2018 IEEE Intelligent Vehicles Symposium (IV)
Changshu, Suzhou, China, June 26-30, 2018

Blockchain Based Provenance for Agricultural Products: A Distributed Platform with Duplicated and Shared Bookkeeping*

Jing Hua[1,2], Xiujuan Wang[1,3], Mengzhen Kang[1,2*], Haoyu Wang[1,4], Fei-Yue Wang[1,5,6]

Abstract— **The provenance (tracing) system of agricultural products is important for ensuring food safety. However, the stakeholders (growers, farmers, sellers etc.) are numerous and physically dispersed, making it difficult to manage data and information with a centralized approach. As a result, the production procedure remains non-transparent and trust is hard to build. In this paper, we propose an agricultural provenance system based on techniques of blockchain, which is featured by decentralization, collective maintenance, consensus trust and reliable data, in order to solve the trust crisis in product supply chain. Recorded information includes the management operations (fertilizing, irrigation, etc.) with certain data structure. Applying blockchain techniques to the provenance of agricultural product not only widens the application domain of blockchain, but also supports building a reliable community among different stakeholders around agriculture production.**

Key words – Blockchain, traceability, trust-building, agricultural product, food safety

I. INTRODUCTION

In recent years, the issues of food safety have attracted great attention. Pesticide and fertilizer residues on various agricultural products have caused wide concern, and safe agricultural products are demanded urgently. To solve this problem, a complete tracing from production, wholesale, logistics to retail, needs to be provided, involving a series of issues including production standards, business reputation, certification, etc.

The traceability management system for agricultural product is supposed to be able to supervise the food quality and safety during the entire process from planting to consumption [1][2]. In China, although such agricultural product quality traceability system has been established, it has not yet been effectively promoted. Different companies and agencies in agricultural supply chain, including those for production, testing, storage, transportation or sales, are establishing their own data recording and traceability systems. This situation leads to many closed and independent databases, which are not worthy of trust-building due to the lack of supervision. Moreover, the information flow is impossible because of the incompatibility among software or data structure, thus it is difficult to achieve a complete tracking and provenance of information.

Besides, the integration of private data into a fully traceable platform is costly for individual stakeholders: it concerns not just the establishment of agricultural product traceability platform, but also the facilities for gathering, transmitting and exchanging information with high efficiency. Currently the automatic identification technology for product information records and inquiries is mainly bar code technology. Radio frequency identification (RFID) technology has good technical performance and high work efficiency. Since huge amount of data are produced by the terminals of Internet of Things for traceability, the construction and management cost for such system is high; open shared and reliable service is desirable.

Blockchain technology originated from the bitcoin technology in 2008 [3], which are mainly characterized by distributed, trustless, asymmetric cryptography, smart contract, and time stamp. The key contribution is that the blockchain allows building mutual confidence and trust among people without centralized authority, by providing mathematical solutions to the problem of trust. The basic idea is to ensure that the information is authentic and not tampered with by establishing a set of "public books" on the internet by "sharing" and "checking" all accounts in the network. The main features of blockchain technology can be summarized as decentralization, consensus trust, collective maintenance and reliable database[4][5].

Inspired by these technologies, agricultural traceability system can be established with a distributed network of the participants, the key component being a replicable and shared data recording system using consensus algorithms[6][7][8]. The problem of data management by scattered participants can be solved with blockchain techniques in a transparent and distributed way thanks to their mentioned characteristics. Starting from the source of the food industry chain (the growers), using Internet and Internet of Things (IoT), all participants including the farmers, food processors, distributors and catering companies, can record the quality-related information on the blockchain, and nobody can tamper once the record is recognized by the whole network. Therefore, not only the credit is guaranteed, but also the cost is decreased and the profitability is enhanced.

The objective of this paper is to propose a system architecture for blockchain based distributed agricultural

*Resrach supported by the National Natural Science Foundation of China (61533019, 31400623, 31700315).

1 State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences (SKL-MCCS, CASIA), Beijing 100190, China

2 Innovation Center for Parallel Agriculture, Qingdao Academy of Intelligent Industries, Qingdao 266000, China

3 Beijing Engineering Research Center of Intelligent Systems and Technology, Beijing 100190, China

4 Qingdao AgriTech Co., Ltd. Qingdao 266000, China

5 School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

6 Research Center for Military Computational Experiments and Parallel Systems Technology, National University of Defense Technology, Changsha 410073, China

traceability system, by constructing a consistent and unchangeable data structure of blockchain nodes.

The structure of this paper is as follows: the Section two introduces the techniques used in this paper, including blockchain, blockchain network, distributed consensus algorithm and digital signature; the Section three presents the basic data structure of agricultural traceability system; the Section four describes the design of the platform, especially focusing on the responsibilities of several roles in the platform. The Section five is the discussion and conclusion.

II. BLOCKCHAIN TECHNOLOGY

A. Blockchain

Blockchain[9][10] is a data structure formed by blocks linked together in chronological order. Each block consists of a block header and a block body, which is a collection of industry data such as bitcoin transaction records, smart contract codes, and agricultural tracing records as in this article, etc. The block header includes the metadata of the block. The most important part includes the timestamp of the block, the hash value of the block, the ID of the block, the ID of the parent block. The existence of the parent block ID makes all the blocks form a chain structure, as shown in Fig.1. The insertion of new blocks is allowed only in the tail, while the existing blocks are not allowed to be modified, which is a key rule of blockchain.

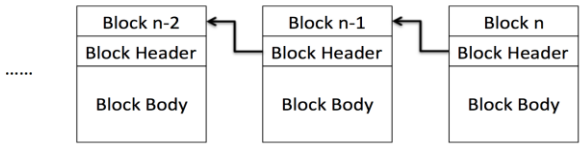

FIGURE 1 STRUCTURE OF A TYPICAL BLOCKCHAIN

B. Blockchain Network

Blockchain network[11][12][13] is a point-to-point network, i.e., the position of each participating node in the network is equivalent. There is no central control node or central router. The nodes transmit data and negotiate with each other by passing messages. Each node can save a complete piece of blockchain data. The more nodes involved, the more backups exist. In this architecture, data is owned and jointly maintained by multiple parties at the same time, which resulting two benefits: firstly, with the redundant backup of data, a single node crash or exit will not affect the overall stability; secondly, under joint management of multiple nodes, the data in blockchain has a higher credibility degree and lower possibility of being tampered than a closed database.

C. Consensus Algorithm

As mentioned, in blockchain network, each node maintains the same data. Thus, maintaining data consistency among all the nodes is a very important issue. Specific to the blockchain structure, each node needs to form a consensus on what is the next candidate block in the blockchain, which requires a distributed consensus algorithm. In the field of distributed computing, there are several classical algorithms for distributed consensus, such as PAXOS[14][15] algorithm and BFT[16] algorithm.

PAXOS algorithm has a great influence in the field of distributed storage. It is the basis of many distributed databases and plays an important role in the field of big data.

While the PAXOS algorithm considers that all nodes will not to send forge messages, the BFT algorithm considers the possibility that there are attacking nodes who would send tampered messages. It ensures that when the number of attack nodes does not exceed the threshold, consensus can still be reached.

In the bitcoin network, a competition mechanism and a prove-of-work (PoW) algorithm have been creatively designed[3][9]. This makes bitcoin a completely open network. The number of nodes does not need to be determined in advance. Anyone can join or leave the network at any time, as long as it complies the protocol.

These consensus algorithms can all be used for agricultural traceability platforms. According to the actual situation of the platform participants (whether credible or open), we can choose the appropriate algorithm.

D. Digital Signature

In any network system, user authentication is very important. Usually this task is done by the server, but there are no central servers in a peer-to-peer network. As a result, designers of bitcoin used the digital signature technology in encryption algorithms as a means of authentication[9].

Asymmetric encryption[17] requires a pair of keys, public and private. Only the private key can decrypt the data encrypted by the public key, and vice versa. At the same time, the private key can also be used for digital signatures. The signed information can be verified with a public key, ensuring that this information is indeed sent by the owner of the private key and has not been modified. Each user of the blockchain system uses an asymmetric encryption method to indicate the identity: the private key is kept by the user himself or herself, the public key can be freely distributed to others, all the information sent by the user is signed by the private key, and the other users use the corresponding public key to verify the authenticity of the information.

III. BLOCKCHAIN FOR AGRICULTURAL PROVENANCE

The target of an agriculture traceability platform is to record information related to production supply chain, including data for the production, processing, storage, transportation and distribution of agricultural products, so that all the process can be supervised by third parties (customers, insurance companies, etc.). As described in the introduction, the characteristics of blockchain technology fit perfectly the needs of agriculture traceability platform. The target of building an agriculture traceability platform based on blockchain technology is to record all related information on blockchain structures. This means to involve different companies and agencies to work together. Therefore, a generic structured representation of the data need to be defined.

In this section we introduce the concepts and definitions of agricultural traceability system. What is stored on the bitcoin blockchain is the transaction history, which is relatively simple. The content of agricultural traceability system is much more complex, it involves companies, seeds, fertilizers,

pesticides, time, agricultural operations, residue testing. It is difficult to cover all these information with a uniform structure, and there is bound to be a lot of redundancy, if possible. So we designed two related structures in the agriculture traceability system: basic planting information and provenance record.

A. Basic Planting Information

Agricultural traceability records include the information of agricultural production (the origin of an agricultural supply chain), processing, storage and other processes. Each record should contain a source production information, typical of which are listed in Table I, as they are common for a same batch of product. A unique label for every set of these information is defined as a global identification (the first item of Table I), which all records must contain.

TABLE I. DATA STRUCTURE OF BASIC PLANTING INFORMATION

Key	Explanations
identity	global identification
name-of-species	seed name
geographical-location	longitude and latitude of the greenhouse
planting-time	time when the planting begins
company-name	name of the planting company
greenhouse-number	label of the greenhouse
grower's name	name of the grower

B. Provenance Record

A provenance record contains the information of an agricultural operation. The structure of data body of the chain is designed as in Table II. The structure is generic in order to contain different kinds of operation.

TABLE II. DATA STRUCTURE OF A PROVENANCE RECORD

Key	Explanations
identity	global identification
date-time	operation time
location	location where this operation happens
company	Name of company who is responsible for this operation
person	Name of person who is responsible for this operation
operation-type	operation type, such as irrigation, fertilizing and spaying pesticides
inputs	Description of inputs in the operation, such as name and quantity of pesticides
memo	some additional information
digital-signature	digital signature of the company

IV. SYSTEM DESIGN OF AGRICULTURAL PROVENANCE

As mentioned earlier, there are many similarities between the technology required by agricultural traceability systems and blockchain technology. However, agricultural traceability systems also have unique characteristics and can not exactly borrow the design of existing blockchain networks. In this section we describe the design of agricultural traceability systems, especially the difference with the blockchain network.

System design diagram is shown in Fig.2. The platform mainly includes three roles: registration center, data node and clients, which are explained in following sections.

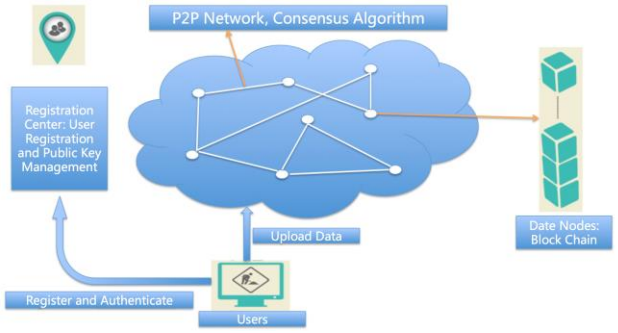

FIGURE 2 STRUCTURE OF AGRICULTURAL PROVENANCE SYSTEM DESIGN

A. Registration Center

Unlike the anonymity of bitcoin networks, agricultural traceability system needs to be clear that who submits the data, who is responsible for the accuracy and timeliness of the data, and therefore must include authentication process, which is the responsibility of the registration center. The registration center does not process any agricultural traceability data, Instead, it is responsible only for user registration. Users here refer to those who need to submit data to the platform, mainly including agricultural companies and institutions such as planting companies, testing organizations, transportation companies, storage companies, sales companies, etc.

The registration center receives user's registration application and verifies the user's true identity either online or offline (usually by ID card, copy of business license, etc.). After the authentication is completed, a pair of keys are generated by the user through the asymmetric encryption algorithm. The private key is kept secretly by the user. The public key is uploaded to the registration center and connected with the account of the company. Then the user's registration process is completed.

When users upload data to the platform, in addition to traceability information, each piece of data needs to be digitally signed with a private key. Since the registration center saves the public key corresponding to each user, anyone can confirm the true identity of the data uploader. This certification process has a dual meaning. For the inquirer, the data source can be confirmed. For the data uploader, it is effective to prevent others from uploading the fake data.

B. Data Nodes

Data nodes are the main parts of the system. They form a point-to-point distributed network and communicate with each other by message passing. Every data node is running in server mode and they have the following responsibilities:

- Accept the user's data upload request, verify the validity of the data, and broadcast legal data to other nodes. Users can initiate requests to any of the data nodes and upload a formatted data. After the node

receives the request, it firstly verifies the correctness of the data format and inquires the registration center about the validity of the digital signature. After passing the verification, the data is saved to the local data cache pool and broadcasted to other data nodes.

- Receive data broadcast from other nodes. Each node needs to receive the data broadcast from other nodes and determine whether it has existed in the blockchain or whether it already exists in the local cache pool. If both are negative, it is confirmed as new data and stored in the local cache pool.
- Tidy up block data. By receiving user uploads and broadcasts from other nodes, each node maintains a cache pool of unarchived data. When a certain amount of data is accumulated, the data nodes organize them into a single block according to a pre-defined format and use this block as the next candidate block in the blockchain.
- Send and receive block data broadcasts. Each time a node prepares a candidate block, it broadcasts the block to other nodes. At the same time it is also preparing to receive alternative blocks broadcast by other nodes. Because all nodes are running in parallel, different next candidate blocks may be generated.
- Run a distributed consensus algorithm. Since there are multiple candidate blocks, consensus algorithm is needed to decide which is the next. As mentioned earlier, several algorithms can be used, according to different situations. After the consensus is formed, each node receives this result, records the new block in the local blockchain, and deletes the data in the cache pool that has entered the blockchain, and be ready to run the next round of consensus algorithms.

C. Customers

Platform users can be divided into two categories: firstly, the users who need to upload data. They should register on the registration center and maintain their private keys for digital signatures, as mentioned; secondly, consumer users, or end users, who can query requests to data nodes and obtain the results. Normally they do not need to register, but being end-users, they are important components of this system.

V. Discussions and Conclusion

In the Introduction, we mentioned two issues in current agricultural traceability systems: the credibility of the data and the difficulty of integrating the subsystems of each company. According to the platform design described in this paper, these two issues can be well resolved. Firstly, the entire platform is a distributed peer-to-peer platform, the data nodes are maintained by various companies and agencies, the consensus and data consistency among nodes can be achieved with the algorithm; there is no single, private database. Therefore, once the data enter into the distributed network and form the consensus, it can no longer be modified, and the problem of data credibility is solved naturally. Secondly, since it was designed to be an open data-sharing platform, all companies are involved from the start, and there is no issue with subsystem consolidation. However, the implementation of this traceability platform designed in this paper may have a resistance: due to the openness of the platform, the data uploaded by the participating companies will be visible to all the participants, which means that some of the data considered as trade secrets will also be available to others. This could be avoided by choosing different data licenses by companies.

The proposed blockchain here can be used to record detailed operations in production and supply chain. Currently, very often limited information is recorded in the provenance system, such as the grower and location of an agricultural product. The procedural information such as which and when fertilizers have been used, and who and when have done test, are unknown. Customers are often confused by whether an agricultural product is really organic or green as claimed. The designed system is suitable for record the dynamic production information, with corresponding operation time. This makes the cost of cheating much higher, while those who are really doing healthy product can benefit.

The blockchain-based traceability of agricultural products can not only change the traditional product traceability system to a cloud platform supported by the underlying protocol of the blockchain, but also provide a decentralized data and information network. Although the application of blockchain technology in the traceability of agricultural products faces unprecedented challenges and uncertainties, its technical feasibility and market need have shown its tremendous potential. Once the blockchain landed on food traceability successfully, blockchain will usher in a new era.

References

[1] F. Lv and S. Chen, "Research on Establishing a Traceability System of Quality and Safety of Agricultural Products Based on Blockchain Technology," Rural Finance Research, vol. 12, pp. 22-26, 2016.

[2] Y. Yang and Z. Jia, "Application and Challenge of Blockchain Technology in the Field of Agricultural Internet of Things," Information Technology, vol. 258, pp. 24-26, 2017.

[3] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Consulted, 2008.

[4] Y. Yuan and F. Y. Wang, "Blockchain: The State of the Art and Future Trends," Acta Automatica Sinica, 2016.

[5] Y. Yuan, T. Zhou, A. Y. Zhou, Y. C. Duan, and F. Y. Wang, "Blockchain Technology: From Data Intelligence to Knowledge Automation," Zidonghua Xuebao/acta Automatica Sinica, vol. 43, pp. 1485-1490, 2017.

[6] Y.-b. Zhang, "The New Ecosystem of Cross-border E-commerce between EU and China based on Blockchain," China Business And Market, vol. 32, pp. 66-72, 2018.

[7] T. Hong, "Accelerating the Application of Blockchain in the Field of Agricultural Products E - commerce in China," Journal of Agricultural Information, pp. 18-20, 2016.

[8] Y. Yuan and F.-Y. Wang, "Parallel Blockchain: Concept, Methods and Issues," IEEE Acta Automatica Sinica, vol. 43, pp. 1703-1712, 2017.

[9] Andreas M A. Mastering Bitcoin: Unlocking Digital Cryptocurrencies. O'Reilly Media, 2014.

[10] Jerry B, Andrea C. Bitcoin: A Primer for Policymakers. Mercatus Center, George Mason University, 2013.
[11] George C, Jean D, Tim K, Gordon B. Distributed Systems: Concepts and Design (5th Edition). Pearson, 2011.
[12] Joshua K. The Mission to Decentralize the Internet. The New Yorker, 2013.
[13] Rudiger S. A Defnition of Peer-to-Peer Networking for the Classfication of Peer-to-Peer Architectures and Applications, Proceedings of the First International Conference on Peer-to-Peer Computing, 2002.
[14] L. Lamport, "Fast Paxos," Distributed Computing, vol. 19, pp. 79-103, 2006.
[15] L. Lamport, The part-time parliament: ACM, 1998..
[16] L. Lamport, R. Shostak, M. Pease. The Byzantine Generals Problem. ACM Transactions on Programming Languages and Systems. 4 (3): 382–401, 1982.
[17] Bruce S. Applied Cryptography: Protocols, Algorithms and Source Code in C. Wiley, 2015.