International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Federated Learning Frameworks for Secure and Decentralized Authentication

Merlin Balamurugan

Vice President, Digital Engineering, Leading Banking Organization

Abstract

Federated Learning (FL) [1] is a cutting-edge machine learning approach that enables multiple or edge devices to train models collaboratively without sharing sensitive data. This approach not only ensures the privacy and security of data by keeping it localized but also promotes the collective improvement of machine learning models across various participants. A systematic literature review explored integrating Blockchain technology with federated learning. Blockchain's potential to address existing security and privacy vulnerabilities in traditional federated learning systems is analyzed in depth. One of the key benefits of combining Blockchain with FL is enhanced protection against potential attacks, such as data tampering or unauthorized access. The study also examines how Blockchain-based federated learning systems can offer better records and rewards management, contributing to fairer and more transparent systems. In addition, Blockchain's role in improving verification and accountability within federated learning frameworks has been critically evaluated. By integrating Blockchain, federated learning can achieve higher levels of trust and security in collaborative machine-learning processes. The latest research highlights innovative Blockchain-based methods that tackle these challenges, ensuring robust privacy and security measures. Overall, this approach represents a significant advancement in distributed machine learning, aligning with contemporary needs for data protection and collaborative efficiency.

Keywords: Federated Learning, Distributed machine learning, Blockchain, Smart contract, Privacy and Security

1. Introduction

Traditional centralized machine learning methods collect and store data in a central data center or cloud service for processing and training. This approach consolidates data in one location, making it easier to manage but potentially exposing sensitive information to greater risks. Federated Learning introduces a significant shift by decentralizing the training process across numerous devices. Rather than moving data to a central location, Federated Learning enables each device to handle its local data independently.

The fundamental concept behind Federated Learning involves deploying the machine learning model directly onto each participant's local device. Each device then trains the model using only the data on that particular device. As a result, data never leaves its original location, preserving privacy and security. After training, the device generates updates to the model based on its local data. These updates are kept confidential and are stored only on the local device.

The next step involves aggregating these local updates to form a global update. The global update is then sent from each device to a central server. The central server collects and merges these global updates with the existing global model. This process helps improve the model by incorporating diverse data insights while maintaining privacy.

Once the global model is updated, the central server returns the revised model to the participating devices. Each device then receives this updated model and continues local training if necessary. This iterative training, updating, and aggregating process continues, gradually refining the model with contributions from all participating devices.

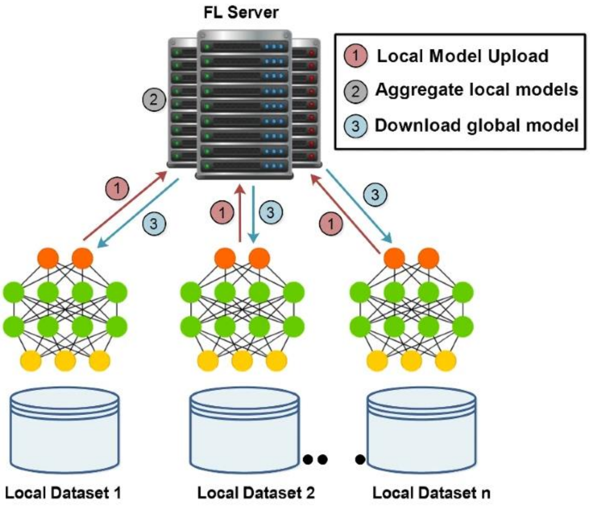

Figure 1: Federated Learning architecture [2]

By decentralizing the training process, Federated Learning enhances data privacy and reduces the risk of exposing sensitive information. It also allows for more efficient use of distributed resources, leveraging the computational power of numerous devices. This approach aligns well with modern data security and privacy needs while enabling practical collaborative model training.

2. Problem Statement

Federated learning (FL) is a promising framework for distributed machine learning that trains models without sharing local data while protecting privacy [3]. FL exploits the concept of collaborative learning and builds privacy-preserving models. This section explains existing attacks in the federated learning architecture:

Single point of failure attack: In various situations, the central server can compromise the security of the FL system, such as (1) instability leading to a system crash, (2) a compromised central server generating a false global model, and (3) maximum consumption of system resources.

Denial of service and distributed denial of service attack: By continuously propagating fake model updates, malicious devices can stress the system so much that it crashes, called a Denial of Service (DoS) attack. Similarly, if an FL server is compromised, it repeats this process and paralyzes the entire FL system, referred to as a Distributed Denial of Service (DDoS) attack.

Free-riding attack: In the FL model training task, high cost induces dishonest participants to gain incentives without contributing to local model updates. For instance, free-riders send fake or similar model updates with minimum noise and can directly upload the untrained model. Hence, this situation in FL systems leads to issues of fairness and trustworthiness.

Poisoning attacks: Poisoning attacks are categorized into data poisoning and model poisoning. The data poisoning attack is launched by changing the model’s training data, and false model updates are propagated. Furthermore, malicious participants can flip the labels of datasets and implement the predefined poisoned model updates, which degrade the performance of the global FL model. Therefore, data poisoning attacks ultimately lead to model update poisoning attacks. Besides, reverse and random model poisoning attacks are also generated in FL systems.

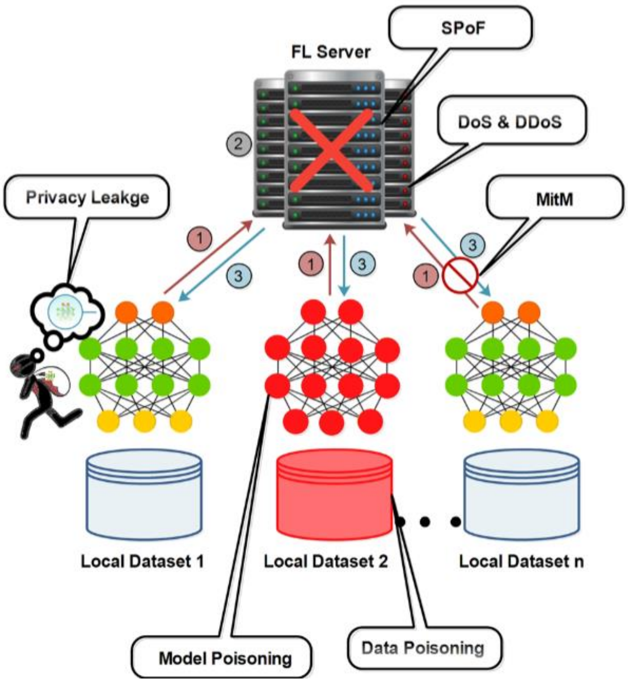

Figure 2: Attacks to Federated Learning [3]

Man-in-the-middle attack: In this attack, the attacker pretends to be an FL server or client to send fake model updates and control the traffic. The common types of MitM attacks are session hijacking and Internet Protocol (IP) spoofing. Meanwhile, the attacker hijacks a legitimate session between a trusted FL client and the FL server in session hijacking. IP spoofing relates to convincing the FL server or clients that they are in connection with a trusted entity; however, in reality, the attacker is acting on the other side.

Eavesdropping Attacks: The eavesdropping attack in the FL system leaks sensitive information about FL participants, such as gender, profession, location, etc. Similarly, an adversary can delete, modify, corrupt, or intercept the broadcasted model between the FL server and participants.

3. Solution

Blockchain technology [4] is a cutting-edge term known for the decentralized ledger technology that keeps an immutable record of transactions. It has a blockchain containing the associated block's transaction record, timestamp, and hash value. The transactions in the blockchain are digitally signed, and the hash is stored to retrieve the information for the next time. In this way, the history of all transactions can be recorded in a tamper-proof manner. Furthermore, the blocks are connected in a Peer-to-Peer (P2P) network and maintain the cloned version of the integral transaction logs.

Blockchain’s essential components:

i. FL participants: Participants work as entities or devices, as in a traditional FL environment. FL participants participate in model training and send local model updates to the next phase for verification and aggregation. At first, the initial model is sent to all participating clients in the FL system. Then, FL participants generate local model updates based on their raw datasets. FL participants and miners directly communicate with each other.

ii. FL integration with blockchain: The integration acts as middleware interacting with FL participants [11] and the blockchain. The authors used the REST-API (Representational state transfer-application Programming Interface) to interact with the Hyperledger Fabric blockchain to record and incentivize gradient uploads. Furthermore, gRPC API facilitates data transfer between FL clients and the Ethereum blockchain network using remote procedure calls (RPC) developed by Google.

iii. Miners working: The miners can be personal computers, standby servers, or cloud-based nodes if they willingly download the mining software. At this step, the FL participants send the local model updates to the miners [12]. Each of the FL participants/data holders is directly connected with the miner and ensures constant communication. The miners are responsible for receiving the local model updates from participating FL devices or participants. Furthermore, aggregation is performed based on the consensus algorithm, and a block is uploaded to the blockchain network.

iv. Smart contract: The Smart Contract (SC) in the blockchain system opens new doors for decentralized applications and automatically executes the program logic when it meets the pre-defined conditions. All conditions are transparent [13] and immutable to participated FL clients, and before they join the FL model training process, they will agree on them. Furthermore, SC allows the clients to codify agreements without any trusted third party. Researchers used smart contact in different ways, such as registering the participants, coordinating the model training, aggregating the local model updates, evaluating the participants' contributions, and awarding rewards. A smart contract is assigned between FL participants and miners.

v. Consensus algorithm: In the blockchain network, the consensus algorithm is the backbone and plays a significant role in validating transactions. All parties establish a joint agreement that defines how a new block is formed, verified, and accepted on a blockchain network. As miners reach the consensus mechanism, such as Proof of Work (PoW), Proof of Stake (PoS), and Byzantine Fault Tolerance (BFT), to name a few, then a new block is appended into the blockchain. By adopting blockchain technology in federated learning, it becomes more flexible. FL participants will start a new FL training process, and through a consensus algorithm, miners reach an agreement to build a fully converged global model. With the successful execution of the consensus algorithm, the block is added to the blockchain network.

vi. Blockchain network: Finally, verified new blocks are added to the blockchain network. The FL model process continues until it reaches the required learning rate. After that, FL clients or other participants can request to download the global model for their purposes. Finally, the global model can be downloaded by the miners, and FL participants can get the model from them.

4. Application of the solution in various organization processes

Integrating Blockchain with Federated Learning combines both technologies' strengths to create more secure, transparent, and efficient decentralized systems. Here is how Blockchain-Federated Learning (B-FL) can be applied to various organizational processes:

Healthcare

- Medical Data Sharing: B-FL can be used for collaborative research among hospitals and medical institutions. Blockchain ensures the integrity and traceability of medical records, while federated learning allows for training models on distributed patient data without compromising [5] privacy.

- Clinical Trials: By using B-FL, organizations can securely share and aggregate clinical trial data from various sources, ensuring transparency in trial results and preventing data tampering.

Finance

- Fraud Detection: Banks and financial institutions can use B-FL to collaboratively train fraud detection models on transaction data from multiple institutions. Blockchain [6] ensures the security and immutability of transaction records, while federated learning helps detect fraudulent patterns across different datasets.
- Credit Scoring: Financial organizations can use B-FL to build more accurate credit scoring models by combining data from various sources without revealing sensitive information. Blockchain can record data provenance and ensure compliance with data protection regulations.

Insurance

- Claims Processing: B-FL can streamline the insurance claims process by analyzing claims data from various insurers while maintaining data privacy. Blockchain ensures that the claims data is tamper-proof and transparent.
- Risk Assessment: Insurance companies can use B-FL to build risk assessment models collaboratively. Blockchain provides a secure and immutable record of risk-related data, and federated learning aggregates insights from different insurers.

Smart Cities

- Traffic Management: In smart cities, B-FL can optimize traffic management systems by analyzing data from various sensors and traffic cameras. Blockchain ensures the data's integrity [7] and security, while federated learning helps make real-time traffic predictions and decisions.
- Energy Distribution: B-FL can be used to optimize energy distribution and consumption in smart grids. Blockchain ensures transparency and security of energy transactions, while federated learning [8] models predict energy demand and supply efficiently.

Legal and Compliance

- Contract Management: B-FL can be used for managing legal contracts and compliance documents. Blockchain ensures that contract terms are immutable and verifiable, while federated learning analyzes compliance patterns across different contracts.
- Regulatory Reporting: Organizations can use B-FL to streamline regulatory reporting processes. Blockchain provides a secure record of compliance data, and federated learning models can analyze data to ensure adherence to regulations.

Telecommunications

- Network Optimization: B-FL can help optimize network performance by analyzing data [9] from various network nodes. Blockchain secures and records network transactions, while federated learning models enhance network efficiency and reliability.
- Customer Experience: Telecommunications companies can use B-FL to improve customer experience by analyzing customer feedback and usage patterns across providers while maintaining data privacy.

Retail

- Customer Insights: Retailers can use B-FL to analyze customer behavior and preferences from data collected across various stores. Blockchain [10] ensures the security and authenticity of transaction records, while federated learning helps in generating actionable insights.

- Inventory Management: B-FL can optimize inventory management by analyzing sales and inventory data from multiple retail locations. Blockchain ensures transparency in inventory transactions, while federated learning predicts inventory needs more accurately.

Government and Public Services

- Public Health Monitoring: Governments can use B-FL to monitor and respond to public health issues. Blockchain secures health data records, while federated learning models predict outbreaks and optimize public health responses.
- E-Governance: B-FL can enhance e-governance systems by securely processing and analyzing data from various government departments. Blockchain provides a transparent and secure record of transactions and decisions.

5. Benefits of solutions

Combining Blockchain with Federated Learning offers some of the following benefits, particularly in terms of security, privacy, and efficiency:

Enhanced Data Privacy and Security:

- Data Privacy: Federated Learning allows models to be trained on decentralized data without sharing it. Blockchain adds an extra layer of security by ensuring the integrity of data interactions and transactions. Sensitive data remains local and private, mitigating risks [14] associated with centralized data storage.
- Data Integrity: Blockchain's immutable ledger guarantees that all interactions and data exchanges are recorded in a tamper-proof manner. Being immutable ensures the accuracy and reliability of the training process, preventing unauthorized modifications.

Improved Transparency and Trust:

- Transparent Auditing: Blockchain provides a transparent and traceable record [15] of all transactions and model updates. This transparency helps build trust among parties involved in the federated learning process, as all actions are logged and can be audited.
- Accountability: With blockchain, each participant's contributions and actions are recorded, which enhances accountability. This recording is crucial in verifiable compliance and integrity scenarios, such as regulatory reporting and collaborative research.

Decentralized Collaboration:

- Secure Collaboration: B-FL facilitates secure collaboration among multiple parties without requiring a central authority. Blockchain supports decentralized governance, allowing various stakeholders to contribute to and benefit from the learning process while maintaining control over their data.
- Scalability: The decentralized nature of B-FL makes it easier to scale collaborative efforts. Organizations and institutions can join the federated learning network and contribute their data and computing resources, expanding the model’s capabilities.

Reduced Risk of Data Breaches:

- Minimized Exposure: Since data remains decentralized and not aggregated in a central repository, the risk of large-scale data breaches is reduced. The attacker cannot access the complete dataset even if one node is compromised.
- Secure Updates: Blockchain’s cryptographic security ensures that model updates and transactions are safe. This security reduces the risk of malicious updates or tampering with the model’s training process.

Regulatory Compliance:

- Data Sovereignty: B-FL supports compliance with data protection regulations, such as GDPR and CCPA, by ensuring that data never leaves its origin and is processed locally. Blockchain's transparency helps demonstrate compliance with regulatory requirements.
- Audit Trails: The immutable nature of blockchain provides a comprehensive audit trail of all data interactions and model updates, making it easier to comply with regulatory requirements and conduct audits.

Enhanced Model Accuracy and Robustness:

- Diverse Data Sources: Federated Learning aggregates insights from diverse data sources, which can improve the accuracy and robustness of the trained models. Blockchain ensures that the contributions from different sources are authentic and verifiable.
- Adaptability: B-FL allows for continuous model improvement by incorporating new data and insights from various nodes in the network, enhancing the model's adaptability to evolving trends and patterns.

Incentive Mechanisms:

- Tokenization: Blockchain can enable token-based incentive mechanisms, rewarding participants for contributing to the federated learning process. This incentive encourages more organizations to participate and share their data or computational resources.

Conflict Resolution:

- Dispute Management: Blockchain's transparency and immutability help resolve disputes by providing a transparent and verifiable record of all transactions and model updates. This transparency can be precious in scenarios involving multiple stakeholders with differing interests.

6. Conclusion

Here is a comprehensive conclusion on choosing the solution [16]:

- Decentralized Security Enhancement: Integrating blockchain technology into Federated Learning (FL) architecture introduces a decentralized approach that enhances security and robustness by linking blocks in a secure chain, mitigating risks associated with centralized data storage.
- Immutable Model Records: The deployment of smart contracts ensures that all model updates are immutable and transparently recorded, preserving the integrity and historical accuracy of the federated learning process.
- Improved Privacy: Blockchain technology significantly boosts privacy within FL by securing data transactions and updates without exposing sensitive information, ensuring that data remains protected throughout learning.
- Efficiency Boost: Blockchain integration optimizes FL efficiency by streamlining data management and synchronization, reducing central bottlenecks, and effectively leveraging decentralized resources.
- Incentive Mechanism Implementation: Blockchain facilitates the implementation of an incentive mechanism that motivates participants to engage and contribute their resources, enhancing the overall performance and effectiveness of the federated learning system.
- Overall System Enhancement: Combining blockchain with FL benefits the system by increasing security, efficiency, privacy, and participant motivation, leading to a more robust and performant federated learning framework.

7. References

1. Abdulrahman S, Tout H, Ould-Slimane H, Mourad A, Talhi C, Guizani M (2021) A survey on federated learning: the journey from centralized to distributed on-site learning and beyon
2. https://link.springer.com/article/10.1007/s10462-022-10271-9
3. https://www.researchgate.net/publication/363613142_Securing_federated_learning_with_blockchain_a_systematic_literature_review
4. Drungilas V, Vaičiukynas E, Jurgelaitis M, Butkienė R, Čeponienė L (2021) Towards blockchain-based federated machine learning: smart contract for model inference
5. Agbo C, Mahmoud Q, Eklund J (2019) Blockchain technology in healthcare: a systematic review
6. Ali O, Clutterbuck Ally M, Dwivedi Y (2020) The state of play of blockchain technology in the financial services sector: a systematic literature review
7. Chai H, Leng S, Chen Y, Zhang K (2021) A hierarchical blockchain-enabled federated learning algorithm for knowledge sharing in internet of vehicles
8. Chen Y, Qin X, Wang J, Yu C, Gao W (2020) FedHealth: a federated transfer learning framework for wearable healthcare
9. Asad M, Moustafa A, Ito T (2020) FedOpt: towards communication efficiency and privacy preservation in federated learning
10. Cheng Y, Liu Y, Chen T, Yang Q (2020) Federated learning for privacy-preserving AI
11. Cong Xie IG, Sanmi K (2019) Asynchronous federated optimization
12. Batool Z, Zhang K, Toews M (2022) Fl-mab: client selection and monetization for blockchain-based federated learning
13. Cui L, Su X, Ming Z, Chen Z, Yang S, Zhou Y, Xiao W (2021) Creat: blockchain-assisted compression algorithm of federated learning for content caching in edge computing
14. Desai HB, Ozdayi MS, Kantarcioglu M (2021) Blockfla: accountable federated learning via hybrid blockchain architecture
15. Fabasoft (2021) Digital contract management made easy
16. Feng L, Zhao Y, Guo S, Qiu X, Li W, Yu P (2021) Blockchain-based asynchronous federated learning for internet of things