# LEACH protocol based on Clustering and Multi-leader Selecting in Wireless Sensor Network

Ping Chen,Yanqiong Zhang,Wenzhan Dai
School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018
E-mail: 314471469@qq.com, yqzhang@zjgsu.edu.cn, dwz@zjsu.edu.cn

**Abstract:** In this paper, we consider the energy consumption problem of LEACH protocol in Wireless Sensor Network. Based on the network clustering analysis and multi-leader selection strategy, we propose applying a simple augmentation to the original protocol in order to solve the energy consumption problem. Finally, some comparative simulations reveal that the proposed approach has good performance in LEACH routing protocol, and can greatly improve the network life cycle and throughput.

**Key Words:** WSN, network partition, multi-leader, LEACH, energy consumption

## 1 Introduction

In recent years, as the important part of Internet of things (IoT), Wireless Sensor Network (WSN) has drawn more and more research attention, and energy consumption is one of the most basic and important problem in WSN for the continued development of smart world[1]. Generally, WSN is composed of many randomly deployed micro sensor nodes, and these nodes can transmit the sensing data to the Base Station (BS), and then BS transmits the related data to the remote control center for further processing. The fundamental energy consumption comes from routing setup and data transmission[2]. However, considering the limited resources of sensor nodes, each node could only get local information, so WSN often establishes a multi-hop routing to forward data. Therefore, how to find a forwarding path with low energy consumption is one of the key concerns in WSN.

Routing protocols have been widely investigated for WSN, and detailed references can be found in [3, 4]. Routing protocols can be divided into flat routing and hierarchical routing according to whether the nodes in the routing process have a hierarchical structure and whether they act as different roles. In hierarchical routing protocols, two-layer is one of the relatively typical structures of WSN. There are two different kinds of nodes accordingly, which are termed as Cluster-Head (CH) nodes and non-Cluster-Head (non-CH) nodes, respectively. The CH node is responsible for the data collection, data fusion and data forwarding in its own cluster. The non-CH node is responsible for collecting the perception data and transmitting the related data to its CH node. Typical hierarchical routing protocols include Low-Energy Adaptive Clustering Hierarchy Protocol (LEACH), Stable Election Protocol (SEP), and so on. In this paper, we just focus on the LEACH protocol [5], which is a well-known and classic dynamic clustering method for sensor network, and could be completely distributed and require no global knowledge of the network.

LEACH protocol randomly selects the CH nodes in a circular method. The energy consumption of the entire network is allocated to each node, so it could reduce the speed of energy consumption and prolong the network life cycle to some extent. The architecture of LEACH is shown in Fig. 1. LEACH is periodically executed, and each execution is called a round. Each round is divided into two different phases, the setup phase and the steady state phase. In the setup phase, WSN generates the CH nodes randomly, meanwhile, the sensor nodes form a number of clusters. In the steady stage phase, CH nodes collect and fuse the data in their own cluster, and then send their data to BS directly. In this protocol, data aggregation is made possible using the clustering scheme, while dynamic re-establishment of the clusters balances the energy consumption over the nodes. However, LAECH selects the CH nodes randomly and does not take into account the network structure. Therefore, the selected CH nodes are likely to be concentrated in an area of the network, it will make there is no cluster head nodes around some sensor nodes. Additionally, some CH nodes might be elected near the network borders or far away from BS, which would lead to a high energy consumption.

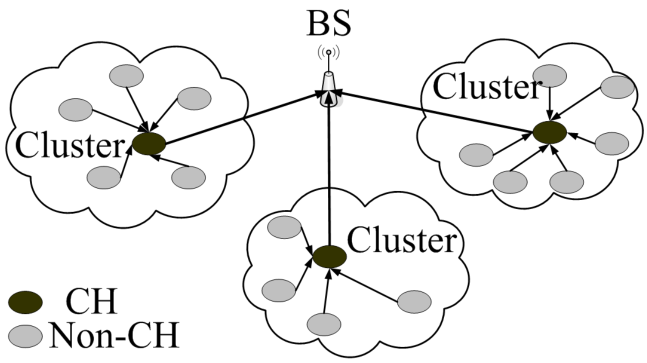


Fig. 1: Architecture of LEACH protocol

In light of the increasing attention to hierarchical routing protocols, some improved routing protocols for LEACH have been proposed. LEACH-C[6] is a centralized routing protocol, which uses a central control algorithm to form clusters in the setup stage. It could generate much better clusters by controlling the location and the number of the CH nodes than LEACH. However, since every node needs to inform its own status, the distance between the node and the BS cannot be too far, which leads to that LEACH-C could not be applied to the large-scale network. SEP[7] extends the network life cycle by introducing some high energy nodes and increases the probability of nodes with higher energy to be

*This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61703368.

selected as the CH nodes, which could improve the network life cycle. However, the CH nodes are selected in SEP without considering the distribution of sensor nodes, and they cannot change with the modifications of the network structure.

In this paper, an improved routing protocol with three levels of network structure is proposed based on network clustering and leader ranking. In the previous discussion, LEACH-C and SEP protocols have been put forward to improve LEACH. However, these protocols do not take into account the distribution of sensor nodes and the network structure, and could not avoid some unreasonable located CH nodes. Motivated by this problem, we study a novel routing solution to improve LEACH under consideration of the network structure. In the new scheme, the network is divided into several sub-areas according to the network structure by using network clustering algorithms, and then the most important node is elected in each sub-areas by using node sorting algorithms, and next LEACH protocol is carried out in each sub-area. There are three kinds of nodes in our new method: Leader nodes, CH nodes and non-CH nodes, only the Leader nodes are responsible for sending data to BS. Simulation results demonstrate that the energy consumption is greatly reduced compared to LEACH, LEACH-C and SEP.

The rest of this paper is organized as follows. Section 2 presents the new scheme with related preliminaries about clustering and sorting algorithms. Section 3 gives some comparative simulations to illustrate the proposed protocol. Finally, the concluding remarks are given in section 4.

## 2 Related work

Consider a WSN composed of $N$ sensor nodes with the following Fig. 2 First Order Radio Model [8] as the wireless communication energy consumption model. Every sensor node can receive and send $k$ bits data, and the consumed energy is related to the communication distance $d$. $E_T$ and

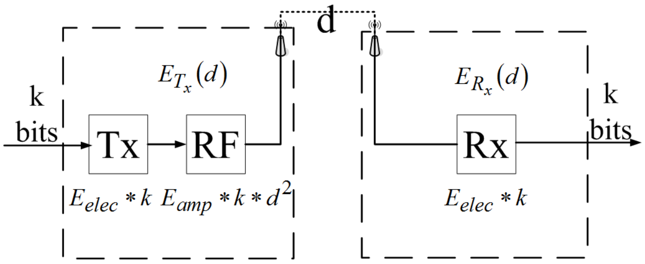


Fig. 2: First Order Radio Model

$E_R$ respectively denote the energy consumed by sending data and by receiving data, which are described as follows:

$$E_T = \begin{cases} E_{elec} * k + E_{amp} * k * d^2, d \leq d_0 \\ E_{elec} * k + E_{amp} * k * d^4, d > d_0 \end{cases} \tag{1}$$

$$E_R = E_{elec} * k \tag{2}$$

where $E_{elec}$ represents the energy consumed by transmitting or receiving 1 bit data, $E_{amp}$ represents the energy consumed of the RF amplifier by sending 1 bit data, and $d_0$ is a given threshold. The sender node determine to choose the free space propagation model or the multi-channel attenuation propagation model according to the communication distance. Note that receiving or sending data is not a low-cost operation, so we need to reduce the number of communication times as far as possible.

In the aforementioned introduction, we know that LEACH protocol is divided into two stages. In the setup stage, the main task is to select CH nodes and form clusters. However most of the existing literature does not take into account the network structure, which would lead to that some CH nodes are far away from BS or some CH nodes are located at the edge of the network or the numbers of non-CH nodes in each cluster are imbalance. In this way, the implementation of LEACH may accelerate energy consumption and reduce the network life cycle. Fig. 3 shows two consecutive selection of CH nodes.

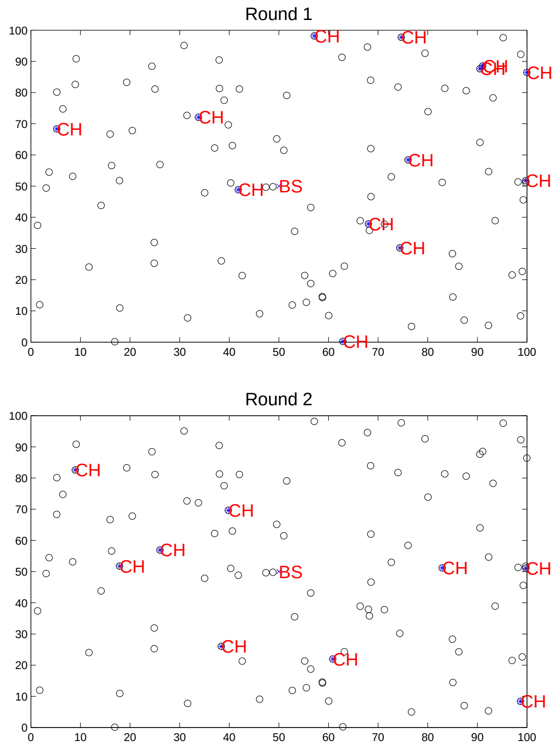


Fig. 3: Contrast diagram of two round of LEACH

In order to reduce the speed of energy consumption and prolong the network life cycle, we propose a method with three levels of network structure based on network clustering and leader node selection strategy. Our proposal needs the following premises:

i BS is fixed with unlimited energy, its position can be broadcast to the whole network.

ii Each node has a unique ID number, the limited energy and fixed position.

iii Every node can perceive its own position and other node's position.

The new scheme architecture is shown in Fig. 4 and the the main idea can be concluded by the following three steps:

1) Divide the whole network into several sub-regions. In this step we use two different partitioning algorithms, which are Spectral Graph Theory and Density and Distance. The purpose is to make the number of nodes in each partition roughly equal.
2) Select an appropriate leader node in each sub-region.

Here we take two sorting methods to select leader nodes, which are Google PageRank and Closeness Centrality. Once the leader node is selected, the node will broadcast its position in this sub-region. The function of leader node is similar to BS, and it is used only for forwarding data.

3) Implement LEACH protocol in each sub-region. First, select a number of CH nodes, and then form several clusters. Non-CH nodes send the perceived data to the CH node and CH node transmits the data to the Leader node. Finally, the Leader node transfers all the data in its sub-region to the BS directly. Repeat multiple rounds in LEACH protocol until energy of all nodes are consumed.

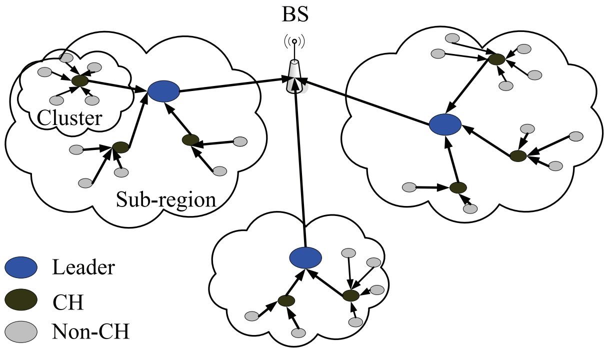


Fig. 4: New architecture

In the following section 3, some comparative simulation experiments are given to illustrate that the new architecture with adding some leader nodes can effectively balance the energy consumption of the network and prolong the network life cycle.

Next, we introduce the network clustering and leader node selection algorithms used in our proposed protocol.

### 2.1 Network Clustering

Clustering is one of the most important problem in unsupervised learning. It involves the division of data structure in the unknown region, which is the basis for further learning. Clustering is essentially the process of dividing data into multiple classes composed of similar data[9]. Here we consider two clustering algorithms: Spectral Graph Theory (SGT) and Density and Distance (DD).

1) Spectral Graph Theory[10]
Denote $A = [a_{ij}]$ as the adjacency matrix of the network, $a_{ij} = 1$ if $d_{ij} \leq k_0$; otherwise, $a_{ij} = 0$, where $d_{ij}$ represents the Euclidean distance between node $i$ and node $j$, and $k_0$ is a given distance threshold. The corresponding Laplace matrix $L = [l_{ij}]$ of $A$ is defined by

$$l_{ij} = \begin{cases} -a_{ij}, i \neq j; \\ \sum_{j=1, j\neq i}^{n} a_{ij}, i = j. \end{cases} \tag{3}$$

The main idea of SGT is as follows:

i) By calculating the eigenvectors and eigenvalues of $L$. The Fiedler eigenvalue is the second-smallest eigenvalue of $L$, and we can use the Fiedler eigenvalue to divide the network into two regions $k_1$ and $k_2$. The nodes corresponding to the non-negative elements are placed in the sub-region $k_1$, others are placed in the sub-region $k_2$.

ii) Compare the normalized Fiedler eigenvalue of each sub-regions, and choose the smallest value corresponding sub-region to carry out the first step. The normalized Fiedler eigenvalue equals Fiedler eigenvalue divided by the number of nodes in this sub-region.

iii) Performs the second step by iteration until the number of sub-regions is $K$.

Fig. 5 shows the partition process from one to eight sub-regions, and there are some uneven sub-regions when $K = 3, 5, 7$.

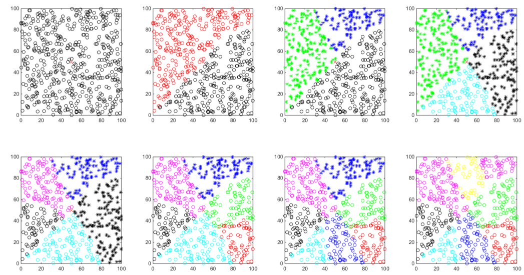

Fig. 5: Network clustering by SGT

2) Density and Distance[11]
There are two main ideas of the DD clustering method:

- The density of cluster center points is higher than its neighbor nodes;
- The distance between the center points is longer.

Each node determines its local density by calculating the Euclidean distance to all other nodes. Then by comparing the local density and distance among a node and its surrounding nodes, the center point nodes can be found, and other nodes cluster through the location of the center point nodes. For node $i$, the local density $\rho_i$ and the distance $\delta_i$ can be calculated as follows:

$$\rho_i = \sum_{j \in N, j \neq i} X(d_{ij} - d_c) \tag{4}$$

$$\delta_i = min_{j:\rho_j > \rho_i}(d_{ij}) \tag{5}$$

where $d_c$ means the distance threshold and $X$ is the indicator function. The nodes with much larger $\delta$ is selected to be the center point. Once the center point is determined, the remaining nodes select the nearest center point to form cluster. Fig. 6 shows the results of seven sub-regions.

SGT is easy to implement, but it needs to assign the number of clusters $K$. If $K \neq 2^{N^*}$, the number of nodes in each sub-region is uneven, then the sub-regions with much more nodes will be relatively faster in energy consumption. Compared to SGT, it's a litter bit more difficult to implement DD, because DD has a more strict prerequisites. Additionally, the distance threshold is a particularly critical parameter in DD, which has a great effect on whether the partition is even. However, because DD algorithm has found the central

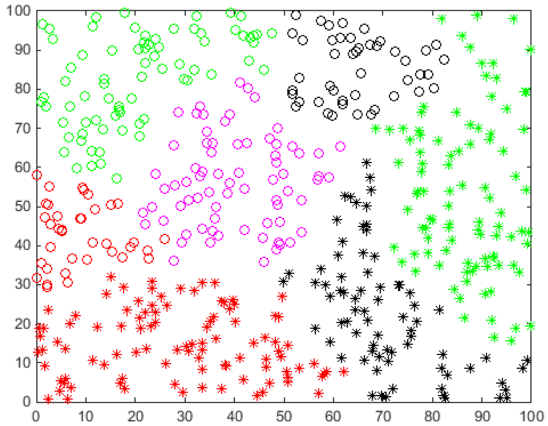

Fig. 6: Network clustering by DD

node in the partition step, it doesn't need to be sorted again. Therefore, in order to achieve the uniform partition for different network structures, the more reasonable clustering algorithm needs to be selected cautiously in practice.

### 2.2 Leader node selection

Recognition and ranking of nodes can help us effectively understand the location and function of nodes, and can also satisfy the upgrading requirements of hierarchical networks and distributed networks in complex networks[12]. In this paper, we consider the following two sorting algorithms: Google PageRank and Closeness Centrality.

1) PageRank algorithm[13]
   PageRank algorithm was firstly used to the ranking of web pages. In WSN, each node can be viewed as a web page and has a corresponding PageRank value. The Leader node can be selected based on the PageRank value. PageRank algorithm has two basic premises:
   i When a node is connected by more nodes, its ranking will be higher.
   ii Higher ranking nodes have higher weights, that is, when a node is connected by a high ranking node, its importance is improved accordingly.

   In order to calculate the PageRank value of each node, we construct a graph $G = (V, E)$, where $V$ is the node set and $E$ is the edge set. For node $i$, the PageRank value is given by

$$P(i) = (1 - d_0) + d_0 \sum_{(j,i)\in E} \frac{P(j)}{O_j} \qquad (6)$$

   where $O_j$ is the out-degree of node $j$, $d_0 \in (0, 1)$ is a damping coefficient and generally can be set to 0.85. The node with the maximum PageRank value is selected as the leader node.
2) Closeness Centrality algorithm (CC)[14]
   In a connected graph $G = (V, E)$, the tightness of a node is a measure of the centrality of the network. The closeness centrality of node $i$ is

$$C(i) = \frac{N - 1}{\sum_{j\in V} d(i, j)} \qquad (7)$$

   where $d(i, j)$ represents the shortest path from node $i$ to node $j$. The larger closeness centrality of a node, the closer it is to all the other nodes. Then the node with the largest closeness centrality value will be selected as the leader node.

In the PageRank algorithm, one node's PageRank value depends on the quantity of its neighbors, so we can choose the most important node based on the network structure. Moreover, the PageRank algorithm is convenient to be implemented and can meet the requirements of massive data processing in big data parallel computing. The CC algorithm is relatively simple, but its computational complexity is too large because it needs to calculate the shortest path between any two nodes, so this algorithm generally could not be applied to the large-scale network.

### 2.3 Performance analysis

Generally, when the energy of a node is decreasing to zero means this node is dead. If all the nodes are dead, which implies the network is not available. In this paper, we compare the death time of the first node and the death time of the whole network among different protocols, and we find that the survival time of nodes in the new scheme is far longer than the traditional way. Additionally, we evaluate the throughput of the whole network, which is defined as the sum of data received by BS in every round, and we find that the throughput of the new architecture is larger and smoother.

In fact, the new architecture essentially extends the original two-layer network structure in LEACH to the three-layer by adding some leader nodes. The sensor nodes are divided into three roles: Leader nodes, CH nodes and non-CH nodes. In the new scheme, the network will be divided into several balanced sub-regions with high probability, which could relatively reduce the communication distance between nodes, thereby reducing the speed of energy consumption. In addition, due to the reduction of the number of nodes in some sub-regions, it could reduce the number of communication times between CH nodes and Non-CH nodes to some extent, which would reduce energy consumption of nodes. Therefore, the proposed new scheme extends the stability period of the network and maintains a balanced energy consumption, thereby prolonging the network life cycle.

## 3 Simulation

In this section, simulation experiments are given to illustrate the performance of the proposed protocols in this paper.

To verify the effectiveness of the new scheme, we give a comparative study about the performance of our method with the traditional routing protocols, such as LEACH, LEACH-C and SEP by 200 times Monte Carlo experiments. In this test, we ignore the effect of the wireless channel interference. Assume the BS node is located in the center position, and the initial energy of each node is set to $0.5J$.

Scene 1: 500 nodes are randomly distributed in the region of $100m * 100m$, and they are divided into 4 sub-regions by Spectral Graph Theory. NEW-1 uses the CC algorithm and NEW-2 uses the PageRank algorithm for leader nodes selection.

Fig. 7(A) shows the variation of the survived nodes in our proposed new scheme and the traditional routing protocols. The first dead node appears when LEACH/LEACH-C/SEP is executed about 700-1300 times. When these three protocols severally run about 1100, 1500 and 2000 times, all of the nodes in the network would be dead. However, the first dead node appears when the new protocol is executed about 1500 times, and all of the nodes die when it runs about after 2300 times. By comparing with the traditional routing protocols, we can obtain that the new scheme is more efficient. In addition, Fig. 7(A) indicates that PageRank algorithm is slightly better than Closeness Centrality algorithm with an increasement about 5 percent of the network lifetime. From Fig. 7(B), we can see the changes of energy consumption. The energy consumption of LEACH protocol is stable in the early stage, and with the changes of CH nodes and their corresponding clusters, the energy is exhausted when it runs about 1000 times, and the energy approaches to zero when LEACH-C and SEP run about after 1500 times. However, the energy consumption curve in the new architecture is relatively more stable and smoother, the network is still available until after about 2000 times of executions.

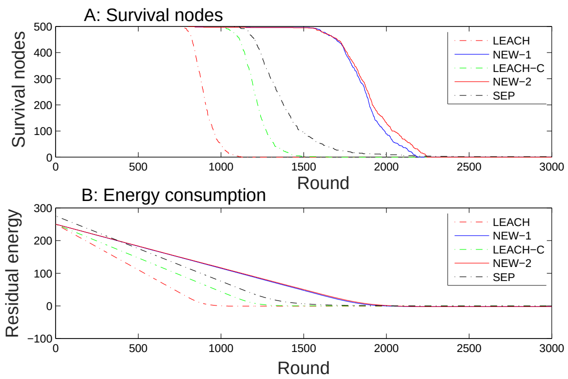


Fig. 7: Comparison of life cycle and energy consumption in SGT

Fig. 8 indicates that the new scheme has a great advantage in throughput. Compared with LEACH or SEP, the throughput in our method has approximatively increased by more than 5 or 3 times. And compared with LEACH-C protocol, although their throughput is roughly equal, the network life cycle in our method is almost increased by 50 percent.

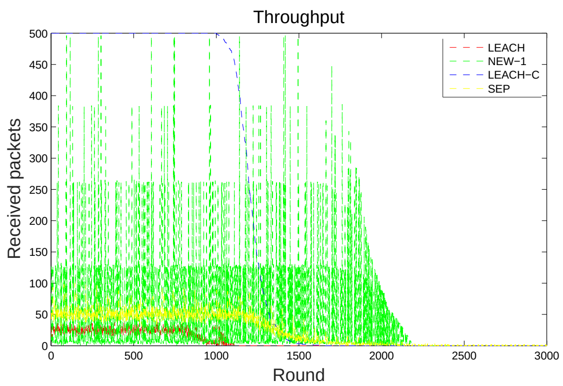


Fig. 8: Comparison of throughput in SGT

Fig. 9 depicts the change of the numbers of the survived nodes as the number of partitions varies from 3 to 5. We can find that the performance of 4 partitions is better than that of 3 partitions and 5 partitions, which implies that the number of partitions is not the key factor, and partition equilibrium is more important in this case.

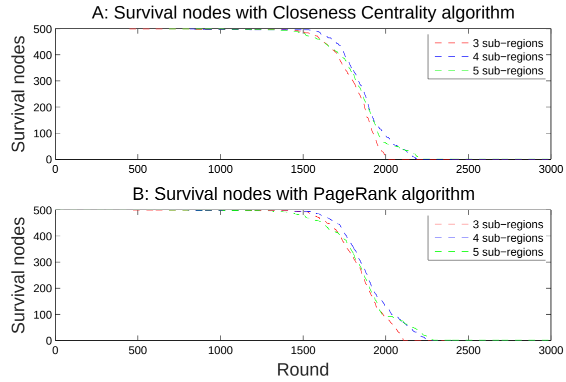


Fig. 9: Comparison of number of partitions in SGT

Scene 2: 500 nodes are randomly distributed in the region of $100m * 100m$, and they are divided into 4 sub-regions by DD algorithm. Since DD algorithm has selected the Leader node when performing network clustering, there is no need to sort the nodes again.

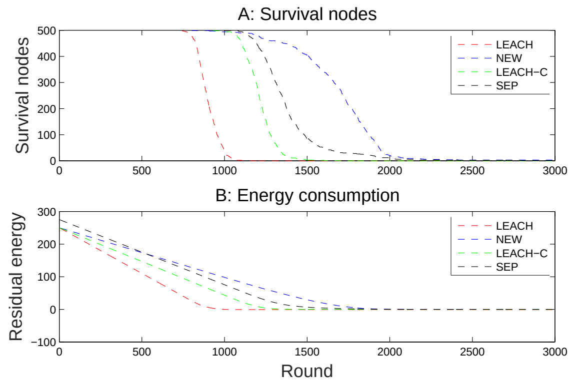


Fig. 10: Comparison of life cycle and energy consumption in DD

Fig. 10(A) shows that the number of alive nodes varies with time. It is seen that the new scheme with DD algorithm is much better than the traditional protocols. In the new architecture, the first dead node appears when the program executes about 1000 times, and all of the nodes die when it runs about 1900 times. Compared with LEACH, LEACH-C and SEP, the performance of the new scheme has great improvement. Although the first death node appears earlier than SEP and LEACH-C, its network life cycle almost increases by 30 percent and 40 percent, respectively. Fig. 10(B) depicts the residual energy of the network at each round. We can see that the residual energy of the new scheme is higher than that of LEACH, LEACH-C and SEP protocols. According to the results of the simulation, we can deduce that the new proposed scheme by using DD algorithm has increased the network lifetime and decreasing the energy consumption compared to LEACH, LEACH-C and SEP.

Fig. 11 exhibits the throughput at each round, which indicates that the new scheme is also better than the traditional protocols. As we can see, the new scheme has exceeded

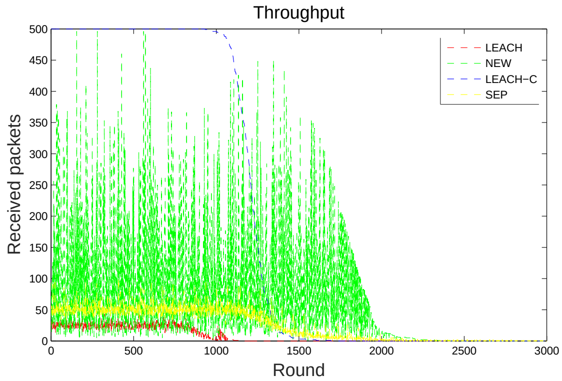


Fig. 11: Comparison of throughput in DD

LEACH and SEP in the aspect of packets received in the B-S, and the throughput has approximately increased by more than 6 or 4 times. Compared with LEACH-C, the network life cycle in our new scheme is almost increased by 40 percent, even though their throughput is roughly equal.

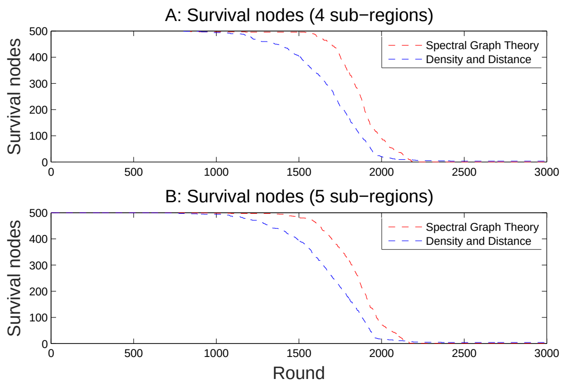


Fig. 12: Comparison of clustering algorithms

Fig. 12 shows the comparison between SGT and DD. Whether the network is divided into 4 or 5 subareas, the network life cycle of SGT is about 15 percent higher than DD. That is because DD algorithm has the more strict prerequisites and the distance threshold parameter plays a key role to affect the performance. If the prerequisites are not satisfied or the threshold parameters are set inappropriately in practice, the network clustering will be unreasonable with a high probability. In fact, DD is not suitable for networks with uneven density distribution. However, SGT adopts the recursive bipartite method based on symbol mode of Fiedler vector, and it no longer depends on the distribution of nodes, so the network can be divided evenly, especially for the case of even sub-regions.

To sum up, the above simulation results illustrate the new protocol has a better performance of the network life cycle and throughout. According to the analysis of the network topology, a large scale WSN can be devided into several small subareas by using clustering algorithms, which can reduce the energy consumption by shortening the communication distance and reducing the frequency of communication. Additionally, the number of nodes in each subarea is roughly equal, which would make the speed of energy consumption in each subarea approximately equal and also avoid excessive energy consumption due to overmuch nodes in a certain subarea. Furthermore, we can use the sorting algorithms to select the most suitable node in each subarea as the leader node, and this node is only responsible for transmitting data to BS, which could improve network lifetime and throughout by avoiding undue data overhead, such as data loss and data retransmission.

## 4 Conclusion

In WSN, many routing protocols are explored to manage the energy consumption, thereby improving the overall life cycle of the network and the network availability. In this paper, we propose a new method to improve the performance of LEACH routing protocol by using network partition technique and leader selection strategy. The simulation shows that the network performance can be significantly improved in the new scheme. However, the proposed scheme has not been tested in a wider range of practical network. If the real distance between the center nodes and BS is too far, that would have a huge impact on energy consumption. Therefore, the future research will focus on the large scale network involving with multi-hop, multi-path and QoS.

## References


[1] Zhu C, Leung V C M, Shu L, et al, Green Internet of Things for Smart World, IEEE Access, 3:2151-2162, 2015.

[2] Dener M, A New Energy Efficient Hierarchical Routing Protocol for Wireless Sensor Networks, Wireless Personal Communications, (3):1-18, 2018.

[3] Akkaya K, Younis M, A survey on routing protocols for wireless sensor networks, Ad hoc networks, 3(3): 325-349, 2005.

[4] Al-Karaki J N, Kamal A E, Routing techniques in wireless sensor networks: a survey, IEEE wireless communications, 11(6): 6-28, 2004.

[5] Heinzelman W R, Chandrakasan A, Balakrishnan H, Energy-Efficient Communication Protocol for Wireless Microsensor Networks, 18:8020, 2000.

[6] Heinzelman W B, Chandrakasan A P, Balakrishnan H, An application-specific protocol architecture for wireless microsensor networks, IEEE Transactions on wireless communications, 1(4): 660-670, 2002.

[7] Smaragdakis G, Matta I, Bestavros A, SEP: A stable election protocol for clustered heterogeneous wireless sensor networks, Boston University Computer Science Department, 2004.

[8] Tang C, Tan Q, Han Y, et al, An Energy Harvesting Aware Routing Algorithm for Hierarchical Clustering Wireless Sensor Networks, Ksii Transactions on Internet & Information Systems, 10(2):684-686, 2016.

[9] Xu D, Tian Y, A comprehensive survey of clustering algorithms, Annals of Data Science, 2(2): 165-193, 2015.

[10] Wang D, Chen W, Qiu L, Multi-leader selection in complex networks, Decision and Control (CDC), 2016 IEEE 55th Conference on. IEEE, 1751-1756, 2016.

[11] Rodriguez A, Laio A, Clustering by fast search and find of density peaks, Science, 344(6191): 1492-1496, 2014.

[12] Barabsi A L, Scale-free networks: a decade and beyond, Science, 325(5939): 412-413, 2009.

[13] Sen T, Chaudhary D K, Contrastive study of Simple PageRank, HITS and Weighted PageRank algorithms: Review, International Conference on Cloud Computing, Data Science & Engineering - Confluence. IEEE, 721-727, 2017.

[14] Saxena A, Gera R, Iyengar S R S, A Faster Method to Estimate Closeness Centrality Ranking, arXiv preprint arXiv:1706.02083, 2017.