Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 7705843, 9 pages
https://doi.org/10.1155/2021/7705843

Research Article

Industrial Internet Federated Learning Driven by IoT Equipment ID and Blockchain

Xu Zhang,[1] **Haibo Hou**,[1] **Zhao Fang**,[2] **and Zhiqian Wang**[2]

[1]*China Academy of Information and Communications Technology, China*
[2]*Guangzhou Institute of Internet of Things, China*

Correspondence should be addressed to Xu Zhang; zhangxu1@caict.ac.cn

Received 9 July 2021; Revised 12 September 2021; Accepted 20 September 2021; Published 8 November 2021

Academic Editor: Bo Rong

With the development of Internet of Things (IoT), 5G, and industrial technology, Industrial Internet has become an emerging research field. Due to the industrial specialty, higher requirements are put forward for time delay, safety, and stability of the identification analysis service. The traditional domain name system (DNS) cannot meet the requirements of industrial Internet because of the single form of identification subject and weak awareness of security protection. As a solution, this work applies blockchain and federated learning (FL) to the industrial Internet identification. Blockchain is a decentralized infrastructure widely used in digital encrypted currencies such as Bitcoin, which can make secure data storage and access possible. Federated learning protects terminal personal data privacy and can carry out efficient machine learning among multiple participants. The numerical results justify that our proposed federated learning and blockchain combination lays a strong foundation for the development of future industrial Internet.

1. Introduction

In recent years, countries worldwide have paid more and more attention to the development of Industrial Internet. Industrial Internet is an important cornerstone of the fourth industrial revolution and a key measure to transform old kinetic energy into a new one [1]. With the development of Industrial Internet in the past few years, a single form of equipment and different types of enterprises have been connected by the Industrial Internet, which allows the resources of different links to be organically combined. The Industrial Internet system architecture consists of four aspects: network connection, platform, security system, and identification analysis system [2]. Among them, the network is used to realize the connection of people, machines, and things, and it is the foundation of Industrial Internet. The security system is responsible for providing security protection and guarantee, and the purpose of the platform is to open up operational data and Internet data to integrate resources. The identification analysis system is an important hub for the realization of Industrial Internet.

The traditional domain name system (DNS) resolution service faces serious challenges in terms of subject identification, resolution methods, security, and service quality, and it cannot meet the needs of industrial networks. The main reasons can be summarized as follows: change of subject identification, mass data and ultralow latency requirements, security and privacy protection, fairness, and reciprocity. It is mainly because the blockchain system has the characteristics of calculation, storage, and scalability, while federated learning (FL) has the characteristics of ensuring information security during big data sharing and exchange and protecting the privacy of terminal personal data. This paper uses blockchain to store important information about devices; the model uses the convolutional neural network (CNN) of FL as the baseline. The technologies of blockchain and federated learning are applied to Industrial Internet identification analysis, which enables identification analysis to play a more important and irreplaceable role in the Industrial Internet field.

The chain storage structure of the blockchain can comprehensively record the data generated by Industrial Internet companies in the production and operation process, which

makes the data nontamperable, thereby ensuring the authenticity and credibility of the data. This is also advantageous for Industrial Internet companies to reduce costs and improve efficiency. The privacy protection of blockchain technology are applied due to the advantages of trusted collaboration; it can be deeply integrated with the Industrial Internet in terms of data confirmation, accountability, and transactions, thereby promoting the transformation of industrial production to digital and intelligent [3]. Feng et al. [4] introduced four core technologies of blockchain: decentralization, consensus mechanism, encryption algorithm, and smart contracts. In the blockchain, users jointly create a public ledger for block verification and transaction records [5]. [6]. Blockchain technology has laid a solid foundation for earning trust and created a reliable cooperation mechanism, and it has a wide range of application prospects [7, 8].

Federated learning is a solution for machine learning and artificial intelligence (AI) to face more stringent data management regulations. In the framework of federated learning, the central server saves global data that can be initially shared, and each client saves local data and trains local machine learning and artificial intelligence models based on the local data. Then, according to a certain communication mechanism, client transmits the model parameters and other data to the central server. The central server collects the data uploaded by each client and conducts training to build a global model; each client has the same role and status in the entire federated learning mechanism [9]. Federated learning effectively solves the problem of the client sharing data between two or more data without contributing data, so it solves the problem of data islands to a large extent.

The sensors and IoT devices deployed in the Industrial Internet of Things generate massive amounts of sensor data, and the analysis of sensor data can promote industrial production and manufacturing. When federated learning analyzes and processes massive amounts of sensor data, there is no need for data interaction between devices, so the privacy of local data can be guaranteed. For example, in the process of anomaly detection in the Industrial Internet of Things [10], the privacy of local data can be guaranteed by using federated learning, and there is no need to interact with local data between devices, which can improve the ability to detect abnormal IoT nodes in the process of anomaly detection. At the same time, data-driven cognitive computing (D2C) faces some important bottlenecks in the Industry 4.0 scenario [11]. In order to solve the problem of privacy leakage in cognitive computing, federated learning can be used in cognitive computing in the Industrial Internet of Things to protect data security and prevention of privacy leaks. However, if the central server fails or has a trust issue, then all computing and information security cannot be guaranteed, and a single point of failure will occur [12]. The decentralized distributed data storage structure of the blockchain can remove the trusted central authority, so it can solve the trust problem of the central server in the federated learning, thereby preventing the single point of failure. Therefore, the use of blockchain technology based on federated learning can prevent single-point failure problems, and the verification mechanism of the blockchain can ensure the authenticity of data while selecting high-quality and credible edge device [11].

The main contributions of this article are summarized as follows:

(1) It is necessary to include privacy protection in the equipment identification of Industrial Internet. We utilize federated learning to guarantee information security during large data sharing and exchange, as well as safeguard the privacy of terminal personal information

(2) We propose a novel framework that applies blockchain and federated learning technology to the research of Industrial Internet identification. The performance of the proposed framework is verified by solid numerical results

The rest of this paper is organized as follows. Section 2 presents the related work. Section 3 introduces blockchain technology as well as its application in industrial Internet. Section 4 proposes our solution of blockchain-based federal learning with numerical results, followed by Section 5 to conclude the paper.

2. Related Work

With the rapid development of Internet of Things (IOT), 5G networks, and industrial technology, some new applications such as smart cities, virtual reality, and industrial intelligent production continue to emerge. The number of wearable devices, industrial machines, and different types of sensors has exploded, which indicates that the future network is transforming from a consumer to a production model. The particularity of industrial production requires that industrial networks can perceive environmental information through intelligent means, support a large number of heterogeneous device access, support massive multisource, multimodal data high-speed transmission, and have stronger security, thus providing better service for enterprise production. This has brought huge challenges to the traditional Internet in terms of architecture, security, and performance [13].

Different from the consumer Internet and the traditional IOT, industrial Internet has diverse communication subjects and higher performance requirements, and traditional DNS resolution services cannot meet their needs. In order to meet the characteristics and requirements of industrial Internet, its identification resolution system must follow these principles: support for multisource heterogeneous communication subjects, security guarantee for identification resolution services in complex environments, fair and equal guarantee for participation of multiple organizations, effectiveness guarantee in scenarios with multiple protocols, high concurrency and differentiated requirements, and providing scalability at the protocol level and system level. In response to the principles, considering the calculation, storage, and scalability characteristics of the blockchain system, federated learning has the characteristics of ensuring information security during big data sharing and exchange and

protecting the privacy of terminal personal data. The technology of blockchain and federated learning are applied to industrial Internet identification analysis, which promote the development of identification analysis in the field of industrial Internet.

Blockchain is a new type of technology that has gradually emerged with encrypted digital currency. It adopts a distributed computing model, uses blockchain to store data, and uses cryptographic principles to ensure the security of transmission and access. Data storage is jointly maintained and supervised by Internet users. It has distinctive features such as decentralization, transparency and openness, and unmodifiable data [14]. It uses distributed storage and calculation to ensure that the entire network node has the same rights and obligations, and the data in the system are essentially maintained by the nodes of the entire network. By this means, the blockchain no longer depends on the central processing node to realize the distributed storage, record, and update of data. Therefore, its application is not limited to currency as an asset type, and its application research in various industries is relatively extensive. Chi et al. [15] summarized the existing blockchain technologies in several typical fields, and then gave the main problems and countermeasures in the development process of blockchain in detail, and finally discussed the prospects and forecasts of blockchain. Blockchain is a distributed ledger technology that relies on logical control functions such as smart contracts to evolve into a complete storage system. Changes in its classification methods, service models, and application requirements have spawned to the diversified core technologies. In order to fully understand the blockchain ecosystem, Bao et al. [16] designed a hierarchical blockchain technology architecture and further analyzed the basic principles, technical associations, and research progress of each layer structure of the blockchain. The frontier application directions of blockchain such as the industrial Internet and smart cities are given at the end.

Federated learning based on client-server architecture and distributed machine learning [17] are both used to process distributed data, but there are differences between them in terms of application fields, data attributes, and system composition [18]. Federated learning algorithms can be divided into machine learning-based algorithms and deep learning-based algorithms. Federated learning has the following characteristics.

(i) Supporting nonindependent and identically distributed data: the federated learning algorithm performs well in nonindependent and identically distributed data. In the actual use of federated learning, the data quality and distribution of the data holder is uncontrollable. The data of the holder cannot be required to meet independent and identical distribution, so the federated learning algorithm needs to support nonindependent and identically distributed data

(ii) Efficient communication: federated learning needs to consider the system heterogeneity of the data holder to improve communication efficiency and reduce communication loss without losing accuracy or loss

(iii) Fast convergence: in the process of joint modeling, it is necessary to ensure the convergence of the model and at the same time increase the convergence speed

(iv) Security and privacy: since data privacy security is an important feature of federated learning, security and privacy are two necessary requirements for federated gradient updates. Security and privacy can be carried out in the aggregation process through encryption and other methods and can also be reflected in the process of stand-alone optimization

(v) Support complex users: complex users refer to the large number of users and the imbalance or deviation of user data. The federated optimization algorithm needs to have good compatibility ability to deal with this situation

Based on the advantages of the above federated learning algorithm, information security can be guaranteed during big data sharing and exchange, and the privacy of terminal personal data is protected. Furthermore, it can carry out high-efficiency machine learning among multiple participants or multiple computing nodes. Zhang et al. [19] made a comprehensive review of recent research and achievements in federal learning and presented future development trends. First, data islands and privacy protection are described to introduce the background of federated learning, and the connotation and mechanism of federated learning are outlined. Then, typical application cases of data sharing and exchange based on federated learning technology are introduced.

3. Blockchain for Industrial Internet

Currently, mainstream object recognition systems include object identifier (OID) recognition system, Ecode recognition system, and handle recognition system. The OID identification system has simple coding rules, good flexibility, and scalability. Therefore, the OID identification system is adopted to apply the traceability system of industrial equipment. We first propose the overall framework design of the system in this section. At the same time, in order to ensure the security of the system, the key and authentication mechanism and the authority management mechanism are proposed, and then, the traceability scheme inside the blockchain and the traceability scheme outside the blockchain are designed in detail.

3.1. System Security Design

(1) Key and authentication mechanism: the key and authentication mechanism mainly include the identification of the traceable company's identity information and the distribution of key pairs. It uses cryptographic techniques such as asymmetric encryption, digital signatures, and public key infrastructure (PKI) authentication systems. A traceable enterprise that successfully performs identity verification can obtain a key pair issued by the key

management center. The key pair includes a public key and a private key. When the user calls the smart contract, the verifier will verify the data on the chain. After confirming that the data is legal, the public key will be used to encrypt the data, and then, the traceability information will be recorded in the blockchain. When reading the information in the blockchain, the smart contract can be called to obtain encrypted data, and the corresponding traceability information can be obtained after decryption with the private key

All traceable companies need to perform identity authentication and key distribution operations before joining the system. The detailed steps of key and authentication are as follows:

(i) The traceability company submits the key pair and digital certificate application to the certification center. During the application process, it is necessary to provide traceable company certification materials, including information such as the company's social credit code and company name

(ii) The certification center reviews the applicant's certification materials. After passing the identity authentication, the certification center submits the applicant's information to the key management center and requests the distribution of key pairs. After receiving the request, the key management center generates a key pair, adds the flag field information, and saves it in the secure database. The marked field is the social credit code of the traceable enterprise, which is used to identify the identities of different applicants

(iii) The key management center sends the key pair to the applicant and at the same time returns the public key to the certification authority center to generate a digital certificate

(iv) The certification center generates a digital certificate based on the applicant's identity information and the public key and returns the digital certificate to the applicant. The certificate adopts the X.509 standard proposed by ITU-T. The key management center will return the key pair, and the certificate authority will return the digital certificate. The applicant can compare them to determine whether the key pair is correct. If there is an error, it needs to reapply

(2) Key and authentication mechanism: in the industrial equipment traceability system, certain transaction information has a certain degree of confidentiality and can only be accessed by specific users. However, blockchain technology has the characteristics of information transparency and information sharing. Any user of the unrestricted blockchain network can obtain the information in the block, which leads to the leakage of users' private information. In this regard, two propose the user rights management. Different types of users have different access rights, thereby ensuring the security of user information and preventing the leakage of user privacy information. We assign permissions based on the tasks and needs of users in traceability companies, regulatory agencies, and industrial equipment traceability systems

(i) Supervision department: it is necessary to supervise all traceability information of industrial equipment. Therefore, the supervision department has the highest access authority in the traceability system and can add, delete, modify, and query traceability information. Adding operations is to call smart contracts to record traceability information on the blockchain. The delete operation will not directly delete the traceability information from the blockchain, but will add the traceability information corresponding to the state of the industrial equipment in the "deleted" state to the block. There are seven industrial equipment statuses in the traceable system, which are transportation, delivery, distribution, sale, return to the factory, and deletion. Similarly, the modification operation also adds traceability information of the state of the industrial equipment to the block. The query operation can query the complete traceability information table of industrial equipment

(ii) Traceability companies: it provides traceability services general access rights and can add, modify, and query traceability information. Among these operations, the addition and modification operations are the same as the highest access authority. For query operations, general access rights can only query the production, circulation, distribution, and supervision of industrial equipment

(iii) User: it needs to query the traceability information of the purchased product and has the lowest access authority. Besides, it can only perform query operations that is the same as the traceability company

3.2. Block and Traceability Information Table. Equations should be provided in a text format, rather than as an image. Microsoft Word's equation tool is acceptable. Equations should be numbered consecutively, in round brackets, on the right-hand side of the page. They should be referred to as Equation (1), and so on in the main text.

The block and traceability information is the most basic data structure in the traceability scheme in the blockchain, which mainly includes three parts: block, transaction table, and traceability information table. Blockchain is a chained data structure composed of multiple blocks. A block is a carrier used to store transaction orders. Each transaction order is a piece of traceability information for industrial

equipment. The user calls the smart contract to obtain the traceability information table of the industrial equipment through the unique OID number of the industrial equipment. Due to different access rights, the traceability information tables obtained by the user are also different.

The block and traceability information table is shown in Figure 1, which mainly includes three parts: block, transaction ticket, and traceability information form. In the block part, the first is the block header, which is composed of the timestamp of the block generated by the hash value of the previous block and the hash value of the root of the Merkel tree. Common hash algorithms include SHA1, SHA2, and MD5. In this paper, we use the SHA-256 hash algorithm, whose reliability and security meet the requirements of the traceability system. The second is the block body, which includes a single transaction number and transaction order. In the transaction ticket part, the transaction table is mainly composed of the product identification code, digital abstract, transaction content, timestamp, public key, and digital signature. The part of the traceability information table mainly includes five links. The product identification code is the identification mark of industrial equipment.

3.3. Blockchain Traceability Scheme. In the device traceability system based on the blockchain, the traceability enterprise first needs to perform identity authentication and key distribution. After the operation is completed, each participant in the system is assigned a different authority. When the equipment circulates in the supply chain, traceability companies, and regulators will call smart contracts to record equipment traceability information.

In the traceability solution outside the blockchain, we apply the OID identification system to the equipment traceability system. We store equipment traceability information in the traceability enterprise identification management server through the OID identification information registration mechanism and obtain detailed equipment traceability based on the OID identification analysis mechanism information to achieve equipment traceability. The external information of capacity equipment traceability blockchain management solves the problem of blockchain data explosion.

The caption can also be used to explain any acronyms used in the figure, as well as providing information on scale bar sizes or other information that cannot be included in the figure itself. Plots that show error bars should include in the caption a description of how the error was calculated and the sample size (see Figure 2).

3.4. Experiment Results Analysis. This section mainly tests the actual operating efficiency of the improved blockchain and analyzes the test results to verify the practicability of the proposed traceability system. First, an improved blockchain on a virtual machine is deployed. After that, sending suggestions and querying requests are used to test the system latency and throughput to activate the blockchain network. In the paper, throughput refers to the number of requests processed per unit time, and its unit is Tx/s. The system delay and throughput are obtained after many tests.

Figure 2 shows the system throughput under the proposal requests and the query requests. The system throughput increases at the beginning and then decreases with the increase of the proposal request. When the proposal request reached 4000, the throughput reaches the highest value. This is because the number of requests exceeds the processing capacity of the node, which will cause thread blocking and reduce system performance. At the same time, we can see from Figure 2, regardless of the query request, the system throughput is basically stable at about 350. This is because during the query request process, the blockchain only performs read operations instead of the write operations, which does not occupy system resources. As a result, the system throughput is relatively stable.

4. Blockchain-Based Federated Learning

The application of FL in the industrial Internet of Things (IIoT) is introduced in this section. We propose a FL framework based on device recognition in IIoT, which takes into account communication efficiency and data privacy.

4.1. The Communication Efficiency of the IIoT. There are two communication modes for the IIoT, including wired communication and wireless communication [20]. Recently, wireless communication technology is widely used in various fields, especially in the IIoT because of its flexibility and scalability. In the IIoT communication system, devices in the network are usually in different environments. In addition, there are a large number of devices in the IIoT, which require the communication system to be flexible and expandable. Therefore, it is a natural trend to regard wireless communication technology as the main method of the IIoT. With FL technology, local devices need to iteratively upload gradients to the central server, which introduces enormous communication overhead. In this case, the top-k algorithm is proposed to reduce communication costs.

4.2. The AI Model Based on Device Identification. Traditional machine learning needs to collect data from multiple devices to a central server for training, which only considers the performance of the central server when an AI model is designed. In this paper, we adopt FL to alleviate the data privacy problems. FL is a distributed machine learning. To consider the performance of all devices in the IIoT, especially in the IIoT, there are a large number of smart devices distributed in the network. The storage and calculation performance of these devices are different (the storage performance of mobile phones is worse than that of notebooks). Therefore, it is essential to consider devices with poor performance to ensures that each device can work successfully. In this paper, the device identification is stored through blockchain technology.

The AI model is designed based on device identification. In this paper, the AI model changes the depth of the convolutional layer according to the performance of the device in the network. We can increase the depth of convolutional layers when devices have superior performance. It is undeniable that the performance of the AI model is better with the

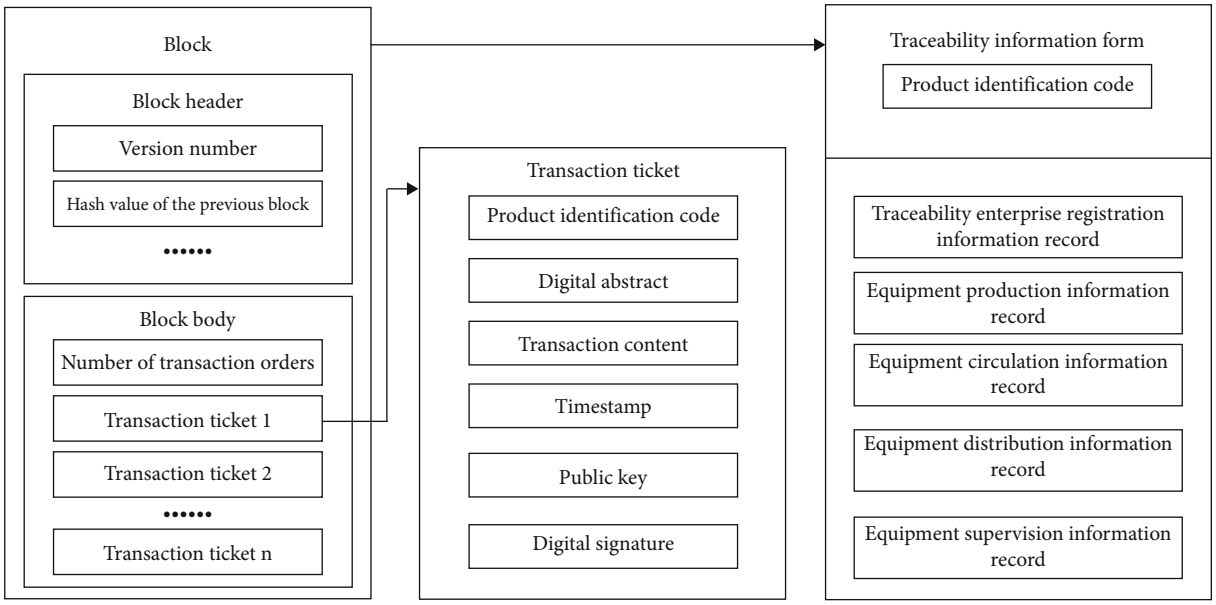

Figure 1: Block and traceability information table.

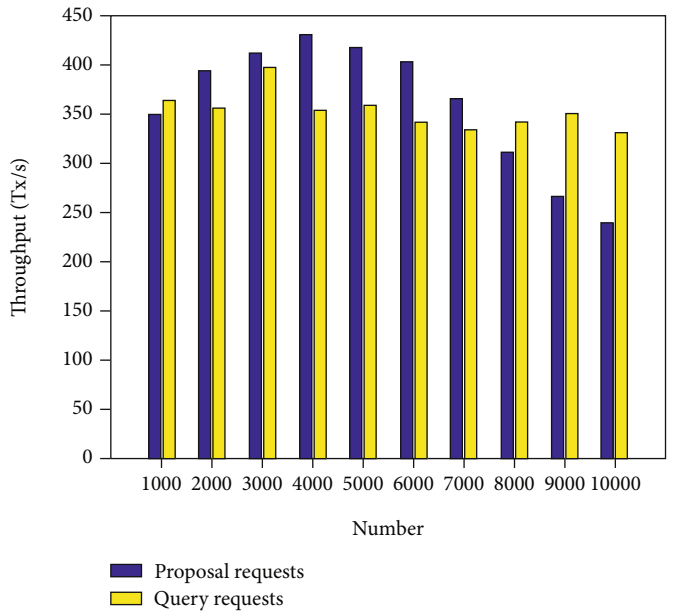

Figure 2: System throughput in case of proposal request and query requests.

increment of convolutional layers. This proposed framework designs the AI model based on the performance of the device, which makes a trade-off between the complexity of the network and the performance of the model.

4.3. Proposed Framework Based on FL. Traditional machine learning aggregates data from different devices to a central server, which may lead to data leakage and privacy infringement. Users are unwilling to share data, which leads to the problem of data isolated island. FL has been studied by many researchers in order to alleviate these problems. FL realizes data sharing in the IIoT and protects data privacy to a certain extent. However, the application of FL technology to the IIoT faces many challenges. For example, devices in wireless communication networks upload local gradients or models, which will bring a lot of communication overhead. It is essential to elaborate methods to reduce communication overhead.

Besides, the attacker can ratiocinate the private data from the gradient or model uploaded by local devices [21, 22]. To alleviate the problem of privacy protection in FL, many methods have been proposed, which are mainly divided into encryption technology [23] and differential privacy methods [24, 25]. Methods based on cryptography protect data privacy at the cost of decreasing communication efficiency. In this paper, we propose an algorithm based

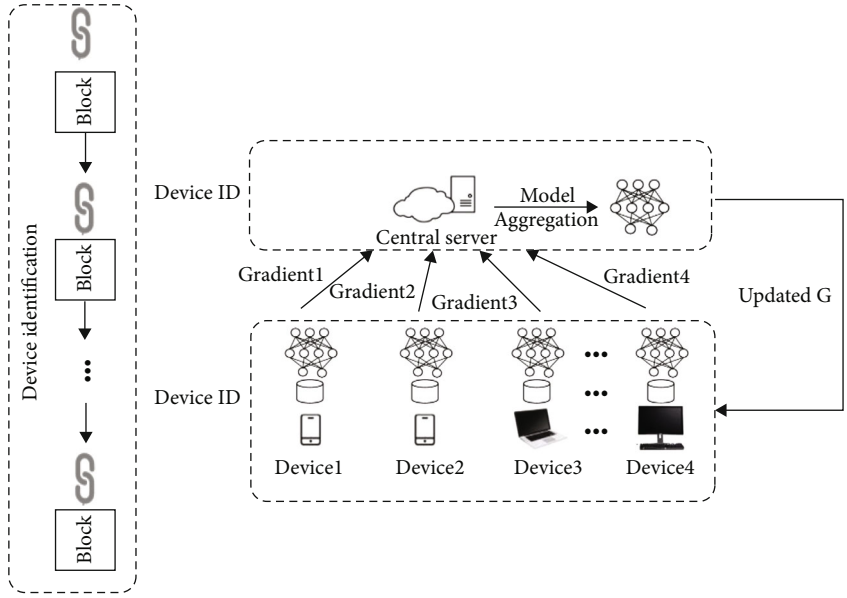

Figure 3: The proposed framework based on FL integrated with blockchain.

on differential privacy technology to alleviate the problem of data privacy.

The proposed framework based on FL integrated with blockchain is shown in Figure 3. It describes the system model applied to the industrial Internet scenario, where the underlying device with different performance is trained based on the local dataset. The training model uses the CNN model, and then, the local equipment uploads the model parameters to the central server for aggregation. After aggregation, the new model parameters are sent to each device. Among them, the blockchain can remove the trusted central authority. Therefore, it can solve the trust problem of the central server in the federated learning, preventing the single point of failure in the federated learning. In order to improve the communication efficiency, the proposed method employs a top-k algorithm when the gradient is uploaded to the central server. Besides, in terms of data privacy, we propose an improved differential privacy technology to realize data sharing. The detailed process of FL is exhibited as follows.

Step 1: local model train. The local device trains the AI model iteratively based on the local database to achieve a linear optimization problem locally. ω_i^l, γ represent the local model and learning rate of the device in the ith iteration, respectively, and $f(\bullet)$ is the loss function.

$$\omega_i^{\mathrm{l}} = \omega_i^{l-1} - \gamma \nabla f\left(\omega_i^l\right). \tag{1}$$

Step 2: local gradient preprocess. The local device trains the AI model based on the local database. To reduce the communication overhead, the proposed framework employs the top-k algorithm. Device D_i calculates the absolute value of the parameters in epoch t. Then, these values are sorted in positive order. We choose the first k values for global aggregation. This paper uses Gaussian mechanism to add noise to the local gradient, which protects the data privacy of users. The gradients in epoch t can be expressed as

$$g_t(x) \longleftarrow OL(\omega). \tag{2}$$

The process of gradient clipping is as

$$g_t(x)' = \frac{g_t(x)}{\max\left(1, \|g_t(x)\|_2 / S\right)}. \tag{3}$$

Step 3: local gradient calculation. The selected k gradients utilize differential privacy for noise disturbance, where S denotes the $L2$-norm threshold of gradient clipping. Note that S is the global sensitively set in advance. The procedure ensures that the $L2$-norm of the gradients of the local device is within the range of S. Then, Gaussian noise is added to the gradients which introduces randomness. The gradients adding Gaussian noise are as

$$g_t(x)'' = \frac{1}{b}\sum\left(g_t(x)' + N\left(0, \sigma^2\right)\right), \tag{4}$$

where b presents batch size and $N(0, \sigma^2)$ represents Gaussiandistribution. The variance σ must meet

$$\sigma = \frac{S}{\epsilon}\sqrt{2 * \ln\left(1.25/\delta\right)}, \tag{5}$$

where δ presents slack factor, s presents privacy budget, and s is set differently based on the data size of the local dataset. Besides, s is modified based on the loss of the model.

Step 4: local model update. Randomness is introduced due to the noise. The local model is updated by

$$\omega_{t+1} = \omega_t - \eta * g_t(x) \tag{6}$$

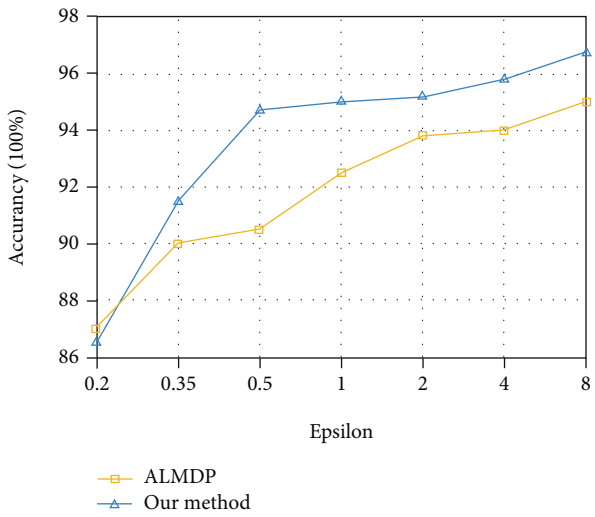

Figure 4: Accuracy of the proposed framework under different epsilons.

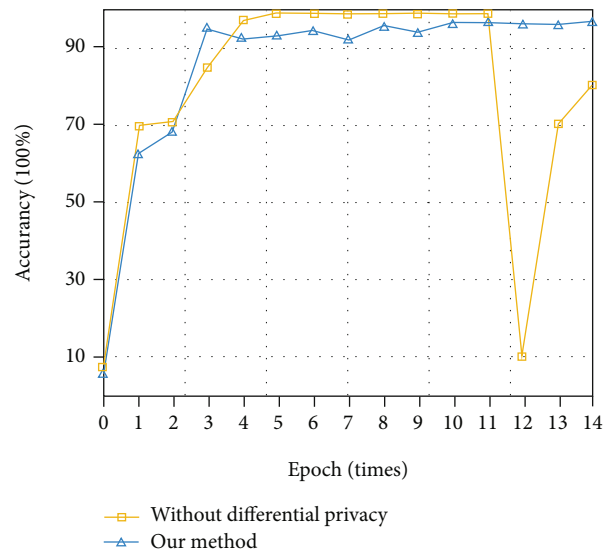

Figure 5: Performance comparisons on our method and FL without differential privacy.

Step 5: local gradient upload. Devices in the industrial Internet upload local gradients to the central server.

Step 6: model aggregate. The central server aggregates the gradients from devices in the IIoT. Then, the server sends the new model to all devices. The global model of epoch $t+1$ is shown as

$$\omega_{t+1} = \omega_t + \frac{1}{m}\left(\sum_{k=1}^{m} \Delta\omega_{t+1}^k\right), \Delta\omega_{t+1}^k = \omega_{t+1}^k - \omega_t^k. \quad (7)$$

4.4. Experiment Result Analysis. To evaluate the performance of the proposed method, a well-known digital classification dataset MNIST is employed. The AI model is designed based on the performance of the devices in the IIoT. Users can get the performance by the device identifier stored in the blockchain. There are 40 devices (15 mobile phones and 25 computers) in the wireless work to verify the proposed framework. We employ the CNN model as the learning model. Mobile phones are given in the network. The CNN model is designed with 2 convolutional layers and 2 fully connected layers. The kernel size is set as 3×3. The gradient clipping threshold S is set to 0.01. The AI model employs the maximum pooling and dropout to alleviate overfitting. Classification accuracy is employed as the evaluation criteria. To realize differential privacy protection, this paper adds Gaussian noise to the gradients in the process of gradient descent of backpropagation. As is shown in Figure 4, we compare the performance of our method with the ALMDP [24] algorithm. The result shows that our method is better. The accuracy of the model can reach 97.6%, which increases with the enlargement of s.

The performance of FL without differential privacy is compared with our proposed framework. As shown in Figure 5, the accuracy of FL without differential privacy is better than our method before epoch 12.

However, the accuracy of FL without differential privacy decreases rapidly in epoch 12 due to malicious modification of one device. Even if the performance of our method is worse, our method makes a trade-off between performance and data privacy.

5. Conclusions

The industrial Internet has received full attention from domestic and international researchers. As an important infrastructure of the industrial Internet, identification resolution technology is a link that must be overcome. The network interconnection of industrial control systems, the intercommunication of industrial data, and system security are issues that require great attention during the development of the industrial Internet. Blockchain technology has the characteristics of decentralization, immutability, and low cost, which can solve the pain points in the development process. Federated learning, as the basic theory of large-scale collaboration in the next generation of artificial intelligence, provides effective solutions to key issues such as small data and privacy in the current development of artificial intelligence, which can further promote the development of the industrial Internet. The numerical results justify that the federated learning and blockchain technology for industrial Internet identification proposed in this paper are practical and effective, which opens up a new research direction for the development of industrial Internet identification.

Data Availability

This is private company data. For confidentiality reasons, all data are not available to the public.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Ministry of Industry and Information Technology of China, 2019 Industrial Internet Innovation and Development Project Device ID Resolution Service Capability Test and Verification Platform Project (20200151).

References

[1] J. Li, F. R. Yu, G. Deng, C. Luo, Z. Ming, and Q. Yan, "Industrial internet: a survey on the enabling technologies, applications, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1504–1526, 2017.

[2] E. P. Yadav, E. A. Mittal, and H. Yadav, "IoT: challenges and issues in Indian perspective," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, pp. 1–5, Bhimtal, India, 2018.

[3] Y. Cheng and H. Shaoqin, "Research on blockchain technology in cryptographic exploration," in *2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, pp. 120–123, Bangkok, Thailand, 2020.

[4] J. Feng, Y. Wang, J. Wang, and F. Ren, "Blockchain-based data management and edge-assisted trusted cloaking area construction for location privacy protection in vehicular networks," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2087–2101, 2021.

[5] A. Kuzmin and E. Znak, "Blockchain-base structures for a secure and operate network of semi-autonomous unmanned aerial vehicles," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, pp. 32–37, Singapore, 2018.

[6] H. Yang, H. Cha, and Y. Song, "Secure identifier management based on blockchain technology in NDN environment," *IEEE Access*, vol. 7, pp. 6262–6268, 2019.

[7] A. Irshad, M. Usman, S. Ashraf Chaudhry, H. Naqvi, and M. Shafiq, "A provably secure and efficient authenticated key agreement scheme for energy internet-based vehicle-to-grid technology framework," *IEEE Transactions on Industry Applications*, vol. 56, no. 4, pp. 1–4435, 2020.

[8] A. Dua, N. Kumar, A. K. Das, and W. Susilo, "Secure message communication protocol among vehicles in smart city," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 5, pp. 4359–4373, 2018.

[9] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: motivation, opportunities, and challenges," *IEEE Communications Magazine*, vol. 58, no. 6, pp. 46–51, 2020.

[10] Y. Liu, N. Kumar, Z. Xiong, W. Y. B. Lim, J. Kang, and D. Niyato, "Communication-efficient federated learning for anomaly detection in industrial internet of things," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, Taipei, Taiwan, 2020.

[11] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A blockchained federated learning framework for cognitive computing in industry 4.0 networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2964–2973, 2021.

[12] J. Passerat-Palmbach, T. Farnan, M. McCoy et al., "Blockchain-orchestrated machine learning for privacy preserving federated learning in electronic health data," in *2020 IEEE International Conference on Blockchain (Blockchain)*, pp. 550–555, Rhodes, Greece, 2020.

[13] F. Li, A. Yang, H. Chen et al., "Towards industrial internet of things in steel manufacturing: a multiple-factor-based detection system of longitudinal surface cracks," in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 4627–4635, Atlanta, GA, USA, 2020.

[14] D. Gräf, M. Friedlein, C. Gänßmantel, J. Franke, and N. Ischdonat, "New concept for the integration of additive manufactured mechanical and mechatronic components in aircraft interior systems," in *2020 Advances in Science and Engineering Technology International Conferences (ASET)*, pp. 1–5, Dubai, United Arab Emirates, 2020.

[15] C. Chi, D. Han, Q. Zhang et al., "Research on distributed new energy spot trading method based on blockchain Technology," in *2020 Chinese Automation Congress (CAC)*, pp. 275–278, Shanghai, China, 2020.

[16] Z. Bao, Q. Wang, W. Shi, L. Wang, H. Lei, and B. Chen, "When blockchain meets SGX: an overview, challenges, and open issues," *IEEE Access*, vol. 8, pp. 170404–170420, 2020.

[17] R. Wakayama, R. Murata, A. Kimura, T. Yamashita, Y. Yamauchi, and H. Fujiyoshi, "Distributed forests for MapReduce-based machine learning," in *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pp. 276–280, Kuala Lumpur, Malaysia, 2015.

[18] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: a survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140699–140725, 2020.

[19] W. Zhang, X. Wang, P. Zhou, W. Wu, and X. Zhang, "Client selection for federated learning with non-IID data in mobile edge computing," *IEEE Access*, vol. 9, pp. 24462–24474, 2021.

[20] Y. Li, Y. Ma, Z. Yin, A. Gu, and F. Xu, "A communication model to enhance industrial wireless networks based on time-sensitive networks," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, pp. 363–367, Chengdu, China, 2020.

[21] H. Yang and Y. Zhou, "AIC-GAN: an auxiliary information classification GAN for learning deep models," in *2020 Chinese Automation Congress (CAC)*, pp. 6106–6111, Shanghai, China, 2020.

[22] L. Melis, C. Song, E. De Cristofaro, and V. Shmatikov, "Exploiting unintended feature leakage in collaborative learning," in *2019 IEEE Symposium on Security and Privacy (SP)*, pp. 691–706, San Francisco, CA, USA, 2019.

[23] L. T. Phong, Y. Aono, T. Hayashi, L. Wang, and S. Moriai, "Privacy-preserving deep learning via additively homomorphic encryption," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 5, pp. 1333–1345, 2018.

[24] N. Phan, X. Wu, H. Hu, and D. Dou, "Adaptive Laplace mechanism: differential privacy preservation in deep learning," in *2017 IEEE International Conference on Data Mining (ICDM)*, pp. 385–394, New Orleans, LA, USA, 2017.

[25] P. C. Mahawaga Arachchige, P. Bertok, I. Khalil, D. Liu, S. Camtepe, and M. Atiquzzaman, "Local differential privacy for deep learning," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 5827–5842, 2020.