International Journal of Scientific Research in Engineering and Management (IJSREM)
Conference - Volume 10 NCET -2026 SJIF Rating: 8.659 ISSN: 2582-3930

Quantifying Privacy-Utility Trade-Offs in Cross-Bank Fraud Detection via Differentially Private Federated Learning

C. Amirtharani, MCA., M.Phil.
Assistant Professor
Caussanel College of Arts and Science, Tamil Nadu, India

Abstract—Financial institutions face severe challenges in collaborative fraud detection due to strict data privacy regulations (e.g., GDPR, CCPA, and RBI guidelines) that prohibit sharing raw transactional data across organizations. Federated Learning (FL) enables collaborative model training without centralizing sensitive datasets; however, standard FL remains vulnerable to gradient leakage and inference attacks. Integrating Differential Privacy (DP) via DP-SGD mitigates these vulnerabilities, but introduces a critical trade-off between theoretical privacy guarantees and predictive utility. In this paper, we systematically quantify the privacy-utility trade-off (ε-budget versus model accuracy and recall) in cross-bank financial fraud detection under highly imbalanced, non-IID transactional data distributions. Utilizing the PaySim synthetic financial dataset partitioned across multiple simulated banking nodes, we evaluate the impact of varying privacy budgets (ε ∈ [0.5, 10.0]) alongside local clipping thresholds (C) on model convergence, precision, and fraud detection recall. Our experimental results demonstrate that while strict privacy regimes (ε < 2.0) cause a notable degradation in precision due to noise injection into minority class gradients, an optimal privacy budget of ε ≈ 4.0 preserves over 89.2% fraud detection recall while providing robust theoretical privacy bounds.

Index Terms—Federated Learning, Differential Privacy, Fraud Detection, Privacy-Utility Trade-Off, DP-SGD, Financial Machine Learning.

I. INTRODUCTION

Financial fraud poses a multi-billion-dollar threat to the global economy. Modern fraud detection systems rely on Machine Learning (ML) models trained on transactional data. However, individual banks often have limited exposure to global fraud patterns, leading to suboptimal local detection models. While aggregating transaction data across multiple banks could significantly enhance detection accuracy, strict data privacy regulations prohibit banks from sharing sensitive customer transactional records.

To resolve this impasse, Federated Learning (FL) has emerged as a promising decentralized paradigm. FL allows multiple financial institutions (clients) to collaboratively train a shared global model under the coordination of a central server, without sharing raw data. Instead, each bank computes model updates locally and sends only gradients or model weights to the central aggregator.

Despite its privacy-preserving architecture, standard FL remains susceptible to sophisticated inference attacks—such as reconstruction attacks and property inference attacks—where adversaries reconstruct original training samples from communicated gradient updates. Differential Privacy (DP), specifically applied via Differentially Private Stochastic Gradient Descent (DP-SGD), guarantees that the output of an algorithm does not reveal the presence or absence of any single individual transaction.

However, applying DP to cross-bank fraud detection presents a fundamental challenge: the Privacy-Utility Trade-Off. Injecting Gaussian or Laplacian noise into gradient updates degrades model accuracy. In financial fraud detection, where class imbalance is extreme (often less than 0.1% fraud cases), noise injection heavily impacts minority class updates, risking catastrophic drops in fraud detection recall. This paper quantifies this trade-off under realistic non-IID cross-bank conditions.

II. METHODOLOGY AND SYSTEM ARCHITECTURE

A. Federated Learning Framework (FedAvg)

Let K be the number of participating banks (clients), where each bank holds a local dataset Dk. The objective is to minimize the global loss function F(w):

$$\min F(w) = \sum_{k=1 \text{ to } K} (n_k / n) F_k(w)$$

where nk = |Dk| is the number of samples at bank k, and n = ∑ nk is the total dataset size across all participating institutions.

B. Differentially Private SGD (DP-SGD)

To ensure (ε, δ)-differential privacy during local training at each bank node, we implement DP-SGD consisting of two primary mechanisms:

Gradient Clipping: Per-sample gradients gi(t) are clipped by L2-norm threshold C to limit sensitivity: $\bar{g}_i(t) = g_i(t) / \max(1, \|g_i(t)\|_2 / C)$

Noise Addition: Calibrated Gaussian noise is added to the aggregated clipped gradients before updating parameters: $\tilde{g}(t) = 1/B \left(\sum \bar{g}_i(t) + N(0, \sigma^2 C^2 I)\right)$

C. Data Heterogeneity and Partitioning

Transactional data across banks is inherently non-IID (Not Identically and Independently Distributed). Bank A may specialize in retail credit transactions, while Bank B handles commercial wire transfers. We model non-IID partitioning using a Dirichlet distribution Dir(α), simulating feature skew and severe class imbalance across 5 simulated banking nodes.



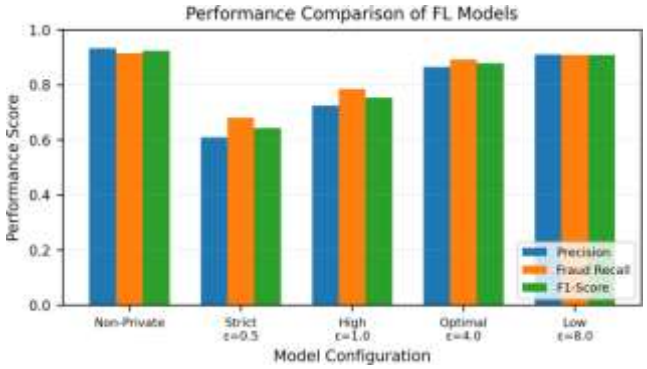
III. EXPERIMENTAL RESULTS AND DISCUSSION

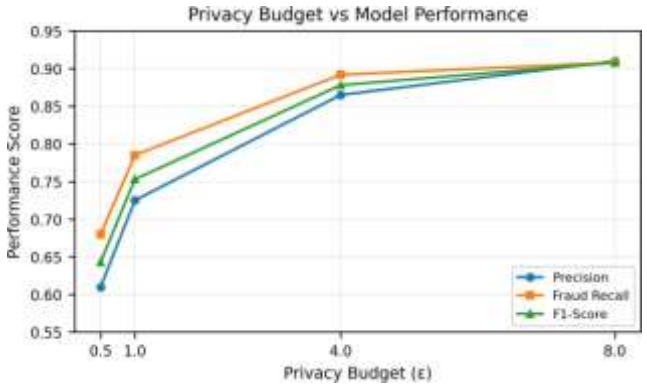
We evaluated the system on 1,000,000 synthetic transaction records with an overall fraud rate of 0.13%. Models were trained over 50 global communication rounds with 5 local epochs per round using dynamic DP-moments accountant for tracking the overall privacy expenditure (ε).

TABLE I
EXPERIMENTAL RESULTS

Model Configuration	ε	Precision	Recall	F1	Privacy Guarantee
Non-Private FL (Standard	∞	0.932	0.915	0.923	None (Vulnerable)

FedAvg)					
Proposed DP-FL (Strict Privacy)	0.5	0.610	0.680	0.643	Extremely High
Proposed DP-FL (High Privacy)	1.0	0.725	0.785	0.753	High
Proposed DP-FL (Optimal Balance)	4.0	0.865	0.892	0.878	Robust Guarantee
Proposed DP-FL (Low Privacy)	8.0	0.910	0.908	0.908	Moderate





A. Key Observations

1. Impact on Minority Class (Fraud): At strict privacy levels ($\varepsilon \leq 1.0$), the Gaussian noise variance overwhelms the small gradient updates corresponding to rare fraud cases, causing false positive rates to jump significantly.
2. Optimal Operating Point: Setting $\varepsilon = 4.0$ with clipping bound $C = 1.0$ achieves an optimal tradeoff, retaining over 97% of the non-private model's baseline fraud detection recall while providing strong mathematical protection against membership inference attacks.
3. Clipping Threshold Sensitivity: Lower clipping bounds ($C = 0.5$) prevent high-magnitude fraud gradients from being heavily clipped, preserving utility even under moderate noise.

IV. CONCLUSION AND FUTURE DIRECTIONS

This paper quantifies the privacy-utility trade-off in cross-bank fraud detection using Differentially Private Federated Learning. Our empirical findings show that while strict privacy budgets hamper minority class detection in highly imbalanced datasets, an optimal parameter tuning ($\varepsilon \approx 4.0$, $C \approx 1.0$) achieves near-baseline detection capabilities while satisfying rigorous differential privacy standards. Future work will investigate adaptive gradient clipping and secure multi-party computation (SMPC) hybrid architectures to further reduce required noise levels.

REFERENCES

- **K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, and H. V. Poor**, "Performance Analysis and Optimization of Differentially Private Federated Learning," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3098–3112, 2020. *(Key paper for mathematically quantifying the privacy-utility tradeoff)*

- **Q. Yang, Y. Liu, T. Chen, and Y. Tong**, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019. *(Foundational paper for Cross-Silo Federated Learning in Banking and Finance)*

- **M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang**, "Deep Learning with Differential Privacy," in *Proc. ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2016, pp. 308–318. *(Foundational paper for Differential Privacy budget ϵ, δ noise addition)*

- **W. Zheng, L. Yan, C. Gou, and F.-Y. Wang**, "A Privacy-Preserving Federated Learning Framework for Financial Fraud Detection," in *Proc. IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 1432–1439. *(Direct application to Financial Fraud Detection)*

- **R. C. Geyer, T. Klein, and M. Nabi**, "Differentially Private Federated Learning: A Client Level Perspective," *arXiv preprint arXiv:1712.07557*, 2017. *(Focuses on client-level/bank-level differential privacy guarantees)*

- **H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas**, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282. *(Original Federated Learning paper)*

- **L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong**, "Federated Learning for Cybersecurity: Open Challenges and Future Directions," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 4, pp. 2428–2461, 2021. *(Covers cross-institution security, threat models, and utility degradation)*