A novel privacy protection method of residents' travel trajectories based on federated blockchain and InterPlanetary file systems in smart cities

Fenghan Liu[1] and Pan Wang[1]

[1]School of Modern Posts, Nanjing University of Posts&Telecommunications, Nanjing 210003, China

Corresponding author:
Pan Wang[1]

Email address: wangpan@njupt.edu.cn

ABSTRACT

The government does have to record and analyze the travel trajectories of urban residents aiming to effectively control the epidemic during COVID-19. However, these privacy-related data are usually stored in centralized cloud databases, which are prone to be vulnerable to cyber attacks leading to personal trajectory information leakage. In this paper, we proposed a novel secure sharing and storing method of personal travel trajectory data based on BC and InterPlanetary File System(IPFS). We adopt the Hyperledger Fabric, the representative of Federated BC framework, combined with the IPFS storage to form a novel mode of querying on-chain and storing off-chain aiming to both achieve the effectiveness of data processing and protect personal privacy-related information. This method firstly solves the efficiency problem of traditional public BC and ensures the security of stored data by storing the ciphertext of complete personal travel trajectory data in decentralized IPFS storage. Secondly, considering the huge amount of information of residents' travel trajectories, the method proposed in this paper can obtain the complete information under the chain stored in IPFS by querying the index on the chain, which significantly improves the data processing efficiency of residents' travel trajectories and thus promotes the effective control of the new crown pneumonia epidemic. Finally, the feasibility of the proposed solution is verified through performance evaluation and security analysis.

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an infectious disease that is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)(cor, 2020). The disease has spread into most nations across the globe, and there is still no specific drug for COVID-19. Non-pharmacological interventions (NPI), such as obtaining the travel trajectories of urban residents and regularly quarantining people who have visited outbreak areas, have been effective in reducing public exposure and thus slowing the spread of the disease(Bootsma and Ferguson, 2007), such as the travel codes used by the Chinese government, which have successfully reduced the number of cases and prevented the mass spread of COVID-19. So for today's city dwellers, trip codes (proof of personal trip trajectory) have become a daily travel pass.

However, the trip codes of urban residents contain a large amount of private personal information, including but not limited to identification information, and personal trip trajectories. These private data are usually stored in centralized databases with the following problems: 1. Inability to address centralized flaws such as server vulnerability to single point of failure SPoF (Single Point of Failure) and DDOS attacks: When the central server is subject to external attacks such as malware or brute force cracking attempts, data are at risk of loss and leakage. There have been several incidents of cyber violence against patients due to the leakage of COVID-19 patient travel track information. 2. Lack of trusted partnerships: Since resident travel trajectory data contains a lot of private information, various departments are reluctant to lose control of this high-value data leading to difficulties in sharing data between different organizations

⁴⁵ in a timely manner, thus greatly reducing the efficiency of cross-departmental collaborative epidemic
⁴⁶ prevention. 3. Transparency issues: The data management and control currently used usually lack
⁴⁷ transparency, and there is a possibility of misusing the data if a malicious user gets hold of the high-value
⁴⁸ resident travel trajectory data.

⁴⁹   Blockchain, as a typical privacy computing method (LU Haitao and ZHOU Qiang, 2022), offers the
⁵⁰ possibility to solve the above problems. Blockchain, as a decentralized distributed ledger, abandons third-
⁵¹ party management, and data are stored on all nodes on the blockchain, and each node jointly supervises
⁵² and maintains this distributed ledger, so the risk of data tampering and loss is almost non-existent. To
⁵³ democratize control of the blockchain, the blockchain provides decentralized, transparent databases that
⁵⁴ give all participants equal access to the stored data and seeks consensus from all participants before
⁵⁵ agreeing to any changes. This decentralized trust model provides a trustworthy partnership among the
⁵⁶ participants to be willing to share data.

⁵⁷   According to different access permissions, blockchain can be divided into the public chain, private
⁵⁸ chain, and alliance chain(Dib et al., 2018). All nodes on the public chain are open to the public, and all
⁵⁹ people can read and write on the blockchain and get the complete blockchain data, and it is difficult to
⁶⁰ reach a consensus for these nodes that come and go at will. Therefore, public chains are limited by the
⁶¹ consensus mechanism and are prone to low throughput rates, and are difficult to scale when applied to
⁶² application scenarios with large amounts of data. For example, the Bitcoin system can process up to 7
⁶³ transactions per second and the Ether platform can process up to 25 transactions per second(Xie et al.,
⁶⁴ 2019). A private chain, on the other hand, completely binds the operation authority to an organization, so
⁶⁵ that data can only be managed and traced within a single organization, which improves the speed of data
⁶⁶ circulation, but its closed nature hinders the sharing of data among multiple parties. A Federated chain is
⁶⁷ a blockchain managed by several organizations that jointly participate, each running one or more nodes,
⁶⁸ in which the data are only allowed to be read and sent by different organizations in the system and jointly
⁶⁹ record the transaction data. The nodes of the coalition chain have a high degree of trust, they can use a
⁷⁰ relatively loose consensus mechanism, so the data processing speed will be greatly improved than that of
⁷¹ the public chain, solving the problem of low efficiency of the public chain, at the same time, the data can
⁷² be shared among the associated nodes, solving the problem of poor data circulation of the private chain.

⁷³   The amount of travel trajectory data generated by city residents every day is huge, and due to the
⁷⁴ limitation of block size in the blockchain and the characteristic that data can only be added but not
⁷⁵ deleted, the blockchain ledger will keep expanding, so it is unrealistic for the data to be completely
⁷⁶ on the chain. In response to this problem, many scholars have proposed a combination of blockchain
⁷⁷ and cloud storage to store the original data under the chain, thus relieving the storage pressure on the
⁷⁸ blockchain. Reference(ZHAO Tian et al., 2021) presents a proposed blockchain based duplicate data
⁷⁹ elimination system for cloud data discrepancy authorization, which optimizes the traditional proof-of-vote
⁸⁰ (PoV) consensus algorithm, simplifies the existing discrepancy authorization process, and enables trusted
⁸¹ management and dynamic update of permissions. However, due to the lack of effective encryption
⁸² algorithms, the confidentiality of the data cannot be guaranteed. (El Ghazouani et al., 2019) proposed
⁸³ a storage model based on blockchain and cloud for medical data management, which uses on-chain
⁸⁴ storage of core data and off-chain storage of real data to ensure the integrity and security of medical
⁸⁵ data and take access control to ensure the security of data, but it cannot achieve efficient sharing of data.
⁸⁶ (Zheng et al., 2018) proposes a solution that combines blockchain, cloud storage, and machine learning
⁸⁷ to achieve medical data sharing, but its complete medical data is stored in the centralized servers of
⁸⁸ the third party under the chain, and there is still a risk of data leakage if the third party is threatened.
⁸⁹ An interplanetary File System (IPFS) is a decentralized, distributed storage system based on content
⁹⁰ addressing. After depositing data into the IPFS system, the system returns a hash value derived based
⁹¹ on that data. According to the principle of the hash function, even if only slight changes are made to
⁹² the information, a completely different hash value will be obtained. This feature will ensure that data
⁹³ on IPFS will not be tampered with, ensuring the security of the data under the chain. So (Wu and Du,
⁹⁴ 2019) proposed a model for the secure sharing of electronic medical records using blockchain combined
⁹⁵ with IPFS to store the complete electronic case data in the IPFS system, but the model uses Data Masking
⁹⁶ to partially generalize the electronic case privacy information that needs to be deposited into the IPFS
⁹⁷ system, which is not secure enough. We need a more secure method to protect the privacy of urban
⁹⁸ residents' travel trajectories.

⁹⁹   Based on the above analysis, the goal of this paper is to protect the security of urban residents' travel

trajectory data while achieving efficient data sharing among departments fighting against the epidemic, so a new approach to protect the privacy of smart urban residents' travel trajectory based on federated blockchain and the interstellar file system is proposed. The key contributions are summarized as follows:

1. Designed a Hyperledger Fabric federated chain framework combined with an IPFS distributed storage model for data sharing while ensuring the security of city residents' travel trajectories. Hyperledger Fabric is a pre-specified party-based permissioned blockchain that allows data sharing without relying on a single permission, while selective disclosure of data can better protect the privacy of sensitive data such as residents' trip trajectories. Secondly, the use of IPFS to store the complete resident travel trajectory data reduces the storage pressure in the blockchain system and ensures the security of the data under the chain.

2. We propose an on-chain lookup and off-chain storage model, in which we store the complete AES-encrypted resident travel traces in IPFS, combine the returned IPFS storage address, AES key, Location_ID, and file's creation time to form an index, and then store the index on the blockchain. Users can query the required data by Location_ID and file's creation time. In this way, it is possible to ensure that personal privacy information will not be leaked and to search all the travel data sent and received and the corresponding transaction details through the index in the massive travel track data, thus improving the efficiency of epidemic prevention and control.

3. The security of our proposed approach is analyzed in detail and the performance of our approach (including data processing throughput and latency) is evaluated using the performance testing tool Hyperledger Caliper.

RELATED WORK

The trajectory data of urban residents can usually reflect the living habits of residents, and there have been related research works that if an attacker analyzes the user's trajectory data, he can not only obtain the user's home address, workplace, and hobbies, but also further infer the user's interest preferences, travel patterns, social habits, and other private information, and the attacker can even use the spatial and temporal correlation in the user's location samples to infer other location information on the user's trajectory(De Montjoye et al., 2015). In the context of the uncontrolled spread of COVID-19, the normal life and work of urban residents have been greatly affected by it. The prevention and control department has taken effective measures to prevent the further spread of COVID-19 by acquiring the travel trajectories of urban residents, while it is important to protect the privacy and security of their travel trajectory data.

Blockchain-based privacy protection for data in smart cities

Blockchain technology can provide data immutability to ensure the authenticity and accuracy of data in smart city management. Considering that users in smart cities require a wide range of services on a daily basis, which leads to a large amount of user personal information being collected and controlled by service providers, (Mohammadinejad and Mohammadhoseini, 2020) introduces an automatic access control manager based on blockchain networks without the presence of a third party, creating a secure channel between users and services and the possibility of their databases being misused or hacked.(She et al., 2019) proposed a homomorphic federated chain for SHS sensitive data privacy protection (HCB-SDPP) based on traditional smart home systems. The model designs a homomorphic encryption algorithm based on pallier encryption to encrypt smart home data, and uploads sensitive data from encrypted all gateway nodes to the federated chain to ensure the security of sensitive data. The convergence of healthcare and smart cities can enable the use of health and medical data from around the world, which also poses challenges to the privacy of patients' health information and the security of nearby mobile health users. (Tripathi et al., 2019) proposes a blockchain-based smart healthcare system framework to provide the inherent security and integrity of the system and ensure that the privacy of patient data is not compromised, (Zhang and Lin, 2018) proposed a PHI (Private Health Insurance) data privacy protection scheme combining federated and private chains. In this scheme private blockchain is used to store the PHI of hospitals and the federated blockchain is responsible for recording the secure index of PHI and using public key encryption and keyword search to achieve data security and privacy protection on the federated blockchain. The Internet of Things (IoT), as a gas pedal for smart cities, requires the extensive use of distributed smart devices to collect and process data in smart city infrastructures. (Kumar et al., 2021) proposes a specialized protection and security framework (ppsf) based on blockchain, which supports Internet-driven smart cities through two key mechanisms: privacy protection mechanism and intrusion detection mechanism.

IPFS in Data Privacy Protection

IPFS, with its high security and efficiency, has also been heavily used in recent years in the area of data privacy protection. (Liang et al., 2022) and (Song et al., 2022) use an improved homomorphic encryption mechanism and SM4 homegrown commercial security algorithm to encrypt raw data, proposing a federal blockchain-based privacy protection scheme for personal data. The federated blockchain combined with IPFS to store encrypted data improves data transfer efficiency, safeguards user privacy and security, and proposes ciphertext policy attribute-based encryption to achieve fine-grained access control. In the IoT research, (Hasan et al., 2022) proposes an advanced security model for IoT multimedia data sharing through blockchain combined with IPFS, which solves the security issues brought by centralized data storage.In the IoT research, (Dhar et al., 2022) proposes an advanced security model for IoT multimedia data sharing through blockchain combined with IPFS, which solves the security issues brought by centralized data storage. (Kumar and Tripathi, 2021) proposes a federated blockchain network with smart contract support to ensure security and privacy in the Internet of Medical Things (IoMT). The network integrates Interplanetary File System (IPFS) cluster nodes to secure and authenticate devices and provide secure storage management through IPFS cluster nodes. (Gupta et al., 2020) proposed an approach called AaYusH (Ethernet Smart Contract (ESC) and IPFS-based TS system) to provide technical support for telesurgery systems, in which IPFS effectively reduces storage costs and provides reliable privacy protection of system data. In order to alleviate the problems of limited scalability and poor storage scalability brought by blockchain, which lead to the inability to directly combine blockchain with IoT under existing conditions. (Li et al., 2022) proposes a three-tier architecture blockchain (TBchain) architecture combined with IPFS to provide high performance privacy protection for IoT data stored on the blockchain. In the (Dwivedi et al., 2021), IPFS is combined with blockchain to achieve car insurance driving data sharing for traffic congestion mitigation by applying it to a vehicle-based self-organizing network (VANET). In this scheme, the decentralized characteristics of blockchain and IPFS are used to achieve fully distributed storage of data and ensure the security of vehicle driving data.Focusing on EHR privacy protection in healthcare management, (Jayabalan and Jeyanthi, 2022) integrates the blockchain framework with the Interplanetary File System (IPFS) to enable healthcare organizations to maintain fail-safe and tamper-proof healthcare ledgers in a decentralized manner.Meanwhile, (Lin and Zhang, 2021) proposes a privacy protection method based on blockchain and improved IPFS to alleviate data privacy issues in IPFS by adding a data access mechanism to prevent potential privacy breaches during IPFS transmission.

Inspired by the above-related work, this paper proposes a privacy protection method for smart city residents' travel trajectories based on federated blockchain and interstellar file system, which assists government departments in effectively controlling the spread of epidemics while ensuring the security of city residents' trajectory data.

PRELIMINARIES

Residents Travel trajectories data

A resident trajectory is a collection of data formed by a resident visiting a location at a certain time. When a resident visits a location at a certain time, the outbreak prevention and control department can obtain the resident's geographic location in real-time through official location service providers including but not limited to GNSS, Bluetooth, cell towers, and WiFi.

As shown in Figure 1, (l_1,t_1) means that the resident visited location l_1 at time t_1, after which all the collected location information of the resident in one day is integrated to form the travel track data of the resident on this day. The complete resident travel trajectory data includes information such as personal name, cell phone number, file's creation time, names of all visited locations, and specific geographic locations of visited locations. For the convenience of discussion, the travel trajectory data is simplified into three parts: user ID, visit location and visit time in this paper. For example, the travel trajectories of residents U_1, U_2, and U_3 for one day are shown in Figure 1, then the travel trajectory of U_1 for that day is expressed as $U_1=\langle (l_1,t_1),(l_3,t_2),(l_2,t_3),(l_4,t_4)\rangle$, the travel trajectory of U_2 is expressed as $U_2=\langle (l_1,t_1),(l_5,t_2),(l_6,t_3)\rangle$, and $U_3=\langle (l_4,t_1),(l_2,t_2),(l_3,t_3),(l_6,t_4)\rangle$. The local epidemic prevention and control department collects the travel trajectories of all residents living in the area in one day, forming the travel trajectory data RTT= $\{U_1, U_2, U_3 \}$ of the residents in the area on this day.

Location	Sign	visit time
Restaurant(l_1)	🍴	t_1:U_1, U_2;
Bank(l_2)	🏦	t_3:U_1, U_3
Company(l_3)	🏢	t_2:U_1,; t_3:U_3
Gym(l_4)	🏋	t_1:U_3; t_4:U_1
School(l_5)	🎓	t_2:U_2
Hospital(l_6)	✚	t_3: U_2; t_4:U_3
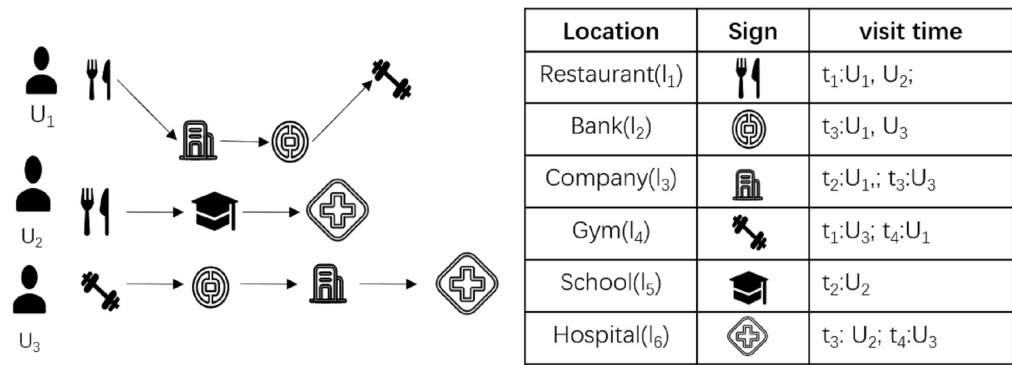
Figure 1. Residents travel trajectories

Hyperledger Fabric

Hyperledger Fabric is a distributed ledger solution platform based on a modular architecture that provides a high degree of confidentiality, resilience, flexibility and scalability. Hyperledger Fabric is a non-fully public blockchain that helps protect the privacy of residential travel data, which is often highly sensitive. Hyperledger Fabric does not require expensive mining calculations to submit transactions through pre-specified membership and strict access policies, so it helps build blockchains that can scale with shorter latency of the blockchain. To ensure the security of communication, Hyperledger Fabric supports TLS for transport layer security for communication between nodes, and the ledger is only shared among organizations participating in the common bookkeeping, while other organizations do not have access to the ledger, setting strict permission control for accessing resources. In terms of privacy protection Hyperledger Fabric implements a private data mechanism to solve the problem of allowing certain private data to be shared among only a small number of organizations in a ledger with multiple participating organizations.

Interplanetary file system(IPFS)

IPFS is a distributed file storage system that assigns unique storage addresses to files uploaded to the system and the hash addresses of IPFS files cannot be changed, so IPFS is a tamper-proof storage mechanism. It differs from the HTTP protocol in that the data is stored in a distributed manner in the system built on IPFS, whereas the HTTP protocol places the data on the server side, which puts tremendous pressure on the server. IPFS addressing is based on the principle of content addressing, and when data is stored in the IPFS system, the system returns a hash value based on that data. According to the principle of the hash function, even if only minor changes are made to the information, a completely different hash value will be obtained. This feature will ensure that data on IPFS will not be tampered with. When requesting data from the IPFS file system, it will use the distributed hash table to find the node where the requested data is located and request the data back in that node. these features of IPFS make the perfect combination of blockchain technology at the application level to provide a distributed storage solution for traditional blockchain programs.

METHODS

In this section, we will first explain the framework of our proposed method and related entities and terms, and then describe in detail the concrete implementation process of the privacy protection method of smart city residents' travel trajectories based on federated blockchain and interstellar file system.

Entities and Nouns

The system model framework of the proposed approach in this paper is shown in Figure 2, and the entities and terms involved in the system framework diagram are explained next in the following:

1. PKI/CA: public key infrastructure (PKI)/certification authority (CA) is a trusted third party (in this system, it is a government department) that generates a CID corresponding to the identity of
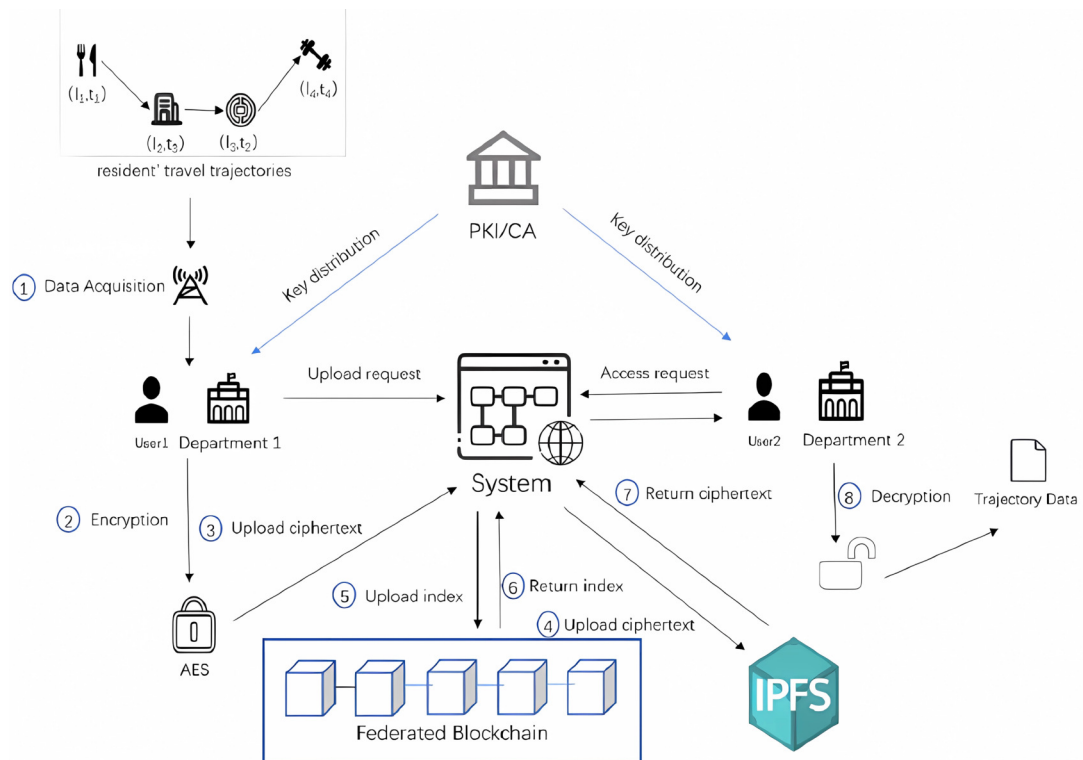
Figure 2. System architecture

each user who joins the system. Based on the CID, the attribute authority sends the identity-related attribute set and key to the user.

2. Departments: Departments are the epidemic prevention and control departments in different regions. The epidemic prevention and control departments in each region collect the travel data of residents through the base stations and WIFI connected to the residents' cell phone signals and share the trajectory data with the epidemic prevention and control departments in other regions in the system.

3. Users: Users in this system refer to the staff of the epidemic prevention and control department, who is responsible for uploading, querying, and downloading the specific operations of the residents' travel trajectory data.

4. Block_Index: To relieve the storage pressure of the blockchain and ensure the security of the travel track data, the system only stores the easily searchable Block_Index on the blockchain. Block_Index consists of the storage address of residents travel trajectory in IPFS, AES encryption key, Location_ID, and CreateTime. The storage address of IPFS and the AES key is encrypted by the public key of the data recipient, and the geographic location of the resident's travel track data and the CreateTime are not encrypted for the convenience of the query.

5. IPFS: Due to the practical limitations of cost, storage capacity, and other factors, a large amount of residents' travel track data is stored encrypted outside the blockchain. IPFS can assist the blockchain to store the cipher text of residents' travel track uploaded from the client by the staff of the epidemic prevention and control department to reduce the storage pressure of the blockchain.

6. Federated Blockchain: Blockchain can be considered a secure distributed database ledger. The federated blockchain network in this system consists of different regional epidemic control authorities, which store the encrypted index on the blockchain to ensure security.

Method Implementation

In general, our method uses AES encryption for residents' travel trajectory data in advance, uploads the encrypted data to IPFS network and obtains the corresponding storage address in order to relieve the storage pressure of blockchain, and stores the address and AES key, Location_ID and file creation time to form an index to the federated blockchain network, through which users can query the data. At the same

267 time, through the private data mechanism of the federated blockchain, it achieves that the travel trajectory
268 data is only shared among a few organizations in the ledger of the participating organizations to exclude
269 the risk of privacy leakage. The proposed approach is described in detail in the following four aspects:
270 data encryption, data upload, data sharing, and data query.To simplify the description, the meaning of
     some special characters are shown as Table 1.

Notations	Descriptions
RTT	Residents travel trajectories
CRTT	Encrypting residents' travel trajectory data
K	AES key
AES_b, AES_d	AES encryption function
PK_R, PK_S	Receiver's public key and Sender's public key
SK_R, SK_S	Receiver's private key and Sender's private key
$Hash_{CRTT}$	Hash address of resident travel trajectory data stored in IPFS

Table 1. Notations description

271
272 Staff U of the epidemic prevention and control department sends an application to the system, the
273 PKI/CA generates an ID corresponding to its identity, and according to the ID, the attribute authorization
274 authority sends the identity-related attribute set S and the public key PK_{id} and private key SK_{id} to the staff.
275 As shown in Step 1 of Figure 2, the local epidemic prevention and control department cooperates with
276 official operators to obtain the location information of residents in real-time through the base stations
277 accessed by residents' cell phone signals and connected WIFI and integrates them to obtain the complete
278 travel trajectory data RTT of all residents in the area (a detailed description of travel trajectory data can be
279 found in 3.1).

280 ***AES encryption***

AES encryption algorithm is a type of symmetric encryption algorithm, which has the advantages of high
security, high efficiency and resistance to traditional password cracking. To ensure that no privacy leakage
occurs during the uploading of residents' travel track data to IPFS, the method proposed in this paper
preemptively encrypts the travel track data to be uploaded to IPFS with AES encryption. Assuming that
the staff of an epidemic prevention and control department in a region uses AES encryption function as
AES_b, the encryption generates encrypted travel track data CRTT:

$$CRTT = AES_b(K, RTT) \tag{1}$$

281 ***Identity verification***

282 Algorithm 1 shows the authentication of the epidemic prevention and control authorities in each region
283 of the federated network for the upload and sharing of resident travel track data. The purpose of this
284 algorithm is to prevent malicious nodes from accessing the federated network and ensures the security of
285 the proposed model. By joining the federated blockchain network, they receive a unique proof of identity,
286 known as a Proof of Identity (PoI). In the Fabric network, the user provides a CA-generated certificate
287 representing his identity and the MSP determines whether the user is trustworthy by identifying him or
     issuing him with a CA for his identity.

Algorithm 1: Algorithm for peer nodes verification in consortium network

Input: input Proof-of-Identity (PoI)
Output: output verification of peer nodes

1 **if** *msg.sender is not valid* **then**
2   Return Invalid peer node;
3 **else**
4   Peer node can upload/share/query/download RTT;
5   Return Valid peer node;
6 **end**

288

289 ***Data Upload***

The process of uploading residents' travel track data is shown in steps 3, 4, and 5 of Figure 2, RTT is pre-encrypted with AES algorithm to form CRTT, and the staff U_S who uploads the data uses their private key SK_S to generate the signature $SigSK_S$(CRTT), and then sends the ciphertext CRTT, POI, and signature $SigSK_S$(CRTT) together to the IPFS node. Algorithm 2 describes the storage mode in which user data is uploaded to the system to form an on-chain storage index and complete data is stored under the chain: After receiving the ciphertext and signature uploaded by the user, the IPFS node verifies the identity of the user. After confirming the identity, the CRTT is uploaded to the IPFS network for storage. At the same time, the IPFS returns the unique hash address $Hash_{CRTT}$ based on the uploaded ciphertext. Each node in the IPFS network uses the public key to sign the received hash address and sends CRTT, $Hash_{CRTT}$, and $SigSK_S$(CRTT) to each IPFS node in the cluster for verification. Each IPFS node in the cluster verifies the user's identity and calculates the CRTT hash locally. If it matches the hash value returned by the IPFS network, a confirmation message is sent to the IPFS node interacting with the user, and the complete resident travel track ciphertext is stored in the IPFS. The stored address $Hash_{CRTT}$ returned by IPFS, AES key K, Location ID where the resident travel track data is located, and file's creation time are formed into an index and encrypted using the public key PK_R of the data recipient i.e. Enc ($Hash_{CRTT}$, K, PK_R) to form an encrypted index Block_Index:

$$Block\_Index = < Enc(Hash_{CRTT}, K, PK_R), Location\_ID, CreateTime > \qquad (2)$$

290 Then Block_Index is sent to the federated blockchain network, the Fabric node receives the relevant
291 information and sends it to other Fabric nodes to jointly verify its identity and content, deploys the
292 transaction after successful verification, and sends it to each node in the blockchain federation in the
293 node group, and the transaction is packaged on the chain after each node reaches consensus and the
Block_Index is stored on the federated blockchain.

Algorithm 2: Data upload

Input: input CRTT,Proof-of-Identity(PoI),K,location_ID,CreateTime
Output: output Algorithm result

1 **if** *Peer node is valid and Each IPFS node in the cluster verifies the identity successfully* **then**
2   CRTT Upload to IPFS;
3   Return $Hash_{CRTT}$;
4   **if** *Consensus nodes reach consensus* **then**
5   Block_Index Upload to Blockchain;
6   Return $Address_{Index}$;
7   **else**
8   Return No consensus;
9   **end**
10 **else**
11   Return Invalid identity;
12 **end**

294

295 ***Data Sharing***

The proposed method uses physical deployment as shown in Figure 3 to share data among each other. The figure mainly includes Fabric nodes, IPFS nodes, WEB Service, and PostgreSQL database for system functions and interactions, and each epidemic prevention and control department joining the network maintains two types of nodes, Fabric nodes, and IPFS nodes. In this paper, the epidemic prevention and control department maintains the endorsement node and the orderer node in the Fabric node in the system. Epidemic prevention and control departments in the network store the index data of residents' travel trajectories at the endorser node, the endorser node runs the smart contract and endorses the result of smart contract execution and stores the data, then the ordering node collects the data and packages the transactions that satisfy the endorsement strategy into blocks to be distributed to each epidemic prevention and control department in the blockchain network. The IPFS cluster consisting of IPFS nodes runs between the epidemic prevention and control departments to store the cipher text of residents' travel track data after

AES encryption to reduce the pressure of blockchain storage. Finally, the WEB server is responsible for communicating with the fabric nodes, IPFS nodes, and business database servers to connect all epidemic prevention and control departments, break the data silos, and realize the sharing of residents' travel track data among all epidemic prevention and control departments, to achieve the purpose of collaborative prevention and control of the epidemic by all departments while effectively protecting the privacy and security of residents' travel track data. Algorithm 3 is the specific implementation of data access in the
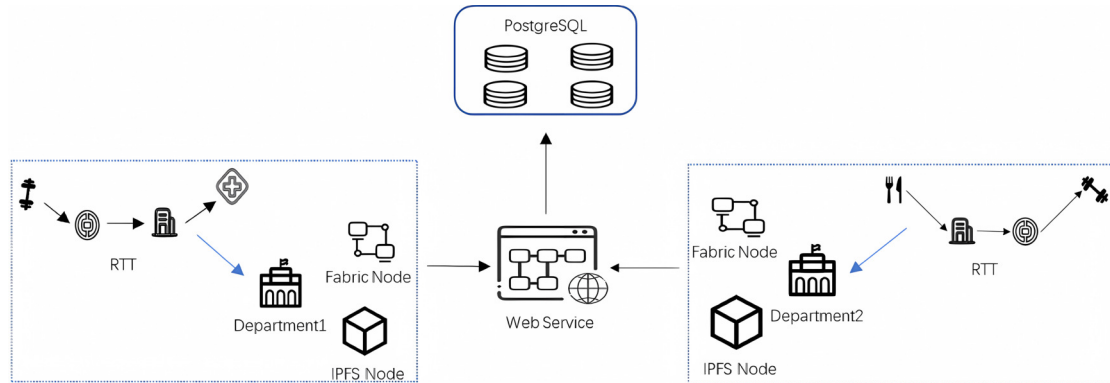
Figure 3. Physical deployment architecture

data-sharing process, the user sends a request to access the data, and the user first authenticates through the POI, and the corresponding Block_Index of the CRTT to be accessed to verify whether the correct trajectory data is being accessed. If the above verification is passed, the user decrypts Block_Index with his private key SK_R and obtains the storage address, AES encryption key, Location_ID, and CreateTime returned by IPFS. Through the returned IPFS storage address, the IPFS network is accessed, and the nodes in the IPFS node group first verify the user's identity and then retrieve the complete resident travel track encrypted data at the corresponding storage location in the system's private IPFS network to send to the user after the verification is passed. After receiving the encrypted resident travel track data, the user decrypts the cipher text by using the AES key obtained when decrypting the Block_Index. Assuming that the AES decryption function used by the staff of a regional epidemic prevention and control department is AES_d and has AES key K, the user obtains RTT by decrypting.

$$RTT = AES_d(K, CRTT). \tag{3}$$

296 RTT is the travel track data of the residents in that geographical location, and the user can download
297 the data for use.

Algorithm 3: Data Access

Input: input Proof-of-Identity(PoI),Block_Index,SK_R
Output: output Algorithm result

1 **if** *PoI is valid and SK_R is right SK* **then**
2   Get $Hash_{CRTT}$,K from Block_Index;
3   M = $Hash_{CRTT}$;
4   N = Get Hash Value from IPFS;
5   **if** *(M == N) and K is right AES_Key* **then**
6   Return RTT;
7   **else**
8   Return Data access failed;
9   **end**
10 **else**
11   Return Access failed;
12 **end**

---

298 ***Data query***

299 The specific process of data encryption, index construction, and data query by users is shown in Figure 4. When the user sends a data query request, the user's identity is first verified through the POI. Algorithm 4
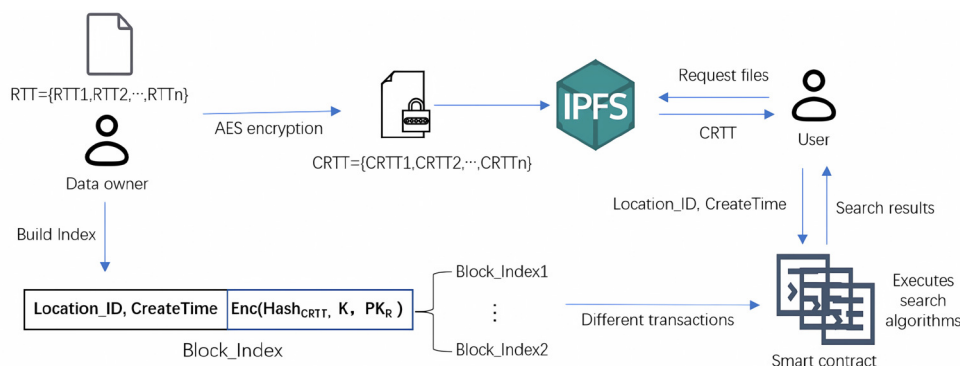
Figure 4. Processes of file encryption, index construction and search

300
301 is the query process of the encrypted index of the residents' travel trajectory data. After the user passes
302 the authentication, he enters his private key SK_q, and the system returns the encrypted index of all the data
303 related to the user through the smart contract deployed for query within the system, which includes the
304 $CRTT_s$ of encrypted residents' travel trajectory data sent by the user to other epidemic prevention and
305 control departments and the $CRTT_R$ of encrypted residents' travel trajectory data received by the user.
306 First, the user receives the Block_Index of these data returned by the system, and the user queries the
307 Block_Index of the travel trajectory data he wants to get by Location_ID and CreateTime, and if the query
308 is successful, the user gets the data index $Block\_Index_{querry}$. The user decrypts $Block\_Index_{querry}$ by SK_q
309 to obtain the storage address, AES encryption key, geographic location of resident travel track data, and
310 CreateTime returned by IPFS, and finally obtain the complete residents travel track data by the algorithm3.

Algorithm 4: Block_Index Query

Input: input Proof-of-Identity (PoI),Location_ID,CreateTime,SK_q
Output: output Algorithm result

1 **if** *Peer node is Valid and SK_q is right SK)* **then**
2   BOD = Get Block_Indexs of $CRTT_s$ and $CRTT_R$;
3   **if** *Location_ID and CreateTime are in the BOD* **then**
4   Return $Block\_Index_{querry}$;
5   **else**
6   Return Query failed;
7   **end**
8 **else**
9   Return Access failed;
10 **end**

311

RESULTS AND ANALYSIS
312

Security Analysis
313

314 Our proposed method ensures the integrity and privacy of residents' travel trajectory data.

315 ***Integrity***

316 The proposed approach in this paper uses a combined blockchain and IPFS storage structure, where the
317 user integrates Block_Index into the blockchain transactions and stores CRTT within IPFS. As shown in
318 Figure 5 either changes to Block_Index or CRTT will result in a change in the hash value, thus ensuring
319 that the stored data is immutable and traceable. If an illegal user wants to modify the hash of the current

320 data stored in the blockchain system at the same time, it must mimic the master chain like the source
321 chain so that the corresponding transaction can be accepted by most nodes, which is almost impossible to
322 achieve due to the limitation of computing power. According to the hash rule, the hash values obtained
323 by hashing two different files are different; therefore, the forged CRTT and Block_Index will not be
324 shared and accessible. For tampering attacks, the data stored in the blockchain is immutable unless the
blockchain is threatened by a 51% attack.
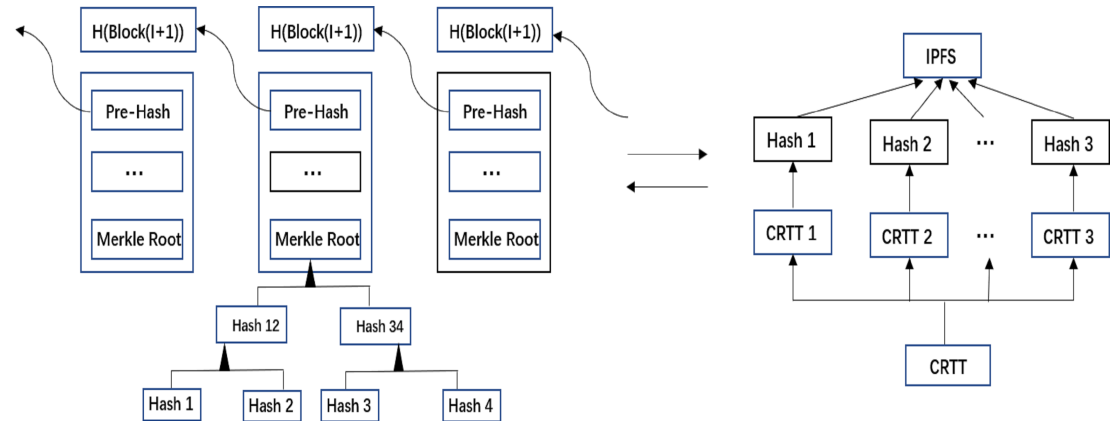
Figure 5. Storage Structure

325

326 *Privacy*
327 The proposed approach in this paper ensures the privacy of RTT from two aspects: 1.Pre-encrypted
328 RTT using AES algorithm to form CRTT stored in IPFS, without the key K and $Hash_{CRTT}$ is unable to
329 access the CRTT. And the Block_Index containing K and $Hash_{CRTT}$ is stored in the blockchain system,
330 and the index cipher cannot be decrypted without the private key SK_R of the data receiver. This ensures
331 that only the designated data receiver can access the resident travel trajectory data sent out by the user2.
332 As shown in Figure 6, the Hyperledger Fabric federated chain architecture used in this paper has a
333 unique channel mechanism, and the data in different channels are isolated from each other and cannot
334 be accessed arbitrarily, which means that all the data in Channel1 can only be accessed by Department1
335 and Department3. This enables the data sharing of residents' travel trajectory data only among related
epidemic prevention and control departments, ensuring the privacy of the data.
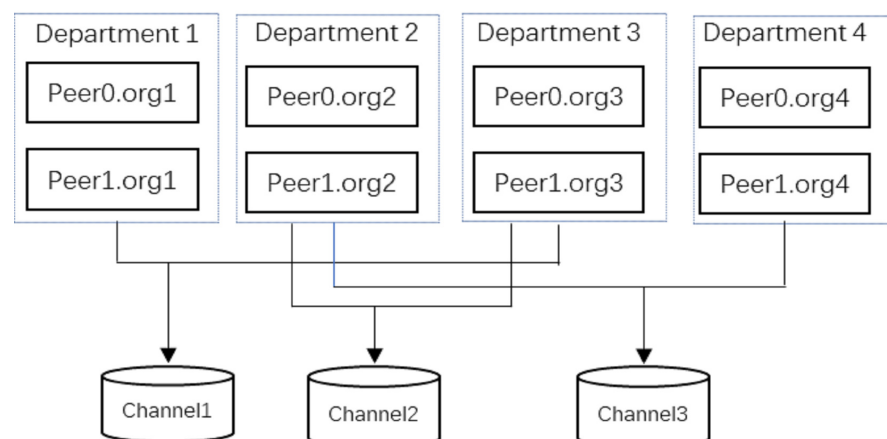
Figure 6. Fabric channel mechanism

336

Performance Analysis
338 Our system is deployed on a virtual machine with a 2-core CPU, 4GB RAM and Ubuntu 18.04 operating
339 system, and we use the travel trajectory data of COVID-19 infected persons in Beijing, China in 2019,

which is legally disclosed, as our test dataset. The performance testing consists of two main parts: 1. We first tested the upload/download time of files on IPFS and compared it with traditional cloud storage regarding the upload speed. 2. Querying and updating in smart contracts are the key operations of this system, and we used the Hyperledger Caliper tool to test the performance performance of system queries and updates regarding the throughput rate and latency. The proposed method refers to the uploading of encrypted index of residential travel trajectory data and the query operation refers to the querying of encrypted index of residential travel trajectory data. We used Hyperledger Caliper tool to test the performance of system queries and updates in terms of throughput and latency.

In the first experiment, we conducted an IPFS storage performance test, we selected files of sizes 1MB, 5MB, 10MB, 50MB, 100MB, and 500MB respectively, and tested the time required to upload files to IPFS and the time required to download files from IPFS, the test results are shown in Figure 7(a), from the figure we can find that the time required to upload files to IPFS is several times longer than the time required to download the file from IPFS. To prove that IPFS has a faster transfer speed, we compared the upload speed with traditional cloud storage. From Figure 7(b), we can see that when the file size is less than 10MB, the upload speed of cloud storage is slightly faster than IPFS, but as the file size keeps getting larger, the upload speed of IPFS far exceeds that of cloud storage, and we have reason to believe that IPFS has better performance than cloud storage when transferring massive files.
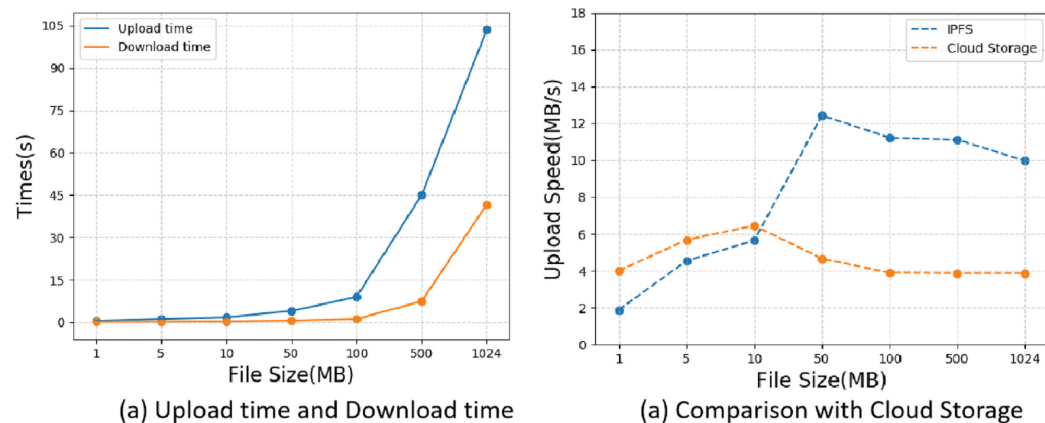
Figure 7. IPFS Performance Testing

Transaction latency is the average time elapsed from the time a transaction is submitted to the time it is added to the ledger and confirmed on the blockchain, and throughput is the number of valid transactions
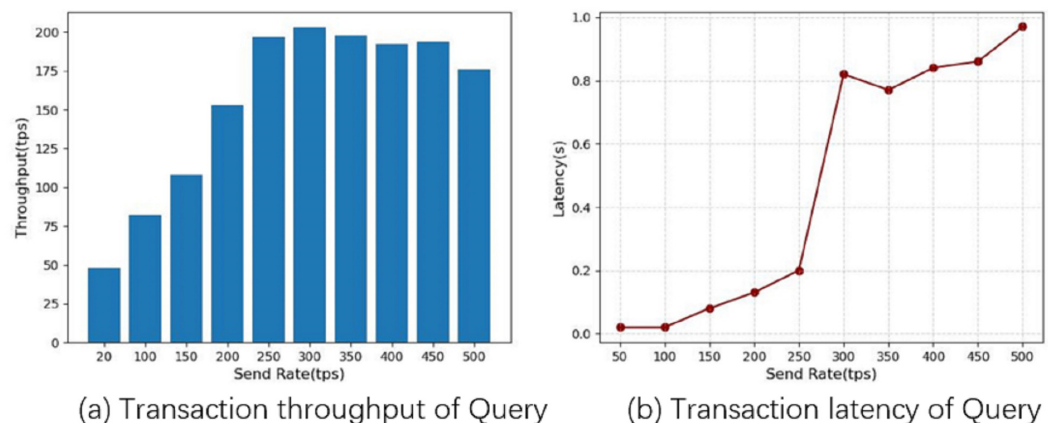
Figure 8. Query operational performance

submitted per second (tps) on the blockchain in the second experiment, we fixed the number of transactions at 5000 and obtained the average latency and throughput of the system by varying the Send Rate, and used this to analyze the performance of the system. We first tested the system query function for 10 rounds, increasing the Send Rate by 50tps each time until 500tps. Figure 8 shows the performance of the system query operation, from which the relationship between the performance of the chain code query function and the frequency of sending can be seen. From Figure 8(a) and Figure 8(b), it can be seen that as the Send Rate value increases, both the throughput and the average latency increase. The system throughput reaches a maximum of 203tps when the send rate reaches 300tps, and the average delay increases with the send rate, and the delay increases significantly above 250tps, indicating that the system can handle more than 250tps, leading to transaction blocking.

Similarly, we test the system update function by increasing the Send Rate by 25tps each time until 200tps: Figure 8 shows the performance of the system update operation. From Figure 9(a), we can see that as the send rate increases, the throughput rate also increases, and the system throughput reaches a maximum of 119tps when the send rate reaches 150tps, after which the system throughput decreases when the send rate increases further, and from Figure 9(b), we can see that the system average latency increases sharply when the Send Rate exceeds 125tps. This situation indicates that the processing capacity of the fabric node reaches its maximum at around 125tps.

(a) Transaction throughput of Update   (b) Transaction latency of Update
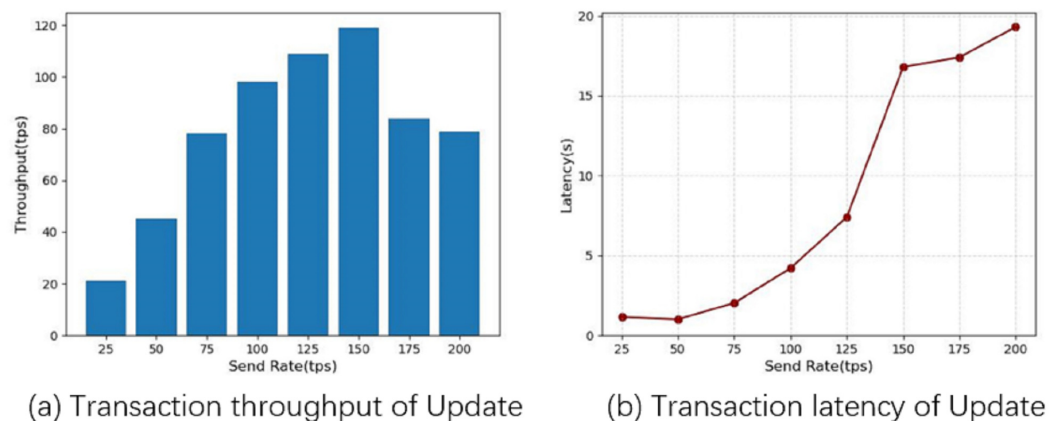
Figure 9. Update operational performance

Finally, we measure the performance metrics as the number of transactions to be processed by the blockchain increases from 200tx to 2000tx. As shown in Figure 10, we find that increasing the number of transactions leads to a linear increase in throughput, which starts to gradually decrease and flatten out when the number of transactions increases to 1200tx, which is the saturation point determined by the capacity of the nodes used in the test. In Figure 11 the response time of the system query operation versus the update operation is shown when the number of transactions goes from 200tx to 2000tx, we can see that in the system the update operation takes more time than the query operation, at 2000tx it takes about 19s, but considering the performance of virtual machines we used for testing, we consider this to be within acceptable limits.

DISCUSSION

In this section, we discuss existing approaches to the privacy protection of trajectory data and compare them with our proposed approach. The currently existing trajectory data privacy protection techniques are as follows:(1). Protecting trajectory data privacy based on suppression techniques: (Chen et al., 2013) distinguishes the sensitivity of the trajectory data in the original trajectory dataset and selectively publishes it. That is, when releasing trajectory data, the locations that do not involve sensitive information are released directly, and the sensitive locations on the trajectory are not released or are released after certain anonymization processes are performed on them. (2). Based on the k-anonymity model to protect user trajectory data: (Chow et al., 2011) constructs the hidden region satisfying k-anonymity by peer-to-peer
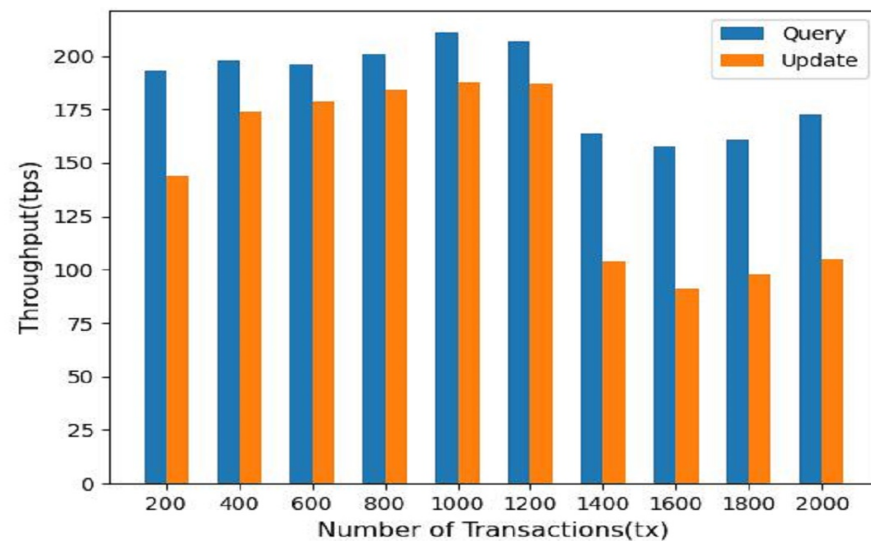
Figure 10. Throughput for different number of transactions

communication, and randomly selects a node in the anonymity region as the initiator of the query to send
the service request, which achieves the effect of hiding the actual query node with the fuzzy region. (3)
Blockchain-based trajectory privacy protection methods: (Jeong et al., 2020)combined blockchain and
cloud storage technology to achieve privacy protection of Telematics dataincluding vehicle trajectory data
and data sharing between different vehicles, and (Zhu et al., 2021) proposed a scheme called B-PPLS
based on Bitcoin system combined with OBE encryption algorithm to protect the privacy of user trajectory
data and achieve data sharing, which guarantees the traceability of location data by blockchain the location
data of different record owners in chronological order, and shares the location data among fully trusted
requestors by trust degree division. A distributed framework combining IPFS and federated blockchain
is proposed in (Kumar and Tripathi, 2020) for secure storage and sharing of reports including patient
trajectory data.

	Privacy protection	Storage mode	Expandability	Sharing
chen and chow	Yes	No discussion	No discussion	No discussion
jeong	No	Cloud Storage+Blockchain	Yes	Yes
zhu	Yes	Blockchain	No	Yes
kumar	No	IPFS+Blockchain	Yes	Yes
Our method	Yes	IPFS+Blockchain	Yes	Yes

Table 2. Comparison with other schemes

As shown in Table 2, this paper compares this paper's scheme with other existing trajectory privacy
protection schemes based on four aspects: privacy protection, storage method, scalability, and shareability.
These two types of trajectory data protection methods proposed in the (Chen et al., 2013) and (Chow
et al., 2011) only discussed the feasibility of the privacy protection aspect of trajectory data and did not
discuss the shareability of data, system scalability, or data storage methods.In the previous paper, we have
discussed the shareability, scalability and the storage method of data of our proposed methods in detail.
The scheme proposed in the (Jeong et al., 2020) uses third-party cloud storage under the chain when
the third party is threatened, the data still has the risk of leakage and cannot ensure the security of the
data under the chain. The approach proposed in the (Zhu et al., 2021) is deployed on the Bitcoin system,
which, as a representative of the public chain, sacrifices scalability to ensure complete decentralization.
The method proposed in the (Kumar and Tripathi, 2020) does not pre-encrypt the patient reports to be
uploaded to IPFS, which can lead to the risk of data leakage during the uploading process. Regarding
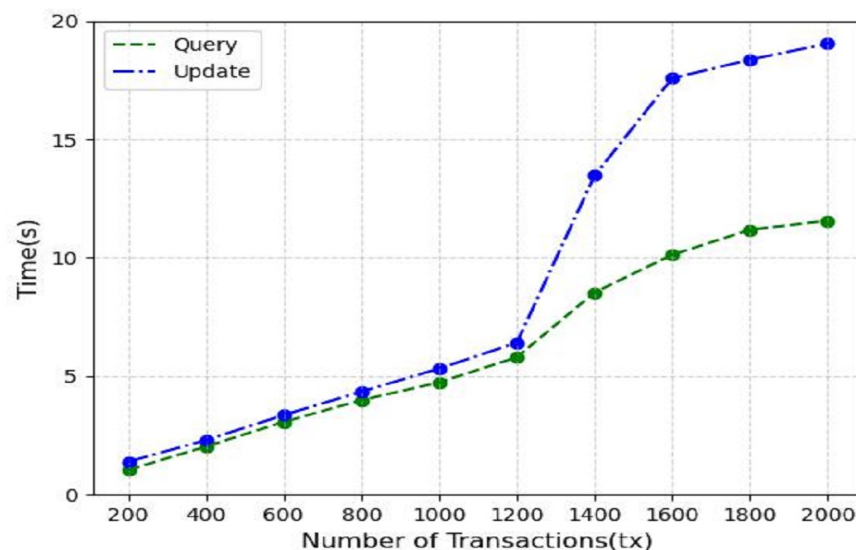
Figure 11. Response Time

the problems in (Jeong et al., 2020) and (Zhu et al., 2021), the approach proposed in this paper uses Hyperledger Fabric federated chain with good scalability and AES encryption technology to provide data privacy protection, and finally data storage through IPFS to relieve the storage pressure of the blockchain system and ensure the security of both on-chain and off-chain data.

In this paper, we have only proposed a rough solution to the problem, and many details are not yet perfect. For example, the contracts involved in this paper all use serial transfers for transactions, and only the serial contracts are stress-tested in the experiments, but in reality, when large-scale data sending and receiving is required, parallel transfers are needed to obtain better throughput rates. And currently this paper implements the simplest one-to-one data sending and receiving, which is not well supported for the more flexible one-to-many case. In the next section, we suggest some potential improvement directions for future research.

CONCLUSIONS

Resident trajectory data plays a huge role in epidemic prevention and control efforts, and its secure sharing will effectively inhibit the spread of COVID-19. In response to the current special demand for sharing resident travel trajectory data, this paper provides a new approach for privacy protection of resident travel trajectories based on federated blockchain and interstellar file system. The proposed method uses the Hyperledger Fabric federated chain framework combined with the storage model of IPFS, which is used to store the cipher text of the complete travel trajectory data, and the blockchain keeps an encrypted index consisting of the AES key used to encrypt the resident travel trajectory data, the returned IPFS storage address, the Location_ID and the creation time of the file. Users can query the corresponding index according to the location_ID and the creation time of the file, and get the travel track data through the index. It protects the privacy and security of residents' travel trajectory data while satisfying the demand of data sharing among various epidemic prevention departments.

In future research, we can improve the universality of the proposed method from the following perspectives. We can try to change the traditional blockchain transaction form into a parallel transaction form based on DAG to improve the throughput rate to adapt to large-scale data transmission, and DAG-based public chain projects such as IOTA and FTM have proved that they can bring more efficient and lighter transaction forms. Secondly, we can use some new blockchain consensus protocols, such as DPBFT protocol(Fu et al., 2021) and Dumbo protocol(Guo et al., 2020), to improve the fault tolerance rate of the system while ensuring the system performance.

REFERENCES

(2020). The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nature microbiology*, 5(4):536–544.

Bootsma, M. C. and Ferguson, N. M. (2007). The effect of public health measures on the 1918 influenza pandemic in us cities. *Proceedings of the National Academy of Sciences*, 104(18):7588–7593.

Chen, R., Fung, B. C., Mohammed, N., Desai, B. C., and Wang, K. (2013). Privacy-preserving trajectory data publishing by local suppression. *Information Sciences*, 231:83–97.

Chow, C.-Y., Mokbel, M. F., and Liu, X. (2011). Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *GeoInformatica*, 15(2):351–380.

De Montjoye, Y.-A., Radaelli, L., Singh, V. K., and Pentland, A. . (2015). Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science*, 347(6221):536–539.

Dhar, S., Khare, A., and Singh, R. (2022). Advanced security model for multimedia data sharing in internet of things. *Transactions on Emerging Telecommunications Technologies*, page e4621.

Dib, O., Brousmiche, K.-L., Durand, A., Thea, E., and Hamida, E. B. (2018). Consortium blockchains: Overview, applications and challenges. *International Journal On Advances in Telecommunications*, 11(1&2):51–64.

Dwivedi, S. K., Amin, R., and Vollala, S. (2021). Blockchain-based secured ipfs-enable event storage technique with authentication protocol in vanet. *IEEE/CAA Journal of Automatica Sinica*, 8(12):1913–1922.

El Ghazouani, M., El Kiram, M. A., and Er-Rajy, L. (2019). Blockchain & multi-agent system: A new promising approach for cloud data integrity auditing with deduplication. *International Journal of Communication Networks and Information Security*, 11(1):175–184.

Fu, J., Zhang, L., Wang, L., and Li, F. (2021). Bct: An efficient and fault tolerance blockchain consensus transform mechanism for iot. *IEEE Internet of Things Journal*, pages 1–1.

Guo, B., Lu, Z., Tang, Q., Xu, J., and Zhang, Z. (2020). Dumbo: Faster asynchronous bft protocols. In *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security*.

Gupta, R., Shukla, A., and Tanwar, S. (2020). Aayush: A smart contract-based telesurgery system for healthcare 4.0. In *2020 IEEE International conference on communications workshops (ICC Workshops)*, pages 1–6. IEEE.

Hasan, H. R., Salah, K., Yaqoob, I., Jayaraman, R., Pesic, S., and Omar, M. (2022). Trustworthy iot data streaming using blockchain and ipfs. *IEEE Access*, 10:17707–17721.

Jayabalan, J. and Jeyanthi, N. (2022). Scalable blockchain model using off-chain ipfs storage for healthcare data security and privacy. *Journal of Parallel and Distributed Computing*, 164:152–167.

Jeong, B.-G., Youn, T.-Y., Jho, N.-S., and Shin, S. U. (2020). Blockchain-based data sharing and trading model for the connected car. *Sensors*, 20(11):3141.

Kumar, P., Kumar, R., Srivastava, G., Gupta, G. P., Tripathi, R., Gadekallu, T. R., and Xiong, N. N. (2021). Ppsf: A privacy-preserving and secure framework using blockchain-based machine-learning for iot-driven smart cities. *IEEE Transactions on Network Science and Engineering*, (8-3).

Kumar, R. and Tripathi, R. (2020). A secure and distributed framework for sharing covid-19 patient reports using consortium blockchain and ipfs. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pages 231–236. IEEE.

Kumar, R. and Tripathi, R. (2021). Towards design and implementation of security and privacy framework for internet of medical things (iomt) by leveraging blockchain and ipfs technology. *The Journal of Supercomputing*, 77(8):7916–7955.

Li, Y., Yu, Y., and Wang, X. (2022). Three-tier storage framework based on tbchain and ipfs for protecting iot security and privacy. *ACM Transactions on Internet Technology (TOIT)*.

Liang, W., Yang, Y., Yang, C., Hu, Y., Xie, S., Li, K.-C., and Cao, J. (2022). Pdpchain: A consortium blockchain-based privacy protection scheme for personal data. *IEEE Transactions on Reliability*.

Lin, Y. and Zhang, C. (2021). A method for protecting private data in ipfs. In *2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 404–409. IEEE.

LU Haitao, Y. X. and ZHOU Qiang, D. J. a. L. R. (2022). Key intrinsic security technologies in 6g networks. *ZTE Communications*, 20(22-31).

Mohammadinejad, H. and Mohammadhoseini, F. (2020). Privacy protection in smart cities by a personal data management protocol in blockchain. *International Journal of Computer Network and Information Security*, (3).

She, W., Gu, Z.-H., Lyu, X.-K., Liu, Q., Tian, Z., and Liu, W. (2019). Homomorphic consortium blockchain for smart home system sensitive data privacy preserving. *IEEE Access*, 7:62058–62070.

Song, J., Li, Y., Lu, H., Wang, J., and Nie, P. (2022). An ipfs privacy storage sharing scheme based on sm4 algorithm. In *2022 2nd International Conference on Education, Information Management and Service Science (EIMSS 2022)*, pages 641–649. Atlantis Press.

Tripathi, G., Ahad, M. A., and Paiva, S. (2019). S2hs- a blockchain based approach for smart healthcare system. *Healthcare*, 8(1):100391.

Wu, S. and Du, J. (2019). Electronic medical record security sharing model based on blockchain. In *Proceedings of the 3rd International Conference on Cryptography, Security and Privacy*, pages 13–17.

Xie, J., Yu, F. R., Huang, T., Xie, R., Liu, J., and Liu, Y. (2019). A survey on the scalability of blockchain systems. *IEEE Network*, 33(5):166–173.

Zhang, A. and Lin, X. (2018). Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain. *Journal of medical systems*, 42(8):1–18.

ZHAO Tian, L. H., YANG Xin, W. H., and ZENG Ming, G. H. a. W. D. (2021). Differentially authorized deduplication system based on blockchain. *ZTE Communications*, 19(67-76).

Zheng, X., Mukkamala, R. R., Vatrapu, R., and Ordieres-Mere, J. (2018). Blockchain-based personal health data sharing system using cloud storage. In *2018 IEEE 20th international conference on e-health networking, applications and services (Healthcom)*, pages 1–6. IEEE.

Zhu, L., Liu, X., Yu, L., Cai, Z., and Zhang, J. (2021). Blockchain-enabled privacy-preserving location sharing scheme for lbsns. *Mobile Information Systems*, 2021.