

Received December 10, 2020, accepted December 22, 2020, date of publication December 25, 2020, date of current version January 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047398

Data Trading Certification Based on Consortium Blockchain and Smart Contracts

WEI XIONG AND LI XIONG 

Department of Information Management, School of Management, Shanghai University, Shanghai 200444, China

Corresponding author: Li Xiong (xiongli8@163.com)

ABSTRACT In this paper, the first data trading certification blockchain solution based on consortium blockchain and smart contracts is proposed to solve the certification data security problem in data trading, so as to realize the auditability, accountability and integrity of data trading. By the proof-of-authority algorithm, a cheap-and-quick consortium blockchain is built. By the consortium blockchain, smart contracts can be deployed safely and conveniently. By the Solidity language, a concise-and-effective certification data smart contract and data trading smart contract are constructed to ensure the certification data security. By deploying the certification data smart contract and the data trading smart contract on the consortium blockchain, the security, transparency and supervisability of certification data is carried out. By utilizing the consortium blockchain and smart contracts, the data trading certification model is established to ensure the certification data security in data trading. By the experiments, the consortium blockchain is successfully established, and the certification data smart contract and the data trading smart contract are successfully deployed, so that the certification data security is effectively guaranteed. Finally, by the Github, the source code of the certification data smart contract and the data trading smart contract is uploaded.

INDEX TERMS Certification data security, consortium blockchain, data trading certification, Ethereum, smart contract.


I. INTRODUCTION

As big data is widely used by enterprises, individuals or governments, their demand for data resources is growing rapidly. The security of data trading has become one of the bottlenecks restricting the circulation of data resources. This bottleneck has led to a contradiction between supply and demand in data trading [1]. The emergence of the data trading market partially alleviated this contradiction. To meet the increasing demand for data resources, the data trading market provides space for the interconnection among DSs (DS: data supplier), DPs (DP: data purchaser) and DTCs (DTC: data trading center).

In data trading, data trading certification not only enables data resources to be traded on demand, but also ensures the auditability, accountability and integrity of data trading. When disputes occur in data trading, the certification data can not only present complete trading details, but also serve as the basis for arbitration of trading disputes. Therefore, the certification data security is not only very important, but

also a key feature of data trading security. However, there is currently no special solution for data trading certification in the data trading market.

On the other hand, blockchain-based solutions have been applied to finance [2], medical [3], intrusion detection [4], food industry [5] and supply chain management [6]. Blockchain [7] not only has security features such as non-tampering, non-repudiation and traceability [8], but also has event system and tamper-proof ordered log. Furthermore, the consortium blockchain is between the public blockchain and the private blockchain. It is a blockchain participated by multiple organizations or institutions. The consortium blockchain [9] has 1) Scalability: The consortium blockchain can be expanded in a short time, and the degree of controllability is high; 2) Low risk: The consortium blockchain belongs to the cryptography system, and its controllability brings the low risk is more easily accepted by users; 3) Fast transaction speed: The consortium blockchain takes a small number of nodes with higher credibility to participate in verification and bookkeeping, which makes the transaction verification process shorter in time, and thus the transaction speed is faster than the public blockchain; 4) Adaptability:

The associate editor coordinating the review of this manuscript and approving it for publication was Mansoor Ahmed .

According to the actual transaction volume and transaction frequency, the transaction cost can be adjusted adaptively. Moreover, smart contract [10] is a programmable computer program. The characteristic of the smart contract is that once deployed to the blockchain, it can only be accessed and cannot be changed [11]. In addition, the public blockchain is completely decentralized, but the consortium blockchain is multi-centralized. And compared with public blockchains, consortium blockchains are supervisable. Therefore, the use of consortium blockchain and smart contracts can achieve a distributed data trading certification system. The consortium blockchain and smart contract technology can store detailed information about the trading between DS and DP, and ensure that the information is not tampered with. The complete trading details can be used as evidence of data trading disputes between DS and DP, so that DTC can make the correct decision.

The use of consortium blockchain and smart contracts in data trading certification faces some challenges. First of all, what kind of consensus algorithm to choose for data trading certification. The difficulty is that the selected consensus algorithm must have the characteristics of fast transaction verification, small amount of calculation and low transaction verification cost. Secondly, what kind of smart contract should be selected for data trading certification. The difficulty is that the selected smart contract must be easy to use, concise and effective. To solve these challenges, by comparing the characteristics of proof-of-work (PoW), proof-of-stake (PoS) and proof-of-authority (PoA), the PoA consensus algorithm is chosen to be used in data trading certification. In addition, Ethereum is an open blockchain platform designed to ensure security, adaptability and flexibility. It pioneered smart contracts. These contracts are decentralized applications that can run in full accordance with the program, act as an “autonomous agent”, and be deployed on the blockchain. Therefore, the Ethereum smart contract is chosen to be used in data trading certification.

In this paper, a blockchain solution for data trading certification is designed by utilizing smart contracts to ensure the security of certification data. The focus of the solution is to propose a data trading certification model (DTCM) based on consortium blockchain and smart contracts. And the content of data trading certification is written into the smart contract (i.e. certification data smart contract) and deployed on the consortium blockchain. This ensures that the certification data will be permanently saved and cannot be tampered with.

The contributions of this paper are summarized as follows:

- (1) The DTCM is proposed. The model has entities participating in data trading. From the perspective of the system designer, the content and method of data trading certification are formulated.
- (2) The data trading certification process is proposed. This process avoids the functions of direct access to data resources and matching supply and demand. These functions are unknown to the system designer.

- (3) In the process of designing the data trading smart contract, various algorithms are designed to automatically execute the payment of Ethereum cryptocurrency, encourage participants to act honestly and solve disputes about data trading.
- (4) The certification data smart contract is designed and implemented, and then deployed on the consortium blockchain to verify the theoretical results of the DTCM. The source code for the data trading smart contract and certification data smart contract has been uploaded to GitHub (<https://github.com/106968687/DTC-SC>).

The road map for this paper is as follows. Related work is provided in Section 2. Data trading certification blockchain solution is proposed in Section 3. Experimental results are introduced in Section 4. Limitations, security features, and performance analysis are discussed in Section 5. Finally, the conclusion is drawn.

II. RELATED WORK

In terms of data trading security research, representative research is mainly divided into the following three categories:

A. RAW DATA PROTECTION SOLUTION

Dai *et al.* [12] proposed a blockchain-based data trading ecosystem to realize that neither the data broker nor the buyer can access the original data of the seller, but can only obtain the required data analysis results. Xiong *et al.* [1] proposed a smart contract-based data trading model solution using blockchain and machine learning to solve the problem that the data trading center has the ability to retain data during the data trading process. He *et al.* [13] proposed a blockchain-based accountability data trading platform to build a distributed, secure and reliable data trading environment. Jung *et al.* [14] proposed a uniqueness index, a new strict data uniqueness measurement method and several accountable trading protocols, so that data brokers can blame dishonest consumers when they detect misconduct. Kokoris-Kogias *et al.* [15] proposed an auditable data management framework to decentralize the sharing and lifecycle management of private data, and enforce fairness.

B. PRIVACY PROTECTION SOLUTION

Zhao *et al.* [16] proposed a new blockchain-based fair data trading protocol in the big data market to ensure the availability of trading data, the privacy of data providers, and the fairness between data providers and data users. Gao *et al.* [17] used the concept of homomorphic cryptography and secure network protocol design to solve the privacy protection problem of data auctions in cyber-physical systems. Sabounchi *et al.* [18] used blockchain technology and contract theory to design a blockchain-based peer-to-peer data trading mechanism to achieve a good compromise between privacy and data utility. Iyilade *et al.* [19] proposed a user data sharing policy framework based on adaptive

purposes to eliminate the risk of personal information privacy leakage.

C. SAFE TRADING SOLUTION

Chen *et al.* [20] proposed a general data trading framework based on blockchain for the Internet of Vehicles to ensure safe and true data trading. Liu *et al.* [21] proposed a secure, decentralized Internet of Vehicles data trading system using blockchain technology, and designed an effective debt credit mechanism to support effective data trading in the Internet of Vehicles. Bajoudah *et al.* [22] envisioned an IoT data stream market that allows any pair of data providers and consumers to exchange data without any prior mutual trust. Guan *et al.* [23] proposed two secure, fair and efficient data trading schemes to solve the problems of complex transaction process, high transaction cost and possible unfair exchange in the existing data trading schemes. An *et al.* [24] proposed a Multi-round False-name Proof Auction scheme to protect the best data auction results from false-name bidding attacks.

To the research of data trading security, scholars have made many contributions to the research of raw data protection, privacy protection and safe trading. On the basis of existing research, it is found that there is another problem worthy of research, namely, the security of certification data in data trading. For this reason, the first data trading certification blockchain solution using smart contracts is proposed to ensure the security of certification data and the on-demand trading of data. In addition, the solution can certificate any type of data trading and ensure the integrity, non-repudiation and tamper-resistance of the trading, as well as the transparency and traceability of trading information. This helps preserve the rights of participants in the trading and build trust for any future events.

III. DATA TRADING CERTIFICATION BLOCKCHAIN SOLUTION

Data trading is the activity of DSs, DPs and data agents trading raw or processed digital information [25]. In this section, the DTCM with multiple DSs, DPs and DTCs will be described in detail. Then, the related data trading certification problem is formulated, and the data trading certification process is summarized. The description of symbols in this paper is shown in Table 1.



FIGURE 1. A data market with multiple DSs, DPs and DTCs.

Consider a data market with m DSs, n DPs, and u DTCs, as shown in **Figure 1**. In the real world, DS corresponds to a data producer, such as an organization with online user data. DP corresponds to an organization or individual who uses data to develop a project or conduct scientific research. Generally, DP does not directly interact with DS. Instead, there is an intermediary between DP and DS, so that both parties can interact. That is, DTC corresponds to a data agent.

TABLE 1. The description of symbols.

Symbol	Description
DS	Data Supplier
DP	Data Purchaser
DTC	Data Trading Center
DTCM	Data Trading Certification Model
CDSC	Certification Data Smart Contract
IC	Initial Contract
PC	Pending Contract
CC	Confirmation Contract
DTSC	Data Trading Smart Contract
PoW	Proof of Work
PoS	Proof of Stake
PoA	Proof of Authority

Here, assume that DTC is a completely trusted entity and will not violate the principle of fairness and justice under any circumstances.

Next, a data trading certification blockchain solution based on consortium blockchain and smart contracts is proposed, so that the certification of on-demand trading data between DSs and DPs is safely and reliably realized. The data trading certification based on the “model + rules” method is expressed, and the certification process is outlined. Furthermore, the DTCM is a six-tuple. And the rules are described from three aspects: trading path uniqueness, script contract progressiveness, and transaction verification method.

A. SMART CONTRACTS

The smart contracts for data trading certification will be introduced in this section.

Definition 1: Certification data smart contract

Certification data smart contract (CDSC) is a contract used to store transaction details. It mainly includes trading participants, resource number, resource purchase purpose, price, data resource purchase agreement, etc. It is divided into three stages: initial contract (IC), pending contract (PC) and confirmation contract (CC). Figure 2 provides a detailed overview of the certification data smart contract functions.

Definition 2: Data trading smart contract

Data trading smart contract (DTSC) is a contract for data trading. It mainly includes functions such as trading participation application, deposit payment, dispute handling, payment settlement and automatic fines. It is divided into three stages: start of trading, dispute resolution and end of trading. Figure 3 provides a detailed overview of the data trading smart contract functions.

B. MODEL

The model for data trading certification will be introduced in this section.

```
Init: DP, DS, DTC, resource name, resource number, purchaser name,  
  
    purchaser purpose, commitment  
  
negotiationPrice:  
  
    require (Caller is DS);  
  
    emit An event in which the trading price has been negotiated;  
  
    return Price;  
  
priceChecking:  
  
    require (Caller is DP);  
  
    emit An event in which the trading price has been checked;  
  
    return Price;  
  
addAgreement:  
  
    require (Caller is DS);  
  
    emit An event in which the data resource trading agreement is added successfully;  
  
    return Agreement content;  
  
checkAgreement:  
  
    require (Caller is DP);  
  
    emit An event in which the data resource trading agreement has been checked;  
  
    return Agreement content;  
  
supplierSignAgreement:  
  
    require (Caller is DS);  
  
    emit An event in which the DS has signed the agreement;  
  
    return The signature of the DS;  
  
purchaserSignAgreement:  
  
    require (Caller is DP);  
  
    emit An event in which the DP has signed the agreement;  
  
    return The signature of the DP;
```

FIGURE 2. Certification data smart contract.

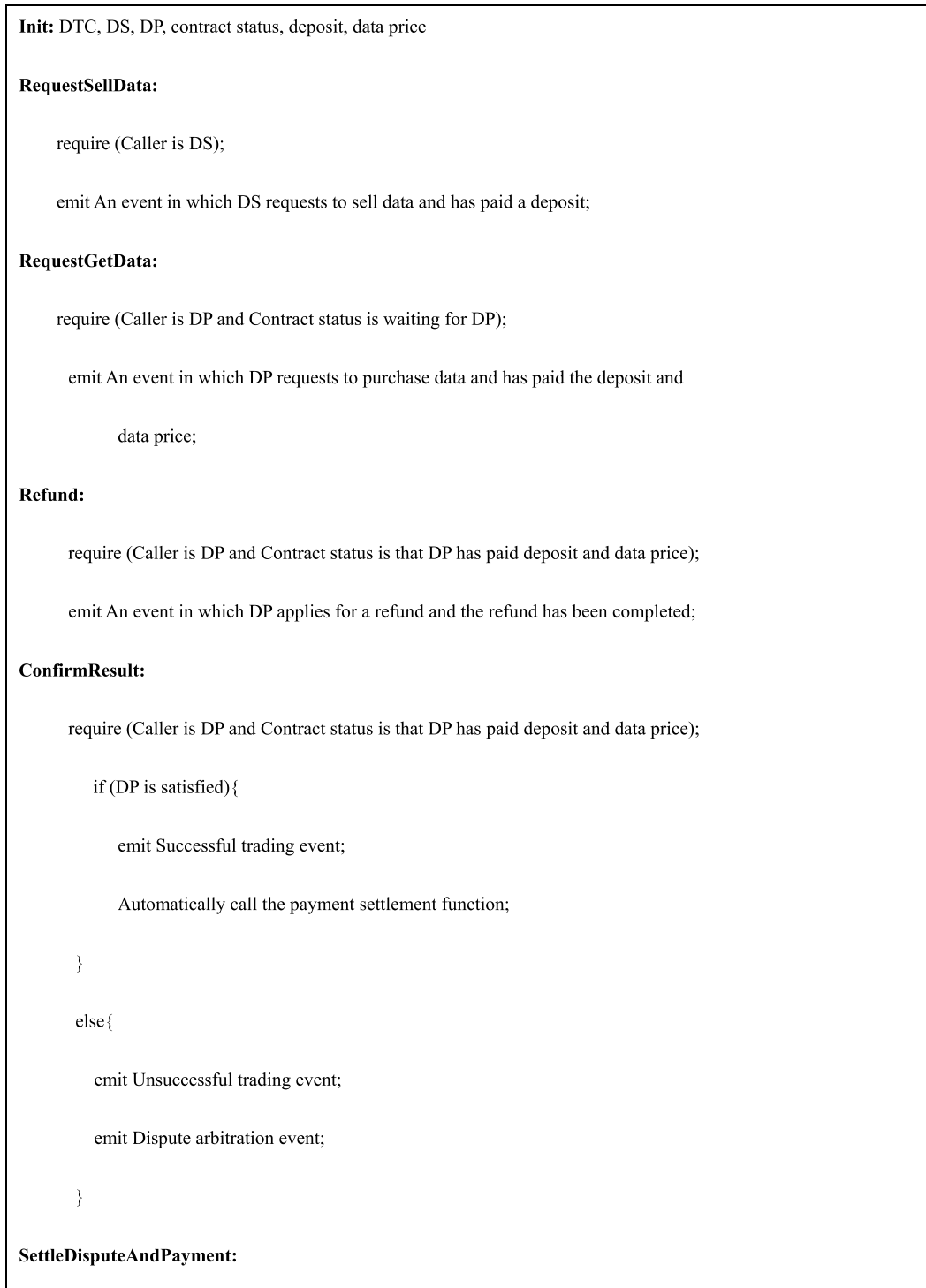


FIGURE 3. Data trading smart contract.

Definition 3: Data trading certification model

The DTCM is a 6-tuple: $DTCM = (DS, DP, DTC, DT, CB, f)$.

This model is simple and effective.

The description of the elements in the tuple is shown in **Table 2**.

The framework of data trading certification is shown in **Figure 4**.

The data trading certification process is outlined below:

- 1) DTC creates and deploys DTSC and can perform DTSC functions. DS and DP apply to DTC for public and private key pairs respectively. Then, DS and DP pay

```

require (Caller is DTC and Contract status is that DP is unsatisfied);

if (DP is right){

    emit An event in which the arbitration result is that DP is right;

    Automatically complete the refund to DP;

    emit An event in which the refund has been completed;

}

else {

    emit An event in which the arbitration result is that DP is wrong;

    Automatically call the payment settlement function;

}

Settlement:

require (Contract status is that the trading is successful or Contract status is that DP is

        wrong);

emit An event in which the payment settlement has been completed;
    
```

FIGURE 3. (Continued.) Data trading smart contract.

TABLE 2. The description of the elements in the tuple.

Element	Description
$DS = \{ds_m \mid m \in N^+\}$	DS represents a finite set of DSs, and ds_m represents the m -th DS;
$DP = \{dp_n \mid n \in N^+\}$	DP represents a finite set of DPs, and dp_n represents the n -th DP;
$DTC = \{dtc_u \mid u \in N^+\}$	DTC represents a finite set of DTCs, and dtc_u represents the u -th DTC;
$DT \subset (DS \times DTC \times DP)$	DT is a trading set, and DS provides data resources to DP through DTC;
$CB = \{cb_v \mid v \in N^+\}$	CB represents a finite set of bookkeeping and verification nodes in the consortium blockchain, and cb_v represents the v -th bookkeeping and verification node;
$f: DTC \rightarrow CB$	f represents the mapping from DTC to CB.

deposits to DTSC to meet the conditions of data trading. DS will obtain the (DSPK, DSSK) key pair. DP will obtain (DPPK, DPSK) key pair. DSPK and DPPK are the Ethereum addresses of DS and DP respectively.

2) DS provides data resource information to DTC. DTC feeds back data resource demand D and price P to DS, and binds resource number $RDigit$. In subsequent

trading, $RDigit$ is locked. DS broadcasts D , while monitoring the demand of DP.

3) DP selects broadcast information by matching requirements, and then uses the consortium chain communication protocol to negotiate with DS. DS and DP form a data resource supply path $dt \in DT$ through DTC.

- 4) After DS and DP reach a consensus, DS writes the complete trading certification information into IC. The parties use the Ethereum signature function to sign the contract. Then DS sends PC to DTC. DTC confirms CC.
- 5) $f(DTC)$ is mapped to CB, that is, DTC is used as a verification and bookkeeping node. After CC is verified by DTC, the transaction information will be broadcast on the entire blockchain network. DTC writes transactions into blocks and deploys CDSC on the blockchain.

The data trading certification algorithm is shown in **Algorithm 1**.

Algorithm 1 Data Trading Certification

Input: $DS, DP, DTC, DTSC$

Output: $dt, CDSC$

- 1: $ds_m \in DS, dp_n \in DP, dtc_u \in DTC$
 - 2: ds_m and dp_n apply for a public-private key pair from dtc_u and pay deposit to $DTSC$
 - 3: dtc_u sends corresponding key pair to ds_m, dp_n
 - 4: dtc_u sends $D_m, P_i, RDigit_w$ to ds_m
 - 5: if ds_m agrees with $D_m, P_i, RDigit_w$
 - 6: ds_m broadcasts $IC(D_m, P_i, DSPK_{ds_m}, RDigit_w)$
 - 7: else
 - 8: ds_m discusses with dtc_u
 - 9: back to step 6
 - 10: end if
 - 11: if dp_n matches IC
 - 12: $dt = (ds_m, dtc_u, dp_n, RDigit_w)$
 - 13: dt locks $RDigit_w$
 - 14: end if
 - 15: when disputes
 - 16: if negotiate
 - 17: ds_m or dp_n revises IC
 - 18: else
 - 19: ds_m publishes dt, IC
 - 20: end if
 - 21: end when
 - 22: dp_n signs $PC(D, P, DSPK_{ds_m}, DPPK_{dp_n}, RDigit_w)$
 - 23: ds_m signs $PC(D, P, DSPK_{ds_m}, DPPK_{dp_n}, RDigit_w)$
 - 24: ds_m sends $CC(D, P, DSPK_{ds_m}, DPPK_{dp_n}, RDigit_w)$ to dtc_u
 - 25: $cb_v \leftarrow f(DTC)$
 - 26: if DTC verified CC
 - 27: cb_v writes the transaction to the block
 - 28: cb_v deploys CC on the consortium blockchain
 - 29: cb_v publishes $dt, CDSC$
 - 30: end if
-

C. RULES

The rules for data trading certification will be introduced in this section.

1) TRADING PATH UNIQUENESS

The blockchain system is a peer-to-peer network, and interconnected distributed nodes can participate in transactions

together [7]. In the DTCM, when the designated DS performs a pre-transaction with the designated DP, the corresponding DP, DTC and DS will form a trading path dt . The blockchain network structure ensures that this trading path is only used for this trading. When DP, DTC and DS select the data resource trading, the resource number $RDigit$ bound to each party will be generated. When a data resource trading forms a trading path, $RDigit$ will also be locked. During the data trading process, $RDigit$ will be continuously written into CDSC to ensure the accuracy of trading participants and participation time by confirming the invariance of $RDigit$. When the data trading certification is performed after the trading is completed, the $RDigit$ at the beginning and end of the data trading must be the same, as shown in **Figure 5**.

2) SCRIPT CONTRACT PROGRESSIVENESS

In the DTCM, DS will generate a preliminary contract and send it to DP. After DP receives the contract, the two parties negotiate the trading details. After the two parties reach a consensus, a contract to be confirmed will be produced. The contract to be confirmed is sent to DTC for confirmation to generate the final contract. The state of CDSC describes the specific data trading certification process in a progressive manner.

- (1) $IC(D_m, P_i, DSPK_{ds_m}, RDigit_w)$ is the initial CDSC. It is generated by DS and waits for the corresponding DP to receive it.
- (2) $PC(D, P, DSPK_{ds_m}, DPPK_{dp_n}, RDigit_w)$ is the CDSC to be confirmed. It is generated after DS and DP negotiate the trading details. The contract contains digital signatures of DS and DP, waiting to be confirmed by DTC.
- (3) $CC(D, P, DSPK_{ds_m}, DPPK_{dp_n}, RDigit_w)$ is the confirmed CDSC. This is a contract for DTC to confirm the certification data of data trading. It contains digital signatures of trading participants. CDSC is deployed on the consortium blockchain by DTC.

3) TRANSACTION VERIFICATION METHOD

The PoW algorithm is reliable and secure, but it does not have real scalability. The PoW-based blockchains have limited performance in terms of transaction processing per second. This limitation is related to the fact that transaction processing relies on a distributed network of nodes. This requires nodes to reach a consensus and agree on the current state of the blockchain. This means that before a new block containing a transaction is confirmed, it needs to be verified and approved by most nodes in the network. Regarding transaction processing per second, blockchains based on PoS generally perform better than blockchains based on PoW. However, this difference is not obvious, and the PoS network does not really solve the scalability problem.

In this case, the PoA algorithm is feasible as a more effective alternative to achieve scalability because it can execute more transactions per second. The PoA is a reputation-based consensus algorithm that introduces a practical-and-effective

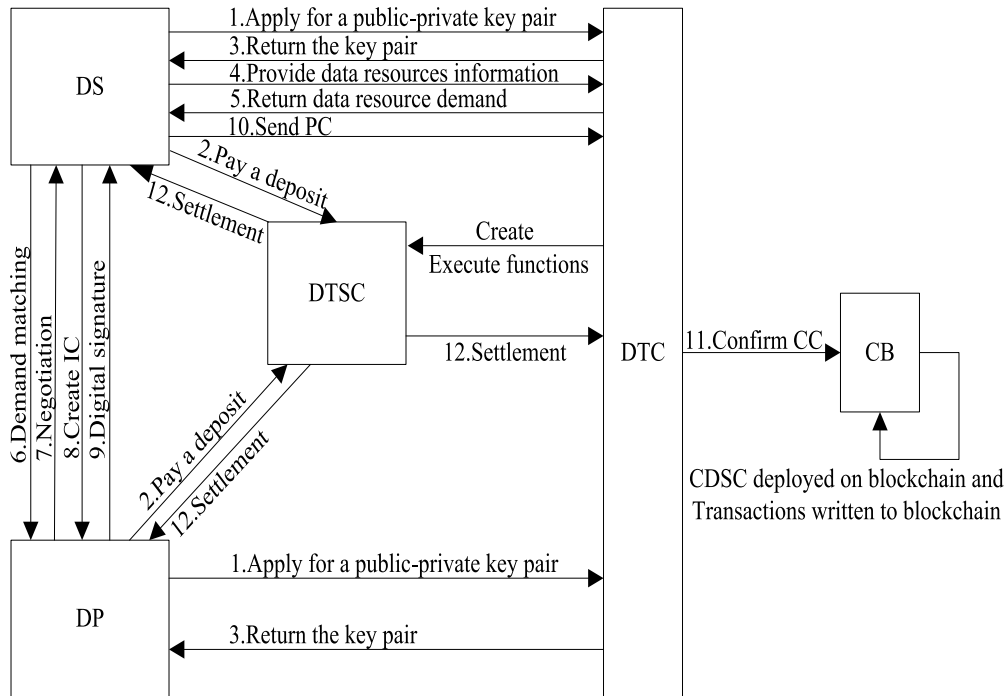


FIGURE 4. The framework of data trading certification.

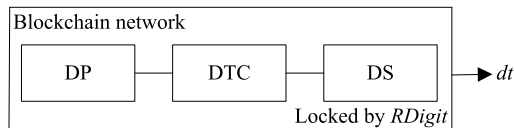


FIGURE 5. The framework of data trading path uniqueness.

consensus mechanism solution for the blockchain network. The characteristic of the PoA algorithm is that the block validator does not rely on the mortgaged cryptocurrency, but on the personal credibility. Therefore, the PoA-based blockchain is protected by an arbitrarily selected verification node that is a trusted entity.

The PoA model relies on a limited number of block validators, which makes it a highly scalable system. Blocks and transactions are verified by pre-approved participants who act as system managers. DTC meets the requirements to become a PoA-based blockchain verification node and will be pre-approved as a verification and bookkeeping node. PoA networks can achieve high throughput and scalability. As an emerging blockchain consensus mechanism solution, it is very suitable for consortium blockchain applications.

IV. EXPERIMENTAL RESULTS

The tool used to implement and test the solution is based on the consortium blockchain of the Parity-Ethereum wallet [26], while DTSC and CDSC are written based on the Solidity [27] language.

A. IMPLEMENTATION DETAILS

The advantages of the consortium blockchain are: (1) The number of custom nodes; (2) The number of mining nodes

can be set as required; (3) The transaction process does not require gas. The advantages of smart contracts are: (1) The contract content is open and transparent; (2) The content of the contract cannot be tampered with; (3) The contract can run permanently.

To reduce the time for waiting for blocks during the test, and to solve the situation that the internal network or external network of the organization cannot synchronize the blocks, a blockchain based on the PoA consensus algorithm is established.

The characteristics of building a consortium blockchain based on PoA (PoA Chain) are:

- (1) The PoA relies on the preset Authority nodes to generate blocks;
- (2) The number of Authority nodes can be set as required;
- (3) The block generation time can be specified, for example, the block is generated 5 seconds after the transaction is received;
- (4) The general Ethereum node can also be connected to the PoA Chain, and normally initiate transactions, contracts, etc.

The DTSC and the CDSC will be deployed on the PoA Chain.

DS applies to DTC to join the consortium blockchain and obtains public and private key pairs from DTC. DS pays a deposit to DTSC to be eligible to sell data resources. DS provides DTC with data resource information to be sold, and DTC feeds back resource demand, resource price, and resource number to DS. DS broadcasts resource requirements to the entire network and simultaneously monitors DP requirements.

DP applies to DTC to join the consortium blockchain and obtains public and private key pairs from DTC. DP pays a deposit to DTSC to qualify for the purchase of data resources. DP chooses DS according to its own needs, and negotiates trading details with DS through the consortium blockchain. DP, DTC, and DS form the data resource supply path.

After DS and DP reach a consensus through negotiation, the complete trading certification content is written into CDSC. CDSC is verified and deployed to the consortium blockchain by DTC. At the same time, DTC is also a book-keeping node for verifying this transaction. If the transaction between DP and DS is not disputed, the payment settlement of the transaction is completed, and DTC writes the transaction into the consortium blockchain. If the transaction between DP and DS is disputed, that is, the data resource provided by DS do not meet the DP requirements, DTC will arbitrate the dispute. According to the arbitration result, DTC will forfeit the deposit to punish DS or DP. In addition, if DS and DP are to exit the trading before the trading begins, the deposit can be refunded. The complete code of DTSC and CDSC can be used for all detailed information.

Figure 6 shows a sequence diagram of function calls and events that occur in three different situations. The sequence diagram shows the complete sequence of events, starting from the application of public and private key pairs to the end of payment settlement. **Case 1** in **Figure 6** indicates that the trading is successful and the data trading certification is successful. On the other hand, in the other two cases, DP points out that the data resources provided by DS can not meet the demand, but DS claims to meet the demand. Thus, **Case 2** and **Case 3** in **Figure 6** show DTC intervening in dispute arbitration. In **Case 2**, DTC determines that the DP is correct through CDSC, so the DP refund event occurs while the deposit of DS is forfeited. However, in **Case 3**, DTC determines that the DP is wrong through CDSC, so the payment settlement event of the trading occurs, and the DP deposit is confiscated.

The main contents of the CDSC code are as follows.

- (1) The information of trading participating entities;
- (2) The purpose of purchasing data resources;
- (3) The data resource number;
- (4) The price of data resource trading;
- (5) The agreement content of data trading;
- (6) The digital signatures of the trading participating entities.

The important algorithms used in the DTSC code are as follows.

1) REQUESTING DATA RESOURCE

Algorithm 2 shows the algorithm of DP requesting data resource. DP first calls the RequestGetData () function in DTSC to pay a deposit for participating in the trading. Then, DP selects DS by matching the data resource information.

After this process, DP and DS reach a consensus through negotiation and sign CDSC. Then, DS sends CDSC to DTC

for verification. After CDSC is verified by DTC, it will be deployed on the blockchain.

Algorithm 2 Requesting Data Resource

Input: EOA, deposit, DTSC status, DP status
1: EOA is the set of Ethereum addresses saved in DTSC
2: Restrict access to any DP \notin EOA
3: if msg.value = deposit then
4: if DTSC status = waiting for DP then
5: Withdraw the msg.value ether from DP.
6: Change DP status to next status.
7: Create a notification about DP.
8: end
9: else
10: Revert DTSC status and show an error.
11: end
12: end
13: else
14: Revert DTSC status and show an error.
15: end

2) SUCCESSFUL TRADING AND PAYMENT SETTLEMENT

After DP obtains the data resource from DS, DP needs to execute the ConfirmResult () function in DTSC so that DTC knows that DP has obtained the data resource. **Algorithm 3** shows the algorithm of the successful data resource trading and payment settlement. When the DP is satisfied, the DTSC will proceed with the payment and settlement of the trading. Then, the trading is successfully completed and both the DS and DP deposits will be refunded. In addition, the DTC and DS will get their share of profit due.

3) DISPUTE RESOLUTION

If the confirmation result of DP is false, that is, the DP is not satisfied with the obtained data resource, DTC will intervene in dispute arbitration. **Algorithm 4** shows the algorithm of the DTC resolving the dispute. The DTC will use the deployed CDSC to arbitrate the dispute. If the arbitration result is that the DP is right, the DP will receive a refund. At the same time, the deposit of DS will be forfeited. If the arbitration result is that the DP is wrong, the deposit of DP will be forfeited. At the same time, the payment settlement is performed. After the trading is successfully completed, the DTC and DS will receive corresponding profit shares.

B. TESTING AND VALIDATION

The established PoA Chain, as well as the compiled and deployed DTSC and CDSC will be introduced in this section. The DTSC and CDSC will be written and debugged in Remix IDE [28]. After the Parity network is configured in Truffle [29], the DTSC and CDSC will be compiled and deployed.

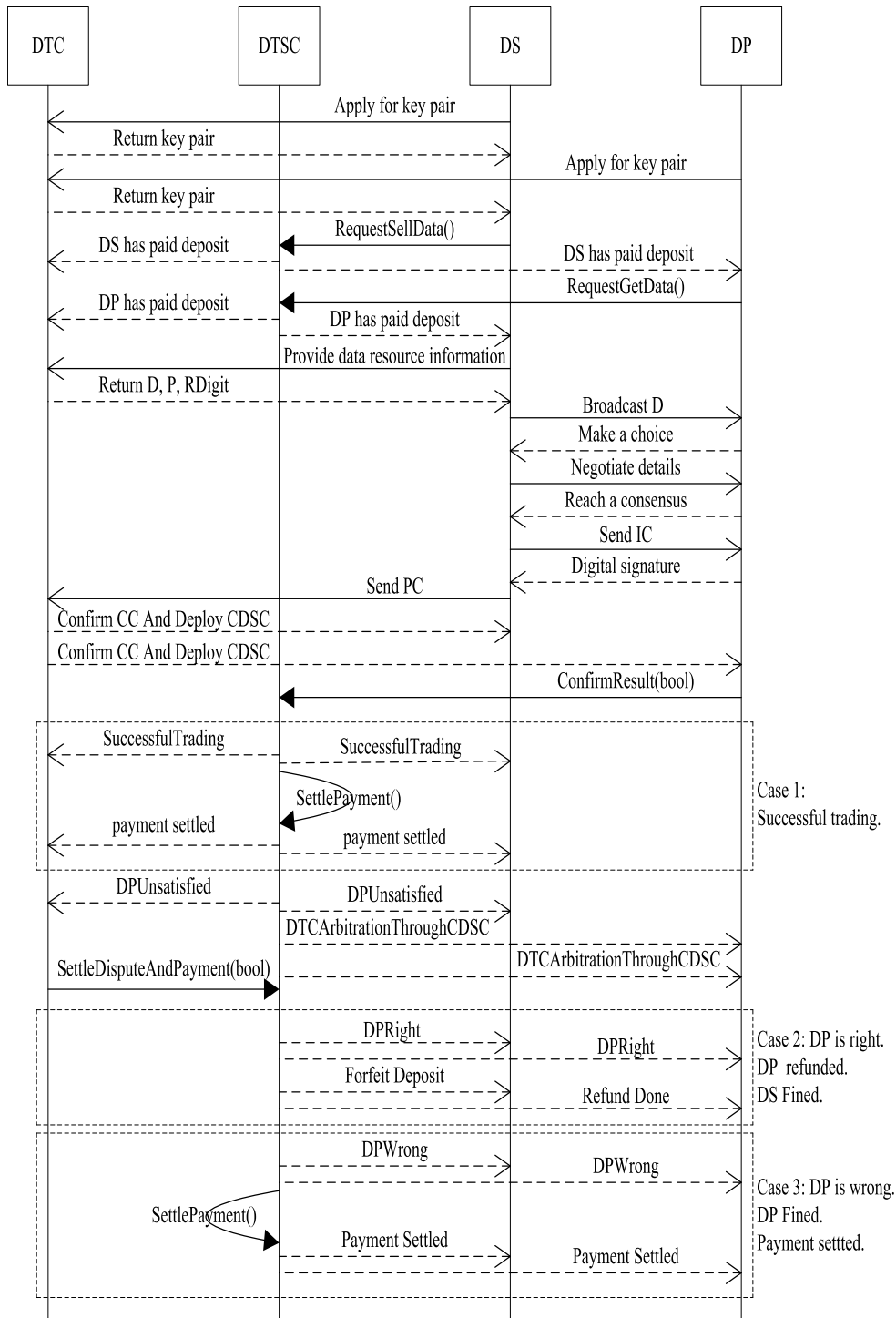


FIGURE 6. The sequence diagram of data trading certification.

1) THE SUCCESSFUL CONSTRUCTION OF THE CONSORTIUM BLOCKCHAIN

a: GENESIS BLOCK

The genesis block of the PoA Chain is constructed. The content to be set in the genesis block mainly includes “name”, “params”, “genesis” and “accounts”. In the genesis block,

the approved verification and bookkeeping nodes are preset, that is, the address of DTC (Authority node). And the nodes that are authorized to join the PoA Chain are preset, that is, the address of DS (User node) and the address of DP (User node). In the genesis block, the Authority node and User node are preset, as shown in Figure 7.

Algorithm 3 Successful Trading and Payment Settlement

Input: EOA, DS address, DTSC status, DP status, DP address, DTC address

- 1: EOA is the set of Ethereum addresses with deposits.
- 2: DTC.EOA \leftarrow DTC address
- 3: DS.EOA \leftarrow DS address
- 4: DPE.OA \leftarrow DP address
- 5: **if** msg.sender = DPE.OA **then**
- 6: Change DP status to Get Data Resource confirmed by DS.
- 7: Create a notification with the result of data resource provided by DS.
- 8: Wait for confirmation from DP.
- 9: **if** DP confirmation result = true **then**
- 10: **foreach** address \in EOA **do**
- 11: Refund the deposit.
- 12: **end**
- 13: Pay DTC.EOA and DS.EOA their share.
- 14: **end**
- 15: **else**
- 16: Execute the Settle Dispute.
- 17: Turn to Algorithm 3.
- 18: **end**
- 19: **end**
- 20: **else**
- 21: Revert DTSC status and show an error.
- 22: **end**

b: POA CHAIN

The PoA Chain is built based on Parity wallet. The three preset Authority nodes correspond to three Parity wallets respectively. When the Authority node is started, the corresponding Parity wallet is also started. As more Authority nodes are started, they will automatically connect to communicate, so that the PoA-based consortium blockchain is successfully constructed, as shown in **Figure 8**.

2) THE SUCCESSFUL DEPLOYMENT OF THE SMART CONTRACTS

a: DTSC COMPILATION AND DEPLOYMENT

After DTSC is compiled successfully, it will be deployed on the PoA Chain by DTC. After the DTSC is successfully deployed, it can receive the deposits from DS and DP, conduct data trading, resolve trading disputes, and perform the payment settlement of trading. **Figure 9** shows the DTSC that has been successfully compiled and deployed.

In the DTSC, Algorithm 2 is an algorithm used by DP to purchase data resources. Algorithm 3 is an algorithm for the DTSC to automatically perform payment and settlement after the DP successfully purchases the data resources. Algorithm 4 is an algorithm used by DTC to resolve data trading disputes.

Algorithm 4 Dispute Resolution

Input: EOA, DTC address, DP address, DS address, DTSC status, DP status

- 1: DTC.EOA \leftarrow DTC address
- 2: DS.EOA \leftarrow DS address
- 3: DPE.OA \leftarrow DP address
- 4: EOA is a set of Ethereum addresses with deposits.
- 5: **if** msg.sender = DPE.OA & DP confirmation result = false **then**
- 6: Create a notification to DTC.
- 7: DTC uses CDSC to arbitrate data resource.
- 8: res \leftarrow DTC arbitration result.
- 9: **if** res = true **then**
- 10: Create a notification that DP is right.
- 11: Refund DP.
- 12: **end**
- 13: **else**
- 14: Create a notification that DP is wrong.
- 15: **foreach** address \in EOA **do**
- 16: Refund the deposit.
- 17: **end**
- 18: Pay DTC.EOA and DS.EOA their share.
- 19: **end**
- 20: Change DP status to Trading Completed.
- 21: **end**
- 22: **else**
- 23: Revert DTSC Status and show an error.
- 24: **end**

```

"validators": {
  "list": [
    "0x62b1e832130672CfeD8CAb10c0919993a505b311",
    "0x000717b7e8ce2cC43E55822339a343CD5F6F84e4",
    "0x00d806880A8CDa0275C378d98D24712578881113"
  ]
}

```

(a) DTC addresses

```

"accounts": {
  "0x0022505aC189405947e027ddd6c5C6CCf9ED8972": {
    "balance": ""
  },
  "0x00bFAA40777963f2324e1f5b3e0f427FEa8f4ac7": {
    "balance": ""
  }
}

```

(b) DS address and DP address

FIGURE 7. The addresses of Authority nodes and User nodes.*b: CDSC COMPILATION AND DEPLOYMENT*

CDSC is deployed on the PoA Chain by DTC and will run permanently. When a contract is written to the blockchain, a timestamp is generated. This timestamp is used as the date of signing the contract. The trading participating entities can use the Ethereum signature function *sign()* to sign the contract. All nodes in the blockchain network can use the function *ecrecover()* to verify the digital signature. **Figure 10** shows the CDSC that has been successfully compiled and deployed.

```
0/25 peers 920 bytes chain 2 KiB db 0 bytes queue 448 bytes sync RPC: 0 conn, 0 req/s, 0 μs
1/25 peers 920 bytes chain 2 KiB db 0 bytes queue 1 KiB sync RPC: 0 conn, 0 req/s, 229705 μs
2/25 peers 920 bytes chain 2 KiB db 0 bytes queue 2 KiB sync RPC: 0 conn, 0 req/s, 229705 μs
```

FIGURE 8. The PoA Chain based on Parity wallet.

```
Replacing 'DTSC'
-----
> transaction hash: 0x73b558647ca3984c914dc74515a23d3a4fbf167566163011dc97b39fb7064b8d
> Blocks: 0 Seconds: 0
> contract address: 0x795655Efd20AF65FB077184e89E18b9CB1D1D5E0
> block number: 3
> block timestamp: 1583897695
> account: 0x62b1e832130672CfeD8CAB10c0919993a505b311
> balance: 99.97053074
> gas used: 1266947
> gas price: 20 gwei
> value sent: 0 ETH
> total cost: 0.02533894 ETH

> Saving migration to chain.
> Saving artifacts
-----
> Total cost: 0.02533894 ETH
```

FIGURE 9. The DTSC compiled and deployed.

```
Deploying 'CDSC'
-----
> transaction hash: 0xa4b2e95b58bcddf36de98bc04eb75388c4ca46f46f408dbc0e7bd62f478deb8
> Blocks: 0 Seconds: 0
> contract address: 0x795655Efd20AF65FB077184e89E18b9CB1D1D5E0
> block number: 3
> block timestamp: 1589347028
> account: 0x62b1e832130672CfeD8CAB10c0919993a505b311
> balance: 99.96605664
> gas used: 1490652
> gas price: 20 gwei
> value sent: 0 ETH
> total cost: 0.02981304 ETH

> Saving migration to chain.
> Saving artifacts
-----
> Total cost: 0.02981304 ETH
```

FIGURE 10. The CDSC compiled and deployed.

The CDSC is deployed on the blockchain, which means that the detailed data trading content has been completely preserved and cannot be tampered with. When a dispute occurs in data trading, trading participants can access the CDSC to present the complete content of the trading, so that the data trading dispute is effectively resolved. In short, the certification data is stored on the blockchain, which not only ensures data security, but also data access and query are very convenient.

V. LIMITATIONS AND DISCUSSIONS

The limitations, security features and performance analysis of the solution will be discussed in this section.

A. LIMITATIONS

The limitations of the solution are summarized as follows.

(1) Trust model: DTC has obtained the data trading license issued by the government, and thus has the qualification of

data trading platform. DTC has been endorsed by the government, thus gaining the trust of DS and DP. The government has established a management and punishment system to limit and control DTC to make it completely credible.

(2) Restrictions on participating entities: The entity participating in the data trading must be an Ethereum user, and the entity knows how to perform the functions of the smart contract and how to access and query the content of the deployed smart contract.

(3) The cost of executing smart contract functions: If an entity needs to execute the functions of a smart contract, it needs to pay. In the real world, only through the use of real money to buy, trading participants can get the gas they need. This means that the trading participating entities need to consume real money to obtain gas in order to perform the corresponding functions in the smart contract, so that they can participate in data trading. Smart contract functions can only be performed by consuming

TABLE 3. Data storage security comparison.

Data storage method	Data storage characteristics					Storage security level
	General storage	Data robustness	Distributed storage	Traceability	Transparence	
Traditional database	Yes	Low	No	No	No	Low
Cloud storage	No	Medium	Yes	No	No	Medium
Blockchain	No	High	Yes	Yes	Yes	High

gas, which slightly increases the cost of entities in data trading.

(4) Selection of verification and bookkeeping nodes: Since the PoA is a reputation-based consensus algorithm, its characteristic is that the block validators do not rely on mortgaged cryptocurrencies, but on their reputation. Therefore, the trusted entity is selected as the verification and bookkeeping node. In the process of setting up the PoA Chain, DTC is considered as a trusted entity with high reputation, so it is selected as a verification and bookkeeping node, that is, a block validator. There can be multiple DTCs with high reputation as verification and bookkeeping nodes, and they can form a consortium blockchain network with multiple DSs and DPs. The motivation for DTC to join and remain honest is that, on the one hand, it can profit from data trading, and on the other hand, it must also consider its reputation. Once fraudulent behaviors lead to poor reputation, DTC will inevitably disappear in the data trading market.

(5) The PoA-based certification model: Although the PoA-based certification model has the advantages of fast transaction verification and zero-cost transaction verification, participating entities can only be added to the consortium blockchain with permission, and cannot participate and withdraw freely. In addition, as long as there is data trading, data trading certification is required. At this time, it is necessary to constantly create and deploy new certification data smart contracts, which will increase the programming workload.

B. SECURITY FEATURES

The key security features of the solution are summarized below.

(1) Auditability: Blockchain is a distributed ledger, it will record all operations and initiators in the process of data trading, and has the characteristics of non-tampering. Therefore, once the data trading is completed, DS cannot deny the operation initiated by it, nor can DP deny the operation initiated by it. Through digital signatures, denials or illegal behaviors can be audited, analyzed, tracked and traced.

(2) Accountability: The PoA-based consortium blockchain is supervisable, which means that the entities involved in the trading can also be supervisable. Therefore, when a dispute occurs in data trading, DS cannot deny his/her behavior, nor can DP deny his/her behavior. For the fraudulent behavior of

the entity, accountability can be pursued by means of account closure or fines.

(3) Integrity: DTSC and CDSC have been deployed on the blockchain. The feature of smart contracts is that they cannot be changed once deployed. The feature of the blockchain is that all information exchanged between nodes cannot be tampered with. This means that all transaction information and certification data will not be tampered with. This not only guarantees the security of transaction information and certification data, but also ensures the integrity of transaction information and certification data.

C. PERFORMANCE ANALYSIS

The performance of the solution will be compared and analyzed from three aspects: storage security, computing load and information transparency. Through comparative analysis, the consortium blockchain and PoA are selected for data trading certification, because among the existing feasible methods, the consortium blockchain and POA are the best choice.

1) SECURITY ANALYSIS OF CERTIFICATION DATA STORAGE

The qualitative evaluation method is used to evaluate the data storage security of traditional database (i.e. centralized database), cloud storage (i.e. distributed database) and blockchain (i.e. decentralized distributed database), as shown in **Table 3**. The database which is easy to tamper with data and easy to single point failure is defined as the database with low security level. The database which is easy to tamper with data but not easy to single point failure is defined as the database with medium security level. The database whose data can not be tampered with and is not easy to single point failure is defined as the database with high security level.

The consortium blockchain is used for DTCM. When the ledger of a node on the chain is damaged, it will not cause data loss, because the distributed ledger of other nodes can be synchronized to the node to recover the data. All transaction data that has been verified and written into the blockchain cannot be tampered with or destroyed. Therefore, the transaction data can be traced back to the genesis block. This is conducive to regulatory agencies to supervise the data trading certification.

TABLE 4. Calculation load comparison.

Feature	PoW-based certification model	PoS-based certification model	PoA-based certification model
Transaction verification time	10 min	< 10 min	<< 10 min
Amount of computation	Complex operation	Simple operation	No calculations required
Transaction verification cost	Higher	High	0

TABLE 5. Information transparency comparison.

Feature	Traditional certification model	Public blockchain based certification model	Consortium blockchain based certification model
Whether it can be supervised	Yes	No	Yes
Whether the user is anonymous	No	Yes	Yes
Whether the user has the right to trace transaction records	No	Yes	Yes
Whether the user has the right to query all account amounts	No	Yes	Yes

2) ANALYSIS OF CALCULATION AMOUNT OF BLOCK GENERATION

The PoA consensus mechanism is used for DTCM. It is different from the traditional PoW consensus mechanism. PoW requires a large number of computational mathematical problems to obtain bookkeeping rights. It is also different from the method that PoS requires a large amount of accumulated equity to obtain bookkeeping rights. Transactions and blocks based on the PoA algorithm are verified by approved accounts (called validators). This verification process is automated and encourages approved and trusted validators to maintain the security and consistency of the network. Compared with PoW and PoS, PoA can provide a faster transaction rate. A more prominent advantage is that transaction verification based on the PoA consensus mechanism does not require gas consumption. This helps reduce transaction costs, as shown in **Table 4**.

3) ANALYSIS OF INFORMATION TRANSPARENCY

Blockchain-based DTCM can solve the mutual trust problem between trading participants in an insecure environment. The blockchain uses asymmetric encryption to protect the security of data trading. The anonymity of the blockchain effectively prevents the disclosure of the identity information of the entities involved in the trading. Compared with other security technologies, the biggest advantage of blockchain is that it can use the data trading certification method in unknown environments.

In the DTCM, all users can trace any single transaction that has occurred, and can query the payment status and balance of any anonymous user who generated the transaction. Because the DTCM is based on the consortium blockchain, the Authority node (DTC) of the verification transaction can control the joining and exit of the User nodes (DS and DP) participating in the trading to a certain extent. Compared with the unregulated public blockchain, the consortium blockchain is better regulated, as shown in **Table 5**.

VI. CONCLUSION

Now, a data trading certification blockchain solution based on consortium blockchain and smart contracts has been presented. At first, the PoA Chain is constructed through the PoA algorithm and Parity wallet. It has the features of fast transaction verification and low transaction cost. Next, the DTSC and CDSC are written through the Solidity language. They have both concise and effective characteristics. Furthermore, the DTSC and CDSC are deployed on the consortium blockchain through the React framework of Truffle. This makes the process of compiling and deploying the DTSC and CDSC be simple and effective. Moreover, the DTCM is created through the PoA Chain and CDSC. Its special advantages include the uniqueness of the data trading path, the progressiveness of the CDSC state, and the scalability of the transaction verification algorithm. Finally, the successful construction of the consortium blockchain and the successful compilation and deployment of smart contracts are demonstrated through the experiments. At the same time, the

comparative analysis shows that the blockchain solution still has the advantages of lower cost, smaller calculation amount and higher supervision.

REFERENCES

- [1] W. Xiong and L. Xiong, "Smart contract based data trading mode using blockchain and machine learning," *IEEE Access*, vol. 7, pp. 102331–102344, 2019.
- [2] P. Treleaven, R. Gendal Brown, and D. Yang, "Blockchain technology in finance," *Computer*, vol. 50, no. 9, pp. 14–17, 2017.
- [3] A. Azaria, A. Ekblaw, R. Vieira, and A. Lippman, "MedRec: Using blockchain for medical data access and permission management," in *Proc. 2nd Int. Conf. Open Big Data (OBD)*, Aug. 2016, pp. 25–30.
- [4] W. Meng, E. W. Tischhauser, Q. Wang, Y. Wang, and J. Han, "When intrusion detection meets blockchain technology: A review," *IEEE Access*, vol. 6, pp. 10179–10188, 2018.
- [5] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain & Internet of Things," in *Proc. Int. Conf. Service Syst. Service Manage.*, Jun. 2017, pp. 1–6.
- [6] K. Toyoda, P. T. Mathiopoulos, I. Sasase, and T. Ohtsuki, "A novel blockchain-based product ownership management system (POMS) for anti-counterfeits in the post supply chain," *IEEE Access*, vol. 5, pp. 17465–17477, 2017.
- [7] S. Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash System*. Accessed: Oct. 6, 2020. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [8] S. Singh and N. Singh, "Blockchain: Future of financial and cyber security," in *Proc. 2nd Int. Conf. Contemp. Comput. Informat. (IC3I)*, Dec. 2016, pp. 463–467.
- [9] W. She and Y. Xiaoyu, "Transaction certification model of distributed energy based on consortium blockchain," *J. Univ. Sci. Technol. China*, vol. 48, no. 4, pp. 307–313, 2018.
- [10] N. Szabo, *Formalizing and Securing Relationships on Public Networks*. Accessed: Oct. 6, 2020. [Online]. Available: <http://www.firstmonday.org/ojs/index.php/fm/article/view/548/469>
- [11] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.
- [12] W. Dai, C. Dai, K.-K.-R. Choo, C. Cui, D. Zou, and H. Jin, "SDTE: A secure blockchain-based data trading ecosystem," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 725–737, 2020.
- [13] Y. He, H. Zhu, C. Wang, K. Xiao, and Y. Zhou, "An accountable data trading platform based on blockchain," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Apr./May 2019, pp. 1–6.
- [14] T. Jung, X.-Y. Li, W. Huang, Z. Qiao, J. Qian, L. Chen, J. Han, and J. Hou, "AccountTrade: Accountability against dishonest big data buyers and sellers," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 1, pp. 223–234, Jan. 2019.
- [15] E. Kokoris-Kogias, E. C. Alp, S. D. Siby, N. Gailly, and L. Gasser, "Verifiable management of private data under Byzantine failures," *Cryptol. ePrint Arch.*, Tech. Rep. 2018/209, 2018.
- [16] Y. Zhao, Y. Yu, Y. Li, G. Han, and X. Du, "Machine learning based privacy-preserving fair data trading in big data market," *Inf. Sci.*, vol. 478, pp. 449–460, Apr. 2019.
- [17] W. Gao, W. Yu, F. Liang, W. G. Hatcher, and C. Lu, "Privacy-preserving auction for big data trading using homomorphic encryption," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 2, pp. 776–791, Apr. 2020.
- [18] M. Sabounchi, J. Wei, and R. Roche, "Blockchain-enabled peer-to-peer data trading mechanism," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul./Aug. 2018, pp. 1410–1416.
- [19] J. Iyilade and J. Vassileva, "A framework for privacy-aware user data trading," in *Proc. Int. Conf. User Modeling, Adaptation, Personalization*, 2013, pp. 310–317.
- [20] C. Chen, J. Wu, H. Lin, W. Chen, and Z. Zheng, "A secure and efficient blockchain-based data trading approach for Internet of vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9110–9121, Sep. 2019.
- [21] K. Liu, W. Chen, Z. Zheng, Z. Li, and W. Liang, "A novel debt-credit mechanism for blockchain-based data-trading in Internet of vehicles," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 9098–9111, Oct. 2019.
- [22] S. Bajoudah, C. Dong, and P. Missier, "Toward a decentralized, trust-less marketplace for brokered IoT data trading using blockchain," in *Proc. IEEE Int. Conf. Blockchain (Blockchain)*, Jul. 2019, pp. 339–346.
- [23] Z. Guan, X. Shao, and Z. Wan, "Secure fair and efficient data trading without third party using blockchain," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (GreenCom) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jul./Aug. 2018, pp. 1395–1401.
- [24] D. An, Q. Yang, W. Yu, D. Li, Y. Zhang, and W. Zhao, "Towards truthful auction for big data trading," in *Proc. IEEE 36th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Dec. 2017, pp. 1–7.
- [25] X. Cao, Y. Chen, and K. J. R. Liu, "Data trading with multiple owners, collectors, and users: An iterative auction mechanism," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 3, no. 2, pp. 268–281, Jun. 2017.
- [26] Parity Ethereum, *The Fastest and Most Advanced Ethereum Client*. Accessed: Oct. 6, 2020. [Online]. Available: <https://www.parity.io/>
- [27] Solidity, *An Object-Oriented, High-Level Language for Implementing Smart Contracts*. Accessed: Oct. 6, 2020. [Online]. Available: <https://solidity.readthedocs.io/en/latest/>
- [28] Remix IDE, *Welcome to Remix Documentation*. Accessed: Oct. 6, 2020. [Online]. Available: <https://remix.readthedocs.io/en/latest/>
- [29] Truffle, *Truffle is the Most Popular Development Framework for Ethereum*. Accessed: Oct. 6, 2020. [Online]. Available: <https://www.trufflesuite.com/boxes/react/>



WEI XIONG is currently pursuing the Ph.D. degree with the Department of Information Management, School of Management, Shanghai University. His research interests include data trading, smart contract, blockchain, and information systems.



LI XIONG is currently a full-time Professor with the Department of Information Management, School of Management, Shanghai University. Her research interests include data trading, information systems, cross-border e-commerce, and collaborative innovation.

...