2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE 2021)

2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) | 978-1-6654-1540-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICBAIE52039.2021.9389859

Research on Anti-tampering Simulation Algorithm of Block Chain-based Supply Chain Financial Big Data

Fan Zhang , Ye Ding*
ShangHai Urban Construction Vocational College, ShangHai 201415，China
*corresponding author
succ@edu.com.cn

***Abstract*—In China, small and medium-sized enterprises are the foundation of overall economic development and are directly related to the employment and livelihood issues of the entire society. However, due to related issues, its financing costs are extremely high. For this reason, supply chain finance came into being. This article summarizes the shortcomings of traditional supply chain finance based on previous work experience. The author discusses the research on block chain-based supply chain financial big data tamper-proof simulation algorithm from four aspects: financial big data feature evaluation criteria, financial big data key feature extraction, supply chain financial big data key feature anti-tampering methods, simulation experiment and data processing content.**

Keywords: block chain; supply chain; big data; tamper-proof

I. INTRODUCTION

In the digital society, data publishing has become an important basic data service. Generally speaking, data needs to rely on a trusted third party to implement a comprehensive storage operation of the data through a centralized server during the release process. However, there are obvious single-point internal and external attacks among third parties. At the same time, centralized servers are also prone to software and hardware failures, which increases the risk of data tampering. The application of big data, blockchain and other technologies makes the anti-tampering work of supply chain financial big data more and more important.

II. SHORTCOMINGS IN TRADITIONAL SUPPLY CHAIN FINANCE

A. The Credit Problem of Core Enterprises

Information asymmetry is the main constraint when banks grant credit to SMEs. Generally speaking, banks mainly rely on the credit of core enterprises, and then derive credit from upstream suppliers and downstream distributors. However, it is often limited to the direct signing of trade contracts between first- and second-tier suppliers and core companies. Banks cannot identify the authenticity of transactions, and even if they can do so, they will have to pay a huge cost. From the perspective of improving economic efficiency and avoiding risks, banks will require companies to provide more collateral, which is very similar to the traditional SME credit model. It is precisely because the credit of core enterprises is difficult to penetrate, this has led to many small and medium-sized enterprises in the supply chain, but their status is obviously marginalized. It cannot enjoy the financing convenience brought by supply chain finance, so its financing difficulties and other problems are becoming more and more obvious[1].

B. The Supply Chain is Not Interoperable and the Transparency of Trade Information is Limited

From the perspective of practical development, the financial links of the supply chain are more complex and involve more companies. Enterprises cannot communicate with each other completely, and trade information cannot be shared, which eventually makes the edge enterprises on the chain appear more obvious. Banks also have no way to verify the authenticity of transaction information. Once the enterprises on the chain have forged purchase and sale contracts, the bank will suffer huge economic losses. Banks, as fund lenders and risk operators, need to fully consider the actual situation of edge enterprises on the chain to reduce the probability of risk.

C. Operation Method has Limitations

In supply chain finance, commercial drafts and bank drafts are the main financing tools. These traditional credit certificates themselves can reflect obvious limitations in operation. Actual commercial drafts and bank drafts cannot be separated, resulting in greatly reduced financing flexibility. In addition, the discount threshold of banks is very high. Many commercial bills cannot meet the bank discount standards due to corporate reputation restrictions. Coupled with the long financing process, it often takes a long waiting time from discounting to applying for funds to the account. There are also various limitations in the application of bills, which increase the difficulty of circulation and hinder the development of the supply chain finance industry.

III. COUPLING ADVANTAGES OF BLOCK CHAIN TECHNOLOGY AND SUPPLY CHAIN FINANCE

The main problem in the development of traditional supply chain finance comes from information asymmetry. Banks do not know about business operations and transaction information, which is more obvious in edge companies on the chain. Besides, the supply chain finance business is generally concentrated on the first- and second-tier suppliers or distributors of core enterprises, and it is difficult to fully develop it, which has restricted the financing capacity of the entire supply chain. Blockchain technology, as a technology that can transmit information, has many characteristics. These characteristics can also be coupled with supply chain finance to help solve the problem of big data tampering in traditional supply chain finance [2].

978-1-6654-1540-8/21/$31.00 ©2021 IEEE

A. Implement Decentralization and Distributed Accounting

From the development process of traditional supply chain finance, it can be seen that core enterprises are in the "central" role, and most chain enterprises will also rely on core enterprise guarantees to obtain bank financial support. For one thing, core companies in a strong position are not willing to provide full trade information. For another, core companies have no time to take care of those marginal companies on the chain, which makes the credit of core companies impossible to penetrate. If block chain technology can be added to it, the central department can be completely replaced, and then transformed into a large number of nodes, that is, a "point-to-point" information network. Among them, each node can participate equally to verify and record data. At this time, the transaction data on the chain will be distributed and recorded in the nodes, and they can maintain each other, and some nodes will lose in time. This will not affect the authenticity and integrity of the entire recorded data. With the help of block chain technology, banks can view the transaction information of enterprises on the chain efficiently and at low cost, and confirm the validity of transactions with core enterprises. Banks can pass on the credit of core enterprises to edge enterprises on the chain, and promote the use of supply chain finance to expand.

B. Block Structure with Time Stamp Based on Block Chain and Immutability

Throughout the entire supply chain operation process, there are many participants. Such as core enterprises, suppliers, distributors, etc., respectively control the contents of logistics, information flow and so on. However, these participants are independent of each other, which affects the generality and transparency of the entire supply chain information, and greatly affects the efficiency of the supply chain and the credit system. The block chain technology itself has a time stamp, which can record the time when a transaction occurs. This can ensure that the entire network authentication operation is carried out after each transaction is formed. Unless hackers can master 51% of the data nodes, they cannot tamper with the data. This can effectively avoid the problem of artificial data changes and strengthen the security of the entire system. All supply chain participants can use block chain technology to understand the effective information of each link. This can promote the interconnection of various links, break the phenomenon of information islands, and strengthen the transparency of this supply chain. Ultimately, banks' concerns about credit risk when implementing supply chain financial services are completely eliminated [3].

C. Achieve Digital Operations and Credit Transmission

Banks consume a lot of manpower and material resources when performing supply chain financial services. For example, in the development of prepaid accounts financing business, staff need to check the status and value of collateral. As for the accounts receivable financing business, the staff also need to confirm the rights of the bills provided by the enterprise. The operating cost incurred in this process is often borne by the enterprise, which further increases the financing cost. Smart contracts under the block chain are mainly based on immutable data, which allows computers to execute preset rules and terms on their own. In addition, the main role in the implementation of the consensus mechanism is to confirm the validity of data records and ensure the smooth execution of the transaction. Moreover, under the action of smart contracts and consensus mechanisms, it provides stable conditions for banks' digital operations. This reduces the frequency of human intervention, significantly improves overall work efficiency, and prevents fraud and operational risks. It is also because of the existence of smart contracts that the additivity of bills on the chain is greatly improved, and the split and transfer operations are completed under the witness of multiple parties. It can solve the problems that cannot be transmitted at multiple levels in traditional supply chain finance and strengthen the financing flexibility of enterprises on the chain.

IV. Research Content of Block Chain-based Supply Chain Financial Big Data Tamper-proof Simulation Algorithm

A. Criteria for Judging the Characteristics of Financial Big Data

In the block chain supply chain financial big data application, before extracting the key features of the distributed big data in the database, the bank needs to effectively set the criteria for judging the characteristics of the distributed data. This can ensure its stable application when evaluating the importance of features, and clarify the ability to distinguish between features. First of all, this type of data is all important data extracted from the decentralized supply chain financial database, and then mapped into a multi-dimensional space. For the convenience of research, the space is represented by R. In the meantime, the staff must also indicate the corresponding spatial dimension, clarify the corresponding set, and finally obtain the calculation formula for the difference degree of the subset with distributed data characteristics. The specific formula is as follows.

$$DFS_i = \frac{\sum_{j=1}^{l} x_i^{(j)} - x_i}{\sum_{j=1}^{l} \frac{1}{n_j - 1} \sum_{j=1}^{l} x_{k,i}^{j} - x_i^{(1)}}$$

This formula can represent all the data in the data set, and for the content of the numbered sample data set, it also involves the feature subset average vector. What'smore, in order to ensure the integrity of the data, the staff can use molecules to represent the characteristics of the data set and the sum of the central data vector of the subset. The binary distance corresponding to the feature center vector of all sample sets is extremely different. Among them, the denominator in the formula represents the internal distribution of each music line. In the actual data set, the difference between data features is further reduced. If the absolute value of DFS is in a large state, in the actual feature subset, it has the ability to identify the data feature category, and the effect is excellent [4].

B. *Key Feature Extraction of Financial Big Data*

From the application of the data score extraction method of distributed scores, it can be seen that distributed data can not only evaluate the standard combination, but also present the key characteristics of big data in the financial chain database. In the actual supply chain supply, it is a method of filtering data features using a split equation, or a feature extraction method of score values corresponding to size features. Generally speaking, the selection of actual data features is mainly based on the value of the sparse score. When extracting key features of actual data, the relevant staff can optimize the distributed supply chain financial big data collection in the database according to the sparse representation idea and the L1 norm minimization method. This can ensure that the acquired data can represent the sparseness of the entire data set and used to represent the reconstruction coefficient. When the actual distributed data is combined and set, it can be expressed as $\left[X_i\right]_{i=1}^{n}$. When calculating the vector dilution of the data, the coefficient matrix can be reconstructed and converted into the minimum linear programming solution of the L1 model between the data. The specific calculation formula is as follows.

$$Y = \min\left\|S_i\right\|, s, t. X_i = X`s_i$$

In the formula, X` represents the specific matrix obtained after deleting the data vector. During reconstruction, one can use s to represent the contribution value of the first data vector in the data set. Therefore, after obtaining a set of data dilution, the reconstruction matrix can also be clear, and common transformation coefficients can be expressed. When the reconstruction coefficient is clear, the difference between the reconstruction feature and the initial data feature of the data set can also be effectively calculated. We can summarize the reconstruction differences of the data samples and clarify the specific characteristics of the data set. If the difference is smaller, the more features the title retains its performance. After specific calculations and analysis, the desired objective function is finally obtained, denoted by S(r), which can meet the expected requirements:

$$S(r) = \frac{\sum_{i=1}^{n}(x_{ir} - (X_{si}^{-})_r)^2}{Var(X\ (r:))}$$

In this formula, the cumulative difference between the data set size feature and the corresponding reconstruction feature is a denominator, and the final Var(X(r:)) represents the dimensional feature dispersion of the data set. In all data sets, we can comprehensively arrange the retention performance of the sparse representation under each feature according to the order from small to large, clarify the smaller feature set, and finally clarify the characteristics of the data aggregated in the financial database of the supply chain. In general, the data function of the data set uses the score of the introduced concept to extract some data with small feature deviation. If the feature deviation is lower than the reconstruction deviation, the actual feature hint can also maintain the rare expression performance, and then become an important feature content of the data [5].

C. *Key Features of Supply Chain Finance Big Data*

In order to check whether the data characteristics have been changed, the relevant staff can use the monitoring code to realize real-time monitoring of the key characteristics of the distributed big data in the database. If the monitoring code can be placed in an appropriate place, the overall structure of the data will not change. With the feature function of encrypted data, we can improve the protection of the anti-tampering mechanism. After obtaining the function decryption key, each function can perform the corresponding decoding operation. At this time, the data function may be modified. The relevant staff can select the appropriate secret key generation function according to the different encryption methods of each part, and meet the following conditions:

$$K2_i = F_i\ (U,\ R,\ H_i\ (CF))$$

In this formula, Fi represents the decoding secret generation function of the first characteristic, and K2i represents the decoding secret generation function of the first characteristic. R represents the first value decoded by the registration code, U represents the first value decoded by the user code, and CF represents the array.

It can be seen from the actual data function that each element decoding key can be calculated by the hidden building function to calculate the data function, and then perform the function, and execute the previous function according to the specific function sequence. Generally speaking, the specified data key function often has some changes, it is difficult to perform the function normally, and finally causes the decoding to produce wrong results, and the decoding function cannot be performed normally. In contrast, the differentiated secret key generation function has obvious diversity characteristics and can protect the distributed data functions in the database. Figure 1 shows the ratio tolerance scheme of the main data functions.

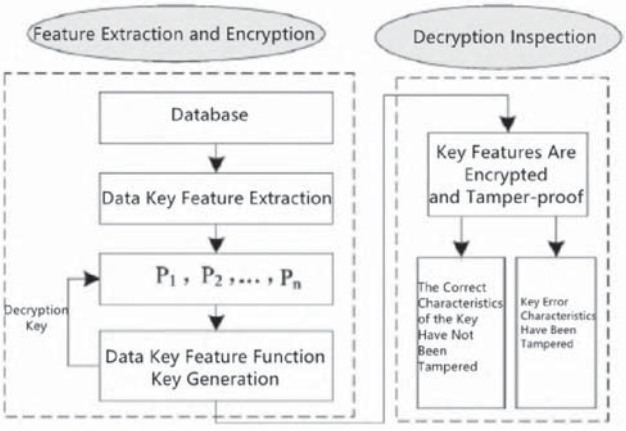

Figure 1. Data Key Features Anti-tampering Program Process

As can be seen from the information in the figure, after the first feature classification appears in the main feature classification of the data, once the encryption structure is calculated, a signature function is generated to check the secret key. Simultaneously, this will create a specific encryption signature function. Otherwise, the system should avoid the

modification of the big data key characteristics in the database to ensure that the data is always in a safe state [6].

D. Simulation Experiment and Data Processing

1) Specific Operation Process

In the supply chain, in order to prevent the forgery of distributed network data, relevant staff need to analyze the anti-tampering performance of distributed network information and data, and judge whether performance needs to be improved through experiments. The experiment is mainly carried out on two PCs, and a 10N/100M switch is set, IIS5.0 is used as the network server software, the operating system is Windows2016, and the information database is SWS Server2016. The calculation data required in the experiment mainly consists of simulation data. After the experiment begins, first of all, what the operator has to do is to reasonably verify the performance of the information in the process, achieve a distributed network protocol and conduct a reasonable encryption performance experiment on the information data. Secondly, companies can use Matlab experiments to evaluate the performance of the RDTP protocol. In the same state of data, people need to carry out experimental operations on the RDTP protocol, PCI and ECDG respectively to understand the specific safety performance under the evaluation results. The final data is shown in Table 1. In general, the information and data security of the decentralized supply chain financial network can be well protected.

Table 1. Specific Experimental Data Results

Cycle	1	2	3
Check the Number of Tampered Information	13643	14513	23440
Number of Error Messages	3	4	6
Error Rate	0.022%	0.039%	0.046%

2) Data Processing

It can be seen from the specific data processing results that the RDTP protocol can better protect the data transmission security of the distributed network of supply chain finance. In contrast with the ECDG protocol and the PCI protocol, the number of information data tampered with in the article RDTP protocol is limited. It can also be seen from this that the article approach can improve the anti-attack capability of the distributed network of supply chain finance. Furthermore, comprehensive prevention of data tampering with distributed network information in supply chain finance is carried out.

In general, by building a block chain accounts receivable financing service platform, the core enterprise accounts receivable can be digitized into payment settlement and financing tools, thereby ensuring that we can conduct external payments or financing operations at any time. From the perspective of enterprise development, block chain can establish a supply chain business circle between core enterprises and downstream suppliers, and achieve "no capital" transactions in the circle. This is also the essence of comprehensive control of transaction costs, which can control the overall cost of the supply chain and greatly improve the convenience of financing. From the perspective of bank development, it can achieve full-process electronic office, and it can also save the time and cost of offline confirmation of the authenticity of accounts receivable. The actual business model can also be replicated and promoted, extending to the entire supply chain financial service category, and the number of actual user groups can also be greatly increased.

V. Conclusion

In summary, in the research on anti-tampering of big data in supply chain finance, researchers need to put forward an encrypted blockchain-based method to prevent the main features of large distributed databases from being tampered with based on the existing deficiencies in existing research methods. In this way, the characteristics of distributed data can be determined and the process can be minimized, which is the essence of accuracy improvement. For code merging and monitoring operations, the block chain can use encryption algorithms to obtain functional keys to maintain the key data functional characteristics unaffected.

References

[1] Wang Feng. Research on auto finance risk control based on big data[J]. China Industry and Economics, 2020(19): 145-146.

[2] Chen Mianmian. Development prospects of supply chain finance under the background of big data [J]. Rural Economy and Technology, 2020, 31(18): 108-109.

[3] Liang Ying. Thoughts on financial innovation in the context of big data[J]. Journal of Liaoning Radio and Television University, 2020(03):77-81.

[4] Bai Xiaosan. Research on anti-tampering simulation algorithm of supply chain finance big data based on blockchain[J]. Journal of Jiujiang University (Natural Science Edition), 2020, 35(03): 65-68.

[5] Li Yang, Zhu Liang. Research on the Application of Big Data and Blockchain Technology in Financial Support for the Real Economy[J]. Jilin Financial Research, 2020(09): 43-45.

[6] Ta Huaishen. The application of big data in the financial risk management of the Internet supply chain[J]. Cooperative Economics and Technology, 2020(18): 74-75.