Research on Data trading Mechanism based on quality inspection and negotiated pricing

Wenqiang Li

Abstract— **With the rapid development of new technologies such as the Internet of Things, artificial intelligence and cloud computing, human beings have entered the era of big data, and with it comes the explosive growth of data. By using new technologies to analyze and process data, the hidden values in the data can be unearthed. Therefore, it is essential to establish a fair and secure data trading bridge between the generation and use of data. Currently, most of the data trading platforms are based on centralized institutions, but centralized trading platforms have problems such as single point of failure, privacy disclosure, transaction opacity, data resale and so on.Blockchain, as an emerging technology in recent years, provides new ideas for data trading due to its decentralized, traceable and tamper-proof technical characteristics. Relevant scholars propose blockchain-based data trading scheme, and although it solves some problems in centralized data trading, the existing scheme still suffers from unguaranteed data quality, privacy leakage, unreasonable data pricing, etc.A data trading model based on quality detection and negotiated pricing is proposed to solve the problems of single point of failure, leakage of user privacy, unguaranteed data quality, and unreasonable data pricing in existing data trading schemes. The model combines smart contracts, cryptography, trusted hardware and other technologies to accomplish fair data transactions while protecting user privacy. By building a test environment to verify the feasibility and performance analysis of the model, the experimental results show that the model not only solves the above problems, but also ensures the fairness and security of the whole transaction process.**

.

Index Terms— **Blockchain; Data Trading; Smart contract; Quality assessment; Price negotiation**

I. INTRODUCTION

With the development and progress of The Times, big data is playing an increasingly important role in today's society. Emerging technologies and applications have entered People's Daily life, such as intelligent systems [1] based on the Internet of Things and social network applications. When people use these applications, they generate a lot of data, which leads to the explosive growth trend of data. According to the data released by the Internet Data Center, the global data volume is expected to grow to 175ZB by 2025 [2]. Because of the important value hidden in the data, it has been widely concerned by the government and enterprises.

New technologies such as the Internet of Things, artificial intelligence and data mining have accelerated the development and innovation of big data application [3]. For example, people can analyze and process data through new technologies to dig out the hidden value in the data [4]. Therefore, it is crucial to build a fair and secure data transaction bridge between the generation and use of data .

At present, more consolidated data trading platforms at home and abroad include Guiyang Big Data Exchange, Shanghai Big Data Trading Center, BDEX, INFOCHIMPS, etc. As the intermediary between the two sides of data trading, these platforms promote data trading to some extent, but in this data trading mode, users must fully trust the data trading platform. However, based on the centralized data trading platform, there are the following problems :

(1) The centralized trading platform has great power and can grasp users' personal information, which may lead to the disclosure of users' privacy information.

(2) The centralized trading platform has the problem of single point of failure. Once it is attacked by malicious attacks, the system will be paralyzed, and both sides of the transaction cannot carry out normal data transaction.

(3) The data trading platform will charge a certain commission fee, and the platform may sell the data without the permission of the seller, which will ultimately affect the seller's earnings.

Blockchain technology is decentralized, immutable, open and transparent, providing new ideas for data transactions. Blockchain is a decentralized distributed ledger, which has been widely used in logistics chain, medical care, finance, Internet of things and other fields. Because of its decentralized characteristics, it can solve the problems existing in the centralized data trading platform. Smart contract can be simply understood as a programmable contract, when the set conditions are met, the program will be automatically executed. The characteristics of blockchain technology and the automatic execution of smart contracts provide good technical support for decentralized trading platforms.

To sum up, most traditional data trading platforms rely on centralized institutions, which not only face high commission fees, but also risk of single point of failure and privacy disclosure [5], and users' data security cannot be well guaranteed. The organic combination of blockchain technology and smart contracts provides new ideas for data trading and ensures that data trading can be completed even if there is no centralized organization in the process of data trading. Although the blockchain-based data trading platform has solved some problems, there are still problems such as data quality cannot be guaranteed, data pricing is

Manuscript received December 11, 2022

Wenqiang Li, School of Software, Tiangong University, Tianjin, China.

unreasonable, and privacy data is leaked.

II. BACKGROUND

A. Research on quality detection of big data

Zhao et al. [6] proposed a fair data transaction protocol based on blockchain. Introduction of marketing manager to solve the problem of data availability. Although this solution can solve the problem of data availability to a certain extent, marketing manager is a centralized organization, which violates the design concept of decentralization. In addition, the centralized organization has too much power, and once attacked, users' privacy information may be leaked.Xiong et al. [7] proposed a data transaction model based on blockchain and machine learning, and introduced an arbitration mechanism. When the data purchaser receives the data, if there is a dispute over the availability of the data, they can initiate arbitration to the arbitration institution, which detects the data availability by using the similarity learning method under the blockchain. When it is found that the data sent by the data provider is inconsistent with the declaration, the arbitration authority can recover the economic losses of the data purchaser. Although the data buyer is protected to some extent, the arbitrator may have access to the plaintext data of the data provider.Zheng et al. [8] proposed a data trading platform based on blockchain, set reward distribution rules to promote users to provide high-quality data, and ensure the fairness of reward distribution through smart contracts. With the same amount of data, the higher the quality of the data, the more rewards the data provider will be allocated.Hu et al. [9] also proposed a big data trading system based on blockchain and proposed and implemented three detailed data quality evaluation schemes to evaluate the quality of data. Users who provide high-quality data can obtain high profits. While this approach encourages users to improve data quality, it does not fundamentally eliminate low data quality.

B. Research on Big data pricing

Liu et al. [10] proposed a two-stage Stackelberg game model for data pricing, which can effectively solve the pricing and purchase problems of data buyers and market agents. However, this model is inefficient and unrealistic when applied to big data transactions. Chen et al. [11] proposed a new model-based pricing framework. The framework differs from other solutions in that instead of pricing data, it directly prices machine learning model instances, selling the machine learning model instead of the raw data to the data buyer. But the study only looked at maximizing returns for data buyers.Shen et al. [12] proposed a big data pricing scheme based on differential privacy. Forward and reverse pricing mechanisms should be designed to reasonably price individual data, provide reasonable compensation to data providers, ensure no arbitrage conditions, and find a balance between data privacy protection and utility. Hu et al. [9] proposed a price negotiation model based on smart contracts, in which data providers and data buyers can negotiate multiple rounds of price and finally get a price satisfied by both parties. Although the model meets the demand of dynamic negotiation pricing, it does not take into account the issue of privacy disclosure.

III. RELATED WORK

A. Overall architecture of FDTM

A new fair data transaction model FDTM based on block chain is proposed. The overall framework of FDTM model is shown in Figure 1.

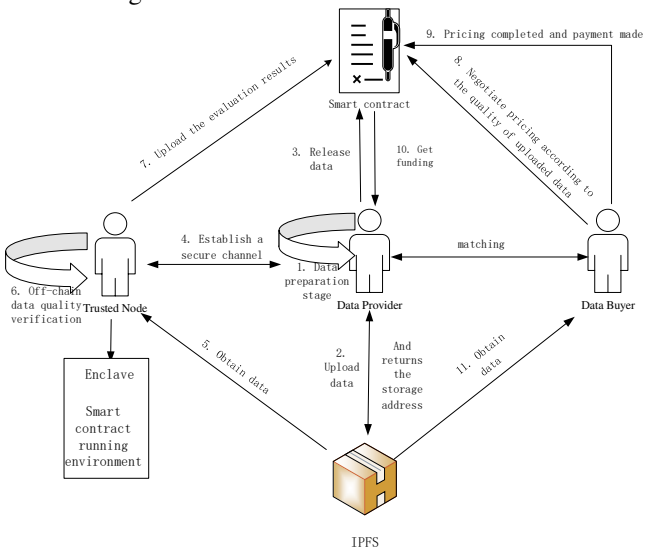

Fig. 1. The overall framework of FDTM

In this model, there are five roles: data provider, data buyer, trusted node, smart contract, and IPFS. The five roles are described in detail below:

Data Provider (DP) : Mainly releases data to meet users' requirements according to the demand of data trading market, and obtains certain profits after trading.

Data Buyer (DB) : it mainly finds commodities that meet its needs in the data trading market, transacts through the data market, and obtains corresponding data by paying a certain price.

Trusted Node (TN) : Provides a trusted execution environment for quality inspection and price negotiation. The trusted environment provided by this node can protect users' privacy data and prevent data leakage. You can get a certain amount of profit.

Smart contract (SC) : As the core of the data trading framework, SC can ensure smooth execution in accordance with the established rules of the contract under the premise of no centralized organization, and all transaction information is publicly visible. Not only manage DP deposit and DB payment funds, but also complete the fund settlement function.

Inter Planetary File System (IPFS) : Mainly used for data storage. DP encrypts the data to be sold and uploads it to IPFS for storage. DP data integrity can be ensured to prevent users from tampering with data. DB or TN can download required data based on phase address.

B. FDTM process analysis

Based on the data transaction model of blockchain, the full text is mainly divided into four stages: data release, data quality detection, negotiated pricing and data transaction.

(1) Firstly, DP divides the data set to be sold into n fragments, and then uses AES algorithm to encrypt the n

fragments respectively. After the encryption is completed, the ciphertext is uploaded to IPFS for storage. IPFS will return the storage address corresponding to the ciphertext, and DP forms a Merkle tree according to the n storage addresses returned by IPFS. Calculate the root hash of the Merkle tree and save it. Subsequently, a deposit of a certain amount of data should be deposited, which should be greater than the sum of the total data price and TN rewards. Finally, call the contract that published the item and look for TN for quality inspection.

(2) TN conducts remote authentication with DP. After successful authentication, a secure channel is established for data transmission. Then TN uses the shuffling algorithm to determine the transaction order of data fragments to be tested for quality. The NIQE algorithm is used to evaluate the quality of the data, and the evaluation results are returned. The average value is calculated and returned through multiple rounds of quality evaluation, and then uploaded to the blockchain.

(3) DB checks the quality assessment results of the data and matches the data that meets its needs to carry out the transaction. Before the formal transaction between the two parties, DB can conduct dynamic price negotiation with DP. Both parties shall meet the negotiation rules during the negotiation process, and finally find a suitable price for subsequent transaction through multiple rounds of negotiation.

(4) When the price negotiation between DP and DB is successful, the final data transaction stage is entered. DB generates a Merkle tree according to the storage address of the data fragment, and then calculates its root hash value for verification to check whether the data is the same as that claimed by DP at that time. When the detection is successful, the final fund settlement stage is carried out, DP obtains the negotiated price, and DB obtains the corresponding key.

C. *FDTM detailed design*

(1) Release of data

First, DP preprocesses its own data set to be sold and splits it into n data fragments. Then, AES algorithm is used to encrypt all the processed data fragments, and all ciphertext is uploaded to IPFS for storage. IPFS will return the corresponding storage address. DP generates a Merkle tree according to the storage address returned by IPFS, calculates the root hash value of the Merkle tree and saves it. Finally, DP publishes the data information by calling publishData function and seeks TN to test the data quality. In response, the TN sends its address and public key to the DP, which uses Intel's official authentication platform to authenticate the TN remotely.

(2) Data quality assessment

After remote authentication is established between DP and TN and a secure data transmission channel is established, DP encrypts the address sequence using TN's public key. After the encryption is complete, DP sends the ciphertext of the address sequence to TN. After receiving the ciphertext of the address sequence, the TN uses its own private key to decrypt the address sequence. After decryption, the Knuth-Durstenfeld Shuffle algorithm is used to conduct N rounds of interaction with the address sequence to generate a new address sequence. Then the public key of DP is used to encrypt the new address sequence and return it to DP. Then DP sends data related information according to the new address sequence. TN obtains the data from IPFS according to the address, and uses the decryption algorithm to decrypt the ciphertext in its Enclave environment. After the decryption is complete, the NIQE algorithm is used to evaluate the quality of the data. Finally, after the evaluation result is returned, the data is destroyed inside the Enclave. The smart contract verifies the returned information, and when the verification is successful, TN gets a certain reward.

(3) Pricing negotiation

After a successful match between DB and DP, DB can check the quality of the data provided by DP and negotiate with DP to determine the final pricing of the data. This scheme is based on the Rubinstein bargaining model for data negotiation pricing. In the process of negotiation, SGX technology is used to ensure the privacy of the bidding information of both parties. Use smart contracts to set negotiation rules. The reason why we choose to use SGX technology is that the smart contract is directly applied to calculate the recommended price, and the bidding privacy of both parties may be leaked. Once the two parties fail to reach the final transaction, the subsequent data sale will be affected to some extent. Although the use of cryptography technology like state encryption to protect the privacy of the bidder, but the use of cryptography technology is very expensive.

(4) Data trading

When the negotiated pricing is completed between DB and DP, both parties conduct data transaction according to the negotiated final price. First, DB submits the data fragment id, public key pk, deposit money and deposit deposit to the smart contract. DP obtains the pk of DB and uses the pk of DB to encrypt k_i through the asymmetric encryption algorithm to obtain ciphertext w_i, where the formula is w_i=RSA.Encpk(k_i). DB obtains wi and decrypts it with its own private key sk to obtain the corresponding k_i, where the formula is k_i=RSA.Decsk(w_i). DB requests data from IPFS according to the address of the data fragment, obtains ciphertext data, and uses k_i to decrypt the data to obtain the data. After success, DB calls the contract to complete the data transaction process, and DP gets the corresponding remuneration.

IV. FEASIBILITY ANALYSIS

This section mainly conducts performance test for FDTM, which is mainly divided into two parts: off-chain function test and on-chain contract performance test.

A. *Off-chain functional testing*

The off-chain functional test is mainly divided into five parts: data encryption, Merkle tree generation, IPFS upload and download, signature and verification, and trusted node response time.

In the data processing stage, AES-256-CBC algorithm is used to encrypt data fragments, respectively for 100MB, 200MB,... Figure 6-24 shows the encryption time for 1000MB data. Through the analysis of the running time results, the time required to encrypt/decrypt data is related to the size of the data file. The larger the file is, the longer the

encryption/decryption time will be. Figure 2 shows that it takes about 6.5 seconds to encrypt 1000MB of data and 2s to decrypt 1000MB of data. In a data transaction scenario, this time consumption meets the requirements.

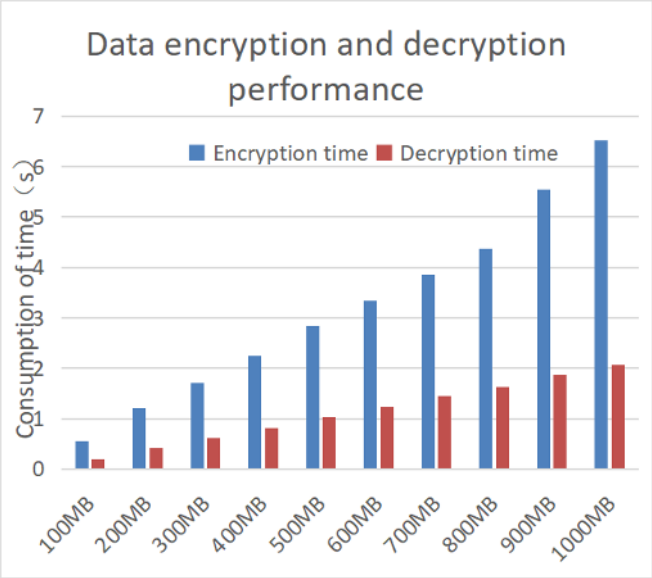

Fig. 2. Data encryption and decryption time

Figure 3 shows the time consumed by the data provider to construct a Merkle tree based on the storage address of the data. In this paper, we measured the time required to construct a Merkle tree with 100 to 1000 pieces of data. When the data pieces were 1000 pieces, the time required to construct a Merkle tree was about 3533μs. When constructing a Merkle tree, it is a process of hashing the data to finally get the root hash value. Therefore, when the number of data fragments increases, the time of constructing a Merkle tree will also increase linearly. In the actual data transaction, the time required for 1000 data fragments is 3533μs, which meets the data requirements.

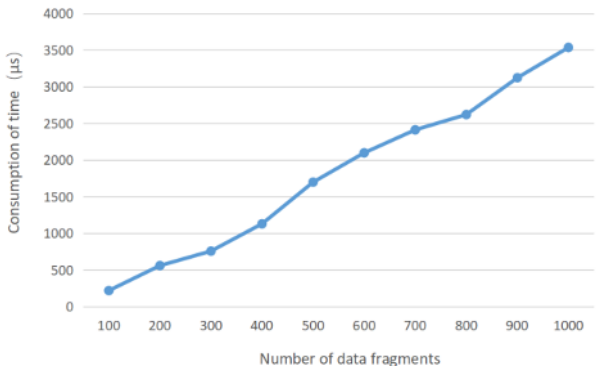

Fig. 3. Merkle tree generation time

In the data transaction in this paper, RSA algorithm is used to sign and verify the information in the key transmission stage and the bid stage of both parties. Figure 4 shows the running time. In a data transaction, the more times the key is transmitted and bid, the more times the signature is authenticated, and the longer time is required. The time for 1000 signatures is about 380ms, and the time for verification is relatively fast, and the time for 1000 authentication is about 35ms. In actual development, we can effectively improve the efficiency of signature/verification by reducing fragmented data.

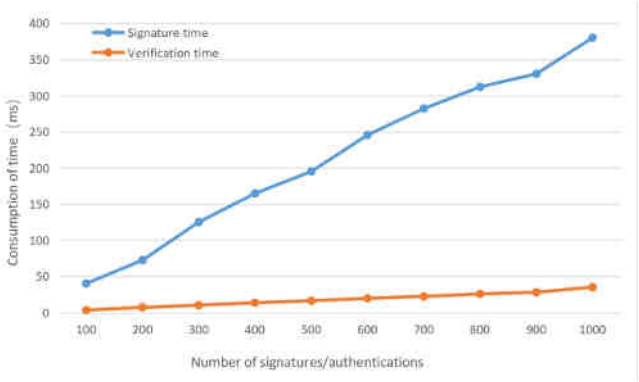

Fig. 4. Signature/verification time

B. Smart contract performance testing

Smart contract is the most critical part of the whole data transaction, and the execution efficiency of its various functions also affects the whole data transaction process. Since there are many functions in the whole data transaction process, this section mainly selects several key functions for testing. For the convenience of description, key functions are numbered, as shown in Table 1.

Table1 Number of key function in smart contract

Function name	Function Description
1	providerDeposits
2	publishDataItem
3	response
4	returnResult
5	nodeReward
6	initProviderPrice
7	initBuyerPrice
8	priceNegotiation
9	requesterDeposits
10	createData
11	paymentOrder

Figure 5 mainly shows the Gas consumption of 11 functions. During the deployment and execution of the contract, a certain amount of gas is required. The gas consumed by the function in this scheme is far lower than the upper limit of gas consumption, and the experimental results show that the gas consumed by the function execution in this scheme meets the requirements.

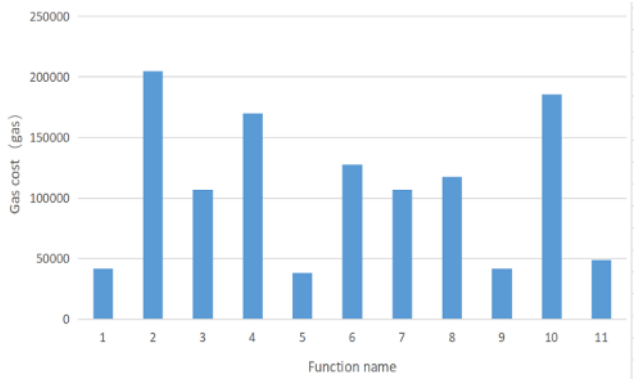

Figure 5 Gas consumption of key functions

V. CONCLUSION

This paper proposes a data transaction model based on quality detection and negotiated pricing, aiming at the problems of data quality guarantee, privacy disclosure, unreasonable data pricing and the lack of a fair data transaction model in existing data transactions. This model combines smart contract, cryptography, trusted hardware and other technologies to realize fair and secure data transaction without relying on the third party, and solves the existing problems in data transaction. Build a development environment and use Solidity language to compile the key contracts in the data trading model for feasibility verification and performance analysis of the model. The experimental results show that the data transaction model not only solves the problems in the data transaction mentioned above, but also ensures the fairness and security of the whole transaction process.

REFERENCES

I. HaddadPajouh H, Dehghantanha A, Parizi R M, et al. A survey on internet of things security: Requirements, challenges, and solutions[J]. Internet of Things, 2021, 14: 100129.

II. Reinsel D, Gantz J, Rydning J. Data age 2025: the digitization of the world from edge to core[J]. Seagate, 2018.

III. Baig M I, Shuib L, Yadegaridehkordi E. Big data adoption: State of the art and research challenges[J]. Information Processing & Management, 2019, 56(6): 102095.

IV. Tanwar S, Bhatia Q, Patel P, et al. Machine learning adoption in blockchain-based smart applications: The challenges, and a way forward[J]. IEEE Access, 2019, 8: 474-488.

V. Adlakha R, Sharma S, Rawat A, et al. Cyber Security Goal's, Issue's, Categorization & Data Breaches[C]//2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, 2019: 397-402.

VI. ZHAO Y Q, YU Y, LI Y N, et al. Machine learning based privacy-preserving fair data trading in big data market[J]. Information Sciences, 2019, 478: 449–460. [DOI: 10.1016/j.ins.2018.11.028]

VII. Wei X, Li X. Smart Contract Based Data Trading Mode Using Blockchain and Machine Learning[J]. IEEE Access, 2019, 7: 102331-102344.

VIII. Zheng S, Pan L, Hu D, et al. A blockchain-based trading platform for big data[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). IEEE, 2020: 991-996.

IX. Hu D, Li Y, Pan L, et al. A blockchain-based trading system for big data[J]. Computer Networks, 2021, 191: 107994.

X. Liu K, Qiu X, Chen W, et al. Optimal pricing mechanism for data market in blockchain-enhanced internet of things[J]. IEEE Internet of Things Journal, 2019, 6(6): 9748-9761.

XI. Chen L, Koutris P, Kumar A. Towards model-based pricing for machine learning in a data marketplace[C]//Proceedings of the 2019 International Conference on Management of Data. 2019: 1535-1552.

XII. Shen Y, Guo B, Shen Y, et al. Personal big data pricing method based on differential privacy[J]. Computers & Security, 2022, 113: 102529.