

Secure and scalable federated learning with blockchain consensus and incentive mechanisms

Santhosh Kumar Gorva *

Department of Computer Science & Engineering (Data Science)

School of Engineering

Dayananda Sagar University

Bangalore 562112

Karnataka

India

Karuna S. Bhosale [†]

Department of Computer Science & Engineering

Pimpri Chinchwad University

Pune 412106

Maharashtra

India

J. Sebastian Nixon [§]

Department of Computer Science & Engineering

School of Engineering

Dayananda Sagar University

Bangalore 562112

Karnataka

India

S. N. Prajwalasimha [‡]

Dilip Kumar Jang Bahadur Saini [@]

Department of Computer Science & Engineering (Cyber Security)

School of Engineering

Dayananda Sagar University

Bangalore 562112

Karnataka

India

* E-mail: drsanthoshg1983@gmail.com (Corresponding Author)

[†] E-mail: karuna.bhosale@pcu.edu.in

[§] E-mail: dr.nixon14@gmail.com

[‡] E-mail: prajwalasimha.sn1@gmail.com

[@] E-mail: dilipsaini@gmail.com

Vijaya Kumbhar [#]

Department of Computer Science & Engineering

School of Computer Studies

Sri Balaji University, Pune

Pune 411033

Maharashtra

India

Abstract

Introduction Federated Learning (FL) empowers disseminated clients to collaboratively train models, while maintaining the privacy of their data. Yet, several security problems including malicious client behavior, poisoned data and model integrity compromise persist in classical FL systems. To this end, in this article, we propose Blockchain-enabled Federated Learning (BFL) system where we can deploy decentralized ledger technology for the tamper-evident model aggregation, transparent client contribution record and verifiable update validation. The designed BFL system uses smart contracts to enforce trust automatically without a need of intermediate party, Zero-Knowledge Proof (ZKPs) technique to keep the identity private with whom participates in activities, and tokenization incentive strategy that improve the reliability of proofs provided. It also presents a tunable blockchain overhead adaptive consensus algorithm to address scalability in various deployment environments including cybersecurity, healthcare, and Industrial Internet of Things (IIoT). Experimental results on benchmark datasets validate the efficiency of the proposed BFL framework by increasing 15-25% resistance to data poisoning and adversarial attacks associated to the state-of-the-art FL methods. The results also place BFL as a secure, transparent and scalable base to the future decentralized intelligent systems.

Subject Classification: *Primary 68M25, Secondary 68M14.*

Keywords: *Blockchain-enabled AI, Privacy-preserving machine learning, Cybersecurity in AI, Industrial IoT (IIoT) security, Decentralized federated learning.*

1. Introduction

Artificial Intelligence (AI) and Machine Learning (ML) are embryonic swiftly, leading to groundbreaking advances across various industries. But all these domains, and others such as health, finance, cyber security and smart cities are all relying on traditional centralized ML methods to collect data in one location. This really raises serious issues in the realm of privacy, ownership and regulatory compliance [1]. FL is gaining presence as a privacy-preserving solution for enabling training of models by distributed clients without sharing raw data, thus decreasing the exposure of sensitive information and adhering to GDPR and HIPAA standards [2].

However, FL still has its vulnerabilities to some security threats like model poisoning, free-riding and inference attack that may introduce the global model

[#] E-mail: veejeya.kumbhar@gmail.com

trust and integrity risk [3]. Furthermore, the absence of trusted coordinator leads to less responsibility accountability and transparency of client's contributions.

This is an optimistic baseline with blockchain; such as a decentralized, unchangeable and transparent ledger to cure these problems [4]. With blockchain support in FL, BFL can engage securely with aggregation, auditable client's behavior and trustless collaboration. Furthermore, both validation and incentive mechanisms can be automated using smart contracts. ZKPs could increase privacy by consenting clients to verify the soundness of appraises deprived of being forced to disclose valuable information [5].

Nevertheless, these advanced BFL frameworks are also limited by scalability, latency and efficiency issues aroused by traditional consensus scheme and large amount of communication overhead [6]. Moreover, finding the tradeoff between transparency and privacy remains a crucial challenge which needs more developed cryptographic algorithms like Homomorphic Encryption (HE), Differential Privacy (DP) etc.

To meet these challenges, we need to design a novel Blockchain-Enabled Federated Learning (BFL) scheme for secure, efficient and privacy-preserving the model aggregation. The main points are condensed here.

- A blockchain-secured FL framework that offers model integrity, verifiable updates, and trustless collaboration.
- An adaptive consensus protocol which maximizes blockchain efficiency with strong security.
- A ZKP-based privacy-preserving authentication scheme that guarantees data integrity without revealing private information.
- An incentive-based reputation mechanism to foster honest client behavior.

Experimental results on various metrics illustrating 15–25% enhanced attack resilience, 30% lower communication overheads and better model convergence as compared to baseline FL systems.

2. Literature Survey

Nowadays, FL has become a new decentralized paradigm which allows the collective training of models deprived of sharing raw data and thus ensures that privacy and regulations are well preserved. Despite these benefits of FL over centralized Machine Learning (ML), the paradigm remains to suffer from several challenges in terms of data leakage, and adversarial poisoning as well as trust among participating clients. Classical security measures such as HE and SMPC retain the privacy but bring huge computational overhead making them impractical on resource restricted environments [7].

2.1 Blockchain for Secure and Trustworthy FL

Blockchain offers a tamper-proof infrastructure-level decentralized architecture and can successfully solve trust and integrity problems in FL [8].

Blockchain-based Federated Learning (BFL) leverages immutable ledgers for secure model aggregation and transparent auditability, while smart contracts automate trust management and client validation processes [9]. However, a majority of BFL state-of-the-art frameworks suffer from suboptimal scalability and latency because model updates are frequent and the consensus process is computationally expensive [10].

2.2 Consensus Protocols for BFL

Consensus methods show a vital part in maintaining the reliability of blockchain networks used for FL. Traditional approaches include Proof-of-Work (PoW) and Proof-of-Stake (PoS), both of which are computationally expensive and unsuitable for real-time updates of the global model [17]. On the other hand, PoA, DPoS, and FBA are lightweight alternatives that have been explored to reduce latency and increase scalability [11-12]. However, how to achieve the best trade-off among decentralization, efficiency, and security remains an open challenge.

2.3 Privacy-Preserving Techniques in BFL

While blockchain increases transparency and accountability, it can compromise metadata and transaction data. Accordingly, a number of privacy-preserving cryptographic techniques have been integrated into the BFL systems, including ZKPs for verifiable privacy, DP for noise-based protection, and HE for secure computation on encrypted data, as summarized in [13-14]. While this is indeed an effective approach toward better confidentiality, it generally introduces extra computational and communication overheads.

2.4 Incentive Mechanisms in BFL

In BFL, client participation and fairness are often compromised due to the free-rider problem, wherein clients benefit without contributing any meaningful updates [17]. To solve this problem, tokenized and reputation-based incentive models on top of blockchains have been proposed that reward honest contributions and punish malicious ones [15-17]. Yet scalable and fair incentive mechanisms are still an open research issue, as a result of the heterogeneity among clients.

2.5 Gaps in Research and Motivation

In the existing literature there are many limitations of BFL, such as expensive computing costs and transaction costs, inefficient consensus algorithms, privacy and transparency tradeoffs, and an unfinished system of rewards for certain exchanges. To remedy these gaps, this article presents a conceptually optimized BFL framework comprising:

- An adaptive consensus protocol to minimize blockchain overhead;
- ZKP-based privacy-preserving authentication;
- An incentive-driven reputation model for honest participation; and

- Lightweight cryptographic primitives for efficient security.

These, together, are contributions towards scalable, secure, and privacy-preserving federated learning, enabling the adoption of robust blockchain-enabled AI in healthcare, IIoT, and cybersecurity domains.

4. Proposed Architecture

The proposed framework integrates FL with Blockchain to allow decentralized, privacy-preserving, and secure AI model training. It consists of the following key components:

- Clients (C): Devices that train local models using private datasets.
- Federated Server (S): Coordinates global model aggregation without accessing raw data.
- Blockchain Network (B): Stores encrypted model updates and ensures secure verification using smart contracts.
- Consensus Mechanism (M): Validates client contributions and resists adversarial attacks.

A. Mathematical Representation of Federated Learning:

Given a set of N clients $C = \{C_1, C_2, \dots, C_N\}$, each client C_i has a private dataset (Eq. 1)

$$D_i = \{(x_i, y_i)\}_{j=1}^{m_i} \quad (1)$$

Where,

- x_j is the input feature and y_j is the corresponding label.

The goal is to minimize a loss function $L(w)$ using a federated optimization framework (Eq. 2):

$$\min_w \sum_{i=1}^N p_i L_i(w) \quad (2)$$

Where,

- w is the global model parameter,
- $p_i = \frac{m_i}{\sum_{i=1}^N m_i}$ is the data proportion of client C_i ,
- $L_i(w) = \frac{1}{m_i} \sum_{j=1}^{m_i} l(w, x_j, y_j)$ is the local loss function for client C_i ,
- $l(\cdot)$ is the chosen loss function (e.g., cross-entropy for classification).

Each client updates the local model $w_i^{(t)}$ at round t using Stochastic Gradient Descent (SGD) (Eq. 3):

$$w_i^{(t+1)} = w_i^{(t)} - \eta \nabla L_i(w_i^{(t)}) \quad (3)$$

Where,

- η is the learning rate.

The server aggregates updates using Federated Averaging (FedAvg) (Eq. 4):

$$w^{(t+1)} = \sum_{i=1}^N p_i w_i^{(t+1)} \quad (4)$$

This prevents raw data from being shared, preserving client privacy.

B. Blockchain Integration for Model Security

To ensure tamper-proof, verifiable model updates, we integrate a private blockchain with FL. Each model update is logged as a transaction on the blockchain ledger.

1. *Model Update Encryption*
2. *Blockchain Transaction Verification*

Each client encrypts model updates before submission using Homomorphic Encryption (HE)(Eq. 5):

$$E(w_i^{(t)}) = w_i^{(t)} + r_i \quad (5)$$

Where,

- $E(.)$ is the encryption function,
- r_i is a random noise term to enhance privacy, ensuring Differential Privacy (DP).

To prevent poisoned updates, the blockchain network enforces a zero-knowledge proof (ZKP) verification before adding transactions (Eq. 6):

$$ZK(E(w_i(t))) \rightarrow \{0,1\} \quad (6)$$

Where,

- If $ZK = 1$, the update is valid and added to the blockchain.
- If $ZK = 0$, the update is rejected.

Each verified transaction is appended to the blockchain, ensuring immutable, tamper-resistant model updates.

C. Lightweight Consensus Mechanism

Instead of traditional PoW or PoS, we implement an optimized Proof-of-Contribution (PoC) consensus to validate model updates efficiently.

Each client C_i receives a trust score $T_i^{(t)}$ (Eq. 7), updated at every round t as:

$$T_i^{(t+1)} = \alpha T_i^{(t)} + (1 - \alpha) S_i^{(t)} \quad (7)$$

Where,

- $S_i^{(t)}$ is the model similarity score (computed using cosine similarity between local and global models).
- α is a decay factor ($0 < \alpha < 10$) to retain historical trust.

Clients with higher $T_i^{(t)}$ are prioritized for participation and incentivized accordingly.

D. Privacy-Preserving Mechanism

To inspire truthful contribution and daunt adversarial behavior, a tokenized incentive mechanism is integrated. Clients earn tokens based on model contribution quality (Eq. 8), defined by:

$$R_i^{(t)} = \lambda_1 T_i^{(t)} + \lambda_2 (1 - \|\nabla L_i(w_i^{(t)})\|) \quad (8)$$

Where,

- $R_i^{(t)}$ is the reward token allocation for client C_i ,
- λ_1 and λ_2 are weighting factors ensuring fairness,
- $\|\nabla L_i\|$ penalizes underperforming clients.

Clients with low contribution scores ($S_i^{(t)} < \theta$) are penalized using stake reduction, discouraging free-riders.

5. Results And Discussion

A complete analysis has been carried out to assess the proposed BFL framework capacity against this type of attack with the same data distribution and training tasks on CIFAR-10, MNIST, IoT (IDS) datasets using TFF 40 clients connected through an ethereum private blockchain. The experimental results demonstrate that BFL consistently outperforms the benchmarks including multiple versions of FedAvg, FedDP and BFL-NT across all performance metrics, thus confirming its scalability, privacy preserving efficacy and model robustness. In particular, their learned BFL obtained 94.3% on CIFAR-10 and 99.2% on MNIST. This also shows 3.2-2.1% improvement over FedAvg and 1.8-0.6% better than BFL-NT. The converged iteration of the proposed BFL model was 85 which is lower than that by FedAvg, FedDP and BFL-NT (120, 110 and 105 rounds respectively). For this reason, the global model stabilization was earlier in BFL as a result of well-tuned PoC consensus mechanism and efficient communication scheduling. Besides, thanks to blockchain-based aggregation and lightweight consensus, the proposed architecture saved around 27% communication overhead and 19% computing cost. Privacy protection has been dramatically improved (52% privacy leakage reduction compared to the baseline FL) owing to the joint application of HE, DP and ZKP. On robustness, the proposed BFL defended against 92.8% model-poisoning defenses better than BFL-NT (by 85.6%), FedDP (by 74.2%), and FedAvg (by 68.5%). Furthermore, BFL achieved 35% higher energy efficiency and 23% lower latency of FL as one can benefit from adaptive client selection, model compression and decentralized aggregation strategies. All of these findings verify that the new BFL approach is a secure and privacy-preserving, energy-efficient FL scheme for IoT, healthcare, finance, and smart cities with scalable and tamper-proof AI model training. Table 1 shows the days needed to train DNN models in different FL settings, and demonstrates the superiority of our proposed Blockchain-enabled Federated Learning (BFL) framework from other baseline approaches no matter what metrics are employed. Extensive experimental results on frequently used datasets show that it can achieve substantial improvement in accuracy, privacy protection,

energy efficiency and robustness, indicating the scalability and security of our proposed approach.

Table 1
Comparative Performance of Federated Learning Models

Metric	FedAvg	FedDP	BFL-NT	Proposed BFL	Improvement over FedAvg
Accuracy (CIFAR-10 / MNIST)	91.1 / 98.3	92.4 / 98.6	93.0 / 99.0	94.3 / 99.2	↑ 3.2 % / ↑ 0.9 %
Convergence Rounds	120	110	105	85	↓ 29 %
Communication Overhead Reduction	—	12%	18%	27%	↑ 27 %
Computational Cost Reduction	—	9%	15%	19%	↑ 19 %
Privacy Leakage Reduction	—	35%	45%	52%	↑ 52 %
Model-Poisoning Attack Robustness	68.50%	74.20%	85.60%	92.80%	↑ 24.3 %
Energy Efficiency Gain	—	10%	15%	35%	↑ 35 %
Latency Reduction	—	10%	18%	23%	↑ 23 %

6. Conclusion

A Blockchain-based Federated Learning (BFL) framework is presented for privacy-preserving, secure, and scalable AI model training across IoT, healthcare, financial, and industrial sectors. The BFL model outperformed (FedAvg, FedDP) and the non-thresholding version of our method on the CIFAR-10 dataset with an accuracy of 94.3%. Moreover, BFL achieved 30% faster convergence, up to 52% less privacy leakage and ensured 92.8% resilience against model poisoning attacks. In addition, it decreases communication overhead by 27% and reduces computation cost by 19%, while improving energy efficiency by up to 35%, through the integration of Proof-of-Contribution (PoC) consensus and privacy-preserving cryptographic schemes like HE, DP, and ZKP. This demonstrates that BFL offers a tamper-resistant, decentralized and energy-efficient framework for secure AI training. Future work will explore in-depth off-chain scalability, lightweight cryptography optimization, and post-quantum security for full security and efficiency improvement in blockchain-enabled FL systems.

References

- [1] S. Pérez Arteaga, A. L. Sandoval Orozco, and L. J. García Villalba, "Analysis of machine learning techniques for information

- classification in mobile applications," *Appl. Sci.*, vol. 13, no. 9, Art. no. 5438 (2023).
- [2] A. Roess, "The promise, growth, and reality of mobile health—Another data-free zone," *N. Engl. J. Med.*, vol. 377, no. 21, pp. 2010–2011 (2017).
 - [3] X. Yao, T. Huang, C. Wu, R. Zhang, and L. Sun, "Towards faster and better federated learning: A feature fusion approach," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Taipei, Taiwan, pp. 175–179 (2019).
 - [4] H. Tuttle, "Facebook scandal raises data privacy concerns," *Risk Manag.*, vol. 65, pp. 6–9 (2018).
 - [5] M. H. ur Rehman, K. Salah, E. Damiani, and D. Svetinovic, "Towards blockchain-based reputation-aware federated learning," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Virtual Conference, pp. 183–188 (2020).
 - [6] P. Mach and Z. Becvar, "Mobile edge computing: A survey on architecture and computation offloading," *IEEE Commun. Surv. Tut.*, vol. 19, no. 3, pp. 1628–1656 (2017).
 - [7] A. Yousefpour, C. Fung, T. Nguyen, K. Kadiyala, F. Jalali, A. Niakanlahiji, J. Kong, and J. P. Jue, "All one needs to know about fog computing and related edge computing paradigms: A complete survey," *J. Syst. Archit.*, vol. 98, pp. 289–330 (2019).
 - [8] S. M. Halim, L. Khan, and B. Thuraisingham, "Next-location prediction using federated learning on a blockchain," in *Proc. IEEE 2nd Int. Conf. Cogn. Mach. Intell. (CogMI)*, Atlanta, GA, USA, pp. 244–250 (2020).
 - [9] M. Ahmadzai and G. Nguyen, "Federated learning with differential privacy on personal opinions: A privacy-preserving approach," *Procedia Comput. Sci.*, vol. 225, pp. 543–552 (2023).
 - [10] C. D. Pop, M. Antal, T. Cioara, I. Anghel, and I. Salomie, "Blockchain and demand response: Zero-knowledge proofs for energy transactions privacy," *Sensors*, vol. 20, Art. no. 5678 (2020).
 - [11] C. Bryant, W. Carvalho, N. Baracaldo, H. Ludwig, B. Edwards, T. Lee, I. Molloy, and B. Srivastava, "Detecting backdoor attacks on

- deep neural networks by activation clustering,” arXiv preprint arXiv:1811.03728 (2018).
- [12] Y. Yao, H. Li, H. Zheng, and B. Y. Zhao, “Latent backdoor attacks on deep neural networks,” in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur. (CCS)*, London, U.K., pp. 2041–2054 (2019).
 - [13] Y. Cheng, Y. Liu, T. Chen, and Q. Yang, “Federated learning for privacy-preserving AI,” *Commun. ACM*, vol. 63, no. 12, pp. 33–36 (2020).
 - [14] N. Bouacida and P. Mohapatra, “Vulnerabilities in federated learning,” *IEEE Access*, vol. 9, pp. 63229–63249 (2021).
 - [15] V. Tolpegin, S. Truex, M. E. Gursoy, and L. Liu, “Data poisoning attacks against federated learning systems,” in *Computer Security—ESORICS 2020*, Darmstadt, Germany, pp. 480–501 (2020).
 - [16] A. Noonina, D. Thakral, P. Mathur, F. Sheth, H. Shaikh, and A. K. Gupta, “A discrete mathematical model and cryptography for secure medical image analysis: Encrypted chest X-ray classification,” *J. Discrete Math. Sci. Cryptogr.*, vol. 28, no. 5-A, pp. 1473–1486 (2025).
 - [17] R. Joshi, P. Mathur, A. K. Gupta, S. Singh, V. Paliwal, and S. Nayar, “Mathematical modeling of intelligent system for predicting effectiveness of premenstrual syndrome,” *J. Interdiscip. Math.*, vol. 26, no. 3, pp. 551–562 (2023).

Received September, 2025