Blockchain and Federated Learning

Anonymous Authors[1]

Abstract

This document provides a basic paper template and submission guidelines. Abstracts must be a single paragraph, ideally between 4–6 sentences long. Gross violations will trigger corrections at the camera-ready phase.

1 Introduction

1.1 Federated Learning

Federated learning (FL) is a distributed learning paradigm that aims to train machine learning models from scattered and isolated data. The special features of FL (compared to data center-based distributed training) are : statistical heterogeneity, system constraints and trustworthiness. To provide an answer to these 3 challenges we need to mix knowledge from different fields, including machine learning, wireless communication, mobile computing, distributed systems, and information security. Therefore FL is a truly interdisciplinary research field. FL has been studied well over the last years and is still developping. However, the traditional FL framework still faces some problems which reduce the reliability of the whole system. These problems detailed in (Wang & Hu) are the following : single point of failure, false data and the lalck of incentives.

- Single point of failure: this problem is based on the fiability of the aggregator which is the central server in a FL system. This aggregator is employed to perform the integration of local training results and to update the global model. But, if the centralized aggregator is compromised, the whole FL system will go down. The reasons for a compromised aggregator are : an intentionally dishonest aggregation, an accidental network connection failure, or even an unexpected external attack.

-False data: despite the predefined protocols, and due to the huge number of clients in a FL mmodel we have to assume that not all the clients are honest and will use the model as expected. As a consequence, it may exist rogue clients who submit false data about their local training results. Therefore, the global performance of the model can be strongly affected. Besides, the whole FL system might be attacked by malicious clients via other means, such as training the local models using partial datasets.

-The lack of incentive methods: The lack of incentives: in most of the traditional FL, clients contribute their computing powers without receiving any payments, this lead to the difficulty of encouraging clients to follow rigorously the protocol. Therefore the model will lack of reliable data. Moreover, the FL model will also lack of clients. Indeed, as FL requires multiple devices to work collaboratively, especially for the data-intensive training tasks where it needs a large number of participants.

1.2 Distributed Machine Learning

The constraints challenges of real-world federated learning settings involve many open problems for the distribution of a federated machine learning. As a consequence, most researchers working on federated learning problems will likely not be deploying production FL systems, nor have access to fleets of millions of real-world devices. Therefore, we need to distinct the practical settings that motivate the work and experiments conducted in simulation which provide evidence of the suitability of a given approach to the motivating problem. Thus FL research can be seen differently from an experimental perspective compared to ML researches in other fields. Hence, we will need additional considerations to conduct FL research as presented in (Kairouz et al.).

1.3 Blockchain

Blockchain, is an emerging technology that come with several attractive properties : decentralization, anonymity, and traceability. These properties has already shown their utilities in different fields. More recently, blockchain begin to be used to address the challenges faced by the acual FL. First, decentralization can be achieved by the deployment of blockchain in FL by replacing the central aggregator by the peer-to-peer blockchain system. On the other hand,

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the Machine Learning and Systems (MLSys) Conference. Do not distribute.

the job of aggregating the global model can be handled by blockchain nodes to avoid the unreliability of the whole FL system usually caused by the centralized server's failure. Moreover, blockchain gives verification mechanisms to FL in the name of transaction verification. These mechanisms allows FL model to remove the unqualified or even malicious local model updates before the global model update. Furthermore, blockchain can sucessffully distribute rewards to FL clients to reward their participation and honest behaviors.

2 Motivation

In the present time, the use of Machine learning (ML) is constantly increasing in every field. Thus, lots of data are generated and gathered from massive end users to train ML models which bring benefits in terms improve the serices offered to people. Therefore, ML framework usually requires end devices to transfer the collected data to the central server for training. But, this process causes two challenges : A large amount of communication resources is required to transfer data. Then, the submission of raw data increases the risk of privacy leakage, which making data owners unwilling to upload data to the central server by fear of security. FL paired with blockchain appeared to be a good solution to solve this problem as well as the actual limits of FL mentionned earlier (single point of failure, false data and the lalck of incentives).

3 Blockchain-based Federated Learning

Although Federated Learning allows for participants to contribute their local data without it being revealed, it faces issues in data security and in accurately paying participants for quality data contributions. (Martinez et al.) propose an EOS Blockchain design and workflow to establish data security, a novel validation error based metric upon which (Martinez et al.) qualify gradient uploads for payment, and implement a small example of the blockchain Federated Learning model to analyze its performance. (Awan et al., 2019) propose a blockchain-based privacy-preserving federated learning (BC-based PPFL) framework, which leverages the immutability and decentralized trust properties of blockchain to provide provenance of model updates. (Korkmaz et al., 2020) propose a decentralized federated learning approach named Chain FL that makes use of the blockchain to delegate the responsibility of storing the model to the nodes on the network instead of a centralized server. (Sharma et al., 2020) propose a distributed computing defence framework for sustainable society using the features of blockchain technology and federated learning. In the work (Li et al., 2020) propose a crowdsourcing framework named CrowdSFL, that users can implement crowdsourcing with less overhead and higher security. This paper (Chen et al., 2020) propose a distributed computing architecture, the federated Learning based on the consortium blockchain. (Kumar et al., 2021) propose a framework that collects a small amount of data from different sources (various hospitals) and trains a global deep learning model using blockchain based federated learning. Additionally (Kumar et al., 2021) collect real-life COVID-19 patients data, which is, open to the research community. The security of local parameters, the learning quality, and the varying computing and communication resources, are crucial issues that remain unexplored in federated learning schemes. (Guo et al., 2020) propose a data sharing mechanism that combines blockchain and federated learning over smart city. The security of local parameters, the learning quality, and the varying computing and communication resources, are crucial issues that remain unexplored in federated learning schemes. (Lu et al., 2021b) present its potential application scenarios in beyond 5G. (Ma et al., 2021) propose a blockchain-based federated learning framework and a protocol to transparently evaluate each participant's contribution. The framework protects all parties' privacy in the model building phase and transparently evaluates contributions based on the model updates. Collaborative model development and privacy protection are critical considerations while training a global deep learning model. To address these challenges (Durga & Poovammal, 2022) propose a novel framework based on blockchain and the federated learning model. Other influential work includes (Lu et al., 2020a), (Lee & Kim, 2021).

4 Research lines

4.1 Security issues: privacy

There are many security problems neglected in federated learning, for example, the participants may behave incorrectly in gradient collecting or parameter updating, and the server may be malicious as well. (Weng et al., 2021) present a distributed, secure, and fair deep learning framework named *DeepChain* to solve these problems. (Nguyen et al., 2020) provide a state-of-art survey on the integration of blockchain with 5G networks and beyond. (Kumar et al., 2020) introduce a secure and decentralized training for distributed data. (Lu et al., 2020b) propose a new architecture based on federated learning to relieve transmission load and address privacy concerns of providers. (Otoum et al., 2020) introduce a solution that integrates both federated learning and blockchain to ensure both data privacy and network security. The security of federated learning is increasingly being questioned, due to the malicious clients or central servers' constant attack to the global model or user privacy data. To address these security issues (Li et al., 2021) propose a decentralized federated learning framework based on blockchain, i.e., a Blockchain-based Federated Learning framework with Committee consensus (BFLC). (Liu et al., 2020b) propose a blockchain-based secure FL framework to create smart contracts and prevent malicious or unreliable participants from involving in FL. (Lu et al., 2021a) introduce the digital twin wireless networks (DTWN) by incorporating digital twins into wireless networks, to migrate real-time data processing and computation to the edge plane. In further (Xu et al., 2021) propose the concept of device's score and use entropy weight method to measure the quality of model update. Other influential work includes (Lu et al., 2020a).

4.2 Quality management: incentive

There are many security problems neglected in federated learning, for example, the participants may behave incorrectly in gradient collecting or parameter updating, and the server may be malicious as well. (Weng et al., 2021) present a distributed, secure, and fair deep learning framework named *DeepChain* to solve these problems. This technique provides a promising privacy preservation for mobile devices while simultaneously ensuring high learning performance (Kang et al., 2019). (Liu et al., 2020a) propose FedCoin, a blockchain-based peer-to-peer payment system for FL to enable a feasible SV based profit distribution. (Lin et al., 2022) propose a novel Wirelessly Powered Edge intelliGence (WPEG) framework, which aims to achieve a stable, robust, and sustainable edge intelligence by energy harvesting (EH) methods. (Wang et al., 2021) propose SFAC, a secure federated learning framework for UAV-assisted MCS. A new ecosystem of ML model trading over a trusted Blockchain-based network is proposed (Nguyen et al., 2021). Problems, however, can arise if there is a lack of quality data for AI-model training, scalability, and maintenance. (Chaabene et al., 2022) propose a data-centric federated learning architecture leveraged by a public blockchain and smart contracts to overcome this significant issue. The proposed approach employs privacy-preserving bidirectional long-short term memory (BiLSTM) and augments the security through the integration of Blockchain technology based on Ethereum smart contract environment (Rahmadika et al., 2022). While several works focus on strategic incentive designs and client selection to overcome this problem, there is a major knowledge gap in terms of an overall design tailored to the foreseen digital economy, including Web 3.0, while simultaneously meeting the learning objectives. To address this gap (Pandey et al., 2022) propose a contribution-based tokenized incentive scheme, namely `FedToken`, backed by blockchain technology that ensures fair allocation of tokens amongst the clients that corresponds to the valuation of their data during model training. Other influential work includes (Lu et al., 2020a).

5 Conclusion

Acknowledgements

Do not include acknowledgements in the initial version of the paper submitted for blind review.

If a paper is accepted, the final camera-ready version can (and probably should) include acknowledgements. In this case, please place such acknowledgements in an unnumbered section at the end of the paper. Typically, this will include thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

References

Awan, S., Li, F., Luo, B., and Liu, M. Poster: A reliable and accountable privacy-preserving federated learning framework using the blockchain. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, CCS '19, pp. 2561–2563, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367479. doi: 10.1145/3319535.3363256. URL https://doi.org/10.1145/3319535.3363256.

Chaabene, R. B., Amayed, D., and Cheriet, M. Leveraging centric data federated learning using blockchain for integrity assurance, 2022. URL https://arxiv.org/abs/2206.04731.

Chen, Y., Chen, Q., and Xie, Y. A methodology for high-efficient federated-learning with consortium blockchain. In *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 3090–3095, 2020. doi: 10.1109/EI250167.2020.9347025.

Durga, R. and Poovammal, E. Fled-block: Federated learning ensembled deep learning blockchain model for covid-19 prediction. *Frontiers in Public Health*, 10, 2022.

Guo, S., Xiang, B., Xia, X., Yan, Z., and Li, Y. Blockchain and federated learning based data security sharing mechanism over smart city. 11 2020. doi: 10.21203/rs.3.rs-104012/v1.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konečný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Raykova, M., Qi, H., Ramage, D., Raskar, R., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. Advances and open problems in federated learning. URL http://arxiv.org/abs/1912.04977. version: 3.

Kang, J., Xiong, Z., Niyato, D., Xie, S., and Zhang, J. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. *IEEE Internet of Things Journal*, 6(6): 10700–10714, 2019. doi: 10.1109/JIOT.2019.2940820.

Korkmaz, C., Kocas, H. E., Uysal, A., Masry, A., Ozkasap, O., and Akgun, B. Chain fl: Decentralized federated machine learning via blockchain. In *2020 Second International Conference on Blockchain Computing and Applications (BCCA)*, pp. 140–146, 2020. doi: 10.1109/BCCA50787.2020.9274451.

Kumar, R., Khan, A. A., Kumar, J., Zakria, Golilarz, N. A., Zhang, S., Ting, Y., Zheng, C., and Wang, W. Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging. *IEEE Sensors Journal*, 21(14):16301–16314, jul 2021. doi: 10.1109/jsen.2021.3076767. URL https://doi.org/10.1109%2Fjsen.2021.3076767.

Kumar, S., Dutta, S., Chatturvedi, S., and Bhatia, M. Strategies for enhancing training and privacy in blockchain enabled federated learning. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pp. 333–340, 2020. doi: 10.1109/BigMM50055.2020.00058.

Lee, H. and Kim, J. Trends in blockchain and federated learning for data sharing in distributed platforms, 2021. URL https://arxiv.org/abs/2107.08624.

Li, Y., Chen, C., Liu, N., Huang, H., Zheng, Z., and Yan, Q. A blockchain-based decentralized federated learning framework with committee consensus. *IEEE Network*, 35(1):234–241, jan 2021. doi: 10.1109/mnet.011.2000263. URL https://doi.org/10.1109%2Fmnet.011.2000263.

Li, Z., Liu, J., Hao, J., Wang, H., and Xian, M. Crowdsfl: A secure crowd computing framework based on blockchain and federated learning. *Electronics*, 9(5), 2020. ISSN 2079-9292. doi: 10.3390/electronics9050773. URL https://www.mdpi.com/2079-9292/9/5/773.

Lin, X., Wu, J., Bashir, A. K., Li, J., Yang, W., and Piran, M. J. Blockchain-based incentive energy-knowledge trading in iot: Joint power transfer and ai design. *IEEE Internet of Things Journal*, 9(16):14685–14698, 2022. doi: 10.1109/JIOT.2020.3024246.

Liu, Y., Ai, Z., Sun, S., Zhang, S., Liu, Z., and Yu, H. *FedCoin: A Peer-to-Peer Payment System for Federated Learning*, pp. 125–138. Springer International Publishing, Cham, 2020a. ISBN 978-3-030-63076-8. doi: 10.1007/978-3-030-63076-8_9. URL https://doi.org/10.1007/978-3-030-63076-8_9.

Liu, Y., Peng, J., Kang, J., Iliyasu, A. M., Niyato, D., and El-Latif, A. A. A. A secure federated learning framework for 5g networks. *IEEE Wireless Communications*, 27(4):24–31, aug 2020b. doi: 10.1109/mwc.01.1900525. URL https://doi.org/10.1109%2Fmwc.01.1900525.

Lu, Y., Huang, X., Dai, Y., Maharjan, S., and Zhang, Y. Blockchain and federated learning for privacy-preserved data sharing in industrial iot. *IEEE Transactions on Industrial Informatics*, 16(6):4177–4186, 2020a. doi: 10.1109/TII.2019.2942190.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Blockchain empowered asynchronous federated learning for secure data sharing in internet of vehicles. *IEEE Transactions on Vehicular Technology*, 69(4):4298–4311, 2020b. doi: 10.1109/TVT.2020.2973651.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Low-latency federated learning and blockchain for edge association in digital twin empowered 6g networks. *IEEE Transactions on Industrial Informatics*, 17(7):5098–5107, 2021a. doi: 10.1109/TII.2020.3017668.

Lu, Y., Huang, X., Zhang, K., Maharjan, S., and Zhang, Y. Blockchain and federated learning for 5g beyond. *IEEE Network*, 35(1):219–225, 2021b. doi: 10.1109/MNET.011.1900598.

Ma, S., Cao, Y., and Xiong, L. Transparent contribution evaluation for secure federated learning on blockchain. In *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*. IEEE, apr 2021. doi: 10.1109/icdew53142.2021.00023. URL https://doi.org/10.1109%2Ficdew53142.2021.00023.

Martinez, I., Francis, S., and Hafid, A. S. Record and reward federated learning contributions with blockchain. In *2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pp. 50–57. doi: 10.1109/CyberC.2019.00018.

Nguyen, D. C., Pathirana, P. N., Ding, M., and Seneviratne, A. Blockchain for 5g and beyond networks: A state of the art survey. *Journal of Network and Computer Applications*, 166:102693, sep 2020. doi: 10.1016/j.jnca.2020.102693. URL https://doi.org/10.1016%2Fj.jnca.2020.102693.

Nguyen, L. D., Pandey, S. R., Beatriz, S., Broering, A., and Popovski, P. A marketplace for trading ai models based on blockchain and incentives for iot data, 2021. URL https://arxiv.org/abs/2112.02870.

Otoum, S., Al Ridhawi, I., and Mouftah, H. T. Blockchain-supported federated learning for trustworthy vehicular networks. In *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, pp. 1–6, 2020. doi: 10.1109/GLOBECOM42002.2020.9322159.

Pandey, S. R., Nguyen, L. D., and Popovski, P. Fedtoken: Tokenized incentives for data contribution in federated learning, 2022. URL https://arxiv.org/abs/2209.09775.

Rahmadika, S., Astillo, P. V., Choudhary, G., Duguma, D. G., Sharma, V., and You, I. Blockchain-based privacy preservation scheme for misbehavior detection in lightweight iomt devices. *IEEE Journal of Biomedical and Health Informatics*, pp. 1–13, 2022. doi: 10.1109/JBHI.2022.3187037.

Sharma, P. K., Park, J. H., and Cho, K. Blockchain and federated learning-based distributed computing defence framework for sustainable society. *Sustainable Cities and Society*, 59:102220, 2020. ISSN 2210-6707. doi: https://doi.org/10.1016/j.scs.2020.102220. URL https://www.sciencedirect.com/science/article/pii/S2210670720302079.

Wang, Y., Su, Z., Zhang, N., and Benslimane, A. Learning in the air: Secure federated learning for uav-assisted crowdsensing. *IEEE Transactions on Network Science and Engineering*, 8(2):1055–1069, 2021. doi: 10.1109/TNSE.2020.3014385.

Wang, Z. and Hu, Q. Blockchain-based federated learning: A comprehensive survey. URL http://arxiv.org/abs/2110.02182. version: 1.

Weng, J., Weng, J., Zhang, J., Li, M., Zhang, Y., and Luo, W. Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive. *IEEE Transactions on Dependable and Secure Computing*, 18(5):2438–2455, 2021. doi: 10.1109/TDSC.2019.2952332.

Xu, C., Qu, Y., Eklund, P., Xiang, Y., and Gao, L. *BAFL: An Efficient Blockchain-Based Asynchronous Federated Learning Framework*. Deakin University, January 2021. ISBN 9781665427449. URL https://dro.deakin.edu.au/articles/conference_contribution/BAFL_An_Efficient_Blockchain-Based_Asynchronous_Federated_Learning_Framework/20622720/1.

A PLEASE ADD SUPPLEMENTAL MATERIAL AS APPENDIX HERE

Put anything that you might normally include after the references as an appendix here, *not in a separate supplementary file*. Upload your final camera-ready as a single pdf, including all appendices.