Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 10, Issue. 10, October 2021, pg.13 – 19

Research on the Deployment Strategy of Blockchain Nodes in the Agricultural Product Blockchain Traceability System

Yanhe Na[1]; Zhan Wen[2]; Haoning Pu[3]; Wenzao Li[3]
[1 2 3]School of Communication Engineering & Chengdu University of Information Technology, China
[1] yanhe_na@163.com; [2] wenzhan@cuit.edu.cn
DOI: 10.47760/ijcsmc.2021.v10i10.003

Abstract— The agricultural product traceability system based on blockchain can monitor the entire growth cycle of agricultural products, trace it at any time, and cannot tamper with information without authorization. However, the energy consumption of the entire system is relatively high due to the introduction of blockchain technology. In order to alleviate the problem of high overall energy consumption, we are trying to reduce the communication power consumption between terminal sensors and the full nodes of the blockchain. We use the K-means algorithm, the DBSCAN algorithm and the improved DK fusion algorithm we proposed to deploy blockchain full nodes to the agricultural products sensors that have been determined to reduce the communication power consumption of the sensor terminals and improve the coverage of the full nodes to the terminal sensors.

Keywords— Agricultural product traceability, Blockchain technology, Deployment strategy, K-means, DBSCAN, Fusion algorithm

I. INTRODUCTION

With the development of economy and society, consumers have higher and higher requirements for the quality of agricultural products such as fruits and vegetables, especially the demand for traceability of organic agricultural products. At present, most of the existing traceability systems are implemented based on traditional centralized databases [1], and data is easily tampered with and attacked. The traceability system using blockchain technology has the characteristics of decentralization, distribution, and non-tampering of data; based on ledger database technology, it can ensure the security and credibility of the data in each link of agricultural products from production, transportation to sales.

At present, blockchain technology is developing fastest in the field of virtual currency, but it is also subject to controversy. The blockchain technology has triggered a lot of technological changes in terms of proof, archiving and traceability. Blockchain technology has gone through three stages: Blockchain 1.0-digital currency; Blockchain 2.0-Programmable blockchain; Blockchain 3.0-Intelligent Internet of Things. The essence of blockchain technology is that it can still make most nodes (devices) agree on a transaction under the conditions of a very poor trust environment, and prevent possible risks, thereby preventing denial, and is based on a special data structure The transaction can be open and transparent, traceable at any time, etc.

Zigbee technology is a new type of short-distance wireless communication technology. The physical layer and MAC layer (data link layer) are defined by the IEEE/802.15.4 standard, the network layer and security layer are formulated by the Zigbee Alliance, and the application layer is determined by the user. Its needs are developed and utilized. The Zigbee device has a transmission output of 0-10dBm, and the actual communication distance is 30-70m. It has energy detection and link quality indication capabilities. The transmission power can be automatically adjusted according to the detection result, which can minimize the energy consumption of the device under the premise of ensuring the link quality. , What is used in our country is the 2.4GHz ISM frequency band.

Li Tianming, Yan Xiang, etc. "Application Research of Blockchain + Internet of Things in Agricultural Product Traceability" [1] discussed some related researches on the integration of blockchain and Internet of Things frameworks into agricultural product traceability by domestic and foreign scholars. Tian Yang, Chen Zhigang, etc. "Overview of the Application of Blockchain in Supply Chain Management" [2] systematically explained the current mainstream application framework and principles of blockchain, and compared the differences between various technologies. Finally, a systematic analysis of the supply chains of different industries is carried out, and the solutions of blockchain in supply chain management of different industries in recent years are comprehensively explained. As for the research on the deployment of blockchain nodes, there have been some related research results. Based on the system model and performance analysis, Yao Sun et al. theoretically analyzed the performance of the blockchain IoT system, and considered the Poisson point process (PPP) modeling in the spatio-temporal domain. The SINR probability density function (PDF) of node-to-full-node transmission is analyzed and calculated, and the transaction success rate and overall communication throughput are calculated. The derived analysis model proposes a search algorithm for finding the optimal blockchain full node deployment under a given IoT node density and blockchain transaction throughput. In terms of terminal sensor energy saving, "a node dormancy scheduling algorithm that effectively prolongs the lifetime of WSN" [4] proposed a sensor dormancy scheduling algorithm to extend the lifetime of WSN networks. Zhang Dejing wrote in "Energy-saving cluster routing algorithm for wireless sensor networks" Research" [5] proposed a routing algorithm based on particle swarm optimization to optimize fuzzy logic to reduce the energy consumption of communication between nodes.

However, there is still a lack of research on how to reduce the energy consumption of the agricultural product blockchain traceability system in transmission. According to Bai Fei’s "Research on Energy-saving Technology for Wireless Sensor Networks Based on Zigbee Technology" [6], we learned that the ratio of communication energy to computational energy consumption for the REM TR100 module is about 190, in other words at a distance of 100m The energy consumption of transferring 1KB of data is equivalent to that of executing 3 million instructions. Therefore, it is necessary for us to study the deployment strategy of the full node. On the one hand, it reduces the energy consumption of terminal sensors in communication, and on the other hand, it improves the coverage of the full node of the blockchain to the terminal equipment.

II. SYSTEM MODEL

We assume that the communication between the agricultural terminal sensor and the blockchain full node is based on the Zigbee/IEEE 802.15.4 protocol, and its full function device node (FFD, Full Function Device) is equivalent to the blockchain full node, and its reduced function equipment (RFD, Reduced Function Device) is an agricultural terminal sensor device. We mainly focus on reducing the communication power consumption of the agricultural terminal sensor to the full node of the blockchain, balancing the load of each node, and maximizing the coverage of terminal equipment.

Agricultural product sensor terminals (such as humidity, temperature, carbon dioxide concentration, air pressure sensors, etc.) are all RFDs. Since RFD cannot store and forward information, it can only communicate with FFD devices, so its power consumption is smaller than FFD. To minimize the power consumption of wireless sensors. FFD is deployed together with the full nodes of the blockchain, and one of the FFDs acts as a coordinator. The system transmission model is shown in Figure 1:

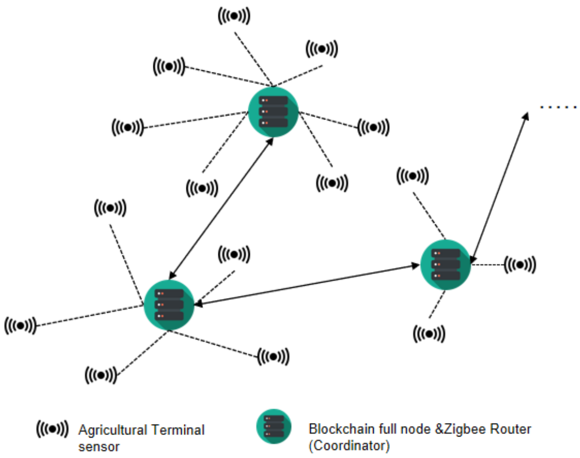

Fig. 1 Transmission model of agricultural product blockchain traceability system

The terminal RFD sensor and FFD use Zigbee protocol to communicate, and the communication frequency is around 2.4GHz; FFD and FFD use P2P (peer-to-peer network technology) for communication, each FFD is a blockchain full node, which combines sensor data After being packaged into blocks and undergoing consensus, they are synchronously recorded to each blockchain node, and it is ensured that the entire data record exists in the form of a single chain, and there will be no fork chains. The system structure diagram is shown in Figure 2:

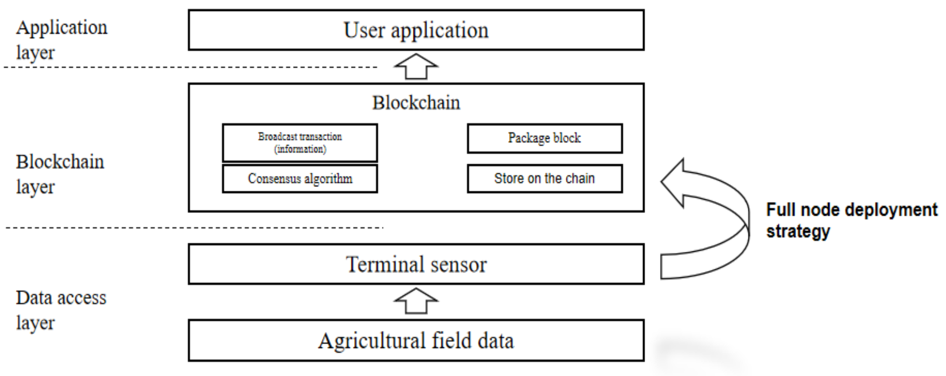

Fig. 2 Blockchain traceability system structure diagram of agricultural products

At present, most of the current blockchain agricultural product traceability systems [1][2] are the three-tier structure shown in Figure 2. Therefore, we have also determined that the system model has a three-tier structure, and the deployment and transmission of the data access layer to the blockchain layer is the direction of this research.

III. Deployment Strategy

This article mainly uses three algorithms to deploy a full node when the position of the agricultural product sensor has been determined. They are k-means algorithm, DBSCAN algorithm and our improved DK fusion algorithm.

A. K-means and DBSCAN algorithm

K-means is an unsupervised clustering algorithm. It usually uses Euclidean distance to characterize the similarity between samples. The smaller the distance, the greater the similarity. The algorithm process is mainly:

Input: database containing n objects, K value;
Output: all generated clusters;
1. Determine the K value (you can start from a certain value, and then use the elbow method to determine the best K value);
2. Randomly select K initial mass points from the sample points;
3. Calculate the distance between each point and the initial point, and divide the initial mass point with the smallest distance into a cluster;

4. Take the average of all points in each cluster and become a new mass point;
5. Repeat 3 steps until the mass point no longer changes.

K-means is a relatively common clustering algorithm, which works well for distance-sensitive clustering problems. The disadvantage is that it is very sensitive to the selection of initial particles.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) defines clusters as the largest collection of tightly connected points, divides sufficiently dense areas into clusters, and searches for clusters of any shape in a noisy spatial database.

However, the algorithm is very sensitive to the radius e and the minimum number of minPts in each cluster. The algorithm process is mainly:

Input: database containing n objects, radius e, minimum number minPts;

Output: All generated clusters meet the density requirement.

1. Detect the object p in the database that has not been checked. If p has not been processed (classified as a cluster or marked as noise), check its neighborhood. If the number of objects contained is not less than minPts, create a new cluster C, and All points in it are added to the candidate set N;
2. Check the neighborhood of all objects q that have not been processed in the candidate set N. If they contain at least minPts objects, add these objects to N; if q does not belong to any cluster, add q to C;
3. Repeat step 2, continue to check the unprocessed objects in N, the current candidate set N is empty;
4. Repeat steps 1-3 until all objects are classified into a certain cluster or marked as noise.

DBSCAN is very suitable for dealing with noisy clustering problems, but it is very sensitive to the selection of parameters e and minPts. Inappropriate selection will make the clustering effect unsatisfactory.

B. Improved DK fusion algorithm

Aiming at the problems of K-means and DBSCAN clustering algorithms, we propose an improved DK fusion algorithm to take the advantages of both, and try to avoid the shortcomings.

First run the K-means algorithm. Then set the DBSCAN parameters according to the k-means results to make the number of clusters close to the preset value (all the points that are not assigned to any cluster are counted as one cluster), and finally run the clustering results with the DBSCAN with the parameters set, according to each The ratio of the number of points in a cluster to the total number of points is randomly selected from each cluster as the initial mass point of the K-means algorithm. The pseudo code of the improved DK fusion algorithm is as follows:

Algorithm 1 Improved DK fusion algorithm

Input: X Coordinate point set, N Number of coordinate points, $minpts$, eps,K

Output: $unite$List of clustering results

```
function IMPROVED DK FUSION ALGORITHM(X, N, minpts, eps)
    Dresult ← DBSCAN(x, eps, minpts)
    K ← len(Dresult)
    presum ← 0
    for i = 0 → len(Dresult) do
        rate ← len(Dresult[i])/N * K
        intnum ← int(rate + 0.5)
        if intnum! = 0 then
            mean[] ← from Dresult[i] randomly draw int points
            presum ← intnum + presum
        else
            temp[] append Dresult[i]
        end if
    end for
    if len(mean)! = K then
        mean[] ← from temp[] randomly draw (K − presum) points
    end if
    unite[] ← K − means(mean[], K)
end function
```

This greatly limits the selection of k-means initial mass points, and the selection of initial mass points will be more uniform. It also avoids the problem that the K-means algorithm will select too many initial particles on the noise points, which leads to poor clustering results.

IV. EXPERIMENTAL RESULTS

We use python to simulate the entire experiment, where the parameters of the terminal sensor refer to the real parameter settings of DIGI's Zigbee chip Xbee Zigbee S2C.

A. Power calculation and parameter setting

The simplified Friis transmission formula under the free space loss model is:

$$\mathrm{Pr} = \mathrm{Pt} + \mathrm{Gt} + \mathrm{Gr} + 20\log_{10}\left(\frac{\lambda}{4\pi \mathrm{R}}\right) \quad (1)$$

To further simplify, separate the constant term in the logarithm and the power is in dBm, then:

$$Loss(dB) = 32.44 + 20\log D + 20\log F \quad (2)$$

$$Loss(dB) = Ps(dBm) - \mathrm{Pr}(dBm) \quad (3)$$

Among them, Ps stands for transmit power, Pr stands for receiving sensitivity, D stands for propagation distance (km), and F stands for frequency (MHz).The conversion between the unit dBm and mW is as follows:

$$(dBm) = 10\log_{10}(mW) \quad (4)$$

From the above formula, we can convert between distance and power.

The technical documents of the DIGI company's Zigbee chip Xbee Zigbee S2C use the enhanced mode of the Xbee Zigbee S2C chip, Pr is -102dBm, and the maximum transmit power of the RF module is 6.3mW.

B. Experimental results and analysis

First, we randomly generated 50 sets of two-dimensional coordinate points in the range of 0-100, using this set of data as a fixed agricultural product detection sensor, and the scale is 1:100 meters. The following figure is the average value obtained after 1000 simulation runs. Three algorithms are used for full-node deployment, and the communication power comparison is shown in Figure 3.

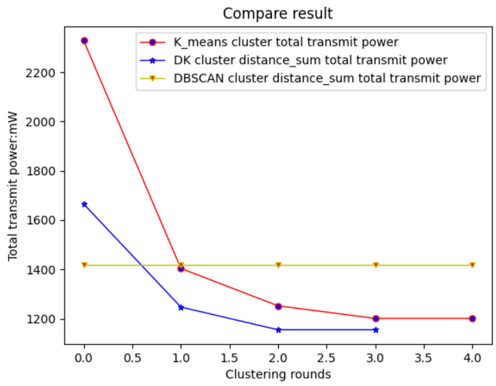

Fig. 3 Blockchain traceability system structure diagram of agricultural products

The abscissa is the clustering round, and the ordinate is the sum of the power of each node communicating with all the nodes after each round of clustering. DBSCAN algorithm clustering does not require iteration, and we will extend it appropriately for comparison. The number of clustering iterations of the improved DK fusion algorithm is lower than that of the K-means algorithm. And the communication power consumption generated by the final clustering is lower than the other two algorithms. The total communication power consumption of the improved DK fusion algorithm is about 18.47% lower than that of the DBSCAN algorithm, and about 3.83% compared to the K-means algorithm.

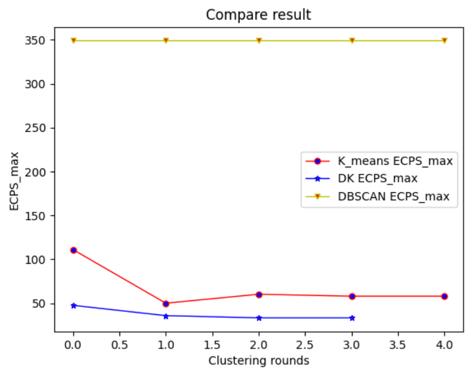

Fig. 4 Curve of maximum communication power consumption in all clusters (ECPS: Each cluster power sum)

Under different clustering algorithms, the maximum communication power consumption curve in all clusters represents the degree of optimization of the cluster with the largest communication power consumption by different algorithms. It can be seen that the DBSCAN algorithm does not optimize the cluster with the largest communication energy consumption. K-means optimizes the cluster with the largest communication energy consumption. However, due to multiple iterations to select the centroid, the optimization may rebound and the effect is unstable. The improved DK fusion algorithm continuously optimizes the cluster with the largest communication energy consumption and the final effect is better than the above two algorithms. Compared with the maximum communication energy consumption of DBSCAN, the cluster power consumption is reduced by about 90.46%, and the cluster power consumption of the maximum communication energy consumption of the K-means algorithm is reduced by about 42.58%. This proves that the improved DK fusion algorithm proposed by us is effective in The optimization effect of the cluster with the largest communication energy consumption is better.

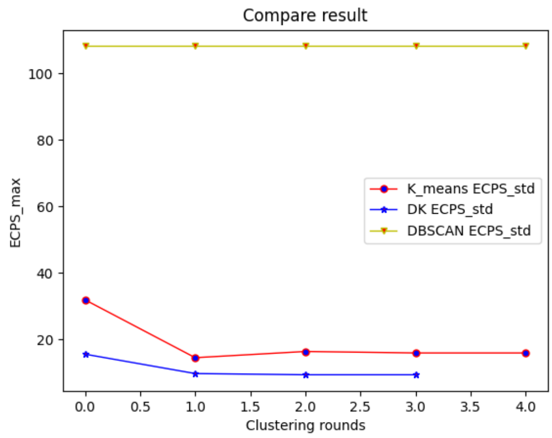

Fig. 5 Standard deviation curve of energy consumption for each cluster

The standard deviation of the power sum of each cluster mainly measures the fluctuation level of the communication energy consumption between clusters. The curve trend is similar to the curve trend of the maximum communication power consumption in all clusters. Mainly because the maximum value has a great influence on the standard deviation, the cluster with the largest communication power consumption can be optimized, and smaller fluctuations in the communication energy consumption between clusters can be obtained, so that the communication energy consumption between clusters can reach balanced.

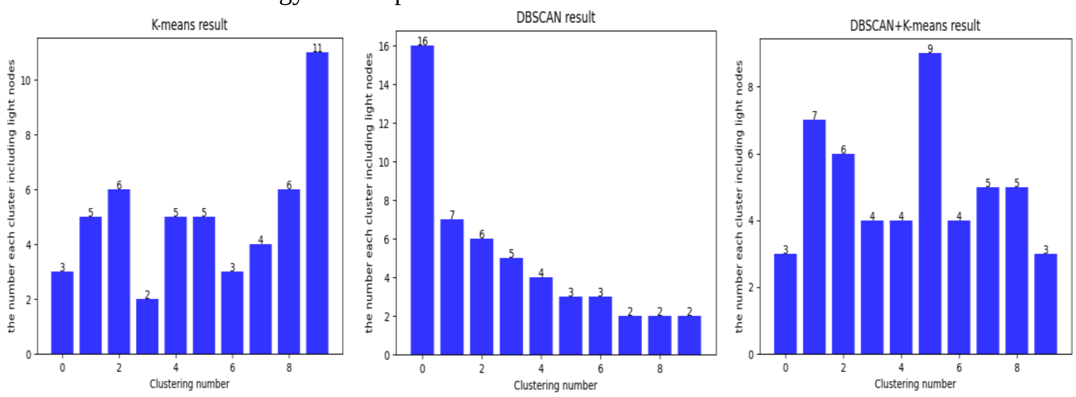

Fig. 6 Standard deviation curve of energy consumption for each cluster

TABLE I

THE STANDARD DEVIATION OF THE FINAL NUMBER OF RFDS CONTAINED IN EACH CLUSTER UNDER DIFFERENT ALGORITHMS

	K-means	**DBSCAN**	**Improved DK fusion algorithm**
Standard deviation	2.366	4.025	1.789
Coverage	96%	78%	99%

From Figure 6 and Table 1, we know that the most balanced load among the three deployment algorithms is the improved DK fusion algorithm. The range is 6, the range of K-means is 9, and the range of DBSCAN is 14.

This is determined by the standard The difference table can also be seen, so the improved DK fusion algorithm we proposed is also the deployment algorithm that contains the most balanced number of RFD nodes.

Coverage calculation is the farthest communication distance calculated from the actual chip parameters when the final clustering is completed. If a certain agricultural product sensor (RFD) exceeds the farthest communication distance, it is judged that the point is not covered, and then the calculation is made The coverage rate of the whole node of the blockchain to all agricultural products sensors, we can see that K-means is not much different from the improved DK fusion algorithm, but the improved DK fusion algorithm can almost achieve full coverage, which proves the improved DK fusion proposed by us Algorithm deployment strategy is better than traditional clustering algorithm deployment strategy under comprehensive comparison.

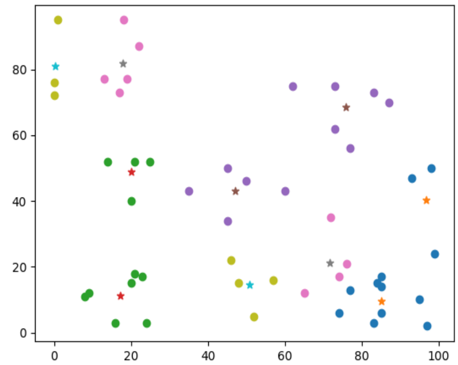

Fig. 7 Standard deviation curve of energy consumption for each cluster

The final deployment result of the improved DK fusion algorithm is shown in Figure 7, where different colors represent different clusters, and the five-pointed star represents the deployment position of all nodes in the cluster.

V. CONCLUSION

We considered the overall power consumption problem in the agricultural product blockchain traceability system. We focused on the reduction of the communication power consumption from agricultural product sensors to the blockchain full node (FFD). We built a system model and designed an experiment. In the experiment, we use actual product parameters and Friesian transmission equations to determine our power consumption after deployment. We use three algorithms to deploy full nodes. Among them, the improved DK fusion algorithm we proposed has the best overall effect, in terms of total energy consumption. It is about 18.47% lower than the DBSCAN algorithm strategy and 3.83% lower than the K-means algorithm. It also has a good effect in optimizing the maximum communication energy consumption cluster and node load. The coverage rate can almost achieve full coverage. In addition, our deployment plan can implement together with the sensor sleep strategy to achieve better energy-saving effects.

REFERENCES

[1] Li Tianming, Yan Xiang, Zhang Zengnian, Tian Yang, Wu Xin, Li Chaoqun. Application research of blockchain + Internet of Things in traceability of agricultural products [J/OL]. Computer Engineering and Application: 1-14 [2021-10- 01].

[2] Tian Yang, Chen Zhigang, Song Xinxia, Li Tianming. Overview of the application of blockchain in supply chain management [J/OL]. Computer Engineering and Applications: 1-18 [2021-10-01].

[3] Y. Sun, L. Zhang, G. Feng, B. Yang, B. Cao and MA Imran, "Blockchain-Enabled Wireless Internet of Things: Performance Analysis and Optimal Communication Node Deployment," in IEEE Internet of Things Journal , vol. 6, no. 3, pp. 5791-5802, June 2019, doi: 10.1109/JIOT.2019.2905743.

[4] Liu Caixia, Li Shude. A Node Dormancy Scheduling Algorithm for Effectively Extending the Survival Time of WSN [J]. Journal of Guilin Institute of Aerospace Industry, 2020, 25(01): 21-24.

[5] Zhang Dejing. Research on Energy-saving Clustering Routing Algorithm for Wireless Sensor Networks [D]. Chang'an University, 2020.

[6] Bai Fei. Research on energy-saving technology of wireless sensor network based on Zigbee technology [D]. National University of Defense Technology, 2008.