Blockchain and Federated Learning Based Data Security Sharing Mechanism Over Smart City

Shaoyong Guo

Beijing University of Posts and Telecommunication

Baoyu Xiang (✉ 473429384@qq.com)

Beijing University of Posts and Telecommunications

Xuwei Xia

State Grid Ningxia Electric Power Company

Zhenhua Yan

State Grid Ningxia Electric Power Compy

Yongliang Li

State Grid Ningxia Electric Power Company

Research

Keywords: Federated learning, Blockchain, Smart city, Data sharing, Privacy protection

Posted Date: November 12th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-104012/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

RESEARCH

Blockchain and federated learning based data security sharing mechanism over smart city

Shaoyong Guo¹, Baoyu Xiang^{1*}, Xuwei Xia², Zhenhua Yan² and Yongliang Li²

*Correspondence:

473429384@qq.com

¹Department of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunication, Tucheng Road, Beijing, China
Full list of author information is available at the end of the article

Abstract

Data is the most important factor in building a smart city. City data is composed of many data islands, such as transportation, industry, and residents. In order to build a smart city, breaking data islands, achieving trusted and collaborative sharing of data, while protecting data privacy are essential. As a distributed ledger, the blockchain can solve the problem of data trust. Federated learning achieves data privacy protection by sharing model parameters instead of original data. However, it still has some problems such as malicious nodes and differential attacks. This paper proposes a data sharing mechanism that combines blockchain and federated learning over smart city. Firstly, the blockchain is combined to ensure the credibility of the performance information of the work nodes, then the work node selection algorithm is designed, and a consensus incentive mechanism IPoQ is proposed for efficient federated learning tasks. Finally, differential privacy technology is introduced to resist differential attack. Experimental results show that the methods proposed in this paper achieves an effective federated learning data sharing mechanism.

Keywords: Federated learning; Blockchain; Smart city; Data sharing; Privacy protection

1 Introduction

The key to building a smart city is to use big data analysis technology to realize the collaborative sharing of various data. Make intelligent decisions based on city data such as transportation, parks, factories, and residents' lives, etc. The problem that needs to be solved is how to ensure the trust of multiple parties and the privacy protection in the process of data sharing [1, 2, 3].

The federated learning method proposed by Google can precisely solve the current privacy security problems after data aggregation. Federated learning means that each terminal device retains its own data and does not share with each other. A central server or an alliance community publishes an initial model. Each device uses its own data to train the model, and then share model parameters. A central server transfers these local models. The model parameters are aggregated into a new global model, and this process is repeated until the model reaches the iterative convergence requirements. This training method changes data sharing to model parameter sharing, which solves the problem of privacy security [4, 5], helps industrial Internet and Internet of Things data security sharing, and accelerates the progress of data sharing collaboration [6]. However, the purpose of model training is to provide users with not bad service under the conditions of ensuring user privacy and security (which can be simply understood as undiminished accuracy and non-increased response

delay). Therefore, there are still the following urgent problems to be solved in federated learning. First, how to choose good work nodes, which have a positive impact on model training. The second is that the amount of data held by each worker node is different, and the work nodes with a large amount of data work together with those who have a small amount of data, but contribute more. Then an incentive method is needed to improve the initiative of the work nodes who contribute more [7]. The third is that federated learning is a problem of multi-party participation. There is a malicious or untrusted party interfering with federated learning. In order to ensure credible collaborative work of multiple parties, blockchain technology can be incorporated into it [8]. In order to solve the above related problems and better protect user privacy, we combined blockchain technology and differential privacy technology into federated learning [9, 10]. Blockchain network can ensure the safe and trusted storage of local model and historical work records of work nodes, and provide corresponding rewards. At the same time, differential privacy technology adds some disturbances to the model to better protect user privacy.

Based on the above problems, we have proposed a data security sharing mechanism based on blockchain and federated learning. In this paper we made the following contributions:

- (1) Combining blockchain technology and federated learning technology, build a data security sharing model over smart city, and design basic data sharing processes.
- (2) To achieve reliable federated learning, design a kind of work nodes selection algorithm based on historical performance and data network transmission quality.
- (3) Modify the consensus method of the blockchain and design an incentive mechanism consensus algorithm based on model training quality, which can achieve the purpose of encouraging excellent work nodes to work and simplify the consensus process to reduce resource consumption and time delay.
- (4) With the goal of balancing data security and model practicality, combine the differential privacy algorithm implemented by Laplace mechanism with federated learning.

2 Related work

Federated learning is proposed to solve the problem of user privacy leakage in machine learning tasks, and it has attracted much attention in academic and industry. Google first proposed the concept of federated learning and applied it to the virtual keyboard application of smartphones [11]. The author in [6] combined the blockchain technology with federated learning to solve the trust problem of multi-party participation, ensure data sharing and collaboration, and specifically analyzed its role in the Industrial Internet of Things. The author in [7] proposed an incentive mechanism combined with contract theory to encourage work nodes to work actively to ensure reliable federated learning. In [8], the author proposed a blockchain federated learning architecture, and analyzed its end-to-end delay, optimized the training speed of each iteration model, and from the three aspects of information exchange, calculation and consensus to optimize the block generation rate. At the same time, in order to better protect user privacy and prevent differential attacks, it is also necessary to integrate differential privacy technology into blockchain-based federated learning. [9] elaborated on the definition and background of differential privacy,

and proposed an individual differential privacy method to improve differential privacy and improve data usability. In [12], the weights are considered according to each customer's local learning accuracy and participation frequency, and two key performance indicators are selected, namely learning speed and standard deviation to optimize the federated learning task. [13] mainly analyzes and optimizes the uncertainties of the federated learning task in the wireless channel and the heterogeneous power constraints of local mobile devices and the difference in local data size. The authors in [14] used a multi-objective evolutionary algorithm to optimize the structure of the neural network model in federated learning to simultaneously minimize communication costs and global model test errors. The scalable method used to encode network connectivity is suitable for federated learning to enhance the efficiency of evolving deep neural networks. Significantly reduce communication costs, and improve the learning performance of federated learning.

The predecessors have done a lot of work in this field to optimize the use of federated learning algorithms in practice. Our main work is to propose a data sharing mechanism over smart city based on blockchain and federated learning, in which a reliable work nodes selection algorithm is designed, and based on the PoQ (Proof of training Quality) consensus algorithm proposed in [6], the incentive mechanism and a slight modification are added to it to propose a consensus incentive mechanism IPoQ (Incentive in Proof of model Quality), while applying differential privacy technology to model interaction. Finally, we described the entire data sharing process in detail.

3 Architecture of system model

Data is one of the most important resources in the information age. The sharing of data not only enables better development of science and technology, but also enables more intelligent applications suitable for different scenarios. But the problem is that data sharing and data protection seem to be difficult to guarantee at the same time. If data is to be shared, data security is easily threatened; and if data is to be strictly protected, it is difficult to break the information islands.

The traditional data sharing model uploads the raw data to the central server, and then this central server uses these data for model training, and finally provides better services. However, there is a big security problem in this way. First, each user uploads his data, which is tantamount to exposing his privacy. Second, once the central server is attacked, there is a risk of information leakage and service paralysis [15]. In response to these problems, we propose a data security sharing mechanism based on blockchain and federated learning, as shown in Fig. 1. Its feature is that it builds a credible and safe data sharing environment to realize the construction of a smart city under the premise of protecting data privacy.

As shown in Fig. 1, the bottom layer is a component of a smart city, including industrial park, smart transportation, smart home and automated factory. They train their local data locally to form a local model, and save the relevant data in the upper blockchain network. At the same time, they also get the aggregated global model from the blockchain layer.

The blockchain layer is composed of base stations connected to various parts of the smart city and communicates through these base stations. It has the characteristics

of decentralization, immutability and traceability, which guarantees the credibility of data storage and the stability and reliability of the system. Using the smart contract technology in the blockchain, we can easily select reliable work nodes based on the historical performance of nodes stored in the chain, and reward them based on their work performance (the quality of the trained model), while solving the problem of mutual trust in the process of multi-party collaborative training [12]. On the other hand, some parks or communities in the city can form a trusted alliance as a sub-chain of the blockchain platform layer under the federated learning task to communicate with it to reduce performance problems caused by communication delays.

The top-level layer is federated learning layer. Each department publishes tasks according to its own needs, and organizes the entire federated learning task through the blockchain layer. At the same time, each department can set its own incentive price to obtain the maximum benefit.

4 Detailed design of system model

In this section, we mainly elaborate the design details of the proposed data sharing mechanism based on blockchain and federated learning. This paper introduces the work node selection algorithm based on historical performance and data network transmission quality, incentive mechanism based on consensus mechanism modification, and differential privacy technology implemented by Laplace mechanism which balances data privacy and model practicability. Finally, the whole data sharing process is introduced in detail

4.1 Reliable work nodes selection

In order to train a better model during the federated learning process, it is necessary to select reliable and efficient work nodes [7]. The performance of these nodes

is to use more local data and more resources for model training, making the global model accuracy rate higher. In our proposed model, the historical work performance of each work node is recorded in the blockchain network, as shown in block content in Fig.1, the block stores the work performance of each work node, local model and final global model. In the process of selecting reliable work nodes, we make a comprehensive score based on the previous work performance stored in the blockchain and their network quality at the time. The parameter n must be determined first, which means that the latest n performances of each work node. Then determine the score weights w_i corresponding to the n times of work performance respectively. Considering that the most recent work performance is more representative of the current working state of the work node, the weights will decay with time, and the recent work performance will have a higher score weight.

We use R_j to represent the overall work performance score of the j -th work node, then has the following formula:

$$R_j = \sum_{i=1}^n w_i r_i \quad (1)$$

Where r_i represents the i -th job performance of the worker node stored in the blockchain, and $\sum_{i=1}^n w_i = 1$.

Each performance of the work node stored in the blockchain is determined by the amount of data trained by the node, the computing resources contributed, and the accuracy of each local model. Using D_i , C_i , and A_i to represent the amount of data contributed, the calculation resource and the accuracy of local model training by the work node at the i -th job, then r_i can be expressed as the following formula:

$$r_i = \alpha D_i + \beta C_i + \chi A_i \quad (2)$$

where $\alpha + \beta + \chi = 1$.

In order to ensure the correct selection of reliable work nodes, the amount of data contributed should be valid data related to the task or beneficial to task training. The computing resources are subject to the CPU and GPU resources contributed by the work nodes. The accuracy of the local model is the arithmetic average of the accuracy of each local training during the training iteration of the work nodes.

When choosing a reliable working node, the quality of its data transmission network is also the focus of consideration. First, we evaluate the packet loss rate during data transmission, and use the packet loss rate to estimate the next possible packet loss.

$$P_{loss} = \frac{1}{m+1} \quad (3)$$

where m is the number of data packets that have successfully assaulted during the period from the last packet loss to the current packet loss. Due to the complex uncertainty in the heterogeneous network, the loss rate measured only with the current data packet loss is not accurate enough and has large jitter. To prevent

this, the weighted average of the current and past packet loss rates can be used to smooth the value to better predict the future packet loss rate.

$$P_{loss} = \delta P_{last} + (1 - \delta) P_{current} \quad (4)$$

When δ is closer to 1, P_{loss} is more affected by the past value. Conversely, the closer δ is to 0, it means that the past value has less influence on P_{loss} , which is mainly determined by the current value. Therefore, in order to better predict the future packet loss rate and evaluate the network transmission quality, the parameter δ need to be carefully considered.

Finally, we express the data transmission network quality Q of the work node as follows:

$$Q = \begin{cases} \frac{1}{rtt} \left\lfloor \frac{1}{P_{loss}} - 1 \right\rfloor, & \left\lfloor \frac{1}{P_{loss}} - 1 \right\rfloor \leq l \\ \frac{l}{rtt}, & \left\lfloor \frac{1}{P_{loss}} - 1 \right\rfloor > l \end{cases} \quad (5)$$

where rtt is the round trip time and l is the size of sliding window of TCP [16].

Then, we set a threshold. When Q is greater than this threshold, we consider that the network transmission quality of the work node meets the requirements.

After calculating the performance score of each working node, sort them in descending order according to the R_j value, and then select the top k nodes with reliable network status as the work nodes participating in the task according to the current network quality of each work node. It will be recorded in the blockchain as the basis for the next selection.

4.2 Incentive in proof of model quality

In order to ensure reliable federated learning, it is necessary for us to give corresponding rewards to the work nodes who perform well in the task training process to motivate them to work better in the future and participate more actively in the federated learning tasks [8]. Therefore, we have designed a consensus incentive mechanism in the context of the alliance chain-Incentive in Proof of model Quality (IPoQ), which applies the incentive mechanism to the consensus of the blockchain. This design has two advantages:

(1) Implemented an incentive mechanism in federated learning, allowing hard-working nodes to receive corresponding rewards, encouraging the work nodes to actively participate in federated learning tasks, and strive to do their best.

(2) Simplify the consensus mechanism of the blockchain. The original blockchain's proof-of-work (PoW) consensus mechanism makes the calculation of the consensus process and the cost of communication resources consume too much. IPoQ effectively simplifies the consensus process and makes the blockchain's resources to be used more efficiently.

When start a new federated learning task, according to the above-mentioned method of work nodes selection, select k reliable work nodes. We call these K work nodes the consensus alliance of this federated learning task. They are responsible for achieving the consensus process of final bookkeeping and rewards.

The purpose of the federated learning task is that when a new task is released into the blockchain, the selected work nodes jointly train a global model M . We can choose various machine learning algorithms for training, such as LG, GBDT, SVM and some deep learning methods.

After determining the model, each work node uses its own local data and resources to train the local model. In order to evaluate the quality of the model, the classification task can adopt Accuracy, Precision, Recall, the harmonic average of accuracy and recall (F1-score); the regression task can take Root Mean Square Error (RMSE), Relative Squared Error (RSE), Mean Absolute Error (MAE), Relative Absolute Error (RAE) and other methods. We use the average absolute error to evaluate the model:

$$MAE(M_i) = \frac{1}{N_{task}} \sum_{i=1}^{N_{task}} |y_i - M_i(x_i)| \quad (6)$$

In the above formula, (x_i, y_i) is the test data provided by the task publisher, $M_i(x_i)$ is the result predicted by the model, and N_{task} is the number of test data provided by the task publisher. The lower the MAE, the higher the prediction accuracy of the model.

The final MAE of each working node is expressed by the following formula:

$$MAE(W_i) = \frac{1}{N_{iter}} \sum_{i=1}^{N_{iter}} MAE(M_i) \quad (7)$$

where N_{iter} indicates the number of iterations.

The consensus and incentive process is as follows:

(1) The consensus alliance (reliable work nodes selected by the work node selection method to participate in this federated learning task) calculates the MAE of each work node based on the transaction information of the previous local model, and then combining formulas (6) and (7).

(2) According to the MAE of each work node, the corresponding reward is given to each node, which is formulated by the task issuer, and the smaller the MAE, the greater the reward.

(3) The work node with the smallest MAE value is also selected as the master node, responsible for implementing the consensus process of the blockchain, packaging all the recorded information into a block, and broadcasting the block to other nodes in the blockchain. After the work node in the consensus alliance verifies that the block information is correct, the block information is recorded in the blockchain.

We consider a simple reward distribution mechanism. First, if the MAE value of a working node participating in a task is large and not within the acceptable range, then it will be removed from the reward list, and even punitive measures can be taken. Then distribute the rewards to the working nodes in the reward list according to the following formula:

$$S_{W_i} = S \frac{MAE_{re}(W_i)}{\sum_{i=1}^{l_{reward}} MAE_{re}(W_i)} \quad (8)$$

$$MAE_{re}(W_i) = \frac{1}{MAE(W_i)} \quad (9)$$

where S_{W_i} is the reward assigned to the i -th work node, S is the reward value provided by the task publisher and l_{reward} is the length of the reward list.

4.3 Differential privacy

In order to protect the user's privacy security, and better carry out the federated learning task, we use the locally uploaded trained model results instead of the original data. This method greatly protects the privacy of users, but there is still the danger of differential attacks. Some attackers can steal the user's private information to a certain extent by deducing the results. Therefore, in the process of federated learning, we combined the differential privacy algorithm to process the local model to prevent illegal attackers from stealing user privacy data [10, 17].

Differential privacy technology has received great attention in both academia and industry. It has been used in Apple's IOS system and Google's chrom browser to protect user privacy [11].

Definition of differential privacy: given two data sets $D1$ and $D2$ that differ by at most one piece of data, select an algorithm A and let $Result(A)$ be the output results of A . If algorithm A has the output result O on $D1$ and $D2$ ($O \in Result(A)$) that meets the following inequality:

$$P(A(D1) = O) \leq e^\epsilon * P(A(D2) = O) \quad (10)$$

which is said that algorithm A satisfies the differential privacy with parameter ϵ [6].

There are two mechanisms to achieve differential privacy, one is the Laplace mechanism and the other is the exponential mechanism. The Laplace mechanism is often used in the mechanism of numerical output. Exponential mechanism is often used for non-numerical output mechanisms (meaningless functions after adding numerical noise).

We use the Laplace mechanism to achieve differential privacy, that is, add the Laplace noise to the original model to obtain a new model for data exchange. So the new model is:

$$M^\Delta = M + Laplace\left(\frac{s}{\epsilon}\right) \quad (11)$$

where s is the sensitivity of the Laplace mechanism, that is defined as follows:

$$s = \max_{D1, D2} |M(D1) - M(D2)| \quad (12)$$

$Laplace\left(\frac{s}{\epsilon}\right)$ is Laplace noise (a random value that satisfies the Laplace distribution). The probability density function of this division is expressed as follow:

$$f(x|\mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}} \quad (13)$$

The probability cumulative function of the distribution from the probability density function is as follow:

$$F(x|\mu, b) = \begin{cases} \frac{1}{2}e^{-\frac{\mu-x}{b}}, & x < \mu \\ 1 - \frac{1}{2}e^{-\frac{x-\mu}{b}}, & x \geq \mu \end{cases} \quad (14)$$

where μ is Expectation and $2b^2$ is the variance. After each working node has trained the local model locally, the local model parameters are subjected to differential privacy protection measures implemented by the Laplacian mechanism (that is, Laplacian noise is added to the model parameters). We share the new model M^Δ with the blockchain layer who aggregates all uploaded local models into a global model and distribute it to all working nodes.

The smaller the parameter ε , the higher the degree of privacy confidentiality, but the data availability will be reduced, so in order to balance the protection of data privacy and the availability of the model, the appropriate ε should be selected.

4.4 Data sharing process

Based on the data sharing mechanism of blockchain and federated learning, aggregation is achieved by transferring locally trained models instead of sharing local original data, which protects user privacy to a certain extent. The storage of relevant data in the blockchain ensures the trusted environment of multi-party participation while protecting privacy [18]. Fig. 2 is the basic process of data sharing based on blockchain and federated learning.

In order to better achieve federated learning and protect user privacy, we have designed the work nodes selection algorithm, an incentive mechanism, and integrated differential privacy technology into it. The detail of data sharing process is as follows:

(1) Those in need can apply to join the data sharing mechanism to become a task publisher. The task publisher publishes a federated learning task through the blockchain platform. The task can be varied according to requirements, such as voice recognition, face recognition, fault detection, behavior tracking, etc. The task publisher also gives rewards for motivating working nodes in the learning task.

(2) When a new task arrives, the blockchain platform retrieves the information in the blockchain. If the task has been processed before, there are two ways to deal with it. One is to use the previous global model as the initial model of the federated task, and then perform federated training. The other is that it can directly return the previous global model as the final result. If it is a new task, a specific initialization method is used to obtain the initial model.

(3) With our designed work nodes selection algorithm, according to the previous work performance of each working node (records stored in the blockchain) and the data network transmission quality at the time, reliable working nodes are selected to participate in the federated learning task.

(4) The selected working nodes perform federated learning tasks and iteratively train local models. Then select the appropriate parameters ε , and on the premise of not affecting the practicality of the model, use differential privacy technology to aggregate and share local models through the blockchain platform to achieve iterative training of the global model.

(5) According to the performance of the local model training of each node, the consensus incentive mechanism-IPOQ that we designed is used to reward the contribution of each work node in this task accordingly. The specific reward value obtained by each working node is determined by the reward distribution mechanism we designed.

(6) The working node with the smallest MAE value is selected as the master node, responsible for packaging relevant information into blocks and broadcasting to other nodes for verification, and finally recording in the blockchain to complete the consensus process. These information serves as the basis for the next federated learning task.

(7) After many iterations of training, a global model that meets the requirements is obtained. Then blockchain platform returns the final global model of federated learning training to the task publisher. Finally, in order to better optimize the federated learning mechanism, feedback will be obtained from the task publisher after each task, mainly in terms of model training time and accuracy.

5 Experiments

In this section, we mainly do some numerical experiments to verify the effectiveness of some of the methods we propose.

We use the public data set MNIST, which contains 60,000 training sets and 10,000 test sets. Each picture in the dataset consists of 28 x 28 pixels, and each pixel is

represented by a gray value. We expand the 28 x 28 pixels into a one-dimensional row vector. These row vectors are the rows in the picture array (784 values per row represent a picture). We use a simple neural network built by keras for model training and prediction, and divide the MNIST data set into small data sets of different sizes as local data of each work node.

Fig. 3 shows the difference in accuracy between the proposed method and the traditional method at different iteration times. It can be seen from the experimental data that under different iterations, the accuracy of the non-federated learning methods in the data set is the highest, and compared with the traditional federated learning methods, the accuracy rate will be reduced. However, the accuracy of the work nodes selection algorithm designed by us has improved. In order to better protect user privacy and prevent differential attacks, after combining differential privacy technology, the accuracy rate has slightly decreased, but after the number of iterations has increased, the decline rate relative to the traditional federated learning method is within an acceptable range. It can be concluded from this that the method we designed still has a higher accuracy rate while better protecting user privacy.

Fig. 4 shows the effect of different differential privacy parameters on the accuracy of the federated learning method we designed on the test set. In order to better protect user privacy, ϵ should be as small as possible, but it can be seen that the smaller the ϵ , the lower the accuracy will be, and the greater the decline. Conversely, when ϵ becomes larger, the accuracy rate increases, and eventually tends to a constant value (that is, when ϵ infinity, the differential privacy technology fails, and the model effect is the same as that without differential privacy protection). Therefore, we can think that in order to achieve the goal of balancing data security

and model practicality, appropriate parameters ε should be selected. From our experimental results, we can choose the value of the inflection point as a parameter, in our experiment ε should take a value of about 1.0.

Fig. 5 shows the effect of different training epochs on the accuracy of federated learning when the value of differential privacy parameter ε is 1.0. It can be seen that when considering the balance between model privacy and model usability, the value of ε is 1.0. The number of model training epochss is 500, the accuracy rate may reach the maximum. When epoch is equal to 1000, the model overfits and the accuracy rate drops.

Fig. 6 shows the effect of different training batch sizes on the accuracy of federated learning when the value of differential privacy parameter ε is 1.0. At this time, the

epoch is set to 100 in order to get the experimental results faster. When the batch size is smaller, the accuracy is higher, but it increases the training time.

6 Results and Discussion

The construction of a virtual city based on informatization—smart city is in full swing and has become a standard feature in almost every city. Its large industry, far-reaching influence, and complex construction have become the common concern of all industries around the world. How to solve a large amount of smart city data sharing, the repeated construction of smart city systems, and how to provide more secure data to city enterprises are all problems that smart city construction has to face. In addition to helping to break information islands, federated learning is also of great benefit in ensuring data security. Therefore, federated learning will become an indispensable necessity in the development of smart cities in the future.

In this paper, we propose a data security sharing mechanism based on blockchain and federated learning over smart city. In response to the shortcomings of the original method, we designed the work nodes selection algorithm to enhance the effectiveness of the federated learning task, and designed a consensus incentive mechanism to encourage the work node to more actively participate in the task, and combined with differential privacy technology, it is a good balance between privacy security and the practicality of the model. The numerical experiment results also verify the superiority and feasibility of the method we designed to a certain extent.

In the future, we will also consider issues related to the specific practice of federated learning to build smart cities, such as better combining federated learning with the cloud, combine work node selection and incentive mechanism with deep reinforcement learning [19] and improving the efficiency of federated learning when the actual data scale is increasing.

Abbreviation

PoQ:Proof of training Quality
 IPoQ:Incentive in Proof of model Quality
 PoW:proof-of-work
 LG:Logistic Regression
 GBDT:Gradient Boosting Decision Tree
 SVM:Support Vector Machine

Competing interests

The authors declare that they have no competing interests.

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Author's contributions

Shaoyong Guo and Baoyu Xiang came up with the proposed scheme and wrote the entire article. All authors read and approved this manuscript.

Acknowledgements

The authors thank the person who provided meticulous and valuable suggestions for improving the paper

Funding

This work is supported by Nation Key R&D Program of China(2019YFB2102302).

Author details

¹Department of State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunication, Tucheng Road, Beijing, China. ²Electric Power Research Institute, State Grid Ningxia Electric Power Co., Ltd., Ningxia, China.

References

- O. B. Mora, R.R., V. M. Larios, J.R.B.-R., R. Maciel, A.O.: A use case in cybersecurity based in blockchain to deal with the security and privacy of citizens and smart cities cyberinfrastructures. In: 2018 IEEE International Smart Cities Conference (ISC2), pp. 1–4 (2018)
- Dattana, V., Gupta, K., Kush, A.: A probability based model for big data security in smart city. In: 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), pp. 1–6 (2019)
- Huerta, J., Salazar, P.: Audit process framework for data protection and privacy compliance using artificial intelligence and cognitive services in smart cities. In: 2018 IEEE International Smart Cities Conference (ISC2), pp. 1–7 (2018)
- Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology* **10**(2), 1–19 (2019)
- Konen, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency (2016)
- Lu, Y., Huang, X., Dai, Y., Maharjan, S., Zhang, Y.: Blockchain and federated learning for privacy-preserved data sharing in industrial iot. *IEEE Transactions on Industrial Informatics* **PP**(99), 1–1 (2019)
- Kang, J., Xiong, Z., Niyato, D., Xie, S., Zhang, J.: Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal* **6**(6), 10700–10714 (2019)
- Kim, H., Park, J., Bennis, M., Kim, S.L.: Blockchain-based on-device federated learning. *IEEE Communications Letters* **24**(6) (2019)
- Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., Megías, D.: Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security* **PP**(99), 1–1 (2016)
- Jiang, L., Lou, X., Tan, R., Zhao, J.: Differentially private collaborative learning for the iot edge. In: EWSN, pp. 341–346 (2019)
- Hard, A., Rao, K., Mathews, R., Beaufays, F., Ramage, D.: Federated learning for mobile keyboard prediction (2018)
- Kim, Y.J., Hong, C.S.: Blockchain-based node-aware dynamic weighting methods for improving federated learning performance. In: 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS) (2019)
- Tran, N.H., Bao, W., Zomaya, A., Nguyen, M.N.H., Hong, C.S.: Federated learning over wireless networks: Optimization model design and analysis. In: IEEE INFOCOM 2019 - IEEE Conference on Computer Communications (2019)
- Zhu, H., Jin, Y.: Multi-objective evolutionary federated learning (2018)
- Majeed, U., Hong, C.S.: Flchain: Federated learning via mec-enabled blockchain network. In: 2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS), pp. 1–4 (2019)
- Fu, W., Gong, X., Dong, P.: Improve mptcp transmission performance based on quality evaluation algorithm in high-loss network environment. In: 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC) (2020)
- Geyer, R.C., Klein, T., Nabi, M.: Differentially private federated learning: A client level perspective (2017)
- Liu, C.H., Lin, Q., Wen, S.: Blockchain-enabled data collection and sharing for industrial iot with deep reinforcement learning. *IEEE Transactions on Industrial Informatics*, 3516–3526 (2019)
- Zhan, Y., Li, P., Qu, Z., Zeng, D., Guo, S.: A learning-based incentive mechanism for federated learning. *IEEE Internet of Things Journal* **7**(7), 6360–6368 (2020)

Figures

Figure 1 Blockchain and federated learning based data sharing model over smart city. The overall architecture and scenario of the model

Figure 2 Basic process of data sharing based on blockchain and federated learning. Describe the workflow of the proposed method in the application scenario.

Figure 3 Accuracy for different methods with different iterations. Experimental results 1.

Figure 4 Impact of differential privacy paramment ϵ for our method. Experimental results 2.

Figure 5 Accuracy of different epoch when $\epsilon = 1.0$. Experimental results 3.

Figure 6 Accuracy of different batch size when $\epsilon = 1.0$. Experimental results 4.

Figures

Figure 1

Blockchain and federated learning based data sharing model over smart city. The overall architecture and scenario of the model

Figure 2

Basic process of data sharing based on blockchain and federated learning. Describe the workflow of the proposed method in the application scenario.

Figure 3

Accuracy for different methods with different iterations. Experimental results 1.

Figure 4

Impact of differential privacy paramant ϵ for our method. Experimental results 2.

Figure 5

Accuracy of different epoch when $\varepsilon = 1.0$. Experimental results 3.

Figure 6

Accuracy of different batch size when $\varepsilon = 1.0$. Experimental results 4.