Blockchain + D-S Evidence Theory Approach to Traceability of Agricultural Products

ZHOU Zihan*[1], WANG Cong[1,2], LIU Chang[1]

[1]School of Shipping Economics and Management, Dalian Maritime University, Dalian, 116026, China
[2]School of Economics and Management, Southeast University, Nanjing, Jiangsu, 210096

Abstract. The establishment of an agricultural product traceability system is of great significance to the supervision of agricultural product quality and safety. In response to the problems of complicated operation and low efficiency of existing agricultural product traceability methods, the agricultural product traceability method of blockchain + Dempster-Shafer(D-S) evidence theory is proposed. Through testing and analysis of accuracy and efficiency, the results show that accurate traceability can be achieved through the efforts of all parties in the market environment and the role of the network; in terms of performance, the write throughput of the Blockchain + D-S evidence theory approach and using a clustered traceability approach are almost identical; the highest query throughputs of the two approaches are 251bit/s and 164bit/s, respectively. The query throughput of the Blockchain + D-S evidence theory approach is higher, whereas the time delay of the data query of the Blockchain + D-S evidence theory approach is 368ms, which is much lower than that of the query of using a clustered traceability approach(1375.6ms). The research results have certain significance for the expansion of the blockchain traceability network and the collaborative development of agricultural product traceability.

Keywords: Blockchain; D-S Evidence Theory; Agricultural Product Traceability; Accurate Traceability

1 Introduction

The rapid progress of Chinese society in recent years has brought about rapid economic development and great improvements in quality of life. However, at the same time, as the structural reform on the supply side of agriculture continues to deepen, the variety of agricultural products in China has become more abundant and diverse, and residents' demand for agricultural products has shifted from the pursuit of quantity to the pursuit of quality. In particular, residents are more concerned about whether there are excessive pesticide and veterinary drug residues and illegally added toxic and harmful substances

---

About the authors:
Zhou Zihan, E-mail: zhouzihan@dlmu.edu.cn; Wang Cong, E-mail: wangcong2017@dlmu.edu.cn; Liu Chang, E-mail： liuchang0817@dlmu.edu.cn.

A. Rauf et al. (eds.), *Proceedings of the 3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, Atlantis Highlights in Engineering 20,
https://doi.org/10.2991/978-94-6463-262-0_42

in agricultural products. The process of agricultural products from cultivation to sale involves farmers, purchasing and processing enterprises, logistics companies, consumers, and other links. The quality control and information traceability involved are particularly complex, and the credibility of quality and safety information needs to be further improved[1].

China has invested many human and material resources into (a) managing the quality and safety of agricultural products, (b) continuously carrying out special rectification actions, and (c) improving the construction of a modernised governance system to monitor the quality and safety of agricultural products, thus forming a positive cycle and a sustainable traceability ecology. China had also invested resources into providing timely and accurate warnings and preventing risks to the quality and safety of agricultural products. In addition, the quality of each farmer's produce varies, making it difficult to effectively promote even high-quality products. At the same time, the quality of crops on the market varies due to legal constraints and insufficient supervision, and some enterprises and individuals are obsessed with maximising their profits, leading to numerous problems in the agricultural market and a lack of trust in the authenticity of agricultural products. In order to monitor the quality and safety of agricultural products in a more reasonable manner, it is important to build a set of efficient and accurate traceability network for agricultural products.

Traceability is the process of tracing the origin of a transaction from downstream to upstream. Current traceability techniques fall into two main categories. One is traceability by product characteristics, which requires the use of special equipment and is very costly. For example, Wang[2] used stable isotope and DNA techniques on lamb to rapidly trace the origin of lamb, and Mamede[3] traced the origin of the seaweed Ulva spp. (commonly known as sea lettuce) to its origin by comparative analysis of three major elements and seven trace elements in the green seaweed Ulva spp. Vanderschueren[4] enableds the traceability of chocolate varieties and origins by comparing the concentration of elements such as cadmium and lead in chocolate. The other category is the use of Internet of Things(IoT) technology for traceability, but this requires the installation of a series of sensors and electronic tags, which is a complex and costly process to implement. Jin[5] studied the traceability of agricultural products through the use of barcode recognition, whereas Ma[6] uses microbial spores designed as barcodes to trace the origin of products. Yan[7] and others used the federated chain model and Hyperledger blockchain platform to trace the origin of steel products; Wang[8] formed a chain that can be traced back to the origin of the product by means of a smart contract permanently recorded in a distributed ledger. The literature[9-11] provides traceability through RFID technology for food, electronics recycling, and electromechanical products, respectively. Scholars[12,13] have also built a traceability system for the quality and safety of pork and wine products through the use of QR codes.

Traceability of agricultural products is a process of tracking agricultural products throughout their life cycle from cultivation to distribution. Compared to other product traceability methods, agricultural traceability can be divided into four categories, the first of which is traceability through near-infrared spectroscopy. This technique can identify the main constituent characteristics of almost all organic matter, and its detection is fast and does not damage the sample in any way. However, infrared spectroscopy reflects the content and composition of organic components in food, which can be altered by processes such as storage and processing, making the results different from the spectrum of the food's origin and thus producing errors. Similar to Cozzolino[14], Guo [15] determined the feasibility of near-infrared spectroscopy for origin traceability of cereals and sea cucumbers. The second category of agricultural traceability is traceability by means of stable isotopes. This method is safe, physically stable and less susceptible to external environmental influences, but it only partially differentiates the product and requires a high level of equipment. Like Zhao[16], Zhao [17]applied the method of stable isotope detection to the traceability of agricultural products by differentiating the stable isotope composition in agricultural products. The third category is the traceability of agricultural products through mineral element fingerprinting. This technique focuses on finding mineral element indicators with significant direct geographical differences, and then screening these elements through a series of methods to identify the elements most conducive to traceability of origin. Chen[18] conducted a traceability study of the origin of wheat by this method. The fourth category of agricultural traceability is the application of electronic nose technology to the traceability of agricultural products. Electronic nose technology has the advantages of simple operation and accurate traceability, but it is susceptible to environmental factors such as external temperature and humidity, as demonstrated in the literature.[19] These systems have achieved the basic needs of agricultural product traceability and solved the problems of traditional traceability systems to a certain extent, but all these methods require highly sophisticated equipment to achieve. Moreover, the cost required is costly, and because the testing results are easily affected by other factors, the traceability results have a certain degree of error. Therefore, it is of great significance to apply blockchain technology with characteristics such as decentralisation and tamper-proof to agricultural product quality and safety traceability systems.

Blockchain was proposed by Nakamoto.[20] There is no precise definition of blockchain in the industry, but it is essentially a combination of two main logical mechanisms: the hashing mechanism[21] and the consensus mechanism.[22] A blockchain can be thought of as a public ledger where all submitted transactions are stored in a list of blocks. This chain grows as new blocks are added.[23] In this way, each block is given the correct timestamp when it is added to the chain.[24] In the early days of blockchain, a consensus structure based on Proof of Work was used, which was highly scalable but had terrible performance[25]; this method is mainly implemented when using Bitcoin. Since then, a large number of new technologies have been innovated to make up for the shortcomings of the blockchain. Luu[26] et al. investigated the introduction of new distributed protocols that improve computational performance by linearly scaling the transaction rate based on the available mining computations. Li[27] achieved a huge breakthrough in storage efficiency as well as trust in the blockchain based on polynomial-encoded sharding. In addition, blockchain is being used in an increasing number of areas. Chelladurai[28]developed a new automated system for healthcare electronic health records that allows for the seamless transfer of medical information from multiple hospitals to patients in parallel. With the development of private chains, federated chains and hybrid blockchains[29], there will be more and more application scenarios for blockchains and people will not be able to live without blockchains.

Because blockchain uses distributed computing and storage, there is no centralized supervisory authority, and power is equally shared by all nodes on the chain, which has decentralised characteristics. Therefore, blockchain has great prospects for development in the field of traceability, and relevant literature on the use of blockchain in the traceability of agricultural products has been discussed. Zheng[30] used mathematical modelling and simulation methods to study the decision to adopt blockchain traceability in agriculture and analyzed the optimal blockchain-based traceability strategy for members of the agricultural supply chain under different scenarios to achieve an intrinsic fit between blockchain technology and agricultural quality and safety traceability. Yan[31] used the technology in the alliance in blockchain to achieve a decentralized and in-depth traceability process by enterprises entering various types of data of agricultural products into the chain and then using the alliance chain to record the circulation data of agricultural products. This blockchain-based traceability system can effectively ensure the authenticity and non-tamperability of traceability data compared with the ordinary verification and retrieval method using a database. These technologies can already be used to trace agricultural products with basic accuracy, but all of the above are characterised by complex implementation and low efficiency. An efficient traceability system can help enterprises shorten traceability time and save time costs, which is of great significance to the development of enterprises. Last, the development of blockchain technology provides a new technical means for efficient and fast traceability of agricultural products. However, the use of blockchain technology for traceability comes with challenges, such as large amounts of data, complex processing and low traceability efficiency due to excessive blockchain storage space occupation.

In response to the above problems, this paper attempts to design a method that is simple to operate and efficient to achieve the goals of authenticity, integrity, safety, and efficiency in the traceability process to ultimately reduce food safety risks.

This paper differs from the above studies in that it innovatively uses blockchain technology in combination with Dempster-Shafer(D-S) evidence theory for traceability seeking. The main contributions are as follows:

1. The design is based on blockchain technology to directly chain the agricultural data collected by IoT devices, which can solve the problem of inaccurate data due to the front-end production information of agricultural products being entered manually, and achieve the non-tamperability of traceability data.
2. Using the D-S evidence-based approach, the data from the sensors are combined to calculate the level of confidence that a problem has arisen at each stage. This method effectively and quickly traces faulty produce to prevent its recirculation, thus ensuring consumer safety; at the same time, it makes information about the entire chain of produce transparent, guaranteeing the safety of produce consumption.

2 Problem Description and Assumptions

This paper provides a complete description of the blockchain + D-S evidence theory approach to agricultural traceability by describing the process of solving a specific example, thus providing a more graphic description of the blockchain + D-S evidence theory approach to agricultural traceability.

2.1 Description of the Problem

Suppose an agricultural product goes through the five stages from planting to purchase by consumers as shown in Figure 1. The farmer grows the product, a company comes to buy and process it after harvest, then it is stored in a warehouse and waits to be sent to distributors around the world. In this trading chain, every link is essential, and this article looks at how technicians can quickly trace the quality of a batch of produce when it becomes problematic?

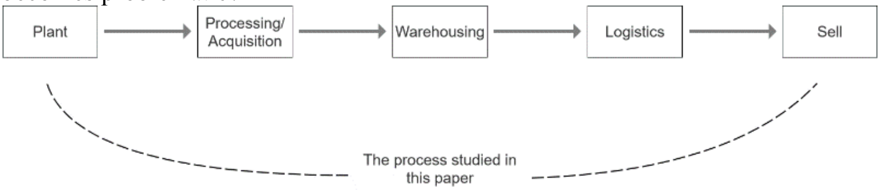

Fig. 1. Five links in the distribution of agricultural products

There are many ways to trace agricultural products alone, such as using a mixture of rough set and decision tree algorithms, as well as researching classification algorithms such as support vector machines, neural networks, plain Bayes and Bayesian belief networks, and using different algorithms for different situations based on their respective strengths and weaknesses to ultimately mine the desired data. This process consumes a lot of time and money and is extremely complex to implement, whereas the blockchain + D-S evidence theory approach allows for the identification of a problem at one point in the process and then the cleaning and processing of the data in this process, which can then be compared to the normal data to reach a conclusion. The problem is therefore transformed into one of identifying where a batch of produce has a quality problem.

2.2 Assumptions

In order to solve the above problem, it is necessary to document the various links in the flow of agricultural products. In this paper, according to the D-S theory of evidence, the setting of evidence sources is the key to solving the problem. Therefore, the sources of evidence need to be comprehensive and independent, and they should be set from multiple sources, such as identifying one relevant piece of evidence from each link and then identifying several sources of evidence from other places outside the link. This paper assumes that the sources of evidence are farmers and regulators.

- Define The Hypothesis Space:

The object under examination in this paper is the five links in the distribution of agricultural products and therefore defines the non-empty set $\theta = \{P_1, P_2, P_3, P_4, P_5\}$. Every two of these elements are mutually exclusive. Thus the power of the non-empty set "θ" has 2^5 elements, of which only a part is taken as the hypothesis space in this paper.

1. Problems in the planting chain,
2. Problems in the processing/acquisition chain,
3. Problems in the storage chain,
4. Problems in the logistics chain,
5. Problems in the sales chain,
6. Problems at every turn, and
7. Empty sets.

3 Constructing a D-S Theory of Evidence

3.1 Determining BPA

Basic probability assignment(BPA) refers to the process of calculating the underlying probability of each piece of evidence in theta. The BPA on the hypothesis space is a function m of $2^{\theta} \rightarrow [0,1]$, called the mass function, and the value of the mass function reflects the degree of confidence in the hypothesis. In contrast, the determination of the value of the mass function requires a combination of multiple data and influencing factors from each source of evidence. The basic probabilities of assignment for each hypothesis space in this paper are shown in Table 1, of which $\mathrm{m_F} \rightarrow \mathrm{m_{Farmer}}, \mathrm{m_G} \rightarrow \mathrm{m_{Regulatory\ Department}}$.

Table 1. Basic allocation probabilities BPA

Suppose	$\mathrm{m_F}$	$\mathrm{m_{RD}}$
A	a_1	a_2
B	b_1	b_2
C	c_1	c_2
D	d_1	d_2
E	e_1	e_2
F	f_1	f_2
G	g_1	g_2

$a_1 + b_1 + c_1 + d_1 + e_1 + f_1 + g_1 = a_2 + b_2 + c_2 + d_2 + e_2 + f_2 + g_2 = 1$, $g_1 = g_2 = 0$。

3.2 Calculate the Normalisation Constant K

According to the formula:

$$K = \sum_{B \cap C \neq \varnothing} m_1(B) \cdot m_2(C) = 1 - \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C) \quad (1)$$

K is the summation of the joint mass functions of the hypotheses whose intersection is not empty, and reflects the degree of conflict of evidence (i.e. the higher the value of

K, the greater the degree of conflict of evidence). According to the hypothesis space there are 16 hypotheses for which the intersection is not empty:

Farmers assume "there are problems at the planting stage;" regulators assume "there are problems at the planting stage".

1. Farmers assume "there are problems in processing/acquisition;" regulators assume "there are problems in processing/acquisition."
2. Farmers assume "there are problems in the storage chain;" regulators assume "there are problems in the storage chain."
3. Farmers assume "there are problems in logistics;" regulators assume "there are problems in logistics."
4. Farmers assume "there are problems at the point of sale;" regulators assume "there are problems at the point of sale."
5. Farmers assume "there are problems at every stage;" regulators assume "there are problems at every stage."
6. Farmers assume "there are problems at the planting stage;" regulators assume "there are problem at every stage."
7. Farmers assume "there are problems at the processing/buying stage;" regulators assume "there are problems at every stage."
8. Farmers assume that "there are problems in the storage chain;" regulators assume that "there are problems in every chain."
9. Farmers assume "there are problems in the logistics chain;" regulators assume "there are problems in every chain."
10. Farmers assume "there are problem at the point of sale;" regulators assume "there are problem at every point."
11. Farmers assume "there are problems at every stage;" regulators assume "there are problems at the planting stage."
12. Farmers assume "there are problems at every stage;" regulators assume "there are problems at the processing/buying stage."
13. Farmers assume "there are problems at every stage;" regulators assume "there are problems in storage."
14. Farmers assume "there are problems at every stage;" regulators assume "there are problems in logistics."
15. Farmers assume "there are problems at every stage;" regulators assume "there are problems at the point of sale."

In summary:

$$\begin{aligned}K = {} & m_F(A)\cdot m_{RD}(A) + m_F(B)\cdot m_{RD}(B) + m_F(C)\cdot m_{RD}(C) + m_F(D)\cdot m_{RD}(D) + \\ & m_F(E)\cdot m_{RD}(E) + m_F(F)\cdot m_{RD}(F) + m_F(A)\cdot m_{RD}(F) + m_F(B)\cdot m_{RD}(F) + \\ & m_F(C)\cdot m_{RD}(F) + m_F(D)\cdot m_{RD}(F) + m_F(E)\cdot m_{RD}(F) + m_F(F)\cdot m_{RD}(A) + \\ & m_F(F)\cdot m_{RD}(B) + m_F(F)\cdot m_{RD}(C) + m_F(F)\cdot m_{RD}(D) + m_F(F)\cdot m_{RD}(E)\end{aligned}$$

$$\begin{aligned}= {} & a_1a_2 + b_1b_2 + c_1c_2 + d_1d_2 + e_1e_2 + f_1f_2 + a_1f_2 + b_1f_2 + c_1f_2 + d_1f_2 + e_1f_2 + \\ & f_1a_2 + f_1b_2 + f_1c_2 + f_1d_2 + f_1e_2\end{aligned}$$

$$= k \tag{2}$$

3.3 Calculate Joint mass

The synthetic formula for the n mass functions is:

$$m_1 \oplus m_2 \oplus \dots \oplus m_n(A) = \frac{1}{K}\sum_{A_1 \cap A_2 \dots \cap A_n = A} m_1(A_1) m_2(A_2) \dots m_n(A_n) \tag{3}$$

Equation (3) shows that the probability of the hypothesis A holding is the probability that all sets intersecting two by two intersect as A. There are two mass functions assumed in this paper, so the formula is as follows:

$$m_1 \oplus m_2(A) = \frac{1}{K}\sum_{A_1 \cap A_2 = A} m_1(A_1) m_2(A_2) \tag{4}$$

Therefore

$$\begin{aligned} m_1 \oplus m_2(\{A\}) &= \frac{1}{K}\{[m_F(A) \cdot m_{RD}(A)] + [m_F(A) \cdot m_{RD}(F)] + [m_F(F) \cdot m_{RD}(A)]\} \\ &= \frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2) \end{aligned} \tag{5}$$

$$m_1 \oplus m_2(\{B\}) = \frac{1}{K}\{[m_F(B) \cdot m_{RD}(B)] + [m_F(B) \cdot m_{RD}(F)] + [m_F(F) \cdot m_{RD}(B)]\} = \frac{1}{k}(b_1 b_2 + b_1 f_2 + f_1 b_2) \tag{6}$$

$$\begin{aligned} m_1 \oplus m_2(\{C\}) &= \frac{1}{K}\{[m_F(C) \cdot m_{RD}(C)] + [m_F(C) \cdot m_{RD}(F)] + [m_F(F) \cdot m_{RD}(C)]\} \\ &= \frac{1}{k}(c_1 c_2 + c_1 f_2 + f_1 c_2) \end{aligned} \tag{7}$$

$$\begin{aligned} m_1 \oplus m_2(\{D\}) &= \frac{1}{K}\{[m_F(D) \cdot m_{RD}(D)] + [m_F(D) \cdot m_{RD}(F)] + [m_F(F) \cdot m_{RD}(D)]\} \\ &= \frac{1}{k}(d_1 d_2 + d_1 f_2 + f_1 d_2) \end{aligned} \tag{8}$$

$$\begin{aligned} m_1 \oplus m_2(\{E\}) &= \frac{1}{K}\{[m_F(E) \cdot m_{RD}(E)] + [m_F(E) \cdot m_{RD}(F)] + [m_F(F) \cdot m_{RD}(E)]\} \\ &= \frac{1}{k}(e_1 e_2 + e_1 f_2 + f_1 e_2) \end{aligned} \tag{9}$$

$$m_1 \oplus m_2(\{F\}) = \frac{1}{K} m_F(F) \cdot m_{RD}(F) = \frac{1}{k} f_1 f_2 \tag{10}$$

Calculating the confidence interval

The confidence function indicates the degree of trust in hypothesis A being true, and the likelihood function indicates the degree of trust in hypothesis A being false. The relationship between the two functions is shown in Figure 2.

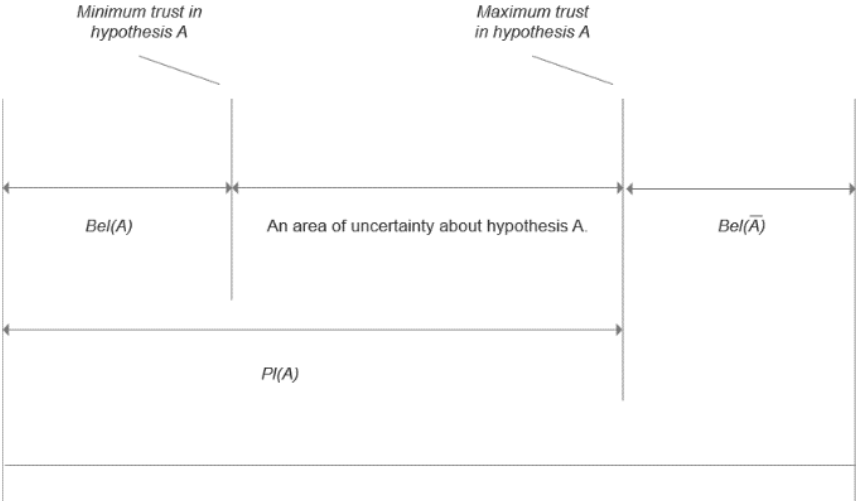

Fig. 2. Visual illustration of the relationship between the confidence function and the likelihood function

The formula for the trust function is:

$$Bel(i) = \sum_{j \subseteq i} m(j) \quad (i, j \subseteq \{A, B, C, D, E, F, G\}) \tag{11}$$

The formula for the likelihood function is:

$$Pl(i) = \sum_{j \cap i \neq \varnothing} m(j) \quad (i, j \subseteq \{A, B, C, D, E, F, G\}) \tag{12}$$

The trust function for A is:

$$Bel(A) = \frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2) \tag{13}$$

The likelihood function of A is:

$$Pl(A) = \frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2 + f_1 f_2) \tag{14}$$

The confidence interval for A is $[\frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2), \frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2 + f_1 f_2)$; that is, hypothesis A is believed to be true with the confidence of $\frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2)$, and hypothesis A is not believed to be true with a confidence of $1 - \frac{1}{k}(a_1 a_2 + a_1 f_2 + f_1 a_2 + f_1 f_2)$.

4 Numerical Analysis

Equation (13) shows that when a_1 and a_2 are the two largest numbers, the highest level of confidence is placed in hypothesis A, and the probability that there is a problem in the growing process is the highest. Similarly when b_1 and b_2 are the two largest numbers, the highest level of confidence is placed in hypothesis B (i.e. the probability that there is a problem in the processing/acquisition process is the highest). However, when the two mass function values for the same hypothesis are not the maximum, it is not possible to determine which hypothesis is trusted the most based on the above, and thus, the values need to be substituted for analysis, as detailed in Table 2.

Table 2. Basic allocation probabilities BPA with values

Suppose	m_F	m_{RD}
A	0.15	0.5
B	0.3	0.1
C	0.15	0.1
D	0.3	0.1
E	0	0.15
F	0.1	0.05
G	0	0

1. $K = a_1a_2 + b_1b_2 + c_1c_2 + d_1d_2 + e_1e_2 + f_1f_2 + a_1f_2 + b_1f_2 + c_1f_2 + d_1f_2 + e_1f_2 + f_1a_2 + f_1b_2 + f_1c_2 + f_1d_2 + f_1e_2$

$$= 0.15 \times 0.5 + 0.3 \times 0.1 + 0.15 \times 0.1 + 0.3 \times 0.1 + 0 \times 0.15 + 0.1 \times 0.05 + 0.15 \times 0.05 + 0.3 \times 0.05 + 0.15 \times 0.05 + 0.3 \times 0.05 + 0 \times 0.05 + 0.1 \times 0.5 + 0.1 \times 0.1 + 0.1 \times 0.1 + 0.1 \times 0.1 + 0.1 \times 0.15$$
$$= 0.295$$

2. $m_1 \oplus m_2(\{A\}) = \frac{1}{k}(a_1a_2 + a_1f_2 + f_1a_2) = \frac{1}{0.295} \times (0.075 + 0.0075 + 0.05) \approx 0.449$

$$m_1 \oplus m_2(\{B\}) = \frac{1}{k}(b_1b_2 + b_1f_2 + f_1b_2) = \frac{1}{0.295} \times (0.03 + 0.015 + 0.01) \approx 0.186$$
$$m_1 \oplus m_2(\{C\}) = \frac{1}{k}(c_1c_2 + c_1f_2 + f_1c_2) = \frac{1}{0.295} \times (0.015 + 0.0075 + 0.01) \approx 0.11$$
$$m_1 \oplus m_2(\{D\}) = \frac{1}{k}(d_1d_2 + d_1f_2 + f_1d_2) = \frac{1}{0.295} \times (0.03 + 0.015 + 0.01) \approx 0.186$$
$$m_1 \oplus m_2(\{E\}) = \frac{1}{k}(e_1e_2 + e_1f_2 + f_1e_2) = \frac{1}{0.295} \times 0 = 0$$
$$m_1 \oplus m_2(\{F\}) = \frac{1}{k}f_1f_2 = \frac{1}{0.295} \times 0.1 \times 0.05 \approx 0.017$$

3. $Bel(A) = \frac{1}{k}(a_1a_2 + a_1f_2 + f_1a_2) = \frac{1}{0.295} \times (0.075 + 0.0075 + 0.05) \approx 0.449$

$$Bel(B) = \frac{1}{k}(b_1b_2 + b_1f_2 + f_1b_2) = \frac{1}{0.295} \times (0.03 + 0.015 + 0.01) \approx 0.186$$
$$Bel(C) = \frac{1}{k}(c_1c_2 + c_1f_2 + f_1c_2) = \frac{1}{0.295} \times (0.015 + 0.0075 + 0.01) \approx 0.11$$

$$Bel(D) = \frac{}{k}(d_1d_2 + d_1f_2 + f_1d_2) = \frac{}{0.295} \times (0.03 + 0.015 + 0.01) \approx 0.186$$

$$Bel(E) = \frac{1}{k}(e_1e_2 + e_1f_2 + f_1e_2) = \frac{1}{0.295} \times 0 = 0$$

$$Bel(F) = \frac{1}{k}f_1f_2 = \frac{1}{0.295} \times 0.1 \times 0.05 \approx 0.017$$

In summary, it follows that $Bel(A) > Bel(D) = Bel(B) > Bel(C)$. Therefore, the highest level of confidence is placed on hypothesis A, which states that the probability of a problem in the growing segment is the highest. It is evident from this analysis that the method using D-S evidence theory is feasible and effective in identifying the links where quality problems with produce occur. It is clear from the practical implications that the basic allocation probability table should be larger in scale; however, the calculation formula of this method has a regularity, and the basic allocation probability table is calculated through the same formula regardless of the scale. Therefore, the calculation efficiency will not change much, and the amount of calculation can be reduced through this method for traceability of agricultural products.

4.1 Accuracy Analysis

In the above solution method, the trust function can eventually be calculated accurately. The trust function affects the degree of trust in the hypothesis, so the size of the trust function value obtained from the solution depends on the size of the mass value determined by the evidence source. However there are many factors influencing the mass value, for example, the evidence source farmer does not want the planting session to be problematic, and thus deliberately determines the mass value of hypothesis A to be very small. Therefore, the following methods can be used to force the source of evidence to determine a more credible mass value.

From equation (11), we know that Bel(i) satisfies $Bel(i) = \sum_{j \subseteq i} m(j)$, $(i, j \subseteq \{A, B, C, D, E, F, G\})$, and that Bel(i) is an increasing function with respect to m(j). Therefore that a large value can be estimated for m(j) when the source of evidence does not disclose a sufficiently correct and credible m(j). For example, making $m(j) = k, (0 \leq k \leq 1)$, where k converges infinitely to 1, results in a larger Bel(i) being found. Publishing such a large value of Bel(i) across the entire market for the distribution of agricultural products (i.e. assuming that i is essentially valid, and assuming that the segment corresponding to i is faulty) would affect the profits of firms in this segment, thus forcing the source of evidence for this segment to disclose a sufficiently credible m(j) to reduce its Bel(i) value. For example, $m_F(A)=0.15$ example shown in Table 2, if the farmer a source of evidence does not correctly disclose $m_F(A)$, it can make $m_F(A)=0.9$. The increase in A is accompanied by a corresponding decrease in the value of the mass function for the other assumptions and the data recreated accordingly are shown in Table 3.

Table 3. Basic allocation probabilities for reallocation BPA

Suppose	m_F	m_{RD}
A	0.9	0.5
B	0.05	0.1
C	0.0075	0.1
D	0.04	0.1
E	0	0.15
F	0.0025	0.05
G	0	0

Final Calculated $Bel(A) = \frac{1}{k}(a_1a_2 + a_1f_2 + f_1a_2) = \frac{1}{0.512125} \times (0.45 + 0.045 + 0.00125) \approx 0.969$

The increase in Bel(A) from 0.449 to 0.969 is significant, and the fact that Bel(A) is close to 1 makes it possible to determine that there is a problem in the growing chain. Specifically the downstream chain will refuse to accept the produce from this farmer, thus putting pressure on the farmer to disclose a credible mass function value in order to reduce Bel(A).

If the value of a hypothetical mass function in the distribution of agricultural products is continuously increased, what will be the result? It is possible to obtain $0 \leq m(j) \leq 1$, $Bel(i) = \sum_{j \subseteq i} m(j) \quad (i, j \subseteq \{A, B, C, D, E, F, G\})$, so if all other players in the agricultural distribution chain take the minimum value of m(j), then m(j)=0. At this point the resulting Bel(i)=1. This demonstrates that accurate traceability is achieved and that it is possible to be completely certain that there is a problem at some point. Furthermore, it is possible to approach this goal through the efforts of all parties in the market environment and the role of the network, thus increasing the accuracy of traceability of agricultural products.

4.2 Comparative Performance Analysis

In this paper, a comparative performance analysis will be conducted with one of the latest cluster-based agricultural traceability models based on blockchain technology to determine the efficient performance of this paper's approach.

Test Environment.

In this paper, an agricultural traceability network was established based on Fabric v1.4.1. The virtual machine was configured with a CentOS7 64-bit Linux system, 1024 MB of RAM, SCSI type LSI Logic disks, and more than 200,000 traceability experimental data from the Kiwi enterprise in Meixian County, Shaanxi. The testing tool was Hyperledger Caliper, and the testing parameters are shown in Table 4.

Table 4. System test parameters

Settings	Value	Description
Number of channels	6	Planting chain, processing chain, storage chain, logistics chain, sales chain, core chain
Number of nodes	26	4 nodes per chain, core chain includes authentication nodes and supervisory nodes

Block generation time	1s	Time interval for generating blocks
Block maximum bytes	50MB	The maximum size of transaction data that can be contained in a block
Maximum number of transactions in a block	10	Maximum number of transactions that can be accepted in a block
Number of transactions	100	Number of transactions per test round

Throughput Analysis.

We tested the throughput of both blockchain + D-S evidence theory and clustered produce traceability models when data were written is shown in Figure 3, and the throughput of both traceability methods when queried is shown in Figure 4.

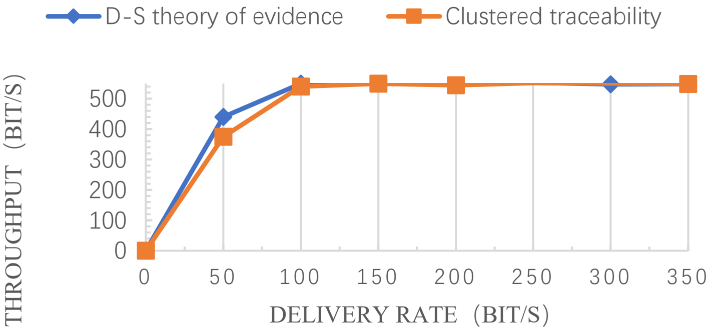

Fig. 3. Data write throughput

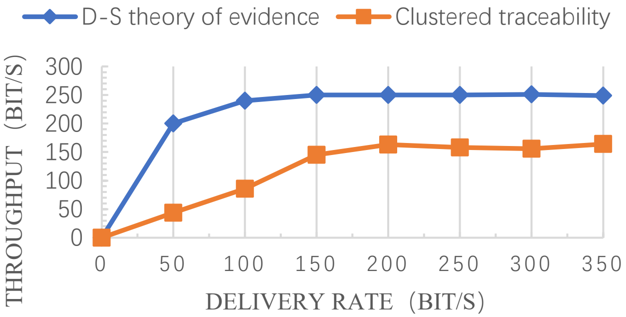

Fig. 4. Data query throughput

From Fig. 3, it can be seen that when the sending rate is from 50 to 100bit/s, the throughput of writing in both methods increases, and tends to level off after 100bit/s. The highest throughput of writing in the blockchain+D-S evidence theory method is 553bit/s, and the highest throughput of writing in the clustered traceability method is 552bit/s. The throughput of writing in both methods is almost the same; from Fig. 4, the query throughput of both traceability methods increases when the sending rate is

50-100bit/s, and tends to level off after 100bit/s, with the highest query throughput of 251bit/s and 164bit/s for the two methods respectively.

Both traceability methods require analysis and tracing of data from all chains when writing data, and the volume of data is extremely large, resulting in high load on a single machine and data writing failures and leading to low overall throughput. However, both traceability methods have the same writing method, resulting in the same throughput in data writing. The throughput of the blockchain + D-S evidence theory method is higher than that of the clustered traceability because the blockchain + D-S evidence theory method can trace back to a certain chain before making a single-chain data query, which requires only one smart contract call and greatly reduces the amount of data analysis. Conversely, whereas the clustered traceability method requires cross-chain queries and therefore needs to perform multiple smart contract invocations.

Query Performance Analysis.

In this paper, the query test was implemented through an external interface and 100 times per round. The time delay of both blockchain + D-S evidence theory and clustered traceability approaches when querying data is shown in Figure 5. As shown in Figure 5, the average time for querying using the blockchain + D-S evidence theory approach is 368ms, and the average time for querying using clustered traceability is 1375.6ms.

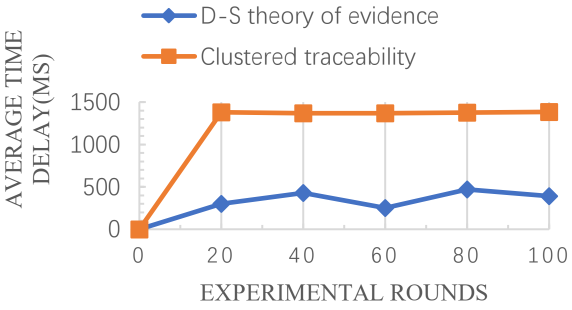

Fig. 5. Average data query time

The main reason for such a large difference in the average latency between the two traceability methods is that when using the blockchain + D-S evidence theory approach for querying, only one blockchain needs to be queried and can be queried directly, resulting in low latency for data queries. In the clustered traceability approach, data from multiple blockchains needs to be queried and data queries need to cross chains, resulting in higher latency.

Storage Performance Analysis.

The storage performance of the D-S evidence theory approach can divide block chains into multiple chains according to the origin regions of agricultural products, thereby enhancing the traceability network storage capacity of agricultural products.

For example, by dividing the region of origin according to latitude and longitude, China's kiwifruit production area can be divided into multiple production areas, with each region of origin maintaining a separate chain and each chain independently storing traceability data within the region of origin. The resulting analysis yields the storage performance of traceability based on the D-S evidence theory approach and clustered traceability, as shown in Figure 6, where the storage capacity consumption of the two approaches is almost the same under the same number of blocks.

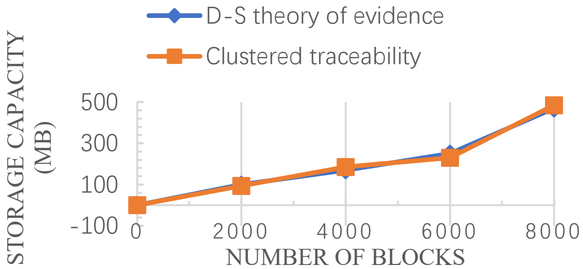

Fig. 6.Storage performance comparison chart

5 Conclusions

The traceability approach based on blockchain + D-S evidence theory proposed in this paper provides reference for the traceability of agricultural products, leading to the following conclusions:

1. By using the technology of blockchain + D-S evidence theory to construct a new type of traceability method, the data of agricultural products collected through IoT + blockchain technology is directly processed on the chain, and then the upstream and downstream enterprises of the industry chain up to consumers are connected by means of the public chain to achieve the whole industry chain tracking of product information. When a problem arises with a product, the blockchain + D-S evidence argument method is used to synthesise sensor data, determine the hypothesis space based on each link of the product from cultivation to sale, and establish a model for distribution to determine the level of trust for problems in each link. The application of this method in the production environment of agricultural products can achieve traceability at a lower cost, which is conducive to improving the quality of agricultural products circulating in the market. Because of the blockchain's unforgeable and decentralised characteristics, it can naturally guarantee the authenticity rate of agricultural products and can effectively reduce the management and supervision costs of market supervision departments for agricultural products.
2. The traceability method based on blockchain + D-S evidence theory was tested and analysed for accuracy and efficiency. The results show that in terms of accuracy, the traceability results of the method may be less accurate at the initial stage of application. However after reaching a certain application time and through the efforts of all

parties in the market environment and the role of the network, the probability value can be infinitely converged to a fixed value and accurate traceability can eventually be achieved. In terms of performance, the write throughput of the blockchain+D-S evidence theory approach and the clustered traceability approach are almost the same; the highest query throughput of the two approaches is 251bit/s and 164bit/s respectively, with the query throughput of the blockchain+D-S evidence theory approach being higher. The time delay of the blockchain+D-S evidence theory approach is 368ms, which is much lower than the 1375.6ms of the query using the clustered traceability approach. The storage performance of the two methods is almost the same. The method can effectively solve the problems of complicated operation and low efficiency of the traceability approach. The research results have certain significance for the expansion of blockchain traceability applications and the collaborative development of agricultural product traceability.

References

1. Jin Y, Wang K, Iop. Research on Traceability Strategy of Agricultural Product Quality and Safety[C]. 4th International Conference on Advances in Energy Resources and Environment Engineering (ICAESEE), 2018.
2. Wang Q, Liu H J, Bai Y, et al. Research progress on mutton origin tracing and authenticity[J]. Food Chemistry, 2022, 373.
3. Mamede R, Ricardo F, Abreu M H, et al. Spatial variability of elemental fingerprints of sea lettuce (Ulva spp.) and its potential use to trace geographic origin[J]. Algal Research-Biomass Biofuels and Bioproducts, 2021, 59.
4. Vanderschueren R, Montalvo D, De Ketelaere B, et al. The elemental composition of chocolates is related to cacao content and origin: A multi-element fingerprinting analysis of single origin chocolates[J]. Journal of Food Composition and Analysis, 2019, 83.
5. Jin H, Liu J. An application study of EAN.UCC bar code identification system in the quality of agricultural products traceability[J]. Second International Conference on Mechanic Automation Control Engineering, 2011: 4719-4723.
6. Ma X L, Ashnest J R. Tracing the Origins of Agricultural Products with Barcoded Microbial Spores[J]. Molecular Plant, 2020, 13(8): 1102-1102.
7. Cao Y, Jia F, Manogaran G. Efficient Traceability Systems of Steel Products Using Blockchain-Based Industrial Internet of Things[J]. IEEE Transactions on Industrial Informatics, 2020, 16: 6004-6012.
8. Wang S, Li D, Zhang Y, et al. Smart Contract-Based Product Traceability System in the Supply Chain Scenario[J]. IEEE Access, 2019, 7: 115122-115133.
9. Zheng M, Zhang S, Zhang Y, et al. Construct Food Safety Traceability System for People's Health Under the Internet of Things and Big Data[J]. IEEE Access, 2021, 9: 70571-70583.
10. Tang L-L, Lu M-T, Hung W C. RFID Application in Waste Monitoring Management of Electronic Products[J]. Int. J. Electron. Bus. Manag., 2008, 6: 161-173.
11. Chen S, Yi J, Zhu X, et al. RFID-based integrated method for electromechanical products disassembly decision-making[J]. International Journal of Computer Integrated Manufacturing, 2017, 30: 229 - 254.
12. Chen T B, Ding K F, Hao S K, et al. Batch-based traceability for pork: A mobile solution with 2D barcode technology[J]. Food Control, 2020, 107.

13. Popovic T, Krco S, Maras V, et al. A novel solution for counterfeit prevention in the wine industry based on IoT, smart tags, and crowd-sourced information[J]. Internet of Things, 2021, 14.
14. Cozzolino D. An overview of the use of infrared spectroscopy and chemometrics in authenticity and traceability of cereals[J]. Food Research International, 2014, 60: 262-265.
15. Guo X H, Cai R, Wang S S, et al. Non-destructive geographical traceability of sea cucumber (Apostichopus japonicus) using near infrared spectroscopy combined with chemometric methods[J]. Royal Society Open Science, 2018, 5(1).
16. Zhao S S, Zhao Y. Application and Preparation Progress of Stable Isotope Reference Materials in Traceability of Agricultural Products[J]. Critical Reviews in Analytical Chemistry, 2021, 51(8): 742-753.
17. Zhao Y, Zhang B, Chen G, et al. Recent developments in application of stable isotope analysis on agro-product authenticity and traceability[J]. Food Chemistry, 2014, 145: 300-305.
18. Chen M M, Fu L X, Li D W, et al. Mineral element fingerprints verified the geographical origin of years and quantities of rice[J]. Journal of Food Composition and Analysis, 2022, 114.
19. Shi Y, Liu M, Sun A, et al. A data fusion method of electronic nose and hyperspectral to identify the origin of rice[J]. Sensors and Actuators a-Physical, 2021, 332.
20. Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System[C], 2008: 1-9.
21. Anagnostakis A G, Giannakeas N, Tsipouras M G, et al. IoT Micro-Blockchain Fundamentals[J]. Sensors, 2021, 21(8): 1-4.
22. Mingxiao D, Xiaofeng M, Zhe Z, et al. A review on consensus algorithm of blockchain[J]. IEEE International Conference on Systems, 2017: 2567-2572.
23. Zheng Z, Xie S, Dai H, et al. An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends[J]. IEEE International Congress on Big Data, 2017: 557-564.
24. Čekerevac Z, Cekerevac P. BLOCKCHAIN AND THE APPLICATION OF BLOCKCHAIN TECHNOLOGY[J]. MEST Journal, 2022: 23-29.
25. Vukolic M. The Quest for Scalable Blockchain Fabric: Proof-of-Work vs. BFT Replication[C]. Open Research Problems in Network Security, 2015: 112-125.
26. Luu L, Narayanan V, Zheng C, et al. A Secure Sharding Protocol For Open Blockchains[J]. Proceedings of the ACM SIGSAC Conference on Computer Communications Security, 2016: 17-30.
27. Li S, Yu M, Yang C-S, et al. PolyShard: Coded Sharding Achieves Linearly Scaling Efficiency and Security Simultaneously[J]. IEEE Transactions on Information Forensics Security, 2018, 16: 249-261.
28. Chelladurai U, Pandian S. A novel blockchain based electronic health record automation system for healthcare[J]. Journal of Ambient Intelligence Humanized Computing, 2021, 13: 693-703.
29. Haque A B, Rahman M. Blockchain Technology: Methodology, Application and Security Issues[J]. ArXiv, 2020, abs/2012.13366: 21-23.
30. Zheng Y, Xu Y Q, Qiu Z G. Blockchain Traceability Adoption in Agricultural Supply Chain Coordination: An Evolutionary Game Analysis[J]. Agriculture-Basel, 2023, 13(1).
31. Yan H-S, Yang J-W, Kim H-H. Application of Blockchain Technology in Agricultural Products' Traceability System[J]. Asia-pacific Journal of Convergent Research Interchange, 2021, 7(12): 55-66.