2017 3rd International Conference on Big Data Computing and Communications

Secured Data Storage Scheme based on Block Chain for Agricultural Products Tracking

Chao Xie
Beijing University of
Posts and Telecommunications
Beijing, China
Email:superceix@outlook.com

Yan Sun
Beijing University of
Posts and Telecommunications
Beijing, China
Email:sunyan@bupt.edu.cn

Hong Luo
Beijing University of
Posts and Telecommunications
Beijing, China
Email:luoh@bupt.edu.cn

Abstract—**The Internet of Things technology has been widely used in the quality tracking of agricultural products, however, the safety of storage for tracked data is still a serious challenge. Recently, with the expansion of blockchain technology applied in cross-industry field, the unchangeable features of its stored data provide us new vision about ensuring the storage safety for tracked data. Unfortunately, when the blockchain technology is directly applied in agricultural products tracking and data storage, it is difficult to automate storage and obtain the hash data stored in the blockchain in batches base on the identity. Addressing this issue, we propose a double-chain storage structure, and design a secured data storage scheme for tracking agricultural products based on blockchain. Specifically, the chained data structure is utilized to store the blockchain transaction hash, and together with the chain of the blockchain to form a double-chain storage, which ensures the data of agricultural products will not be maliciously tampered or destructed. Finally, in the practical application system, we verify the correctness and security of the proposed storage scheme.**

I. Introduction

From planting to consumers to buy, agricultural products such as rice, soybeans, etc., need to go through the production and processing, transportation, storage, grading sales and a series of processes, but it will produce serious food safety risks if any of these parts is forged. Fortunately, with the development of Internet of Things technology, we can automatically trace the entire process to ensure that no man-made food safety issues if we can achieve real-time tracking of this series of processes. However, we have to store the data after tracking the process, there will still be possibilities that someone tampers with or destroys the data in the data storage phase if we just use the traditional storage method to storage the data. It is difficult for us to fundamentally eliminate man-made problems.

The blockchain is essentially a distributed account database, which is composed of a chain of data blocks generated through cryptographic correlation. Each data block contains information which is valid for multiple network transactions. Once the data has been verified and added to the block chain, it will be permanently stored unless someone can control more than 51% of the nodes at the same time. The modification of the database on a single node is invalid, this property makes the data stability and reliability of the block chain excellent. The first blockchain was conceptualized by Satoshi Nakamoto in 2008 and is currently used primarily in digital currencies similar to Bitcoin. The distributed and non-tampering features of blockchain make it favored in many industries, especially in the financial sector. However, until now, only a small number of applications have been put into use in the realistic environment.

Therefore, in the face of the tracking data about agricultural products, we hope that we can use blockchain technology to prevent tampering to ensure data security. However, compared with other industries, the internet of things has a greater amount of data and storage pressure. In order to achieve the consensus of distributed nodes, the block generation speed and transaction processing capacity of blockchain are limited, so it is not possible to directly apply the blockchain technology to store a mass of sensor data. On the other hand, the blockchain system, which focuses on the transactions of users, we can store the data by writing it into transactions, but with the transaction rate limit of blockchain, there is a bottleneck in data throughput.

More importantly, relying on the blockchain itself, it is hard to quickly store or query all the hash bulk data in different blocks through the identity defined by us.

To solve the above problems, we design a secured data storage scheme based on blockchain for agricultural products tracking. Agricultural products bound with a iot sensor module, so that the sensor can acquire the data of the products and upload it to the sever real-time. The server uses a double-chain storage structure to automatically store the data in the blockchain, concurrently, the system can also efficiently query the data and provides it to the upper application. In this way, we can use blockchain to safely and effectively store agricultural tracking data, which ensures the better food safety.

The contribution of this paper includes: (1) We design the overall scheme of agricultural tracking data storage based on blockchain. (2) We propose a double-chain storage structure, which does not seriously affect the efficiency of read-write, to more effectively ensure the security of data storage.

The remainder of this paper is organized as follows. Section II presents related works, Section III provides an overview structure of the storage system, and Section IV describes the detail of storage and query scheme, subsequent We present experiments and analyses in Sections V, and conclude the paper in Section VI.

978-1-5386-3349-6/17 $31.00 © 2017 IEEE
DOI 10.1109/BIGCOM.2017.43

II. Related Work

At present, there are some attempts to trace the food, hoping to solve the problem of food safety through technical means ([1], [2]), Especially in recent years, with the rapid development of Internet of things technology it is possible to connect food producers, transportation and hospitality/retail companies [3]. As we can see, researchers have developed related applications to trace food safety [4], in this paper, the author use cloud computing and cloud storage technology to supply the application, but as described in [5]. Traditional security issues are still present in cloud computing environments, and even the traditional security mechanisms are no longer suitable for applications and data in the cloud. It is still a serious challenge to store data safely while we track our agricultural products.

S Nakamoto proposed the concept of Bitcoin in 2008 and achieved a decentralized point-to-point network through the proof-of-work mechanisms [6], moreover, the following system, blockchain, is widely concerned because of its decentralization, distributed storage and no-tampering features. Researchers began to try to use blockchain to store and protect data [9]. Recent research has provided a solution how can blockchain technology combines swarm robotics, they use the robots as nodes in blockchain, but they do not mention how they store the mass data in their solution.

III. system structure

A. system structure

We design a secured data storage system based on blockchain system and sensor network, with real-time monitoring of products through sensors, effectively storing and protecting data from maliciously tamper, we can track products to ensure food safety and reliability.

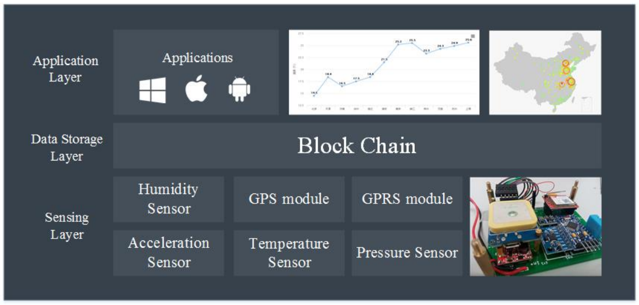


Fig. 1. Overview of system structure.

As shown in Fig.1, there are three layers in this system: Sensing Layer, Data Storage Layer and Application Layer. Sensing Layer contains a variety of iot sensor modules designed by us, which collect data in real time and upload them to the server. Then the server writes it into blockchain in Data Storage Layer. Application Layer is related with applications designed for users based on data system services.

The iot module of the Sensing Layer mainly includes temperature sensor, humidity sensor, acceleration sensor, pressure sensor, GPS module and GPRS module. We can get the temperature, humidity and location changes. According to the acceleration sensor, we can also speculate the transport status of products. The pressure sensor perceives changes in pressure. Once someone open the sealed package of the products, the pressure sensor can catch the exception and upload it to the server.

In the Data Storage Layer, data storage is based on the blockchain, we design a data storage model for iot sensor, and we will discuss it in the next section.

B. data storage model

We build the blockchain system with the open source blockchain framework, ethereum, and the data storage system is designed based on it. The system automatically encapsulates and analyses the data uploaded by the sensor, and writes it into blockchain. At the same time, for the transaction hash, we specially protect its security in storage, and use auxiliary database to store necessary data to improve the efficiency of data I/O.

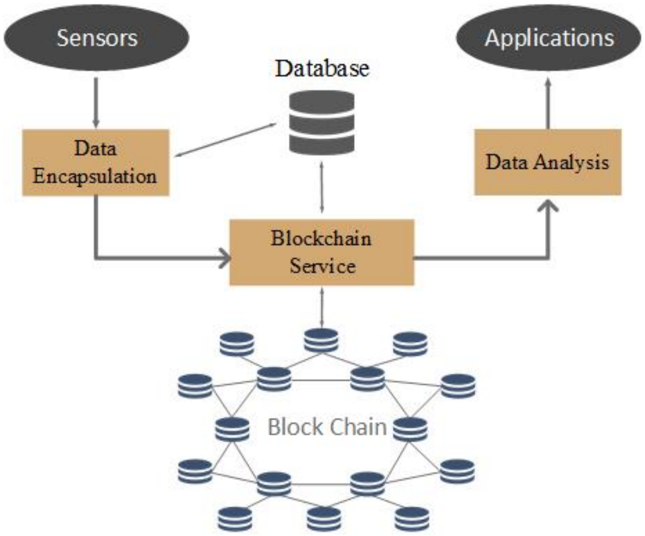


Fig. 2. Overview of data storage model.

As shown in Fig.2, based on the blockchain, the storage model includes Data Encapsulation, Blockchain Service and Data Analysis modules. The Data Encapsulation module receives data from sensors and reassembles the data. The Blockchain Service interact with the blockchain, which writes data into and query data from blockchain. The Data Analysis analyzes the data obtained from the block chain and responses to the Application Layer to build applications.

After receiving the trace data from the sensor, the system queries the corresponding account of blockchain according to the identity in the auxiliary database. Then reassemble the sensor data with the historical transaction hash. The Blockchain Service writes the data into blockchain. The process of querying data is the reverse process of storing data. When the Data Analysis module get a query request, it queries the transaction hash from auxiliary database, and then the Blockchain Service uses these hashes to query the transactions from the blockchain, the Data Analysis module analyzes the data from the transactions and response it.

In Sensing Layer, according to the existing design of the iot sensor, we can get agricultural data from the sensor's GPRS module like this:"860719023995818;32;251;102451;0.25; 1482799635;3957.57214;N;11620.99005;E", the data above is separated by ';' to represent the data obtained from the different sensors. In our pre-designed structure, these data represent the values of identity, humidity, temperature, pressure, acceleration, datetime, latitude, and longitude. After receiving the data, the server verifies the format of the data, the identity and the format of the data must be valid, then the data is formatted into the transaction data in json like this,

```
{”jsonrpc”:”2.0”,
”method”:”eth_sendTransaction”,
”params”:[{
”from”:
”0xb60e8dd61c5d32be8058bb8eb970870f07233155”,
”to”:
”0xd46e8dd67c5d32be8058bb8eb970870f07244567”,
”gas”:”0x76c0”,”gasPrice”:”0x9184e72a000”,
”value”:”0x9184e72a”,
”data”:”0x38363037313930323339393538313 83
b33323b3235313b3130323435313b302e32353b313
4383237393936333353b333935372e35373231343b4
e3b31313632302e39393030353b45”}],”id”:1}
```

The data is sent to the blockchain node in json format, and the node executes the transaction to write the data into blockchain. During this period, all the nodes in blockchain reach a consensus and synchronize the block data, and then the data will be never changed.

IV. DATA STORAGE SCHEME

A. double-chain data structure

The data of blockchain is stored in a series of blocks and transactions are the carrier of data storage. Namely, it can be said that these data blocks are actually composed of a series of transactions. Therefore, in this case, if we want to write a custom data of agricultural products into the block, we can only make the data an extension of the transaction, and the blockchain does not know what its specific this structure is. However, in most of the time, we may need to query the data based on the identity of our custom data. Obviously, if the data is only an extension of the transaction, we cannot solve our query requests through the blockchain.

At the very beginning, we intend to write the corresponding transaction hash each time we execute a transaction. When we need to query the data, we can query all the transaction hash of the identity in database first, and use these hash to get all the data we need from the blockchain. However, unlike the applications in the financial and other industries, the financial industry is more concerned about the security of accounts and tokens in the course of the transaction; they tend not to pay much attention to historical transactions a long time ago. Nevertheless, in the application of agricultural tracking, what we are concerned about is historical transactions. We are more often need to store and manage a large number of block chain transactions automatically. So if a transaction hash is maliciously deleted, we may not be able to detect such a situation, and it is difficult to retrieve this data record in the mass data. This makes it possible that even if all of the data we store in the blockchain is safe. We may not be able to find a piece of data because of the loss of the address.

Therefore, we design a double-chain storage structure based on the block-chain, by adding the parent transaction hash in to the data, the transactions form a data chain so that we can get all the data of the identity only use the last transaction hash, and the historical transaction data is only stored in the blockchain, will never be tampered.

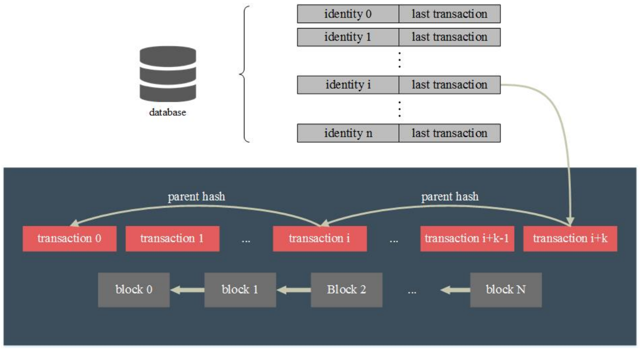


Fig. 3. Structure of the double-chain.

In the blockchain, each block contains the information of the previous block to form a chained structure. In Figure 3, there is no connection between the transactions originally. The last transaction hash for each identity is stored in the database, when there is a new data write request next time. The server queries the last transaction hash and writes it as a parent hash in this transaction, which forms the second chain. We can find its parent transaction in each transaction, and layer by layer to find all the transaction up. The define of the *i*-th data D_i:

$$D_i = \begin{cases} f(D_i, h_{i-1}) & i > 0 \\ D_i & i = 0 \end{cases}$$

where h_i means the hash of *i*-th transaction, and f is a function reassembles D_i and h_{i-1} into a new data. The procedure of building transaction chain and query data from blockchain by the transaction chain.

Algorithm 1 BuildTransactionChain //procedure of building transaction chain

```
Input: Id, // Sensor Id
       D // Sensor Data
if exists(ID) && isRightFormat(D) then
    var lastTrHash = GetLastTransactionHash(Id);
    if lasttrhash is not null then
        var input ={data:D,parentHash:lastTrHash};
    else
        var input ={data:D};
    end if
    var transactionHash = SendTransaction(input);
    UpdataLastTransactionHash(Id,transactionHash);
else
    return false;
end if
```

Algorithm 2 QuerySensorData

```
Input: Id, // Sensor Id
var trhash = GetLastTransactionHash(Id);
var dataList;
while trhash is not null do
    var result = GetTransactionByHash(trhash);
    if result is null then
        return dataList;
    else
        dataList.add(result.data);
        trhash = result.parentHash
    end if
end while
return dataList;
```

B. allocation of blockchain resources

On the issue of agricultural products tracking, now we are facing the Internet of things sensor and the amount of data is much larger compared to the blockchain. The production speed of block and transaction is not high enough to supply the agricultural tracking, it can be said that the blockchain is a limited resource. Therefore, when the blockchain is combined with the Internet of things data we have to improve the storage scheme to support the large amounts of data and not waste the network and storage resources at the same time.

In fact, the server will filter the data before writing it into the blockchain. In our tracking application, the sensors data may not change significantly in a short time, we cannot use the limited blockchain resources to store a mass of repetitive data. So that in the storage scheme, the uploaded data of each sensor is selected one to write into the blockchain per hour, at the same time, a non-normal data judgment strategy is added to compare the data with the historical data. The abnormal data is written into the blockchain, and if parameters of the normal data has no obvious changes with the recent data, it is selected one to write into blockchain per hour, and the others is written into the database. The function *AddNewData* is illustrated in Algorithm 3.

Algorithm 3 AddNewData //Add New Data Of Sensors

```
Input: Id, // Sensor Id
       D // Sensor Data
var T = GetTime(D); // T is the datetime of data
var LT = GetLastTime(Id); // LT is last data's datetime of Id
if LT-T >3600s then
    AddToBlock(Id,D); // add Data to blockchain
else
    var[] R = GetFromBLock(Id); // R is related data array of Id
    var E = CheckData(D,R); // E is whether D is abnormal data for R
    if E then
        AddToBlock(Id,D);
    else
        AddToDB(Id,D); // add Data to database
    end if
end if
```

V. EXPERIMENT AND RESULTS

This experiment uses go-ethereum 1.9 as a blockchain platform and build the system with jdk-8u101. We deployed the blockchain node in five machines, each machine possesses a 3.4 GHz core Intel processor with 8GB memory. And the nodes are deployed in Ubuntu 14.04 OS, one of the machines is deployed with the storage system and the other 4 machines is the node to create blocks.

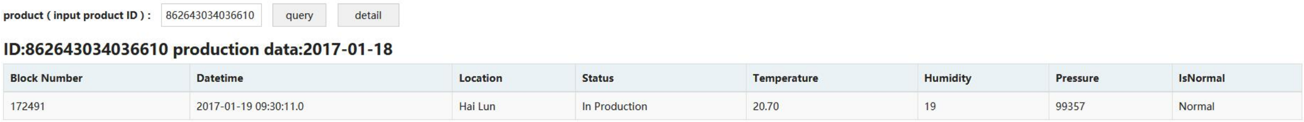
A. functional verification

We designed the demo application of agricultural products tracking system and the iot sensors are binding with agricultural products, the sensor is designed to upload data every ten minutes to the storage system. Users can use product id to query the sensor data in the blockchain.

User can query the data stored in blockchain about the identity when they enter it in the system, Figure 5 shows that there is an abnormal data in the third row which is marked in red to help users identify whether the product is normal.

We design both the chain-type mode and non-chain-type mode to store transaction hash. The system inserts a new record to store the transaction hash in non-chain-type mode, so if the record of the abnormal data is deleted, the program is difficult to find and retrieve the data even if the data is still in the blockchain. The chain-type mode solves this problem, there is only one record in the database, this record is the last transaction hash, if there is an abnormal data of this identity, what can be modified is only this record, but the abnormal data is in the transaction chain cannot be deleted.

The worst case is that the last transaction hash record is deleted, but the system can find this exception if there is no data can be queried, in this exception, users can choose not to buy this product. Any action that can tamper or destroy the data can be noticed by our system in chain-type mode. As shown in Fig.6, the system displays an abnormal about

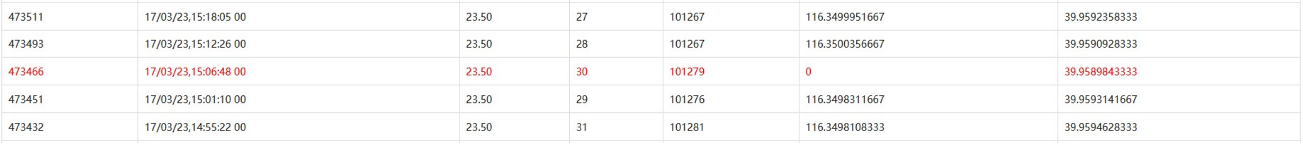


product (input product ID) : 862643034036610

ID:862643034036610 production data:2017-01-18

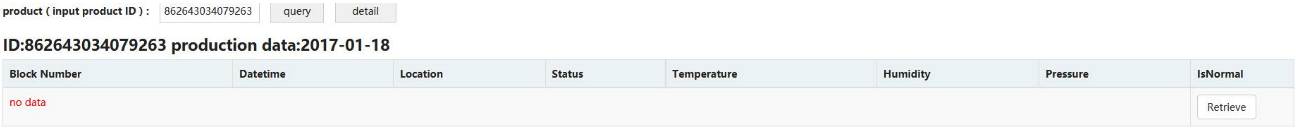
Block Number	Datetime	Location	Status	Temperature	Humidity	Pressure	IsNormal
172491	2017-01-19 09:30:11.0	Hai Lun	In Production	20.70	19	99357	Normal

Fig. 4. The data about the identity 862643034036610.



473511	17/03/23,15:18:05 00	23.50	27	101267	116.3499951667	39.9592358333
473493	17/03/23,15:12:26 00	23.50	28	101267	116.3500356667	39.9590928333
473466	17/03/23,15:06:48 00	23.50	30	101279	0	39.9589843333
473451	17/03/23,15:01:10 00	23.50	29	101276	116.3498311667	39.9593141667
473432	17/03/23,14:55:22 00	23.50	31	101281	116.3498108333	39.9594628333

Fig. 5. The detail sensor data about the identity 862643034036610,it contains abnormal data.

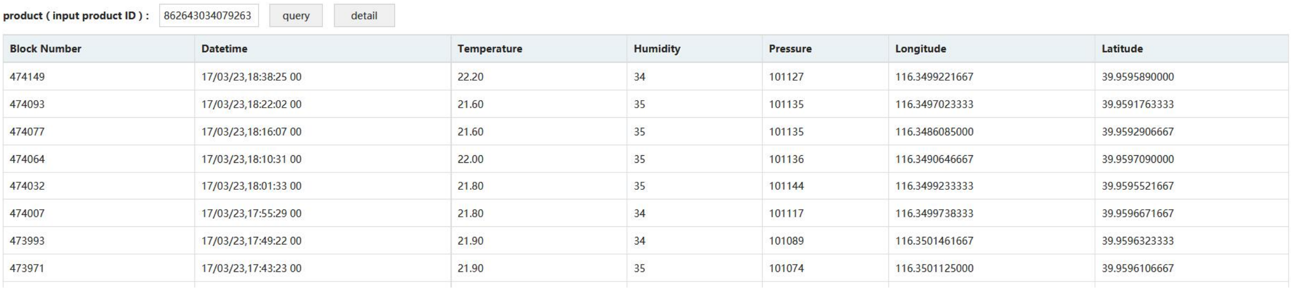


product (input product ID) : 862643034079263

ID:862643034079263 production data:2017-01-18

Block Number	Datetime	Location	Status	Temperature	Humidity	Pressure	IsNormal
no data							Retrieve

Fig. 6. The no data worning about the identity 862643034079263.



product (input product ID) : 862643034079263

Block Number	Datetime	Temperature	Humidity	Pressure	Longitude	Latitude
474149	17/03/23,18:38:25 00	22.20	34	101127	116.3499221667	39.9595890000
474093	17/03/23,18:22:02 00	21.60	35	101135	116.3497023333	39.9591763333
474077	17/03/23,18:16:07 00	21.60	35	101135	116.3486085000	39.9592906667
474064	17/03/23,18:10:31 00	22.00	35	101136	116.3490646667	39.9597090000
474032	17/03/23,18:01:33 00	21.80	35	101144	116.3499233333	39.9595521667
474007	17/03/23,17:55:29 00	21.80	34	101117	116.3499738333	39.9596671667
473993	17/03/23,17:49:22 00	21.90	34	101089	116.3501461667	39.9596323333
473971	17/03/23,17:43:23 00	21.90	35	101074	116.3501125000	39.9596106667

Fig. 7. The retrieved sensor data about the identity 862643034079263.

no data after we test to delete the last record of identity 862643034079263, the system provides a retrieve button for users, users can request to retrieve the data, the system traversal all of the blocks and find the last transaction to retrieve the data, the retrieved data is shown as Fig 7.

B. efficiency comparison

We tested I/O efficiency of the system in different environments, there is six groups of experiments:

a. one thread P, the double-chain structure, single thread to write data.

b. one thread B, the normal structure, single thread to write data.

c. five thread P, the double-chain structure, five thread to write data.

d. five thread B, the normal structure, five thread to write data.

e. one thread ten lines P, the double-chain structure, single thread to write data, and add ten data in one transaction.

f. one thread ten lines B, the normal structure, single thread to write data, and add ten data in one transaction.

We set six groups of experiments in the same network environment, as shown in Fig.8, comparing with the normal structure, the double-chain structure provides higher data security, at the same time, this scheme spends about 130% the time of non-chain mode in data writing. The iot sensor upload data every 10 minutes, for the normal data, considering about the performance of the blockchain, the system chooses one data to write into the blockchain per hour. In the experiment above, it can write about 200-300 sensor data per second, it means that the system can support 75,000-100,000 sensors writing its data into the blockchain. This amount is enough to achieve an agricultural tracking application.

And the data querying experiment is shown as Fig.9, we set 2 groups of experiment, the query speed of double-chain mod is twice slower than the non-chain mod, it can get about 300 data of an identity from the blockchain per second. It means users have to take a few seconds to wait for the data, it is acceptable.

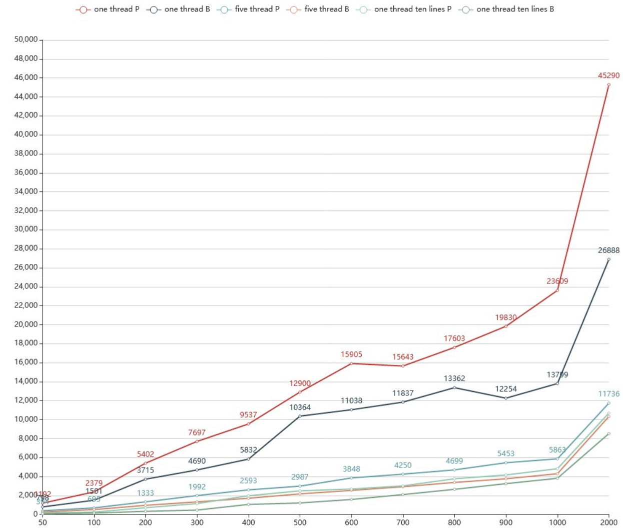


Fig. 8. Efficiency comparison of writing data.

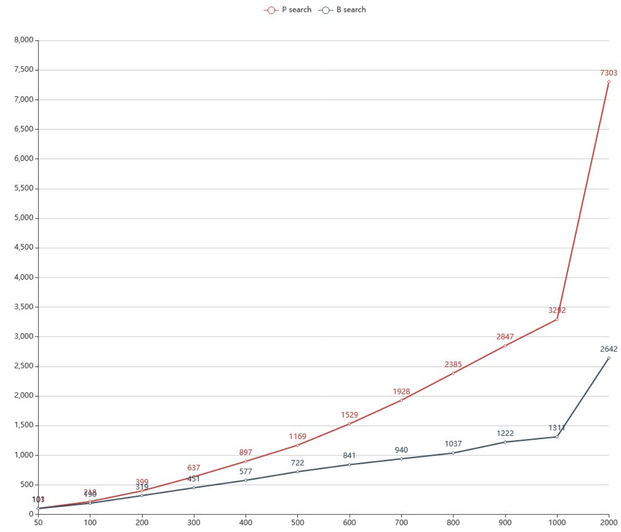


Fig. 9. Efficiency comparison of querying data.

VI. Conclusion

The tracking on agricultural products is mainly to monitor the environment of the products, to address issues of food safety. The tracking data should be securely stored, especially the abnormal data, if the system cannot protect it from being maliciously tampered, the data will be a lack of credibility, users will never believe it. Combine with the blockchain technology, the tracking data is written into the block unchangeably, the safety of the data is guaranteed, and the safety of the food is also guaranteed.

In this paper, we design a scheme to store the tracking data based on blockchain, in order to solve the problem of storing custom data in blockchain, we propose a double-chain storage structure and tested its efficiency. In the experiments, we find that the scheme is efficient enough to achieve the tracking application. We are working on optimizing the I/O efficiency for wider situations. Our future work will be extending and constructing the more efficient scheme.

References

[1] Li H, Xiao H, Qiu T, et al. Food safety warning research based on internet public opinion monitoring and tracing[C]// International Conference on Agro-Geoinformatics. 2013:481-484.

[2] Schuster E W, Albrigo L G, Ehsani R. Agricultural supply chains: track and trace for improved food safety.[J]. Acta Horticulturae, 2009, 824(824):113-120.

[3] Maksimovic M, Vujovic V, Omanovic-Miklicanin E. A Low Cost Internet of Things Solution for Traceability and Monitoring Food Safety During Transportation[C]//HAICTA. 2015: 583-593.

[4] Fu Y, Li F. Application of Internet of Things to the Monitoring System for Food Quality Safety[C]// Fourth International Conference on Digital Manufacturing and Automation. IEEE, 2013:296-298.

[5] Chen D, Zhao H. Data security and privacy protection issues in cloud computing[C]//Computer Science and Electronics Engineering (ICCSEE), 2012 International Conference on. IEEE, 2012, 1: 647-651.

[6] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system[J]. 2008.

[7] Crosby M, Pattanayak P, Verma S, et al. Blockchain technology: Beyond bitcoin[J]. Applied Innovation, 2016, 2: 6-10.

[8] Swan M. Blockchain thinking: The brain as a dac (decentralized autonomous organization)[C]//Texas Bitcoin Conference. 2015: 27-29.

[9] Zyskind G, Nathan O, Pentland A '. Decentralizing Privacy: Using Blockchain to Protect Personal Data[C]// Security and Privacy Workshops. IEEE, 2015:180-184.

[10] Ali M, Nelson J, Shea R, et al. Blockstack: A global naming and storage system secured by blockchains[C]//2016 USENIX Annual Technical Conference (USENIX ATC 16). USENIX Association, 2016: 181-194.

[11] Wood G. Ethereum: A secure decentralised generalised transaction ledger[J]. Ethereum Project Yellow Paper, 2014, 151.

[12] Conoscenti M, Vetr A, Martin J C D. Blockchain for the Internet of Things: a Systematic Literature Review[C]// International Symposium on Internet of Things: Systems, Management and Security. 2016.

[13] Ferrer E C. The blockchain: a new framework for robotic swarm systems[J]. 2016.

[14] Kokoriskogias E, Jovanovic P, Gailly N, et al. Enhancing Bitcoin Security and Performance with Strong Consistency via Collective Signing[J]. Applied Mathematical Modelling, 2016, 37(8):5723-5742.

[15] Yang K, Jia X. An Efficient and Secure Dynamic Auditing Protocol for Data Storage in Cloud Computing[J]. IEEE Transactions on Parallel & Distributed Systems, 2013, 24(9):1717-1726.

[16] Akyildiz I F, Su W, Sankarasubramaniam Y, et al. A survey on sensor networks[J]. IEEE Communications magazine, 2002, 40(8): 102-114.

[17] Botta A, De Donato W, Persico V, et al. Integration of Cloud computing and Internet of Things[J]. Future Generation Computer Systems, 2016, 56(C):684-700.