Article | Received 31 May 2024; Accepted 12 September 2024; Published 23 September 2024
https://doi.org/10.55092/blockchain20240006

Blockchain-based privacy-preserving framework for multi-party data sharing in emerging IoT

Pingchuan Xu, Junxi Liang, Yuxuan Luo, Xinyi Peng and Zhe Sun*

Cyberspace Institute of Advanced Technology, Guangzhou University, Guangzhou, China

* Correspondence author; E-mail: sunzhe@gzhu.edu.cn.

Abstract: With the arrival of the Internet of Things era, the information world can perceive every corner of the real world, providing a steady stream of data collection for its data analysis. However, multi-party data sharing and security analysis have become a challenge that must be solved as Internet of Things technology evolves to the next level. Currently, most data sharing must introduce a trusted entity to perform cross-domain data fusion and protect privacy from a global perspective. Unless trusted entities can ensure that the privacy preserving mechanism is implemented properly, once the data is leaked, it will cause great harm to the public and to private interests. In this paper, we propose a new privacy preserving framework to tackle the problem of trusted entities in multi-party data sharing, which is based on a distributed consensus mechanism on blockchain. We analyze and abstract the privacy and availability requirements of all participants in multi-party data sharing and design a collaborative on-chain-off-chain privacy protection mechanism. To ensure that no one participant's private data is not accessed by each other, promote fair transaction of smart contracts, and achieve a balance among privacy preserving effects, transaction fairness, service quality, and execution efficiency from a global perspective.

Keywords: consortium blockchain; privacy-preserving framework; multi-party data sharing

1. Introduction

As Internet of Things (IoT) technology evolves, various sensors can easily and quickly collect various real-world data such as images, sound, temperature, humidity, light intensity and so on. The total amount of real-world data (RWD) collected through IoT devices every day has reached an astronomical number [1]. Using this data; users can perform application analysis and provide feedback to the real-world via services. The IoT devices and IoT-data-based analysis systems have brought the traditional perception-oriented IoT into a new stage of "Real-World to Information-World to Real-World". Medical IoT equipment and its analysis system can support physician and experts to predict a patient's susceptibility to potential illnesses, which save the time and cost of the cure [2]. Transportation sensors and its analysis

applications can help users match the best travel plans [3]. However, as data aggregation has become increasingly common demand in various fields [4], the issue of privacy preserving in multi-party data sharing has also become a major latent hazard [5]. Failure to implement the data privacy preserving mechanism once the data is leaked will cause great harm to personal privacy and even public safety.

At present, privacy-preserving technologies in data analysis platforms are mainly focused on desensitization technologies to filter and protect private information when data is released. The privacy-preserving data publishing (PPDP) mainly involves anonymity [6], perturbation [7–9], encryption [10–12] and other methods. Anonymity technology [6] mainly reduces the probability of an adversary's guessing by replacing identifications and using fake nodes; Perturbation technology [13] refers to the practice of intentionally adding random noise or disturbances to data or systems in order to protect them from malicious attacks. This technique is designed to obscure or mask the true underlying information, making it difficult for attackers to accurately analyze or exploit the data. is to increase random noise to prevent attacks. Such methods have a certain adverse effect on the usability of data. Homomorphic encryption [12] is a type of encryption that enables mathematical operations to be carried out on ciphertext, producing an encrypted result that, when decrypted; matches the result of operations performed on the plaintext. This means that data can remain secure and private throughout the entire computational process, as it never needs to be exposed in an unencrypted form. The service requester can obtain the calculation results while preventing the disclosure of the plaintext of the data. However, this method supports fewer types of computing and it has a huge impact on computing performance, making it is difficult to generalize to real-world applications [14]. Furthermore, when data comes from multiple data owners, it is usually necessary to introduce a trusted third party to fuse data from different sources and perform some privacy preserving operations at a global level. How to ensure that the behaviors of third parties are authentic and credible has become a difficult problem that many cross-domain data privacy- preserving solutions need to solve. And the next generation of Internet of Things (NGIoT) networks [15] may have wireless deception attack vulnerabilities that can lead to catastrophic consequences for mission-critical applications. Traditional solutions may bring additional signal processing, protocols, and latency overhead.

Blockchain technology [16–21], as an important technology for establishing a mechanism of mutual trust, possesses high application value in the realization of privacy preserving and oversight of multi-party data sharing within IoT data analysis. Using a smart contract in a consortium blockchain [22] can facilitate the completion of data transactions quickly so that both parties in a data transaction can agree on their rights while agreeing on the transaction amount. It has become a research challenge to ensure that the data sets of the data sellers and the trained model of the data buyer are not obtained by the other party. In addition, the existing blockchain technology still has major limitations in processing performance, and it is difficult to directly perform large-scale numerical calculations on the consortium blockchain [23]. In light of this, traditional blockchain design solutions with multi-point storage backup, throughput and performance bottlenecks will be difficult to apply

directly to multi-party mass data sharing scenarios. It is urgent to conduct systematic research on the above issues to ensure user privacy and the rational use of data.

The main innovations of this paper are as follows:

(1) propose a privacy protection framework that combines on-chain and off-chain approaches, and design a smart contract transaction security mechanism that balances privacy protection and transaction fairness.

(2) We design a multi-party data security sharing mechanism to achieve secure data sharing on the blockchain through smart contracts and encryption algorithms.

(3) We validate the feasibility of the proposed framework in the emerging IoT multi-party data sharing environment through case illustrations.

In this paper, we comprehensively analyze and abstract the privacy and availability requirements of each participant in a multi-party data sharing scenario, and propose a blockchain-based privacy protection framework. The rest of the paper is organized as follows. Section 2 reviews some previous studies related to the proposed framework. Sections 3 and 4 formally introduce the motivation and the systematic process and design details of our proposed framework and conduct a case study. Section 5 concludes and outlines future work.

2. Related work

2.1. Privacy-preserving approaches for data sharing

The goal of privacy-preserving approaches for data sharing is preventing the leakage of confidential data when uploading and handling sensitive data on the network. Existing privacy-preserving approaches mainly involve anonymity [6], perturbation [7–9], encryption [10–12], and so on.

The k-anonymity method provides privacy protections by ensuring that each posted record will involve at least k individuals. Sweeney [6] proposes to combine generalization and suppression to understand the formal representation of k-anonymity, and uses the minimum generalization algorithm (MinGen) to provide k-anonymity protection with minimal distortions. The disturbance-based methods are to modify the original data before sending it to a service provider, such as adding random noise [13] and differential privacy [10]. But these approaches may adversely affect the accuracy of the analysis.

Encryption-based approaches do not affect the data itself, but they can place a significant performance burden. But most of them need trusted third parties to perform data fusion. Banu *et al.* [24] propose a technique based on principal component analysis to protect sensitive information in multi-party aggregation scenarios. Mohammed [25] proposed a Secure and Private Blockchain-based Data Sharing (SPB-DS) scheme for Mobile Edge Computing (MEC) systems by encrypting predefined searchable keywords and exploiting the properties of blockchain for data sharing, while two smart contracts based on an optimal payment mechanism are used to reward both providers and requesters for participating in data sharing.

2.2. Blockchain approaches for multi-party data sharing

In multi-party data sharing, a third party is typically introduced to fuse data from different sources and perform some privacy preserving operations at the global level. The decentralized nature of blockchain technology solves the untrustworthiness of traditional sharing on third-party platforms.

Many blockchain-based scenarios have been proposed for the past few years for data access control and trusted transactions in multiparty data sharing [16–21]. Gai *et al.* [25] proposed a blockchain-based privacy-preserving location data sharing scheme, which uses blockchain technology to achieve privacy-preserving location data sharing and uses zero-knowledge constraints to prevent double-payment attacks based on account-model blockchain. Dai *et al.* [16] propose a new blockchain-based data transaction system SDTE, aimed at reducing the limitations of the existing data transaction market. Under this system, none of the data broker and buyer can approach the seller's raw data, they can only access the analysis results they need. At the same time, by embedding access control regulations in smart contracts to control the access to user data in the channel. In terms of data security sharing, Huang *et al.* [18] propose a block chain-based multi-group data sharing system with anonymity and traceability to enable secure data sharing between different groups of users. Salah *et al.* [26] propose a solution that uses smart contracts and IPFS technology to offer a safe, visible, and credible platform for online review systems. Nosouhi *et al.* [27] proposed a novel Information Center Network (ICN) system model based on CPS to process data from IIoT devices and enhance the security and performance of the system.

Nevertheless, anonymity and identity concealment influence the motivation to share data. Jing *et al.* [28] presented proposed a telemedicine data sharing scheme that combines blockchain and cloud services. By employing an anonymous and traceable authentication protocol, it protects the privacy of users' identities and reveals the identity of malicious nodes when necessary. The problem of anonymous sharing of telemedicine health data on blockchain is achieved.

Weng *et al.* [29] propose Deep Chain, which incents multiple participants to perform a common training work by asking them to upload their gradients obtained in local training. Shen *et al.* [30] propose a data-sharing model based on blockchain and smart contracts, aiming to build a privacy-preserving data-sharing platform for the participants. The model analyzes the relationship between different participants to use the Shapley value in game theory to adjust the revenue distribution as an incentive mechanism.

However, blockchain technology was originally designed primarily for disclosure and auditability, without taking into account the privacy information contained in data sharing.

2.3. Blockchain-based privacy-preserving approaches

Since every participant in the blockchain jointly maintains an open and transparent ledger, which includes all transactions, sending addresses, receiving addresses, and data content. Malicious participants can easily steal transaction data on the chain. The lack of a privacy-preserving mechanism restricts the blockchain applications in data security protection fields.

This transparency is crucial for trust and verifiability, allowing participants to audit and verify transactions without a central authority. However, it also exposes sensitive information to all network participants, including potential malicious actors who can exploit this data for unauthorized access or misuse. The core challenge is reconciling this transparency with privacy protection. While the openness of blockchain supports auditability and decentralized trust, it inherently compromises user privacy by making all transaction data publicly accessible. This limitation restricts blockchain's applicability in fields where confidentiality is crucial, such as financial services, healthcare, and personal data management. How to strike a balance between the openness, transparency, and verifiability of the blockchain and the need of preserving user privacy has attracted the attention of many researchers [31–34].

To protect the privacy of IoT data collected from multiple parties, Chen *et al.* [35] proposed a privacy-preserving data sharing blockchain system based on the Internet of Health Things, aiming to solve the privacy preservation and data security problems in healthcare data sharing. The system adopts the Content Extraction Signature (CES) scheme in order to achieve fine-grained privacy protection for patient data. Shen *et al.* [36] proposed a privacy-preserving machine learning method called Secure SVM, which trains encrypted IoT data on the blockchain by using SVM classification algorithms. To address the problem of dynamic increase in data volume, Zheng *et al.* [23] proposed a scalable co-analysis method that dynamically encrypts data using Paillier cryptosystem. To address the current problem that blockchain cannot protect the privacy of transaction records, Kosba *et al.* [22] proposed the Hawk system, which supports the encryption of smart contracts using protocols such as zero-knowledge proofs. This approach effectively protects the privacy of transaction records within the blockchain. Keshk *et al.* [37] proposed a framework to protect security and privacy in smart grids. The framework uses a privacy module to prevent data poisoning attacks and inference attacks, while a malicious conflict detection module uses LSTM to train and verify the output of the privacy module. Ensuring the security and privacy of the data on the chain is extremely critical, the solution proposed in [32] does not consider system security and privacy, Lin *et al.* [38] proposed a new architecture called SecBCS with incentives and privacy protection.

In summary, the current blockchain-based privacy protection technology is still in its infancy, and various technologies are relatively scattered, and a systematic and instructive research framework is urgently needed.

3. Overview of the framework

In this section, we introduce the structure of the framework that will be discussed in this paper in order to quickly get into the details of the problem. In order to better articulate our privacy-preserving scenarios, we will further discuss specific scenarios in the next section to justify our proposal.

As a general blockchain-based multi-party data sharing scenario (shown in Figure 1), our scenario involves two roles: data owner and data buyer.

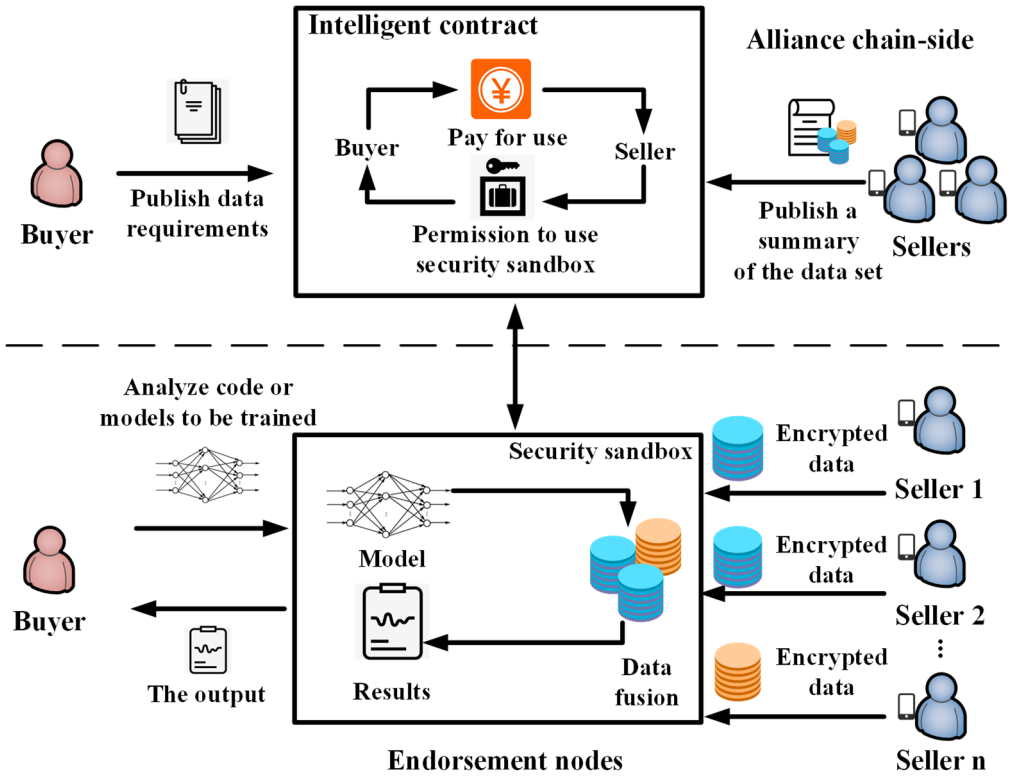

Figure 1. Blockchain-based multi-party data sharing scenario.

The data owner collects data from end devices, like industrial equipment sensors, smart devices, *etc.*, and provides the collected data for further data analysis to the data buyer. As payback, data buyer pays for the data used for analysis as transactions. Data analysis may involve data fusion from multiple fields, thus our scenario may involve multiple data sellers.

To preserve the data privacy of the data owner, data buyer cannot directly get access to the data, but provide analyzing code. And a trustable third party is necessary to conduct the analysis where all information involved during the analysis is kept privately. Since on blockchain, all data and code are promised, and could be trusted, we decide to solve the data analysis problem on the blockchain. In our scenario, we focus on the consortium blockchain since it enables us to restrict the data analysis within a limited node list, which will be talked about in the following sections.

To reach an agreement between the data owners and data buyer, the data owners and data buyers need to respectively upload their data and analyzing code up to the blockchain with corresponding description like data set, price, analyzing behavior, *etc.* The data is kept as private data on a set of selected and trusted nodes by the data owner. And the analyzing code is uploaded as chain code. When the data buyer has analyzing demand, he proposes transaction indicating his analyzing chain code. Nevertheless, the scenarios described above still raise one key i4yussues that need to be urgently resolved.

3.1. Privacy problem analysis: an on-chain-off-chain collaborative privacy preserving mechanism

In multi-party mass data sharing scenarios, data buyer can analyze the data resources of different data sellers, train or optimize their own service models, and enjoy the benefits of enormous IoT data. However, in this process, an entity must undertake the work of data fusion. For data security and privacy preserving considerations, data sellers and data buyer need to ensure that their data and trained model parameters are not exposed to each other. From a privacy protection perspective, traditional PPDP technology can have a negative impact on the accuracy of data analysis results (such as anonymity, disturbance), or affect the feasibility of the entire framework in terms of computational performance (as encryption).

If the data privacy intensity is not strong enough, data buyers can still steal a certain degree of original data set information by analyzing and reconstructing the output results, calling additional chain codes, and monitoring the status of a consortium blockchain. Therefore, we propose a consortium blockchain-based privacy preserving framework for multi -party data sharing to ensure the authenticity and credibility of the on-chain fusion analysis node. Specifically, we design a smart contract transaction guarantee mechanism that takes into account privacy and fairness, and combines the encryption algorithm with the private data management of the consortium blockchain. By ensuring that the access rights of the various participating entity nodes are not tampered with, the smart contract can execute the Data analysis as agreed. In the meantime, we are also designing a code review mechanism in the security sandbox to prevent malicious attackers from stealing other participants' private data by adding additional function codes to smart contracts, generating analysis results and monitoring status. By combining on-chain and off-chain approaches, our proposed framework can provide more comprehensive privacy guarantees for data sharing on the blockchain.

3.2. Blockchain-based privacy-preserving framework for multi-party data sharing

Our general objective of the framework is to provide privacy preserving which covers the entire data processing procedure, and try our best to promote the transaction speed. To that end: (1) we switch the raw data to equivalent parameters to secure the off-line data transmission; (2) we monitor the online communication behaviors of the chain code to secure the on-line transactions; and (3) we separate the original data analyzing mission into pieces, which could be endorsed separately to improve the data analyzing speed and efficiency.

We first give an overall framework description (shown in Figure 2), then focus on the key challenges respectively. The data sellers keep collecting the data respectively as the input of our framework. And conduct deep learning on such data for model training with recent sliding window data. We take the trained model as the extracted data for further data analysis, uploading and updating such data onto the blockchain.

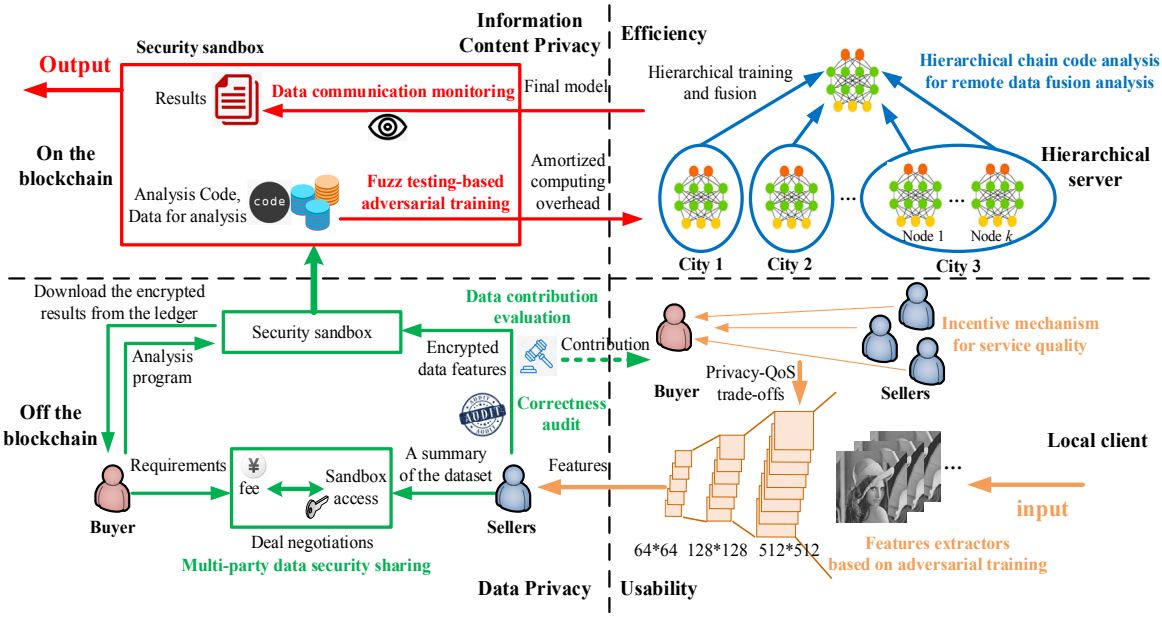

Figure 2. Blockchain-based privacy preserving framework for multi -party data sharing.

The data uploaded on to the blockchain is taken as private and encrypted data on the peers which are selected by the data seller. Such data cannot be visited without authorization of its owner. The data owner also provides a summary of the dataset for the data buyer to purchase. The transaction engine of our framework is a deal negotiation chain code, which take transaction requirement from the data buyer. When receiving such transaction proposal, it searches for trade matching among the dataset description, and return the result match and sellers' authorization.

The data analyzing engine also works as a chain code on the block chain. When the data buy gets the result of match trade. It proposes to the data analyzing engine with sellers' authorization, and the corresponding analyzing chain code address. The data analyzing engine then selects the peers which storing the sellers' private data as endorsing peers to conduct the data analysis. With the authorization of all data sellers, the endorsing peers are able to require the private data from the other endorsing peers, and call the decryption chain code of the data sellers for data decryption. During the data analysis process, we monitor the data communication of the data buyer's chain code, including its returning value and co-occurrence of world state changes as suspicious data privacy transmission.

To improve the usability and efficiency of our framework, we evaluate the contribution of the sellers' data to the analyzed result, and feedback to the sellers to encourage them to upload more valuable data. The data seller may involve distributed data storage and uploading problem, thus our framework also enables parallel data analysis conducted on multiple set of endorsing chain code within various channels. The generated results are further fused as the final result.

3.3. Off-chain privacy for smart contract transactions

Given that the consortium blockchain relies on a multi-point back-up to ensure that data is not tampered with, any plaintext data can be accessed by all parties for review, which is

difficult to meet the privacy requirements of data sharing. Encryption mechanisms make it difficult to ensure that users are providing data in the way they claim, thereby undermining the fairness of transactions.

We store user data as private data of alliance nodes to protect the privacy and security of user data during data transactions and mining between buyers and sellers. Upon completion of the transaction match between the buyer and the seller, the sellers' authorized node which stores the private data specified by the transaction approves the transaction. Authorized nodes encrypt their private data and exchange the encrypted data with each other, so as to ensure that a buyer chain code can access all the sellers' encrypted data during the node endorsement process.

To access the seller's data content, the seller must provide a decryption chain code and ensure that the chain code has been installed and deployed on all possible approval nodes in the chain before the transaction occurs (shown in Figure 3). Then they can authorize the buyer's data analysis chain code to access the seller's data by calling the seller's decryption chain code. Since the sellers' data decryption key is updated in real time through the chain code status, the sellers can change their decryption chain code status when the buyer and seller reach the transaction. As a result, encrypted data can only be accessed during transaction training and in the security sandbox.

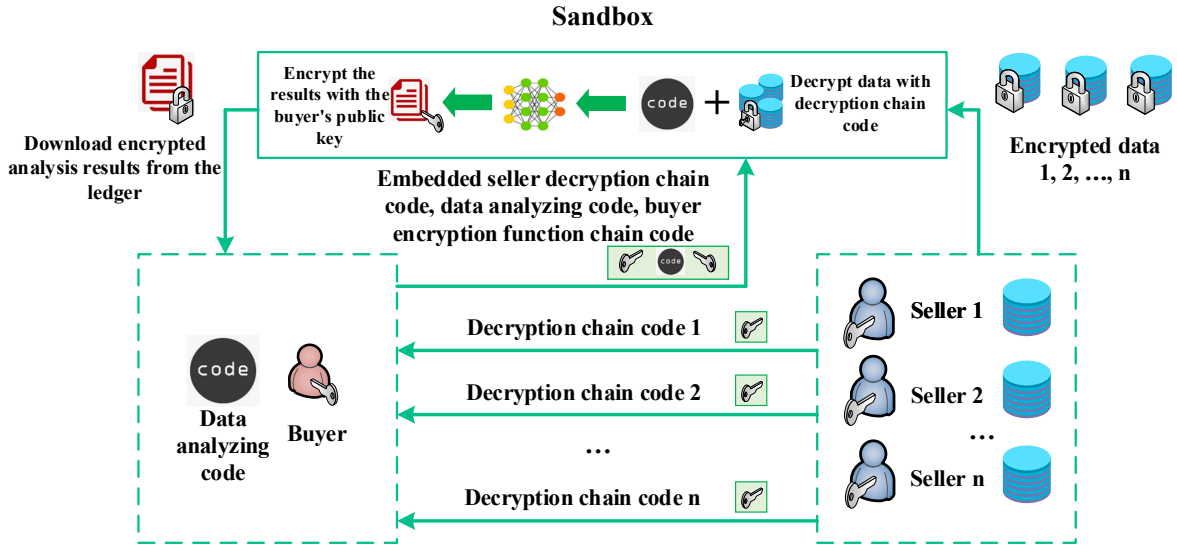

Figure 3. Multi-party data security sharing for consortium blockchain.

3.4. Data privacy preserving on blockchain

As a trusted third party, the behavior of data analysis on the blockchain is promised because the analyzing code could be installed and deployed on the blockchain nodes as chain code. However, such chain still cannot satisfy our requirements of data privacy preserving for two reasons. First, chain code is binary code to blockchain users, which cannot be interpreted or audited, thus may leave a backdoor to the chain code owner. Second, data utilized for analysis should be unreadable to all users but the data owner. However, there is still no appropriate methods which can perfect meet our requirement (methods like differential privacy and

homomorphic encryption could do harm to the analysis accuracy and analysis efficiency, which is critical for blockchain-based applications).

To fill the gap between chain code implementation and data privacy preserving requirements, our idea is to keep the data uploaded to the blockchain encrypted as private data, and merely decrypt such data during the chain code analyzing. Considering the chain code owner is able to steal the data privacy, and the key point is to avoid that.

(1) Customized endorsement based on data sellers

Since the buyer has the source code, it is feasible to monitor the running statement of the VM endorsing the transaction, and learning the data being analyzed. Therein, the endorser peer should be a trusted peer which should not collude with chain code owner. In our scenario, we believe the peers of the data sellers could be trusted, because any collusion (the seller helps the buyer to monitor the endorsement) will also do harm to the data privacy of the data seller itself.

When a transaction among the data sellers and the data buyer is decided, we just select the peers which stores the private data of each data sellers as the endorsing peers (According to our scenario description, such peers are trusted by the data sellers). Then enable the data access of the private data among the endorsing peers. Then each endorsing peer has the entire set of all sellers' data for analysis, which could be provided to the analyzing chain code for endorsement.

(2) Chain code communication behaviors monitoring

All involving chain code need to be robust enough to reverse engineering, so that no seller could conduct reverse engineering to learn the data privacy of the other data sellers. Fortunately, such countermeasures are well investigated and proposed. Chain code developers are able to make the reverse engineering very difficult with acceptable efforts. Due to the page limit of the paper, we do not talk about that in this paper.

Since the chain code is robust to reverse engineering, the sellers also have the problem to learn whether the buyer is stealing their data privacy. Our point is to monitor the data transferring behaviors of the buyer's chain code, and report the suspicious behaviors for further offline countermeasures [39]. In our framework, we focus on 2 data behaviors: returned results, and call code behaviors. To monitor the returned results, we predefine a set of rules and keywords.

(3) Key management for the private data

Since the data analysis requires to fuse all sellers' data, a seller has the chance to get access to the other sellers' data. Therein, no decryption key should be transferred among the buyer and all the sellers. As an alternative, our solution is to ask each seller to provide their own decryption chain code which could be used to get access to their data if the transaction is authorized.

When the data buyer's chain code needs to access the data of A, it simply calls the corresponding method of a and authorizes the access control through a. After the analysis, the data seller can generate a new set of encryption and decryption keys and update the decryption keys in the decryption chain code.

4. Security and privacy analysis

(1) Privacy Protection

To enhance the strength of privacy protection, the framework integrates both on-chain Layer L_i and off-chain Layer L_t privacy protection mechanisms. The transparency of on-chain operations facilitates auditing and tracking of data analysis processes, but it may expose the actions of participants. Therefore, privacy-sensitive operations and result generation are executed in a secure off-chain sandbox environment S, reducing the risk of leaking state information from the blockchain. This ensures that only verified code can run, effectively preventing malicious node from using on-chain monitoring or injecting malicious code to steal private dat. This combination of on-chain and off-chain mechanisms enhances the framework's privacy protection capabilities and ensures the security of the system.

(2) Fairness Assurance

The automation and immutability of smart contracts are fundamental to ensuring fairness within the framework. Through the smart contract, the data seller D_s, with the assistance of authorized nodes N, encrypts their private data and exchanges encrypted information with the data buyer D_b. All steps of the transaction are executed transparently within the smart contract, ensuring that neither party can manipulate the process to gain an unfair advantage. Additionally, the seller controls the timing of data decryption through decryption chain code, ensuring that the data cannot be accessed or misused during the transaction. The immutability of the smart contract guarantees the fairness of the transaction process, ensuring that the order and results of each transaction are traceable and impartial. Ultimately, this ensures fairness and security in multi-party data sharing and transactions.

5. Case study and scenario analysis

5.1. The comparison of our framework

As shown in Table 1, Okegbile *et al.* [40] proposed a PBFT-based BeDS framework and conducted a detailed analysis of its communication process, validation process, and block grading process, thereby investigating the impact of latency and data age on the performance of data-sharing systems. Cui *et al.* [41] proposed a solution for cross-domain medical data sharing using distributed cloud storage to address data isolation issues. Compared to the aforementioned solutions, our approach exhibits significant advantages in privacy protection, cross-domain sharing, and offline privacy. By integrating both on-chain and off-chain privacy protection mechanisms, our solution achieves more stringent data protection, employing dynamic permission control and encryption algorithms to ensure high confidentiality during data transmission and storage. Additionally, real-time monitoring technology enables us to promptly detect and prevent data leakage or unauthorized access. This comprehensive approach effectively enhances data privacy, security, and controllability, rather than relying solely on cloud storage or individual privacy measures. Our solution offers a more thorough and detailed approach to handling cross-domain sharing and offline privacy protection.

Table 1. Technical comparison.

	On-chain Privacy-preserving	Cross-domainsharing	Blockchain	Off- chain privacy-preserving	Transaction Fairness
BeDS [40]		√	√		
SCMDS [41]	√	√	√	√	
Our framework	√	√	√	√	√

5.2. Multi-party medical data sharing in IoT

With the advent of the IoT era, an extensive amount of real-world data collected by users is available for developing treatment plans and conducting pharmacological analysis, leveraging deep learning techniques. As a notable example, the U.S. Food and Drug Administration (FDA) has established the FDA Real-World Evidence Program Framework, which evaluates the use of real-world evidence (RWE) throughout the drug development process. Real-world data is typically collected via personal monitoring devices (such as heart rate monitors, blood glucose meters, blood pressure monitors) and hospital diagnostic equipment (including magnetic resonance imaging, CT scans, *etc.*). This collected data can be transmitted seamlessly to analytical services in real time, facilitating prompt feedback on treatment plans.

As depicted in Figure 4, the data collected involves patient privacy and is under the control of patients and medical institutions, including hospitals and clinics. In contrast, new drug research and dosage determination are predominantly conducted by pharmaceutical companies and research institutions. Traditionally, data obtained through cooperation with small-scale medical institutions and volunteers in clinical trials have a limited sampling scope and yield insufficient data. It is critical for research institutions to obtain comprehensive data in a timely manner to overcome new diseases and develop medicines as soon as possible, such as patients with Covid-19.

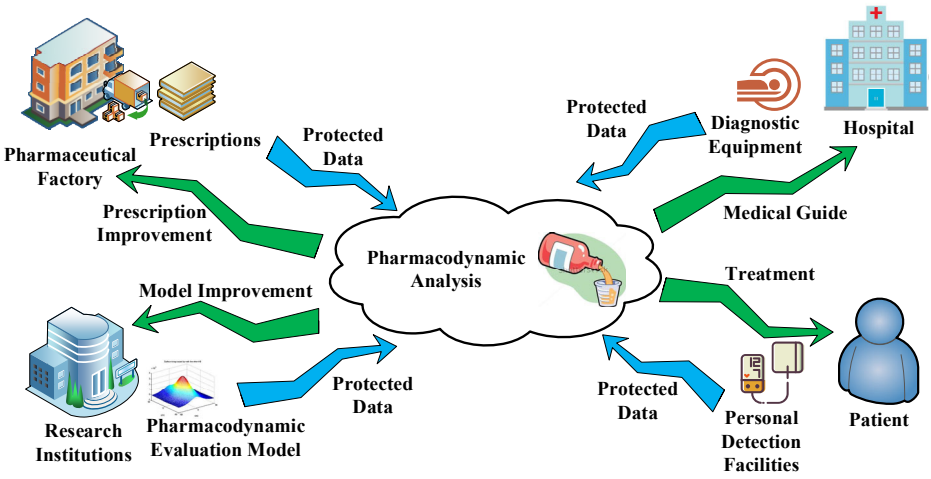

Figure. 4. Collaborative training in pharmacological analysis.

The private information in medical data is closely related to the rights and interests of patients. If not effectively protected, this information might be exploited by malicious attackers

for targeted insurance recommendations and medical fraud. Therefore, medical institutions are obliged to prevent patients' private data from being accessed by unauthorized participants, while also addressing the demand for effective disease treatment. Additionally, medical institutions compensate users for the data provided, aiming to ease the financial burden associated with disease treatment. Consequently, the protection of patient privacy and the fair assessment of the contribution of patient-submitted data have become critical agenda items.

5.3. Blockchain-based privacy protection framework for IoT multi-party healthcare data analytics

Traditional approaches to medical data sharing typically rely on long-term cooperation and established reputations. The advent of a collaborative analysis framework on the blockchain for private data introduces a novel approach. As illustrated in Figure 5, the integration of on-chain and off-chain privacy protection mechanisms serves to prevent the leakage of sensitive user data during collaborative processing.

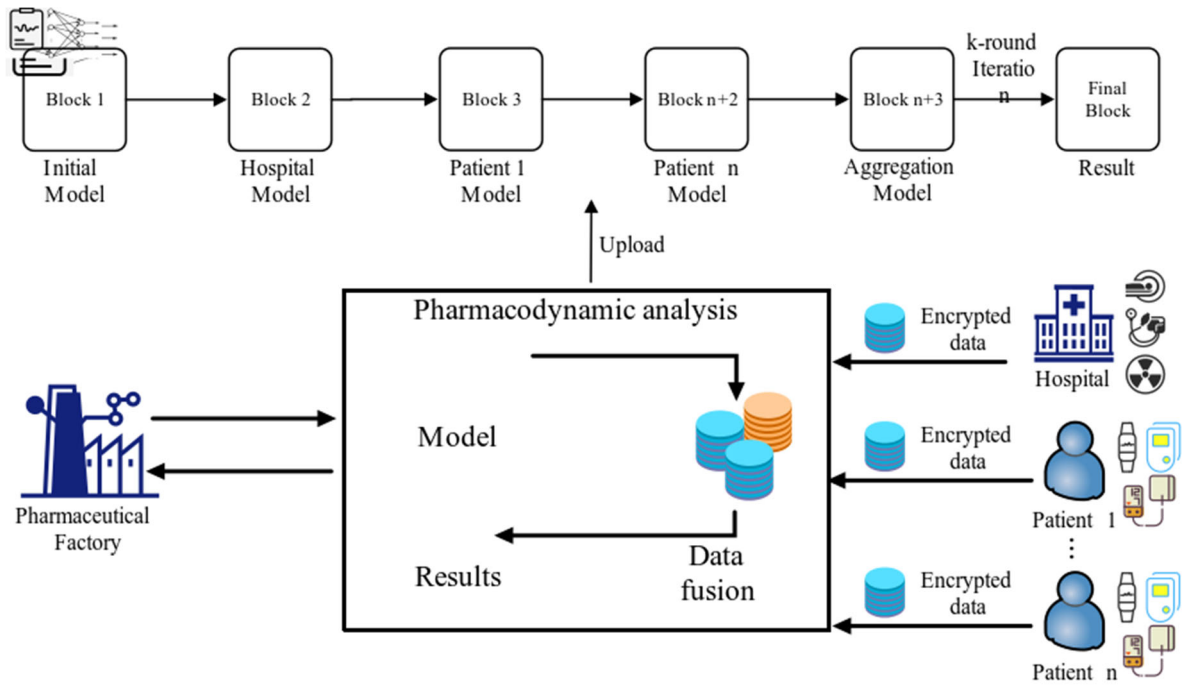

Figure 5. Blockchain-based medical data analysis.

(1) Privacy preserving

To safeguard the private data uploaded by patients to the distributed blockchain ledger from attackers, patients can employ secure multi-party computing protocols, such as functional encryption, to encrypt their locally trained models. Pharmaceutical entities or supporting nodes performing aggregation are only able to decrypt the aggregated parameter model, without accessing the specific data of individual patients. To secure medical service models, the blockchain meticulously records all data-flow information and oversees the utilization of patient information.

(2) Fairness guarantees

The contribution evaluation mechanism is utilized to gauge the precision improvement in the global pharmacological analysis model, thereby ascertaining the contribution of each data set. The degree of contribution is linked to proofs of correctness, source, and data quality, enabling any third party to obtain these proofs for verification purposes. Should the quality of the data model uploaded by a patient significantly deviate from the required standards, the verification process will identify such discrepancies, thus thwarting malicious attempts at data poisoning or free-riding. The fairness guarantee mechanism incentivizes patients to contribute high-quality data and reap benefits from their participation.

(3) Usability tradeoff

Prior to uploading medical data to the blockchain, patients can configure settings against multiple privacy attacks, defining the minimum tolerable leakage threshold, and employ adversarial training to extract data features or model parameters. By calibrating the weights of the data extractor, patients can control the service quality of the uploaded model and the extent of privacy preservation. Pharmaceutical companies may implement a game theory-based incentive mechanism to motivate patients to upload high-quality data.

(4) Efficiency optimization

Given the accelerated research and development of new Covid-19 related drugs, the vast global data volume overwhelms the endorsement nodes of traditional blockchain systems in terms of computational load and time consumption. Pharmaceutical companies can employ a hierarchical chain code analysis architecture to decompose learning models, enabling more efficient processing. Initially, lower-level analysis nodes can be used to train local pharmacodynamic models from patient data, followed by the utilization of top-level nodes to aggregate these local models. This process can be implemented using a multi-chain structure. In this structure, each sub-blockchain generates a partial model, and the resultant analysis serves as input for the comprehensive blockchain model, effectively distributing the computational burden across multiple nodes.

6. Conclusion

Over the last decade, machine learning technologies have made breakthroughs that have driven the expansion of enormous IoT data applications, making multiparty data sharing an irreversible trend. However, multiparty data sharing requires a trusted entity to perform data fusion. Blockchain-based multi-point endorsement technology provides a new way to solve this problem, but there are details of data privacy, information content privacy, usability and efficiency that still need to be perfected. Motivated by this need, we propose a blockchain-based privacy preserving framework to ensure the privacy and usability of multiparty data sharing. Specifically, we design a smart contract transaction guarantee mechanism that takes into account privacy and fairness; these methods can collectively protect data off-chain and on-chain parameters in cross-domain fusion analysis of IoT users' data. Finally, the proposed framework can provide a systematic theory and technical basis for multiparty data sharing.

Limitations and Future Work: the blockchain-based privacy protection framework proposed in this paper effectively addresses the issue of trusted entities in multi-party data sharing, but certain limitations remain in practical application. While the on-chain and off-chain collaborative privacy protection mechanism strikes a balance across multiple aspects, it may encounter performance bottlenecks and challenges in execution efficiency when dealing with more complex and dynamic application scenarios. This is particularly relevant in large-scale, multi-dimensional data sharing environments, where the framework may require further optimization to ensure adequate performance under high concurrency and massive data processing. In our future work, we will focus on improving the scalability and execution efficiency of the framework, particularly in large-scale data sharing scenarios. To address potential performance bottlenecks that could arise in high concurrency and big data environments, we plan to optimize the framework's design. Moreover, we will enhance the flexibility of smart contracts to better accommodate diverse privacy protection and fair transaction requirements in different application contexts. Additionally, we will explore cross-chain data sharing, investigating how to enable secure data interactions between different blockchain platforms while ensuring privacy and fairness in cross-chain operations.

Acknowledgments

Funding: This research was funded in part by the Major Research plan of the National Natural Science Foundation of China, grant number 92167203; in part by Guangdong Basic and Applied Basic Research Foundation, grant number 2024A1515011492; in part by Guangzhou Science and Technology Plan Project, grant number 2023A03J0119.

Conflicts of interests

The authors declared that they have no conflicts of interests.

Authors' contribution

Conceptualization, S.Z.; Investigation, L.J.X., X.P.C., L.Y.X.; Writing - original draft, S.Z., L.J.X., X.P.C.; Writing - review & editing, S.Z., L.J.X.; Project administration, S.Z., L.J.X., P.X.Y; Supervision, S.Z.; All authors have read and agreed to the published version of the manuscript.

References

[1] Cave A, Kurz X, Arlett P. Real‐world data for regulatory decision making: challenges and possible solutions for Europe. *Clin. Pharmacol. Ther.* 2019, 106(1):36.

[2] Silva BM, Rodrigues JJ, de la Torre Díez I, López-Coronado M, Saleem K. Mobile-health: A review of current state in 2015. *J. Biomed. Inform.* 2015, 56:265–272.

[3] Lu N, Cheng N, Zhang N, Shen X, Mark JW. Connected vehicles: Solutions and challenges. *IEEE Internet Things J.* 2014, 1(4):289–299.

[4] Li R, Harai H, Asaeda H. An aggregatable name-based routing for energy-efficient data sharing in big data era. *IEEE Access* 2015, 3:955–966.

[5] Lv Z, Song H, Basanta-Val P, Steed A, Jo M. Next-generation big data analytics: State of the art, challenges, and future research topics. *IEEE Trans. Ind. Inf.* 2017, 313(4):1891–1899.

[6] Sweeney L. Achieving k-anonymity privacy preserving using generalization and suppression. *Int. J. Uncertainty Fuzziness Knowledge Based Syst.* 2020, 10(5):571–588.

[7] Beaulieu-Jones BK, Wu ZS, Williams CJ, Lee R, Bhavnani S, *et al*. Privacy-preserving generative deep neural networks support clinical data sharing. *Circ. Cardiovasc. Qual. Outcomes*. 2019, 12(7).

[8] Wang M, Xu C, Chen X, Hao H, Zhong L, *et al*. Differential privacy oriented distributed online learning for mobile social video prefetching. *IEEE Trans. Multimedia* 2019, 21(3):636–651.

[9] Wang T, Zheng Z, Bashir AK, Jolfaei A, Xu Y. FinPrivacy: A Privacy-Preserving Mechanismfor Fingerprint Identification. *ACM Trans. Internet Technol.* 2020. (In Press)

[10] Prasad KD, Reddy K, Vasumathi D. Privacy-Preserving Naive Bayesian Classifier for Continuous Data and Discrete Data. In *First International Conference on Artificial Intelligence and Cognitive Computing*, 2019, pp. 289–299.

[11] Li MH, Chow S, Hu SS, Yan YJ, Shen C, *et al*. Optimizing Privacy-Preserving Outsourced Convolutional Neural Network Predictions. *arXiv* 2020, arXiv:2002.10944.

[12] Phong LT, Aono Y, Hayashi T, Wang LH, Moriai SH. Privacy-preserving deep learning via additively homomorphic encryption. *IEEE Trans. Inf. Forensics Secur.* 2017, 13(5):1333–1345.

[13] Dwork C, Smith A, Steinke T. Exposed! a survey of attacks on private data. *Annu. Rev. Stat. Its Appl.* 2017, 4(1):61–84.

[14] Wu J, Chen L, Feng Y, Zheng Z, Zhou MC, *et al*. Predicting quality of service for selection by neighborhood-based collaborative filtering. *IEEE Trans. Syst. Man Cybern.* 2012, 43(2):428–439.

[15] Muraro Gularte K, Ruiz Vargas J, Javidi da Costa J, Santos da Silva A, Almeida Santos G, *et al.* Safeguarding the V2X Pathways: Exploring the Cybersecurity Landscape Through Systematic Review. *IEEE Access* 2024, 12:72871–72895

[16] Dai W, Daia C, Choo K, Cui C, Zou D, *et al*. SDTE: A Secure Blockchain-Based Data Trading Ecosystem. *IEEE Trans. Inf. Forensics Secur.* 2019, 15(99):725–737.

[17] Chen J, Wu J, Liang H, Mumtaz S, Li J, *et al*. Collaborative Trust Blockchain Based Unbiased Control Transfer Mechanism for Industrial Automation, *IEEE Trans. Ind. Appl.* 2019, 56(4):4478–4488.

[18] Huang H, Chen X, Wang J. Blockchain-based multiple groups data sharing with anonymity and traceability. *Sci. China Inf. Sci.* 2020, 63(3):130101.

[19] Huang L, Zhang G, Yu S, Fu A, Yearwood J. SeShare: Secure cloud data sharing based on blockchain and public auditing. *Concurr. Comput. Pract. Exp.* 2019, 31(22):1–15.

[20] Sajid A, Khalid B, Ali M, Mumtaz S, Qamar F. Securing Cognitive Radio Networks using blockchains. *Fut. Gener. Comput. Syst.* 2020, 33(10):1–10.

[21] Połap D, Srivastava G, Jolfaei A, Parizi R. Blockchain Technology and Neural Networks for the Internet of Medical Things. In Proc. *IEEE Conference on Computer Communications* (INFOCOM), 2020.

[22] Kosba A, Miller A, Shi E, Wen Z, Papamanthou C. Hawk: The blockchain model of cryptography and privacy-preserving smart contracts. In *IEEE symposium on security and privacy* (SP), 2016, pp. 839–858.

[23] Zheng B, Zhu L, Shen M, Gao F, Zhang C, *et al*. Scalable and privacy-preserving data sharing based on Blockchain. *J. Comput. Sci. Technol.* 2018, 33(3):557–567.

[24] Banu R, Nagaveni N. Evaluation of a perturbation-based technique for privacy preservation in a multi-party clustering scenario. *Inf. Sci.* 2013, 232:437–448.

[25] Alamer A. A secure and privacy blockchain-based data sharing scheme in mobile edge caching system. *Expert Syst. Appl.* 2023, 221:121572.

[26] Salah K, Alfalasi A, Alfalasi M. A Blockchain-based System for Online Consumer Reviews. In IEEE INFOCOM 2019-*IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2019, pp. 853–858.

[27] Jiang X, Zhang G, Tian H, Xiang H, Zhang X, *et al*. A Blockchain-enabled Secure Access Management Method in Edge Computing. *2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS)*, 2023, pp. 2596–2603.

[28] Liu J, Zhang X, Chen Y, Wang W, Wang L, *et al*. Conditional Anonymous Remote Healthcare Data Sharing Over Blockchain. *IEEE J. Biomed. Health Inform.* 2023, 27(5):2231–2242.

[29] Weng JS, Weng J, Li M, Zhang Y, Luo W. DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive. *IEEE Trans. Dependable Secure Comput.* 2019, 18:2438–2455.

[30] Shen M, Duan J, Zhu L, Zhang J, Du X, *et al*. Blockchain-Based Incentives for Secure and Collaborative Data Sharing in Multiple Clouds. *IEEE J. Sel. Areas Commun.* 2020, 38(6):1229–1241.

[31] Ren Y, Zhu F, Qi J, Wang J, Sangaiah A, *et al*. Identity management and access control based on blockchain under edge computing for the industrial Internet of Things. *Appl. Sci.* 2019, 9(10):2058.

[32] Lu Y, Tang Q, Wang G. Zebralancer: Private and anonymous crowdsourcing system atop open Blockchain. In Proc. *ICDCS*, 2018, pp. 853–865.

[33] Makhdoom I, Zhou I, Abolhasan M, Lipman J, Ni W. PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities. *Comput. Secur.* 2020, 88:101653.

[34] Chen C, Zhao H, Qiu T, Hou R, Sangaiah A. A multi-station block acknowledgment scheme in dense IoT networks. *Comput. Commun.* 2018, 119:179–190.

[35] Chen S, Fu X, Si H, Wang Y, Gao S, *et al*. Blockchain for Health IoT: A privacy-preserving data sharing system. *Softw. Pract. Exper.* 2022, 52(9):2026–2044.

[36] Shen M, Tang X, Zhu L, Du X, Guizani M. Privacy-preserving support vector machine training over blockchain-based encrypted IoT data in smart cities. *IEEE Internet Things J.* 2019, 6(5):7702–7712.

[37] Keshk M, Turnbull B, Moustafa N, Vatsalan D, Choo K. A Privacy-Preserving-Framework-Based Blockchain and Deep Learning for Protecting Smart Power Networks. *IEEE Trans. Ind. Inf.* 2019, 16(8):5110–5118.

[38] Lin C, He D, Zeadally S, Kumar N, Choo K. SecBCS: a secure and privacy-preserving blockchain-based crowdsourcing system. *Sci. China Inf. Sci.* 2020, 63(3):1–14.

[39] Wu L, Wu S, Zhou Y, Li R, Wang Z, *et al*. EthScope: A Transaction-centric Security Analytics Framework to Detect Malicious Smart Contracts on Ethereum. *arXiv* 2020, *arXiv*:2005.08278.

[40] Okegbile S, Cai J, Alfa A. Practical Byzantine Fault Tolerance-Enhanced Blockchain-Enabled Data Sharing System: Latency and Age of Data Package Analysis. *IEEE Trans. Mob. Comput*, 2024, 23(1):737–753.

[41] Cui J, Duan L, Ni W, Li C. Secure Cross-domain Medical Data Sharing based on Distributed Cloud and Blockchain Services. *2023 IEEE International Conference on Web Services (ICWS)*, Chicago, IL, USA, 2023, pp. 698–700.