DRM-FL: Cross-Silo Federated Learning 접근법의 프라이버시 보호를 위한 분산형 랜덤화 메커니즘

무함마드 필다우스[1], 초느에진랏[2], 마리즈아길랄[2], 이경현[3]
[1]부경대학교 인공지능융합학과
[2]부경대학교 정보보호학과
[3]부경대학교 컴퓨터공학부

{mfirdaus, chocho1612, marizaguilar}@pukyong.ac.kr, and khrhee@pknu..ac.kr

DRM-FL: A Decentralized and Randomized Mechanism for Privacy Protection in Cross-Silo Federated Learning Approach

Muhammad Firdaus[1], Cho Nwe Zin Latt[2], Mariz Aguilar[2], Kyung-Hyune Rhee[3]
[1]Dept. of Artificial Intelligence Convergence, Pukyong National University
[2]Dept. of Information Security, Pukyong National University
[3]Divison of Computer Engineering, Pukyong National University

Abstract

Recently, federated learning (FL) has increased prominence as a viable approach for enhancing user privacy and data security by allowing collaborative multi-party model learning without exchanging sensitive data. Despite this, most present FL systems still depend on a centralized aggregator to generate a global model by gathering all submitted models from users, which could expose user privacy and the risk of various threats from malicious users. To solve these issues, we suggested a safe FL framework that employs differential privacy to counter membership inference attacks during the collaborative FL model training process and empowers blockchain to replace the centralized aggregator server.

1. Introduction

Federated learning (FL), a more advanced variant of distributed machine learning, has rapidly grown popular to improve data security and user privacy in various industries with big data-driven artificial intelligence (AI). FL facilitates shifting trained models instead of user data, where data is usable but not visible in the system. FL's key benefit is that it allows multi-party model learning to be done cooperatively without revealing sensitive personal information.

The majority of contemporary FL approaches, on the other hand, still rely on a central server to aggregate the trained collaborative models into a global model result. In this case, the traditional server collects all uploaded models from participants using a centralized method, resulting in a single point of failure, thus, causing the entire FL system to fail. Furthermore, during the collaborative training process, the hostile users (i.e., attackers) must be considered since they may launch numerous attacks on the system's security flaws, including membership inference, poisoning attacks, and reverse engineering of the trained FL models, to expose the privacy of other users and they even cause the global model's development to fail on purpose.

To overcome the concerns mentioned earlier, we presented a safe FL framework that

capitalizes on the benefits of blockchain and differential privacy. Blockchain with a smart contract, as a distributed ledger technology, is used to provide fully decentralized federated learning by replacing the centralized aggregator server. Likewise, blockchain can create a decentralized incentive system that motivates users to share their learned models. Moreover, during the collaborative FL model training process, we apply differential privacy (DP) to address the problem of participant linkability. As a result, by injecting noise into all users' updated models, DP can mitigate the membership inference attack from malicious users.

The remainder of this article is organized as follows: In Section 2, we provide background information on federated learning, blockchain, and differential privacy. Our proposed paradigm, a secure federated learning architecture, is shown in Section 3. Finally, Section 4 concludes this paper.

2. Related Works

The FL concept was introduced by Google [1] to solve the critical limitations of traditional machine learning algorithms, which may be vulnerable to privacy and security concerns. FL intends to make it more comfortable for participants to collaborate on training models without revealing their personal information; as a result, personal information is preserved private and never leaves their devices [2]. Nonetheless, a centralized aggregator server controls the entire orchestration system in a traditional FL approach. As a result, there is a possibility of a single point of failure and the risks of various attacks on the system's security flaws. Therefore, several works proposed a secure and reliable FL framework to tackle these challenges.

In [3], the authors designed a blockchain-empowered secure FL framework by leveraging smart contracts to prohibit adversary and untrustworthy users from entangling in FL. Here, the centralized server distinguishes hostile and untrustworthy users by automatically enforcing smart contracts to oppose data or model poisoning attacks. On the other hand, the authors [4] suggested DeepChain as a fair, secure, and distributed protocol that uses a blockchain-based incentive mechanism to encourage clients to behave correctly in the system. Moreover, DeepChain substitutes the centralized server with blockchain and ensures that each participant's data is kept private and the entire training process can be audited. Further, in [5], the study proposed a novel framework, NbAFL, that uses a differential privacy method by adding noise before aggregating FL to evade data leakage. In the NbAFL, they also construct a theoretical convergence bound for the trained FL model's loss function.

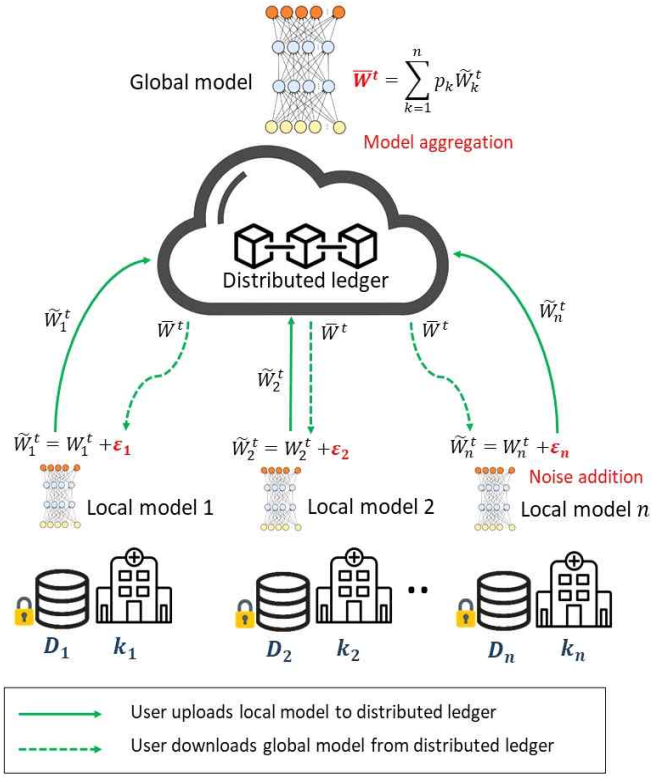
<Table 1> The characteristics of cross-silo FL and cross-device FL [6]

Characteristics	Cross-silo FL	Cross-device FL
Distribution scale	Typically 2-100 users	Massively parallel, up to 10^{10} users
Setting	Users are different organizations (e.g., medical or financial)	The users are a very large number of mobile or IoT devices
Data availability	All users are almost always available	Only a fraction of users are available at any one time
Client statefulness	Stateful, each user may participate in each round of the computation	Generally stateless, each user will likely participate only once in a task
Reliability of clients	Relatively few failures	Highly unreliable

3. Decentralized and Randomized Mechanism for Cross-Silo FL Privacy Protection

Basically, the FL system is categorized into cross-device FL and cross-silo FL based on data availability. Adapted from [6], the detailed

difference characteristics of these two FL categories can be seen in Table 1. In our proposed model, we use a cross-silo federated learning approach since the users are almost always available and participate in each round of the computation.



(Figure 1) Proposed Model.

Figure 1 shows our proposed model to form a secure FL framework by leveraging blockchain and differential privacy techniques. Here, blockchain is utilized to address the bottleneck of a centralized FL server [7] and provides a decentralized rewarding mechanism that motivates users to share their learned models. On the other hand, we apply differential privacy to address the problem of participant linkability by injecting noise into all users' updated models before it uploaded to the distributed ledger. The proposed model is started with the initial learning model process, where the initial model w^t and other parameters, such as a reward amount, evaluation criteria, and a training datasets, are uploaded into the smart contract blockchain. Then, in iteration t, each user downloads w^t and trains its local model w_k^t using its local datasets D_k based on the following formula.

$$F_k(w) = \frac{1}{n_k} \sum_{x_i \in D_k} f_i(w) \tag{1}$$

After user k generates w_k^t, then differential privacy noise ε_k is added to w_k^t using Gaussian mechanism as follows [8].

$$f(D) + N(0, S_f^2 \sigma^2) \tag{2}$$

where $N(0, S_f^2 \sigma^2)$ is the normal distribution with mean 0, and standard deviation $S_f \sigma$. In this case, ε_k is added to achieve ε-differential privacy [8], defined by:

$$\Pr(M(D_1) \in S) \leq e^{\epsilon} \Pr(M(D_2) \in S) + \delta \tag{3}$$

where the term δ is denote to allows for a small probability of failure. Thus, k updates its trained local model with noise $\widetilde{w}_k^t$ to protect against reverse engineering attacks, such as membership inference attacks while uploading the trained local model to distributed ledger blockchain.

Once all users upload $\widetilde{w}_k^t$ to blockchain, a particular consensus mechanism verifies and aggregates $\widetilde{w}_k^t$ to obtain a new global model $\overline{w}^t$ for next iteration through a weighted aggregation [1] as follow.

$$\overline{w}^t = \sum_{k=1}^{n} p_k \widetilde{w}_k^t \tag{4}$$

where n is the number of users, $p_k \geq 0$ and $\sum_k p_k = 1$ [9]. When, the consensus process is passed, the verified $\overline{w}^t$ will be stored in the blockchain. Thus, all users can downloads a new global model $\overline{w}^t$ from distributed ledger for the

next iteration until the model reaches a precise accuracy or the number of iterations exceeding the upper limit

4. Conclusion and Future Direction

We presented the concept of a decentralized and randomized for privacy protection to form a secure cross-silo federated learning framework by leveraging blockchain and differential privacy method. Blockchain is utilized to address the bottleneck of a centralized FL server and provides a decentralized rewarding mechanism that motivates users to share their learned models. On the other hand, we apply differential privacy to address the problem of participant linkability by injecting noise into all users' updated models before it uploaded to the distributed ledger. Even though the global model aggregation via blockchain enhances security and privacy, traditional consensus algorithms among blockchain nodes, on the other hand, can add latency to the aggregation time. Therefore, it is suggested that consensus techniques with minimal latency be developed for future direction.

Acknowledgment

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2022-2020-0-01797) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation) and Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2021R1I1A3046590)

References

[1] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). PMLR.

[2] Hard, A., Rao, K., Mathews, R., Ramaswamy, S., Beaufays, F., Augenstein, S., ... & Ramage, D. (2018). Federated learning for mobile keyboard prediction. arXiv preprint arXiv:1811.03604.

[3] Liu, Y., Peng, J., Kang, J., Iliyasu, A. M., Niyato, D., & Abd El-Latif, A. A. (2020). A secure federated learning framework for 5G networks. IEEE Wireless Communications, 27(4), 24-31.

[4] Weng, J., Weng, J., Zhang, J., Li, M., Zhang, Y., & Luo, W. (2019). Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive. IEEE Transactions on Dependable and Secure Computing, 18(5), 2438-2455.

[5] Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., ... & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 15, 3454-3469.

[6] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends® in Machine Learning, 14(1 - 2), 1-210.

[7] Firdaus, M., & Rhee, K. H. (2021). An Incentive Mechanism Design for Trusted Data Management on Internet of Vehicle with Decentralized Approach. Journal of the Korea Institute of Information Security & Cryptology, 31(5), 889-899.

[8] Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci., 9(3-4), 211-407.

[9] Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60.