Federated Machine Learning and Blockchain

Ponnusamy Vijayakumar[1,*], Nemanja Zdravković[2], Emilija Kisić[2] and V Saraswathi[1]

[1]SRM IST, ECE Department, Kattankulathur, Chennai, India
[2]Belgrade Metropolitan University, Tadeuša Košćuška 63, Belgrade, Serbia

Abstract
Federated Learning (FL) is a distributive machine learning approach for privacy-preserving sensitive data. FL uses a server to coordinate, initiate distributive learning and generate the aggregated model from the outcome of distributed nodes. Each collaborative distributive node generates a local model or portion of the model by tanning from its locally available data. The locally trained model weights are communicated to the server for generating the aggregated final model. The server shares the aggregated final model with the collaborating nodes to use. During those weight communication and model communication, FL algorithms suffer from the attack of malicious data. Generally, federated learning can use encryption or other defined techniques to ensure data privacy or security. But encryption makes more computational complexity and implementation complexity in the large distributed system. A blockchain is a distributed public digital ledger that generates blocks for every transition and stores them in all the computers in the network. The block of record cannot be altered without altering the content in all computers in the network's consensus. This distributive nature and immutability of transactions of blockchain enable the Integration of Blockchain with FL for securing the trained models' integrity. This research presents a comprehensive survey of blockchain-integrated Federated Learning.

Keywords
blockchain, big data, blockchain-integrated federated learning, distributive learning, federated learning, machine learning

1. Introduction

Federated Learning is alternatively referred to as collaborative learning is a kind of machine learning that involves the training of an algorithm over a network of dispersed edge devices or servers that store local data slices without swapping them. Federated learning allows several parties to construct a strong machine learning model with no information sharing, there by addressing crucial concerns such as data security, data privacy, access rights and access to diverse data. Its use cases are spread in a wide variety of business, including pharmaceutics, defense, IoT, and telecommunications. The attractive characteristics of Federated learning [1] are:

- Universality for cross-organizational scenarios-make cross-organizational enterprise set up by vertical organizational participation. For example, banks that possess clients' purchasing data could cooperate with online purchase applications for the recommendation system of the products.
- Enormously non-identically independent distribution (Non-IID)-client applies unbalanced (different amount, size and nature) and non-IID data (different distribution) due to the heterogeneity of distributed device resources, which makes the machine learning model more robust
- Decentralized technology - the client completely autonomously uses the local data and trains the model without the intervention of the server. Because of this, FL is a unified technology that facilitates the cooperative integration of machine learning models and data amalgam in a distributed setting.
- Parity in standing for each node-client with small data sets also given equal weightage.

There are a few research challenges [2] for implementing federated machine learning:

- Imbalanced Data,
- Statistical Heterogeneity ,
- Expensive Communication Systems ,
- Heterogeneity,
- Privacy Concerns.

Among those challenges, this article deals with the last challenges of privacy concerns.

In federated learning, only model updates (like gradient information) are shared to the central server without sharing the data. But, transmitting local model updates during the training process may result in a third-party attack and remapping the data is possible by the received training update. Differential privacy and secure multiparty computation provide privacy at the expense of reduced accuracy or system efficiency. This trade-off can be avoided by the use of blockchain technology. Thus this research article intended to review those block chain

BISEC'22: 13th International Conference on Business Information Security, December 03, 2022, Belgrade, Serbia
*Corresponding author.
vijayakp@srmist.edu.in (P. Vijayakumar); nemanja.zdravkovic@metropolitan.ac.rs (N. Zdravković); emilija.kisic@metropolitan.ac.rs (E. Kisić); saraswav@srmist.edu.in (V. Saraswathi)

based frameworks and approaches for the federated machine learning.

2. Cutting edge of blockchain-based federated machine learning

Federated learning distributive learning uses sensitive data such as patient health data, safety-related industrial data, and personal banking information collected via the Internet of Things (IoT) applications which need to be secured and the privacy of data to be maintained while training and testing of the machine learning models.

Figure 1 shows the typical federated machine learning model. The model shows that the local information are habituated to train the model, and the model update of new weights are gradient values are updated to the central server, aggregating the local model updates and creating a global machine learning model. This global protype is shared with all clients to use. This communication of model updates from the collaborative client node to the server and the global model to the client collaborator has proven to attack where the attacker can retrieve the data from the model. This reverse data recovery by the attacking node raises the security and privacy issue.

The blockchain is a distributed ledger that keeps records like a series of blocks. The blockchain is created by the cryptographic hash of the prior block. This blockchain is immutable because the data can't be modified without affecting the existing block data, which in turn cannot be altered without changing its prior block data. Every time data is added in the form of tractions, it will be updated on each and every peer node.

Figure 2 shows the blockchain-enabled federated machine learning. The weight update from the client will be verified by the miner using the consensus algorithm and some unique thrust test mechanisms. If the node passes the thrust test and is only allowed to update to the global model, the poisoning attack can be avoided through this mechanism. Adaptation of blockchain technology between the server and client node can ensure sustaining information security and privacy [3].

There are many blockchain-based frameworks for federated machine learning proposed in the literature. Few of the recent literature is discussed in the following paragraphs. Industry 4.0 automation is advancing cognitive computing, an AI concept miming human cognition. AI and machine learning technologies have enhanced decision-making and data-driven intelligent manufacturing. The challenges, including poisoning assaults, performance, and limited data resources, must be handled.

Recent studies minimally examined the topic, resulting in unpredictable performance, inefficiency, and privacy

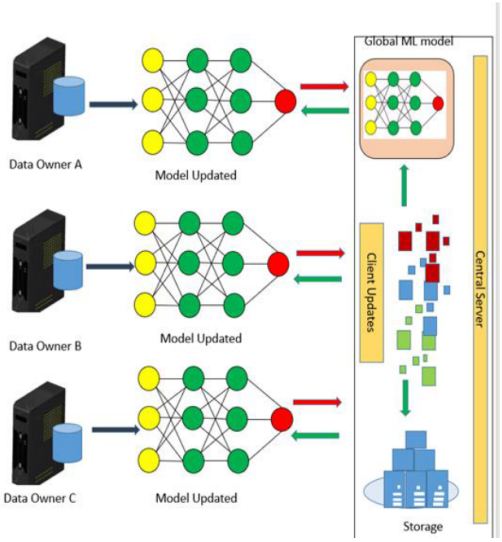

Figure 1: Federated machine learning framework.

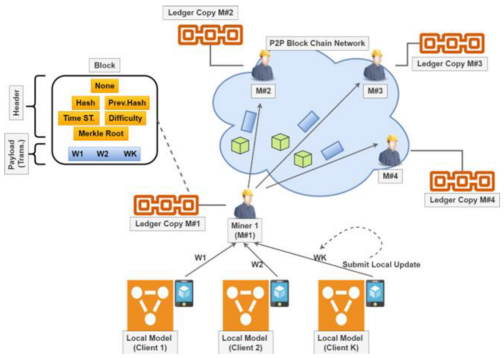

Figure 2: Blockchain-enabled Federated machine learning.

leaks. Federated learning with blockchain created a distributed model for big data-driven cognitive computing (D2C). Federated learning solves "data island" with data security and processing speed, while blockchain offers incentive mechanisms, decentralization, and poisoning resistance. Blockchain-enabled federated learning accelerates member selection and verification. D2C outperforms leading models in extensive review and assessment. Industry 4.0 manufacturing performance improved by federated learning model for big D2C improves performance and privacy issues related to cognitive computing improved significantly.

By integrating blockchain into federated learning, cognitive computing may increase accuracy, poisoning attack resistance, and incentive mechanism for Industry

4.0 automation. An optimization model employs a modified Markovian decision process to increase poisoning resistance and accuracy. The CIFAR – 10 dataset is used for global accuracy prediction with the time setting of 4 rounds with an arrival time 0.7; the global accuracy is 0.82 [4].

As a cutting-edge innovation, the digital twin has tremendous potential. with the fast growth of the Internet of Things (IoT), which links physical components with digital space to optimize physical systems. Limited cellular resources and security concerns prevent IoT digital twin implementation. Blockchain-based digital twin edge network architecture was proposed for pliable and safe digital twin creation. First joint federated learning was established using an access point (AP) to let resource-limited smart devices build digital twins at mobile network operator network edges (MNOs). A directed acyclic graph (DAG) blockchain-based model was proposed to protect local and global model modifications. An iterative double auction-based local model update verification and cooperative federated learning system incentivizes APs to help train local models for resource-limited smart devices and verify local model updates. A 5x5 convolutional neural network (CNN) model is constructed as a collaborative federated learning approach for machine learning and local model update verification with a learning accuracy of [0.1, 0.8] [5].

Federated learning (FL) allows user data privacy-preserving machine learning (ML) model training. Existing FL techniques can enable collaborative ML, but secure sharing of training data is difficult, especially with adversarial FL clients. FL security problems and severe privacy rules require a solid and trusted FL infrastructure, such as blockchain. Scalability and client engagement issues limit blockchain-based solutions' industrial strength. Block chain integration Classic FL, Hyperledger Fabric, and gamification are all intended to be integrated via FL. It is advised that FL asynchronous and synchronous collaborative activities use a secure application for closing and signing off. The immutable ledger of an enterprise-level blockchain network may be implemented at many FL levels to improve both industrial security and auditable traceability. We evaluate three datasets to demonstrate how the enhanced security of the FL process allows for a more precise global ML model to converge to its optimum performance. FabricFL's overall training time is 0.17 s longer than BasicFL's. FabricFL has a minimum and maximum training round timestamps of 46.93 and 58.82, while BasicFL has 48.22 and 59.28 [6].

For multi-party machine learning, federated learning is the answer to the problem of protecting the confidentiality of training data. There are potential security risks associated with using this method. A trustworthy and secure federated learning platform shared by many data owners is being developed using Hyperledger Fabric, a permissioned blockchain architecture. Local modifications are encrypted using homomorphic threshold encryption and then added to a distributed ledger. The privacy and security flaws in the federated learning system have been exposed, but the suggested approach has been shown to fix them in the security analysis. The suggested device runs on a 2.10GHz CPU and 16GB of RAM running Ubuntu 20.04.1 LTS. To train machine learning models, PyTorch is employed. Model accuracy is little impacted by homomorphism encryption [7].

Federated learning is used in speech recognition, image classification, healthcare and smart cities. Numerous investigations have shown that current federated learning methods are susceptible to attacks and fail to fulfill practical security requirements. Scheming a safe federated learning system to ensure training accuracy is yet unsolved. VFChain is proposed, a blockchain-based auditable and verifiable federated learning architecture. Blockchain-selected committee aggregates models and records verifiable proofs. To offer audibility, a blockchain-based on the use of a verified data structure is advocated to verifiable proof search efficiency and secure committee rotation. An optimization technique for multiple-model learning problems improves search efficiency. VFChain was built using well-known deep learning models and tested on a public dataset with actual data [8].

With federated learning, several users may train the same machine learning model in parallel with no sharing of information. Due to the hub of the framework and the untrusty of clients, traditional federated learning systems are accessible to poisoning attacks from malicious clients and servers. A blockchain-based Privacy-preserving Byzantine-robust Federated Learning (PBFL) strategy was proposed to reduce the influence of the malicious clients and central server. Cosine similarity is employed to evaluate Clients with malicious intent uploading malicious gradients, and completely homomorphic encryption is used for safe aggregation. This system has a trustworthy root and uses blockchain for transparency. The scheme's calculation in ciphertext settings is explained. The poisoning-resistant technique can obtain comparable outcomes to the non-attack FedSGD scheme even with a minimal root dataset. A balanced client data distribution evaluates the technique [9].

The Internet of Vehicles (IoV) became essential for creating Smart Transportation Systems once the Internet of Things (IoT) was integrated into transportation (STS). Due to software and wireless connectivity, STS vehicles and equipment are vulnerable to cyberattacks. A federated deep learning-based intrusion detection system (FED-IDS) is proposed that offloads server learning to dispersed vehicle edge nodes to effectively identify threats. FED-IDS uses a factors-aware transformer network to acquire spatial-temporal vehicle traffic flows for attack classification. Blockchain-managed federated training

lets numerous edge nodes provide distributed, safe, and reliable training without a centralized authority. As a safety measure, miners check distributed local updates from participating vehicles before they are added to the blockchain. FED-IDS outperformed state-of-the-art methods on Car-Hacking and TON IoT datasets. It validates the cyber-protection of intelligent transportation system networks. The proposed FED-IDS with Accuracy: 1.69%, F1- score: of 2.3% [10].

Blockchain-enabled Federated Learning (BFL) securely stores model changes on the blockchain. Mining delays training. Mobile devices have CPU and energy limits. The machine learning model owner (MLMO) must choose the energy and data used by devices for training and the block production rate to reduce the mining costs and system latency while attaining the goal accuracy. Deep reinforcement learning is suggested for MLMO decision-making under BFL uncertainty. The proposed DRL scheme provides up to a 12% improvement in latency over the Greedy approach [11].

To train a global deep learning model from a little amount of data from several hospitals, blockchain-based federated learning is recommended. Blockchain technology verifies data, and federated learning trains the model worldwide while protecting organization privacy. A data normalization method was proposed to handle data heterogeneity from hospitals with diverse CT scanners. Patients with COVID-19 were identified using a Capsule Network-based segmentation and classification approach. Cooperatively training a global model with federated learning and privacy using blockchain technology. COVID-19 patient data is collected for the study. Updated data enhances CT image recognition in the proposed framework. Federated blockchain and capsule network with the performance of 98.68% of improved detection of COVID – 19 [12].

IIoT devices are widely used in intelligent industries. Blockchain is strong yet susceptible to hacks. Smart factory blockchain-based IIoT networks must detect abnormalities to prevent attacks. Federated learning is employed to construct block hunter, a threat-hunting framework that automatically hunts for threats in blockchain-based IIoT networks. Block hunter uses federated machine learning models and a cluster-based architecture to discover anomalies. Block hunter IIoT network federated threat hunting approach that detects unusual activity while protecting privacy. The block hunter can accurately identify unusual behaviors with low bandwidth and FedAvg approach is applied with the detection of 95% [13].

3. Conclusion

Federated learning is a distributive machine learning approach that is a promising technology to uses a massive amount of data to make a reliable machine learning model. Bock chain is the distributive ledger mechanism that uses the hashing principle and consensus mechanism to maintain immutable records. Since both technologies are based on the distributed computing mechanism, both can be merged to make more advancements. This article reviewed such various merged schemes in various domains of the field and applications. The review shows that merging those two technologies makes the more promising solution for their reliable deployment.

Acknowledgment

This paper was supported by the Blockchain Technology Laboratory at Belgrade Metropolitan University, Belgrade, Serbia.

References

[1] L. Li, Y. Fan, M. Tse, K.-Y. Lin, A review of applications in federated learning, Computers & Industrial Engineering 149 (2020) 106854.

[2] M. Shaheen, M. S. Farooq, T. Umer, B.-S. Kim, Applications of federated learning; taxonomy, challenges, and research trends, Electronics 11 (2022) 670.

[3] D. Li, Z. Luo, B. Cao, Blockchain-based federated learning methodologies in smart environments, Cluster Computing 25 (2022) 2585–2599.

[4] Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, Y. Xiang, A blockchained federated learning framework for cognitive computing in industry 4.0 networks, IEEE Transactions on Industrial Informatics 17 (2020) 2964–2973.

[5] L. Jiang, H. Zheng, H. Tian, S. Xie, Y. Zhang, Cooperative federated learning and model update verification in blockchain empowered digital twin edge networks, IEEE Internet of Things Journal (2021).

[6] V. Mothukuri, R. M. Parizi, S. Pouriyeh, A. Dehghantanha, K.-K. R. Choo, Fabricfl: Blockchain-in-the-loop federated learning for trusted decentralized systems, IEEE Systems Journal (2021).

[7] J. Sun, Y. Wu, S. Wang, Y. Fu, X. Chang, Permissioned blockchain frame for secure federated learning, IEEE Communications Letters 26 (2021) 13–17.

[8] Z. Peng, J. Xu, X. Chu, S. Gao, Y. Yao, R. Gu, Y. Tang, Vfchain: Enabling verifiable and auditable federated learning via blockchain systems, IEEE Transactions on Network Science and Engineering 9 (2021) 173–186.

[9] Y. Miao, Z. Liu, H. Li, K.-K. R. Choo, R. H. Deng, Privacy-preserving byzantine-robust federated learning via blockchain systems, IEEE Transactions on Information Forensics and Security 17 (2022) 2848–2861.

[10] M. Abdel-Basset, N. Moustafa, H. Hawash, I. Razzak, K. M. Sallam, O. M. Elkomy, Federated intrusion detection in blockchain-based smart transportation systems, IEEE Transactions on Intelligent Transportation Systems 23 (2021) 2523–2537.

[11] N. Q. Hieu, T. T. Anh, N. C. Luong, D. Niyato, D. I. Kim, E. Elmroth, Deep reinforcement learning for resource management in blockchain-enabled federated learning network, IEEE Networking Letters (2022).

[12] R. Kumar, A. A. Khan, J. Kumar, N. A. Golilarz, S. Zhang, Y. Ting, C. Zheng, W. Wang, et al., Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging, IEEE Sensors Journal 21 (2021) 16301–16314.

[13] A. Yazdinejad, A. Dehghantanha, R. M. Parizi, M. Hammoudeh, H. Karimipour, G. Srivastava, Block hunter: Federated learning for cyber threat hunting in blockchain-based iiot networks, arXiv preprint arXiv:2204.09829 (2022).