
Akant K. Raghuwanshi, Virendra S. Chaudhary
Department of Electronic and Communication Engineering, RKDF University, Bhopal, Madhya Pradesh, India

Abstract

Recently, cluster distribution and selection have been made in the most advanced protocols based on Low Power Adaptive Clustering Hierarchical Protocol (LEACH), followed by cluster head election. Conversely, it is not thought that data redundancy is sufficient. This study suggests a modified Leach approach incorporating relative information entropy (Leach-CIE). The relative entropy value of the data's probability distribution in the node's two subsequent periods is calculated in conjunction with the close entropy model. The information is then compared with a threshold value to determine the degree of redundancy. The network accomplishes reducing energy consumption by refusing to transmit redundancy data. Instead, a multi-hop routing (MFRCRE) considers its residual energy, the energy consumption of each node in the network during the data transmission phase, by optimizing node forwarding conditions and considering factors such as communication distance and energy consumption. It is proposed to balance the rate—account ratio. The results of the experiments demonstrate that this approach can successfully reduce and balance the energy consumption of network nodes to increase the life of the network.

Keywords: Wireless Sensor Network, LEACH, LEACH-CIE, MFRCRE, Optimise Node, Multi-Hop Routing, Relative Entropy, Probability Distribution.

Introduction

The wireless sensor network (WSN) has limited node energy and redundant data traits when the cost and application value are considered. The energy consumption of each network node must be as low as possible in order to extend the WSN working cycle. The energy consumption rates of each node must be balanced in order for WSN to provide comprehensive and efficient monitoring of the targeted region. The traditional low-energy adaptive clustering hierarchy (LEACH) was cited in Literature study from 2017 in that year. They clarified how the clustering concept was initially represented in the protocol. The energy consumption of network nodes is significantly reduced via network-layer data transfer. However, as the application system is upgraded, the protocol can no longer satisfy requirements, leading to the emergence of numerous new, enhanced LEACH protocols as needed.

Literature added a cost function in the cluster head election strategy to maintain the energy balance of the nodes, and those closer to the sink and with higher residual energy were selected first. Still, the paper did not consider the total energy in the forwarding process. The energy consumption of multi-hop transmission may be higher than that of a single-hop direct message, resulting in higher overall network energy consumption. Literature designed an efficient cluster head election strategy based on the particle swarm optimization algorithm. Although the network energy consumption is reduced compared to the random election cluster head protocol, this paper only optimizes the energy consumption within the cluster. The problem is that the energy consumption between sets and routing levels has not been discussed too much and needs to be further optimised.

Corresponding Author: Akant K. Raghuwanshi, Department of Electronic and Communication Engineering, RKDF University, Bhopal, Madhya Pradesh, India, e-mail: akantthakur7@gmail.com


Source of support: Nil

Conflict of interest: None

© The Author(s). 2022 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and non-commercial reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
heuristic algorithm, which can obtain the optimal cluster head in WSN. The method is used to balance the network energy consumption, but there is no other design of the node energy consumption balance strategy at the routing level. Literature\(^6\) believed that intermittent dormancy of nodes could save energy well, so they designed a node-adaptive sleep-wake scheduling algorithm. Still, the paper only reduced the energy consumption of a single node to reduce the overall energy consumption of the network. There is not much consideration from the perspective of balancing network energy consumption. Literature\(^7\) used the bidirectional selection of standard and cluster head nodes to optimise cluster head election. They proposed a multi-hop data distribution method that comprehensively considered data transmission volume and transmission distance in the data transmission stage. Although this method can reduce network energy consumption to a certain extent, this method does not consider the remaining energy of nodes, which is not conducive to the balance of energy consumption of nodes. Literature\(^8\) used the K-means algorithm and data regression strategy to achieve reasonable clustering for the network with uneven distribution of nodes. The paper achieves energy consumption balance between clusters by uniform clustering for the energy consumption at the routing level. Literature\(^9\) uses the cuckoo algorithm to find the optimal path but ignores clustering and data redundancy problems. Based on the LEACH protocol, Literature\(^10\) set the node clustering factor as the basis for the node to be elected as the cluster head and then designed a variable data transmission period within the cluster to achieve the energy balance of the network nodes. Still, this strategy’s scope of use is limited, and it is challenging to implement the network with a fixed data collection period. Literature\(^11\) uses data compression sensing to reduce the amount of data transmission as much as possible to save energy while accurately acquiring and analyzing data. The paper reduces the amount of data transmission through data fusion but does not discuss data transmission—energy consumption balance. Literature\(^12\) first gridded the monitoring area, used the double-chain genetic algorithm to design a reasonable path, and used the convenient and movable sink node to follow the path to collect the entire data, reducing the communication distance of other nodes and saving the network. Energy consumption, but this strategy is challenging to implement for a fixed sink node network.

Because the energy consumption of WSN nodes is mainly reflected in data transmission, most of the traditional low-power WSN algorithm research and improved protocols based on LEACH are carried out from the transmission path, cluster head election and clustering strategy, etc., while ignoring the WSN data is redundant according to the characteristics of data redundancy in WSN.

**Improved LEACH Protocol**

**LEACH Protocol**

The core idea of the LEACH protocol is that the cluster head is elected regularly, and the remaining nodes select the cluster to which they belong based on the principle of the smallest distance. The cluster head collects the data of all cluster nodes and then forwards it to the sink node. The cluster head is randomly selected and held in turn. Each node in the network needs to generate a random number in the \((0,1)\) interval and compare the random number with the threshold \(T(n)\).

\[
T(n) = \begin{cases} 
\frac{p}{1-p(n \mod \frac{1}{p})}, & n \in G_r \\
0, & n \notin G_r 
\end{cases}
\]

Among them, \(p\) represents the percentage of cluster heads in the network, \(n\) represents the number of rounds that have been successfully completed, \(n\) represents the node number, and \(n \in G_r\) denotes that node \(n\) is a member of the set \(G_r\). The number of games returned in the most recent \(r \mod (1/p)\) round is known as \(G_r\). If the random number produced by the nodes in \(G_r\) is smaller than the threshold \(T(n)\), the set of nodes that have not already been elected as cluster heads within a round will be chosen.

The following are flaws in the LEACH protocol: (1) The LEACH protocol uses single-hop direct routing, which is detrimental to the energy balance of the nodes in the overall network since long-distance communication takes more energy. (2) The energy-saving effect of the LEACH protocol needs to be further improved because it does not perform duplicate operations on the data of the wireless sensor network.

The modified centralized low-energy adaptive clustering hierarchy protocol, or LEACH-CM (modified centralized low-energy adaptive clustering hierarchy protocol), is an advancement over LEACH in that it not only optimizes the cluster head election strategy but also ensures that the cluster head is more significant from the residual energy than the network average as a whole. The cluster head nodes (CHs) are where the point is formed and where the network’s total energy is stored. In addition, LEACHCM dynamically alters the data transmission path of select nodes to achieve a more excellent energy-saving effect. The number of live nodes is alive multiplied by the cluster head ratio \(p\). To confirm the efficacy of this approach, the experimental component will pass a comparison with LEACH-CM.

**Energy Consumption Model**

According to the literature,\(^13\), the node data transmission model is as follows. Sensor sending and transmission (\(I\) the unit is a bit) the energy consumed by the data:

\[
E_{Tx}(l, d) = \begin{cases} 
LE_{elec} + LE_{f_{sd}}d^2, & d < d_0 \\
LE_{elec} + LE_{f_{sd}}d^3, & d \geq d_0 
\end{cases}
\]
Figure 1: 10 sampling cycles are one round of Data Transmission cycle.

Similar to long-distance freight commodities, loading and unloading fees are also required in addition to transportation fees. Equation (2) consists of two parts: the energy consumption of the transmission circuit and the transmission energy consumption. $E_{elec}$ represents the energy consumption of the unit bit data sent by the sensing circuit, and its size is related to encoding, modulation, filtering, etc. During the transmission process, according to the transmission distance $d$ and the threshold $\epsilon = \frac{d_{th}}{\sqrt{n_{am}}}$ relationship, subdivide the model into free-space models (transmission energy consumption vs distance is proportional to the square of).

In the multipath attenuation model (transmission energy consumption versus distance is proportional to the fourth power), the energy consumption parameters under the two models are $\epsilon_{fs}$ and $\epsilon_{amp}$. Since $d$ is greater than 0, the energy consumption under both models varies with the distance increases and monotonically increases, compared to the free-space model, the multi-channel transmission model's energy consumption under the radial attenuation model is more, so for better energy saving, reduce the number of the transmission distance is particularly important. The number of $l$ (unit is a bit) received by the sensor According to the energy consumed:

$$E_{RX} = lE_{elec} \quad (3)$$

Similar to the energy consumption of the signal transmission circuit, the node is receiving the energy consumed by the data is only related to the size of the data packet, so the formula (3) The amount consumed by the sensor node to receive $l$ (bit) data The energy is $lE_{elec}$.

Relative Entropy

Since the sensor collects the data value at a certain time, the relative entropy calculation method of discrete variables is used when calculating the relative entropy of periodic data. Suppose there are two probability distribution functions, respectively, $f(x) = \{f(1), f(2), \ldots\}$ and $g(y) = \{g(1), g(2), \ldots\}$, and

$$\sum_{i=1}^{n} f(i) = 1, f(i) \in [0,1]$$
$$\sum_{j=1}^{n} g(j) = 1, g(j) \in [0,1]$$

Then the relative entropy of the two probability distributions is:

$$KL(f||g) = \sum f(i) \log \frac{f(i)}{g(i)} \quad (4)$$

The value of the relative entropy $KL(f||g)$ can be used to describe the difference between the two probability distribution functions $f(x)$ and $g(y)$. In general, the greater the difference between the probability distributions, the farther the ratio of $f(i)$ to $g(i)$ is from 1, and the farther the value of $KL(f||g)$ is from 0. In the extreme case, when the two probabilities are $\delta = 0$, the value of $KL(f||g)$ is $\delta$.

De-redundancy Method

The data set used in this paper is the environmental data collected by Intel Berkeley Research Lab in a real environment using multiple heterogeneous sensors for a continuous month, including data information such as date, time, sensor number, temperature, humidity, etc. This paper uses temperature value data $T_i$, the node completes the collection of environmental data every 30 s, and sends data once every 10 data collection processes. Each data sending period is represented by $T_i$, as shown in Figure 1.

The calculation method and de-redundancy method of node period data probability distribution are as follows:

1. First, the current data transmission cycle $T_i$ of the sensor node and the data column vectors $X$ and $Y$ of the previous cycle $T_{i-1}$ are reconstituted into a new $10 \times 2$-dimensional vector $x = [X, Y]$.

2. Traverse the maximum value $x$ and the minimum value $x$ of the elements in the vector $x$, find the data span interval as $(\max x, \min x)$, and divide the interval into $d_{uan}$ segments ($d_{uan} = 10$ in this article), as follows shown in which:

$$\min x \leq x < \max x$$

Then the probability distributions of $T_i$ and $T_{i-1}$ period data are:

$$P = \frac{N_1}{10}, Q = \frac{N_2}{10}$$

3. Calculate the relative entropy value $KL(P||Q)$ of the data distribution between the current cycle $T_i$ and the previous cycle $T_{i-1}$ of the sensor node according to formula (5). The smaller the relative entropy value, the greater the probability distribution of the data in the two cycles.

$$KL(P||Q) = \sum_{j=1}^{d_{uan}} P_j \times \log \frac{P_j}{Q_j} \quad (5)$$

According to the practical application design of this article, the following will briefly analyze whether $Q_j$ is 0:

* $\exists j \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ such that $Q_j = 0$, then $\frac{P_j}{Q_j} = 0$. When $KL(P||Q) = \infty$, which means that there is no data in the previous cycle, but there is in the current cycle, so there is a big difference in the data of the two cycles.

* $\forall \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, all have $Q_j \neq 0$, then $Q_j = \frac{1}{10}$, and $\sum_{j=1}^{10} P_j = \frac{1}{10}$, Hence $Q_j = 1/10$ for all $j \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Similarly, $P_j \in \{110,120,130,140,150,160,170,180,190,1\}$ and $\sum_{j=1}^{10} Q_j = 1$, hence $Q_j = 1/10$ for all $j$. Therefore, it can be concluded that $\frac{P_j}{Q_j}$ is non-negative, and the value of $KL(P||Q)$ increases as the ratio of $P_j$ to $Q_j$ increases. That
is, the greater the difference between the two periods of data, the greater the value of KL(P||Q).

From this conclusion, the relative entropy value KL(P||Q) of the data probability distribution P of the current period \( T_i \) of the node and the data probability distribution Q of the previous period \( T_{i-1} \) is closer to 0, indicating that the correlation between the two periods of data is higher. The stronger it is, the higher the data redundancy in two cycles is, so it is necessary to set a reasonable threshold Do as the standard for measuring and judging data redundancy, and only when KL(P||Q) ≥Do will node data be sent out.

**Multi-hop Forwarding Routing Improvements**

Communication is the primary energy consumption path of nodes in WSN, so a proper routing protocol needs to be designed to save energy. Moreover, the communication distance of nodes is limited; to balance the communication energy consumption of long-distance nodes, it is necessary to design multi-hop forwarding routes further.

Geographical adaptive fidelity (GAF) is an energy-efficient routing algorithm in WSN. It utilises the dense distribution of network nodes and reduces the number of redundant nodes in the network by setting redundant nodes to sleep. Network energy consumption, but the algorithm has a significant packet loss rate for the mobile node network. Literature proposed a Cluster Routing Algorithm for Virtual Grid based on a virtual grid, CRVB; the algorithm first divides the network into virtual in each area, and the cluster head is selected according to the principle of maximum residual energy. The communication between sinks adopts multi-hop forwarding to reduce network energy consumption while balancing the network load. Literature proposed a hybrid multi-hop routing algorithm for cellular meshes (improved hybrid clustering routing algorithm, IHCRA), which monitors. The area is divided into several regular hexagons, which are set based on factors such as angle and distance. Set the cost function to find the centre with the most minor cost from the neighbour nodes. The successor node acts as its next hop node, extending the network life, and has better results than traditional LEACH, GAF and CRVB protocols. A multi-hop forwarding route that considers its residual energy proposed in this paper (multi-hop forwarding routing considering its residual energy, MFRCRE) is divided into two parts: intra-cluster multi-hop forwarding and cluster Maximum step size transfer. Intra-cluster transmission adopts the idea of splitting into zeros. Turn high-energy long-distance single-hop transmission into low-energy multi-hop information. It reduces the network’s overall energy consumption and balances each node rate’s energy consumption. The inter-cluster message is based on the principle of high efficiency, and the maximum step size is used. Transmission until the data is transmitted to the sink node at a single hop distance (sensor within the farthest transmission distance of the node). This paper will be used in conjunction with the IHCRA method in the comparative simulation experiments, verifying the MFRCRE method’s effectiveness.

**Intra-cluster Multi-hop Forwarding**

Usually, the cluster head and each node in the cluster is transmitted in a single hop within a distance, or although the source and destination nodes do not belong to the same clusters, but their distances are within a single hop distance, as shown in Figure 2.

Show: The circle in the Figure is a certain cluster in the area, and the S1 node is the distance within the cluster. The node with the farthest cluster head, the energy consumption of this node is larger, if it raises front failure, which may cause the upper left corner of the square monitoring area Loss of effective monitoring, to prevent this from happening, it is necessary to the direct communication between S1 and the cluster head is transformed into indirect communication, that is, finding a suitable the relay node S2 helps to forward data.

Assumption because \( d_s = \frac{d_0}{\sqrt{E_{amp}}} \) So \( E_{fs} = E_{amp} = \text{ const } \) when \( d = d_0 \) since both \( E_{fs} \) and \( E_{amp} \) are monotonic increasing function and \( E_{fs} \) is proportional to \( d^2 \), and \( E_{amp} \) is proportional to \( d^0 \) ratio, so for better energy saving, it needs to be larger than the long-distance under \( d_0 \) the single-hop straight hair is transformed into a multi-hop jump at a shorter distance, and at the same time, to balance the energy consumption rate between nodes, this paper establishes the energy consumption ratio (\( E_{amp} \): turn the energy consumption of the development process is divided by the remaining energy), and the lower energy consumption is satisfied. On the premise, it is necessary to meet a lower energy consumption ratio to avoid excessive energy consumption and premature failure.

- **Distance formula (guarantee lower energy consumption)**
  Distance formula: when the source node S1 and the relay node S2 and the target nodes together form an obtuse triangle, and the connection between S1 and the target node constitutes an obtuse triangle. When forming the longest side of an obtuse triangle, there must be \( d_1^2 + d_2^2 < d_3^2 \), according to the energy consumption model of the sensor node shows that the

![Figure 2: Schematic diagram of maximum step size transmission](image)

Table 1: Experimental parameter settings

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experimental settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Square network monitoring area, area/m²</td>
<td>100×100</td>
</tr>
<tr>
<td>Sink node coordinates/m</td>
<td>(150,51)</td>
</tr>
<tr>
<td>Initial energy/J</td>
<td>0.11</td>
</tr>
<tr>
<td>/J(bit)</td>
<td>50×</td>
</tr>
<tr>
<td>/J(bit m2))</td>
<td>10×</td>
</tr>
<tr>
<td>/J(bit m4))</td>
<td>0.0013×</td>
</tr>
<tr>
<td>Energy consumption of data fusion EDA/J(bit)</td>
<td>5×</td>
</tr>
<tr>
<td>Sensor packet size/bit</td>
<td>4000</td>
</tr>
<tr>
<td>cluster head ratio p</td>
<td>0.21</td>
</tr>
<tr>
<td>The farthest transmission distance of a sensor node in one hop is dr/m</td>
<td>91</td>
</tr>
</tbody>
</table>

Table 2: 30 groups rERR and rAVG values under LEACH-CIE method when Do = 0.05

<table>
<thead>
<tr>
<th>Index</th>
<th>Value</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rERR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Energy consumption of the sensor node is related to the distance is proportional to the square of the space, according to which “long-distance, high-power single-hop direct”transmission” is converted to “close-range, low-power multi-hop transmission”. Therefore d, d1, d2 must satisfy the following distance formula:

\[
\begin{align*}
\text{d} & < d_0, \quad d^2 - d_1^2 - d_2^2 \geq 2 \frac{E_{\text{source}}}{E_{\text{sink}}}, \quad d_1 < d_0, d_2 < d_0 \\
\text{d} & \geq d_0, \quad \varepsilon_{\text{source}}(d_1^2 + d_2^2) \geq 2E_{\text{source}}, d_1 < d_0, d_2 < d_0 \\
\text{d} & \geq d_0, \quad \varepsilon_{\text{source}}(d_1^2 + d_2^2) \geq 2E_{\text{source}}, d_1 < d_0, d_2 < d_0 \\
\text{d} & \geq d_0, \quad \varepsilon_{\text{source}}(d_1^2 + d_2^2) \geq 2E_{\text{source}}, d_1 < d_0, d_2 < d_0 \\
\end{align*}
\]

Among them, when \( d < d_0 \), if \( d_1 \geq d_0 \) or \( d_2 \geq d_0 \), that means the distance forwarded by the relay node is longer than that of the single-hop direct transmission. Longer and consumes more energy. At this time, the source node S1 will directly communicate with the target node point-to-point without forwarding by the relay node S2.

- Condition of energy consumption ratio (guarantee energy consumption balance) The way to control energy balance should not be to control the energy consumption of nodes. The value is the same, and the node energy consumption ratio should be held to be as the same as possible. Energy consumption is the ratio of the node’s current communication energy consumption to the current remaining energy. The energy consumption ratio condition transforms a single-hop direct transmission with a high energy consumption ratio into a low energy consumption ratio of multi-hop forwarding.

\[
\frac{E_{\text{source}}}{E_{\text{sink}}} \geq \frac{E_{\text{source}}}{E_{\text{sink}}} \begin{cases}
    d < d_0, & d^2 - d_1^2 - d_2^2 \geq 2 \frac{E_{\text{source}}}{E_{\text{sink}}} \frac{d_1^2 + d_2^2}{E_{\text{sink}}} \\
    d \geq d_0, & \frac{E_{\text{source}}}{E_{\text{sink}}} \frac{d_1^2 + d_2^2}{E_{\text{sink}}} \geq 2 \frac{E_{\text{source}}}{E_{\text{sink}}} \\
    d \geq d_0, & \frac{E_{\text{source}}}{E_{\text{sink}}} \frac{d_1^2 + d_2^2}{E_{\text{sink}}} \geq 2 \frac{E_{\text{source}}}{E_{\text{sink}}} \\
    d \geq d_0, & \frac{E_{\text{source}}}{E_{\text{sink}}} \frac{d_1^2 + d_2^2}{E_{\text{sink}}} \geq 2 \frac{E_{\text{source}}}{E_{\text{sink}}} \\
\end{cases}
\]

Maximum Step Size Transmission Between Clusters

In the random distribution of nodes, the nodes closer to the sink are the point that plays the role of a bridge, not only transmitting their data but also responsible for transferring Sending data from distant nodes so that the energy consumption will be relatively higher. Therefore, the multi-hop forward routing (MFRCRE) method considers its residual energy when the distance between the source node and the target node is greater than the sensor’s single.
Each sensor node has the capabilities of environmental perception, data collection, data fusion processing, computing, wireless communication, etc. The environment will not cause damage to the sensor node and will not affect the regular operation of the node. Once the position of the sensor node in the network is determined, it will not change, and each node knows its position coordinates; the battery power of each node and the initial energy is not zero but limited. The energy consumption of sensor nodes mainly occurs in the transmission, reception, transmission and fusion of data, and the energy consumed by sensor environment monitoring and calculation is ignored. The experimental parameters are set as shown in Table 1.

The calculation model of the experimental performance comparison index:

- The number of failed nodes in the entire network

\[ rDEAD = \sum_{r=1}^{r_{\text{max}}} \text{dead}(r) \]  

Among them, \( \text{dead}(r) \) is the number of newly added failed nodes in each round period, and the sum of the number of failed nodes in all rounds is superimposed to obtain the total number of failed nodes in the network. The trend of the comparison index over time can be used to characterize the speed of the network inactivation. When the network’s overall number of failed nodes is equal to the total number of nodes, the network is completely deactivated.

- Average residual energy of the network

\[ rAVG = \text{avg}(200) = \frac{1}{n} \left( \sum_{i=1}^{n} E_{i299} - e_{200i} \right) \]

Among them, \( e_{i} \) is the energy consumption value of node \( Si \) in the \( r \)th round period, \( E_{i} \) is the remaining energy value of the node \( Si \) after completing the \( r \)th round of the data transmission period, and \( \text{avg}(r) \) is the net energy at the end of the \( r \)th round period. The network average residual energy comparison index can be used to compare the energy-saving effect of different methods. The faster the index parameter decreases with time, the higher the network energy consumption.

- Data error

\[ err(i) = \frac{1}{10} \sqrt{\sum_{k=0}^{\infty} (x_k - y_k)^2} \]

\[ rERR = \frac{r_{\text{err}}^{\max}(\sum_{i=1}^{r_{\text{max}}} err(i))}{n_{\text{err}}^{\max}} = \frac{r_{\text{err}}^{\max}(\sum_{i=1}^{r_{\max}} \sqrt{\sum_{k=0}^{\infty} (x_k - y_k)^2})}{n_{\text{err}}^{\max}} \]

Where \( err(i) \) refers to the data error value of node \( i \) in each
cycle, because this paper assumes that the node performs 10 data collection processes in each cycle, the denominator under this calculation model is 10, rERR is the data value obtained by the sink node after the network runs for cycles, and the actual value of the n sensor nodes. The average error of the collected data. Experiments must be carried out within the allowable error range to be effective.

- The mean square error of the residual energy of the network

\[
D_e = \sqrt{\frac{\sum_{i=1}^{n_{\text{alive}}} (E_i - \bar{E})^2}{n_{\text{alive}}}}
\]  

(12)

Among them, \(n_{\text{alive}}\) represents the number of surviving nodes in the network in the current cycle, and \(E_i\) represents the remaining energy of node Si in the current process; the smaller the mean square deviation of the remaining power of the network, the stronger the energy balance capability of the network.

**Relative Entropy**

Since the data value of the sensor node is non-negative, the relative entropy value is non-negative, and the closer KL(P||Q) is to 0, the higher the data redundancy. In terms of the design of the relative entropy, threshold \(D_0\), due to the extensive Do design, too much data is judged to be redundant and discarded, resulting in too large errors in the data obtained by the sink node. According to preliminary experiments, \(D_0<0.5\) is limited. The influence of \(D_0\) changes on the error, set \(D_0\) as the step size of 0.01, increase from 0.01 to 0.5, and calculate the network under the LEACH-CIE (improved LEACH protocol combining relative information entropy) method. After running for 200 cycles under different \(D_0\) conditions, the node average residual energy (rAVG) and the average error of the network rERR change. The graph of the relationship shown in Figure 4 is obtained as shown in Figure 4, as the relative entropy threshold, \(D_0\), increases continuously.

The average residual energy (rAVG) of the network is growing, and the average error (rERR) is also growing. Therefore, to better solve the energy consumption, it is necessary to make \(D_0\) as large as possible; however, to reduce errors, it needs to be as small as possible. In the practical application assumed in this paper, to monitor and warn the environment better and longer, it is necessary to set a reasonable error range and choose the optimal energy-saving strategy. Combined with the literature Xu Chi et al.\[23\] the system test accuracy is ± 0.3°C, and the error of the alarm temperature measurement value of the train on-board equipment set by Yin Keqiang et al.\[24\] does not exceed 0.4°C, and the FCDA (precision configurable the average absolute error of the mean query of the data aggregation algorithm) method is 0.136 6~0.303 0. The error is much smaller than that of the B-based method. Therefore, this paper sets the allowable data error of WSN to 0.300. As can be seen from Figure 4, taking \(D_0 = 0.05\) can satisfy the ideal error condition. To further verify that \(D_0 = 0.05\) is the excellent set value, now keep \(D_0\) unchanged and carry out 30 groups of experiments based on the LEACH-CIE method for 200 cycles of network operation.

Each experiment’s positions of sensor nodes are randomly generated, and 30 groups are obtained. The values of the data average error rERR and the network average residual energy rAVG are shown in Table 2. The data in Table 2 shows that the errors of the 30 experiments are all within the ideal range. According to Figure 4 and Table 2, after determining \(D_0 = 0.05\), in the average case, the network under the LEACH-CIE method still has an ideal energy saving after running for 200 cycles, and the average data error value of 30 experiments is 0.280 and less than the error threshold of 0.300. In a similar application, Literature\[20\] designed an abnormal data-driven data aggregation method (DA-ADD) in WSN based on anomalous data. The adopted strategy is to calculate the absolute value of the difference between the two data, the data more significant than the threshold value is judged as shock data, and only the shock data is transmitted. In addition, when the cataclysm data is generated, an algorithm with a complexity of n2 is used to calculate the support degree between nodes to further determine the cataclysm data’s validity. Next, the DA-ADD method is compared with the LEACH-CIE way (setting \(D_0 = 0.05\)) in terms of error and energy consumption. Figures 5 and 6 are the comparison experiments of 30 groups of average error and average residual energy between the DA-ADD and the LEACH-CIE methods, respectively.

It represents the average error of the network, and the ordinate in Figure 6 depicts the average residual energy of the network. The x-axis of Figure 7 illustrates the average error, the y-axis represents the number of cycles, and the z-axis

![Figure 4: Graph of mean error and average remaining energy of LEACH-CIE varying with Do.](image)

![Figure 5: Comparison of average error between DA-ADD and LEACH-CIE](image)
represents the average residual energy. Figure 7 compares the error and energy consumption between DA-ADD and LEACH-CIE from a three-dimensional perspective. It is not difficult to see from it. This is because average; the LEACH-CIE method has higher average residual energy and lower average error.

Since the DA-ADD method uses the comparison between individual data, the absolute value of the difference reflects the abnormality of the data. In contrast, the LEACH-CIE method in this paper uses the relative entropy value of the probability distribution of each 10 data in adjacent periods to reflect the abnormality of the data. Reflecting the exception of data, both methods achieve the purpose of energy saving through limited transmission (only abnormal data is transmitted). The DA-ADD method dramatically reduces the amount of data transmission in the process of de-redundancy but also significantly increases. The data error obtained by the sink node is reduced. In the 30 sets of simulation experiments, the average error value of the LEACH-CIE method is 0.280 4, the average residual energy value is 0.268 4, while the average error of the DA-ADD process is 0.286 5, the average residual energy value is 0.264 1, combined with Figure 5, Figure 6 and Figure 7, it is not difficult to see that the LEACH-CIE method has higher average residual energy and lower average error than the method in the literature,[24] so it is reasonable to set Do=0.05. Subsequent experiments will Continue this parameter setting value.

**Comparative Experiment**

**Comparison of different performance indicators**

- Comparison of the number of failed nodes in the network after the end of each cycle. As shown in Figure 8: (1) With
the continuous growth of the network operation cycle, the number of sensor node failures under the LEACH, LEACH-CM and LEACH-CIE methods are increasing. However, the LEACH-CIE method proposed in this paper is better than other methods. In the two methods, the number of failed nodes increases more slowly; (2) LEACH-CIE appears as the first failed node at the latest. Table 3 records the performance of the three methods under the same environment and dataset. Line 10 times, each time the first cycle of the failed node occurs. Table 3 Number of cycles failed nodes first appear in 3 methods. From the data in Table 3, we can calculate the average number of cycles of the first failure node in the network under the LEACH, LEACH-CM and LEACH-CIE methods, which are 19.9, 30.1 and 32.8, respectively. On average, the way in this paper has the latest failure node. Therefore, the number of cycles is in line with extending the network life cycle.

**Figure 12**: Distribution of two types of nodes after 100 rounds of network operation under LEACH and LEACH-CM protocols

![Distribution of surviving and failing nodes after 100 rounds of LEACH](image)

**Figure 13**: Comparison of 3 methods under different number of nodes

![Comparison of 3 methods](image)

**Fig.13**: Comparison of 3 methods under different number of nodes
After each round of data transmission, the average value and mean square error of the remaining energy of the network are compared. It can be seen in Figure 9:

1) Under the same conditions, LEACH-CIE has a slower energy consumption rate, so the energy saving effect is more significant;
2) When the average residual energy of the network is as low as close to 0, it means that the network has been paralyzed. It can be seen from the Figure that LEACH-CIE has a longer network survival time. It can be seen from Figure 10 that in the same period, the method LEACH-CIE in this paper has a smaller mean square error value of residual energy in the network, so the energy consumption rate of nodes is more balanced.

• Distribution of “surviving” nodes in the network
  The rectangles in Figures 11 and 12 are both 150×100 rectangles shape area, the abscissa and ordinate in Figure 11 and Figure 12 represent the section.

  The coordinate value of the point in the network, where the coordinates of the sink node are (150, 50). Figure 11 shows the distribution of surviving and failed nodes in the network after running for 100 cycles under the method LEACH-CIE in this paper. As seen from the Figure, compared with LEACH and LEACH-CM, LEACH-CIE has more nodes surviving in the same period, and the distribution is more uniform, so the energy consumption of network nodes under LEACH-CIE is lower and more equal.

  In Figure 12, “.” represents the node still alive, and “×” represents the failed node. It can be seen that after the networks under the LEACH protocol and the LEACH-CM protocol run for 100 cycles, the distribution of surviving nodes is uneven, and the nodes farther away from the sink will fail first.

  The distance formula finds the candidate relay nodes in some areas. By changing the distribution density of nodes due to too fast energy consumption. Through these two points, the purpose of reducing the network’s overall energy consumption and balancing each node’s energy consumption rate is achieved.

**Comparison of Multi-hop Forwarding Routes**

Based on the mean square error of the residual energy in the network, the multi-hop forwarding routing method (MFRCRE), which takes into account the residual energy of the network, is compared with the IHCRA routing algorithm, and the results are shown in Figure 14.

Figure 14 compares the MFRCRE algorithm with the IHCRA algorithm under the same network arrangement, the same clustering structure, and the same network data. It can be seen from the Figure that the multi-hop routing algorithm that takes into account its residual energy has lower residual energy mean square error than IHCRA, which means that in the same period, the network node energy consumption under the multi-hop routing algorithm that takes into account its residual energy is higher balanced.

**Conclusions**

The method in this paper combines the information entropy theory. It uses the relative entropy model to calculate the relative entropy value of the data distribution of the adjacent two node cycles as the basis for judging the data redundancy. The operation of blocking transmission can achieve the purpose of de-redundancy and energy saving of WSN data. In the data transmission process, the node communication distance and energy consumption ratio factors are comprehensively considered, and the node forwarding conditions are set: distance condition and energy consumption ratio condition. Only when the two needs are met simultaneously will the node perform the forwarding task, which balances the energy consumption of each node in the network and avoids the rapid failure of nodes in some areas. By changing the distribution density of nodes
and conducting multiple sets of simulation experiments, compared with the LEACH and LEACH-CM protocols, the network under the LEACH-CIE method in this paper has a longer life cycle, and the node energy consumption is more balanced.

**Reference**


