

Trends in Blockchain and Federated Learning for Data Sharing in Distributed Platforms

Haemin Lee and Joongheon Kim

Department of Electrical and Computer Engineering, Korea University, Seoul, Republic of Korea

E-mails: haemin2@korea.ac.kr, joongheon@korea.ac.kr

Abstract—With the development of communication technologies in 5G networks and the Internet of things (IoT), a massive amount of generated data can improve machine learning (ML) inference through data sharing. However, security and privacy concerns are major obstacles in distributed and wireless networks. In addition, IoT has a limitation on system resources depending on the purpose of services. In addition, a blockchain technology enables secure transactions among participants through consensus algorithms and encryption without a centralized coordinator. In this paper, we first review the federated learning (FL) and blockchain mechanisms, and then, present a survey on the integration of blockchain and FL for data sharing in industrial, vehicle, and healthcare applications.

I. INTRODUCTION

Internet of things (IoT) technologies have been applied in various areas, and thus they generate a massive volume of data. Rapidly increasing IoT network size and accumulated data volume broaden the performance of artificial intelligence (AI) and create new values in various industry applications. The collected data should be processed with low latency while assuring security and reliability. Distributed multiple devices can work collaboratively to show better performance [1]. Therefore, effective deployment and intelligent utilization of IoT are needed. However, classical centralized learning algorithms in the IoT are vulnerable to security [2]. Centralized machine learning methods which involve data sharing from end devices to the centralized third-party server have privacy and data leakage issues. In addition, when data size is large, centralized machine learning algorithms have difficulties in terms of accuracy and efficiency. With these challenges, federated learning (FL) can be one of alternative solutions that enables on-device machine learning without directly transfer the private end-user data to a central server [1].

Despite the benefits, there are still remaining issues in FL computation. Since blockchain has been widely used to address security issues in distributed scenarios, recently, there are many approaches to enhance the privacy and security problems by replacing the federated learning server with blockchain [3], [4]. In this paper, we first present the existing blockchain and FL frameworks. Then, we present the privacy-preserving data sharing for a distributed industrial, vehicle, and healthcare IoT applications. Finally, we present several open research challenges with their possible solutions. The rest of the paper is organized as follows. Section II discusses the background of blockchain and federated learning technologies and how blockchain helps in existing FL techniques.

Section III describes blockchain applications for AI. Open research challenges and issues are discussed in Section IV. Section V concludes the paper.

II. RELATED WORK

A. Federated Learning (FL)

FL is a technique in which multiple local clients, i.e., IoT devices, smartphones, and one central server work together to learn a global model in a decentralized context. Each local client computes an update to the current global model maintained by the server, and only this update is communicated instead of the whole dataset. The FL is definitely useful in twofold: data privacy and communication efficiency. First, FL has less communication cost as it simply transfers its updates while transferring the local data to a central server put more burden on network traffic and storage costs. Second, the dataset of the local client is never uploaded to the server, which reduces the data leakage. The advantage is decoupling of model training from the need for direct access to the raw training data [5]. For applications where the training objective can be specified on the basis of data available on each client, FL can significantly reduce privacy and security risks by limiting the attack surface to only the device, rather than the device and the cloud [5]. However, there are still other issues associated with FL. Malicious actors may attempt to compromise the global model or access the client's dataset. In addition, the lack of incentive or motivation for the client to cooperate in the FL system. To this end, adopting and leveraging blockchain features for FL applications can be a solution.

B. Blockchain

Blockchain is a distributed, open-source, immutable, public digital ledger which is distributed among networked peers. Fundamentally, blockchain is a chain of blocks that make up the ledger with consensus algorithms and encryption. This ledger holds a permanent record of transactions and interactions that took place among participants accessing the distributed and decentralized blockchain network [6]. This network configuration that consists only of agreed blocks between users is robust to malicious attacks. Consensus protocols such as proof-of-work (PoW), proof-of-stake (PoS) is the method to reach a consensus in a distributed environment that can impact the performance. Blockchain can also be applied to financial services, smart contracts, IoT, and security services.

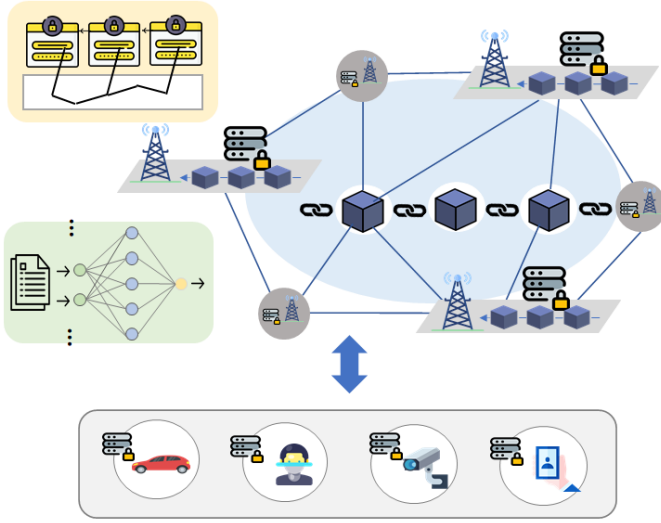


Fig. 1. Blockchain Federated learning application

Businesses that require high reliability and honesty can use blockchain to attract customers. Besides, it is distributed and can avoid the single point of failure situation [7]. With these characteristics, the combination of AI and blockchain technologies has paved the way for many stable systems that support the interaction of multiple devices while providing confidentiality, authentication, and integrity [6].

C. Framework

In the blockchain-based FL (BCFL) framework, the decentralized nature of blockchain can replace the central server [8]. The functions of the centralized server can be implemented by the smart contract (SC) instead, and be actuated by transactions on the blockchain. That is, the participating nodes perform FL via blockchain, which maintains the global models and local updates. The mechanism consists of blockchain storage, committee consensus mechanism, and model training part. The storage of BCFL is a blockchain system that only authorized devices can access the FL training contents. In the committee consensus mechanism, a few honest nodes constitute a committee that validates the updates and assigns a score on them. Only the qualified updates will be packed onto the blockchain. And at the beginning of every round, a new committee is formed. For the model training, the nodes other than committees perform local training each round.

III. DATA SHARING IN DISTRIBUTED PLATFORMS

In this section, we describe the integrated blockchain and FL for various applications as Fig. 1. The authors in [9] propose a mechanism for distributed multiple parties in IIoT applications, which incorporates federated learning into permissioned blockchain. In [10], a decentralized approach for the vehicle to relieve transmission load and address privacy concerns of providers is proposed. In [11], the combination of federated learning and blockchain technology adds progressive value to the healthcare sector.

A. Industrial Internet of Things (IIoT)

IIoT devices have some different requirements, unlike the consumer IoT that focuses more on user's comfort and convenience. IIoT is deployed in mission-critical industries and must carry out an advanced task, not simply feeding sensed data to control systems. IIoT-enabled industry requires additional security implementations throughout the operation process, and it must be highly reliable even in harsh conditions. Moreover, effective decision-making is required through sophisticated data analysis. As a sequence, the distributed IIoT can utilize collaborate data sharing to meet the quality of service. However, in the existing data sharing scheme with the distributed multiple parties has the risk of data leakage. Data leakage may take place during data storage, data transmission, and data sharing, which may lead to severe issues for both owners and providers.

Authors in [9], propose a collaborative architecture that incorporates federated learning into permissioned blockchain to share data to reduce the risk of data leakage. Permissioned blockchain establishes secure connections among all the end IoT devices through its encrypted records, which are maintained by the entities equipped with computing and storage resources. Different from the existing consensus such as PoW, they applied federated learning empowered consensus, i.e., proof-of-quality (PoQ) which makes better use of the node's computing resources.

B. Internet of Vehicles (IoV)

With the development of communication technologies in 5G networks and IoT, the automotive industry continues to accelerate towards connected cars [12], [13]. The Internet of Vehicles (IoV) is a distributed network that supports the use of data created by connected cars and vehicular networks. In IoV settings, a massive amount of diverse types of data is constantly generated by the moving vehicles, which includes additional data such as trajectories, traffic information, and multimedia data. Through the data sharing, each contributor can make better use of their data by utilizing them together to implement a collaborative task like traffic prediction. However, data sharing in IoV faces two critical challenges [10]. First, the vehicles need to share data efficiently despite unreliable inter-vehicle communications. Second, data providers, i.e., IoV are increasingly concerned about data security and privacy issues that can discourage them from providing the data available with them for analysis.

The presented research results in [10] integrate blockchain and FL, which can support reliable and secure data sharing among distributed users. The vehicular networks in [10] consist of vehicles, road side units (RSUs), macro base stations (MBSs). The MBS first distributes the global model to the blockchain. Then the participating vehicles download the global model from blockchain and train their local models. A certificate authority performs the identification of participants to access the permissioned blockchain. Node selection stage formulates and solves an optimization problem using the deep reinforcement learning (DRL) algorithm to select participating

vehicles [14], [15]. Then the selected vehicles perform local training on their data by computing the local gradient descent. Vehicles then send the parameters of the trained local model to the nearby RSU and upload it to the blockchain for further verification and aggregation. Then the MBS retrieves the parameters and executes global aggregation.

C. Internet of Health Things (IoHT)

With the advent of the Internet of Health Things (IoHT), such as wearable and implantable medical devices, distributedly produced medical data supports real-time analytics and medical tracking, remote monitoring. However, the Internet of Health Things (IoHT) collects very sensitive health and ambience data that requires privacy. In addition, there are some restrictions on the centralized data collection process.

The presented research results in [11] present a lightweight hybrid FL framework in blockchain which smart contracts manage the authentication of participating federated nodes, distribution of global or locally trained models. The IoHT interfaces with the edge nodes and the edge nodes have their local, private data for local training and inferencing. The edge nodes are also capable of acting as a local blockchain client or FL worker. The communication module of the edge nodes first performs differential privacy and then securely shares the encrypted model and training data to the blockchain App for further processing. The blockchain client processes the block creation and shares the smart contract of the hybrid blockchain node for global analysis.

IV. OPEN ISSUES

A. FL

In FL, various communication and networking technologies can be additionally considered in order to improve the performance. First, caching-related technologies can be considered for local training and aggregation [16]–[20]. In addition, traffic-aware [21] and video/contents-aware [22]–[24] application-specific learning should be additionally considered in order to improve the performance.

B. Blockchain

In this section, we present the challenges to guarantee data privacy by applying blockchain technique efficiently. First, blockchain itself has security issues such as 51% attack, forking attack and, double spending attack. Moreover, AI-specific consensus protocols can be designed considering proofs based on the quality of learning models or provenance of data [6]. Another challenge is the delay and errors of both communications and blockchain on FL. Therefore, some directions need to be taken to minimize the latency.

V. CONCLUSIONS

In this paper, we surveyed and reviewed the privacy-preserving data sharing related to the use and applicability of blockchain features for Federated learning. We gave a short overview of blockchain and federated learning and showed how blockchain technology can enhance and solve privacy

issues. Moreover, we presented an application of a blockchain federated learning framework in the industrial, vehicle network, and healthcare sectors. Our paper concludes by showing that adopting blockchain for federated learning can definitely leverage the performance of traditional federated learning, but there still exist many research challenges to be addressed.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation of Korea (NRF-Korea, 2021R1A4A1030775). J. Kim is a corresponding author of this paper.

REFERENCES

- [1] J. Park, S. Samarakoon, A. Elgabri, J. Kim, M. Bennis, S.-L. Kim, and M. Debbah, "Communication-efficient and distributed learning over wireless networks: Principles and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 796–819, May 2021.
- [2] N.-N. Dao, T. V. Phan, U. Sa'ad, J. Kim, T. Bauschert, D.-T. Do, and S. Cho, "Securing heterogeneous IoT with intelligent DDoS attack behavior learning," *IEEE Systems Journal*, pp. 1–10, 2021.
- [3] M. Saad, J. Choi, D. Nyang, J. Kim, and A. Mohaisen, "Toward characterizing blockchain-based cryptocurrencies for highly accurate predictions," *IEEE Systems Journal*, vol. 14, no. 1, pp. 321–332, March 2020.
- [4] E. Boo, J. Kim, and J. Ko, "LiteZKP: Lightening zero-knowledge proof-based blockchains for IoT and edge platforms," *IEEE Systems Journal*, pp. 1–12, 2021.
- [5] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [6] K. Salah, M. H. U. Rehman, N. Nizamuddin, and A. Al-Fuqaha, "Blockchain for ai: Review and open research challenges," *IEEE Access*, vol. 7, pp. 10 127–10 149, 2019.
- [7] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *2017 IEEE International Congress on Big Data (BigData Congress)*, 2017, pp. 557–564.
- [8] Y. Li, C. Chen, N. Liu, H. Huang, Z. Zheng, and Q. Yan, "A blockchain-based decentralized federated learning framework with committee consensus," *IEEE Network*, vol. 35, no. 1, pp. 234–241, 2020.
- [9] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Blockchain and federated learning for privacy-preserved data sharing in industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4177–4186, 2019.
- [10] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, and Y. Zhang, "Blockchain for secure and efficient data sharing in vehicular edge computing and networks," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4660–4670, 2018.
- [11] M. A. Rahman, M. S. Hossain, M. S. Islam, N. A. Alrajeh, and G. Muhammad, "Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach," *IEEE Access*, vol. 8, pp. 205 071–205 087, 2020.
- [12] M. Shin, J. Kim, and M. Levorato, "Auction-based charging scheduling with deep learning framework for multi-drone networks," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4235–4248, May 2019.
- [13] S. Jung, W. J. Yun, M. Shin, J. Kim, and J.-H. Kim, "Orchestrated scheduling and multi-agent deep reinforcement learning for cloud-assisted multi-UAV charging systems," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 6, pp. 5362–5377, June 2021.
- [14] S. Jung, J. Kim, M. Levorato, C. Cordeiro, and J.-H. Kim, "Infrastructure-assisted on-driving experience sharing for millimeter-wave connected vehicles," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2021.
- [15] M. Choi, A. No, M. Ji, and J. Kim, "Markov decision policies for dynamic video delivery in wireless caching networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 12, pp. 5705–5718, December 2019.

- [16] M. Choi, J. Kim, and J. Moon, "Wireless video caching and dynamic streaming under differentiated quality requirements," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 6, pp. 1245–1257, June 2018.
- [17] —, "Dynamic power allocation and user scheduling for power-efficient and delay-constrained multiple access networks," *IEEE Transactions on Wireless Communications*, vol. 18, no. 10, pp. 4846–4858, October 2019.
- [18] A. Malik, J. Kim, K. S. Kim, and W.-Y. Shin, "A personalized preference learning framework for caching in mobile networks," *IEEE Transactions on Mobile Computing*, vol. 20, no. 6, pp. 2124–2139, June 2021.
- [19] M. Choi, A. F. Molisch, and J. Kim, "Joint distributed link scheduling and power allocation for content delivery in wireless caching networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 12, pp. 7810–7824, December 2020.
- [20] M. Choi, A. F. Molisch, D.-J. Han, D. Kim, J. Kim, and J. Moon, "Probabilistic caching and dynamic delivery policies for categorized contents and consecutive user demands," *IEEE Transactions on Wireless Communications*, vol. 20, no. 4, pp. 2685–2699, April 2021.
- [21] S. Jung, J. Kim, and J.-H. Kim, "Intelligent active queue management for stabilized QoS guarantees in 5G mobile networks," *IEEE Systems Journal*, pp. 1–10, 2021.
- [22] J. Kim, G. Caire, and A. F. Molisch, "Quality-aware streaming and scheduling for device-to-device video delivery," *IEEE/ACM Transactions on Networking*, vol. 24, no. 4, pp. 2319–2331, August 2016.
- [23] J. Koo, J. Yi, J. Kim, M. A. Hoque, and S. Choi, "Seamless dynamic adaptive streaming in LTE/Wi-Fi integrated network under smartphone resource constraints," *IEEE Transactions on Mobile Computing*, vol. 18, no. 7, pp. 1647–1660, July 2019.
- [24] J. Yi, S. Kim, J. Kim, and S. Choi, "Supremo: Cloud-assisted low-latency super-resolution in mobile devices," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.