Design and Research of a New Clustering Algorithm for Wireless Sensor Network

Yuanjing Zhu\textsuperscript{1}, Xiaolong Chen\textsuperscript{2,\*}
\textsuperscript{1}Yunnan University Dianchi College, Kunming 650228, China
\textsuperscript{2}Yunnan Minzu University, Kunming 650031, China

\*Corresponding author: chenxiaolong@minzu.edu.cn

Abstract. At present, the mainstream method to extend the life cycle of wireless sensor networks is to use clustering technology. An improved method based on multi-attribute decision making is designed for the defect of PPTC algorithm. The path cost detection is performed periodically, and the ordered weighted average operator is used to calculate the path consumption weight value and the node weight value. Finally, the Prim algorithm is used to establish the optimal cluster tree. Thus, keeping the best path is always in the optimal state of dynamic updates. The simulation results show that the improved algorithm mechanism is more reasonable than the PTTC algorithm. The network topology and transmission path are more reasonable. Each node can better share the network energy consumption and make the network run longer.

Keywords: Wireless sensor network, clustering, optimal path.

1. Introduction
Wireless Sensor Networks (WSN) is a distributed sensing network. The main function is to sense and transmit the monitored environmental information through nodes deployed in the environment. Due to the influence of many factors such as spatial location, energy limitation and volume of the monitoring point, each node in the WSN uses wireless communication to transmit information. Different protocol stacks can be used to form different types of multi-hop ad hoc networks. The network configuration is flexible. It supports sensor node location movement, and can also access the Internet through a gateway or IP-based wireless sensor network protocol stack [1] to meet the remote and intelligent monitoring and control requirements of the information age. The core task of the wireless sensor network is to collect, process and transmit environmental information. Together with communication technology and computer technology, it constitutes the three pillars of the information age. Since the wireless sensor network node can no longer obtain new energy supply after deployment, how to reduce energy consumption and improve network life begins as the main direction of WSN research [2-4].

The LEACH (Low Energy Adaptive Clustering Hierarchy) based on the literature [5] is a classical cluster structure algorithm, but does not take into account the remaining energy information of the node when electing the cluster head. Literature [6] proposed that EDACH (Energy Driven Adaptive Clustering Hierarchy) adopts a proxy node strategy. The EECH (Energy Efficient Clustering Hierarchical) algorithm proposed in literature [7] that a node with higher energy has a higher probability to be selected as a cluster head. The EECHS algorithm proposed in literature selects the cluster header (CH) based on
the LEACH algorithm and takes into account the remaining energy and distance information of the node. The Priming Tree Topology Clustering Algorithm for Wireless Sensor (PTTC) proposed in literature takes into account the remaining energy of the node and the probability of electing cluster heads to elect cluster heads. However, these algorithms do not consider the energy left by the sensing device and the energy lost by the path transmission in the calculation of the path weight.

This paper explores an improvement measure based on the PTTC algorithm. The number of cluster heads and the cluster head selection are in accordance with the PTTC algorithm, but the residual energy of the node, the cluster head, and the path transmission loss are comprehensively considered when calculating the weight of the sensor node and the path weight. According to the calculated weights, the prim algorithm is used to establish the optimal cluster tree structure and the timing update keeps the cluster tree structure optimized. The simulation results show that compared with the improved algorithm of PTTC algorithm, the network energy consumption can be more evenly distributed on each node, so that the network can run for a longer time.

2. Related work

2.1. Based on multi-attribute decision making
Multi-attribute decision-making refers to determining the optimal solution strategy or arranging various solutions when the results are determined by various influencing factors. The specific idea is to collect data, arrange the scheme, and select the best. In order to obtain more objective and accurate results, the well-known scholar Yager published the idea of Ordered Weighted Average Operator, which ranks the influence factors on the results, and assigns different weights according to the degree of influence. In this paper, the operator is used to focus on various factors affecting the weight of the path, to reduce the adverse effects caused by subjective guessing, and to maximize the entropy as the final criterion.

3. System model

3.1. Algorithm Description
After the optimal number of cluster head CH is calculated according to the PTTC algorithm and the probability that each sensor node becomes the cluster head CH, the clustering network topology generation phase is started. After the network is powered on, the base station broadcasts the probe message to the node within the single-hop path to determine the single-hop path weight between the node and the neighbor node, and updates the path weight routing table after receiving the correct reply. The node that receives the probe message continues to send probe messages to neighbor nodes other than the sender of the probe message, and updates the path weight routing table after receiving the correct reply. After traversing all the nodes, the Prim algorithm is used to construct the cluster tree, which will form a preliminary, transmission path optimization network structure.

In this algorithm, the planned information transmission path takes into account the remaining energy of the node, the number of nodes passing through, and the energy consumption on the path. The energy consumption of the nodes in the network is more evenly averaged, and the information transmission path is kept in an optimized state, thereby improving the survival life of the network.

3.2. Optimal number of cluster heads
The PTTC algorithm firstly uses the multipath fading channel model to establish the energy consumption model of wireless communication between nodes. According to the data that the cluster head needs to be fused, the energy expression formula of the cluster head consumed by the network is obtained. Then, according to the distance from the node to the cluster head and the information to be transmitted, the expression of energy consumption of each child node is derived, and then the relation expression between the total energy consumption in the cluster and the number of cluster heads is obtained. Finally, we take the derivative of the number of cluster heads by the expression of the total energy consumption to calculate the optimal number of cluster heads $m_{\text{best}}$. 


\[ m_{\text{best}} = \frac{kh^2\varepsilon_{fa}}{0.342h^4\varepsilon_{fa} - E_{\text{node}}} \]  

\[ k = \lambda H = \lambda(4h^d) \] is the number of nodes, \( H \) is the area of the sensing monitoring area, \( h = \left(\frac{H}{4}\right)^{\frac{1}{d}} \), \( E_{\text{node}} \) is the energy consumption of the node when transmitting and receiving information, \( \varepsilon_{fa} \) is the amplifier factor of multipath fading, and \( \varepsilon_{fa} \) is the free space amplifier factor.

Based on this, the probability \( p_{\text{clu}} \) of the node as the cluster head is derived:

\[ p_{\text{clu}} = \frac{m_{\text{best}}}{k} = \frac{h^2\varepsilon_{fa}}{k(0.342h^4\varepsilon_{fa} - E_{\text{node}})} \]  

3.3. Node and path weights value

After the system is started, the base station starts to send network parameter information to the surrounding broadcast, including the number ID, the remaining energy \( node(i) \cdot Er \), and its own parameters \( status \). The node stores the received broadcast information from the neighbor in the neighbor information table, and counts the total number of neighbor nodes. Combining the parameters recorded in the neighbor table, comprehensively consider the degree of influence of multiple factors to determine the proportion of various influencing factors using the operator. Finally, the weight of the node is calculated. Formulated as follows:

\[ V = \nu_1F_1 + \nu_2F_2 + \nu_3F_3 \]  

In the formula (1): \( F_1 \) is the total number of other nodes in the single-hop range of the node; \( F_2 \) is the remaining energy parameter; \( F_3 \) is the parameter of the sensor from the base station; \( \nu_1, \nu_2, \nu_3 \) indicates the magnitude of the influence degree of the above factors. Normalize the following three key influencing factors using the following formula:

\[ F_1 = \frac{\text{spot}(m) \cdot \text{neimum} - \text{Nmin} \cdot \text{neighbor}}{\text{Nmax} \cdot \text{neighbor} - \text{Nmin} \cdot \text{neighbor}} \]  

\[ F_2 = \frac{\text{spot}(m) \cdot P_w}{\text{spot}(m) \cdot P_{\text{max}}} \]  

\[ F_3 = \frac{\text{spot}(m) \cdot \text{lentoCS} - \text{Lmin} \cdot \text{toCS}}{\text{Lmax} \cdot \text{toCS} - \text{Lmin} \cdot \text{toCS}} \]  

In the formula: \( \text{spot}(m) \cdot \text{neimum} \) is the number of nodes in the single-hop path of network node \( i \); \( \text{Nmax} \cdot \text{neighbor} \) is the number of nodes with the most neighbors; \( \text{Nmin} \cdot \text{neighbor} \) is the number of nodes with the least number of neighbors; \( \text{spot}(m) \cdot P_w \) is the remaining energy parameters of the network node \( m \); \( \text{spot}(m) \cdot P_{\text{max}} \) is the maximum available energy parameter before the network node \( m \) starts; \( \text{spot}(m) \cdot \text{lentoCS} \) is the length of the distance between the network node \( m \) and the base station; \( \text{Lmax} \cdot \text{toCS} \) is the distance from the farthest node of the base station to the base station; \( \text{Lmin} \cdot \text{toCS} \) is the distance from the nearest node of the base station to the base station.

The weighting coefficient of the weight set \( \{ \nu_1, \nu_2, \nu_3 \} \) of each attribute is determined by the decision based on the maximum entropy, and the optimistic parameter \( \alpha \) in the operator is set to \( \alpha = 1/2 \). Therefore, the unreasonable judgment of the decision makers on the choice of results is eliminated, and the weights of the influences of various influencing factors on the cluster head selection can be more accurately reflected after performing the normalized calculation of various influencing factors.

When determining the path weight, the PPTC algorithm only uses the hop count as the path weight to measure the pros and cons of the path, and does not take into account the total energy consumption when the information is transmitted on the path. At the same time, even if the same path varies with the
remaining energy of the node on the path, the weight of the path will change accordingly. Therefore, in order to make the energy consumption of all nodes as balanced as possible, it is necessary to periodically find the best path. The path weight evaluation function is defined as follows:

\[ k(m) = \frac{v \cdot \text{spot}(m)}{v \cdot \text{spot}(n)} \cdot \frac{1}{\text{spot}(m) \cdot \text{consume}(n)} \]  

(7)

In this formula: \( v \cdot \text{spot}(m) \) is the weight parameter of the node \( m \) that receives the path detection message; \( v \cdot \text{spot}(n) \) is the weight parameter of the node \( n \) that sends the path detection message; \( \text{spot}(m) \cdot \text{consume}(n) \) is the energy consumed by the path of the sender node \( n \) to the receiver node \( m \). The algorithm takes into account the key factors of the path weight: the remaining energy, the number of hops transmitted, the number of neighbors, and the energy consumption of the information from the start point to the end point.

3.4. MST construction

After determining the optimal number of cluster heads and selecting the node that serves as the cluster head, the generation phase of the cluster tree structure is started. The process of constructing a Minimum Spanning Tree (MST) network using the Prim algorithm is shown in Figure 1. The purpose of building a minimum spanning tree network is to minimize the overall network overhead on the premise of ensuring communication between any nodes in the network.

![Figure 1 The process of building MST based on prim algorithm](image)

1. Set the initial state of all nodes are clustered in H. Start from any point s in H and put point s into set K.
2. Find a node t in the remaining nodes of the H-K to minimize the weight of all points in the K set. Then put the t node into the set K.
3. Repeat step 2 until all nodes are added to the set K. At this point, a minimum spanning tree MST can be generated. If H contains M nodes, the corresponding M-1 edges will be generated.

The overall cluster tree network structure is shown in Figure 2 after all cluster heads' minimum spanning tree network generation. In the network operation phase, the sensor node at the last end transfers the collected information to its upper-level node, which is its parent node, and the upper-level node then passes the information to its upper-level node, which is the parent’s parent node. The data is finally converging to the root node of the cluster tree network through one-way flow, that is, the data fusion processing is performed at the cluster head node. When the overall network topology is generated by the Prim algorithm, in order to achieve information exchange between cluster head networks and between cluster head networks and base stations, each cluster head will form a minimum spanning tree network and serve as the final data aggregation point of the MST for data fusion processing.
Figure 2 Complete cluster tree network structure

4. Algorithm performance analysis

4.1. Initial conditions
Simulate in the MATLAB environment. The simulation parameters are consistent with the PPTC algorithm. The nodes are randomly placed in the field of 100m*100m, and the base station is deployed in the center of the field. The initial settings of the simulation parameters are shown in Table 1.

| Parameter name             | Size       |
|----------------------------|------------|
| Simulation area/m²         | 100*100    |
| Number of sensing points   | 500        |
| Node startup energy/J      | 1.0        |
| Sensing information size/B | 500        |
| Control instruction size/B | 20         |
| $E_{\text{elec}}$/(nJ/bit) | 50         |
| $\varepsilon_{fs}$/(pJ/bit/m²) | 10    |
| $\varepsilon_{mp}$/(pJ/bit/m⁴) | 0.0013 |
| $E_{\text{DA}}$/(nJ/bit/signal) | 50       |
| Influencing factor $\alpha$ | 0.2       |
| Influencing factor $\beta$ | 0.8       |

Under the same initial conditions, the LEACH, PTTC algorithm and the improved PTTC algorithm were selected to perform 50 simultaneous simulations and the average value was calculated, with a single run time of 500 s. And record the comparison results.

4.2. Energy consumption analysis
The running rounds and energy consumption records of the three algorithms are shown in Table 2. From Table 2, when the energy consumption is 50%, LEACH, PTTC, and improved PTTC run 478, 1080, and 1368 rounds respectively. Comparing the energy consumption from the same number of rounds, it can be seen that LEACH, PTTC, and improved PTTC consume 48.67%, 22.43%, and 18.52% of energy respectively when running to 800 rounds. It can be seen that the improved PTTC algorithm has a 30.15% lower energy consumption than the LEACH algorithm and a 3.91% lower energy consumption than the PTTC algorithm.
Table 2 Three algorithms running rounds and energy consumption

| Serial number | LEACH 50% energy consumption | PTTC 50% energy consumption | Improved PTTC 50% energy consumption | LEACH 800 rounds energy consumption percentage | PTTC 800 rounds energy consumption percentage | Improved PTTC 800 rounds energy consumption percentage |
|---------------|-----------------------------|-----------------------------|-------------------------------------|---------------------------------------------|---------------------------------------------|---------------------------------------------|
| 1             | 476                         | 1124                        | 1333                                | 48.28                                      | 22.91                                      | 19.92                                      |
| 2             | 480                         | 1045                        | 1358                                | 48.64                                      | 23.51                                      | 17.76                                      |
| 3             | 482                         | 1058                        | 1377                                | 48.60                                      | 23.93                                      | 18.23                                      |
| 4             | 485                         | 1074                        | 1390                                | 48.82                                      | 20.55                                      | 18.61                                      |
| 5             | 470                         | 1087                        | 1406                                | 49.00                                      | 21.06                                      | 19.08                                      |
| 6             | 472                         | 1095                        | 1419                                | 49.17                                      | 21.56                                      | 19.36                                      |
| 7             | 474                         | 1112                        | 1350                                | 49.35                                      | 22.06                                      | 19.74                                      |
| 8             | 477                         | 1125                        | 1343                                | 48.12                                      | 22.57                                      | 17.11                                      |
| 9             | 479                         | 1033                        | 1356                                | 48.25                                      | 22.94                                      | 17.49                                      |
| 10            | 481                         | 1049                        | 1368                                | 48.48                                      | 23.45                                      | 17.86                                      |
| average       | 478                         | 1080                        | 1368                                | 48.67                                      | 22.43                                      | 18.52                                      |

4.3. Packet loss rate
Figure 2 shows that the packet loss rate of the three algorithms increases with time, but the packet loss rate of the improved PTTC algorithm is always at a minimum. The highest packet loss rate is 0.12% before the end of the simulation, indicating that this improvement can ensure good communication quality. The LEACH algorithm has the highest packet loss rate. It also reflects that the random selection of cluster heads will cause some cluster head load imbalance to aggravate the cluster head energy consumption, causing some cluster heads to die early, resulting in network structure breakage, resulting in a large number of packet loss phenomena.

Figure 3 The packet loss rate for 3 algorithms

5. Conclusions
Aiming at the problem that the information transmission path is solidified in the PTTC algorithm and the energy consumption of the node is not balanced, a perfect optimization algorithm is designed. The operator is used to centrally deal with various factors affecting the weight of the path, such as the remaining energy, the communication energy consumption between the two points, and the weight of the node itself. The entropy value is maximized as the weight value. The algorithm can maintain the optimal path of node data transmission for a long time. The simulation results show that the improved algorithm can balance the energy consumption of the network and make the network run longer.
Acknowledgements
This work was financially supported by the Science Research Fund of Yunnan Provincial Department of Education, China (Project Number.2021J0798,2020J0364,2019J0827) and Teaching Research Project of Yunnan Minzu University, China (ProjectNumber:2019JWC-JG-77)

References
[1] Li Fengguo. Research and implementation of wireless sensor network based on 6LoWPAN[D]. Nanjing University of Posts and Telecommunications, 2013.
[2] Qin Junxiang, Xu Xiaofeng, Yi Kefu, et al. Clustering algorithm for wireless sensor networks using weight function timing, September 8, 2015 [J]. Computer Engineering and Applications, 2015.
[3] Zhang Huanan, Li Shijun, Jin Hong. Energy-saving research on clustering routing in wireless sensor networks [J]. Computer Engineering and Science, 2015, 37(10): 1869-1876.
[4] Jiang Shen, Wang Dawei. An Energy Balanced Clustering Routing Algorithm in Wireless Sensor Networks [J]. Computer Technology and Development, 2014(1): 113-117.
[5] HEINZELMANWR, CHANDRAKASANA, BALAKRISHNANH. Energy-efficient communication protocol for wireless micro-sensor networks [C]. In Proc. HICSS, 2000:1-10.
[6] KIMKT, YOUNGHY. Energy driven adaptive clustering hierarchy (EDACH) for wireless sensor networks [J]. LNCS, 2005, 38 (23): 1098-1107.
[7] HUY, LIW, KANGZ. Study on energy efficient hierar chical routing protocols of wireless sensor network [C]. In Proc. ICIE. 2009: 325-328.