Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 5338876, 9 pages
https://doi.org/10.1155/2022/5338876

Research Article

Trusted Data Sharing Mechanism Based on Blockchain and Federated Learning in Space-Air-Ground Integrated Networks

Da Li,[1] Qinglei Guo,[1] Chao Yang,[2] and Han Yan[3]

[1]*State Grid Digital Technology Holding Co., Ltd., Beijing 100053, China*
[2]*State Grid Liaoning Electric Power Supply Co., Ltd., Shenyang 110004, China*
[3]*Beijing Jingan Yun Xin Technology Co., Ltd., Beijing 100193, China*

Correspondence should be addressed to Han Yan; yanhan@jingantech.com

Received 28 July 2022; Revised 31 August 2022; Accepted 15 September 2022; Published 7 October 2022

Academic Editor: Haitao Xu

Network data is distributed data on electricity, with the explosive growth of network data, and it has become an inevitable trend of network development to synergized and shared crossdomain scattered data and enhances the value transmission of network data. Federated learning, as a technology that combines data value delivery and data privacy security, is widely concerned in the process of data sharing. However, currently federated learning is used within a single business system. In the process of crossdomain data sharing, how to ensure the data trust, model trust, and result trust of federated learning is still an urgent problem to be solved. To this end, we designed to use blockchain structure to record each behavior of data sharing. Based on its tamper-proof and traceability, combined with cryptography technology, we constructed an endogenous trusted architecture for crossdomain data sharing. In addition, a reverse auction node incentive mechanism based on high credit preference is designed to solve the common problems in data sharing, such as low enthusiasm of users in sharing, unstable data quality of contributions, and unreasonable distribution of data sharing benefits. Through theoretical analysis and experimental verification, it can be seen that the incentive mechanism designed in this paper can meet the authenticity, user rationality, and budget feasibility. On this basis, it can motivate users to participate in data sharing, improve the average quality of data shared by users, and ensure security and trustworthiness and resist malicious attacks to a certain extent.

1. Introduction

International Data Corporation (IDC) issued a white paper entitled Data Age 2025 and predicted that the global data volume would grow dramatically in the future, from 45 ZB in 2019 to 175 ZB in 2025 [1]. The explosive growth of network data promotes the development and prosperity of data industry. Take electric power data as an example. In the process of production, transportation, distribution, and electricity consumption, power grid companies and consumers will generate power data of various types and huge volume with high authenticity and value. And the power data information will gradually develop with the development of power Internet of Things and the intelligence of power system. However, in the current development process, there are many isolated islands of data, resulting in limited scope of data use and difficulty in realizing repeated utilization, and data value is not fully utilized [2–4]. Although the data sharing mechanism with centralized structure is simple in design and easy to implement, it still has many problems, such as prone to single point of failure, difficulty in ensuring data privacy, inability to verify data authenticity, difficulty in equitable income distribution and ensuring data quality, and low enthusiasm of users for sharing [5–7].

In order to solve the above problems and explore a more secure and efficient data sharing mechanism, many relevant studies have been published recently. In [8], the author proposed a distributed machine learning technology—federated learning technology. By using federated learning technology, the global data model can be learned without uploading the original data, which not only protects the user's data privacy but also greatly reduces the communication cost. It has become an inevitable trend to solve the problem of data privacy protection in the process of data sharing through

federated learning technology [9–11]. But at present, federated learning is used in single business system, and there is no mechanism to guarantee data reliability, model reliability, and result reliability in the crossdomain untrusted data sharing environment. Blockchain has attracted wide attention due to its decentralized, traceable, and tamper-proof characteristics, and many scholars are working on using blockchain to solve the trusted problem of federated learning data sharing across domains [12–14].

In [15], the author proposed BlockFL, a federated learning structure combined with blockchain, which uses blockchain network to replace the central server to realize the exchange and update of the device's local model. In [16], the author proposed the use of federated learning technology combined with differential privacy to further strengthen the protection of data privacy and designed a consensus mechanism based on the quality of data model training, which combined the consensus mechanism with federated learning process to save computing resources and improve work efficiency. At the same time, the theory of zero trust has been gradually enriched and formed a standardized framework, and the concept of endogenous security has also been gradually developed. The endogenous security based on mimicry [17], endogenous security based on organic immunity [18], endogenous security architecture based on NEW IP, and other ideas and concepts have been formed [19]. These advanced security ideas and architectures are also affecting the development direction of mobile communication network security. Paper [20] analyzes the application of endogenous security theory in mobile communication network and gives the application method and practical scheme.

The above research has laid a good foundation for solving the privacy security problem in data sharing and designing a reliable data sharing environment. However, although the protection of privacy security has been further improved, the enthusiasm of data provider nodes to participate in data sharing and provide high-quality data is still not high. This is because they often consume certain resources in the process of participating in data sharing, which makes most nodes unwilling to provide free data sharing services [21, 22].

In the incentive mechanism of the study, the author of the paper [23] for networked data sharing scenario designed a data quality driven the auction model of incentive mechanism. Through the expectation maximization algorithm to evaluate data quality and actual task condition, guarantee the quality and credibility of data on the chain. However, the credibility of the algorithm depends on the data provider to truthfully upload and share the cost, and the system is vulnerable to attacks by malicious nodes. In [24], the author designed an incentive mechanism based on data quality for the mobile crowd perception scene. And it encourages the data provider to upload higher quality data through the additional reward mechanism based on credibility. However, the evaluation mechanism of node credit value is not fully considered. For example, it did not consider the changing range of credit value and the changing trend of the speed of credit value.

Based on the existing research, this paper designs a node incentive mechanism in an endogenous trusted sharing environment. With the decentralized, traceable, and tamper-proof characteristics of license chain, a decentralized and trusted trading environment is established. And the federal learning technology combined with differential privacy is used to strengthen the protection of user privacy security and realize data security sharing. In addition, this paper designs a reverse auction incentive mechanism based on high credit preference, which can not only improve the enthusiasm of users to participate in data sharing but also encourage users to submit higher quality data.

The arrangement of the remainder of this paper is as follows: chapter 2 introduces the building of the endogenous trusted environment, chapter 3 introduces evaluation mechanism of node credit value, chapter 4 introduces the reverse auction incentive mechanism with high credit preference, chapter 5 introduces the performance test results of the incentive mechanism designed in this paper, and chapter 6 summarizes and prospects for this paper.

2. System Model

This article is written for a common distributed data sharing scenario, the system involved in the data sharing nodes has certain types of data in the local store, and by participating in data sharing can help participants gain a more complete, but due to the problem of resource depletion and the risk of privacy [25], in the system node, it tends not to be active and free of charge for data sharing. Therefore, on the one hand, the system needs to strengthen security, and on the other hand, it needs to reward nodes for their contributions in data sharing [26]. This scenario is more common in practical applications. For example, in intelligent traffic, the actual traffic situation can be described more completely through information such as location and running speed shared by all vehicles. However, most vehicles are not willing to actively and freely share their own data due to concerns about privacy disclosure and resource consumption [23]. The goal of this paper is to design a reverse auction node incentive mechanism based on high credit preference in an endogenous trusted sharing environment. Virtual currency is used as a reward to motivate nodes in the system to actively participate in data sharing and provide high-quality data.

According to existing research work [27, 28], we design a two-tier endogenous trusted sharing architecture based on license chain. We build it as follows.

The upper level is the license chain layer, which is composed of the license chain nodes. They only store the license chain, but do not store original data, do not directly participate in data sharing, and do not conduct federated learning. The license chain stores two types of transactions, one that records the registration information of device nodes and the other that records the transaction information generated during the various steps of data sharing. The license chain node is selected during the system initialization and will maintain a relatively stable state in the follow-up work and will not exit at will; so, it is considered to be credible.

The lower layer is the device layer, which is composed of device nodes. Generally, device nodes store certain types of

data locally and do not store license chain nor can they carry out license chain retrieval. According to different demands in data sharing tasks, device nodes can be divided into data request nodes and data supply nodes. Device nodes must be registered and verified by all license chain nodes before joining the system. The registration information includes the IP address and data of the node, including locally stored data names and keywords of the node. This information is recorded on the license chain for retrieval. The structure diagram of the system is shown in Figure 1.

When a node generates a data sharing request, it will first send the request information to its superior license chain node to retrieve relevant nodes and data. The nodes that meet the requirements through screening become federated learning nodes of this data sharing task, and they together form a federated learning node set. The federated learning node will train the local data model based on the local relevant data according to the training requirements and send the local data model to other federated learning nodes after the training is completed. After receiving the local data model trained by other nodes, the federated learning node obtains the global data model through iterative training and sends it to the data requesting node. By avoiding direct transmission of raw data, the data provider firmly maintains ownership of the data.

Take the system with N data providing nodes as an example and assume that the number of data providing nodes in the system is sufficient. Each data provider node P_i has a local data set D_i, and all agree to participate in data sharing with compensation on the premise of not disclosing data privacy. The basic flow of a data sharing process is as follows.

(1) Data sharing request: when a node in the system generates a data sharing request, it will send the sharing request and related information to its superior license chain node, including the data type requirements and task indicators requirements of the data sharing task (such as data source diversity, working time, etc.).The license chain node receiving the data sharing request will first retrieve the request through the license chain to determine whether there is already a data sharing for the request. If the retrieval is successful, the data sharing request will be sent to the original federated learning node set, the global model will be returned to the data request node after the agreement is obtained, and the income will be distributed. If the retrieval fails or the federated learning node does not agree to share data, the multiparty data retrieval process will be entered

(2) Multiparty data retrieval process: after entering the multiparty data retrieval process, the license chain node receiving the data sharing request will conduct relevant node retrieval through the license chain to obtain the node set T_r related to the data sharing task. The system will select some nodes from set T_r to participate in this data sharing task according to the node selection mechanism designed in this paper, and the selected nodes form the federated learning node set T_f of this task

(3) Node selection based on node credit: according to the node selection mechanism with high credit preference designed in this paper, the system selects some nodes from the relevant node set T_r to participate in the data sharing task and form the federated learning node set T_f of this task

(4) Federated learning process: the federated learning process will unfold between federated learning nodes. Taking node P_i in set T_f as an example, local data model $\widehat{m}_i$ can be trained through the following steps: first, select relevant data sets for training. The federated learning node P_i will select the local related data subset d_i in the local data set D_i according to the requirements of the data sharing task. Second, data encryption is based on differential privacy. By adding Laplacian noise, we can obtain the encrypted local relevant data subset $\widehat{d}_i$ based on the local relevant data subset d_i. Third is local model training. The local data model $\widehat{m}_i$ can be obtained by using machine learning algorithms on the encrypted local data subset $\widehat{d}_i$. The above training process is performed locally by node P_i. After the training, node P_i will send the local data model $\widehat{m}_i$ to other federated learning nodes. At the same time, node P_i will also train other local data models it receives with $\widehat{m}_i$ until it receives local data models of all federated learning nodes and trains global data model M

(5) Incentive mechanism: after the federated learning process, the credit value of federated learning nodes can be updated and rewarded according to the work quality of the nodes. The revenue c_i obtained by the node consists of two parts, namely, basic revenue c_i^b and additional reward c_i^+. c_i^b is the quotation of nodes, which is available to all nodes involved in this work. c_i^+ is a reward based on the quality of work earned by nodes, and only nodes with high quality of work are awarded

(6) Consensus mechanism: after the consensus process begins, the node with the highest credit in the committee will be appointed as the leader of the committee through the voting mechanism based on credit. If there are multiple nodes with the same credit value in the committee at the same time, one of the nodes will be randomly selected as the leader of the committee. Leaders need to drive consensus mechanisms, collect all transaction records, and write them into a temporary block. The leader then broadcasts the provisional block to all members of the committee. In members of the commission after receiving a temporary block to block in the record of the transaction information verification, validation includes not only the content of the transactions

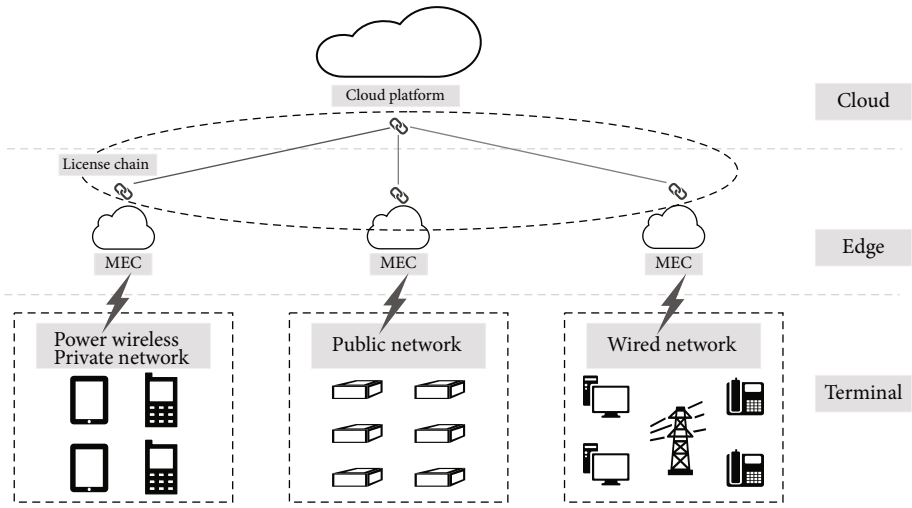

Figure 1: System structure diagram.

of the basic information (e.g., block head) but also the quality of the model (i.e., local test results and the records in the trade gap is within a certain range), if the verification through, trading argues that this is legal. If the provisional block is approved by more than two-thirds of the committee, the leader sends the block with its signature to all the nodes, and the block is then recorded on the blockchain

3. The Evaluation Mechanism of Node Credit Value

The credit value of nodes is a long-term observation index which can reflect the past behavior of nodes and predict the future performance of nodes to a certain extent. It is accumulated from the work quality of nodes in each data sharing task [16, 29].

In this paper, it is considered that in a data sharing task, and the variation of the credit value of node P_i depends on two factors [22]: the data model quality contributed by nodes in this data sharing task q_i and the initiative of nodes to participate in data sharing w_i. The evaluation results of the above two indicators are expressed as values within the range of [0, 1].

In this paper, the accuracy of model prediction results is used to measure the quality of the model. For example, absolute mean error MAE can be used to measure the training quality of the model for regression tasks.

In a specific data sharing process, a subset T_f of device nodes can be obtained through retrieval, which is called the federated learning node set of the data sharing. Each node in T_f will train the local data model and participate in the consensus process.

After the federated learning process, node P_i will take the subset of local related data set d_i as the test set, test the received local data model, and score the model quality according to the test results. Taking node P_i as an example, the local data model $\widehat{m}_j$ is tested, and the score $q_{i,j}$ can be obtained. Each federated learning node P_i will obtain a model quality evaluation matrix Q_i through testing:

$$Q_i = \left[q_{i,1}, q_{i,2}, \cdots, q_{i,n_f}\right]. \tag{1}$$

Since the node cannot test the local data model trained by itself, let $q_{i,i} = 0$. The average model quality of each node can be calculated based on the model quality q_i evaluation matrix of each node, which can be used as the model quality score of node P_i in this data sharing task. The calculation formula is as follows:

$$q_i = \frac{\sum_{m=1,2,\cdots n_f} Q_m[i]}{n_f - 1}. \tag{2}$$

In a data sharing task, the credit value variation of node P_i can be expressed as

$$r_i^f = f(q_i, \bar{q}, w_i, \bar{w}). \tag{3}$$

In order to facilitate analysis and calculation, the range of node credit value is set as [0, 1], where 0 represents absolute distrust, 1 represents absolute trust, and 0.5 is a neutral value. By default, the initial credit value for joining device nodes to the system is set to 0.5.

The credit updating algorithm of the node is as follows:

$$r_i^m = \max\left(\min\left(r_i^{m-1} + \beta \cdot r_i^f, 1\right), 0\right), \tag{4}$$

$$r_i^{tmp} = \begin{cases} r_i^f \wedge \left(1 - r_i^{m-1}\right), & r_i^f \geq 0, \\ -\left(-r_i^f\right) \wedge r_i^{m-1}, & r_i^f < 0, \end{cases} \tag{5}$$

$$r_i^f = \frac{4}{\pi} \times \arctan\left(\frac{1}{2} \times \left(\frac{q_i}{\bar{q}} + \frac{w_i}{\bar{w}}\right)\right) - 1. \tag{6}$$

The range of r_i^f is (-1, 1), and the range of r_i^f is [0, 1].

After the end of a data sharing task, the credit value r_i^m of node P_i will be affected by two factors, one is the credit value r_i^{m-1} of node P_i in the previous stage, and the other is the credit value variation r_i^f of node P_i in this data sharing task. As the data requester is generally more likely to forgive the errors of good nodes than malicious nodes, the larger the node credit value is, the greater the impact of the incremental in credit value of the node on the node credit value is, and the smaller the impact of the decrement in credit value of the node on the node credit value is. The smaller the node credit value is, the smaller the impact of the increment in credit value of the node on the node credit value is, and the greater the impact of the decrement in credit value of the node on the node credit value is.

4. The Reverse Auction Node Incentive Mechanism with High Credit Preference

The credit value of a node is constantly updated based on its performance in data sharing tasks [30]. The better the quality of the data model provided by the node, the higher the initiative of participating in the data sharing, and the higher the credit value of the node. On the contrary, the worse the quality of the data model provided by the node, the lower the initiative of participating in data sharing, and the lower the credit value of the node [31, 32]. The credit value of nodes is a value accumulated according to the long-term performance of nodes, which can reflect the past performance of nodes and predict the future performance of nodes to a certain extent. Although the work quality of high credit value nodes is not necessarily higher than that of low credit value nodes in a particular data sharing process, statistically speaking, high credit value nodes are more likely to provide high quality work. Based on this, a reverse auction node incentive mechanism with high credit preference is designed.

In a specific data sharing task, the data request node will send the data sharing request to its superior license chain node, which contains the requirement D for the diversity of data sources.

Through the multiparty data retrieval process, the related node set T_r can be obtained, and the nodes in T_r are divided into two groups according to the size of the credit value, namely, the high credit value node set T_r and the low credit value node set L.

$$P_i \in \begin{cases} H, & \text{if } r_i^{m-1} \geq r_{\text{med}}, \\ L, & \text{otherwise}, \end{cases} \tag{7}$$

$$r_{\text{med}} = 0.5, \tag{8}$$

$$F = \begin{cases} H, & \text{if } H >= D, \\ H + L', & \text{otherwise}, \end{cases} \tag{9}$$

where L' is the $D - H$ nodes randomly selected from the set of low credit value node set. Set F is a set of federated learning nodes, and only the nodes in F will receive the data sharing task sharing request.

The node receiving the data sharing request chooses whether to participate in the data sharing according to the specific requirements of the task and its own situation. If the node agrees to participate in the data sharing, it will make a quotation. According to the quoted price c_i^b of the node and the requirement of the data-requesting node for the diversity of data sources, the node set that will participate in federated learning can be screened out, and the cost C of this data sharing task can also be calculated. The cost of data sharing task consists of two parts: base cost C^b and reward cost C^+.

$$C = C^b + C^+, \tag{10}$$

$$C^b = \sum_{i=1,2,\cdots,n_h} c_i^b, \tag{11}$$

$$C^+ = \beta \cdot C^b. \tag{12}$$

β is the ratio of reward cost C^+ to base cost C^b, where $\beta = 1 - r_r^m$, which means that the ratio of reward cost to base cost in this data sharing task is negatively correlated with the credit value of data request node. The higher the credit value of the data request node is, the lower the reward cost of the data sharing task will be, otherwise, the opposite, but it will not exceed the base cost at most. Since nodes in the system generally need to share data, this mechanism to some extent encourages nodes to provide high-quality work in order to obtain a higher credit value.

After the data provider P_i participates in the data sharing task issued by the data request node P_r, it can obtain a reward c_i. c_i consists of two parts, namely, the basic reward c_i^b and the additional reward c_i^+.

$$c_i = c_i^b + c_i^+. \tag{13}$$

c_i^b is the quote of the node, which can be obtained by all nodes participate in the data sharing task. c_i^+ is the reward assigned to the node according to the work quality of the data provider, which only the well-behaved nodes can get. The reward c_i^+ obtained by a node in a data sharing task is related to two factors, one is the reward cost C^+, and the other is the work quality of P_i in this data sharing task. The service quality evaluation standard of the node in this data sharing task is determined by the specific requirements of the data request task. For example, the quality of the model trained by the node, the working time of the data model of the node, etc.

The higher the quality of the work of the federated learning node P_i in this data sharing task, the more rewards, and vice versa.

5. Experiment Analysis

5.1. Mechanism Performance Analysis. This paper mainly focuses on a data sharing scenario in which a data requesting

node and multiple data providing nodes jointly complete a data sharing task. There are three types of objects in the scenario, namely, the platform side, the data requester, and the data provider. Both the data requester and the data provider are considered untrustworthy.

(1) Authenticity: due to the decentralized structure based on the permissioned chain, each step of the mechanism can be automatically executed through smart contracts, and any member of the system can check and trace the transaction information on the chain, ensuring that the platform works honestly according to the request of the data requester and the data provider

(2) Individual rationality: according to the reverse auction incentive mechanism based on high credit preference designed above, the income that the data provider can obtain during a data sharing process is

$$c_i = c_i^b + c_i^+. \tag{14}$$

c_i^b is the quote of the node, which can be obtained by all nodes participate in the data sharing task. c_i^+ is the reward assigned to the node according to the work quality of the data provider. Since the node cannot make an offer that is lower than the cost, it is guaranteed that the node's benefit in a data sharing process cannot be lower than its cost.

(3) Budget feasibility: according to the reverse auction incentive mechanism based on high credit preference designed above, the cost of the data requester in a data sharing process is

$$C = C^b + C^+, \tag{15}$$

$$C^b = \sum_{i=1,2,\cdots,n_h} c_i^b, \tag{16}$$

$$C^+ = \beta \cdot C^b. \tag{17}$$

The cost of the data sharing task consists of two parts, namely, the basic cost C^b and the reward cost C^+. The basic cost is the quotation made by the data provider, which can be known before data sharing. The reward cost is a variable related to the credit value of the data requester. In this paper, let

$$\beta = 1 - r_r^m. \tag{18}$$

That is, the total cost of the data requester will not exceed twice the basic cost, and the data requester can start data sharing after agreeing to this budget.

(4) Security and reliability: since both the data requester and the data provider are untrustworthy, it is necessary to consider some attack behaviors that the node may take to maximize its own interests. This paper focuses on three types of attacks: first, the data requester may speculate on the private data of the data provider through the shared data, resulting in the leakage of the privacy of the data provider. By the differential privacy-based federated learning technology, the data model can replace the original data to participate in the sharing, which greatly strengthens the data privacy protection for the data provider. Second is data spoofing attacks, that is, deliberately providing some irrelevant data in the process of data sharing, disrupting the normal work of the system. By adding high credit preference to the incentive mechanism, nodes that constantly provide high-quality service can be selected, thereby avoiding the occurrence of data spoofing attacks. Third is slander attack, that is, malicious evaluation of other users, deliberately uploading unfair evaluation, etc. By listening to and adopting the evaluation of the majority of nodes, the impact of slander attacks can be reduced to a certain extent

5.2. Simulation Design. This paper mainly simulates data spoofing attacks and observes the impact of attack behavior on the normal operation of the mechanism. Since it is not necessary to observe the influence of node enthusiasm on node credit value, it is assumed that all nodes actively participate in the sharing task.

In order to facilitate the calculation, the model quality, working time, and quotation range of the node are set to [0, 1]. Honest nodes train local data models based on locally relevant data; so, the quality range of their models is set to [0.5, 1]. The malicious nodes intend to conduct data spoofing attacks; so, the real data model is not obtained through model training, which leads to a large change in its model quality; so, the range of its model quality is set to [0, 1]. Similarly, the working time of the honest nodes includes the communication time and the training time; so, the variation range of working time is set to [0.2, 1]. The working time of the malicious node only includes the communication time; so, the variation range of working time is set to [0.1, 0.5]. In the same way, the work cost of honest nodes includes communication cost and training cost; so, the variation range of price is set to [0.2, 1]; the work cost of malicious node is only related to communication cost; so, the variation range of the quotation is set to [0.1, 0.5].

In this simulation, the malicious node will not adjust the attack strategy intelligently and will definitely make the attack behavior.

5.3. Simulation Results and Analysis. First, the initial credit value of the node in the system is set to a neutral value 0.5, and the initial account balance of the node is set to 1. The number of data providing nodes is set to 1000, of which malicious nodes account for 10%, that is, 100. The variation range of the credit value of the data requester is set to [0.5, 1], the variation range of the expected data model quality is set to [0.6, 1], and the variation range of the diversity of data sources is set to [50, 100]. 500 data sharing tasks are

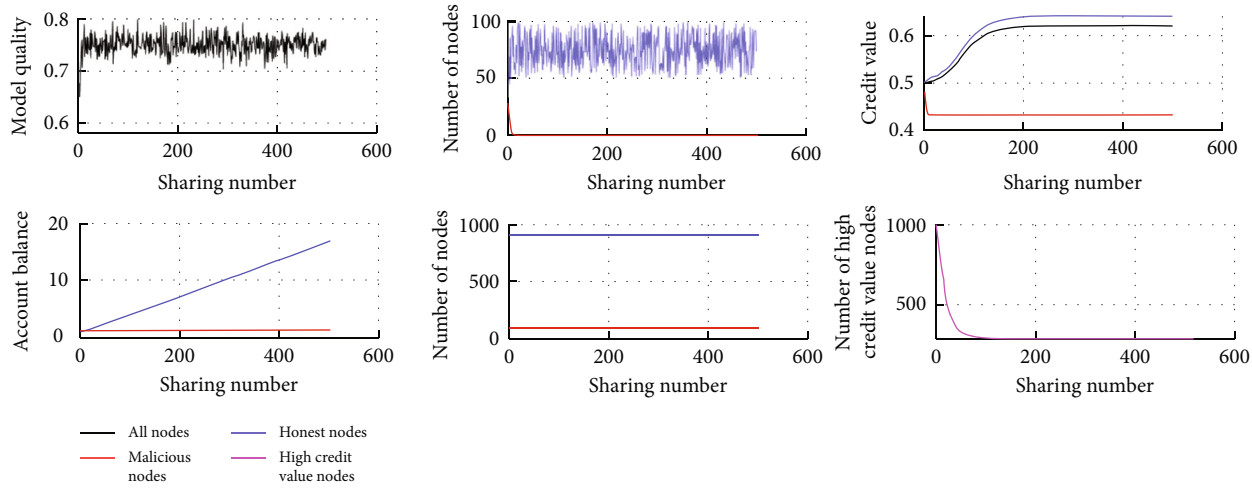

Figure 2: Simulation result 1.

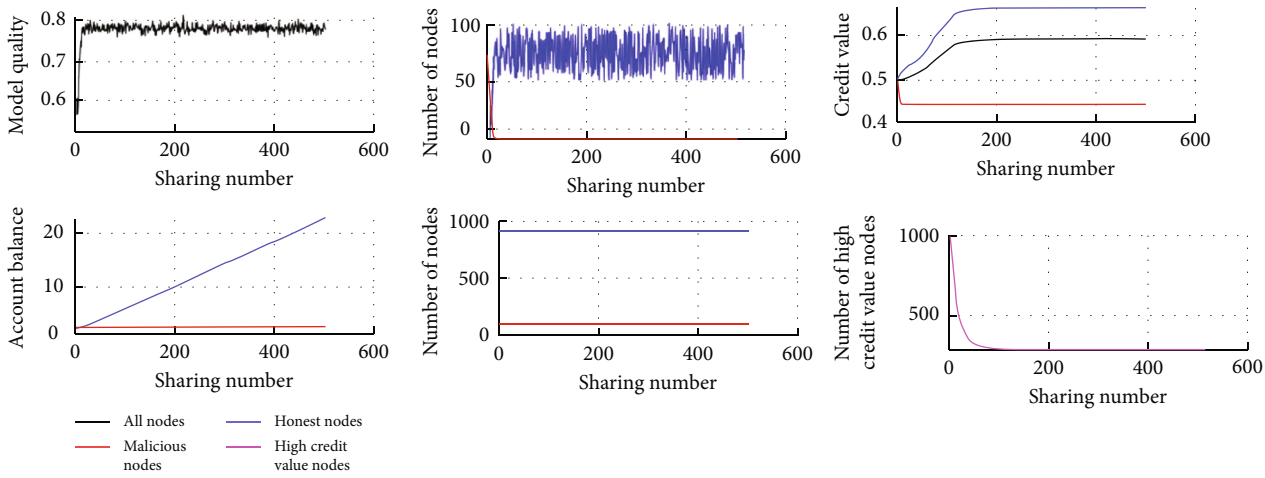

Figure 3: Simulation result 2.

simulated in MATLAB, and the simulation results are shown in Figure 2.

Compared with the traditional data trusted federated learning system model, the data trusted sharing mechanism proposed in this paper can reduce the number of malicious nodes and the credit value of malicious nodes with the increase of sharing times, improve the average quality of the model, and is relatively stable within a certain range.

Figure 2(a) shows the variation curve of the average value of node model quality, which reflects the change of the average value of local model quality trained by nodes in different data sharing tasks. Figure 2(a) demonstrates that the average value of node model quality is not high at the initial stage due to the participation of malicious nodes, but it increases rapidly and changes within a certain range.

Figure 2(b) shows the variation curve of the number of nodes in the data sharing task, reflecting the number of honest nodes and malicious nodes in federated learning nodes in each data sharing task. Because different data haring tasks have different requirements for the diversity of data sources, the sum of the number of honest nodes and malicious nodes is not a fixed value. However, it can be seen that malicious nodes only participate in the first few data sharing tasks. This is because the quality of the data model provided by malicious nodes is unstable; so, the credit value of malicious nodes will gradually decline, and the possibility of being selected to participate in data sharing tasks will become lower.

Figure 2(c) shows the variation curve of node credit value, which reflects the change of the average value of credit value of all nodes including honest nodes and malicious nodes. It demonstrates that the credit value of a malicious node will decline rapidly and eventually tend to a certain value. After the credit value of a malicious node decreases, it is less likely to be selected to participate in the data sharing task and not participate in the data sharing task; so, the credit value will not change.

Figure 2(d) shows the variation curve of node average account balance, reflecting the changes of account balance at honest nodes and malicious nodes. Malicious nodes only

participated in the previous data sharing tasks; so, their account balance only increased slightly at the beginning. The account balance of honest nodes will continue to grow with their continuous participation in data sharing tasks.

Figure 2(e) shows the variation curve of the number of effective nodes, reflecting the change of the number of effective nodes and the number of honest nodes and malicious nodes with the progress of data sharing task. It demonstrates that both honest nodes and malicious nodes do not produce invalid nodes because there are enough data providing nodes. When the credit value of the node drops to a certain value, it will not be selected to participate in the data sharing task. After that, it will maintain the fixed value; so, it will not become invalid nodes.

In order to further test the ability of the mechanism designed in this paper to resist data spoofing attack, keep other initialization values unchanged, increase the number of malicious nodes to 33%, that is 333, and simulate 500 data sharing tasks in MATLAB. The simulation results are shown in Figure 3. Compared with the simulation results in Figures 2 and 3, the change curve of the variation value of node model quality is more stable in simulation result 2, that is, when the number of malicious nodes reaches 33%. The variation curve of the number of nodes in the data sharing task in simulation result 1 and simulation result 2 is basically the same. It suggests that even the number of malicious nodes is increased, the mechanism designed by this paper can also be very good to exclude malicious nodes in the data sharing tasks, also can quickly reduce the chances of malicious nodes participate in data sharing tasks, and can deceive attack resistance data. The variation curve of the average value of node model quality of the two simulation results is basically the same, because simulation result 2 sets more initial malicious nodes; so, the average value of all nodes is lower than simulation result 1. The variation curve of node average account balance shows the same trend in Figures 2 and 3, indicating that the increase in the number of malicious nodes will not affect the continuous growth of the account balance of honest nodes as they keep participating in data sharing tasks.

At the same time, compared with the traditional incentive mechanism and the paper [33], which did not consider the malicious nodes and data storage security and privacy, the incentive mechanism designed by this paper on the basis of improving the nodes involved in data sharing is able to quickly reduce the chances of malicious nodes participate in data sharing tasks. And it can share data security and privacy through the endogenous and reliable security system.

6. Conclusion

Based on the previous research results on data sharing mechanism, this paper designs a node incentive mechanism in the endogenous trusted sharing environment. Through the credit value evaluation mechanism based on the work quality of nodes, the work performance of nodes is accumulated and quantified, which provides a reference index for predicting the future work performance of nodes. By adding high credit preference into the incentive mechanism, the work quality of nodes is guaranteed to a certain extent, and it is helpful to eliminate malicious nodes. Through the reverse auction mechanism, the data-providing node is prompted to make a reasonable quotation, which ensures that the data requester can complete the data sharing at a lower price. In addition, reward the data providing nodes based on their performance in the current data sharing task and promote the nodes to provide high-quality data as much as possible. Through theoretical analysis and simulation experiments, the rationality and feasibility of the incentive mechanism designed in this paper are verified.

With the development of Internet of Things, big data, and other technologies, data sharing will become the key technology to meet these challenges. The data security sharing method based on blockchain and federated learning has broad development prospects. However, due to the need to use data model to replace the original data to participate in sharing, how to improve the effect of mapping the original data to the data model is a key problem. In addition, most incentive methods need to take into account the data model quality of the data provider. How to more reasonably quantify the quality of the data model is still a problem to be solved. In addition, the incentive mechanism in the data sharing scenario mostly involves the transaction between the two sides of data sharing. How to realize the transaction safely and efficiently remains to be studied.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in the Research and Application of Key Technologies of Multi-Level and Distributed Collaborative Interaction of Source-Network-Load-Storage Based on Blockchain (5700-202272179A-1-1-ZN).

References

[1] "International Data Corporation (IDC)," https://www.seagate.com/cn/zh/our-story/data-age-2025/, 2020-5.

[2] Z. Yang and W. Xiaoting, *Big data sharing and its obstacle analysis*, vol. 37, no. 4, 2017Library Work in Colleges and Universities, 2017.

[3] B. Wang, B. Li, and H. Li, "Oruta: privacy-preserving public auditing for shared data in the cloud," *IEEE transactions on cloud computing*, vol. 2, no. 1, pp. 43–56, 2014.

[4] S. Liu, Q. Qu, L. Chen, and L. M. Ni, "SMC: a practical schema for privacy-preserved data sharing over distributed data streams," *IEEE Transactions on Big Data*, vol. 1, no. 2, pp. 68–81, 2015.

[5] Y. Jingang, Z. Hong, L. Shu, L. S. Mao, and P. X. Ji, "Data sharing model for internet of things based on blockchain," *Journal*

of Chinese Mini-Micro Computer Systems, vol. 40, no. 11, pp. 2324–2329, 2019.

[6] X. Q. Dong, B. Guo, Y. Shen, X. L. Duan, Y. C. Shen, and H. Zhang, "An efficient and secure decentralizing data sharing model," *Chinese Journal of Computers*, vol. 41, no. 5, pp. 1021–1036, 2018.

[7] F. Song, M. Zhu, Y. Zhou, I. You, and H. Zhang, "Smart collaborative tracking for ubiquitous power IoT in edge-cloud interplay domain," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6046–6055, 2020.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," *International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, 2017.

[9] H. Yang, J. Zhao, Z. Xiong, K. Y. Lam, S. Sun, and L. Xiao, "Privacy-preserving federated learning for UAV-enabled networks: learning-based joint scheduling and resource management," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 10, pp. 3144–3159, 2021.

[10] J. Weng, J. Weng, J. Zhang, M. Li, Y. Zhang, and W. Luo, "Deep chain: auditable and privacy-preserving deep learning with blockchain-based incentive," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2438–2455, 2021.

[11] L. Lyu, J. Yu, K. Nandakumar et al., "Towards fair and privacy-preserving federated deep models," *IEEE Transactions on Parallel and Distributed Systems*, vol. 31, no. 11, pp. 2524–2541, 2020.

[12] Y. Qu, L. Gao, T. H. Luan et al., "Decentralized privacy using blockchain-enabled federated learning in fog computing," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5171–5183, 2020.

[13] Z. Chen, S. Chen, H. Xu, and B. Hu, "A security authentication scheme of 5G ultra-dense network based on block chain," *IEEE Access*, vol. 6, pp. 55372–55379, 2018.

[14] H. Zhi, H. Ge, and Y. Wang, "Cooperative communication method based on block chain for a large number of distributed terminals," *IEEE Access*, vol. 10, pp. 11679–11695, 2022.

[15] H. Kim, J. Park, M. Bennis, and S. L. Kim, "Blockchained on-device federated learning," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1279–1283, 2020.

[16] 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.

[17] K. Song, Q. R. Liu, and S. Wei, "Endogenous security architecture of Ethernet switch based on mimic defence," *Journal on Communications*, vol. 41, no. 5, pp. 18–26, 2020.

[18] Y. Liu and M. G. Peng, "6G endogenous security: architecture and key technologies," *Telecommunication Science*, vol. 36, no. 1, pp. 11–20, 2020.

[19] W. Y. Jiang, B. Y. Liu, and C. Wang, "Network architecture with intrinsic security," *Telecommunication Science*, vol. 35, no. 9, pp. 20–28, 2019.

[20] Z. Chen, H. W. Meng, and Z. Guan, "Research on intrinsic security in future internet architecture," *Journal of Cyber Security*, vol. 1, no. 2, pp. 36–45, 2016.

[21] Y. He, H. Li, X. Cheng, Y. Liu, C. Yang, and L. Sun, "A blockchain based truthful incentive mechanism for distributed P2P applications," *IEEE Access*, vol. 6, pp. 27324–27335, 2018.

[22] F. Song, Z. Ai, H. Zhang, I. You, and S. Li, "Smart collaborative balancing for dependable network components in cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 6916–6924, 2021.

[23] W. Chen, Y. Chen, X. Chen, and Z. Zheng, "Toward secure data sharing for the IoV: a quality-driven incentive mechanism with on-chain and off-chain guarantees," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1625–1640, 2020.

[24] H. Gao, C. H. Liu, J. Tang, D. Yang, P. Hui, and W. Wang, "Online quality-aware incentive mechanism for mobile crowd sensing with extra bonus," *IEEE Transactions on Mobile Computing*, vol. 18, no. 11, pp. 2589–2603, 2019.

[25] F. A. Al-Zahrani, "Subscription-based data-sharing model using blockchain and data as a service," *IEEE Access*, vol. 8, pp. 115966–115981, 2020.

[26] A. Fu, S. Yu, Y. Zhang, H. Wang, and C. Huang, "NPP: a new privacy-aware public auditing scheme for cloud data sharing with group users," *IEEE Transactions on Big Data*, vol. 8, no. 1, pp. 14–24, 2022.

[27] Y. Tao, P. Xu, and H. Jin, "Secure data sharing and search for cloud-edge-collaborative storage," *IEEE Access*, vol. 8, pp. 15963–15972, 2020.

[28] F. Song, Z. Ai, Y. Zhou, I. You, K. K. R. Choo, and H. Zhang, "Smart collaborative automation for receive buffer control in multipath industrial networks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 2, pp. 1385–1394, 2020.

[29] F. Song, L. Li, I. You, and H. Zhang, "Enabling heterogeneous deterministic networks with smart collaborative theory," *IEEE Network*, vol. 35, no. 3, pp. 64–71, 2021.

[30] V. Jaiman and V. Urovi, "A consent model for blockchain-based health data sharing platforms," *IEEE Access*, vol. 8, pp. 143734–143745, 2020.

[31] L. Zhang, "Reverse auction mechanism design with quality preference," in *2015 12th International Conference on Service Systems and Service Management*, pp. 1–5, Guangzhou, 2015.

[32] E. B. Sifah, Q. Xia, K. O.-B. O. Agyekum, H. Xia, A. Smahi, and J. Gao, "A blockchain approach to ensuring provenance to outsourced cloud data in a sharing ecosystem," *IEEE Systems Journal*, vol. 16, no. 1, pp. 1673–1684, 2022.

[33] B. J. Zhang, Y. C. Guo, and Z. K. Wang, "Research on data sharing incentive mechanism based on smart contract," *Computer Engineering*, vol. 48, no. 8, pp. 37–44, 2022.