Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/40/20230639

Research advanced in FL systems based on blockchain

Bangxiang Wang[1,4], Jingjie Ma[2], Yingruige Wang[3]

[1]School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing, Jiangsu province, 210000, China
[2]Portland Institute, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu province, 210000, China
[3]College of Communication and Information Technology, Xi'an University of Posts and Telecommunications, Xi'an, Shanxi province, 710000, China

[4]xianggozi@njust.edu.cn

Abstract. Federated learning (FL), as an excellent distributed machine learning paradigm, has gradually entered the public eye. FL's purpose is to solve the difficulty of centralized management of distributed data in real life and the prominent issues of data privacy, as well as data security. In FL, the central server distributes training tasks to clients to facilitate the training process. The client trains data locally and only uploads the updated model of local training to the central server, which is able to protect local data security effectively. In spite of that, FL still have disadvantages over high single-point failure, as well as lacking incentive mechanism. Thus, due to excellent decentralized nature of blockchain, a positive and feasible solution appears because of the technology combined with blockchain. In this article, we will introduce FL systems based on blockchain (BFL) and the current status of this field in detail. Specifically, we will first discuss the system coupled structure of BFL. Then, we will provide more details on challenges in BFL and corresponding solutions and explain main applications of BFL in various industries. Finally, we will discuss the difficult problems faced by BFL and give promising research directions in the future.

Keywords: Federal Learning, Blockchain, Coupling Structure, Incentive Mechanism, Privacy Protection.

1. Introduction

Nowadays, the complexity and data volume of computer-based problems are increasing, and Machine learning has proposed a feasible solution for solving such problems. In classical machine learning, models' learning efficiency and accuracy relies on the computing power of centralized servers, as well as high-quality training data. Nevertheless, during training this model, there are situations where the data owner is unwilling to share complete and accurate data because the data contains privacy-sensitive information. For researchers, obtaining sufficient amounts of high-quality data to train large models or high-precision models is difficult. Data owners are isolated and disconnected to form islands. To cope with the problems associated with privacy breaches and isolated data storage, the implementation of the federal learning law has emerged as a prominent solution and garnered increasing research interest in recent times.

DOI: 10.54254/2755-2721/40/20230639

FL combines the concept of decentralization to bridge different data sources, enabling training of statistical models on isolated data centers or remote devices. This method establishes a cohesive data ecosystem, uncovering the full potential and significantly magnifying the value of extensive data resources. FL does not require data sharing. The central server only accepts the parameters such as gradients transmitted upstream from the client to make the model updated. And after the central server completes the update of the global model, it sends the updated global model to the client to start the next round of training. Each of the client maintains data localization, and the feature of using locally calculated model updates to do training on a shared global model, which reduces the exposure risk of customers' private data effectively. Since Google first introduced the FL algorithm, it has significantly improved the performance of various applications. It has applications in the construction and analysis of medical information databases, privacy-preserving collaborative research, computer vision, education and industry.

Although this method effectively solves some existing privacy issues, its one-to-many communication method also creates new problems and challenges. First, there is a failure of single-point and malicious data. Given that the centralized federated learning framework requires a central node responsible for aggregating and incorporating the local training parameters provided by participants and subsequently updating the global model, any malicious attack targeting the central node has the potential to render the entire framework inoperable. Secondly, there may be malicious nodes in the child nodes to initiate poisoning attacks. Since the original data cannot leave the local client, poisoning attack detection based on the original data also fails, and the security of each client and the entire system is threatened. At the same time, the entire system lacks an incentive mechanism. Clients in traditional FL are seen as contributing their data and resources on computing free of charge. This assumption is unreasonable in actual scenarios. How to motivate each node remains to be explored. Communication cost, incentive mechanism, and privacy protection cannot be ignored during the training process.

People from all walks of life are trying to jointly optimize the FL framework and different algorithms. Combining blockchain technology may become an excellent direction. The consensus mechanism plays a crucial role in facilitating the exchange of parameters and validating the update function of the learning model. In the absence of any coordination or training data centrally managed, authentic and traceable transactions can be performed by users in the absence of a centralized intermediary. Replacing the function of a single central server with a chain structure and combining the elimination mechanism can also effectively make improvement on the security and privacy of the overall framework, as well as solving the centralization problem in FL. Combining blockchain with FL is gradually attracting the research interest of scholars. Partially blockchain-based decentralized and stable platforms have been established to power FL systems, and this form of ensemble algorithms has enormous potential waiting to be tapped.

The blockchain-based FL algorithm has shown broad application prospects. Aiming at this task, this paper aims to provide a systematic research progress report by introducing the FL systems rooted in blockchain technology (BFL) and the current status of this field in detail. Specifically, we will first discuss the system coupled structure of BFL. After that, we will provide more details on the problems in BFL and corresponding solutions, and explain main applications of BFL in various industries. Finally, we will discuss the difficult problems faced by BFL and gain a promising future in research directions.

2. Research status

The research objects of this article are papers published in journals, conferences and preprints from 2019 to 2023. The research scope includes the latest research results rooted in BFL. When collecting these papers, this research follows the following principles: First, use keywords to search in the existing search engines to find relevant literature; secondly, according to the references in the selected papers, further search for literature that meets the needs of this research. Papers; finally, use literature expansion tools such as connected papers to find literature related to the desired topic to ensure comprehensive and relevant research is included.

DOI: 10.54254/2755-2721/40/20230639

In this article, we uniformly call the combination structure of BFL to facilitate subsequent explanation. In recent years, people's attention to BFL has been on the rise. It can be seen from the search frequency of Zhihu keywords that from 2019 to 2022, the number of relevant documents increased from less than 20 articles per year to more than 100 articles and shows a continuous upward trend. The BFL architecture is in a state of continuous development. These BFL research works usually adopt different BFL structure designs, which are aimed at important problems in different FL. For instance: failure of single-point, poisoning attack and a lack of incentive mechanism. And for different application fields (such as: Internet of Things, Healthcare, Business and Finance, etc.). Researchers work hard to design and improve the BFL structure to improve the system in security, robustness and scalability. In addition, for specific application fields, like IOT and health caring, researchers are also working on combining blockchain with FL aiming at solving specific issues like privacy and security of data.

Literature [1] is concerned about the structural design of BFL and the deployed blockchain platform and discusses the effective rewards and the transparent contribution recognition for customers in BFL, and summarizes the feasible application of BFL. Literature [2] discusses the coupling type of BFL, the type of blockchain and the learning equipment used. Literature [3] summarize the application of BFL on the Internet of Things and edge devices. Although the challenges and applications of BFL have been relatively comprehensively summarized in recent years, a lack of strict and detailed classification of the framework and review that combines the BFL framework with the system still exist currently. For example, under the classic BFL framework, there is still a lack of in-depth discussion of the coupling methods of BFL. In this paper, Section 3 will introduce the four system coupling structures of BFL in sequence firstly. Section 4 will discuss in detail the current challenges BFL faces and existing solutions. Section 5 will detail the practical applications of BFL.

3. General System Architecture

Blockchain technology has many excellent properties such as decentralization, difficulty in tampering, and traceability. Its application has produced very fruitful results in many other fields. However, traditional FL structures are faced with challenges such as high single point failure rates, no incentives, susceptibility to poisoning attacks, and insufficient data privacy. To address these systemic flaws, blockchain technology has emerged as an effective solution. Through combining with blockchain, FL has been wildly used in numerous advanced fields, like Intelligent transportation [3-5], intelligent medical care [6], [1-2] and other fields already have relatively mature frameworks and applications. In different BFL system structures, four different types of system models are essentially adopted, and we divide them into four types: fully coupled type, flexible coupling type, overlapped coupled type, and loose coupling type. The following is a preliminary description of the definitions of these four structures:

(1) Fully coupled. the blockchain directly replaces the server in FL, and the nodes in FL work as both FL clients and nodes in the blockchain.

(2) Flexibly coupled type. The client in FL is still responsible for local data training and updated data uploading to the miner node. The miner node is in the blockchain system. FL and the blockchain have no shared nodes, and the global model is aggregated. Happens in the blockchain.

(3) Overlapped coupled type. Flexible coupling is similar. But in overlapped coupled type, Certain nodes work in both blockchain and FL simultaneously, and the roles of these nodes can be dynamically adjusted.

(4) Loosely coupled type. The blockchain does not directly participate during learning traditional FL, and may play the role of transmitting and storing data [7] and performing reputation calculation incentive rewards [8], etc.

3.1. Fully coupled model

The node of the FL client is the node of the blockchain in the fully coupled model [1]. At this time, the node plays two roles: (1) Own and train the local data model and broadcast it to other nodes. (2) Maintain the blockchain and reach a new consensus, verify the local model and generate new blocks to help the

DOI: 10.54254/2755-2721/40/20230639

global model update quickly. The diagram of fully coupled model is shown in Figure 1 below. It can be observed that the model is decentralized. The original centralized server is replaced by the blockchain. Each node in the group may participate in the training of local models and through methods such as smart contracts aggregate the global model.

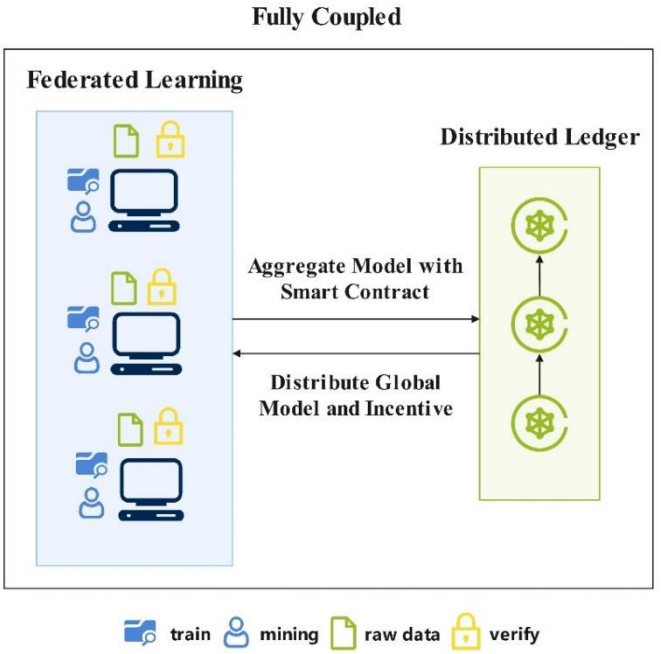

Figure 1. The diagram of fully coupled model.

The general training process of a fully coupled model is: (1) The local client node performs model training based on the data it holds. (2) Each node encrypts and broadcasts its latest training model to other nodes. (3) Other nodes receive the sent model and perform verification including digital signature and model itself. (4) Model aggregation can be performed through smart contracts or by selected nodes, which provide blockchain with aggregated model and update the global model. (5) The smart contract analyzes the results of the last round of updates and sends data rewards or block rewards to the corresponding nodes through the incentive mechanism.

In basic typical frameworks, single point of failure in FL has always been a difficult problem to solve, the classical method aims to replace the central aggregator by using blockchain directly. BAFFLE is a FL model without a central aggregator and driven by blockchain [9]. In the calculation process, effective model division and serialization analysis are adopted to collect local update models and realize independent or parallel global model updates. In the end, the performance of the results trained by BAFFLE is almost the same as that of the traditional FL model and consumes less computing resources. The DAG-FL framework, introduced in Literature [10], utilizes a directed acyclic graph (DAG) as its underlying structure. And this framework is composed of a three-layer asynchronous architecture. Combined with DAG Ledger technology, this model can better overcome system asynchronous problems and detect abnormal nodes. The literature assumes that all nodes are legal nodes and introduces an incentive mechanism FL model based on repeated competition. Under this model, the more skilled legal nodes will contribute more energy to the model training task, which solves the existing problem of malicious nodes.

DOI: 10.54254/2755-2721/40/20230639

3.2. Flexible coupling model

For the BFL framework, we separate the nodes that exercise blockchain functionality from the FL system. Researchers propose a flexible coupling model. In this model, the nodes within the FL system have the sole responsibility of gathering and training data specific to their local environment, as well as packaging the trained model for submission. Within the blockchain system, miner nodes are responsible for updating the global model directly on the blockchain. As shown in Figure 2, BFL's flexible coupling model is as follows: (1) Client nodes within the FL system collect local data and train the model. (2) The client node transfers the locally encrypted data model to the miner node within the blockchain system. (3) The miner node verifies the received trained model. (3) The smart contract aggregation model in the blockchain is written into the blockchain and sent to the client node. (4) The smart contract distributes data rewards or block rewards to client nodes based on the results of the previous round of training.

Literature [11] provides a relatively reliable and standard flexible coupling architecture BLADE-FL. The client node trains data and models locally. After uploading to the blockchain miner node, the miner passes the designed intelligent. The contract aggregation model can realize the self-motivation of the system and effectively solve the problem of node failure in the system. Literature [12] adopts a deep reinforcement learning method, which can minimize system delay, energy consumption and incentive cost while achieving the model target accuracy and provides a feasible way for model owners to find the global optimal solution. solution.

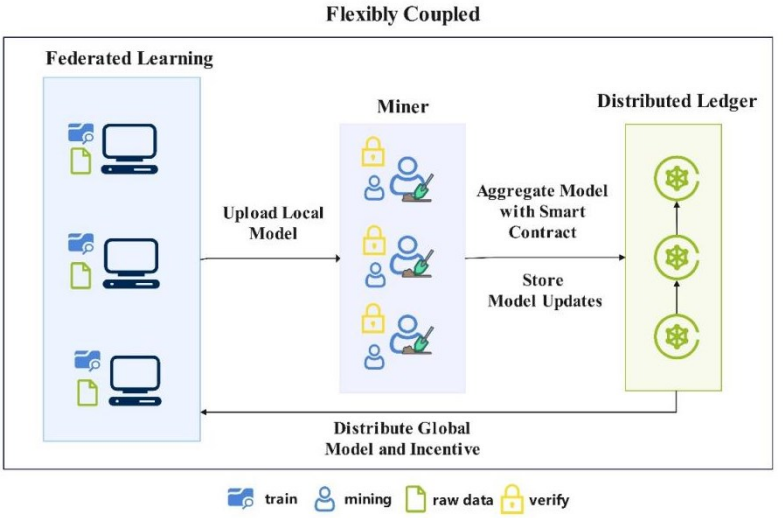

Figure 2. The diagram of flexible coupled model.

3.3. Overlapped coupled model

The client nodes in the overlapped coupled model are neither closely integrated with the blockchain nodes like the fully coupled model, nor are they completely separated from the blockchain nodes like the flexible coupling. In the overlapped coupled model, there are some composite nodes that simultaneously act as Client node and blockchain node, that is, nodes with three roles in the model: client node, blockchain node and composite node. Figure 3 presents the structural graph of the presented model, which combines the strengths of both the fully coupled model and the flexible coupling model. However, due to the intricate network structure, configuring this model can be challenging.

The general process of training the overlapped coupled model is shown below: (1) Client nodes and composite nodes collect local data and train models. (2) Blockchain node receives the uploaded encrypted model from the client node and is then broadcasted to the entire blockchain network by the composite node. (4) The smart contract aggregation model is written to the blockchain and distributed

DOI: 10.54254/2755-2721/40/20230639

to client nodes and composite nodes. (5) Data rewards or block rewards are distributed to the corresponding client nodes and composite nodes based on the last round of training results through smart contrast.

In the context of the IIOT, the mentioned literature [13] considers the challenges arising from communication channels which is unreliable and lacking in trust among users. The article also provides a blockchain-supported FL architecture which runs in the Digital Twin Unlimited Network (DTWN), significantly improving the reliability, security and privacy of the entire system. Literature [14] proposes an overlapped coupled model for heavy-haul railways, which can implement distributed learning in the absence of a reliable central server. By using the blockchain smart contracts, it manages the entire FL and improves intelligent control over powerful-haul railway systems efficiently and accurately.

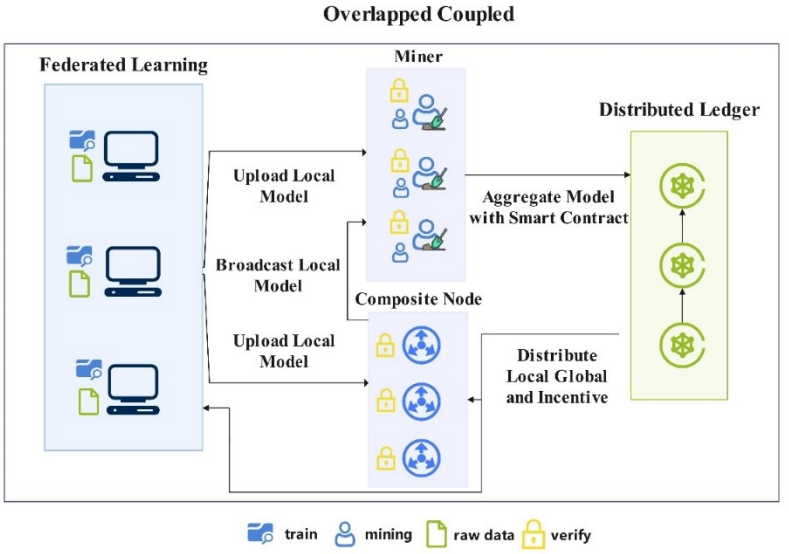

Figure 3. The diagram of overlapped coupled model.

3.4. Loosely coupled model

The blockchain in this model generally assists FL models to implement other functions instead of participating in the FL training process directly. For example, blockchain is introduced for reputation calculation. Literature [15] proposes a fine-grained FL framework, in which the FL training process occurs on cloud nodes and fog nodes. This method uses blockchain to calculate and store the reputation of each client, which makes reputation data an important criterion for measuring client reliability and trustworthiness. In addition, there is also literature [7] that uses blockchain to achieve data sharing. Literature [16] uses Merkle Tree to store client data in the blockchain in a trustworthy manner and is used in the IIoT framework. In the future, disputes can be better handled to ensure that data can be checked, difficult to tamper with, and traceable.

The following uses reputation calculation to illustrate the general training process of loosely coupled models: (1) The client node collects local data and trains the model. (2) The local model is uploaded by the client node to the miner node, resulting in an update of the local model. (3) The blockchain node verifies the uploaded model and generates the corresponding reputation opinion and writes it to the blockchain. (4) A central aggregator collects validated model data and model aggregation. (5) Issue corresponding rewards based on reputation information.

DOI: 10.54254/2755-2721/40/20230639

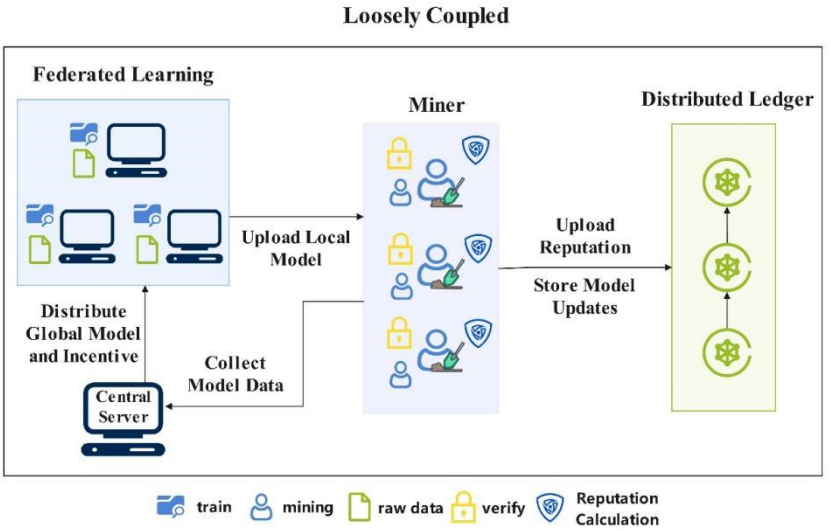

Figure 4. The diagram of loosely coupled model.

After the above summary, we can draw the pros and cons of the four coupling methods, as shown in the Table 1 below.

Table 1. Pros and cons of the four coupling methods.

	pros	cons
Fully coupled	Decentralization, effectively avoid Single-point-failure 2. Decentralization, avoid privacy leakage	The client is responsible for both the blockchain node and the FL node, which requires high computing resources
Flexibly coupled	1. The original data is local and will not be subject to malicious attacks from the blockchain 2. The blockchain can provide FL with more efficient data sharing	BFL are in different systems, improving difficulty in coordinating and managing
Overlapped coupled	Balance full coupling and flexible coupling, optimize the node's position	1. complex configuration network 2. complex incentive mechanism and formula design
Loosely coupled	(Take reputation as an example) It can enhance FL training model's quality and prevent malicious attacks	Loose structure (not decentralized), Single-point-failure and risk of privacy leakage exist

4. Challenges and existing countermeasures in BFL

4.1. Incentive mechanism design

The traditional FL model assumes that each client is reliable, uses its own data selflessly, and trains honestly according to the protocol, but this assumption is obviously not consistent with the actual situation. For companies, businesses, organizations and individuals, they tend not to share their information, such as medical information, financial information and other data of great value. Therefore, in the BFL model, we hope to use incentive mechanisms to foster selfless participation from all groups during the learning process. The incentive mechanism in BFL will allow clients with more excellent data and computing power to receive more rewards to encourage all parties to come up with better data

DOI: 10.54254/2755-2721/40/20230639

for reliable training. Furthermore, the use of incentive mechanisms can also punish malicious clients to a certain extent and detect and filter out malicious clients.

A growing number of studies have taken incentive mechanisms into consideration. The incentive mechanism designed in [17] includes two parts: data rewards and mining rewards. When miners complete model aggregation and generate blocks, they will receive mining rewards from the blockchain. This reward is the same as the customers owned by the miners themselves, similar to the customers possessed by the miners themselves. Then the miners allocate it to each client according to the number of sample data of each client. Literature [18] proposes an incentive mechanism that combines the reputation mechanism and Multi-KRUM [19] to address the problem that home appliance manufacturers have difficulty in obtaining feedback from users. Customers use mobile phones and mobile edge computing ([19]EMC) server training model, and the relevant data information of the home is gathered using a mobile phone. After the client transfers the model to the blockchain, the verifier uses the Multi-KRUM algorithm to calculate the reputation. If the submitted model update passes the verification, the corresponding client reputation will increase; if it fails the verification, the corresponding reputation will be Reduce, and finally reward according to different reputations. Literature [20] uses a competition mechanism to help global model updated. In every training round, all workers will choose the best model presented by workers from the round previously and add their own model to start training and update, and in this round The worker's income is determined by the worker's choice vote in the next round, so each worker will choose a better model to train together in order to train a better model to ensure that they can get more in the next round. Choose to vote, so as to get more benefits. Similarly, the income of the next round of miners is determined by the vote of the next round of miners, thus forming a logical closed loop, and each worker will train a better model for higher income.

4.2. Single-point-failure and poisoning attacks

In traditional FL, a Single-point-failure generally means that the central aggregator is easily affected by malicious nodes when aggregating the global model, causing errors in global model update. Central aggregators are also more vulnerable to attacks than regular clients. Hence, we typically employ blockchain to substitute the central aggregator, effectively mitigating the issue of single-point failure during the learning process. Full coupling model, flexible coupling, and overlapped coupled model are all without The network structure of the center can better resist the single-point failure problem. For the papers in this aspect, please refer to the third part of this article. However, the loose coupling model cannot solve the Single-point-failure problem of traditional FL because the blockchain performs tasks such as reputation calculation indirectly. Literature [17] Considering in the edge computing scenario, the proposed FL-Block model combining BFL realizes the non-centered privacy protection and anti-poisoning attack, and refrain from the single-point failure problem.

Poisoning attacks generally refer to behaviors that reduce the skills of the global model. Traditional FL assumes that each node is legal and actively participates in training. But in fact, the client is likely to have problems and become a source of poison that destroys the global model. Problems like destroying the local training data and maliciously modifying the client update model will damage the global model greatly [21] . If a malicious client submits an incorrect model update, it is likely to directly cause the FL model update to fail. Even if only a low-quality model is submitted, after aggregation by the central aggregator, the global model accuracy will be substantially reduced or even invalid. This kind of attack will bring great losses in traditional FL. But in the BFL model, we can avoid poisoning attacks through consensus algorithms and incentive mechanisms. In each training round, we can choose a client node as the master node to collect and update the model. And the client node can also broadcast other miner nodes with the messages, while verifying the miner model. With the blockchain saving the verified model update, the unverified model will be discarded. Moreover, in the current BFL model, many of them have designed an incentive mechanism or a reputation mechanism [18], which can well verify the results submitted by the client and calculate the reputation, and can effectively avoid poisoning attacks.

DOI: 10.54254/2755-2721/40/20230639

4.3. Data privacy and security

The main attack methods that undermine the privacy of FL data are inference attacks and inference reconstruction. We can generally use methods such as differential privacy to resist attacks. Privacy-preserving FL framework called BC-based PPFL was introduced in the literature [22]. This framework utilizes the characteristics of blockchain technology, such as immutability and decentralized trust, to facilitate model updates while also effectively identifying and excluding malicious clients. Consequently, it offers resistance against inference attacks. Literature [23] aims at the problem of leakage of private data in model training and verification due to the separation of data income rights and ownership in the blockchain and proposes a data transaction mechanism based on reverse game and a privacy protection verification mechanism. This mechanism implements privacy protection on test data and submit models to make verification on training model accurately, protecting data privacy effectively. The EFL chain proposed in [24] can also withstand inference attacks. In addition, there are also models proposed in [25] that cannot block inference attacks, which is related to the FL algorithm they designed.

This section discusses the challenges and existing countermeasures in BFL, and the following Table 2 summarizes the problems encountered in this section, the literature used, and the contributions made by the literature used.

Table 2. Problems Solved by Part BFL.

question	literature	contribution
excitation	[7][22]	Allocation according to the number of client sample data
	[21]	Crowdsourcing BFL
	[18]	Reputation-based Multi-KRUM
	[13]	competitive update model
Single-point-failure and poisoning attacks	[7]	FL-Blockmodel in edge computing scenarios
	[26]	Multi-KRUM public knowledge
	[27]	Verification Proof Consensus Mechanism
Data and Privacy Security	[22]	BC-Based PPFL frame
	[23]	Privacy Protection Verification
	[24]	EFLChain

5. Existing applications of BFL

At present, BFL have been widely used in all walks of life. For areas where BFL can be better applied, they often have the following characteristics: (1) Data is scattered and difficult to collect and concentrate. (2) Data has high value and privacy. (3) Problems in the field usually involve multiple parties, and multi-node systems and machine learning are often used to solve problems. The following will be divided into three parts: industrial Internet of Things, smart medical care, and smart transportation to illustrate the wide range of application fields of BFL.

5.1. Industrial Internet of Things

Training equipment is dispersed most of the time in the field of Industrial Internet of Things (IIoT), and the collected data is also complex and changeable. This requires the relevant BFL to have timeliness and high model generalization capabilities. And can effectively protect data security. For example, we can use the industrial data of a certain factory to infer the cost and output of the factory. In some industries or fields, these related data will be highly sensitive and confidential. Literature [28] proposed a distributed multi-party data sharing model, ensuring the security and confidentiality of data and facilitate retrieval with differential privacy technology. When the scale of the Internet of Things keeps expanding, ensuring high-quality services while managing limited resources has emerged as a significant challenge in the network environment. By integrating digital twins and edge networks, a novel approach called Digital Twin Edge Network (DITEN) is introduced. This method, in conjunction with BFL, offers enhanced stability of connections, bolstered communication security, and improved

DOI: 10.54254/2755-2721/40/20230639

data privacy protection within the DETEN network. Moreover, it enhances communication efficiency and effectively addresses resource constraints of IoT devices by leveraging the BFL framework. The utilization of Blockchain-Based Federated Learning (BFL) for data sharing, combined with privacy-preserving techniques, is extensively discussed in the referenced literature [3]. This approach effectively transforms obstacle of data sharing into a problem in machine learning. Additionally, it seamlessly integrates Federated Learning (FL) into the consensus mechanism of a permissioned blockchain, guaranteeing robust data security. The cited literature [5] delves into the realm of cognitive computing and addresses challenges such as poisoning attacks, performance issues, limited resources, and privacy breaches in the context of the Internet of Things (IoT). To tackle these challenges, the literature proposes a cognitive computing approach driven by big data, leveraging the combination of BFL technology. This decentralized paradigm not only ensures the privacy and security of IoT data but also accelerates the convergence speed through advanced verification and member selection mechanisms. Moreover, BFL demonstrates significant applicability in fault detection within the Industrial Internet of Things (IIoT) domain. Previous studies [4][29] addressed the issue of heterogeneous data in fault detection for IoT equipment. These studies proposed a federated averaging algorithm known as Centroid Distance Weighted Federated Averaging. This algorithm effectively solves the problem of data heterogeneity in fault detection for IoT devices.

5.2. Smart Healthcare

In medical scenarios, whether it is hospital data or patient medical records, they are highly sensitive and private data. The leakage of these data will bring great harm to society, so there are few hospitals or Individuals are willing to share this data. Blockchain-based FL has become an excellent solution to solve such problems, allowing hospitals or individuals to share data without revealing privacy. Literature [15] believes that the protection of patient privacy data limits related scientific research. Development, proposed a healthcare alliance based on BFL, emphasizing new secure aggregation protocols and privacy protection audit trails, but did not propose specific feasible solutions. Literature [30] proposed a specific solution for detecting COVID-19 based on BFL model, and proposed data normalization and capsule neural network technology to remedy difficulty of model training and data heterogeneity. The hospital trains on its real data and only shares weights and gradients to the blockchain. This article provides a data platform that can run among various dispersed hospitals, ensuring data privacy and security, and enabling automated detection of patients with COVID- 19. This year, due to IoT devices used widespread in various aspects of daily health management, the Internet of Healthy Things (IoHT) has also been proposed as a new concept. Aiming at protecting data privacy issues in IoHT , the literature [31] proposed a method through blockchain A lightweight hybrid FL model managed by smart contracts, which supports complete privacy and anonymity of IoHT data.

5.3. Intelligent Transportation

Intelligent transportation refers to the use of wireless communication, sensor equipment, machine learning and other technologies to monitor traffic conditions in real time and manage safety to form a safe and coordinated traffic network. FL is widely used in vehicle management in intelligent transportation. Call it Internet of Vehicles (IoV). Literature [26] proposes a blockchain-based autonomous FL design, which realizes local vehicle machine learning (oVML) model update through distributed exchange and verification, improves automated vehicle performance and high efficiency of vehicle communication network. In the heavy-duty railway system, due to the frequent changes of the train model, the traction or braking operation of personnel is not timely or in place, and dangerous situations are likely to occur. Therefore, it is an ideal solution to replace manual operations with intelligent systems. In literature [14], the BFL algorithm was suggested as a solution for enabling asynchronous collaborative machine learning among distributed agents who possess their own datasets, the solution is decentralized, using blockchain smart contracts to achieve comprehensive governance of the FL process. Considering the security and reliability of the model, literature [27] proposes an architecture of hybrid blockchain, consisting of a blockchain with permission and an acyclic graph

DOI: 10.54254/2755-2721/40/20230639

which is directed locally and achieved by asynchronous FL. In this scheme, deep reinforcement learning (DRL) serves as node selection, which realizes efficient data sharing in IoV.

5.4. BFL Future Research Direction

The development of traditional federal learning is hindered by its centralization, lacking in incentive mechanism and the difficulty of ensuring data security. The introduction of blockchain can just solve the above problems and form a BFL system with excellent nature. BFL is a good solution to some problems that traditional federal learning cannot solve, but BFL is not perfect.

It can be seen from our discussion that data security and model security are a very important topic in the field of BFL. Most blockchain systems themselves lack sufficient privacy protection methods. For BFL, we should introduce privacy protection technologies such as differential privacy, homomorphic encryption, and secure multi-party computation to protect data on the blockchain. Although there are much cutting-edge research on security in the respective fields of blockchain and FL, security is still a topic worthy of research for BFL.

In BFL, choosing the appropriate client is a research-worthy issue. The system hopes that the selected client has richer real data, a stable network, and better training equipment performance. At the same time, the legality of the client is also one of the factors that we must consider. Literature [32] proposed a client selection scheme FedCS. This scheme can actively select and manage clients based on their resource status, which is different from traditional Compared with FL, the efficiency of this scheme is significantly improved. However, it's worth noting that this solution does not address the critical aspects of ensuring the reliability and legality of the clients involved in the training process. This aspect presents an important area for further investigation and exploration.

FL shifts the focus from data sharing to model sharing. Yet, under existing consensus mechanisms, model sharing can incur substantial communication costs and computational overhead. Therefore, how to achieve efficiency in consensus mechanism to reduce communication costs and computing overhead has also become a challenging question.

We have been discussing BFL before, but we hope that the combination of FL and blockchain technology can bring progress to the blockchain technology itself. We can consider whether miners can use deep learning and reinforcement to Learning and other methods to choose the optimal mining strategy to enhance the performance of the blockchain system.

6. Conclusion

This paper studies BFL model and elaborates on the current research status of BFL through extensive investigation and reading of relevant scientific research papers in the field of BFL. A complete and comprehensive introduction to the existing BFL model is given from three perspectives: system architecture, challenges and existing solutions in BFL, and the current practical application of BFL. Finally, through our research, we discuss the current challenges faced by BFL and provide four promising directions for research. For current difficult problems in the respective fields of BFL, they can often be solved by combining the advantages of the two. BFL can effectively remedy privacy security problem in distributed systems and has huge advantages in training larger-scale high-precision models. This article will help researchers quickly understand the knowledge in this field and assist researchers in conducting more efficient scientific research work.

Authors Contribution

All the authors contributed equally.

Acknowledgments

This work was supported Undergraduate research training at Nanjing University of Science and Technology (202210288077).

DOI: 10.54254/2755-2721/40/20230639

References

[1] Yunlong Lu, Xiaohong Huang, Ke Zhang, Sabita Maharjan, and Yan Zhang. 2020. Blockchain empowered asynchronous FL for secure data sharing in internet of vehicles. IEEE Transactions on Vehicular Technology (TVT) 69, 4 (2020), 4298–4311.

[2] AR Faridi, A. Hafeez and F. Masood, "FL with Blockchain: A Study of the Latest Decentralized Couple," 2022 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), Greater Noida, India, 2022, pp. 91-96, doi: 10.1109/ICCCIS56430.2022.10037647.

[3] Yunlong Lu, Xiaohong Huang, Yueyue Dai, Sabita Maharjan, and Yan Zhang. 2019. BFL for privacy-preserved data sharing in industrial IoT. IEEE Transactions on Industrial Informatics (TII) 16, 6 (2019), 4177–4186.

[4] Weishan Zhang, Qinghua Lu, Qiuyu Yu, Zhaotong Li, Yue Liu, Sin Kit Lo, Shiping Chen, Xiwei Xu, and Liming Zhu. 2020. Blockchain-based FL for device failure detection in industrial IoT. IEEE Internet of Things Journal (IoT-J) 8, 7 (2020), 5926–5937.

[5] Youyang Qu, Shiva Raj Pokhrel, Sahil Garg, Longxiang Gao, and Yong Xiang. 2020. A blockchained FL framework for cognitive computing in industry 4.0 networks. IEEE Transactions on Industrial Informatics (TII) 17, 4 (2020), 2964–2973.

[6] Gaofeng Hua, Li Zhu, Jinsong Wu, Chunzi Shen, Linyan Zhou, and Qingqing Lin. 2020. Blockchain-based FL for intelligent control in heavy haul railway. IEEE Access 8 (2020), 176830–176839.

[7] LU YL, HUANG XH, DAI YY, et al. BFL for privacy - preserved data sharing in industrial IoT[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 4177-4186.

[8] UR REHAMN MH, SALAH K, DAMIANI E, et al. Towards blockchain - based reputation - aware FL [C] // Proceedings of the 2020 IEEE Conference on Computer Communications Workshops. Piscataway: IEEE, 2020: 183-188.

[9] Paritosh Ramanan and Kiyoshi Nakayama. BAFFLE: Blockchain Based Aggregator Free FL. 2019.

[10] Mingrui Cao, Long Zhang, and Bin Cao. Towards on-device FL: A direct acyclic graph-based blockchain approach. arXiv preprint arXiv:2104.13092, 2021.

[11] Chuan Ma, Jun Li, Ming Ding, Long Shi, Taotao Wang, Zhu Han, and H. Vincent Poor. When FL Meets Blockchain: A New Distributed Learning Paradigm. pages 1–8, 2020.

[12] Nguyen Quang Hieu, Tran The Anh, Nguyen Cong Luong, Dusit Niyato, Dong In Kim, and Erik Elmroth. Resource Management for Blockchainenabled FL: A Deep Reinforcement Learning Approach. 2020.

[13] Hyesung Kim, Jihong Park, Mehdi Bennis, and Seong-Lyun Kim. 2019. Blockchained on-device FL. IEEE Communications Letters 24, 6 (2019), 1279–1283.

[14] Liu, Ji, et al. "From distributed machine learning to FL: A survey." Knowledge and Information Systems 64.4 (2022): 885-917.

[15] Muhammad Habib Ur Rehman, Khaled Salah, Ernesto Damiani, and Davor Svetinovic. Towards blockchain-based reputation-aware FL. IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops, INFOCOM WKSHPS 2020, (February): 1 83–188, 2020.

[16] ZHANG WS, LU QH, YU QY, et al. Blockchain-based FL for failure detection in industrial IoT[J]. IEEE Internet of Things Journal, 2021, 8(7): 5926-5937.

[17] QU YY, GAO LX, LUAN TH, et al. Decentralized privacy using blockchain-enabled FL in fog computing[J]. IEEE Internet of Things Journal, 2020, 7(6): 5071-5083.

[18] ZHAO Y, ZHAO J, JIANG LS, et al. Privacy - preserving blockchain - based FL for IoT devices[J]. IEEE Internet of Things Journal, 2021, 8(3): 1817-1829.

[19] M. Shayan, C. Fung, CJ Yoon, and I. Beschastnikh, "Biscotti: A ledger for private and secure peer-to-peer machine learning," arXiv preprint arXiv:1811.09904, 2018.

DOI: 10.54254/2755-2721/40/20230639

[20] Hong Liu, Shuaipeng Zhang, Pengfei Zhang, Xinqiang Zhou, Xuebin Shao, Geguang Pu, and Yan Zhang. 2021. BFL for collaborative intrusion detection in vehicular edge computing. IEEE Transactions on Vehicular Technology (TVT) 70, 6 (2021), 6073–6084 .

[21] Jiawen Kang, Zehui Xiong, Dusit Niyato, Yuze Zou, Yang Zhang, and Mohsen Guizani. 2020. Reliable FL for mobile networks. IEEE Wireless Communications 27, 2 (2020), 72–80.

[22] Aggelos Kiayias, Alexander Russell, Bernardo David, and Roman Oliynykov. 2017. Ouroboros: A provably secure proof-of-stake blockchain protocol. In Annual International Cryptology Conference (CRYPTO). Springer, Berlin, Germany, 357–388.

[23] Xidi Qu, Shengling Wang, Qin Hu, and Xiuzhen Cheng. 2021. Proof consensus of FL: A novel energy-recycling algorithm. IEEE Transactions on Parallel and Distributed Systems (TPDS) 32, 8 (2021), 2074–2085.

[24] Umer Majeed and Choong Seon Hong. 2019. EFLChain: Ensemble learning via FL over blockchain network: A framework. In Proceedings of the KIISE Korea Computer Congress. 845–847.

[25] Jiawen Kang, Zehui Xiong, Dusit Niyato, Yuze Zou, Yang Zhang, and Mohsen Guizani. Reliable FL for Mobile Networks. IEEE Wireless Communications, 27(2):72–80, 2020.

[26] SHAYAN M, FUNG C, YOON CJM, et al. Biscotti: a blockchain system for private and secure FL[J]. IEEE Transactions on Parallel and Distributed Systems, 2021, 32 (7): 1513-1525.

[27] KANG JW, XIONG ZH, JIANG CX, et al. Scalable and communication **-** efficient decentralized federated edge learning with multi **-** blockchain framework [C]// Proceedings of the 2020 International Conference on Blockchain and Trustworthy Systems, CCIS 1267. Singapore: Springer, 2020: 152-165.

[28] Yunlong Lu, Xiaohong Huang, Ke Zhang, Sabita Maharjan, and Yan Zhang. Communication-efficient FL and Permissioned Blockchain for Digital Twin Edge Networks. IEEE Internet of Things Journal, 4662(c):1–1, 2020.

[29] China University of Petroleum (East China). A failure detection method for Internet of Things equipment based on BFL: 202011087722. 7[P]. 2021-01-05. (China University of Petroleum (East China). A failure detection method for Internet of things device based on BFL: 202011087722. 7[P]. 2021-01-05.)

[30] Rajesh Kumar, Abdullah Aman Khan, Sinmin Zhang, WenYong Wang, Yousif Abuidris, Waqas Amin, and Jay Kumar. Blockchain-FL and Deep Learning Models for COVID-19 detection using CT Imaging. 14(8):1–12, 2020.

[31] RAHMAN MA, HOSSAIN MS, ISLAM MS, et al. Secure and provenance enhanced Internet of health things framework: a blockchain managed FL approach [J]. IEEE Access, 2020, 8: 205071-205087.

[32] Ismael Martinez, Sreya Francis, and Abdelhakim Senhaji Hafid. Record and reward FL contributions with blockchain. Proceedings - 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2019, pages 50–57, 20 19.