Wireless Internet of Things System Optimization Based on Clustering Algorithm in Big Data Mining

Jing Guo
Shaanxi Institute of International Trade & Commerce, Xi'an 712046, China

***Abstract*—The rapid development of the Internet of Things (IoT) has highlighted the importance of Wi-Fi sensor networks in efficiently collecting data anytime and anywhere. This paper aims to propose an optimized routing protocol that significantly reduces power consumption in IoT systems based on clustering algorithms. The paper begins by introducing the architecture of Wi-Fi sensor networks, sensor nodes, and the key technologies needed for implementation. It distinguishes between cluster-based and planar protocols, noting the advantages of each. The proposed protocol, DKBDCERP (Dual-layer K-means and Density-based Clustering Energy-efficient Routing Protocol), utilizes a two-layer clustering approach. In the first layer, nodes are clustered based on density, while in the second layer, first-level cluster heads are further grouped using the K-Means algorithm. This dual-layer structure balances the responsibilities of cluster heads, ensuring a more efficient distribution of data reception, fusion, and forwarding tasks across different levels. Simulation results demonstrate that the DKBDCERP protocol achieves optimal performance, with the smallest curve value and the most stable amplitude. It significantly reduces energy consumption, with the total cluster-head power consumption recorded at 0.1J and a variance of 0.1×10^{-4}. The introduction of two election modes during the clustering stage and the adoption of a centralized control mechanism further contribute to reduced broadcast energy loss. This research introduces an innovative two-layer clustering scheme that enhances the energy efficiency of wireless sensor networks in IoT environments. By leveraging clustering algorithms and a network routing protocol optimized through big data mining techniques, the proposed DKBDCERP significantly reduces energy consumption while maintaining communication stability in large- scale wireless Internet of Things (IoT) systems. The optimized routing protocol provides a novel solution for reducing power consumption while maintaining network stability, offering valuable insights for future IoT applications.**

Keywords—Wireless sensor; network routing protocol; clustering algorithm; two-layer clustering; Internet of Things

I. Introduction

The supporting technology of the Internet of Things integrates RFID (radio frequency identification), Sensor technology, Wireless Sensor Networks, intelligent service, and other technologies. The research on the Internet of Things technology is the research on these supporting technologies. In the supporting technologies of the Internet of Things, the emergence of transmission networks with features such as short distance and low power consumption makes it possible to build a ubiquitous network connecting things [1]. Therefore, the lookup of the Wi-Fi sensor community has an irreplaceable role in the area of the Internet of Things. With the non-stop improvement of a variety of conversation technologies, Wi-Fi sensor community technology, which can pick out and attain the records statistics wanted by way of humans at any time, somewhere, and in any environment, has laid a strong basis for the improvement of the contemporary Internet of Things [2]. The wireless sensor community is a multi-hop self-organizing network, which is composed of a giant variety of sensor nodes, which are randomly allotted in monitoring areas and can speak with each other, and is a very necessary technical shape of the underlying community of the Internet of Things [3].

At present, Zhang Lingling et al. proposed a totally dispensed dimension and conversation approach primarily based on match triggering, which permits every node to reap a balance between estimation error and power consumption barring world facts [4]. Luo et al. proposed an intrusion detection algorithm for Wi-Fi sensor networks based totally on laptop learning, which brought the nearby density of information and the distance of record points into fuzzy clustering, enhancing the clustering effectiveness whilst decreasing the clustering convergence time [5]. Liu et al. proposed a quantization strategy for the propagation characteristics of ultrasonic sources. The nodes calculated quantization information according to the quantization strategy and measured values and transmitted the quantization information to the base station, which then estimated the location of the sound source according to the proposed multi-source positioning method based on the possibility mean clustering algorithm [6]. Aiming at this feature of agricultural monitoring, Gao Hongju et al. proposed to apply a clustering algorithm to cluster head nodes for spatial data fusion, and reduce data transmission and energy consumption through clustering [7]. Sun Dayan et al. brought the K-means clustering technique into the positioning hassle of Wi-Fi sensor networks, screened the distance data with giant blunders via cluster analysis, and used the multilateral positioning approach to stumble on and remedy the ultimate distance statistics as the ultimate end result [8].

Despite extensive research on clustering-based WSN routing protocols, most existing approaches suffer from high energy consumption due to single-layer architecture and lack of centralized control. Our work addresses this gap by proposing a two-layer clustering scheme combined with centralized optimization and immune-algorithm-based routing, which significantly enhances energy efficiency and network longevity. Based on the clustering algorithm, this paper proposes an optimized routing protocol for Wi-Fi sensor networks, which notably reduces the strength consumption of IoT systems. Firstly, the shape of Wi-Fi sensor networks, sensor nodes, and the key applied sciences wanted for implementation are introduced. It is pointed out that the two kinds of protocols have

*Corresponding Author

distinctive utility emphases. Cluster-based and planar protocols have their personal advantages, however, cluster-based networks are simpler to study from the benefits of planar protocols. Then, an excessive electricity effectivity WSN protocol DKBDCERP (Dual-layer K-means and Density-based Clustering Energy-efficient Routing Protocol) based totally on two-layer clustering is designed. In the first layer, the nodes of the Wi-Fi sensor community are clustered based totally on density, and in the 2nd layer, the first-level cluster heads are clustered primarily based on K-Means. This double-layer cluster shape can stabilize the tasks undertaken via the cluster heads more, make the center of attention of the first-level cluster head and the second-level cluster head different, distribute the statistics receiving, fusion, and forwarding amongst the cluster heads at all levels, and decrease the power loss of the cluster heads. The remainder of this paper is organized as follows. Section II reviews the architecture and routing protocols in WSN. Section III details the design of the DKBDCERP protocol, including clustering mechanism and routing optimization. Section IV presents the simulation setup and performance evaluation. Section V discusses the implications and limitations of our findings. Finally, Section VI concludes the paper and outlines future research directions.

II. Wireless Sensor Network Routing Protocol

A. Wireless Sensor Network Architecture

From the perspective of the whole network, the structure of WSN consists of three parts: a common sensor node, a sink node (base station), and a user management node. The network structure is shown in Fig. 1.

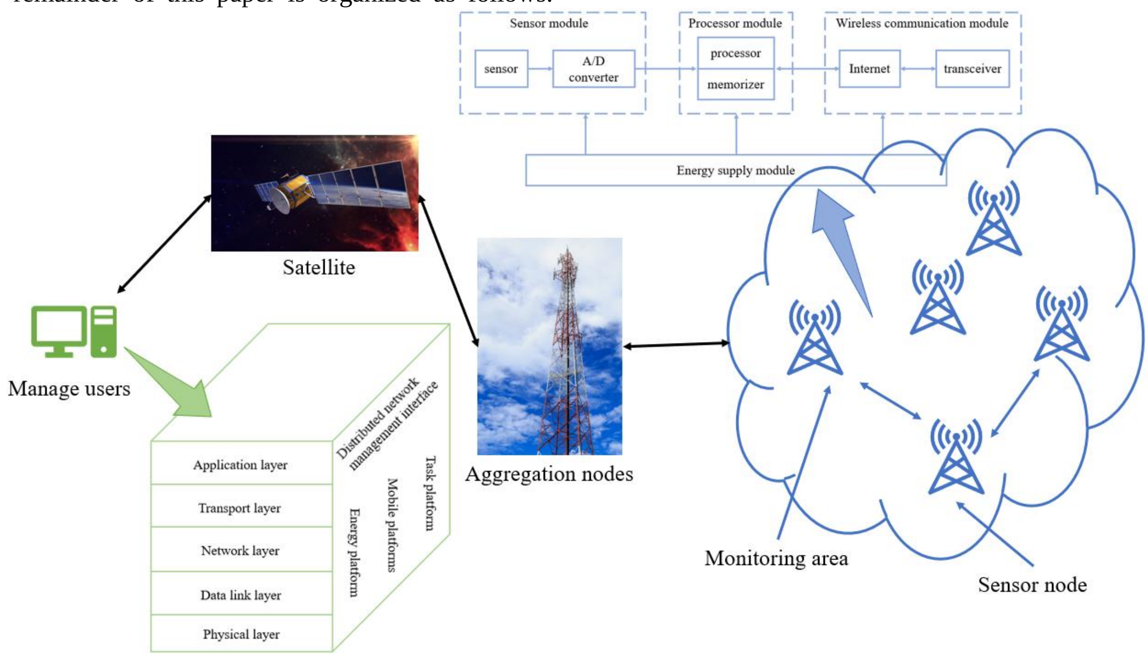

Fig. 1. Network structure of wireless sensor.

As can be seen from Fig. 1, there are sensor nodes in the monitoring area, which are randomly deployed in the monitoring area without any infrastructure support and form a network through their own wireless communication function [9]. Sensor nodes perceive and collect relevant information according to the user's instructions, and each node has the same function. They can communicate with each other, share data information, perform simple fusion processing of the collected data, and then transmit the data to the sink node in the form of a single-hop or multi-hop. Aggregation nodes have no perception ability, but they also have the ability to calculate, store, and transmit data, and this ability is much stronger than that of ordinary nodes [10]. The sink node connects the detection area with the user terminal through satellite and the Internet, transmits the data obtained by the sensor node to the user, receives the user's request for information query, network management, and other tasks, and transmits the command to the sensor node so that the node can collect data according to the user's requirements.

Each sensor node is an embedded design, and the functions of different node designs are different, but the basic structure of the node remains unchanged, mainly composed of four modules: sensor, processor, wireless communication process, and energy supply process [11].

The sensor in the sensor module is responsible for sensing and collecting the information about the monitored object in the monitoring area, and the A/D converter converts the collected analog signal into a digital signal that can be processed by the processor to complete the data acquisition. The processor module mainly processes the information collected by nodes and the information received from other nodes, and then temporarily stores it in the memory [12]. The Wi-Fi conversation module is accountable for the verbal exchange between nodes in the network, and via the Wi-Fi conversation function, the nodes can alternate manage statistics with every other and obtain and ship the gathered environmental information. The power-provided module gives the strength required for the operation of the different three modules via a surprisingly small battery so that the nodes in the community can work normally. Each of the 4 modules is interrelated.

The protocol stack of WSN is similar to the layered structure of other network protocols. The protocol stack is divided into physical layer, data link layer, network layer, transport layer, and application layer from bottom to top. In addition, it also includes three management platforms and distributed network management interfaces, so that sensor nodes can collaborate with each other, so that limited energy can be used efficiently, so that the network can complete the work smoothly in multiple tasks.

B. Routing Protocol

1) Plane protocol: The Information Negotiation Protocol (SPIN) consists of a collection of adaptive protocols. This protocol takes gain of the truth that neighboring nodes have comparable records and honestly distributes facts that different nodes do not. Nodes are assigned a frequent identity to describe the records they collect and are negotiated earlier than statistics are transferred between any nodes [13]. Also, SPIN is in a position to comprehend the modern power degree of the node and regulate the protocol in accordance with the closing energy. These protocols work in a time-driven mode, walking facts throughout the community even if the person has not requested any information. SPIN is designed to tackle the shortcomings of traditional flooding, the use of negotiation, and useful resource adaptation [14]. The SPIN protocol is based totally on two simple ideas: transmitting records that describe perceptual facts is extra environmentally friendly and saves greater electricity than transmitting all data.

SPIN's meta-data negotiation solves the simple flooding problem, which saves a lot of energy. SPIN is a three-step protocol in which sensor nodes exchange three types of messages: ADV (advertisement), REQ (request), and DATA (data transmission). ADV is used to broadcast new DATA, REQ is used to request data, and facts are the data itself. Its special work waft is confirmed in Fig. 2.

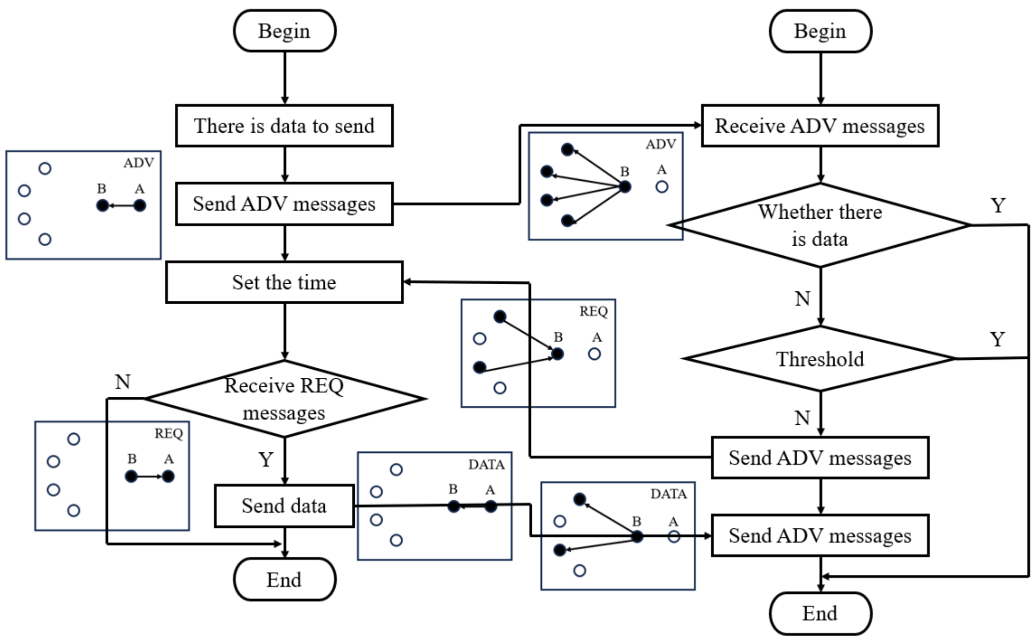

Fig. 2. SPIN workflow.

Directional Fusion Protocol (DD) is a data-centric and application-specific mechanism in the region all facts generated by way of sensor nodes are named by using the capability of attribute fee pairs. The quintessential notion of DD is to mix information from unique sources to restrict redundancy and restrict the variety of required transfers, as a final result saving electrical energy and prolonging the life cycle [15]. Sensors measure occasions and create gradient facts for their respective neighbor nodes. The base station requests statistics through broadcasting interest. Interest refers to the duties that the community is required to accomplish. Interest spreads around the network, hop after hop, every node spreading to its neighbors. As Interest spreads in the course of the network, a gradient is hooked up to describe the statistics that satisfy the question directed to the preferred node [16]. Multiple paths of data drift are formed, and then the fine paths are strengthened to forestall flooding. To decrease conversation consumption, facts are aggregated alongside the path. The intention is to discover a top aggregation course to get the facts to the base station.

2) Cluster routing protocol: By dividing the network into clusters, a hierarchical network topology can be obtained. The so-called cluster is a set of a couple of nodes with positive identical properties, which is composed of the cluster head node and cluster participants [17].

The LEACH routing algorithm is the fundamental algorithm of the clustering algorithm. In order to shop community strength consumption, the LEACH algorithm makes use of the cluster as a unit to transmit records to the base station, which reduces the strength consumption of member nodes in the cluster and prolongs the survival time of nodes. The protocol takes "rounds" as the cycle period, and every spherical consists of the cluster formation stage and the steady transmission stage.

At the beginning of a new round, each node in the neighborhood randomly generates a variation between zero and one. Then the neighborhood compares the price of the node with the threshold to pick out the cluster head node. If the threshold is smaller, it can come to be the cluster head node. After the node is conscious that it is the cluster head node, it informs the neighborhood of the information and then waits for distinctive nodes to maintain the records and select whether or not no longer be a phase of the cluster as a cluster member node in accordance with the distance rule. After that, the community is divided into a couple of clusters, which is the cluster formation stage. This is the stable transmission stage. After a certain amount of time, the network will start a new round of repetitive work.

Borrowing the clustering idea of the LEACH algorithm, the PEGASIS algorithm is clustered in the form of a chain and is divided into a unique cluster containing all nodes in the network. Before forming a chain, nodes in the network broadcast power signals to identify their nearest neighbor nodes, which is essential for constructing an energy-efficient chain topology. Using the greedy algorithm, the node farthest from the base station is regarded as the beginning provider of the chain, and the node closest to the node is decided as the subsequent node. The node already in the chain cannot be chosen over and over [18]. And so on, till the chain consists of all the nodes in the network, after which, the cluster head node is chosen and the cluster ends. In the secure transmission stage, the token ring mechanism is used to manipulate the transmission of data, and the frequent node in the chain transmits the amassed facts to its subsequent hop node, and then the subsequent hop node fuses the facts with its personal facts and continues to transmit it downward in this way. Finally, the cluster head node sends all the information to the base station after fusion processing, till a node dies, the spherical ends, and the community at once enters the formation stage of the subsequent spherical chain.

III. WSN Energy-Efficient Routing Protocol based on Two-Layer Clustering

A. Wireless Communication Model and Energy Consumption Calculation

The Free-space mannequin assumes perfect propagation surroundings with solely one straight, barrier-free direction between the sender and the receiver. H. Friis proposed to use the following equation to calculate the power depth of the obtained sign in free area at a distance of d from the transmitter.

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} \tag{1}$$

In the Free-space model, the communication range is defined by concentric circles centered around the transmitter. A receiver located within this range can successfully receive all packets; otherwise, the transmission fails.

Between two moving nodes, a single straight path is not the only way for a signal to propagate. In the Two-ray ground reflection model, both the linear propagation path and the reflection path of the ground are considered [19]. Over long distances, this model makes more accurate predictions than the Free-space model. When the distance is d, the energy received is approximately:

$$P_r(d) = \frac{P_t G_t G_r h_t^2 h_r^2}{d^4 L} \tag{2}$$

As the distance increases, the strength consumption in the above formula (2) is quicker than in formula (1). However, the two- ray floor reflection mannequin is now not advantageous when dealing with brief distances due to the jitter triggered by the introduction and destruction of the aggregate of two lines. The Free-space mannequin is nevertheless used when the distance d is small.

The electrical energy consumption of the sender is the sum of the strength consumption of the digital computing device of the sending circuit and the strength consumption of the power amplifier, and the energy consumption of the receiver is the strength consumption of the digital machine of the receiving circuit. In the experiment, each the free house mannequin and the double course propagation mannequin are used. When the transmission distance is much less than the threshold, the strength amplification loss adopts the free area model. When the transmission distance is higher than or equal to the threshold, the multipath attenuation mannequin is adopted. The electricity bumps off through the sensor node to transmit Kbit is as follows:

$$E_{Tx}(k,d) = E_{Tx-\text{elec}}(k) + E_{Tx-amp}(k,d) \tag{3}$$

$$E_{Tx}(k,d) = \begin{cases} E_{\text{elec}}{}^{*}k + \varepsilon_{fs}{}^{*}k^{*}d^2 & d < d_c \\ E_{\text{elec}}{}^{*}k + \varepsilon_{mp}{}^{*}k^{*}d^4 & d \ge d_c \end{cases} \tag{4}$$

The energy consumed per Kbit received by the sensor node is:

$$E_{Rx}(k) = E_{Rx-\text{elec}}(k) \tag{5}$$

$$E_{Rx}(k) = E_{\text{elec}} * k \tag{6}$$

In addition, records fusion additionally consumes a sure quantity of energy, and EDA is used to characterize the electricity fed per unit bit of information fusion. We anticipate that the information amassed by way of the neighboring node has excessive redundancy, and the cluster head can fuse its member information into a constant-size packet and then ship it to the sink node.

B. Clustering Algorithm Analysis and Comparison

LEACH is a classic routing protocol designed in 2000. Based on the idea of clustering, the protocol divides all nodes in the network into multiple clusters, which are composed of multiple sub-nodes and a single head node. Each node randomly generates a decimal between 0 to 1, and whether a node can be selected as the cluster head depends on the comparison between the random number generated and the threshold [20]. When the random count is below the thickness threshold, the node is selected as the head node. The calculation of the threshold value T(n) is as below:

$$T(n)=\begin{cases}\dfrac{p}{1-p^{*}\left(r \bmod \dfrac{1}{p}\right)} & n\in G \\ 0 & \text{otherwise}\end{cases} \tag{7}$$

Here, p is the desired percentage of cluster heads, r is the current round number, and G is the set of nodes that have not been selected as cluster heads in the last 1/p rounds. This ensures that every node has an equal chance of becoming a cluster head over time, and prevents the same nodes from being selected repeatedly, which would lead to premature energy depletion.

However, this randomness can still result in suboptimal energy distribution. In this paper improved protocol, an energy- aware election mechanism is introduced to dynamically adjust the thresholds based on the residual energy of each node, distance from the base station and local density. This multi- factor approach ensures that cluster heads are selected more precisely, improving load balancing and prolonging network lifetime.

In LEACH protocol, the basis unit receives the data sent by the first nail, and the first nail senses and fuses the data sent by the child nail. The whole process drains a lot of power, thus making the task heavier for the first node. To balance the energy loss of the network, the deal rotates the opposite nodes. The time cycle is divided into several rounds, one cycle is a cycle, and each cycle contains two processes, namely: cluster head election, and data stable transmission. However, there are some problems with the LEACH protocol [21]. First of all, the head node is frequently selected, resulting in increased network energy loss. Secondly, each and every node has an identical chance of being chosen as the head node, so it is not possible to pick out the high-quality node as the head node primarily based on the cutting-edge proper situation. Finally, the frequent node transmits data to the cluster head, and the cluster head transmits records to the base station, all of which undertake the mode of single-hop transmission, resulting in too lengthy transmission distance and decreasing the community life.

The k-means algorithm is a classical clustering algorithm primarily based on partition in cluster analysis, which was proposed by means of J.B. MacQueen in 1967. The predominant concept of the K-means algorithm is as follows: given the statistics set containing N statistics pattern points and the variety of clustering classes k, ok facts objects are chosen randomly as the preliminary clustering middle point, and the "distance" (similarity) between the unselected facts pattern factors and the core factor of every category (cluster) is calculated, and clusters with shut distance are added. The common of every cluster (class) is then recalculated as the core point. The distance between every facts pattern factor and the core of the new cluster (class) is calculated once more for comparison, the contributors in the cluster (class) are re-adjusted, and the cluster is up to date iteratively. The manner is repeated till the criterion feature converges or the quantity of iterations ends. The intention is to divide the records set containing *n* pattern factors to be labeled into ok training (class), to which every pattern factor belongs, and the distance between every pattern factor and the category middle of the category to which it belongs is minimal in contrast to the middle factors of different classes.

The formula for calculating the cluster center (mean) of each cluster is defined as follows:

$$z_j=\frac{1}{c_j}\sum_{i\in n_j}x_i, i=1,2,\cdots,n,\ j=1,2,\cdots,k \tag{8}$$

The formula for calculating the criterion function (objective function) of the K-means algorithm is defined as follows:

$$J=\sum_{j=1}^{k}\sum_{\substack{i=1\\ x\in c_j}}^{n_j}\ \| x_i - z_j \|^2 \tag{9}$$

It can also be expressed as the formula (10):

$$J=\sum_{j=1}^{k}\sum_{i=1}^{n_j} d_{ij}\left(x_i, c_j\right) \tag{10}$$

Formula (10) is pretty general, representing the distance characteristic between the statistics pair and the center. The targets of formulation (9) and components (10) are the same, and each J can converge to the smallest point, at which time the clustering impact is the best. To sum up, J is a criterion feature expressed through the sum of squares of errors, additionally recognized as the goal feature of clustering. The smaller the cost of the function, the smaller the classification error and the higher the clustering effect.

C. *DKBDCERP, a Dual-Layer Clustering High-Performance WSN Protocol*

DKBDCERP operates in rounds, and each round basically consists of six phases: first-level cluster creation, first-level cluster head election, second-level cluster creation, second-level cluster head generation, inter-cluster route creation, and message delivery. The cluster construction, cluster head determination, and inter-cluster routing are all remembered on the base station calculation control. In the cluster development stage, the first layer of cluster development consists of the usage of DPC-MND to cluster sensor nodes. The 2D layer cluster development consists of the usage of K-Means to cluster the first-degree cluster heads [22]. The 2nd layer of the cluster selects the secondary cluster head, which is used to get hold of the facts from the principal cluster hair and fuse it. According to the distance, power, and deflection Angle of secondary cluster heads, the international most desirable inter-cluster routing is installed with the aid of an elevated immune algorithm.

If nodes are grouped in every round, the ordinary effectivity will be decreased due to giant aid consumption, so in DKBDCERP, the institution of the first layer of clustering is carried out as soon as each and every ten rounds. The institution of the 2d layer cluster is carried out as soon as per round.

DPC-MND is a density-height clustering algorithm based mostly on mutual proximity-MND computes the proximity intensity of the node spread and explores the density top factor of the node spread based totally on the K-nearest neighbor idea. Then, the identified nodes adjacent to the density top factor are grouped into this cluster, and the nests in the clip are accelerated and regressively grouped according to their proximity, resulting in dynamic clustering based on the density of the node distribution.

In large-scale WSNs, the number of members in each cluster is generally very large. If only one cluster head is selected in the cluster, the cluster head will undertake multiple tasks such as receiving, fusion, and forwarding at the same time, and the burden is too heavy, resulting in severe and premature energy loss of the node. If two cluster heads are used in a cluster, energy consumption will be balanced to a certain extent [23]. But it still doesn't minimize energy consumption. Therefore, it is considered to set up a first-order cluster head in the first layer cluster and a second-order cluster head in the second layer cluster. In this way, the overburden of single cluster heads can be reduced, and the advantages of double cluster head theory can be integrated. The first-level cluster head is accountable for receiving packets from regular nodes in the first-layer cluster. The secondary cluster head is accountable for receiving the information packets dispatched by way of the most important cluster hair, and then forwarding them to the base station through multi-hop interclassed routing.

As mentioned above, the first layer clustering in the DKBDCERP protocol is performed once every ten rounds. However, the node energy will be lost each round due to the corresponding task, if the first node is selected once every ten rounds, there will be a round in which the node energy cannot be used up to perform the corresponding task. Therefore, this chapter selects a new first-level cluster head in the first-layer cluster according to the cluster head selection algorithm after dividing the first-layer cluster every ten rounds.

One-level cluster head election factor algorithm:

$$\lambda_{\text{firCH}} = \mu \frac{E_{\text{res}}}{\overline{\overline{E_{\text{res}}}}} + v\left(1 - \frac{d_{\text{tolodes}}}{\overline{\overline{d_{\text{tollodes}}}}}\right) \tag{11}$$

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} \| x - \mu_i \|^2 \tag{12}$$

As an unsupervised clustering algorithm, the K-means algorithm adopts the execution mode of the initial stage and adjustment stage. In the adjustment phase, the center of each cluster is constantly updated, using the mean value of the data object as the center. This is constantly adjusted until the criterion function (objective function) converges without any significant change and the result is output. The overall algorithm framework is relatively simple and practical, with strong expansibility and scalability, and good robustness. It only needs to provide the number k of data sets and classifications. In the case of suitable data amount and obvious quantification of similarity between data, it is relatively fast and efficient.

Input: first-class cluster head set Cfirst, cluster centroid number k

Output: Clustering result C, each cluster corresponds to centroid coordinates (x, y)

Step l: Determine the initial cluster centroid number k;

Step 2: Based on the relationship between the first order clause head Cfirst and the clause centroid, categorize it into the closest centroid;

Step 3: After all first-level cluster heads Cfirst are classified, the center of mass is recalculated according to the formed class;

Step 4: Repeat steps two and three till the algorithm converges, and then output the clustering end result C and centroid coordinates (x, y). At this point, the 2d layer cluster is complete.

In this chapter, after the first layer cluster is divided each ten rounds, a new first-stage cluster head is chosen in the first layer cluster in accordance with the cluster head decision algorithm. Then the first-order cluster heads are grouped and the second- order cluster heads are chosen in accordance to the centroid of the clustering. The clustering of the major cluster head and the decision of the secondary cluster head are carried out in every round.

The secondary cluster head is more often than not used to get hold of the records from the important cluster head and fuse it and in the end multi-hop ahead to the base station. Therefore, it is crucial to consider the remaining strength of the migrant node, the relationship between the migrant node and the base station, and the relationship between the migrant node and the center of mass in the two-stage cluster head election algorithm. A candidate node is an ordinary node except for the first-level cluster head, which is represented as follows:

$$C_{\text{candidate}} \in \text{Nodes} \cap C_{\text{candidate}} \notin C_{\text{first}} \tag{13}$$

Two-level cluster head election factor algorithm:

$$\lambda_{\text{secCH}} = \mu \frac{E_{\text{res}}}{E_O} + v\left(1 - \frac{d_{\text{tocen}}}{d_{\text{diagonal}}}\right) + \gamma\left(1 - \frac{d_{\text{tosink}}}{d_{\text{diagonal}}}\right) \tag{14}$$

When sending records between clusters, the secondary cluster head sends the documents to the base station in a multi- hop manner. In the intercluster routing algorithm, each second-level cluster head selects exclusive second-level cluster heads as the subsequent hop, and the impact on the following parameters needs to be comprehensively evaluated, namely, the residual electrical energy of the candidate cluster head, the distance between the candidate cluster head and the nearby cluster head, and the route of the candidate cluster head. The path is represented via the connection between this cluster head and the candidate cluster head and the Angle between this cluster head and the base station [24]. Using energy, distance, and Angle to pick out the subsequent hop can effectively reduce the range of forwarding hops and verbal trade conflict. Based on the above three factors, this paper constructs the route weight matrix between the secondary cluster head and the base station. Then, in accordance with the matrix, the Dijkstra algorithm is used to generate the preliminary suboptimal path, and the global most environment-friendly route is calculated in accordance with the prolonged immune algorithm.

A numerical two-dimensional routing matrix D[i][j] is described to characterize the directional power from factor i to factor j. i,j= 1,2,... n,n+1. The n cluster heads are sorted according to their distance from the base station, and the base system is in the first place. The directional weights among the factors are expressed as follows:

$$D[i][j]=\begin{cases} 0, i=j \\ 1e-6, i<j\wedge i=0 \\ \mu\left(1-\dfrac{E_{j\,res}}{Eo}\right)+ve^{(d-d_0)/100}+\gamma\dfrac{\theta_{ij}}{\pi}, i<j\wedge i\neq 0\vee i>j\wedge j>0 \\ 1e-6, i=1\wedge j=0\wedge d_{itos}<d_0\vee i>1\wedge j=0\wedge d_{itos}<d_0 \\ ve^{(d-d_0)/100}, i=1\wedge j=0\wedge d_{itos}\geq d_0\vee i>1\wedge j=0\wedge d_{itos}\geq d_0 \end{cases} \quad (15)$$

According to the immune algorithm principle, Dijkstra was used to generate the initial suboptimal path when generating the initial suboptimal path. Finally, according to the distance of each starting node, the next hop is modified, and the globally optimal path satisfying the lowest energy consumption is obtained.

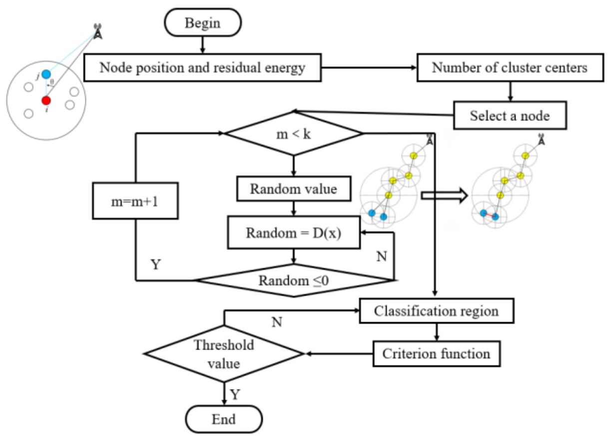

Fig. 3. Algorithm flow chart.

As proven in Fig. 3, when discovering the route from the cluster head to the base station, the route from the cluster head to the base station with a shut distance between beginning nodes solely satisfies the shortest course from this factor to the base station. However, from the perspective of energy consumption of all nodes, it is not the optimal path. Therefore, the immune algorithm is improved in this paper, where the distance between adjacent starting nodes is less than the threshold γ (γ=45 in this paper). When modifying the next hop, check whether the high-energy node already points to the low-energy node. If yes, do not modify the next hop. Otherwise, modify the next hop node to avoid loops.

IV. Simulation Analysis

This chapter uses Python to simulate WSN and analyzes and compares related protocols. In the comparison experiment, LEACH and recent routing protocols EBCRP, KBECRA, and DBSCAN cluster-based routing protocols were used. The reasons for comparison are as follows: LEACH is more classic. The head node selection of EBCRP takes into account the distance between the node and the base station, so as to reduce the phenomenon of energy holes around the base station, which has the same purpose as the protocol in this chapter. KBECRA uses a double cluster head similar to this chapter and also uses K-means clustering. DBSCAN and DPC-MND used in this chapter both belong to density clustering algorithms.

The test in this chapter runs most of 2,000 rounds. If the variety of surviving nodes in the community is much less than 20 per cent, the community shape is viewed to be significantly damaged. At this time, the simulation ends. See Table I for analog settings.

TABLE I Simulation Parameters

Argument	Value
Number of nodes	1000
Network detection range	600*600
Base station coordinates	300, 650
Initial node energy	0.5
Rf energy consumption coefficient	50
Signal amplification energy consumption in the free space model	10
Signal amplification energy consumption under multipath attenuation model	0.0013
Data fusion energy consumption	50
Data fusion ratio	0.6
Control message length	200
Data message length	4000

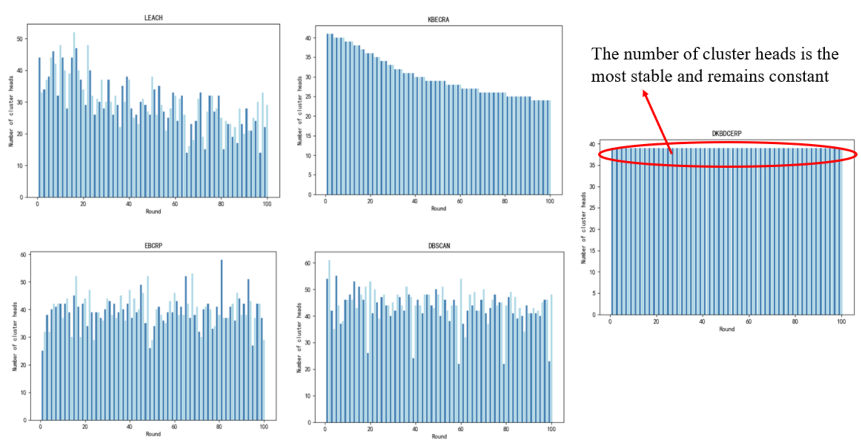

Fig. 4. Cluster head adjustment.

As you can see from Fig. 4, the wide variety of every head varies radically in LEACH. The motive is that the cluster head decision is random in LEACH. Moreover, due to the fact the wide variety of cluster heads is now not optimal, the electricity consumption of the community is increased, the quantity of surviving nodes is reduced, and the quantity of cluster heads is much less and less. EBCRP and DBSCAN reflect on the consideration of the insurance of cluster heads and successfully manage the wide variety of cluster heads. Therefore, the wide variety of cluster heads is very concentrated. Although the quantity of EBCRP clusters is dispensed around 40, it nevertheless fluctuates greatly. The distribution number of DBSCAN cluster heads ranges from 46 to 47 and fluctuates greatly, so the performance of DBSCAN cluster heads in energy consumption is relatively average. Although the wide variety of clusters in KBECRA fluctuates little, due to the speedy power consumption, the range of surviving nodes and the number of clusters additionally reduce rapidly, so the impact is now not good. However, the number of DKBDCERP clusters suggested in this article is steadily assigned to the most suitable ones, with good cluster effects, slow power consumption, and no longer having a wide variety of clusters as soon as KBECRA opens the game. Therefore, the DKBDCERP proposal has superb dependability.

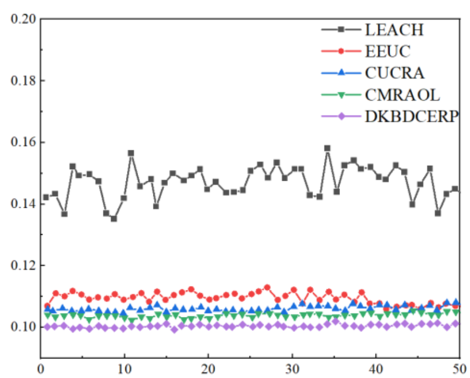

Fig. 5. Cluster head energy consumption.

As seen in Fig. 5, the curve value of the DKBDCERP protocol in this paper is the smallest, and the curve amplitude is the most stable, for the following four reasons. 1) The DKBDCERP protocol is based on grid minimum energy consumption. In the cluster establishment stage, the election mode is changed twice, and the additional broadcast energy consumption is reduced by using a centralized control mechanism. 2) The election of cluster heads tends to be constant cluster centroid coordinates, that is, the distribution of cluster heads is greater uniform and reasonable. 3) The wide variety of participants in every cluster location is fixed, that is, earlier than the node death, the electricity consumption in every cluster head is nearly the same. 4) The hierarchical mechanism is adopted when the cluster head transmits data, which reduces the conversation distance required for transmission. The simulation effects exhibit that the multiplied protocol can limit the electricity consumption of every cluster head.

As can be seen from Fig. 6, the amplitude of the LEACH protocol is the largest, and the curve fluctuation is the most obvious, indicating a large difference in energy consumption between cluster heads. However, the variance of electricity consumption of the EEUC protocol is considerably smaller than that of the LEACH protocol, and its curve fluctuation is smaller than that of the LEACH protocol, due to the fact the cluster dimension of the EEUC protocol is different. Therefore, the strength consumption of all cluster heads in the community is successfully balanced. CMRAOL protocol improves the cluster head election mechanism of EEUC, making it take into account the distance of the Sink. However, in contrast with the EEUC protocol, the cluster head power consumption is reduced. However, the protocol did not enhance the cluster structure, so the electricity consumption hole between cluster heads used to be no longer appreciably improved. It indicates that the strength consumption variance overall performance of the DKBDCERP protocol in this paper is the best, indicating that the power consumption amongst cluster heads is extra uniform, and the purpose is comparable to the strength consumption evaluation consequences of the identical cluster head. In this paper, the non- uniform cluster partition shape is adopted to make the strength consumption of every layer cluster head extra balanced. Moreover, the constant quantity of member nodes makes the strength eaten up by means of intra-cluster conversation extra stable, so the curve amplitude of the protocol in this paper is the lowest and the amplitude adjustments the least.

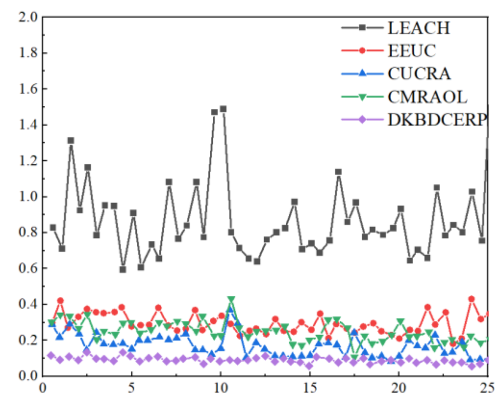

Fig. 6. Variance of energy consumption at cluster head.

V. Discussion

The proposed DKBDCERP protocol demonstrates high energy efficiency and reliable network longevity. Its dual-layer architecture reduces the burden on individual cluster heads and enhances routing flexibility. However, the model assumes homogeneous sensor capabilities and does not account for node mobility or dynamic environmental conditions. Future research should explore heterogeneous networks, adaptive parameter tuning, and machine learning-based optimization to further generalize the protocol's applicability.

VI. Conclusion

A variety of routing protocols for the Internet of Things have been proposed, but most of these protocols are for specific application environments, can only improve some specific performance in the network, and cannot take into account the energy consumption of the network. Based on the K-Means algorithm, DKBDCERP, a dual-layer clustering high-performance WSN protocol, is proposed in this paper, which greatly reduces the energy consumption of Internet of Things systems. Specific conclusions are as follows.

An excessive electricity effectivity WSN protocol DKBDCERP based totally on double-layer clustering is designed. In the first layer, the DPCMND algorithm is elevated via mutual proximity and distance, which can dynamically cluster in accordance with the node distribution density, and make the range of first-level cluster heads nearer to the base station increase, which avoids the power gap around the base station to a sure extent. At Layer 2, mass, residual energy, distance, and intermediate values of different elements are utilized to select the secondary cluster heads, and then a novel course right weight is developed to replace the Euclidean proximity, and the Dijkstra algorithm is augmented by adding the concept of image immunization algorithms to it to get the world's most cutting-edge direction.

Simulation results show that the number of DKBDCERP clusters suggested in this article is assigned gradually at the highest quality worth, the effect of clamping is good, and the intensity is consumed slowly, so that the clusters are now varied, and do not show a decreasing style as KBECRA did at the beginning. Therefore, the DKBDCERP protocol has high reliability.

In this paper, the power consumption variance performance of the DKBDCERP protocol is the best. The whole strength consumption of cluster heads is 0.1J, and the electricity consumption variance is 0.1×10-4, indicating that the strength consumption amongst cluster heads is surprisingly uniform. In this paper, the non-uniform cluster partition shape is adopted to make the electricity consumption of every layer cluster head greater balanced, and the constant quantity of member nodes makes the electricity consumption of intra-cluster verbal exchange extra stable.

In future research, we plan to investigate the integration of reinforcement learning techniques to dynamically optimize clustering and routing decisions. Additionally, we aim to extend the current framework to support heterogeneous sensor networks and adapt to real-time environmental changes. A prototype implementation in a physical IoT testbed will also be considered.

FUNDING

Key R&D Plan Project of Shaanxi Provincial Department of Science and technology; fund number: 2022GY-022.

REFERENCES

[1] Zhong Y, Chen L, Dan C, et al. A systematic survey of data mining and big data analysis in internet of things[J]. The Journal of Supercomputing, 2022, 78(17): 18405-18453.

[2] Lv X, Li M. Application and research of the intelligent management system based on internet of things technology in the era of big data[J]. Mobile Information Systems, 2021, 2021(1): 6515792.

[3] Lee, Joong Ho. “Energy-Efficient Clustering Scheme in Wireless Sensor Network.” International Journal of Grid and Distributed Computing (2018): n. pag.

[4] Zhang Lingling, Zhang Ya. Multi-target estimation triggered by Distributed events in sensor networks [J]. Control Theory & Applications/Kongzhi Lilun Yu Yinyong, 2020, 37(5).

[5] Luo Fucai, Wu Fei, Chen Qian et al. Intrusion detection algorithm of wireless sensor networks based on Machine Learning [J]. Journal of Harbin Engineering University, 2019,41(03):433-440. (in Chinese)

[6] Liu Yunting, Jing Yuanwei, Zhang Siying. Research on multi-source localization in wireless sensor networks based on quantitative information [J]. Journal of University of Electronic Science and Technology of China,2017,46(04):530-533.

[7] Gao Hongju, Liu Yanzhe, Chen Sha. WSN cluster-head node Data Fusion based on improved K-means Algorithm [J]. Transactions of the Chinese Society for Agricultural Machinery,2015,46(S1):162-167.

[8] Sun Dayang, Qian Zhihong, Han Mengfei et al. Improved Cluster Analysis algorithm for multilateral positioning in wireless sensor networks [J]. Acta Electronica Sinica,2014,42(08):1601-1607.

[9] Li, Kai and Kien A. Hua. “Mobility-assisted Distributed Sensor Clustering for energy-efficient wireless sensor networks.” 2013 IEEE Global Communications Conference (GLOBECOM) (2013): 316-321.

[10] Kalla, Neeharika and Pritee Parwekar. “A Study of Clustering Techniques for Wireless Sensor Networks.” (2018).

[11] Wang F, Wang H, Ranjbar Dehghan O. machine learning techniques and big data analysis for Internet of Things applications: A review study[J]. Cybernetics and Systems, 2024, 55(1): 1-41.

[12] Abusaimeh, Hesham et al. “Balancing the Network Clusters for the Lifetime Enhancement in Dense Wireless Sensor Networks.” Arabian Journal for Science and Engineering 39 (2014): 3771-3779.

[13] Pal, Vipin et al. “Optimizing Number of Cluster Heads in Wireless Sensor Networks for Clustering Algorithms.” International Conference on Soft Computing for Problem Solving (2012).

[14] Andronie M, Iatagan M, Uță C, et al. Big data management algorithms in artificial Internet of Things-based fintech[J]. Oeconomia Copernicana, 2023, 14(3): 769-793.

[15] Mitra A, Bera B, Das A K, et al. Impact on blockchain-based AI/ML-enabled big data analytics for Cognitive Internet of Things environment[J]. Computer Communications, 2023, 197: 173-185.

[16] Xie, Qing Yan and Yizong Cheng. “K-Centers Mean-shift Reverse Mean-shift clustering algorithm over heterogeneous wireless sensor networks.” 2014 Wireless Telecommunications Symposium (2014): 1-6.

[17] Qi Q, Xu Z, Rani P. Big data analytics challenges to implementing the intelligent Industrial Internet of Things (IIoT) systems in sustainable manufacturing operations[J]. Technological Forecasting and Social Change, 2023, 190: 122401.

[18] Dutta, Raju et al. “Efficient Statistical Clustering Techniques for Optimizing Cluster Size in Wireless Sensor Network.” Procedia Engineering 38 (2012): 1501-1507.

[19] Rakhshan, Noushin et al. “Energy Aware Clustering Algorithms for Wireless Sensor Networks.” (2011).

[20] Sharma, Ishant and Balpreet Singh. “Energy efficient fault tolerant and clustering algorithm using alternative backup set for wireless sensor network.” 2015 International Conference on Advances in Computer Engineering and Applications (2015): 649-653.

[21] Hou R, Kong Y Q, Cai B, et al. Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning[J]. Neural Computing and Applications, 2020, 32(10): 5399-5407.

[22] Dai H N, Wang H, Xu G, et al. Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies[J]. Enterprise Information Systems, 2020, 14(9-10): 1279-1303.

[23] Fan, Jiande et al. “A Robust Multi-Sensor Data Fusion Clustering Algorithm Based on Density Peaks.” Sensors (Basel, Switzerland) 20 (2019): n. pag.

[24] Bohan Li, Lin Yue, Chuanqi Tao et al. “Web and Big Data.” Lecture Notes in Computer Science (2018).