International Journal of Future Generation Communication and Networking
Vol. 8, No. 3 (2015), pp. 179-190
http://dx.doi.org/10.14257/ijfgcn.2015.8.3.17

The Cluster Head Preferred Hierarchical Clustering Routing Protocol Based on G-Means in Wireless Sensor Networks

Tianshu Wang, Gongxuan Zhang, Xichen Yang and Yang Lv

Department of Computer Science and Engineering, Nanjing University of Science and Technology, China
wangtianshu122@163.com; gongxuan@ njust.edu.cn;
yxcwudi@ gmail.com; lvyang1128@gmail.com

Abstract

In hierarchical structures determined by traditional routing protocols of wireless sensor networks, clustering is not structured and the networks prone to generate scatters, leading to some nodes die quickly. So this paper presents a cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC). It uses G-Means algorithm to detect each clustering structure, so that each cluster is Gaussian distribution and it avoids the generation of scatters. At the same time, the cluster head preferred algorithm proposed in this paper gives the corresponding weight for each node of all clusters, thus select a suitable cluster head node for every cluster. Experimental results show that compared to traditional routing protocols, the death rate of nodes in a wireless sensor network which uses GHPHC protocol is more slow, and has a longer life cycle.

Keywords: *Wireless Sensor Networks, hierarchical clustering, routing protocol, G-Means, Gaussian distribution, and Cluster Head Preferred*

1. Introduction

Wireless sensor networks (WSNs) are defined as wireless networks consist of randomly distributed tiny, large quantity and cheap nodes through self-organization [1]. These tiny nodes are some stationary or mobile sensors. They cooperate with each other to collect process and transmit information [2]. Because of the defects of hardware and software in the wireless sensor networks, the sensor nodes can't provide an adequate amount of energy for computing, storage and communication functions, constantly. Therefore, it's necessary to in-depth research routing protocols, use efficient routing algorithms and optimize allocation method of nodes for slowing the dead speed of a single node and then maximizing the survival time of the entire wireless sensor network [3].

Currently, the study of routing protocols which aim to maximize the extension of the network lifetime has achieved some results. Heinzelman et al. [4] proposed the LEACH (Low Energy Adaptive Clustering Hierarchy protocol). LEACH is one of the most typical hierarchical routing protocols. Its main idea is to select the head node of each cluster randomly in the form of round. Then the energy load of the entire network will be distributed equally between all of the nodes. So it can reduce energy consumption and extend the life cycle of the network. However, the selection of the cluster head nodes is random in the protocol. This selection may cause uneven distribution of clusters in the network and then resulting energy imbalance. Against this shortcoming, Younis et al [5] proposed HEED (Hybrid Energy—Efficient Distributed) protocol. The selection of the cluster head nodes mainly depends on residual energy and communication cost within a cluster these two parameters. The protocol divides all the nodes into several levels according to the proportion P of residual energy in initial energy. Higher-level nodes consider themselves as the cluster heads and announce other low-level nodes to join the

ISSN: 2233-7857 IJFGCN

corresponding clusters by sending a broadcast message. However, the HEED can't completely avoid uneven distribution of network clusters. After determining the cluster head nodes, the remaining nodes select the appropriate cluster head node based on the strength of the signal. If there are too many nodes near a certain head, this head node will cost too much energy and then speed up the death of this node. If a certain head has too little neighboring nodes, it will cause excessive waste of energy.

Considering these shortcomings, Zhong Shan et al [6] presented a low energy consumption clustering routing protocol based on K-Means. Through K-Means clustering algorithm, all nodes in the network are divided into several clusters of similar size based on the distance similarity. This method optimizes topology of the network, reduces communication power consumption intra-cluster and avoids too much energy consumption of some nodes due to excess load. However, the K-Means algorithm has randomness in the choice of initial centers. If the initial value selection is not appropriate, effective clustering results can't be got. Thus affecting the energy distribution of the round and shortening the lifetime of the network.

Based on the analysis above, this paper proposed a cluster head preferred hierarchical clustering routing protocol based on G-Means [7] (GHPHC). G in G-Means algorithm represents the Gaussian distribution [8] which is also called normal distribution. G-Means clustering algorithm is adding the statistical judgment thinking of Gaussian distribution on the basis of K-Means algorithm. The statistical test of normal distribution can determine whether all K-Means clusters are in line with the Gaussian distribution, and then determine the clustering structure of the network. GHPHC protocol makes nodes clustering in the way of normal distribution every round and takes the cluster head preferred algorithm proposed in this paper to select the appropriate cluster head in each cluster. This method optimizes the network topology, realizes the nodes full coverage, avoids scatter generation, reduces energy consumption of each node, and thus extends life cycle of the network.

2. Prior Knowledge

2.1. Network Topology

Routing protocols can be divided into planar routing protocols and hierarchical routing protocols according to the network's topology, the status and functions of each node [10].

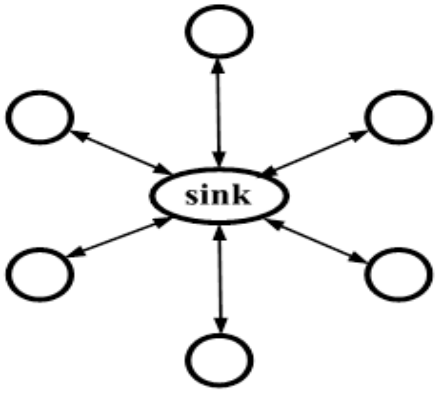


Figure 1. A Wireless Sensor Network based on Planar Routing Protocol

The wireless sensor network based on Planar routing protocol is shown in figure 1, it contains a sink node and some other same nodes. These other nodes in the network can only communicate with the sink node, and they have the same status and functions. Typical planar routing protocols contain Flooding protocol, Crossing protocol, SPIN

protocol and so on. These planar routing protocols are simple and have a good robustness. But sink node has excessive computing and communication pressure, leading to the rapid depletion of the battery energy, and eventually death of the sink node will result in destruction of the entire network.

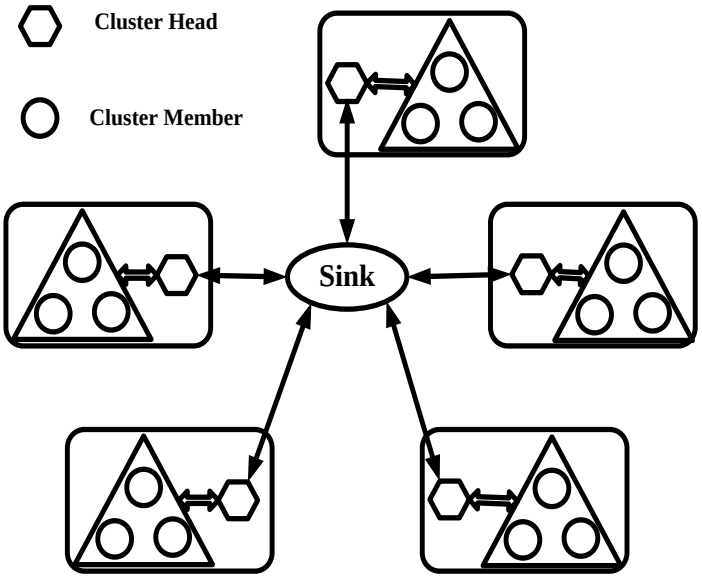


Figure 2. A Wireless Sensor Network based on Hierarchical Routing Protocol

The wireless sensor network based on hierarchical routing protocol is shown in figure 2. Its main idea is dividing the whole network into a plurality of relatively small collection of network nodes. Each collection is called a cluster. In general, each cluster has a node which is selected as the cluster head according to certain rules. The other nodes within the cluster are called member nodes [11], which is only responsible for gathering information. In each cluster, the head node is used to manage or control the collection of nodes and in charge of collecting the data which are gathered by member nodes and doing aggregation. Finally, it sends the results to the base station or sink node. Typical hierarchical routing protocols contain LEACH protocol, TEEN protocol, PEGASIS protocol [12] and so on. These protocols have the characteristics of scalability and efficiency and put forward the idea of data fusion [13] in cluster heads, reduce power consumption of sink node and decrease the transmission distance between each peer nodes. Finally, it greatly improves network stability and energy efficiency. Therefore, hierarchical routing protocol is more advantageous than planar routing protocol [14]. The cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC) proposed in this paper is built on the hierarchical routing protocol. And combining G-Means clustering algorithm, GHPHC protocol divides the network into multiple clusters in a most reasonable way. Finally, it realizes the network optimization allocation, extends the life cycle of the whole network.

2.2. LEACH based on the "Round" Thinking

The cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC) is based on the "round" thinking of LEACH. In fact, the most appropriate nodes are selected as the cluster heads each round. After determining the distribution of each cluster, cluster members send perception information to the cluster heads and

consume some energy. After receiving information and perform corresponding data aggregation algorithm, cluster heads consume a certain amount of energy.

2.3. Gaussian Distribution Test Algorithm: Gaussian(X)

The Gaussian distribution test algorithm can be used to measure whether the cluster X obeys Gaussian distribution. X is d-dimension and c represents the center of X. In this paper, X is the set of nodes to be tested in the network. Because every node has the coordinate information in the network, d is equal to 2. The process of Gaussian distribution test is as follows:

1 Choose a critical value α;

2 Initialize two centers and call these two centers as the sub-center of c;

3 Adopt K-Means algorithm ($k = 2$) [15] to calculate the center of cluster X, down them as c_1, c_2;

4 Calculate the d-dimension vector which connects c_1, c_2 and down it as $v = c_1 - c_2$. Project d-dimension data of X in v, the process can be expressed as $x_i' = x_i, v / v^2$. X' is the one-dimensional expression of projecting data to v. Convert X' to standard normal distribution whose expected value is 0 and variance is 1. Then sort out them;

5 $A_*^2(Z)$ represents the i-th order value, calculate the formula $z_i = F(x'(i))$, F is the cumulative distribution function of $N(0,1)$;

6 According to (1), if $A_*^2(Z)$ is in the range of critical value α, X obeys Gaussian distribution. Otherwise, X doesn't obey Gaussian distribution and use $\{c_1, c_2\}$ to replace c.

$$A^2(Z) = -\frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log(z_i) + \log(1 - z_{n+1-i})\right] - n$$

$$A_*^2(Z) = A^2(Z)\left(1 + 4/n - 25/\left(n^2\right)\right) \quad (1)$$

2.4. G-Means Clustering Algorithm

Set X as the collection to be clustered, α as the critical value. G-Means clustering algorithm divides X into several clusters which are equilibrium distribution and obey Gaussian distribution, based on the critical value α. The specific steps are as follows:

1 Define C as the center set, calculate the mean value of all elements in X and set this value as X's center, down it as $\overline{X}$, $\{\overline{X}\} \to \mathrm{C}$.

2 Adopt G-Means clustering algorithm to get new center, and update new center set C, $C \leftarrow$ K-Mean(C, X).

3 Determine the set N_j which is allocated to C_j, $\{x_i \mid \mathrm{class}(x_i) = j\}$. And class() is the judge function.

4 Use the statistical test method to determine whether every node set N_j obeys Gaussian distribution, *Gussian*$(N_j) < \alpha$.

5 If a set N_j obeys Gaussian distribution, keep this set, otherwise, replace C_j to two centers and go to step 2 until no more centers generate.

3. GHPHC Algorithm

This paper proposed a cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC), and the protocol flow chart is shown in Figure 3.

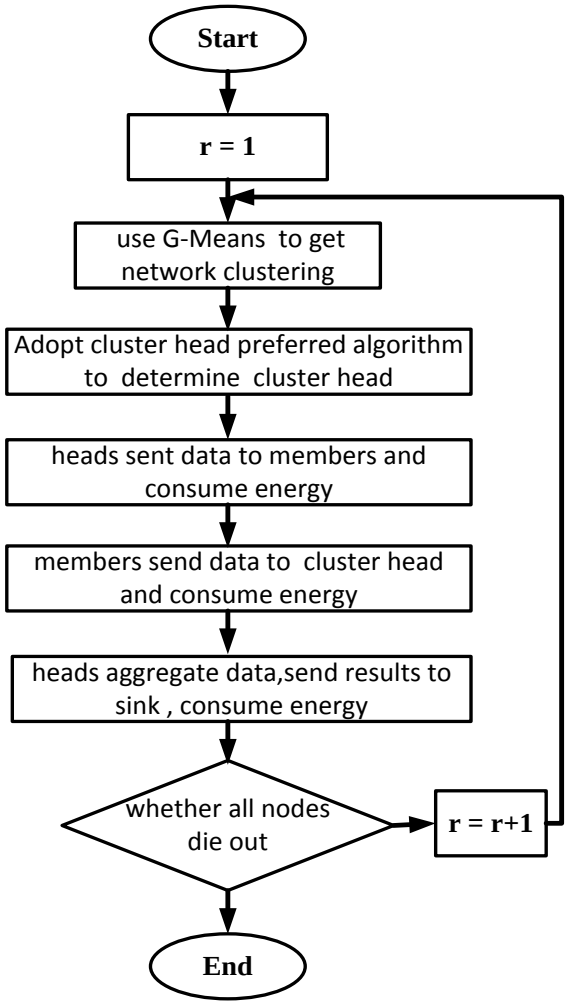


Figure 3. GHPHC Protocol Flow Chart

GHPHC considers the "round" thinking of LEACH protocol. In this figure, r represents the round number. In each round, G-Means clustering algorithm is used to get the network clustering. And the cluster head preferred algorithm proposed in this paper is adopted to determine the cluster head node of each cluster. After determining the hierarchical clustering structure of the network, cluster head nodes first send information to cluster members. In this process, the cluster head nodes will consume some energy. After cluster members receive messages, they send gathered information to their cluster head nodes. This process also needs to consume energy. Finally, the cluster head nodes implement data aggregation, sent results to sink node and consume some energy.

3.1. Use G-Means to Get Network Clustering

In every round of the cycle, GHPHC protocol uses G-Means clustering algorithm to get the network clustering structure. There are n nodes in the network, the coordinate of every node is (x_i, y_i). If the coordinates of any two nodes A, B are (x_A, y_A) and (x_B, y_B), then the Euclidean distance[16] between A and B is d_{AB}. The calculation method is as shown

in (2). According to (3), calculate mean distance $\overline{d}$ of the Euclidean distances between all nodes and put $\overline{d}$ into the center C.

$$d_{ij} = \sqrt{(x_A - x_B)^2 - (y_A - y_B)^2} \quad (2)$$

$$\overline{d} = \sum_{i=1,\ j=1}^{n} d_{ij} / n(n-1) \quad (3)$$

Set C as the initial center set of K-Means algorithm, call K-Means function to get new center set C'. Select the center for every node in the network, basing on the distance to every center in C'. That is to say allocate corresponding set of nodes for every center in C', down every set of nodes as N_j.

Adopt the Gaussian function to determine whether every set of nodes N_j obeys Gaussian distribution. If a set N_j obeys Gaussian distribution, keep this set and its corresponding center C_j. Otherwise if a set N_j don't obey Gaussian distribution, call K-Means function ($k = 2$) for N_j to get new centers C_{j1}, C_{j2} and replace C_j in C' to update center set to C'. Return to last loop until all the sets of nodes N_j obey Gaussian distribution and determine the N_j as the clustering in network.

3.2. Determine Cluster Heads by Cluster Head Preferred Algorithm

By G-Means algorithm, the network is divided into clusters of specific structures. How to choose a suitable cluster head node for each cluster set is the key factor of extending the network life cycle. The cluster head preferred algorithm proposed in this paper can efficiently select cluster heads, optimize network structure as far as possible, and then prolong the network lifetime.

The Figure 4 shows the process of cluster head preferred algorithm. Before running the cluster head preferred algorithm, the cluster nodes sort according to the merits order. The merits order here is based on the sum of Euclidean distance value which is between the node and all other nodes in the cluster. And if the smaller the sum value, the better this node is. Otherwise this node is worse. The 'Store' matrix stored each node's weight information of some cluster in the network. The first column of this matrix represents the nodes' numbers of this cluster. And the second column indicates the corresponding weights of these nodes. In the first cycle, theoretically the first node is the worst selection to be chosen as cluster head node, so the weight of this node is assigned to 1. Then enter the second cycle, remove the worst node, get the worst node in the remaining cluster nodes, add 1 to weight of this node, and until the end of the cycle by parity of reasoning. In addition, if a node is selected as cluster head node too many times, the node needs to be lowered weight. If a node's energy is less than the threshold, then lower its weight. And if a node's energy is less than the half of the threshold, continue lowering the weight.

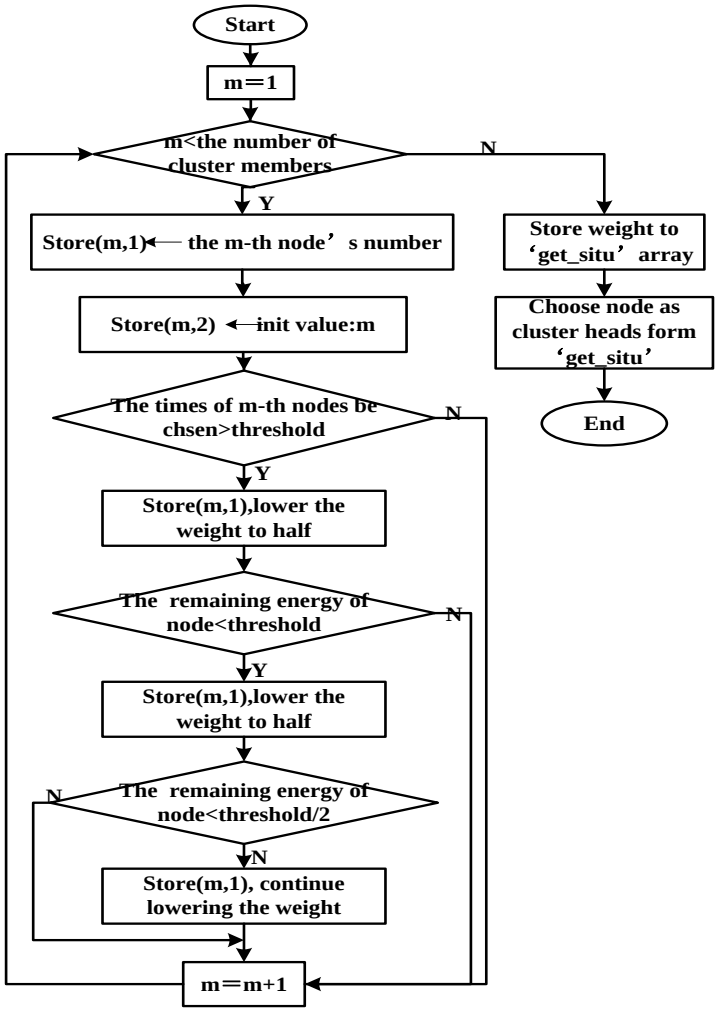


Figure 4. Process of Cluster Head Preferred Algorithm

4. Algorithm Simulation

This paper uses MATLAB to simulate a wireless sensor network. And the simulation environment is an area of 100*100m2 where 100 nodes are randomly distributed. Each node has 0.5J energy. The coordinate of sink node is (50, 50).

The Figure 5 shows the cluster distribution in one of the rounds in traditional LEACH protocol. The empty circles represent cluster member sensor nodes and the solid circles represent cluster head sensor nodes. There are only three clusters in the network of Figure (a). One of these three cluster heads is connected to more than 40 nodes, and the other two heads is also connected to at least 20 nodes. Figure (b) is the cluster distribution of another round in LEACH network. The network contains 19 clusters and two clusters contain only the cluster head node. There are several cluster head nodes connect only 1-3 cluster members. Therefore, the traditional LEACH protocol has great randomness about selecting cluster head nodes. In the case that the number of cluster heads is selected too much, as shown in (b), at least five cluster heads connect only one, two or even no cluster members. These nodes assume duties as the cluster head nodes, but did not play the data forwarding and data fusion capabilities, resulting in energy waste. If the number of cluster heads is selected too little, then individual cluster head node will connect excess nodes. In this case, a substantial energy will be loss and nodes will die out rapidly.

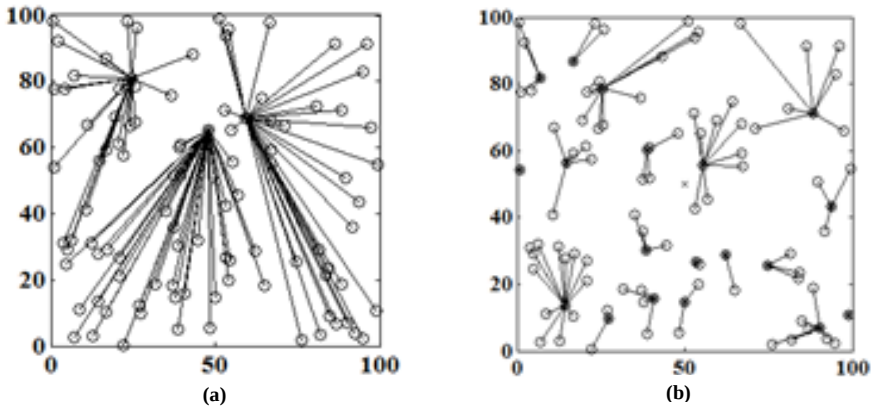


Figure 5. Cluster Distribution of Traditional LEACH

The Figure 6 shows the cluster distribution in cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC). In the network there are 100 nodes, the number of dead nodes is 0 and the number of network clusters is 10. Each cluster head node connects with almost an equal number of cluster members. The number of selected clusters per round will not change to selected by cluster head preferred algorithm can better forward the information and do data aggregation. Then it ensures that there will not be nodes which will have too much energy dissipation because connecting too many nodes every round.

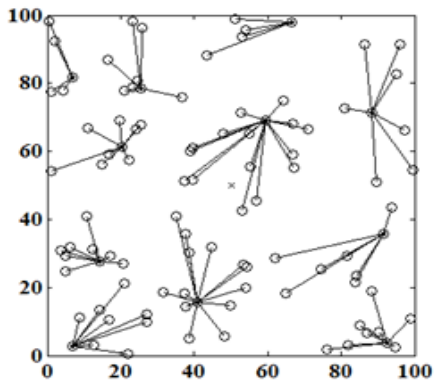


Figure 6. Cluster Distribution of GHPHC

From Figure 5 and 6 it can be seen that compared to traditional LEACH protocol, the clustering distribution is more structured and the distribution of nodes is more uniform in GHPHC proposed in this paper. It is based on G-Means clustering algorithm. Each network clusters are in line with Gaussian distribution. Then it avoids the generation of scattered nodes to prevent the waste of energy, and each cluster contains almost the same number of nodes to ensure the energy balance to prevent individual cluster head node from losing substantial energy. Meanwhile, selecting out cluster heads according to the cluster head preferred algorithm proposed in this paper, allocates the optimal cluster head node for each cluster based on the calculated weight. So this method optimizes the clustering distribution of network, reduces the energy dissipation of nodes when communicating with each other, reduces power consumption of network, thereby extending the lifetime of the network.

The experimental results shown in Figure 7 represent that in the network which used traditional LEACH protocol, some nodes began to die from 1000 round. From 1100 round, the curve began a sharp decline and nodes began to die rapidly. Finally at 1400 round, there are almost no living nodes in the network. While in the network which adopted cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC), some nodes also began to die from 1000 round. Then the curve has been dropping very gently, and the nodes slowly die. At about 1150 round, the number of survived nodes in GHPHC network has been more than LEACH network. When all nodes die out in LEACH network, there were nearly half of the nodes working in the GHPHC network. Lifetime of LEACH network is only 1510, while the lifetime of GHPHC network is up to 2,000 rounds. The simulation results show that the proposed GHPHC protocol greatly extended life cycle of network, improved the efficiency of the entire sensor network.

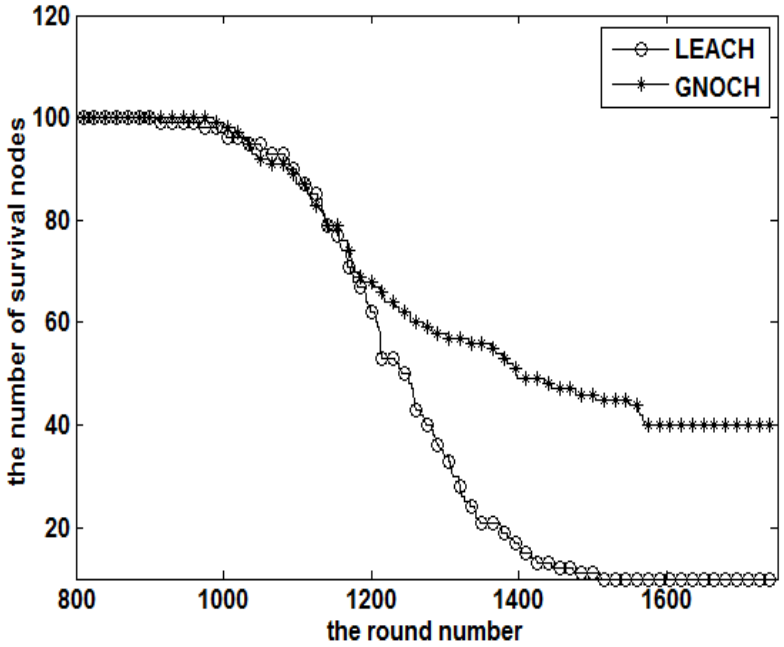


Figure 7. The Round Number and the Number of Survival Nodes

4. Conclusions

This paper proposed a cluster head preferred hierarchical clustering routing protocol based on G-Means (GHPHC).It uses G-Means clustering algorithm to determine the structure of network, and uses cluster head preferred algorithm to assign appropriate cluster head for each cluster, to maximize the network lifetime. Compared with traditional methods, this protocol adopts the idea of Gaussian distribution to optimize the cluster distribution and prevent the generation of scatter. Simulation results show that GHPHC improves efficiency of data-aware and forwarding in wireless sensor networks, and greatly extends network lifetime.

Compared with the traditional protocol, the energy efficiency of GHPHC has certain improvement. However, demise speed of nodes is still a little fast at the initial stage. Therefore, next research will consider joining some other algorithms, to slow down the demise speed of nodes at the initial stage.

Acknowledgements

This paper is supported by the National Natural Science Foundation funded project No.61272420 and the Natural Science Foundation of Jiangsu Province project No.BK2011022.

References

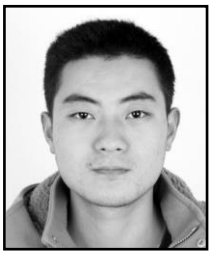
[1] Ren, Feng-Yuan, H.-N. Huang, and C. Lin, "Wireless sensor networks", Journal of software, vol.7, no.14, **(2003)**.
[2] Yick, Jennifer, B. Mukherjee and D. Ghosal, "Wireless sensor network survey", Computer networks, vol.12, no.52, **(2008)**.
[3] Singh, S. Kumar, M. P. Singh, and D. K. Singh, "Routing protocols in wireless sensor networks–A survey", International Journal of Computer Science & Engineering Survey (IJCSES), vol.1, **(2010)**.
[4] W.R. Heinzelman, A. Chandrakasan and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks", Proceedings of the 33rd Hawaii International Conference on System Sciences, **(2000)**; Hawaii, America.
[5] O. Younis and S. Fahmy, "HEED: a Hybrid, Energy-Efficient, Distributed Clustering Approach for Ad Hoc Sensor networks", IEEE Trans. Mobile Computing, vol.4, no.3, **(2004)**.
[6] S. Zhong, G. Wang, X. Leng, X. Wang, L. Xue and Y. Gu, "A Low Energy Consumption Clustering Routing Protocol Based on K-Means", Journal of Software Engineering & Applications, vol.5, no.12, **(2012)**.
[7] G. Hamerly and C. Elkan, "Learning the k in k-means", Proceedings of the Seventeenth Annual Conference on Neural Information Processing Systems, **(2003)**.
[8] Novey, Mike, T. Adali, and A. Roy, "A complex generalized Gaussian distribution—characterization, generation, and estimation", IEEE Transactions on Signal Processing, vol.3, no.58, **(2010)**.
[9] M. M. R. Mozumdar, G. Nan, F. Gregoretti, L. Lavagno and L. Vanzago, "An efficient data aggregation algorithm for cluster-based sensor network", Journal of Networks, vol.7, no.4, **(2009).**
[10] L. Jin, G. Chang, D. Sun, N. Zhou, J. Jia and C. Liu, "Research on hierarchical routing protocol based on optimal clustering head for wireless sensor networks", International Journal of Modeling, Identification and Control, vol.4, no.15, **(2012)**.
[11] H.-Y. Yang, S.-Q. Geng, W.-C. Wu, "A Wireless Sensor Network Routing Protocol for Emergent Event Monitoring", Proceedings of IEEE 11th International Conference on Solid-State and Integrated Circuit Technology, **(2012)**.
[13] W. Linping and C. Zhen, "Improved algorithm of PEGASIS protocol introducing double cluster heads in wireless sensor network", Proceedings of the 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering, **(2010)**.
[14] Y. Feng, S. Tang, G. Dai, "Fault Tolerant Data Aggregation Scheduling with Local Information in Wireless Sensor Networks", TSINGHUA SCIENCE AND TECHNOLOGY, vol.5, no.16, **(2011)**.
[15] Waware, Sanjay, Dr N. Sarwade, and P. Gangurde, "A review of power efficient hierarchical routing protocols in wireless sensor networks", International Journal of Engineering Research and Applications (IJERA), vol.2, no.2, **(2012)**.
[16] X. Yang, "Wear State Recognition of Drills Based on K-means Cluster and Radial Basis Function Neural Network", International Journal of Automation and Computing, vol.3, no.7, **(2010)**.
[17] J. Yang, M, Xu, B and A. Xu, "Multipath routing algorithm based on parametric probability for wireless sensor networks", Proceedings of 2nd International Conference on Intelligent Computation Technology and Automation, **(2009)**.

Authors

Tianshu Wang, she was born in Jiangsu Province, China, in 1989. She is a Ph. D. candidate of School of Computer Science and Engineering in Nanjing University of Science and Technology in China. Her main research interests include embedded system, wireless sensor network, Internet of Things, trusted computing and cloud computing.

Gongxuan Zhang, he was born in Jiangxi Province, China, in 1961. He is a Ph. D., professor, and Ph. D. supervisor of School of Computer Science and Engineering in Nanjing University of Science and Technology in China. His main research interests include Client/Server computing, CORBA technology, Web Service, information security, embedded system, sensor network and distributed computing.

Xichen Yang, he was born in Jiangsu Province, China, in 1989. He is a Ph. D. candidate of School of Computer Science and Engineering in Nanjing University of Science and Technology in China. His main research interests include Image quality assessment, wireless sensor network.



Yang Lv, he was born in Jiangsu Province, China, in 1987. He is a Master candidate of School of Computer Science and Engineering in Nanjing University of Science and Technology in China. His main research interests include embedded system, wireless sensor network, Internet of Things, trusted computing and cloud computing.