EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

A blockchain-based spatial data trading framework

Hui Liu[1], WeiPeng Tai[2], Yaofei Wang[1] and Shenling Wang[1*]

*Correspondence:
slwang@bnu.edu.cn

[1] School of Artificial Intelligence, Beijing Normal University, Beijing, China
[2] Anhui Gongda information technologyco. ltd., Maanshan, China

Abstract
With the increasing utilization of space related data, the demand for spatial big data sharing and trading is growing rapidly, which promotes the emergence of spatial data market. However, in conventional data markets, both data buyers and data sellers have to use a centralized trading platform which might be dishonest. Blockchain is a decentralized distributed data storage technology, which uses the traceability and unforgeability to confirm and record each transaction, and can solve partial disadvantages of the centralized data market; unfortunately, it also introduces the problems of security and privacy. To address this issue, in this paper, we propose a blockchain-based spatial data trading framework with trusted execution environment to provide a trusted decentralized platform, including data storage, data query, data pricing, data reputation and security computing. Based on this framework, we use an auction pricing mechanism to ensure data trading authenticity and efficiency. At last, a spatial data trading framework was implemented and its effectiveness and security were verified.

Keywords: Spatial data, Data trading, Data pricing, Blockchain

1 Introduction

With the increasing popularity of all kinds of sensors and the wide application of mobile positioning technology, the amount of spatial data is growing rapidly. Spatial data have become a key asset in current economy; the utilization of spatial data can bring huge economic benefits or help users make better decision. In order to ensure the normal circulation and use of data, many newly-established institutions about spatial data sharing and trading have emerged in recent years. In addition to the traditional way of data circulation, there is also a big data trading market, which facilitates data transactions by matching data demand with data sources.

In a conventional data market [1–4], data seller sends the data to a centralized trading platform. Although the data trading platform accelerates the circulation of data, there are still many problems at present:

1. Malicious data buyers may resell the seller's data by caching the source data at a lower price than the data seller does after obtaining the data, thus damaging the seller's interests.

2. The centralized trading platform may be unreliable who may cache and resell the source data without the permission of the data seller, thus undermining the interests of both parties.
3. The centralized trading platform with low fault tolerance may be relatively easily paralyzed owing to attacker's invasion, which will affect the data transaction between the data buyer and the data seller, and even cause the loss of key data.

In addition, the centralized trading platform lacks effective information communication channel between data buyer and data seller, which leads to low efficiency of data transaction [5].

In order to avoid the disadvantages in centralized data market, such as data security and privacy, data copyright protection and data sharing performance bottlenecks, decentralized data market was born. Decentralized data market architecture can improve the transparency and credibility of data transactions by getting rid of both single point of failure and performance bottleneck. Nevertheless, its system design and security assurance will be more difficult than that of centralized data market due to the lack of centralized management. For example, double payment has always been the difficulty of distributed system [6, 7].

Blockchain, first introduced by Nakamoto in 2008 [8], is a distributed public ledger. It maintains a continuously growing list of ordered records called blocks. Each block in the chain is linked to the previous block through a cryptographic hash. All nodes in the network share the same copy of digital ledger. Information stored in the blockchain is open to everyone, making the actions of nodes transparent.

The introduction of blockchain into the data market system will enable data sellers to enter into transactions with the data buyers directly without relying on any third party, so that sellers can strengthen the ownership of data and ensure the openness and transparency during the transaction process, which can solve partial disadvantages of the centralized data market. However, it also introduces the problems of data storage and privacy protection [9–12], as explained below:

1. Data storage. The blockchain system is only suited for storing a small amount of transaction data, not for a large number of data files. Thus, data files should not be stored directly in the blockchain. Furthermore, the data file stored module cannot disclose identity information, as well as the buyer and the seller cannot obtain the real identity of the other party.
2. Privacy protection. It is necessary to design privacy protection scheme in data transaction process, so as to protect the user's identity from being disclosed and ensure the security of transaction.

To address this issue, we propose a blockchain-based spatial data trading framework which takes advantages of the blockchain to build a decentralization platform for data trading, including data storage, data query, data pricing, data reputation and security computing.

To tackle the first challenge, we use IPFS (InterPlanetary file system) to store data files, which is a distributed file storage system. When a file is uploaded to IPFS, it is available to all peers in the IPFS network. The user will receive a hash index, which will allow the user to retrieve the file later. This index will replace the data stored in the smart contract, lowering the burden of the entire system.

To address the second challenge, we adopt smart contract to realize the data transaction, furthermore the combination of cryptography technology and Ethereum account mechanism is to solve the privacy protection and data security problems in the transaction.

In our design, data buyers and data sellers conduct transactions directly on the blockchain, which avoids risks caused by centralized trading platforms. All payment records and reviews generated by data buyers are faithfully recorded in the blockchain through consensus protocol and can be protected securely with trusted execution environment (TEE). The rest of this paper is organized as follows: In Sect. 2, we introduce the background and related work. In Sect. 3, we design the framework of spatial data trading based on blockchain. In Sects. 4 and 5, we verify effectiveness and security of a spatial data trading system. At last, we summarize our results and make some concluding remarks.

2 Background knowledge

Due to its decentralized immutable and traceable characteristics, blockchain technology is used in data trading market in recent years, which has attracted great attention of the industry. For example, Shanghai Data Trading Center [13] uses alliance chain to store transaction related information in blockchain nodes to ensure data transaction security, efficiency and credibility.

Wang et al. [13] applied blockchain technology to the data market, which improved the transparency and security of data transactions, but did not take into account the long-term sustainability of data market. Zyskind [14]used blockchain to protect the privacy of personal data, transforming the blockchain into an automatic access control manager to strengthen the ownership of data, which realized data storage and data access control without third parties.. Crowdbc [15] is a crowdsourcing system constructed by blockchain in which a requester's task can be solved by a crowd of workers without relying on any third trusted institution. The author focuses on and processing of image data in transaction. Baig [16] constructed a data market based on blockchain and introduced a trusted intermediary in the transaction between the buyer and the seller. Although this makes the transaction between the buyer and the seller easier, it also reduces the security of the system a lot. Dai proposed SDTE [17], a blockchain-based data trading ecosystem. In SDTE, buyers of data cannot directly access the original data they purchased, but can only obtain the analysis results of the data, which is generated from Intel SGX (Software Guard Extensions). However, if there are too many malicious data sellers in a transaction, honest data sellers will not be able to get reasonable compensation.

Spatial data in this paper include [18]: remote sensing, mapping and other raw data, such as low-resolution satellite images, medium resolution satellite images, high-resolution satellite images, sub-metre high-resolution satellite images, aerial photogrammetry data, series scale vector data, terrain data and other types of original

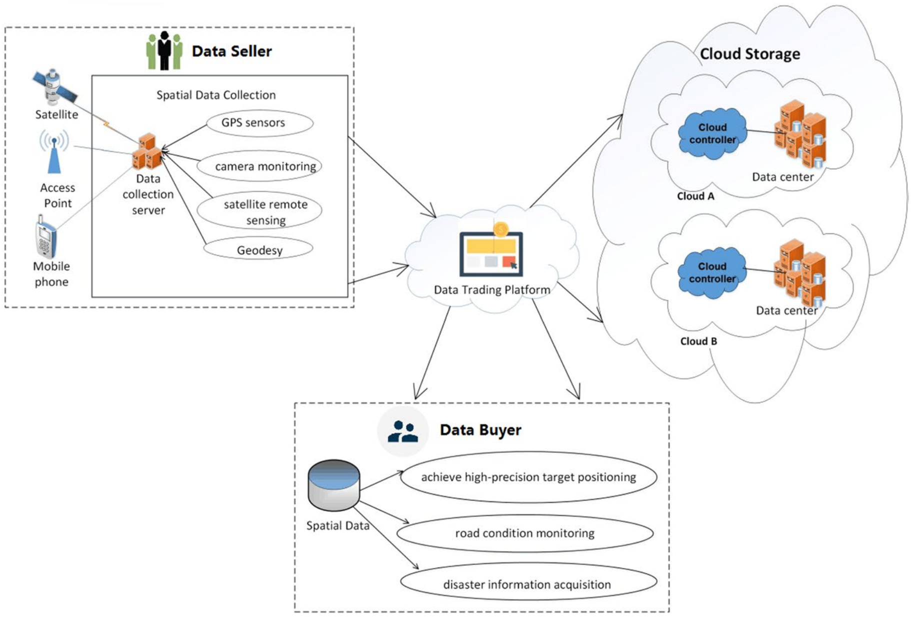

Fig. 1 Spatial data trading system

spatial information data. As shown in Fig. 1, after obtaining spatial data through various devices such as satellite, access point and mobile phones, the data owners upload them to the cloud storage platform through network and then sell them to the data buyers through data trading platform who may use the data to achieve high-precision target positioning, make road condition monitoring or realize disaster information acquisition.

The spatial data trading system based on blockchain designed in this paper consists of three main components: a smart contract in Ethereum blockchain, trading application system for users and a point-to-point data transmission network. As shown in Fig. 2, after collecting some data for sale, the data seller registers and adds data digest information in the smart contract while data itself is stored in IPFS using data storage module. When needing some data to calculate a task, a data buyer will use the data query module and issue an order containing data requirements through the smart contract. Then, the system queries for qualified data in the way of security calculation. Subsequently, using auction method, the data pricing module will determine the price of the data. As the payment has been completed, the system delivers the data to the buyer securely. With the data delivered and used, data reputation module is used to evaluate the data quality and the reputation of the seller.

2.1 Data query

Data buyers can actively locate spatial data resources through different query methods - keyword query, region query, nearest neighbor query, anti nearest neighbor query, etc.- and view the details of spatial data through data digest, combining with thumbnails and metadata, so as to specify their own requirements; In addition, data buyers can clarify their specific data requirements into a logical expression or a

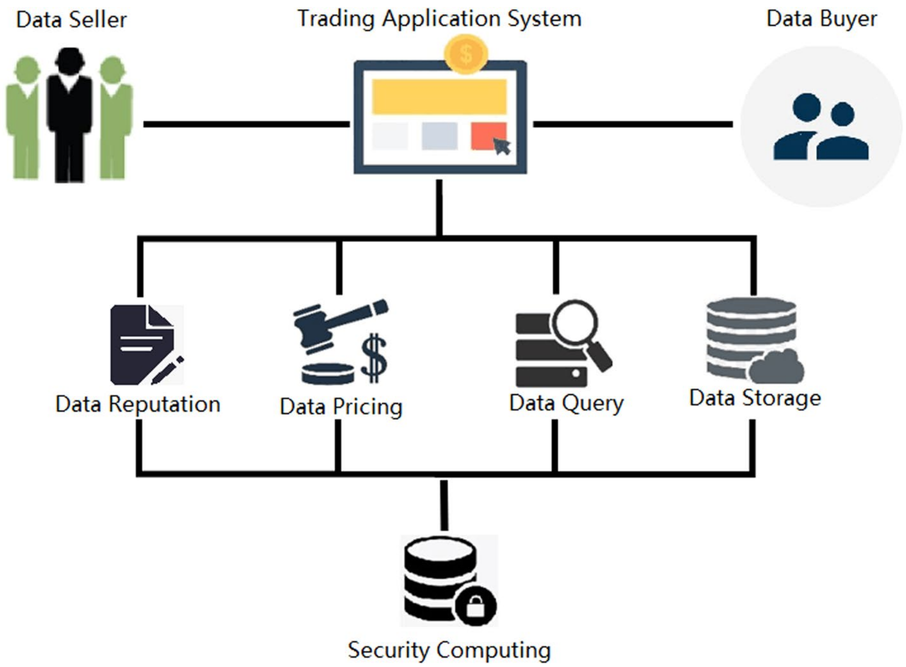

Fig. 2 Data trading platform

mathematical function and store them in the blockchain for the seller to query [19, 20]. Thus, according to the buyer's data requirements, the seller can also query the data to confirm whether the data is consistent with the buyer's demands.

Although the query conditions of spatial data are relatively simple, it is inappropriate to upload the buyer's data query requirements directly to the blockchain, otherwise, it is easy for attackers to infer the buyer's data requirements and obtain the privacy of both the buyer and the seller, which obviously increases the risk of privacy disclosure. Therefore, in order to protect the privacy of system users, we should query the data while obfuscating the query conditions. The most common solution is function encryption, by which data buyer with decryption key, when the query conditions are encrypted, can obtain the data query value of cipher text data, instead of any information about original data.

2.2 Data storage

Most public blockchain systems have restrictions on the number and space of transactions in the block due to the limited size of the block (such as bitcoin system) or the "Gas" consumed in the block (such as Ethereum system). Therefore, it is infeasible to store massive data directly on the blockchain in the spatial data trading system. Recently a distributed data storage InterPlanetary file system (IPFS) is introduced for storing the shared data in various other domains like health care, cloud computing, IoT, etc.

IPFS [21–23] is a peer-to-peer, content-addressable, distributed file storage system, using a swarm of computers connected. When a file is uploaded to IPFS, it is available to all peers in the IPFS network. The uploaded file is divided into chunks, which are assigned as a unique cryptographic hash. Thus, data added to IPFS are addressed by using this unique cryptographic hash, which makes it content addressable. It uses distributed hash tables (DHTs) to locate files. In this case, traded data files are stored in IPFS while the hashes correlated with IPFS are stored in blocks, thus reducing the huge cost of storage space. In summary, IPFS provides high throughput with secure storage model that supports concurrent access of data with high storage capacity.

The cloud storage service of the spatial big data trading platform is mainly to establish a storage space station, which uses computer, Internet, Internet of things and other technologies to carry out daily storage management on the products and services traded by the platform and various information generated during the data trading, and can quickly and accurately complete the statistical summary of product and service transaction information. Taking the merits of cloud storage service, such as rapid retrieval, high reliability, large amount of storage and good confidentiality, a great deal of data can be safely saved [24, 25].

Although it is impractical to store the original data in the blockchain, the data digest correlated with specific data can be stored in the blockchain, while the original data can be stored in IPFS. When the data are successfully stored in IPFS, the user will receive a hash index, which helps to retrieve the file later. This index will replace the data stored in the smart contract, reducing storage bottlenecks of the entire system.

We represent spatial data as T1= {hashid, time, space, other attributes }, where hashid is the hash value generated by hashing spatial data, i.e. hashid = hash (time, space, other attributes); time with the time attribute indicates the time when the data is generated; space with a spatial attribute is usually expressed in the form of longitude and latitude coordinates. In order to facilitate the query and storage of spatial data, the space here is converted by GeoHash(Geographical Hash) algorithm, which can hash the longitude and latitude information. Other attributes symbolize other traits of spatial data besides temporal and spatial attributes, including signature information.

2.3 Data pricing

Undoubtedly, data, as a commodity, has some unique properties [26, 27], contributing to data pricing needs considering more issues. First, the marginal cost of data is extremely low, among which the marginal cost refer to the cost of copying a product, so that once the data buyer obtains the data from data seller, he may resell the data. Second, the value of data is not only related to the amount of data, but also to its content and quality. For example, a pile of face images is of little value to a person who needs remote sensing images. Third, the quantification of data value is difficult to estimate. Moreover, valuations vary greatly among different users.

Many studies [13] have found that most data buyers only need some statistical results such as calculating the average value of data sets, or extracted features by training data for machine learning, rather than the data itself. Consequently, data buyer merely purchases the right to use the data rather than the ownership of the data. In this way, data is isolated from the data consumer.

Although the value of the data itself is difficult to quantify, it is essential to evaluate the value of data, for buyers often need data from multiple sellers whose valuation of data may vary greatly. When calculating the value of data, Shapley value, which is a solution to distribute benefits and costs fairly to multiple participants, can be used to calculate the contribution of each data [28]. Since the computation complexity of Shapley value enlarges exponentially with the increase of data quantity, approximate algorithms or distributed algorithms are often employed in practical application.

In this paper, we use auction theory to design pricing mechanism in spatial data trading system. Each data seller has a private valuation including the risk assessment of privacy leakage for their data; comparatively, for each data buyer, he also has a valuation for the data he will buy. Because of unsymmetrical information of their valuation, both buyer and seller may dishonestly report their valuation of data. To solve the problem, we use auction theory to design incentive compatible mechanism so that two parties can get the highest return when reporting their real valuation supplemented by Shapley value calculation algorithm, which makes it much easier to design pricing mechanism.

2.4 Data reputation

It is hard for data buyers to estimate the quality of seller's data instantly before they use it. Due to the particularity of online data trading, sellers usually have more obvious information advantages than that of buyers. In the case of asymmetric information, malicious sellers cannot be prevented from cashing out through lower quotations and selling low-quality data to seek improper interests.

To solve the problem, in this paper, we introduce a reputation mechanism to ensure the long-term sustainability of data trading. As a significant signal of data quality, reputation is an important intangible asset of data seller. When the quality of goods such as data cannot be directly observed, reputation plays a decisive role and can effectively reduce the uncertainty of data trading. The reputation of a data seller is computed as reputation score by aggregating the subjective feedback provided by data buyers after they acquire and use the data, which accurately represents the data quality of a seller.

2.5 Security computing

Generally, information in the blockchain system is open to all users, and the execution of all transactions or scripts is transparent because of the openness and transparency of the Blockchain technology. Besides, the computing power of the blockchain smart contract is relatively weak due to the limitation of block size, therefore, it is neither safe nor feasible to run the trading transactions issued by the buyer directly in the smart contract, for the computing cost is too high. Similarly, such problems also exist in the process of data query and data pricing. Furthermore, public chain nodes in Bitcoin, Ethereum and other blockchain are not trusted with each other, which makes the privacy protection work more challenging, since they are also completely open and transparent.

Trusted hardware, such as trusted execution environment (TEE) [29–31], is a common method to implement secure computing. TEE ensures that the code and data loaded in it are protected in terms of confidentiality and integrity. The data seller will send his data to the TEE equipment of the secure buyer who has some TEE hardware, thus the calculation task will be executed in the TEE and the result will be returned to the buyer in a safe way. Hardware isolation in the TEE protects data and computing service from applications running on the operating system, while trusted applications running in TEE can access the full functionality of the device's main processor and memory. The typical hardware technologies supporting TEE are ARM TrustZone and Intel SGX(Software Guard Extensions).

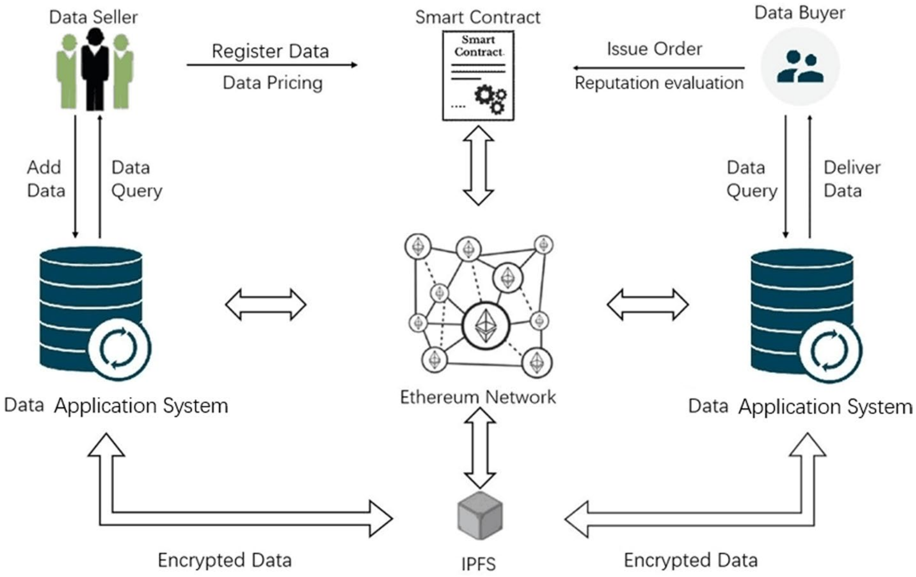

Fig. 3 Data trading framework based on blockchain

Intel Software Guard Extensions (SGX) [32] provides a widely used TEE implementation for general-purpose computation, which is known as enclaves in SGX. Code running inside an enclave has a protected address space. When data in an enclave moves off the processor to memory, it is transparently encrypted with keys only available to the processor. Thus the operating system, hypervisor and other users cannot access the enclaves memory. In the enclave, the code and data are measured at the startup stage and the measurement is signed into an attestation report based on a hardware-based root of trust. The report can be verified to show the unmodified enclave code logical, by which users can confirm the security of enclave.

In this paper, we use Intel SGX as the exemplary implementation to build a trusted exchange by using trusted execution environment, assisting in the fair payment of the transactions.

3 The proposed framework

We have implemented a spatial data trading system based on Ethereum private chain, including trading application system for users, smart contracts in Ethereum network and data transmission network. Specifically, the application system is written by JavaScript, Ethereum smart contract is written by solidness, and the interface of IPFS is used for data transmission.

We simplify the system into one buyer and multiple sellers to trade. The pricing mechanism adopts the second price auction method in which the transaction data adopts the lowest bid data of the seller, while the data are paid at the second lowest price.

The data transaction process is shown in Fig. 3 First, the data seller registers and adds a new data information in the smart contract while data itself is stored in IPFS which is deployed in trusted execution environment (TEE) and the data buyer issues an order containing data requirements. After matching the data requirements condition of data

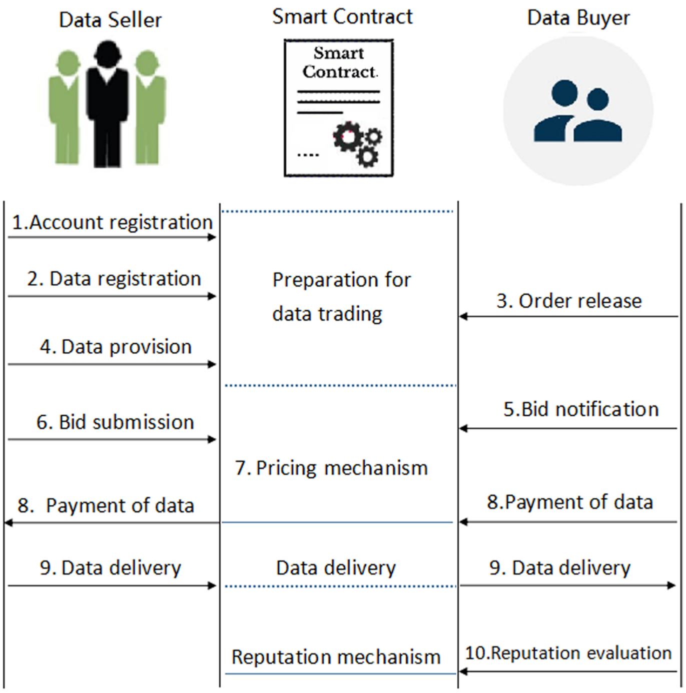

Fig. 4 Workflow of the system

buyer in the order, the seller is allowed to bid in the transaction. Then, the system will run the pricing mechanism to complete the transaction between the buyer and seller and deliver the result data to buyer through the security calculation method. Since the data has been delivered and used, the buyer shall score the seller and its data with the data quality. Based on the scoring results, the seller's reputation value is adjusted to ensure the reliability of data trading.

To describe how the framework works, we introduce the main workflow of the system in Fig. 4.

Step 1 Account registration. To make a deal in the data trading system, the seller should first register an Ethereum account in the corresponding Ethereum account management module. Besides, the seller is required to register an account in the smart contract.

Step 2 Data registration. After collecting some data for sale, the data seller can encrypt and upload the data to IPFS, and register the data in the smart contract account, including the storage address, hash value and registration time of the data, data digest concerning thumbnails and metadata, etc.

Step 3 Order release. If a data buyer wants to buy specific data in the trading system, he will add an order to the smart contract. The order contains the buyer's demand for data, including data type, data query condition, data registration time, data price, data quantity, data seller's reputation value, etc. Data buyers are permitted to set requirements of data registration time and reputation value of seller when adding data orders to purchase data. Afterwards, data sellers who meet the data registration time and reputation value required by buyers are eligible to bid for the order, so as to improve the timeliness and reliability of data trading.

Step 4 Data provision. Data sellers select the data that meet the needs of data buyers and offer a preset price which has been sealed in a hash value as their bid.

Step 5 Bid notification. The data buyer notifies the smart contract to prepare for auction bid of the data seller. The data buyer notifies smart contract to prepare for auction bid of the data seller. For these accepted sellers who have participated in the transaction, they're notified to go to next step to bid; while new sellers are prevented from participating in the transaction.

Step 6 Bid submission. The data seller submits individual bid, which needs to match the previous hash value.

Step 7 Pricing mechanism. We use Vickrey reverse auction mechanism for data pricing, which guarantees authenticity, so each seller has an incentive to make a true evaluation of his personal data.

Step 8 Payment of data. The transaction is achieved at the lowest bid of the seller, who is paid at the second lowest price.

Step 9 Data delivery. The encrypted data of the seller who wins the bid are delivered to the buyer to complete the order.

Step 10 Reputation evaluation. After the data has been delivered and used, the buyer shall score the seller and its data with the data quality. Based on the scoring results, the seller's reputation value is assessed under the guarantee of reputation evaluation mechanism.

Smart contract is a set of digital commitments deployed in a specific address of the blockchain. The data trading platform creates a smart contract for data transactions. All participants in data transactions can register after they agree to the terms and conditions of the contract. The trigger conditions of the contract are specified in the smart contract. At a given time, when the conditions meet the specified trigger conditions, the contract will begin to be executed automatically.

The main algorithm of data resources sold by the seller is shown in Algorithm 1. This procedure is executed from the data seller's perspective. Firstly, the data seller generates data information *Data_info* including storage address of data,hash value of data etc. Then data seller encrypts and uploads the data to IPFS, offering the quantity and bid of data $\{S\_num, S\_bid\}$ to the smart contract. After that, the data seller publishes the bid, which needs to match the previous hash value.

Algorithm 1 : data sold by the seller

```
Input: Address of data seller $S\_addr$; Status of smart contract $SC\_sta$; Information of data $Data\_info$.
Output: Quantity and bid of data available for sale $\{S\_num, S\_bid\}$; Timestamp of sell $S\_ts$.
if $SC\_sta ==$ Runstatus then
    Instantiate objects of contract;
    if $S\_addr ==$ Seller.addr then
        Set $Data\_info$; //including storage address of data,hash value of data
        Add $Data\_info$;
        Publish $Data\_info$;
        $S\_ts =$ block.timestamp;
        Return $\{S\_num, S\_bid\}$ ;
    end if
else
    Return “transaction denied”;
end if
```

The main algorithm of data buyer is shown in Algorithm 2. This procedure is executed in the perspective of data buyer. Firstly, the data buyer receives data information *Data_info* to determine his demand for data (including data type, quantity and bill of data, etc.). Then data buyer sends the quantity and bill of data {*B_num*, *B_bill*} to the smart contract.

Algorithm 2 : data bought by the buyer

```
Input: Address of data buyer $B\_addr$; Status of smart contract $SC\_sta$; Information of data
    $Data\_info$.
Output: Quantity and bill of data available for buyer $\{B\_num, B\_bill\}$; Timestamp of buyer $B\_ts$.
if $SC\_sta ==$ Runstatus then
    Instantiate objects of contract;
    if $B\_addr ==$ Buyer.addr then
        Receive $Data\_info$;
        Add data bill $\{B\_num, B\_bill\}$;
        Publish data bill $\{B\_num, B\_bill\}$;
        $B\_ts =$ block.timestamp;
        Return $\{B\_num, B\_bill\}$;
    end if
else
    Return “transaction denied”;
end if
```

4 System analysis

4.1 Privacy analysis

The identity privacy and data privacy of system participants are protected in the system. Both data buyers and sellers use their own accounts for each transaction, which can reduce the associated information between multiple transactions. In the payment process, both parties are required to provide the account address and related key signature to complete the transaction. After the payment, the buyer obtains data from the IPFS file system according to the address where the data file is stored and the corresponding hash value, without directly contacting with the data seller. All of these allow data buyers and sellers to hide their identities from each other in the process of data transactions, eliminating the risk of identity privacy disclosure. Therefore, identity privacy is based on the anonymity feature of blockchain, while data privacy is fully protected with data

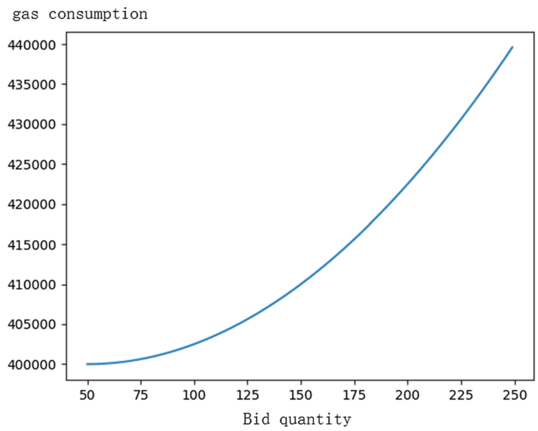

Fig. 5 Relationship between gas consumption and bid quantity

encryption and secure computing mechanism thus attackers unable to launch an attack by seeking any additional information of data during the trading.

4.2 Pricing analysis

Since data pricing is implemented on the Ethereum blockchain through smart contracts, its computational cost should be highlighted due to the limited computing power of the smart contract. The cost of data pricing based on Vickrey reverse auction is evaluated with the gas consumption in Ethereum blockchain Fig. 5 shows the relationship between the gas consumption of data pricing and the number of bids for data orders. We find that when a large number of data sellers bid for the same data order, the data pricing cost will be evidently high, which will become limitation of application in the trading system based on Ethereum blockchain. Therefore, in the follow-up study, reducing the cost of data pricing is one of the future research directions.

5 Experimental methods

We have implemented a spatial data trading system based on Ethereum private chain, smart contracts in Ethereum network and data transmission network. All the tests of this system are completed on PC with 18 vcores(3 GHz Intel Xeonr Platinum 8124M). The online transaction module of data transaction system is basically implemented in Python language.

This paper mainly tests the system throughput, storage space occupation and fault tolerance. Among them, for the system throughput, we test the amount of data that can be processed per unit time under the condition of number of nodes and different data through concurrency test; for the storage space, we test the storage capacity required for data under different number of nodes; for the fault tolerance performance, we compare the ratio of node failure under malicious attack in different modes. To improve the eciency of system transactions, partition concurrency and parallelization optimization technology [33, 34] are applied in the experiment.

In Consortium BlockChain such as fabric, node throughput is restricted by communication delay, transaction verification time and hash operation time. In a system with N

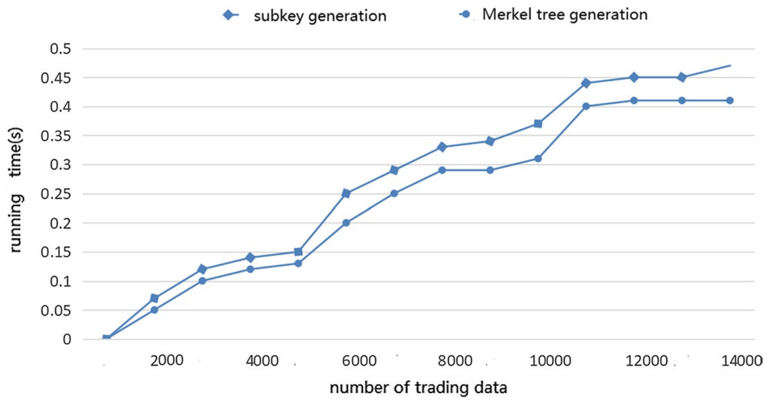

Fig. 6 Running time tests

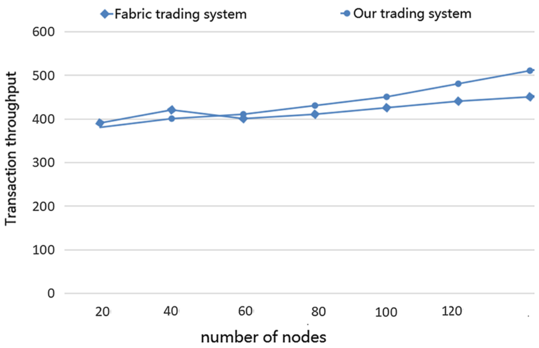

Fig. 7 Transaction throughput tests

nodes and an average block of T transactions, assuming the node communication delay t_c, transaction verification time t_v, and one hash operation time t_o, the throughput of the whole system is shown in equation:

$$TPS = \frac{T}{N\times t_c + T\times t_v + T\times t_o}.$$

In order to test the storage requirements of transactions in the system, this paper compares the storage variation with nodes. In an N-node blockchain system, the number of transactions is N_T, assuming that the average storage capacity of each transaction is s_t, and the storage capacity of a hash result is s_h. then we can estimate the storage capacity as $S = N_T \times N \times (S_t + S_h)$.

6 Experimental evaluation

To test the running speed of the system, we note down the running time results of subkey generation, and Merkel tree generation for data encryption, which is closely related to the depth of the tree since the generation of subkey and Merkel tree are based on binary tree. We use $D = \lceil \log_2 n \rceil + 1$ to calculate the depth of the tree, where n represents the number of transaction data. The test results are shown in Fig. 6.

With the growing number of transaction data n, the depth D of the tree increases, similarly, the generation time of the subkey and Merkel tree also increases. The test results in Fig. 6 show that when the number of transaction data is 5000, the depth of the subkey generation tree and Merkel tree is 13, while the depth of the tree reaches 14, the number

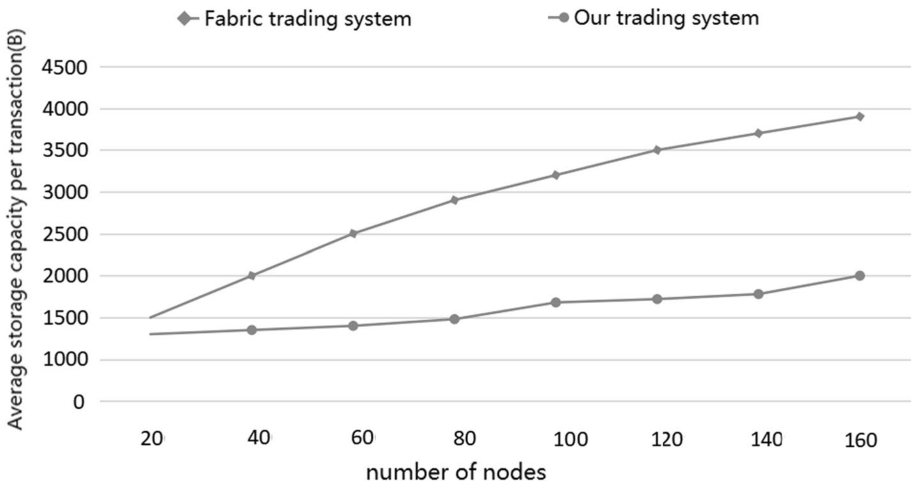

Fig. 8 Storage tests

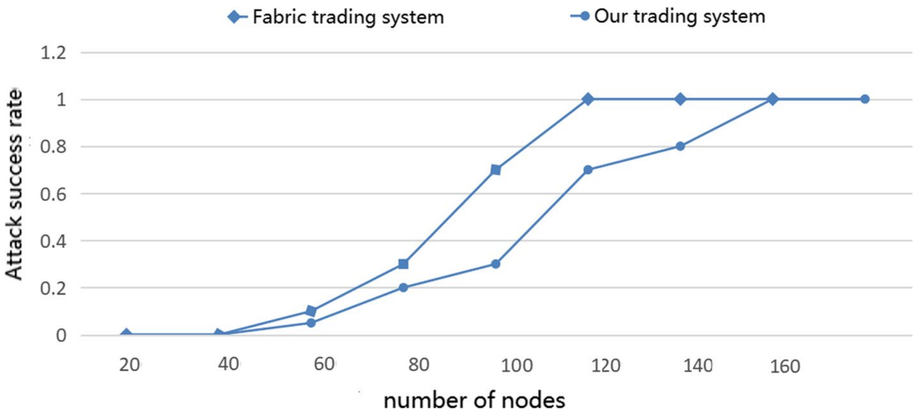

Fig. 9 Fault tolerance tests

of transaction data is. As the depth of the tree increases, the running time rises by about 0.1 seconds.

With the expanding number of nodes in one partition, transaction throughput of the system increases as well, which is higher than that of fabric system. As shown in Fig. 7, with nodes increasing in one partition, the transaction throughput can be improved due to partition concurrency, parallelization optimization and the reliable security setting which can reduce the transaction verification cost. Thus, this system can effectively improve the transaction throughput.

In order to test the storage requirement per transaction in the system, we compare the storage variation with nodes.

As shown in Fig. 8, when the number of nodes in in fabric system increases, the average storage capacity per transaction rises obviously, while average storage capacity per transaction adds slowly in the trading system described in this paper, along with nodes increasing. Therefore, it can reduce storage costs when dealing with large amounts of data in our trading system.

By comparing the success rate of malicious attack on the fabric system and the system in the paper under different ratio of malicious nodes, as shown in Fig. 9, in the traditional PBFT-based blockchain system [33, 35], if the number of malicious nodes is less than 1/3, fault tolerance can be achieved breezily, but if it exceeds 1/3, it will be affected. This system can tolerate more node failures.

7 Results and discussion

We have implemented a blockchain based spatial data trading system. The experiments results results show that the system described in this paper has better throughput, less storage capacity and better fault tolerance, adapting to the high-frequency and real-time requirements of spatial data trading. The sustainability of the trading system is the guarantee for the long-term effective operation of the system. The introduction of reputation mechanism into the trading system is a useful supplement to the auction mechanism and improves the reliability of transactions; there may also be another situation in the auction, where potential high-quality bidders who participate in the auction may not be selected for many times, which will seriously affect their enthusiasm to participate in the transaction as they would see little chance to win. To solve this problem, we will further combine the reputation mechanism and ``virtual participation credit"[36] algorithm to improve the possibility of the loser being selected, so as to build a long-term effective trading system and ensure the sustainability of the trading system.

Additionally, “Scalability Triple Difficulty” refers to the unavoidable contradiction among scalability, decentralization and security of block chain systems. Given that the system is a completely decentralized and does not rely on any trusted third party, scalability and security will be more challenging [13, 37]. In order to ensure security, the system proposed in this paper sacrifices scalability, to some extent. Though partition concurrency and parallelization optimization technology can improve the efficiency of system transactions, which is verified both in [33, 34] and our experiments, however, it is hard to achieve very high security. In the future deployment of the data trading system, we will further explore the trade-off between scalability, decentralization and security [38].

8 Conclusion

In this paper, we propose a blockchain-based spatial data trading framework , where we design a decentralization platform with trusted execution environment(TEE) using the framework.

The platform, characterized with huge remote sensing data storage and processing capabilities, is designed to deal with not only spatial data transaction, but also related services in the future. It will provide service interface, upon which users, without prior installation, can obtain the integrated services of remote sensing data, software and computing resources via the cloud service terminal, can conveniently and economically use high resolution remote sensing information, effectively reducing the threshold of remote sensing in terms of cost and maintenance.

Abbreviations

IPFS	InterPlanetary file system
SDTE	Secure blockchain-based data trading ecosystem
SGX	Software Guard Etxtensions
DHTs	Distributed hash tables
TEE	Trusted execution environment
PBFT	Practical byzantine fault tolerance

Author contributions

LH carried out the design of the trading system and experiments in this paper and drafted the manuscript. WT collected spatial data, YW participated in the design of smart contract, SW carried out embellishment of the paper, performed the grammatical analysis and financial support. All authors read and approved the final manuscript.

Availability of data and materials

Data sharing is not suitable for this paper, because a lot of data in this paper is personal privacy data, which needs relevant legal protection.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 14 January 2021 Accepted: 29 May 2022
Published online: 19 August 2022

References

1. K. Misura, M. Zagar, Data marketplace for internet of things, in *2016 International Conference on Smart Systems and Technologies (SST)* (2016)
2. D. Zhao, C. Ye, B. Zhang, Data marketplace and its values in data trading. Libr. Inf. Serv. **61**(13), 5–12 (2017)
3. J.Z.F. Pang, H. Fu, W.I. Lee, A. Wierman, The efficiency of open access in platforms for networked cournot markets, in *IEEE Infocom-ieee Conference on Computer Communications* (2017)
4. G.S. Ramachandran, R. Radhakrishnan, B. Krishnamachari, Towards a decentralized data marketplace for smart cities, in *2018 IEEE International Smart Cities Conference (ISC2)* (2019)
5. D.D. Nguyen, M.I. Ali, Enabling on-demand decentralized iot collectability marketplace using blockchain and crowdsensing, in *Global IoT Summit* (2019)
6. G. Su, W. Yang, Z. Luo, Y. Zhang, Y. Zhu, Bdtf: a blockchain-based data trading framework with trusted execution environment (2020)
7. A. Sadiq, N. Javaid, S. Omaji, A. Khalid, M. Imran, Efficient data trading and storage in internet of vehicles using consortium blockchain, in *16th International Wireless Communications and Mobile Computing Conference (IWCMC), 2020* (2020)
8. S. Nakamoto, Bitcoin: a peer-to-peer electronic cash system. https://bitco.in/pdf/bitcoin.pdf (2008)
9. L. Ruinian, S. Tianyi, M. Bo, L. Hong, C. Xiuzhen, S. Limin, Blockchain for large-scale internet of things data storage and protection. IEEE Trans. Serv. Comput. **12**, 762–771 (2018)
10. S. Zheng, L. Pan, D. Hu, M. Li, Y. Fan, A blockchain-based trading platform for big data, in *IEEE INFOCOM 2020—IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 991–996 (2020)
11. X. Liu, H. Huang, F. Xiao, Z. Ma, A blockchain-based trust management with conditional privacy-preserving announcement scheme for vanets. IEEE Internet Things J. **7**(5), 4101–4112 (2020)
12. L.D. Nguyen, S.R. Pandey, B. Soret, A. Broering, P. Popovski, A marketplace for trading AI models based on blockchain and incentives for iot data. CoRR arXiV:2112.02870 (2021)
13. J. Wang, Z. Zheng, F. Wu, G. Chen, Blockchain based data marketplace. Big Data **6**(03), 25–39 (2020)
14. G. Zyskind, D.M.S. Zekrifa, P. Alex, O. Nathan, Decentralizing privacy: using blockchain to protect personal data, in *IEEE Security and Privacy Workshops* (2015)
15. M. Li, J. Weng, A. Yang, W. Lu, Y. Zhang, L. Hou, L. Jia-Nan, Y. Xiang, R. Deng, Crowdbc: a blockchain-based decentralized framework for crowdsourcing. IEEE Trans. Parallel Distrib. Syst. 1 (2018)
16. F. Baig, F. Wang, Blockchain enabled distributed data management—a vision, in *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)* (2019)
17. W. Dai, C. Dai, K.K.R. Choo, C. Cui, D. Zou, H. Jin, Sdte: a secure blockchain-based data trading ecosystem. IEEE Trans. Inf. Forensics Secur. **15**, 725–737 (2020)
18. L. Anselin, I. Syabri, Y. Kho, Geoda: an introduction to spatial data analysis. Geograph. Anal. **38**, 5–22 (2006)
19. J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, Y. Zhang, Blockchain for secure and efficient data sharing in vehicular edge computing and networks. IEEE Internet Things J. **6**(3), 4660–4670 (2019)
20. L.D. Nguyen, I.L. Mayorga, A.N. Lewis, P. Popovski, Modeling and analysis of data trading on blockchain-based market in iot networks. IEEE Internet Things J. **8**(8), 6487–6497 (2021)
21. Y. Chen, H. Li, K. Li, J. Zhang, An improved p2p file system scheme based on ipfs and blockchain, in *2017 IEEE International Conference on Big Data (Big Data)* (2017)
22. S. Vimal, S.K. Srivatsa, A new cluster p2p file sharing system based on ipfs and blockchain technology. J. Ambient Intell. Hum. Comput. (2) (2019)

23. X. Wu, Y. Han, M. Zhang, S. Zhu, Secure Personal Health Records Sharing Based on Blockchain and IPFS (2020)
24. A. Sadiq, M.U. Javed, R. Khalid, A. Almogren, M. Shafiq, N. Javaid, Blockchain based data and energy trading in internet of electric vehicles. IEEE Access **9**, 7000–7020 (2021)
25. J. Zhang, S. Zhong, J. Wang, X. Yu, O. Alfarraj, A storage optimization scheme for blockchain transaction databases. Comput. Syst. Sci. Eng. **36**(3), 521–535 (2021)
26. S. Dziembowski, L. Eckey, S. Faust, Fairswap: How to fairly exchange digital goods, in *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018*, Toronto, ON, Canada, October 15–19, 2018, pp. 967–984 (2018)
27. C. Xu, K. Zhu, C. Yi, R. Wang, Data pricing for blockchain-based car sharing: a Stackelberg game approach, in *IEEE Global Communications Conference, GLOBECOM 2020, Virtual Event*, Taiwan, December 7–11, 2020, pp. 1–5 (2020)
28. W. Shi, C. Wu, Z. Li, A Shapley-value mechanism for bandwidth on demand between datacenters. IEEE Trans. Cloud Comput. **6**(1), 19–32 (2018)
29. N. Eltayieb, R. Elhabob, A. Hassan, F. Li, A blockchain-based attribute-based signcryption scheme to secure data sharing in the cloud. J. Syst. Archit. **102**, 101653 (2019)
30. Q. Feng, D. He, S. Zeadally, K. Liang, Bpas: Blockchain-assisted privacy-preserving authentication system for vehicular ad-hoc networks. IEEE Trans. Ind. Inform. **16**, 4146–4155 (2019)
31. B.G. Jeong, T.Y. Youn, N.S. Jho, S.U. Shin, Blockchain-based data sharing and trading model for the connected car. Sensors **20**(11), 3141 (2020)
32. M. Hoekstra, R. Lal, P. Pappachan, V. Phegade, J.D. Cuvillo, Using innovative instructions to create trustworthy software solutions, in *Proceedings of the 2nd International Workshop on Hardware and Architectural Support for Security and Privacy* (2013)
33. Tao, H. Research on key algorithms and technologies of blockchain for distributed energy trading scenarios. PhD thesis, University of Electronic Science and Technology of China (2020)
34. Wenlin, L. Ethereum throughput bottleneck analysis and optimization research. PhD thesis, Xiangtan University (2020)
35. G. Wang, M. Nixon, Randchain: practical scalable decentralized randomness attested by blockchain, in *2020 IEEE International Conference on Blockchain (Blockchain)*. IEEE (2020)
36. Luo, T., Kanhere, S.S., Huang, J., Das, S.K., Wu, F. Sustainable incentives for mobile crowdsensing: Auctions, lotteries, and trust and reputation systems. IEEE Commun. Mag. 55(3), 68–74 (2017)
37. D. Kraft, Difficulty control for blockchain-based consensus systems. Peer-to-Peer Netw. Appl. **9**, 397–413 (2016)
38. J. Tang, Y. Cui, Q. Li, K. Ren, J. Liu, R. Buyya, Ensuring security and privacy preservation for cloud data services. ACM Comput. Surv. **49**(13), 13 (2016)

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.