Hindawi
Security and Communication Networks
Volume 2023, Article ID 4211453, 11 pages
https://doi.org/10.1155/2023/4211453

Research Article

Design and Optimization of Blockchain-Based Distributed Data-Sharing System for Urban Rail Transit

Mo Chen[1], **Hailin Jiang**[1], **Hongli Zhao**[1], **Huijun Zuo**[2] and **Qiang Zhang**[3]

[1]*State Key Laboratory of Traffic Control and Safety, Beijing Jiaotong University, Beijing, China*
[2]*Unit 96901 of PLA, Beijing 100094, China*
[3]*Unit 61741 of PLA, Beijing 100094, China*

Correspondence should be addressed to Hongli Zhao; hlzhao@bjtu.edu.cn

Received 28 July 2022; Revised 4 September 2022; Accepted 13 September 2022; Published 15 April 2023

Academic Editor: Jianbo Du

With the wide application of data-driven intelligence technology, the management efficiency and quality of urban rail transit improve considerably. However, due to data privacy and security protection, a large amount of data is still scattered among operators or rail transit departments in the form of data islands. The data barriers between urban rail departments severely hinder the intelligent development of urban rail transit systems. In this study, we design a distributed data-sharing system based on blockchain for urban rail transit system. Data production node in urban rail department, such as servers in data centers, wayside equipment, and onboard equipment, can share their data in a distributed way. Blockchain provides audit and check functions to guarantee data-sharing security. We explore the data-sharing incentive mechanism which has an impact on sharing willingness of data production nodes. To maximize the individual utility of nodes, we formulate an evolutionary game model for data nodes with bounded rationality to adapt their sharing strategies. The uniqueness and stability of the equilibrium of the game are also analyzed theoretically. Extensive experiments are conducted and illustrated that data production nodes can share their data efficiently and safely in our proposed data-sharing system. Our proposed evolutionary game model can determine the most effective data-sharing incentive mechanism.

1. Introduction

A large amount of log data is generated during the daily operation of railways. Rail transit will generate train travel log data, passenger flow log data, fault alarm log data, line data, and so on. With the rapid development of rail transportation, the effective storage and analysis of the large amount of log data become a very urgent need. The analysis and processing of these data are of great significance to improve rail transportation capacity, strengthen rail transportation safety management, analyze accident causes, and reduce rail transportation operation costs. However, the data of rail traffic logs are stored in each independent department, which has strong confidentiality. The use and supervision of log data are very strict. If you want to use log data for fault analysis, accident tracing, passenger flow prediction, image recognition, and other scenarios, you will face many data security problems. Therefore, it is very important to ensure the safe sharing of log data and the confidentiality and privacy of the data.

Data-sharing technology will effectively solve the problem that the development of big data cannot be further promoted in the urban rail transit system. The traditional centralized data sharing provides a way to break the data barriers in the urban rail transit system industry. However, for the traditional data sharing, the data information of different application systems is often managed centrally. This information storage and access structure will bring a series of information security issues, such as the vulnerability of data center servers, which will lead to the risk of data leakage, and there is no unified standard for data management and data authorization. The problems of information security and computational efficiency exposed by the

traditional centralized data sharing and processing model have paid extensive attention.

To solve the above problems, we consider applying blockchain-based distributed data sharing. Nowadays, blockchain has been applied to digital finance, Internet of Things, intelligent manufacturing, supply chain management, digital asset trading, etc. The distributed data sharing based on blockchain has also been applied to some fields such as medical care. In terms of urban rail transit data sharing, a lot of research has not been carried out at home and abroad. Studying the urban rail transit data-sharing scheme based on blockchain can effectively promote the realization of more data-driven urban rail transit intelligent application scenarios and improve the intelligence level of urban rail transit.

We first design a distributed urban rail transit data-sharing system based on blockchain; then, an incentive model for users to participate in data sharing is defined using evolutionary game theory (EGT) [1]. EGT has been used widely to many studies, such as using prisoner's dilemma to study blockchain mining in bitcoin system. However, little research has been carried out on modeling proportion of users participating in data sharing. In our proposed model, each user has the choice to share data or not, leading to users to be divided into two groups: users who participate in data sharing and users who do not participate in data sharing, and users will play games with each other. Therefore, the game has four different strategies: participating user A, nonparticipating user A, participating user B, nonparticipating user B, and A and B are two sides of the game. The main contributions of this study are as follows:

(1) We design an urban rail transit blockchain-based distributed urban rail transit data-sharing system for users to share data

(2) We design an incentive mechanism to study the proportion of users' data sharing

(3) From the perspective of EGT, taking a pair of users as an example, we propose a model to study the impact of different incentive parameters on the final trend of data-sharing proportion of both sides of the game in this system.

The rest of this study is organized as follows. In Section 2, we discuss the related work. The design and implementation of the system are discussed in Section 3. In Section 4, we use the proposed game model to analyze the evolution of the proportion of users sharing data theoretically and numerically [2]. Then, in Section 5, we summarize this study.

2. Related Work

With the advent of the era of big data, the application of data-sharing technology has been extended to many fields such as machine learning, the Internet of Things, and medical care [3–5]. However, in the area of rail transit data sharing, a lot of research has not been conducted at home and abroad, and it is necessary to improve and innovate it by referring to data-sharing technology in other fields and combining its characteristics and needs.

2.1. Data Sharing. The traditional data-sharing system adopts a centralized service model, and data are generally centrally managed by a central server, which is not conducive to efficient and open data sharing. There are high security risks in the process of data exchange, leading to great concerns in the process of data sharing among various departments. Scholars at home and abroad have carried out a lot of research on how to build an efficient, secure, and private data-sharing system. Azaria et al. [5] proposed the MedRec medical data-sharing system, which is the first prototype data-sharing system in the medical scene. The big data-sharing platform presented by Yan [6] is composed of three major segments: management and operation system, data resource platform, and data application platform, which provides ideas for building a big data-sharing platform in China. Meanwhile, Poline et al. [7] proposed a data-sharing method in neuroimaging, which makes it easy for researchers to share raw, processed, and derived neuroimaging data, as well as appropriate metadata and provenance records and improves the reproducibility of neuroimaging studies. Feng et al. [8] proposed the use of federated learning to protect user privacy during model training. And the data-sharing approach proposed by Wang et al. [9] provides a solution for data manipulation based on structured data descriptions rather than raw data files. However, all the above data-sharing systems are designed based on distributed storage, and the rise of blockchain technology provides a secure privacy option for distributed data storage and sharing.

2.2. Data Sharing and Management Based on Blockchain. Zyskind et al. [10] proposed a blockchain-based personal data management system to prevent user privacy leakage due to personal data collection by third parties. A Hawk protocol based on blockchain smart contracts was designed by Kosba et al. [11] to guarantee the confidentiality of blockchain transactions and solve the problem of transaction privacy leakage. A blockchain application that can be used for IoT devices was presented by Dorri et al. [12] to address the data security and privacy issues in IoT scenarios. Linder [13] proposed to protect private files through public-private key encryption and use smart contracts to provide audit trails. Kostić and Tang [14] applied blockchain technology and big data analytics to the audit procedure. Ali et al. [15] offered a secure blockchain-based global named storage system. In terms of data management, Huang et al. [16] offered a blockchain-based IoT data sharing solution to solve three types of trustworthy problems in the process of IoT data sharing and management. Pinno et al. [17] designed an access control architecture for IoT and suggested a more resilient management approach. Di Francesco Maesa et al. [18] designed a blockchain technology-based access control approach for resources. Xu et al. [19] applied blockchain to a range of data management projects such as data valorization and sensitive information sharing. Krawiec et al. [20] presented the use of blockchain for healthcare information interaction. Xiong et al. [21] surveyed the latest schemes on secure and privacy-preserving medical data sharing in the

last decade and classified them into unlicensed blockchain-based and licensed blockchain-based approaches. Dubovitskaya et al. [22] later offered a permissioned blockchain-based sharing system for electronic medical data management and further improved the security and privacy protection mechanisms in the data-sharing system. Kuo et al. [23] proposed ModelChain, which combines privacy-preserving online machine learning with a private blockchain, for cross-institutional healthcare, addressing the security and robustness vulnerabilities of centralized architectures. Zhang et al. [24] investigated a blockchain-based data sharing framework for AI-driven network operations, using blockchain technology to establish a mutually trusted data-sharing framework to bridge the data barriers between different operators and implemented a prototype system based on a hyperledger structure. A secure data-sharing environment is created by combining access control and monitoring through data chains and behavior chains. Liu et al. [25] suggested a secure data-sharing scheme in the blockchain-enabled MEC system using an asynchronous learning approach.

Zhu et al. [26] proposed a blockchain-enabled distributed security scheme on communication-based train control system. On March 18, 2019, the first blockchain-based electronic invoice for rail transportation in China was issued at Futian Station of Shenzhen Metro, China, further securing the security and validity of metro invoice issuance by using the open and tamper-evident function of blockchain in data sharing and opening the precedent of blockchain technology application in the metro field.

3. Design of Blockchain-Based Data-Sharing Scheme for Rail Transit

To solve the above problems, we propose a blockchain-based data-sharing platform for rail transit, which can protect the security, privacy, and confidentiality of data sharing, and provide data sharing services to designated data requesters, while other participants cannot access unauthorized data content, this platform can promote the sharing of rail transit data, enhance data processing capability, and analyze data value, which is of great practical significance to the safe operation of rail transit and increase speed and efficiency.

3.1. Key Technologies. This platform is designed mainly using blockchain technology, which is used to record information and control access, and distributed storage technology, which is used to store files The platform is designed mainly using blockchain technology, which is used to record information and control access, and the Inter Planetary File System, which is used to store files.

(i) Blockchain: in 2008, Satoshi Nakamoto suggested the concept of "blockchain" in the "Bitcoin White Paper," and in 2009, he founded the Bitcoin social network and developed the first block, which is called the "genesis block" [27]. Blockchain system is a decentralized system. Each blockchain node sends messages to all other nodes, and each node decides the final policy based on all the messages it receives. Due to the high network latency in peer-to-peer networks, the order of transactions observed by each node cannot be exactly the same, so the blockchain system needs to set up a mechanism to agree on the order of transactions that occur at about the same time. This algorithm to reach consensus on the order of transactions within a time window is the consensus mechanism. And each block consists of a block header and a block body. The block header contains information such as the current block hash, the hash of the previous block, the version number of the block chain, the verification code of the block chain, and the timestamp of the block. The most critical information in the blockchain is called transaction, which is recorded by the data of the block. If node consensus fails, the blockchain rejects the transaction. Each device of rail transportation can store the log information summary in the blockchain, and complete the secure and trustworthy sharing of data through smart contracts.

(ii) The InterPlanetary File System: IPFS (InterPlanetary File System) is a decentralized file sharing platform that supports identifying files by their content. IPFS uses Distributed Hash Tables (DHT) to retrieve file locations and use them to communicate with nodes for connectivity. When a file is uploaded to IPFS, it is split into blocks, each containing up to 256 KB of data or links to other blocks, and each block is identified by a cryptographic hash, also known as a content identifier, which is calculated based on its content. Because IPFS uses content identifiers to identify, authenticate, and transfer blocks and files, it is particularly well suited for use with blockchains. In addition, a second, different files with the same hash cannot be easily created, so it is not possible to flood IPFS with files with a given target file identifier. In short, files provided via IPFS are easily verified, and it is difficult to block a compute node by providing a different file with the same name or identifier.

3.2. Blockchain-Based Data-Sharing Model for Rail Transit. The blockchain-based rail transit data-sharing platform includes four parts: rail transit data requester, rail transit data owner, rail transit decentralized data-sharing platform, and rail transit data source. As shown in Figure 1, the data requestors refer to relevant rail transportation departments, research institutes, upstream and downstream companies supplying rail transportation equipment, universities, etc., which have demand for data. They can be divided into several categories according to the application areas, such as train operation control, rail transportation equipment maintenance, and rail transportation passenger flow prediction. The data owners refer to the various rail transit departments that own rail transit operation data. They are responsible for the daily operation of rail transit, and the rail transit equipment of each department will generate a large amount of data. For example, onboard equipment generates

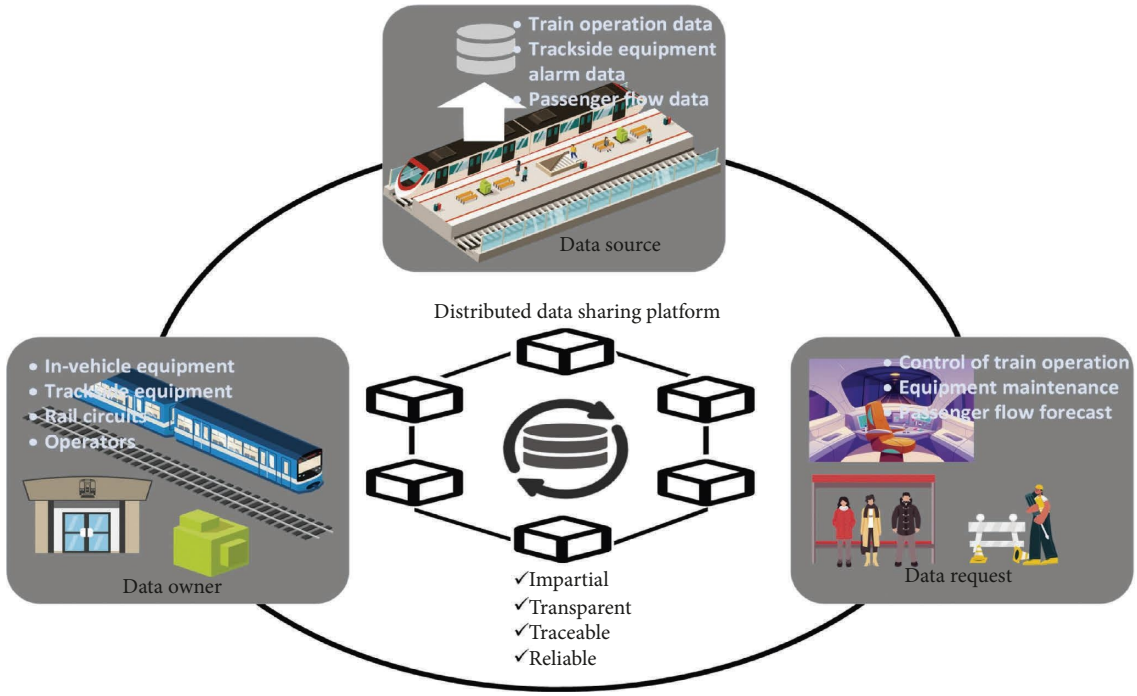

Figure 1: Blockchain-based rail transit data-sharing model.

train operation data, trackside equipment generates trackside equipment alarm data, and operators have daily new rail traffic flow data. Data source indicates the source of data, and the computer server provides data storage. It is held by the data owner.

In this model, the rail transit data owner releases data-sharing information through the blockchain data-sharing platform, and the rail transit data requestor initiates a data access request to the corresponding rail transit data owner according to its needs. After the rail transit data owner approves the access request from the data requester, the data requester can access the specified rail transit data content through the decentralized blockchain data-sharing platform. In addition, both parties can also complete data retrieval and data quality evaluation through the blockchain-based data-sharing platform. More importantly, the decentralized data-sharing platform can ensure the trustworthiness, transparency, and equality of the interaction between the two sides of rail transit data sharing. It creates a good data-sharing environment and protects the rights and interests of both parties sharing rail transit data.

Relying on blockchain technology, the decentralized data-sharing platform can meet the needs of rail transit data sharing for data trust, system security, and trustworthy expansion and has advantages that traditional data-sharing platforms cannot have.

3.3. Blockchain-Based Data-Sharing Access Control Method for Rail Transit. The meaning of access control is that when a requestor of rail traffic data requests access to the rail traffic log data, the data owner can give the specified data requestor access to his shared data by means of attribute encryption. Unauthorized users will not be able to access the data content. The data owner can protect the security and confidentiality of data sharing by keeping the data through the rail traffic data sharing platform.

The traditional public key encryption method is a coarse-grained access control with underground efficiency, which is difficult to adapt to the demand of selective sharing of rail transit data by data owners in rail transit data sharing. Therefore, this study proposes a data-sharing access control scheme based on blockchain and CP-ABE design. This scheme uploads the access policy with temporal attributes to the blockchain, and only data requesters whose node attributes satisfy the access policy within a specific time can obtain the decryption key. This scheme is suitable for solving the fine-grained control problem that exists in the rail transit data sharing environment. It can track the action records of rail transit data requesting nodes to publish request information and obtain access rights to rail transit data.

This access control method consists of three main participants: a track data owner, a track data requester, and a trusted key management center.

(1) DO: the track data owner, which is the unit or department that owns the track log data, is mainly responsible for publishing the metadata of the track dataset, setting the data-sharing access policy, and publishing the key and the cipher text based on the attribute encryption.

(2) DR: data requestor of rail traffic data is only the data requestor whose attribute set matches the data-sharing access policy and can access the specified dataset.

(3) AC: trusted key management center is responsible for generating public parameters and generating and

distributing keys for data owners and data requesters.

The CP-ABE algorithm in the blockchain-based data shared access policy for rail transit is divided into four parts.

(1) System initialization: the initialization is executed on the Trusted Key Management Center. The algorithm inputs the security factor λ and the attribute space U to generate the system public key PSK and the system master key MSK:

$$\text{Setup}(\lambda, U) \longrightarrow (\text{PSK}, \text{MSK}). \quad (1)$$

(2) Key generation: key generation is executed on the Trusted Key Management Center and provides the attribute association key USK for the data requestor based on the system public key PSK, the system master key MSK, and the attribute set A submitted by the data requestor:

$$\text{KeyGen}(\text{PSK}, \text{MSK}, A) \longrightarrow \text{USK}. \quad (2)$$

(3) Plaintext encryption: plaintext encryption is executed by the rail transit data owner to generate the ciphertext CT for attribute encryption by using the system public key PSK, the information to be encrypted T, and the access control structure A_{cp} associated with the access policy:

$$\text{Encrypt}\left(\text{PSK}, T, A_{\text{cp}}\right) \longrightarrow \text{CT}. \quad (3)$$

(4) Ciphertext decryption: decryption is performed by the data requester to obtain the plaintext T based on the system public key PSK, the attribute association key USK, and the attribute-encrypted ciphertext CT:

$$\text{Decrypt}(\text{PSK}, \text{USK}, \text{CT}) \longrightarrow T. \quad (4)$$

3.4. Blockchain-Based Distributed Data-Sharing Solution for Rail Transit. We combine blockchain and IPFS to design a distributed data-sharing scheme. The data are stored in IPFS in a distributed manner, and the data access control function is realized by deploying special functional blockchain smart contracts, which in turn guarantees the absolute control of data holders over the data.

As shown in Figure 2, the rail transit data-sharing platform is composed of two parts: on-chain and off-chain. The on-chain part is mainly a blockchain platform, which is responsible for data-sharing information on-chain storage, data-sharing access control, and on-chain identity information registration update. The off-chain part includes IPFS and trusted third-party key distribution center, which are mainly responsible for rail transit data storage, key distribution, identity registration, and e-verification.

The data-sharing information stored in the upper part of the chain includes information of data owners, information of data visitors, and some operation information of visitors accessing data, which includes reading and downloading of rail transit data. Based on these information, the rail transit data-sharing platform can strictly supervise the data access behavior of data requesters, and the rail transit data will be stored in the IPFS.

The blockchain-based rail transit data-sharing process is divided into three parts as follows:

(1) Rail transit data release: Figure 3 shows the process of publishing rail transit data. The owner of rail transit data needs to join the blockchain and IPFS first. After passing the identity authentication, the data source can be built or hosted. Then, it needs to write the detailed description of the rail transit data and the sharing related protocol to IPFS and publish the file. Then, the name and hash value of rail transit data are written to the blockchain, and after broadcast by the blockchain, the blockchain consensus is completed and synchronized across the network. This is the end of the rail transit data publishing process.

(2) Rail transit data request: Figure 4 shows the request process of rail transit data sharing. The node requesting data first needs to join the blockchain after authentication and then retrieve the required dataset according to the data name and hash value. It also obtains files such as data information and sharing protocols from IPFS. Then, select the eligible files to initiate rail transit data sharing, followed by downloading the relevant files to the local area, and finally, the blockchain performs consensus and stores them in the whole network. All operations during its data-sharing process are recorded.

(3) Rail transit data-sharing authority interaction: Figure 5 shows the interaction process of rail transit data-sharing authority. The owner of the rail transit data needs to review the identity information and request content of the data requestor. If the request does not meet the requirements, it is rejected directly. If it meets, the data access method is encrypted with the public key of the data requestor and written to IPFS according to the access control policy. Then, the data requestor will decrypt the downloaded file with the private key and obtain the access method. After accessing the dataset, the data-sharing interaction is completed and the permission interaction process is finished.

4. Evolutionary-Based Urban Transit Data-Sharing Strategies Design and Optimization

Take rail transit passenger flow data prediction as an example; the realization of passenger flow prediction needs the support of data sharing. By analyzing passenger boarding and alighting information and passenger flow etc., with machine learning methods, the prediction of passenger flow at the next time can be realized, thus helping operators to make reasonable operation plans, effectively alleviating traffic congestion, and improving the operation efficiency of urban rail transit systems. In the process of forecasting, the size of passenger flow data determines the effectiveness of passenger flow forecasting. However, each operator can only

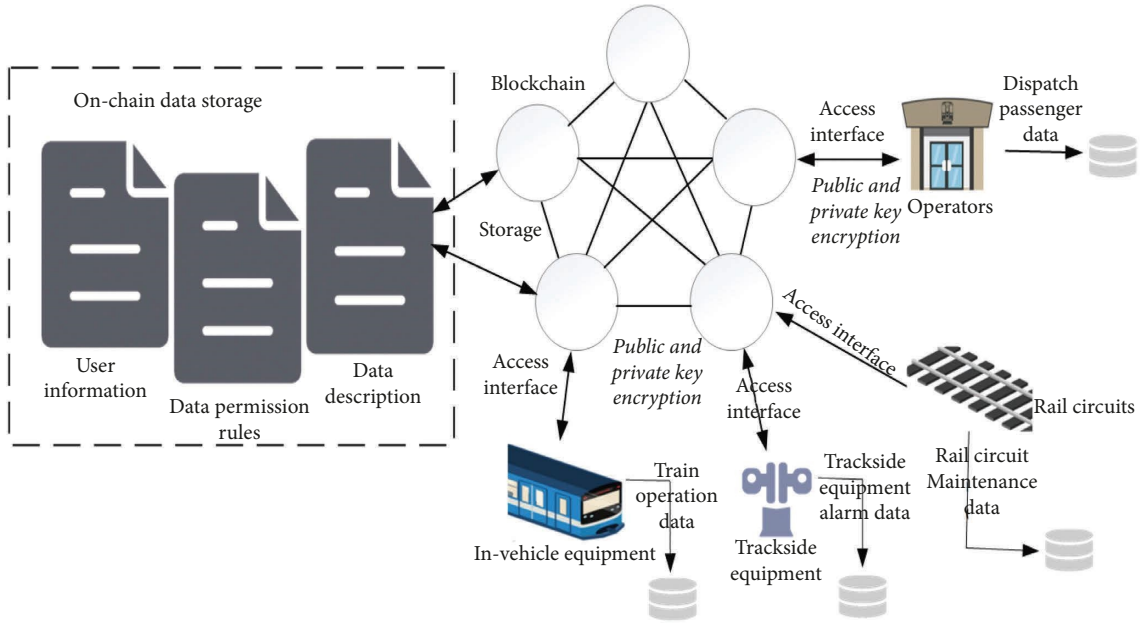

Figure 2: Blockchain-based distributed rail transit data-sharing platform.

get the passenger data of its own operation section, which is not shared among the operators, and the lack of data affects the prediction accuracy of passenger flow. Therefore, we need to design an effective data-sharing incentive mechanism to encourage operators to share data to achieve accurate passenger flow prediction.

Therefore, we create an evolutionary game payoff matrix and use EGT to analyze the evolutionary game process of the rail transit data generation node A and B (i.e., system user A and user B), analyze the final evolutionary direction of the game, and design an effective incentive mechanism to break the data barriers among rail transit operators based on the evolutionary results. Basically, the game is a two-player game where user A and user B generate their own data and share it with each other. After that, the data receiver has to analyze the data and reward the data sharer based on his progress.

4.1. Problem Formulation and Assumptions. In our proposed data-sharing framework based on blockchain, users are playing games with each other most of the time. Accurately, users share their data to get reward from other users of the trading program (aka data requesters), and each user can decide whether to engage in the data sharing or not [28]. Obviously, users' proportion to share data is an ongoing process affected by environment, incentive mechanism, and interactions among users, so users' proportion to share data is an evolutionary process [29]. Therefore, after constructing the EGT model, the influencing factors of the spread of sharing intention among users can be analyzed. The EGI model is based on the following concepts:

(1) Players: we consider the situation where two users play against each other in our proposed model, i.e., user A and user B. The participants are all self-interested, the information of everyone's strategic choices is symmetrical, and decisions are made in order [30].

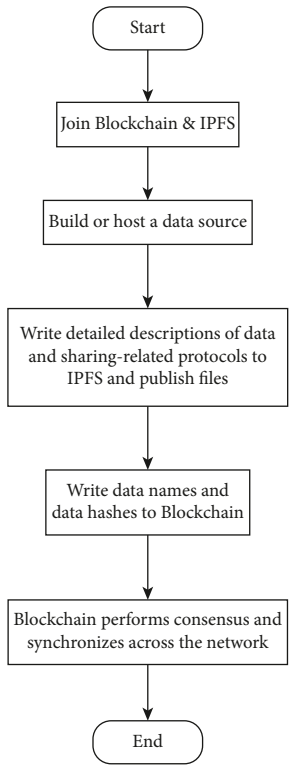

Figure 3: Rail transit data release process.

(2) Strategy space: users can choose from a strategy space of [P, N], where P refers to the participation strategy and N is the opposite [31].

(3) Incentive reward: to increase the proportion of users' sharing data, we assume that there is a mechanism

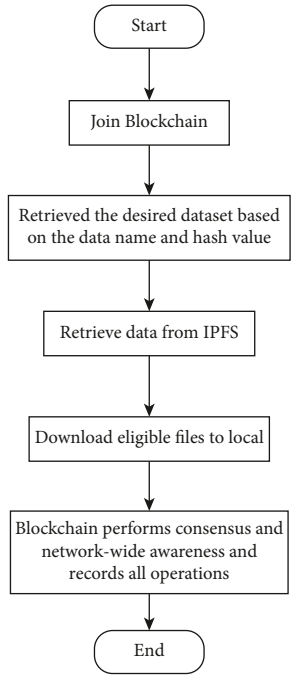

Figure 4: Rail transit data-sharing request process.

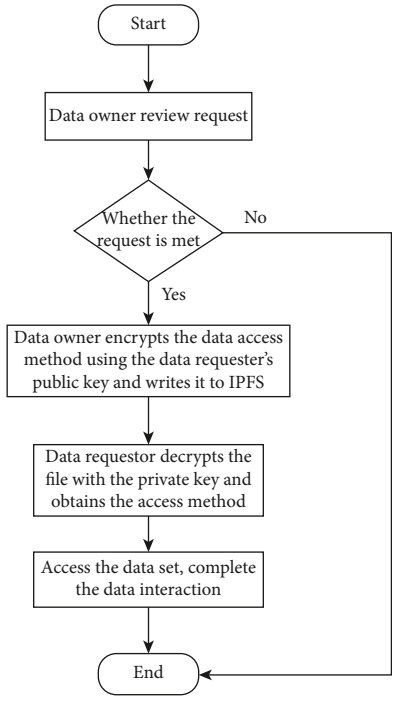

Figure 5: Rail transit data-sharing permission interaction process.

that allows users to obtain a reward R_i for one data sharing.

(4) Shared income magnification: if both parties participate in data sharing, both parties can obtain the data shared by other users and help the other party obtain more benefits. Therefore, we assume that the user's income is multiplied by $\gamma(\gamma > 1)$ times at this time.

(5) Parameters: other parameters and concepts used in the model are listed in Table 1

Based on the above assumptions and defined parameters, we can deduce that there are four combinations of strategies that both users may choose in this model, namely, [P, P], [P, N], [N, N], and [N, P], and the utility matrix of the game is listed in Table 2.

Setting the proportion of data shared by user groups $X(t) = x$, with $x \in [0, 1]$ based on Table 2, we infer the expected income of users who make strategy (P) as

$$I_U(P) = x[\gamma R_s - C_s + R_i] + (1 - x)[R_s - C_s + R_i]. \quad (5)$$

The expected income of users with strategy (U) is

$$I_U(N) = x[R_n - C_n] + (1 - x)[R_n - C_n], \quad (6)$$

and the expected income of total users is

$$I_U = xI_U(P) + (1 - x)I_U(N). \quad (7)$$

4.2. Evolutionary Game-Based Solution Method and Steady State Analysis. According to the Malthusian growth model, the growth rate of the proportion of users' sharing data is positively related to the utility and expected return of the incentive strategy. Hence, according to (1) or (2), the RDE of the proportion of users sharing data (noted later as RDE_x) is expressed as

$$F(x) = \frac{dx}{dt} = x(I_U(P) - I_U). \quad (8)$$

Simplify (4) to obtain

$$F(x) = \frac{dx}{dt} = x(1 - x)[(x\gamma + 1 - x)R_s + R_i - C_s - R_n + C_n]. \quad (9)$$

Thus, according to (5), we can get three equilibrium points:

$$\begin{aligned} x_1^* &= 0, \\ x_2^* &= 1, \\ x_3^* &= \frac{C_s + R_n - C_n - R_i - R_s}{R_s(\gamma - 1)}. \end{aligned} \quad (10)$$

According to differential equation stability theory, if x^* is a steady state, it must satisfy $F\prime(x^*) < 0$.

Case I: if $R_s + R_i - C_s < R_n - C_n < \gamma R_s + R_i - C_s$ meaning that the income from choosing strategy (N) is higher than the income from choosing strategy (P) when the other party chooses strategy (N) and lower than the income from choosing strategy (P) when the other party also chooses strategy (P), the stability of these three points is analyzed in Table 3. The results show that x_1^* and x_2^* are both evolutionary stable strategies (ESS) [32], which means that achieving a specific ESS depends on the initial proportion of users participating in data sharing. When $0 < x < C_s + R_n - C_n - R_i - R_s/R_s(\gamma - 1)$, ESS tends to be strategy (U),

Table 1: Parameters.

Symbols	Definition
R_s	Rewards from participating in data sharing
C_s	Cost from participating in data sharing
R_n	Rewards from not participating in data sharing
C_n	Cost from not participating in data sharing
R_i	Incentive rewards from participating in data sharing
γ	Scaling parameters for users to gain from shared data

Table 2: Payoff matrix.

User A and User B	Participative	Nonparticipative
Participative	$\gamma R_s - C_s + R_i, \gamma$, $R_s - C_s + R_i$	$R_s - C_s + R_i$, $R_n - C_n$
Nonparticipative	$R_n - C_n$, $R_s - C_s + R_i$	$R_n - C_n$, $R_n - C_n$

and when $C_s + R_n - C_n - R_i - R_s/R_s(\gamma - 1) < x < 1$, ESS tends to be strategy (P) meaning that if the initial percentage of users involved in data sharing is higher than a certain threshold (i.e., $C_s + R_n - C_n - R_i - R_s/R_s(\gamma - 1) < x < 1$), the rest of the users will choose strategy (P) as time goes by. Similarity, if the initial percentage of users involved in data sharing is lower than the threshold, ESS tends to be strategy (N).

Case II: if $R_n - C_n > \gamma R_s + R_i - C_s$ meaning that the income from choosing strategy (N) is higher than the income from both parties choosing strategy (P). The stability of these three points is analyzed in Table 4. The results show that x_1^* is the only ESS, which means that whatever the initial percentage of users involved in data sharing and the users will choose strategy (N). Because, in this case, strategy (N) brings the most income, the initial group adopting strategy (P) can easily be invaded by a small group of groups adopting strategy (N).

Case III: if $R_n - C_n < R_s + R_i - C_s$ meaning that the income from choosing strategy (N) is lower than the income from choosing strategy (P) when the other party chooses strategy (N), the stability of these three points is analyzed in Table 5. The results show that x_2^* is the only ESS, which means that whatever may be the initial percentage of users involved in data sharing, the users will select strategy (P).

Above all, the evolutionary stable strategy trend of EGI incentive model is shown in Figure 6, and the dynamic phases of the evolution of the proportion of users participating in data sharing of all cases are discussed in Section 5.

5. Performance Analysis

5.1. Numerical Analysis. To verify the stable strategy of users participating in data-sharing evolutionary games in the above three cases and show how the strategy stability influenced by certain parameters in the game, we satisfy three different cases by setting different values of R_s, C_s, R_n, C_n, R_i, and γ as shown in Table 6 and use Matlab to simulate the evolution process; the simulation results are shown in Figure 7.

In Figure 7(a), we set two initial proportion of users participating in data sharing to satisfy the first case: one is $x = 0.1$, lower than the threshold x_3^*, and the other is $x = 0.4$, higher than the threshold x_3^*. Hence, when $x = 0.1$, the user group will eventually evolve into a group that chooses strategy (U) in the game, and when $x = 0.4$, it will evolve towards strategy (P) which verifies the theoretical results.

To satisfy the second case, we start with $x = 0.5$ in Figure 7(b). The result shows though most of the initially

Table 3: Stability of equilibrium points (case I).

Equilibrium point	$F'(x^*)$	Stability
x_1^*	−	Stable
x_2^*	−	Stable
x_3^*	+	Unstable

Table 4: Stability of equilibrium points (case II).

Equilibrium point	$F'(x^*)$	Stability
x_1^*	−	Stable
x_2^*	+	Unstable
x_3^*	—	—

Table 5: Stability of equilibrium points (case III).

Equilibrium point	$F'(x^*)$	Stability
x_1^*	+	Unstable
x_2^*	−	Stable
x_3^*	—	—

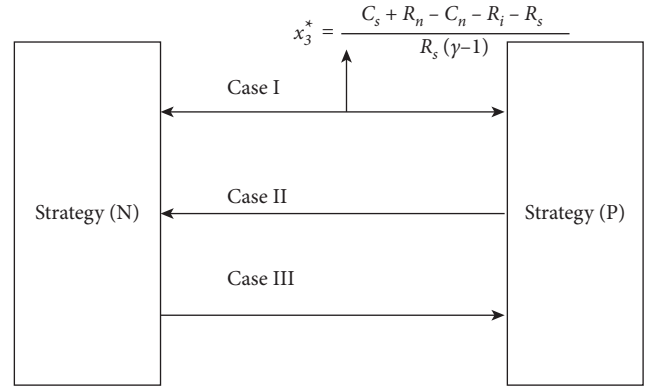

Figure 6: Evolutionary stable strategy trend of the EGI incentive model.

Table 6: Evolutionary game parameter settings.

	Parameter					
Case	R_n	C_n	R_s	C_s	R_i	γ
I	9	4	6	5	3	2
II	16	4	6	5	3	2
III	7	4	6	5	3	2

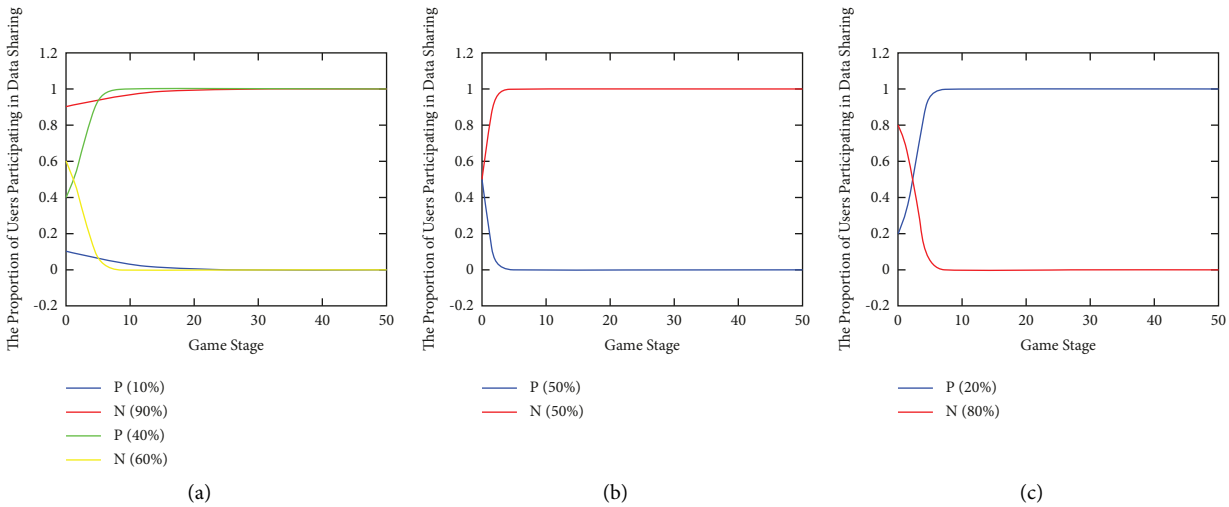

FIGURE 7: Dynamic stages in the evolution of the percentage of users involved in data sharing. (a) Case I, (b) Case II, and (c) Case III.

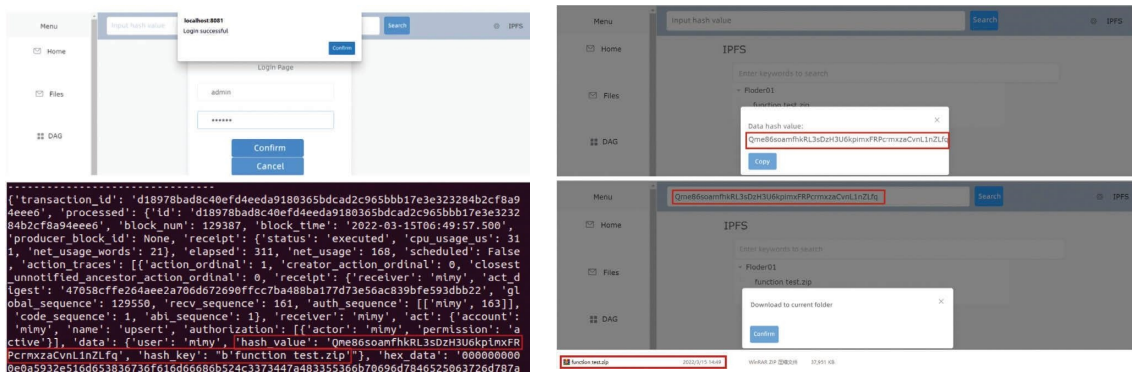

FIGURE 8: The system functions test. (a) Upload rail transit data. (b) Download rail transit data.

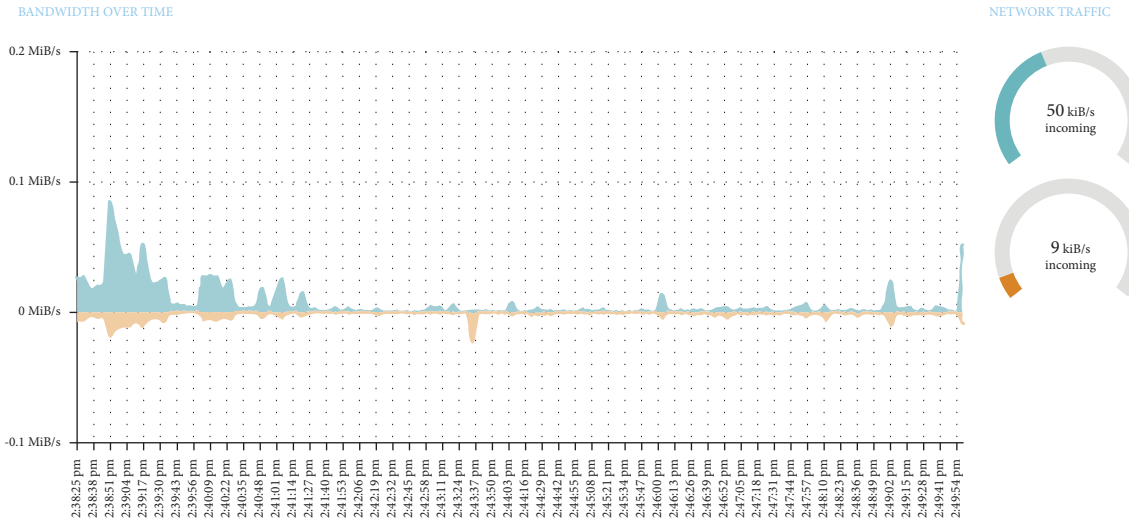

FIGURE 9: The system performance test.

users choose strategy (P) at first, the game eventually moves towards an nonparticipative group.

Finally, the evolution in the third case is presented in Figure 7(c), even if we only start the game with $x = 0.2$, the game will still evolve towards strategy (P), which matches the above theoretical analysis results [33].

Therefore, to control the final direction of evolution, we only need to set the parameter values to satisfy certain conditions based on the above parameters. That is, in order to encourage data sharing among operators, we can set the parameter values to satisfy Case I so that the final evolutionary trend will move in the direction of strategy (P) and the data barriers will be broken, regardless of the initial decision on the proportion of operators participating in data sharing.

5.2. System Functions and Performance Test. To ensure the system functional integrity and the efficiency of data transmission, we test the system's functions and simulate the data upload process of the data generation node and measure the data upload rate. The system function test results are shown in Figure 8. We implement blockchain-based data upload and download, which means the system has a complete prototype. We also test the real-time rate of rail traffic data upload. Since a large file (for example, a 1G file) will be broken when uploading to IPFS, forming multiple file fragments, each fragment is 256 KB, the uploading process of fragments is performed synchronously, and the storage node will store it according to its want-list table. In the other words, the data storage method of IPFS is that many storage nodes start to store fragments of large data files synchronously. When all fragmented files are stored, the storage work is completed. We measure the upload rate of a single fragment. As shown in Figure 9, the data upload rate reached 50KiB/s, which can meet the daily needs of urban rail transit data generation nodes.

6. Conclusions

In this study, we presented the design of a distributed data-sharing system based on blockchain to provide an efficient and secure data-sharing environment to promote more users to involve in data share. We also utilized EGT to analyze the evolution of user data-sharing behaviour in our designed system. The process of the relevant parameters affecting data-sharing behavior under three cases is studied in detail, and numerical simulation analyses are conducted, which are carried out by controlling the relevant parameters and the initial proportion of users involved in data sharing. We also verified the percentage of users involved in data sharing can be improved by adjusting the incentive parameters. Finally, we conducted functional and performance tests on the designed blockchain-based distributed data-sharing system, and the results showed that the upload rate of a single file fragment of size 256 KB can reach 50KiB/s in this system, and each file is broken into multiple file fragments and uploaded at the same time, and the performance of the system can meet the daily demand of rail transit data generation nodes. It can be seen that this distributed file transfer system provides a solution to enhance the data-sharing rate of rail transit.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This paper was supported by Beijing Natural Science Foundation (L201002), the Natural Science Foundation of China under Grants (61973026), Beijing Municipal Education Commission Funding (I20H100010, I19H100010), in part by the Beijing Jiaotong University Project under Grant (RCS2021ZZ005), and Fundamental Research Funds for the Central Universities (2021CZ107).

References

[1] R. Akkaoui, X. Hei, and W. Cheng, "An evolutionary game-theoretic trust study of a blockchain-based personal health data sharing framework," in *Proceedings of the 2020 Information Communication Technologies Conference (ICTC)*, pp. 277–281, Nanjing, China, May 2020.

[2] Y. H. Shan, L. I. Zhong-Fu, and S. O. Management, "Evolutionary game analysis of knowledge-sharing mechanism of construction supply chain towards construction industrialization," *Journal of Engineering Management*, vol. 29, 2015.

[3] X. Chen, J. Ji, C. Luo, W. Liao, and P. Li, "When machine learning meets blockchain: a decentralized, privacy-preserving and secure design," in *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, Seattle, WA, USA, December 2018.

[4] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2020.

[5] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: using blockchain for medical data access and permission management," in *Proceedings of the International Conference on Open Big Data*, Vienna, Austria, August 2016.

[6] L. Yan, "Study on the System Structure and Construction Thinking of Big Data Public Platform," *Library Theory and Practice*, 2017.

[7] J.-B. Poline, J. L. Breeze, S. Ghosh et al., "Data sharing in neuroimaging research," *Frontiers in Neuroinformatics*, vol. 6, p. 9, 2012.

[8] J. Feng, L. Liu, Q. Pei, and K. Li, "Min-max cost optimization for efficient hierarchical federated learning in wireless edge networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 11, pp. 1–2700, 2022.

[9] J. Wang, M. Chen, G. Lü et al., "A data sharing method in the open web environment: data sharing in hydrology," *Journal of Hydrology*, vol. 587, Article ID 124973, 2020.

[10] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing privacy: using blockchain to protect personal data," in *Proceedings of the 2015 IEEE Security and Privacy Workshops*, pp. 180–184, IEEE, San Jose, CA, USA, May 2015.

[11] A. Kosba, A. Miller, E. Shi, Z. Wen, and C. Papamanthou, "Hawk: the blockchain model of cryptography and privacy-preserving smart contracts," in *Proceedings of the 2016 IEEE Symposium on Security and Privacy (SP)*, pp. 839–858, IEEE, San Jose, CA, USA, May 2016.

[12] A. Dorri, S. S. Kanhere, and R. Jurdak, "Towards an optimized blockchain for iot," in *Proceedings of the 2017 IEEE/ACM Second International Conference on Internet-Of-Things Design and Implementation (IoTDI)*, pp. 173–178, IEEE, Pittsburgh, PA, USA, April 2017.

[13] P. Linder, "Decryption Contract Enforcement Tool (Decent): A Practical Alternative to Government Decryption Backdoors," *Cryptology ePrint Archive*, 2016.

[14] N. Kostić and X. Tang, "The Future of Audit: Examining the Opportunities and Challenges Stemming from the Use of Big Data Analytics and Blockchain Technology in Audit Practice," *Accounting and Finance*, 2017.

[15] M. Ali, J. Nelson, R. Shea, and M. J. Freedman, "Blockstack: a global naming and storage system secured by blockchains," in *Proceedings of the 2016 USENIX Annual Technical Conference (USENIX ATC 16)*, pp. 181–194, Denver, CO, Colorado, April 2016.

[16] Z. Huang, X. Su, Y. Zhang, C. Shi, H. Zhang, and L. Xie, "A decentralized solution for iot data trusted exchange based-on blockchain," in *Proceedings of the 2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 1180–1184, IEEE, Chengdu, China, December 2017.

[17] O. J. A. Pinno, A. R. A. Gregio, and L. C. De Bona, "Controlchain: blockchain as a central enabler for access control authorizations in the iot," in *Proceedings of the GLOBECOM 2017-2017 IEEE Global Communications Conference*, pp. 1–6, IEEE, Singapore, December 2017.

[18] D. Di Francesco Maesa, P. Mori, and L. Ricci, "Blockchain based access control," in *IFIP International Conference on Distributed Applications and Interoperable Systems*, pp. 206–220, Springer, Berlin, Germany, 2017.

[19] X. Xu, C. Pautasso, L. Zhu et al., "The blockchain as a software connector," in *Proceedings of the 2016 13th Working IEEE/IFIP Conference on Software Architecture (WICSA)*, pp. 182–191, IEEE, Venice, Italy, April 2016.

[20] R. Krawiec, D. Housman, M. White et al., "Blockchain: opportunities for health care," *Proceedings of the NIST Workshop Blockchain Healthcare*, pp. 1–16, 2016.

[21] H. Xiong, H. Zhang, and J. Sun, "Attribute-based privacy-preserving data sharing for dynamic groups in cloud computing," *IEEE Systems Journal*, vol. 13, no. 3, pp. 2739–2750, 2019.

[22] A. Dubovitskaya, Z. Xu, S. Ryu, M. Schumacher, and F. Wang, "Secure and trustable electronic medical records sharing using blockchain," in *AMIA annual symposium proceedings*, vol. 2017, p. 650, American Medical Informatics Association, 2017.

[23] T.-T. Kuo and L. Ohno-Machado, *Modelchain: Decentralized Privacy-Preserving Healthcare Predictive Modeling Framework on Private Blockchain Networks*, 2018, https://arxiv.org/abs/1802.01746.

[24] G. Zhang, T. Li, Y. Li, P. Hui, and D. Jin, "Blockchain-based data sharing system for ai-powered network operations," *Journal of Communications and Information Networks*, vol. 3, no. 3, pp. 1–8, 2018.

[25] L. Liu, J. Feng, Q. Pei et al., "Blockchain-enabled secure data sharing scheme in mobile-edge computing: an asynchronous advantage actor–critic learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2342–2353, 2021.

[26] L. Zhu, H. Liang, H. Wang, B. Ning, and T. Tang, "Joint security and train control design in blockchain-empowered cbtc system," *IEEE Internet of Things Journal*, vol. 9, no. 11, pp. 8119–8129, 2022.

[27] Bitcoin, *A Peer-To-Peer Electronic Cash System*, Decentralized Business Review, 2008.

[28] H. Abbass, G. Greenwood, and E. Petraki, "The n -Player Trust Game and its Replicator Dynamics," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 3, pp. 470–474, 2016.

[29] X. Liu, W. Wang, D. Niyato, N. Zhao, and P. Wang, "Evolutionary game for mining pool selection in blockchain networks," *IEEE Wireless Communication Letters*, vol. 99, 2017.

[30] C. Hao, Q. Du, Y. Huang, L. Shao, and Y. Yan, "Evolutionary game analysis on knowledge-sharing behavior in the construction supply chain," *Sustainability*, vol. 11, no. 19, p. 5319, 2019.

[31] C. Lei, D. H. Ma, and H. Q. Zhang, "Optimal strategy selection for moving target defense based on Markov game," *IEEE Access*, vol. 5, no. 99, pp. 156–169, 2017.

[32] J. Li and G. Kendall, "The effect of memory size on the evolutionary stability of strategies in iterated prisoner's dilemma," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 6, pp. 819–826, 2014.

[33] A. Ghoneim, G. W. Greenwood, and H. Abbass, "Distributing Cognitive Resources in One-Against-many Strategy Games," *in Evolutionary Computation*, 2016.