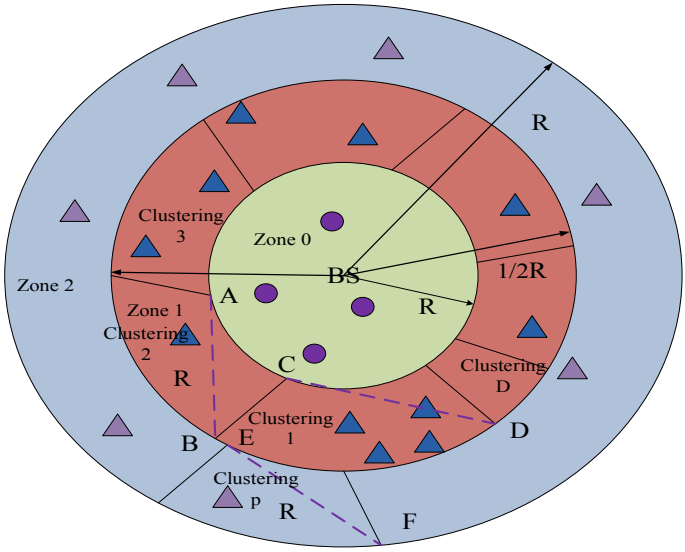

Fig. 1. Division of clusters

3.2 Cluster Head Selection Based on Genetic Algorithm

WSN clustering is a multi-objective optimization problem. This paper proposes a genetic algorithm based on the geographical distribution of nodes to solve the clustering optimization model [13].

Step 1: population initialization
Geographical Distribution Genetic Algorithm (GDGA) is a process to obtain the optimal chromosome through the process of chromosome selection, crossing, and mutation based on the initial population [14].

Geographical Distribution Genetic Algorithm (GDGA) is a process to obtain the optimal chromosome through the process of chromosome selection, crossing, and mutation based on the initial population. The algorithm uses two-dimensional gene coding. The number of chromosomes is first defined as Q, and each chromosome has P bits (the entire WSN is pre-divided into P clusters) in binary code. As shown in Table 1:

Table 1. Chromosome coding

Node number	1	2	3	4	5	6	….	N
CM or CH	1	0	0	0	0	0	…	1

One dimension of a chromosome is represented by node numbers 1 to N. Two-dimensional indicates whether this node is selected as the cluster head. 0 indicates

that the node is a common node (CM), as represented by a cluster head (CH). For example, the first digit of the gene in Table 1 indicates that node 1 is selected as the cluster head [15].

3.3 Fitness Function

The calculation formula for the fitness function of the CH node selection needs to define some important parameters as the basis for the CH selection.

The fitness function of the CH node can be expressed as

$$F_t = \mu_1 \times E_{residual} + \mu_2 \times \frac{1}{D_{toCH_i}} + \mu_3 \times \frac{1}{D_{to\text{Relay}}} \tag{4}$$

$$\mu_1 + \mu_2 + \mu_3 = 1 \tag{5}$$

where μ_1, μ_2, and μ_3 are the control parameters of the three parts of the fitness function. In this algorithm, the equivalent values of the three control parameters represent the same effect of the three influencing factors. $E_{residual}$ represents the remaining energy of the node. The more the remaining energy, the more likely it is to be selected as the cluster head. D_{toCH_i} represents the distance from other nodes in the cluster to that node. The smaller the distance, the more likely it is to be selected as the cluster head. $D_{to\text{Relay}}$ represents the distance from this node to the previous relay node. The smaller the distance, the more likely it is to be selected as the cluster head [16].

3.4 Operation of Genetic Algorithm

Crossover and mutation operations are performed on the current genes to generate new genes. The operations are shown in Fig. 2 and Fig. 3.

Node number	1	2	3	4	5	6	·····	N
CM or CH	1	0	0	0	0	0	·····	1

Gene crossover

Node number	1	2	3	4	5	6	·····	N
CM or CH	1	0	0	0	0	0	·····	1

Fig. 2. Gene mutation

4 Simulation Experiment and Analysis

In order to verify the effectiveness of the GDGC algorithm, we used Matlab 2016 to compare its three algorithms, LEACH, ERP, and HEED. Assume that 100 wireless

Fig. 3. Gene mutation

sensor nodes are randomly arranged in a 50 × 50 area. The position of BS is at the center coordinate of (100, 100). We compare the overall network energy consumption, node life, remaining energy (the number of remaining nodes) and other indicators [17].

The simulation parameters are shown in Table 2.

Table 2. Simulation input parameters

Parameters	Value
Sensor field region (m2)	(100 * 100)
BS location	(50, 50)
Number of sensors	100
Initial energy of the node (J)	200
Data packet length (bits)	2048
Number of iterations	100
E_{elec} (nJ/bit)	50
E_{amp} (pJ/bit)	0.0012
E_{fs} (pJ/bit)	10
d_{th} (m)	30

The simulation results are shown in Fig. 4. After the four protocols run the same 100, 300, and 500 rounds, the remaining energy of the WSN running the GDGA algorithm is higher than the other three algorithms.

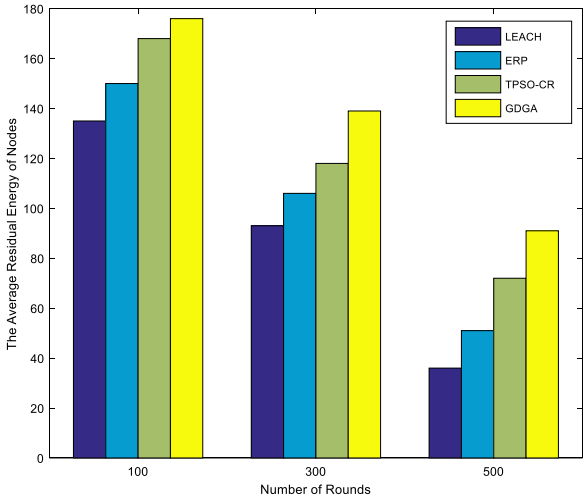

Fig. 4. The average remaining energy of the nodes when running the same round

As shown in Fig. 5, the number of nodes surviving is compared after running the four algorithms for several rounds. We can see that the GDGA algorithm has more nodes than the other three algorithms. This proves that our algorithm can extend the maximum life of WSN.

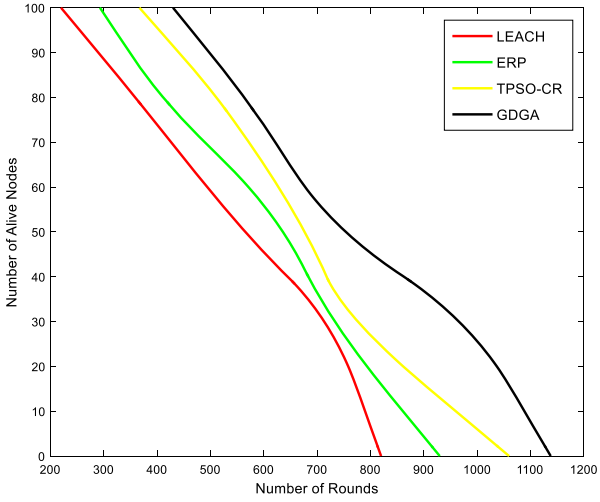

Fig. 5. Comparison of number of alive nodes

We cluster the four algorithms and then execute the greedy algorithm as the routing algorithm. This way we can compare the data transmission capabilities of the four algorithms (Fig. 6).

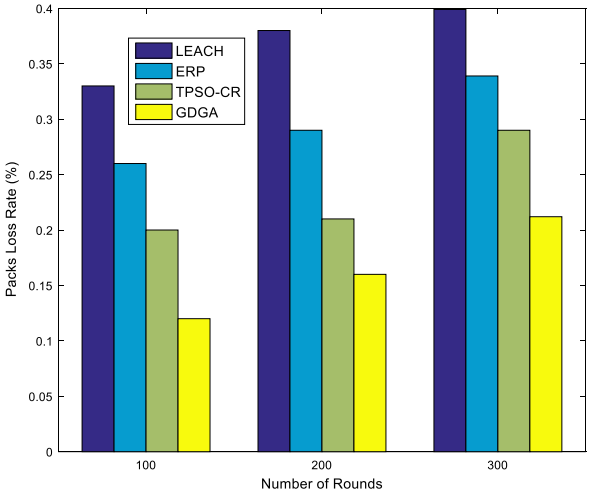

Fig. 6. Packs loss rate

5 Conclusion

In this paper, in the traditional clustering algorithm, a genetic algorithm based on the geographical distribution of sensor nodes is introduced to obtain the GDGA algorithm, which can cluster faster. At the same time, a fitness function based on multiple factors is constructed to balance the overall network performance Power consumption, to a certain extent, can avoid the situation that the node's transmission information exceeds the node threshold, and extend the network life as much as possible. The simulation results show that the GDGA algorithm and the classic Compared with LEACH, ERP, and TPSO-CR, it can not only reduce the average energy consumption of the network, increase the number of remaining nodes, extend the network life cycle, but also be superior to the other three protocols in terms of performance.

Acknowledgments. This work was partially supported by the National Natural Science Foundation of China (No. 61771410, No. 61876089), by the Postgraduate Innovation Fund Project by Southwest University of Science and Technology (No. 19ycx0106), by the Artificial Intelligence Key Laboratory of Sichuan Province (No. 2017RYY05, No. 2018RYJ03), by the Zigong City Key Science and Technology Plan Project (2019YYJC16), by and by the Horizontal Project (No. HX2017134, No. HX2018264, No. E10203788, HX2019250).

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**(17), 393–422 (2002)
2. Cardai, M., Du, D.Z.: Improving wireless sensor network lifetime through power aware organization. Wireless Netw. **3**, 333–340 (2005)
3. Chen, M.-T., Tseng, S.-S.: A genetic algorithm for multicast routing under delay constraint in WDM network with different light splitting. J. Inf. Sci. Eng. **21**(8), 85–108 (2005)

4. Bhondekar, A.P., Vig, R., Singla, M.L., Ghanshyam, C., Kapur, P.: Genetic algorithm based node placement methodology for wireless sensor networks. Proc. Int. Multiconf. Eng. Comput. Sci. **1**, 18–22 (2009)
5. Wang, P., He, Y., Huang, L.: Near optimal scheduling of data aggregation in wireless sensor networks. Ad Hoc Netw. **4**, 1287–1296 (2013)
6. Kulkarni, R.V., Forster, A., Venayagamoorthy, G.K.: Computational intelligence in wireless sensor networks: a survey. IEEE Commun. Surv. Tutor. **13**(1), 68–96 (2011)
7. Gazen, C., Ersoy, C.: Genetic algorithms for designing multihop lightwave network topologies. Artif. Intell. Eng. **13**, 211–221 (1999)
8. Jiang, H., Zhang, T., Zhao, X., et al.: Large data based anomaly detection mechanism for power information network traffic. Telecommun. Sci. **33**(3), 134–141 (2017)
9. Han, W., Tian, Z., Huang, Z., Zhong, L., Jia, Y.: System architecture and key technologies of network security situation awareness system YHSAS. Comput. Mater. Continua **59**(1), 167–180 (2019)
10. Li, R., Zhang, L., Li, H., et al.: Summary of network anomaly traffic detection based on entropy. Appl. Comput. Syst. **26**(6), 36–39 (2017)
11. Gu, Y., He, T.: Dynamic switching-based data forwarding for low-duty-cycle wireless sensor networks. IEEE Trans. Mob. Comput. **10**(12), 1741–1754 (2011)
12. Xu, G., Wang, Z., Zang, D., et al.: Data center network anomaly detection algorithm based on link state database. Comput. Res. Dev. **55**(4), 815–830 (2018)
13. Rout, R.R., Ghosh, S.K.: Adaptive data aggregation and energy efficiency using network coding in a clustered wireless sensor network: an analytical approach. Comput. Commun. **40**, 65–75 (2014)
14. Hong, M., Bei, Y.X.: Network anomaly data detection model based on intrusion feature selection. Modern Electron. Technol. **40**(12), 69–71 (2017)
15. Ying, W.: Wireless network traffic anomaly data detection simulation. Comput. Simu. **34**(9), 408–411 (2017)
16. Kaiwartya, O., Kumar, S., Abdullah, A.H.: Research on time synchronization method under arbitrary network delay in wireless sensor networks. Comput. Mater. Continua **61**(3), 1323–1344 (2019)
17. Zhang, H., Yi, Y., Wang, J., Cao, N., Duan, Q.: Analytical model of deployment methods for application of sensors in non-hostile environment. Wireless Person. Commun **97**, 389–399 (2017)