Journal of Computers Vol. 29 No. 6, 2018, pp. 254-263
doi:10.3966/199115992018122906025

An Efficient Storage and Query Scheme Based on Block Chain for Agricultural Products Tracking

Ya-Dong Liu*, Yan Sun, Hong Luo

Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 100876, China
{lydliu, Sunyan, luoh}@bupt.edu.cn

Received 9 January 2018; Revised 22 January 2018; Accepted 22 January 2018

Abstract. As agricultural products safety has a significant impact on peoples lives, it is of great significance to establish and improve the safety regulation system. Meanwhile, the rapid development of blockchain technology provides a new regulation method for the agricultural products safety issue. However, the exist blockchain query methods can not be apply to agricultural IoT directly. Besides, block generation speed of the exist methods are difficult to improve when memorizing massive IoT data. To solve the above problem, firstly, a blockchain query scheme based on hash database is proposed. The database stores the mapping among IoT data, transaction hash and block hash. Then, we put forward the anomalous data detection method based on voice prediction model. The memory depth and normal range of various products is set by domain expert to detecting abnormal data. To this end, we improve the ethereum workshop block generation mechanism and presented an adaptive adjustment strategy for mining difficulty based on data variation. The experi-mental results show that the proposed approach can outperforms the existing methods.

Keywords: agricultural products safety, anomalous data detection, blockchain, query scheme

1 Introduction

The adoption of future Internet of Things (IoT) technology provides excellent benefits for agricultural products movement process from the countryside to the dining table. Thus, the quality of agricultural products is guaranteed. For these products from production to transportation and sales, the IoT equipments such as sensors would generate mass of data in each production process. These generated data are often used to identify agricultural safety levels in application. Twodimensional code security traceability technology is still using the traditional storage methods to store IoT data. In this way, the tamper cost is low. Therefore, the credibility of the evidence which is used to identify the safety level of agricultural products would decreases.

After the generation of Bitcoin in 2008, more and more researchers pay attention to block-chain. It is a secure, decentralized distributed storage technology. A blockchain is connected by a series of blocks in the order of timestamps. In this pattern, each block contains multiple records. When the block is produced, these records are verified by the entire blockchain system before they are recorded. Once the blocks are generated, the stored data is tamper-resistant.

Considering the particularity of agricultural data, tampering and traceability is the main requirement for users. In this case, we introduce blockchain technology to store agricultural data for preventing tampering and ensuring data security. When user queries the product's provenance data, the matching method based on content from a large number of blocks is inefficient. Meanwhile, in the entire process of logistics, whether a reader or sensors would generate mass of data. If these data are stored as additional data in the transaction, the block generation rate would have a great impact on the effecting transaction rate.

* Corresponding Author

Based on discussed above, in order to improve query speed of provenance data, we proposed a storage method based on hash database. This method establishes visible mapping relationship among data id, blocks and transactions to accelerate data query. In addition, blockchain throughput is low per unit time. To solve the problem, we design a Anomalous Data Detection Method Based on Voice Prediction Model. The memory depth and normal range of various products is set by domain expert to detect abnormal data. If the data is abnormal, then, it is written to the blockchain; otherwise, the normal data would not be written to the distributed storage system. This greatly reduces the blockchain storage overload. However, in this case, the stored data is still large when our provenance platform having plenty of farm products. Hence, we built a provenance platform for agricultural products based on alliance chain. Also, we improve the ethereum workshop block generation mechanism and present an adaptive adjustment strategy for mining difficulty based on data variation. With the increasing data variation, the mining difficulty and block generation rate increases, and vice versa.

The remainder of this paper is organized as follows. Section II surveys the related works from the literatures. Section III overviews our system. In Section IV, we propose data storage and query method and present in detail block optimization scheme. Experiment and simulation results are described in Section V while conclusions are drawn in Section VI.

2 Related Work

Satoshi Nakamoto first proposed the Bitcoin in [1]. The Bitcoin is an electronic cash system which is based on blockchain. Blockchain technology is a distributed database that maintains a continuously-growing list of ordered records called blocks. The emergence of blockchain technology has attracted the attention of a large number of researchers. They have made an important contribution to the development of blockchain.

Zhu et al. [2] believe that the transaction can not be verified is one of the drawbacks in the blockchain. Therefore, they suggested an accurate transaction authentication mechanism which called Interactive Incontestable Signature (IIS). Since it can confirm the transaction between the dealer and the owner, the scheme has unforgeability and non-repudiation. Using the IIS can speed up transaction confirmation and provide the viability for business applications. Taylor [3] described the algorithms and hardware for Bitcoin systems. In particular, it described the use of computer hardware from the view of mining. It suggested that customized hardware may be an alternative to generalized hardware facilities. Huckle et al. [4] explores how the Internet of Things and blockchain technology can benefit from shared economic applications. The focus of this paper is to learn how to use blockchain to create decentralized, shared economic applications which allow people to monetize, safely, and create more wealth for their stuff. It describes a vision by using the IoT and blockchain to build a more convenient life. However, the problem of massive redundancy which caused by blockchain technology has not been solved. In the context of Eyal and Sirer's" selfish-mine", Gōbel et al. [5] studied the effect of communication delay in the blockchain. They use a simplified Markov model to track the block-chain state and use the discrete-event simulation to study the behavior of a network of Bitcoin miners. It provides a theoretical model for detecting communication delays in the blockchain. Tian [6] proposed an agricultural traceability system based on RFID and blackchain. However, it didn't give the system design and implementation program. It just describes the techniques and specifications which may be used in the traceability system.

With the development of blockchain technology, the Ethereum proposed the smart contract system based on blockchain. Smart contract is a program which is automatically executed on the blockchain. The basic idea of them is to require data from the outside of the blockchain. Reliable data feed is a key link in the smart contract system [7]. It established a reliable data feed system, called Town Crier (TC).TC has established a bridge between the smart contract and the existing web site. It combines the front end of the blockchain with a trusted hardware backend to collect reliable data from HTTPS-based websites and services to smart contracts. They deploy the data into the Ethereum smart contract system through Intel's Software Guard Extensions (SGX). However, due to the impact of transaction confirmation speed in the Ethereum, it is difficult to apply this technology on a large scale.

3 System Overview

3.1 System Framework

Fig. 1 shows the framework of efficient query system based on blockchain and IoT technology. This system can store large amounts of data securely. Moreover, it provides the safety of agricultural products by using blockchain characteristics which is not changeable.

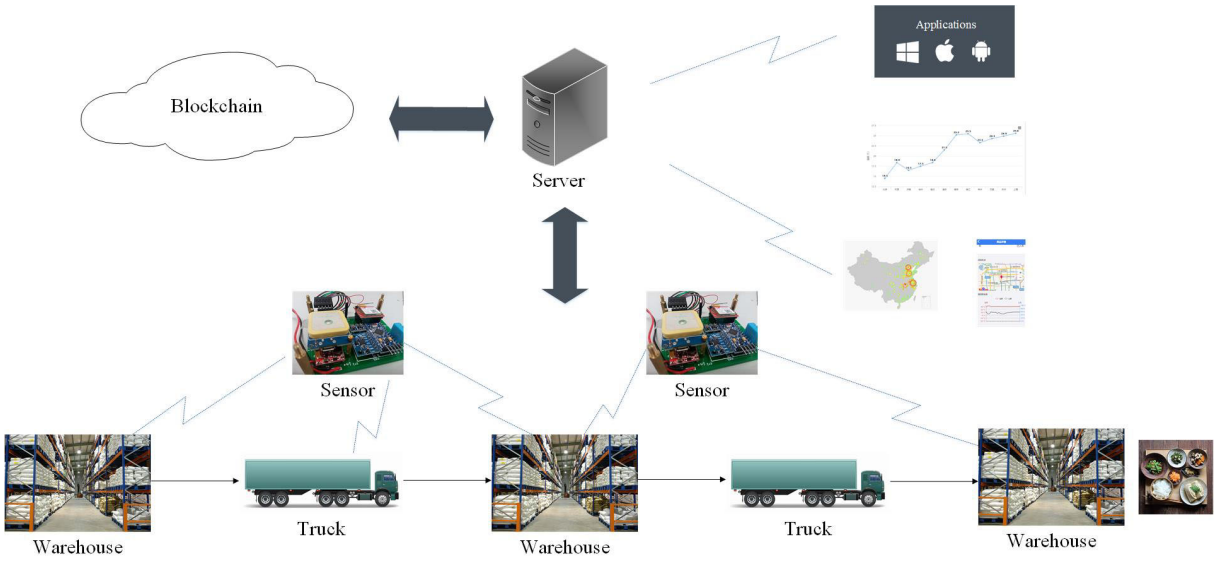

Fig. 1. A graph for the DSC node operating mode problem

As depicted in Fig. 1, agricultural products may be stored and transported multiple times from source to destination. We deploy sensors in food packaging, warehouses and transport vehicles to perceive related food quality data. The sensing module mainly includes temperature sensor, humidity sensor, acceleration sensor, pressure sensor, GPS, gprs module. Sensors can perceive the surroundings, temperature, humidity, and location. Besides, the acceleration sensor is used to perceive the goods transport status. Pressure sensor is used to perceive the pressure. Once the product seal is opened, the pressure would change. Thus, we can obtain the abnormal value. The data generated from sensing module would be uploaded to server by gprs. Then, after receiving the uploaded data, the server parses it and writes it to the blockchain. Consequently, we can build Safe Trace System and different applications based on the stored data.

3.2 Data Storage Model

In this paper, the open source framework Ethereum is used to develop data storage. To solve the search problem of agricultural IoT data, we design a data storage mechanism. The data storage module is responsible for parsing uploaded IoT data and storing them in the blockchain as additional data for the transaction. For the hash code generated in transaction, it is regarded as an unique certificate to read the data from blockchain. We store hash code with respect to blocks into the secondary database to improve search efficiency. Fig. 2 shows data storage model of the system. It consists of a sensor module, an RFID reader, an outlier detection module, a secondary database, and a blockchain that are described hereafter. The sensor module uploads the sensor-aware data to the server. Then, the outlier detection module performs outlier detection and stores it in the auxiliary database and the blockchain. The RFID reader is responsible for storing the logistics information directly into the auxiliary database and the blockchain.

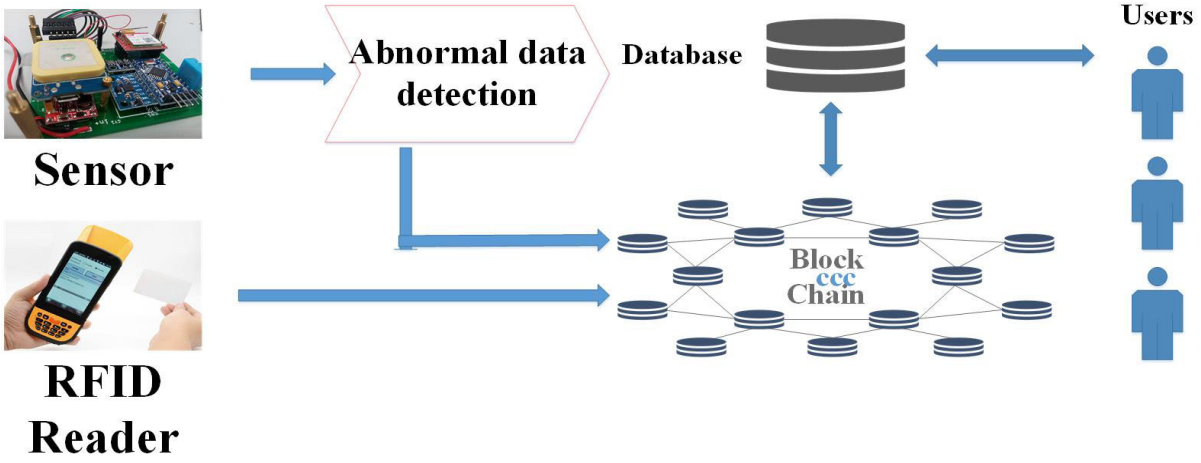

Fig. 2. Data storage model

The uploaded IoT data include humidity, temperature, barometric pressure, acceleration, position information (longitude, latitude) and timestamp. Before these data are written permanently to the blockchain, the outlier detection module compares the historical data and writes the result to the blockchain. RFID reader would upload agricultural logistics information including the library and storage information, which will be written directly to the blockchain.

The data is written into blockchain in the form of additional data. Each transaction would generate a unique hash code to query it. In order to query provenance easily, these hash codes would be stored in the auxiliary database along with the data.

4 Data Storage Scheme And Blockchain Optimization

4.1 Data Storage Scheme

Due to the particularity of the block-chain storage form, the sensor data and the logistics information data can only stored in the form of transactions. In this case, the only way to fetch data is querying by transaction hash. Without transaction hash, the query procedure should traverse the content according to the data content, which is very slow. Thus an external database can help store the block hash and each transaction's hash, so users can fetch data quickly.

Study on storage structure based on block-chain, the data of the block chain is logically stored in the hash of the block, each block contains several transactions. Each block is logically linked to form a block chain as Fig. 3 shown. Each block header has a block hash and a block number that identifies the entire block, according the block hash or block number can fetch the entire information of the block.

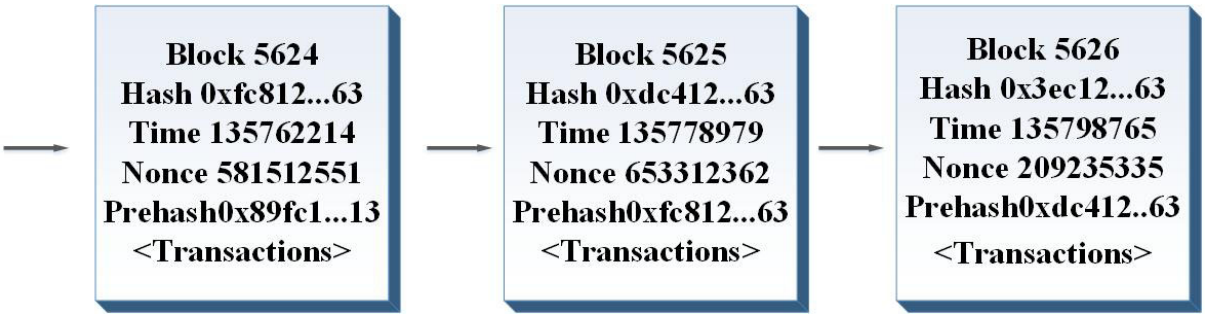

Fig. 3. Blockchain logical structure

Based on the structural characteristics of block-chain, after the transactions with data write into block-chain successfully, the block hash is saved together with transaction hashes and data identity in external database, as Fig. 4 shown. While querying data stored in block-chain, select transaction hash based on data identity in external database then use this transaction hash to fetch data from block-chain. When the

transaction hashes are lost caused by human factors in external database, according to block hash the block information can be fetched from block-chain and compared with items in external database which has the same block hash then the lost data can be found quickly.

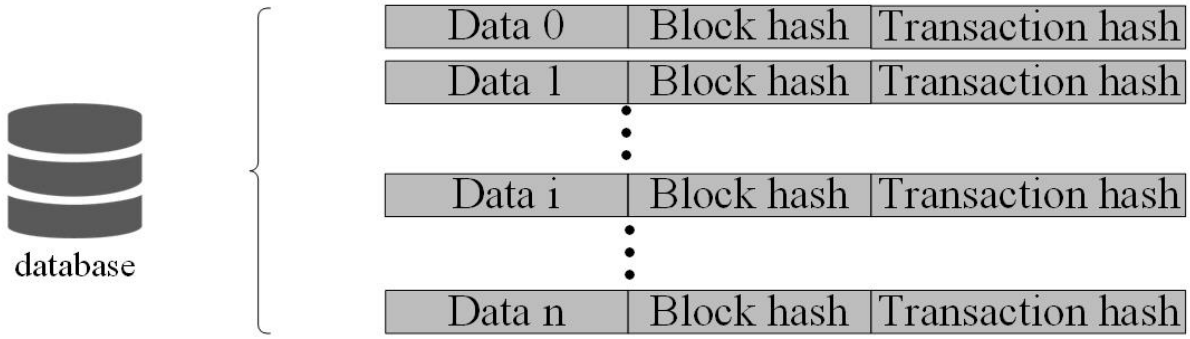

Fig. 4. Hash database storage structure

Algorithm 1. *Data Save Procedure*

```
Input: Sensor Data
Output: Success
1: identity ←analyze sensor data;
2: HexData ←encode the data to hex string;
3: Transaction ←ethCreateTransaction(HexData);
4: TransactionHash ←ethSendTransaction (Transaction);
5: while TransactionIsNotValid(Transaction) do
6:  waiting for block generate;
7: end while
8: BlockHash ←GetBlockHashByTransaction(Transactionhash);
9: write identity, BlockHash, TransactionHash to database;
10: return Success;
```

Algorithm 2. *Data Query Procedure*

```
Input: identity
Output: sensor data
1: TxHash ←SelectT xHash(identity);
2: if T xHash is not NULL then
3:  sensorData ←GetDataByHash(T xHash);
4:  return sensorData;
5: end if
6: BlockHash ←SelectBlockHash(identity);
7: Txs ←GetTxsByBlockHash(BlockHash);
8: for txhash in T xs do
9:  if SelectT xHash(txhash) is NULL then
10:       Write txhsh to database;
11:  end if
12: end for
```

4.2 Abnormal Data Detection

Agricultural IoT system based on block-chain faces massive IoT data every day, it will bring a lot of pressure to the blockchain storage and affect the query efficiency if all the sensor data and logistics information are stored.

In response to this problem we have designed an anomaly data detection method. According to the quality parameter range defined by the product provider for each product, we define the sensor data which not in the range is abnormal data. Instead of writing massive all the sensor data to the blockchain, little abnormal data will be wrote to the block-chain. While the massive normal data can be wrote to distributed file storage system, for example IPFS.

The quality-related parameters (temperature, humidity, barometric pressure, etc.) of agricultural

products during storage and transportation generally have an acceptable normal range. The quality-related parameters' range depend on the manufacturer. In the process of product transport, the sensor perceives the external environment and uploads the data in real time. If sensor data is in the normal range, it will be wrote to distributed file storage system; otherwise it will be wrote to the block-chain.

As for quality parameters without normal range or quality parameters that can't set normal range, for example the readtime video streaming of production environment. We use historical data to predict current data, if the difference between the predicted value and the current collection value exceeds the threshold then the collection value will be wrote to blockchain as abnormal data.

$$\hat{x} = \sum_{k=1}^{k=\mu} \alpha_k x_{i-k} \tag{1}$$

Using voice prediction model for reference, as formula 1 defined. While predicting data number i, data before number i data will be selected to predict the prediction value of current collection value. defines the depth of memorization, it is selected according to the smoothness level of each quality parameter. and the threshold depend on the experts of the product.

4.3 Block-chain Optimization

Based on the consensus principle of block-chain, miners of the block-chain are always preparing to mine the potential block. Once a miner mined a new block, the transactions in the block are valid. While there can be several miners mined a new block at the same time then made the block-chain fork which affects the efficiency of the block-chain. In order to make the probability of block-chain fork in a reasonable interval, a suitable mining difficulty is necessary. Limited by the difficulty, the block rate can't be very fast. Bitcoin has a block rate of 10 minutes which is much slower than ethereum with a 15 seconds average block rate. However still far enough to support the mass IoT data written to the block-chain.

We design the product quality traceability and tracking system based on alliance chain. Based on the characteristics of the alliance chain, we optimized the ethereum block chain. Compared with public network block-chain, the alliance chain's network is more stable and controllable, the network consumption is much lower than public network. We design an adaptive adjustment strategy of difficulty coefficient which is suitable for alliance chain in order to improve the throughput of block-chain to meet the demand of massive IoT data by enhancing the block rate.

Current difficulty control algorithm of ethereum has a fixed factor of 10 seconds. Due to the network consumption the public chain's block rate is 15 seconds on average. The mining difficulty of next block is generated by the difficulty of parent block and current block. When the interval of current block time and parent block time lower than, it is considered current mining difficulty is low and a higher difficulty is expected. Otherwise there should be a lower difficulty.

$$difficulty_{next} = difficulty_{parent} + \left(\frac{difficulty_{parent}}{2048} \times \max\left(1 - \frac{blocktime_{current} - blocktime_{parent}}{\Delta t}, -99\right)\right) \tag{2}$$

In our system, in order to adapt massive IoT data, an adaptive throughput that can adjust by the data increment per unit time is expected. When the data increment increasing, the mining difficulty become lower and the block rate become fast then the throughput increase; When the data increment decreasing, the mining difficulty should slightly higher or unchanged in order to decrease the blank block.

$$\Delta t = \max\left(10 - INT\left(\frac{data[i] - data[i-1]}{data[i-1]}\right), 6\right) \tag{3}$$

We define data [i] represent the data amount of number i unit time. Because can't infinitely reduced, a minimum value of 6 seconds is expected.

5 Experiments and Analysis

5.1 Experiments Setup

In this section, experiments are designed and analyzed. We run our experiments on five computers. The proposed systems are implemented with the Java programming language and Go language with version 1.9. The experiments are conducted on these computers with an Intel(R) Core(TM) i7 6700 CPU processor, 16 GB memory, and the 64-bit Windows7 system. In these five computers, a storage system is deployed on a computer, which is responding for receiving sensor data. The others are used as the billing node for mining records. The configurations of server simulator are windows server 2012 R2 operating system with Inter Core Intel Xeon E5620 CPU processors, 128 GB memory.

We designed a blockchain system model based on agricultural IoT, where the sensor attached agricultural product uploads each data every two minutes. The user can query the corresponding sensor data from the blockchain using the product id.

5.2 Data Query Efficiency Comparison

To evaluate the performance of our method in query efficiency, we compare improved data query strategy with primitive blockchain data query policy.

When a user queries a sensor data in the original blockchain, it needs to traverse all blocks from end to source to search the required data. However, improved query strategy obtain the corresponding transaction hash according to the id from the secondary database, and then search the corresponding data according to the transaction hash from the blockchain. In this experiment, we query the transaction data in the top 1000. Results are shown in Fig. 5, where the horizontal coordinate represents the blockchain number and the vertical coordinate represents the query time. As the blockchain number increase, the query time of improved query strategy increase continuously. However, the query time of original query strategy do not increase any more.

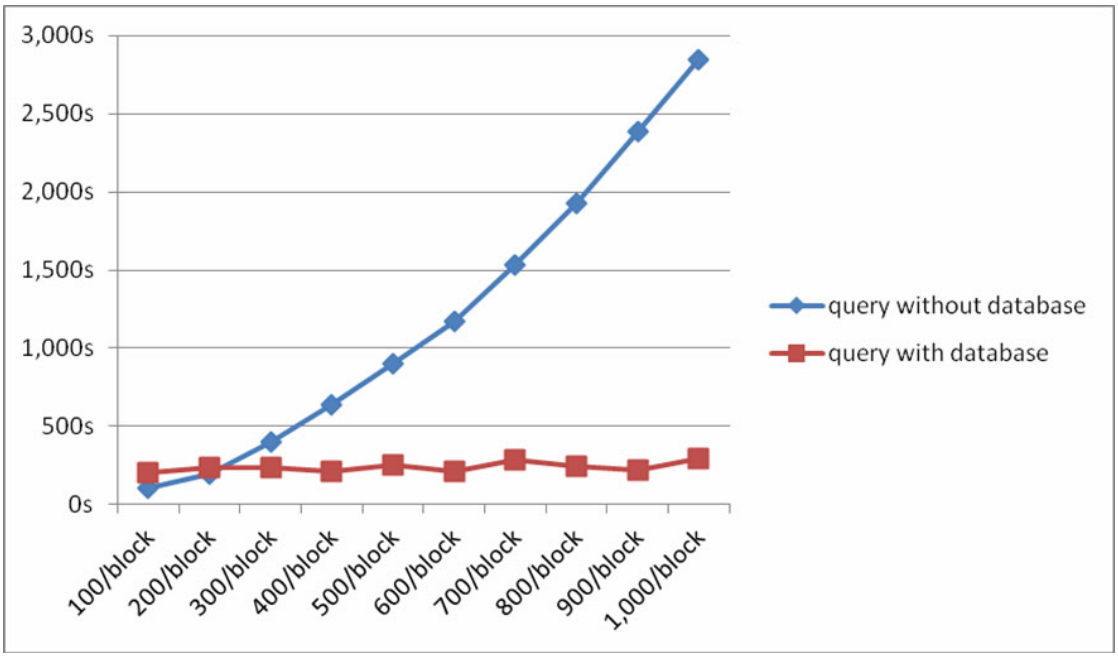

Fig. 5. Data query efficiency comparison

5.3 Block generation Rate Comparison

In order to optimize performance of our methods when writing large amounts of data, we improve the Ethereum mining difficulty control strategy. To evaluate the performance of optimized method in write

speed, we compare optimized method with original method. In this experiment, we set time unit as 10 seconds, and randomly generate number of concurrent threads for each time unit. The results are shown in Fig. 6. With the number of concurrent threads increasing, the data volume increases. In tests, when only creating a thread, 840 bytes could be written and five transactions could be sent in each 10 seconds. Fig. 7 shows the target block time and actual block time. From Fig. 7, we observe that the actual block time is basically the same as the calculated target block time. The reason is that our optimized method is used. Fig. 8 shows the blockchain write speed of two methods. Before optimization, blockchain write speed of the method keeps steady. When the amount of data suddenly increased, data being not written in time may be occurred. After adopting optimization method, blockchain write speed can change dynamically. With the data volume increasing, the blockchain write speed increases.

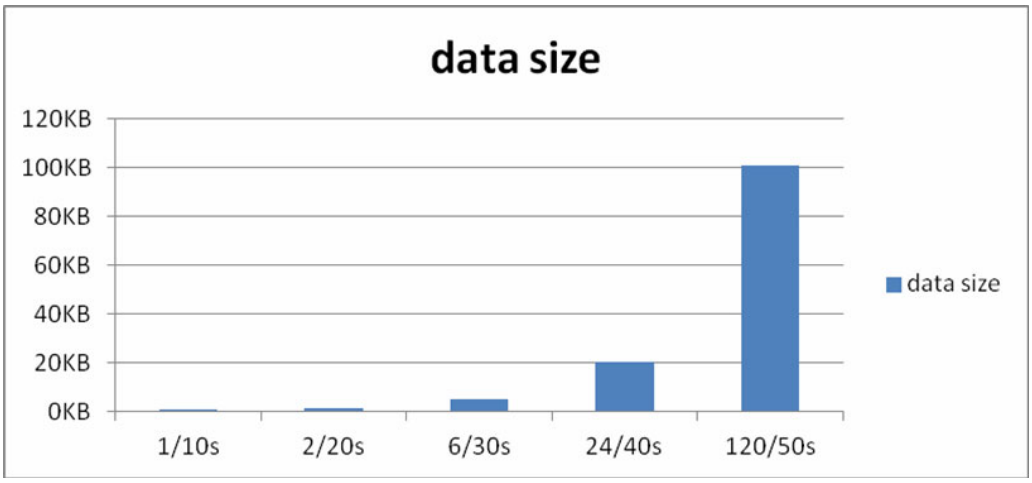

Fig. 6. Data change chart

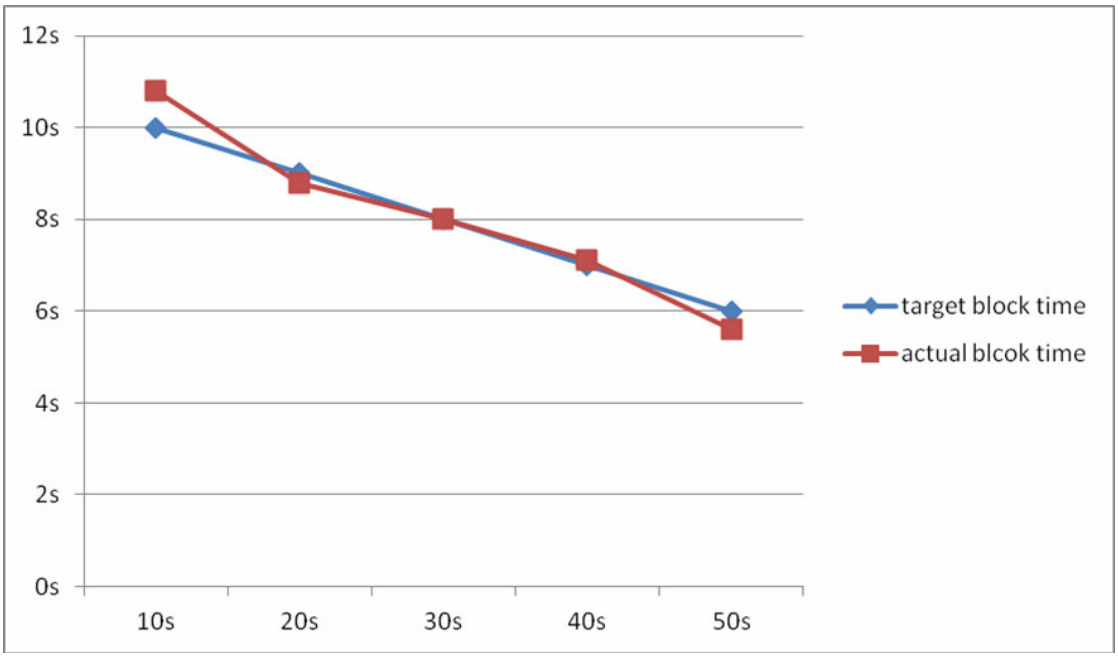

Fig. 7. Target and actual block time

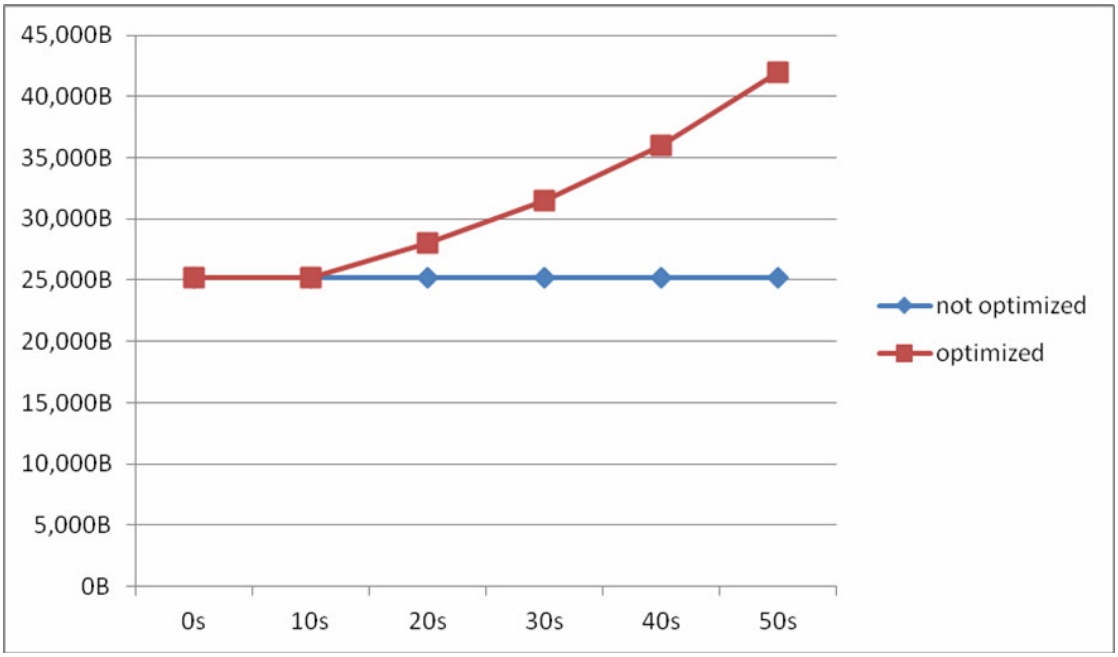

Fig. 8. Blockchain write speed

6 Conclusion

This paper explores how the IoT and blockchain technology can benefit agricultural products safety. The focus of the research is showing how blockchain technology can be exploited in agricultural IoT. In this work, we proposed a blockchain query scheme based on hash database, which can improve query efficiency. In addition, anomalous data detection method based on voice prediction model is presented. It reduces storage overload of blockchain. Finally, we present an adaptive adjustment strategy for mining difficulty based on data variation which improve the throughput of blockchain.

In our future work, we will extend our work to improve the full node synchronization mechanism, which decreases synchronization precess overhead while ensuring blockchain security.

Acknowledgements

This work is partly supported by the National Natural Science Foundation of China under Grant 61672109, 61772085, 61370196.

References

[1] S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system. <https://bitcoin.org/en/bitcoin-paper>, 2018.

[2] Y. Zhu, R. Guo, G. Gan, W.-T. Tsai, Interactive incontestable signature for transactions confirmation in bitcoin blockchain, in: Proc. Computer Software and Applications Conference, 2016.

[3] M.B. Taylor, Bitcoin and the age of bespoke silicon, in: Proc. International Conference on Compilers, Architecture and Synthesis for Embedded Systems, 2013.

[4] S. Huckle, R. Bhattacharya, M. White, N. Beloff, Internet of things, blockchain and shared economy applications, Procedia Computer Science 98(C)(2016) 461-466.

[5] J. Gōbel, P. Keeler, A.E. Krzesinski, P.G. Taylor, Bitcoin blockchaindynamics: the selfish-mine strategy in the presence of propagation delay. <https://arvix.org/abs/1505.05343>, 2016.

[6] F. Tian, An agri-food supply chain traceability system for China based on RFID blockchain technology, in: Proc.

International Conference on Service Systems and Service Management, 2016.

[7] F. Zhang, E. Cecchetti, K. Croman, A. Juels, E. Shi, Town crier: an authenticated data feed for smart contracts, in: Proc. ACM Conference on Computer and Communications Security, 2016.