Federated Learning Using Blockchain-based Marketplace

Boyun Eom, Sunhwan Lim, Young-Ho Suh, Sungpil Woo, Chanwon Park
Autonomous IoT Research Section,
Electronics and Telecommunications Research Institute
Daejeon, South Korea
Email : {eby, shlim, yhsuh, woosungpil, cwp}@etri.re.kr

***Abstract*—** Federated Learning (FL) is an emerging machine learning paradigm aimed at collaboratively training AI models with distributed multiple clients in a privacy-preserving manner. However, the performance of FL can be poor, especially in cross-silos FL, without sufficient participants who train models with their data. Therefore, finding relevant datasets and engaging their owners in FL can be one of the key important factors in reality. Motivated by this, in this paper, we introduce a system equipped with efficient ways to search for data owners who possess suitable datasets and to inspire them for cooperation. In our approach, all participants who perform the learning process in FL can receive transparent and automated incentives from every gain of the final output, which can work as a motivation for participation in federated learning.

Keywords—federated learning; blockchain; incentive; AI-data commons

I. INTRODUCTION

In general, the quantity and quality of data used to train an AI model have a crucial impact on its performance in making predictions. Therefore, many organizations have started to construct their own systems to gather and analyze data. However, due to concerns such as privacy protection or the misuse of data, data owners hesitate to share their data, which can hinder the development of AI services. As a result, the need for systems such as data commons [1, 2] has been growing to obtain qualitative data while preserving privacy. Federated learning has also gained attention owing to its strength in data security. Indeed, it has more advantages than general machine learning methods, such as data diversity and hardware efficiency. However, to achieve better effectiveness in federated learning by promoting cooperation, a fair incentive mechanism needs to be provided. In this paper, we present a system that enables federated learning to proceed more effectively with participants who have published relevant data on the marketplace in AI-Data Commons. In our work, we have also designed the system to support virtualized infrastructure to fulfill the process of local training if the data publisher requests computational resources.

The remainder of this paper is organized as follows; In section 2, we briefly review related topics such as federated learning and blockchain. After explaining the design of the system in section 3, we discuss future works and conclude in section 4.

II. RELATED WORK

A. Federated Learning

Although it is relatively new setting of distributed learning, federated learning shows its growing attraction by expanding applied domains. In federated learning, training processes occur on several clients with local data and only the model weights are sent to a server as shown in Fig.1. The server, then aggregates all weights from clients, updates its model and sends new model to clients. These processes are iterated until the federated learning is finished.

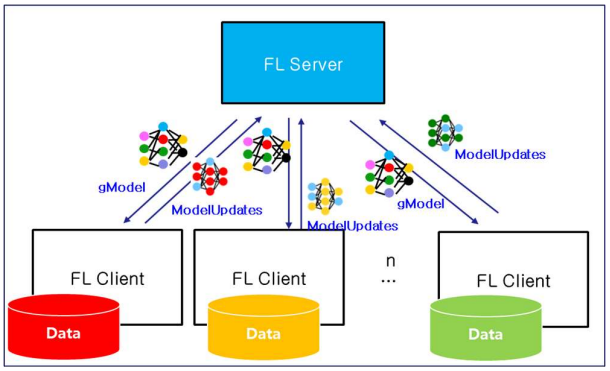

Fig. 1. Federated Learning

Depending on data partitioning over sample and feature space, federated learning can be categorized into three: horizontal, vertical and hybrid federated learning[3]. Also, based on the participating clients and the scale of federation, federated learning can be divided into two types: cross-device and cross-silo federated learning. In cross-device federated learning, typically large number of mobile devices participate and selection of the participants for every round of federated learning needs to be considered. Meanwhile, in cross-silo federated learning, organizations which have data on their purpose are usually participated as clients. The contribution of clients and incentives for that could be critical factor in the performance of federated learning[6-7].

Well-known core challenges of federated learning are communication efficiency throughout the federated network and heterogeneity of data caused in Non-IIDness [8].

There are approaches of decentralized, i.e. server-less, federated learning framework such as SimFL and swarm to overcome the stability and reliability of server-based federated learning.

B. Blockchain

Ethereum has overcome the limit of blockchain-based application which was considered that blockchain technology was only applicable for payment system [9]. In fact, Ethereum network acts as a world computer by sharing global resources among participating nodes. Every transactions on blockchain network are recorded on blocks and all participant nodes keeps synchronization of those blocks. There have been many applications designed and developed leveraging this technology; medical, public utilities, identity management, asset registration, smart home and etc. [10-12]. It is smart

contracts which bring up blockchain to these various business. Smart contracts on Ethereum are collections of codes and these are deployed on a blockchain. When the pre-defined conditions are met, that smart contract runs immediately on the participant node. Although there exist some drawbacks in this innovative distributed and decentralized paradigm such as the latency, performance, energy consumptions, privacy issue and etc., transparency and accuracy can be important advantages we can achieve from blockchain-based applications.

III. AI-DATA COMMONS FOR FEDERATED LEARNING

The primitive purpose of AI-Data Commons framework is to create an ecosystem to spur innovative AI services while ensuring the sovereignty of data in selling or sharing it [2,13-15]. A high-level architecture, functional modules, P-C-I (Participation-Collaboration-Incentives) workspace and a PoC of AI-Data Commons have been explained in [13-15] and we have extended AI-Data Commons to provide federated learning.

A. Components of System

In AI-Data Commons, data, algorithms, or AI models can be published on the marketplace and these are called as ADC asset. For federated learning support, we have extended AI-Data Commons so that it can provide functionalities for the collaboration of machine learning more efficiently.

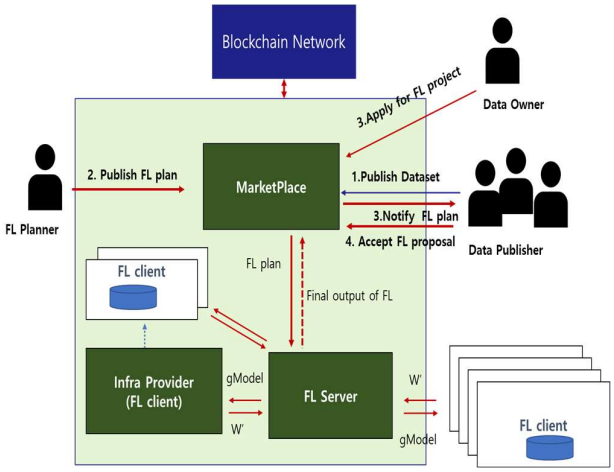

Fig. 2. Interaction of components between blockchain-based marketplace and federated learning in our system

In our work, there needs five system components of the advanced AI-Data Commons as below.

- Market Place(MPSS) : As a frontend, MPSS is a web-portal which provides UI for publishing a federated learning plan file as well as ADC assets. Users also can search and buy or invest ADC assets through MPSS.
- Dapp Server(DSSS) : DSSS is one of nodes which forms of blockchain networking. It stores data on blocks and runs smart contracts under EVM(Ethereum Virtual Machine).
- FL Server(FLS) : FLS provides an initial global model to the FLCs. For every federated rounds, it receives updated model weights from FLC, aggregates those weights for global model and send the global model to FLCs, repeatedly.
- FL Client(FLC) : FLC module can be downloaded and run on the data provider's machine. Using local data, FLCs train the global model received from FLS and send updated model weights to FLS.
- Infrastructure Provider(InfraP) : In case that a data provider does not have any computation resources, AI-Data Commons provides infrastructure as a form of sandbox. For this, InfraP creates a virtual machine to execute FL client which uses data in a secure way. After finishing the federated learning, all data used in training on the virtualized infrastructure are removed.

System components and their interactions are illustrated in Fig. 2. and more detailed sequence is explained in next section.

B. Overall Sequence

Fig. 3. Shows overall procedures of the proposed system, which can be explained into three steps: 1) organizing federated learning, 2) launching federated learning project and 3) providing incentives transparently.

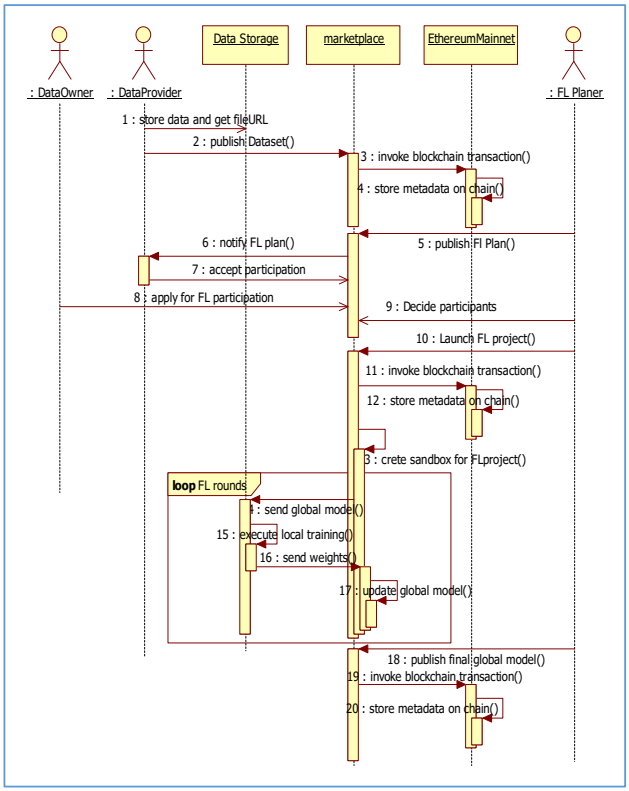

Fig. 3. Sequence Diagram

1) Organizing Federated Learning Coalition using marketplace

- [Seq No.1~2] Data owner publishes a dataset which is stored on a primitive storage to the marketplace and becomes data provider.
- [Seq No.5] FL planner searches for data in need for training AI models through federated learning on the

marketplace. For the registration of a FL plan on the marketplace, the FL planner specifies the task, dataset description in need for local training, relevant data providers who the FL planner wants to participate with the possessed data and incentive information which will be provided to the data providers.

- [Seq No.6] Once the FL plan is published on the marketplace, the system sends notification to the data provider to provide information about the federated learning.
- [Seq No.7] Data provider reviews the FL plan including incentive conditions suggested by FL planner and accepts the proposed task.
- [Seq No.8] A user who does not publish datasets on AI-Data Commons reviews the FL plan on the marketplace and publishes possessed dataset to apply for participating federated learning.
- [Seq No.9] FL planner decides the members of the coalition for the collaborative learning.

2) Launching Federated Learning Project on marketplace

- [Seq No.10~12] FL planner designs the federated learning method in more detail and register this on marketplace as a FL project. Metadata on the project and every transactions are recorded on the blocks.
- [Seq No.13] The system creates a sandbox to perform the federated learning as FL planner designed
- [Seq No.14~17] The processes for the federated learning round proceed repeatedly.

3) Providing incentives transparently

- [Seq No.18] FL planner publishes the final result of the federated learning which is an AI model on the marketplace.
- [Seq No.19-20] Since the result of collaboration is registered as a digital asset in the blockchain-based marketplace, whenever there is a consumption of the digital asset, the profit is automatically divided and incentives are transferred into every participant's wallet as contracted.

IV. Concluding Remark

AI-Data Commons is blockchain-based framework for trustful data trade while preserving privacy. In this paper, we have presented the design of enhanced AI-Data Commons to support efficient methods for federated learning. Considering that sufficient participants who own relevant datasets are significantly important in federated learning in real, we have equipped AI-Data Commons with salient features for facilitating in finding and promoting participants. A blockchain-based incentive mechanism, which guarantees transparency and non-fabrication, plays a role in inspiring more participation in our system. Blockchain-based incentive mechanism which guarantees transparency and non-fabricability plays a role of inspiring more participations in our system.

The proposed system provides an interface for executing federated learning by allowing users to publish federated learning projects and form sandboxes with a FL server and FL clients. If a participant of the collaborative learning can utilize their ample computation resources, the FL client module can be downloaded from AI-Data Commons and communicate with the FL Server. In cases where there is a need for additional infrastructure to perform the FL client, AI-Data Commons can provide virtualized infrastructure while ensuring the protection of sensitive data.

Acknowledgment

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT)(2022-0-01032, Development of Collective Collaboration Intelligence Frame Work for Internet of Autonomous Things).

References

[1] https://datacommons.org

[2] S. Lim, Y. Suh, D. Park, S. Woo and C. Park, "Design of SW Framework for Trustworthy AI-Data Commons," 2020 International Conference on Information and Communication Technology Convergence(ICTC), 2020, pp. 1883-1885, doi: 10.1109/ICTC49870.2020.9289370

[3] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. "Federated machine learning: Concept and applications", ACM Transactions on Intelligent Systems and Technology (TIST), 10(2):12, 2019

[4] David Leroy, Alice Coucke, Thibaut Lavril, Thibault Gisselbrecht, and Joseph Dureau, "Federated learning for keyword spotting", In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 6341–6345. IEEE, 2019

[5] Zhuang Yan, Li Guoliang, and Feng Jianhua, "A survey on entity alignment of knowledge base", Journal of Computer Research and Development, 1:165–192, 2016.

[6] M. Tang and V. Wong, "An Incentive Mechanism for Cross-silo Federated Learning: A Public Goods Perspective," in Proc. of IEEE INFOCOM, 2021.

[7] H. Yu, Z. Liu, Yang Liu, T. Chen, M Cong, X. Weng, D. Niyato, and Q. Yang, "A Fairness-aware Incentive Scheme for Federated Learning," in Proc. of AAAI/ACM Conference on AI, Ethics, and Society (AIES), 2020.

[8] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith," Federated learning: Challenges, methods, and future directions", 2019.

[9] https://ethereum.org/

[10] Y. Abuidris, R. Kumar, T. Yang, J. Onginjo "Secure large-scale E-voting system based on blockchain contract using a hybrid consensus model combined with sharding", ETRI Journal, Vol. 43, Issue 2, Nov. 2020.

[11] W. Mougayar, "The business blockchain: Promise, practice, and application of the next internet technology", John Wiley & Sons, 2016.

[12] S. Brotsis, K. Limniotis, G. Bendiab, N. Kolokotronis, S. Shiaeles, "On the suitability of blockchain platforms for IoT applications: Architectures,security, privacy, and performance, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271‑350.

[13] Y. Suh, S. Woo, B. Eom, D. Park, S. Lim and C. Park "PCI Workplace: A decentralized collaboration framework for AI based problem-solving", 2021 International Conference on Information and Communication Technology Convergence(ICTC), 2021, pp. 1838-1840.

[14] S. Woo, Y. Suh, M. Zubair, B. Eom, D. Park, S. Lim and C. Park "Metadata Modeling and Operation Flow of Problem Solving Ecosystem for ECG Data Engineering and Arrhythmia Diagnosis", 2021 International Conference on Information and Communication Technology Convergence(ICTC), 2021, pp. 1835-1837.

[15] B. Eom, S. Lim, Y. Suh, S. Woo, D. Park and C. Park "Artificial Intelligence-Enabled Data Value Curation on AI-Data Commons", 2021 International Conference on Information and Communication Technology Convergence(ICTC), 2021, pp. 1316-1318.