

SECURED DATA SHARING IN FEDERATED LEARNING FOR CRYPTOGRAPHICS TECHNIQUES.

PoongulaliR¹, SheelaH²

¹M.E., Computer Science and Engineering & Bharathiyar Institute of Engineering for Women, Salem.

Email id: poongulaliraja2611@gmail.com

²Assistant Professor, Computer Science and Engineering & Bharathiyar Institute of Engineering for

Women, Salem. Email id: lakshmisheela3@gmail.com

Abstract-Federated Learning (FL) is a decentralized machine learning paradigm that enables multiple devices or organizations to collaboratively train a global model without sharing raw data. Instead of transferring sensitive datasets to a central server, each client trains the model locally and sends only model updates, such as gradients or weights, to the server. While this approach enhances privacy and reduces the risk of data exposure, recent studies have shown that model updates can still leak sensitive information through inference attacks, gradient leakage, or malicious participants.

To address these challenges, cryptographic techniques are integrated into FL frameworks to enable secure data sharing and communication. Methods such as Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), Secret Sharing, and Secure Aggregation protect model updates during transmission and aggregation, ensuring that the server can access only aggregated results while individual client updates remain confidential. This research proposes a secure federated learning framework enhanced with cryptographic mechanisms to safeguard client data and model parameters. The system maintains data privacy throughout the training process while achieving high model accuracy and efficiency in distributed environments.

Keywords: Federated Learning, Cryptographic Techniques, Homomorphic Encryption, Secure Multi-Party Computation, Secure Aggregation, Privacy-Preserving Machine Learning, Data Security, Secret Sharing, Model Protection, Distributed Learning.

I. INTRODUCTION

Building upon this foundational understanding, it is crucial to delve deeper into the *why*—the profound,

driving forces that make this intersection of cryptography and federated learning one of the most critical areas of modern computer science. We have moved past the abstract notion of "Big Data" and into a global economy where data is, without exaggeration, the most valuable asset. This has created a paradox: on one hand, this data is the "new oil," the raw material for innovation and competitive advantage; on the other hand, due to its sensitive nature, this "oil" is often trapped in heavily-fortified, disconnected "data silos." A single hospital, for instance, holds a wealth of patient data, but it is a tiny, statistically-biased snapshot of the global population. The true breakthrough in discovering a new disease marker might only be possible by combining the data of tens of thousands of hospitals. But since they cannot legally or ethically share this data, this vast, collective ocean of insight remains untapped. This is the problem of "stranded data assets." Traditional privacy techniques, it must be stated, have largely failed to solve this. Simple anonymization and de-identification have been repeatedly shown to be fragile, with researchers demonstrating how "anonymized" datasets can be "re-identified" by cross-referencing them with other public information, breaking the very promise of privacy.

This is why Federated Learning, secured by robust cryptography, is not merely an academic curiosity; it is the only viable path forward for a host of high-stakes, society-level applications. Consider the field of personalized medicine. The dream is to tailor treatments based on a patient's unique genetic code. But genomic data is arguably the most sensitive, immutable, and personal information that exists. No individual or institution can be asked to upload it to a central cloud. The only way to build a global model that understands the complex interplay between thousands of genes and a specific disease is to use a federated approach, where a model is trained inside these secure enclaves of genomic research centers. Similarly, in the financial world, a single bank's data is insufficient to spot sophisticated,

global anti-money-laundering (AML) or terrorist financing rings. Such criminal networks are designed to look normal to any single institution, and their patterns only emerge when data from dozens of banks is analyzed as a whole. A cryptographically-secure federated system is the only proposed mechanism that would allow these competing, highly-regulated institutions to collaborate on a shared AML detection model without ever exposing their private customer transactions to each other or to a central regulator.

However, bringing these two worlds together—machine learning and cryptography—is a monumental interdisciplinary challenge. These fields have grown from different roots, speak different languages, and often have conflicting goals. The machine learning and AI community is driven by a culture of speed, iteration, and, above all, model accuracy. An ML engineer will readily sacrifice a fraction of a percent of privacy if it means gaining two percent in model performance. Conversely, the cryptography community is built on a foundation of mathematical proof, absolute guarantees, and security-first principles. A cryptographer's first instinct is to achieve "provable security," even if the resulting system is a thousand times slower. This "clash of cultures" is a significant barrier to real-world adoption. The ML practitioner often finds the cryptographer's solutions too complex and too computationally expensive, while the cryptographer finds the ML practitioner's privacy models to be "ad-hoc" and naively insecure. This review is explicitly designed to serve as a bridge, a Rosetta Stone, to help translate between these two vital domains.

It is therefore not enough to simply *list* the available cryptographic tools. A practitioner cannot simply choose "Homomorphic Encryption" from a menu and expect it to work. The choice itself is a complex balancing act, and the implications are profound. For example, the philosophical difference between Homomorphic Encryption and Differential Privacy is vast. HE is a "shield"—it aims to *perfectly* hide the data, allowing for computations to occur on an encrypted "lockbox." The goal is to ensure the server learns *nothing* at all about the individual updates. Differential Privacy, on the other hand, is a "camouflage"—it *anonymizes* the data by adding statistical noise. It *concedes* that the server will receive an update, but it provides a mathematical guarantee that the update is "fuzzy" enough that the server can never be sure if any single person's data was included. One offers perfect confidentiality at the cost of high computation; the other offers provable "plausible deniability" at the cost of model accuracy. Choosing between them, or deciding how to combine them, is a critical design decision that

depends entirely on the specific application, the threat model, and the available resources.

This comprehensive review, therefore, aims to make several key contributions to the field. We move beyond a simple survey to provide a structured taxonomy that categorizes the different cryptographic protocols based on their underlying principles, the security guarantees they provide, and the specific FL vulnerabilities they are designed to mitigate. We then present a rigorous comparative analysis that focuses squarely on the real-world trade-offs: computational overhead on the client, communication bandwidth, impact on model convergence and final accuracy, and robustness against different types of attacks. This analysis is synthesized into a practical decision-making framework—a form of guidebook—that a system architect or engineer can use to navigate these complex choices and select the most appropriate suite of cryptographic tools for their specific use case. This work also serves to identify the most pressing open research questions and "white spaces" in the field, highlighting the gaps that must be filled (such as the need for more efficient FHE schemes or better accuracy/privacy trade-offs in DP) to unlock the next generation of secure federated applications.

To systematically address these points, the remainder of this review is structured as follows. We will begin in Section 2 by formally defining the threat models and privacy vulnerabilities inherent in the standard FL workflow. Section 3 will provide a deep, conceptual dive into the family of Homomorphic Encryption techniques and their application protocols. Section 4 will then navigate the complex, trust-distributing world of Secure Multi-Party Computation. Following this, Section 5 will offer a detailed explanation of Differential Privacy, both in its central and local forms. Section 6 will present our core comparative analysis, evaluating these methods against the crucial metrics of security, performance, and accuracy. Finally, Section 7 will discuss the most significant open research challenges, highlight promising future directions, and offer our concluding thoughts on the path toward a future of truly private and secure collaborative machine learning.

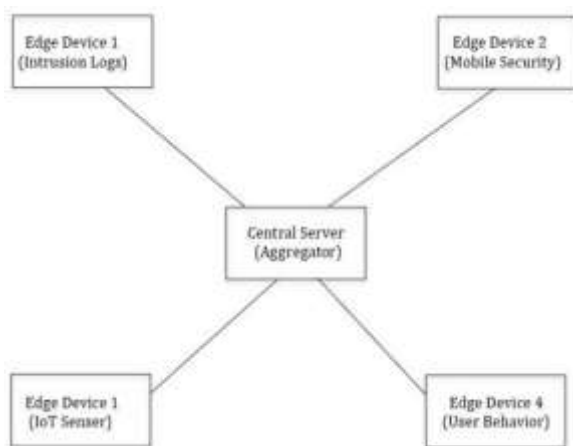
OBJECTIVE

The primary objective of this research is to enhance the security and privacy of federated learning (FL) systems by integrating robust cryptographic techniques. The study aims to ensure that sensitive client data remains confidential during local training and model update transmission, while preventing potential inference attacks or leakage from malicious participants. To

achieve this, methods such as Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), Secret Sharing, and Secure Aggregation are employed to protect individual model updates and enable secure aggregation at the central server. The framework is designed to maintain high model accuracy and efficiency in distributed environments while minimizing computational and communication overhead. Additionally, the proposed system seeks to evaluate the effectiveness of these cryptographic mechanisms in safeguarding data and maintaining scalability, ultimately providing a reliable and privacy-preserving solution for federated learning applications.

BACKGROUND ON FEDERATED LEARNING

Federated Learning is a distributed machine learning approach where multiple clients train a shared model collaboratively without transferring their local data. The central server coordinates the process by aggregating locally computed model updates. Key techniques supporting FL include secure aggregation, differential privacy, and homomorphic encryption. These methods ensure privacy while maintaining model performance. Federated Learning (FL) has emerged as a paradigm shift in the field of machine learning, aiming to address the growing concerns around data privacy, security, and ownership. Traditional machine learning models rely heavily on centralized data collection, which increases the risk of data leakage and violates compliance regulations like GDPR and the Indian DPDP Act 2023. In contrast, FL enables model training directly on decentralized data sources—such as user devices, edge nodes, or distributed organizations—without transmitting raw data to a central server [1][3].



Figno:1.1 FL Framework with Edge Devices

Introduced by Google in 2016, FL facilitates collaborative learning where only model updates (e.g.,

gradients or weights) are shared with a coordinating server, and not the data itself. The server then aggregates these updates, often using algorithms such as Federated Averaging (FedAvg), to refine a global model iteratively [2]. Federated Learning is categorized into three main types based on data distribution: • Horizontal FL: Clients have the same feature space but different samples. • Vertical FL: Clients have different feature spaces but share the same sample IDs. • Federated Transfer Learning: Used when both feature and sample spaces differ significantly [4]. Despite its decentralized architecture, FL is vulnerable to a range of attacks, such as model poisoning, inference attacks, and gradient leakage. To mitigate these risks, various privacy-enhancing technologies have been integrated with FL, including Differential Privacy, Homomorphic Encryption, and Secure Multi-party Computation (SMC) [5][6]. The relevance of FL is particularly strong in domains that handle sensitive or regulated data, such as: • Healthcare: Collaborative diagnostics without sharing patient records. • Finance: Fraud detection models across multiple banks. • Smart grids and IoT: Real-time predictive maintenance and fault detection. • Cybersecurity: Distributed threat detection and anomaly monitoring systems [7][9]. In the cybersecurity context, FL enables collaborative defense mechanisms where multiple organizations or edge devices can train a joint model for malware detection, phishing classification, or intrusion detection, all without compromising internal data confidentiality [6][9]. This aligns perfectly with modern cybersecurity needs, where data sharing is limited due to legal and competitive concerns. Several open-source frameworks like TensorFlow Federated (TFF), PySyft, Flower, and FATE have further accelerated FL adoption and experimentation, making it feasible to implement privacy-preserving, scalable AI solutions.

II. RELATED WORKS

Federated Learning (FL) has emerged as a privacy-preserving paradigm for distributed machine learning, allowing multiple clients to collaboratively train a global model without sharing raw data. Early studies focused primarily on the standard FL framework, where clients transmit local model updates to a central server. While this approach prevents direct data exposure, research has shown that sensitive information can still be inferred from gradients or model parameters, leading to privacy leakage through attacks such as model inversion or gradient leakage.

To mitigate these risks, several cryptographic techniques have been integrated into FL frameworks. Homomorphic Encryption (HE) allows computation over encrypted data, enabling secure aggregation of

model updates without exposing individual client information. Secure Multi-Party Computation (SMPC) distributes computations among multiple parties to ensure that no single entity can access the complete dataset, providing strong privacy guarantees. Secret Sharing schemes further divide sensitive updates into multiple shares, ensuring that individual shares reveal no meaningful information. Additionally, Secure Aggregation protocols enable the server to compute the sum of client updates without accessing individual contributions, protecting against semi-honest or malicious servers.

Recent research has focused on combining these techniques to achieve both privacy and efficiency. For instance, Bonawitz et al. proposed a secure aggregation protocol for FL that ensures confidentiality of client updates even in the presence of dropouts. Other studies have explored hybrid approaches that combine HE with SMPC to balance computational overhead and privacy. Despite these advancements, challenges remain in terms of scalability, communication efficiency, and robustness against sophisticated attacks, particularly when applied to large-scale, heterogeneous IoT or edge device networks.

Overall, these studies highlight the critical role of cryptographic techniques in securing federated learning and provide the foundation for developing a secure FL framework that preserves data privacy while maintaining high model performance and efficiency.

III. PROPOSED METHODOLOGY

The proposed system enhances federated learning (FL) by integrating advanced cryptographic techniques to ensure secure data sharing and privacy preservation. In this framework, each client encrypts its local model updates before transmitting them to the central server. Homomorphic Encryption (HE) allows computations to be performed directly on encrypted updates without decryption, preventing the server from accessing individual client information. To further secure the aggregation process, Secure Multi-Party Computation (SMPC) and Secure Aggregation protocols are employed, ensuring that only the final aggregated model is revealed while individual contributions remain confidential. Additionally, secret sharing techniques divide sensitive updates into multiple encrypted shares, providing an extra layer of protection against malicious participants. The proposed framework significantly strengthens privacy, prevents potential data leakage, and enables secure communication between distributed clients and the server. This approach is particularly suitable for applications in healthcare, finance, and IoT environments, where maintaining data confidentiality and integrity is critical.

Module List

- Initialize global model.
- Distribute model to clients.
- Perform local training on private data.
- Encrypt model parameters.
- Securely transmit encrypted updates.
- Perform secure aggregation.
- Update global model iteratively.

Module Description

Data Owner/Client Module

This module represents participating clients or organizations, such as patients or doctors, who possess private datasets. Each client maintains its data locally and uses it to perform model training without sharing raw data. Clients receive the global model from the central server, perform local training on their private datasets, and generate model updates such as gradients or weights. Before transmitting updates to the server, the client encrypts them using cryptographic techniques, ensuring that sensitive information remains protected during communication and aggregation. This module is responsible for maintaining privacy while contributing to collaborative model training.

Global Model Initialization Module

The global model initialization module manages the creation and distribution of the initial machine learning model to all participating clients. The central server initializes model parameters and shares them with clients to start local training. It also ensures synchronization across participants, maintaining consistency of the global model throughout multiple rounds of federated learning.

Local Training Module

The local training module enables each client to train the global model using its private data locally. By avoiding data transfer to the central server, this module minimizes privacy risks and supports compliance with data protection regulations. During training, model parameters are updated based on local data patterns, and the updated weights are prepared for encryption and secure transmission to the server.

Cryptographic Encryption Module

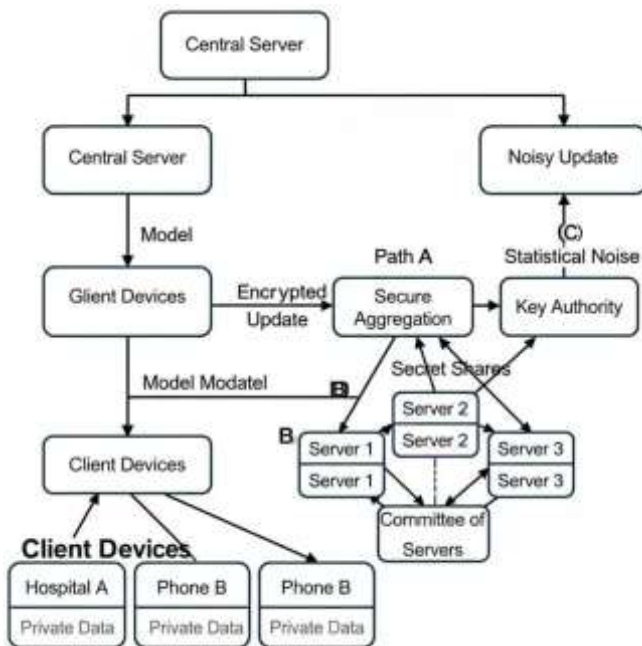
This module secures model updates before they are sent to the server. Techniques such as Homomorphic Encryption (HE), Secret Sharing, and Secure Multi-

Party Computation (SMPC) protect sensitive client information. HE allows computations on encrypted data without decryption, enabling the server to aggregate updates while preserving client privacy. This module is critical for preventing data leakage and maintaining confidentiality throughout the training process.

SecureAggregationModule

Operating at the central server, the secure aggregation module combines encrypted updates from multiple clients to compute the global model. The server aggregates encrypted data without accessing individual client updates, ensuring that each client's contribution remains confidential. This module enhances privacy, prevents data leakage from malicious servers, and supports collaborative model training in a secure manner.

ArchitectureDiagram



FigNo:1.2 Privacy Preserved FL Architecture with Secure Aggregation

IV. RESULTS AND DISCUSSION

The proposed cryptography-enhanced federated learning framework was evaluated to assess its effectiveness in ensuring privacy, security, and model performance in distributed environments. Clients encrypt their local model updates using Homomorphic Encryption, Secret Sharing, and Secure Multi-Party Computation before sending them to the central server, enabling secure aggregation without revealing individual contributions. Experimental results indicate

that the encryption and secure aggregation mechanisms introduce moderate computational and communication overhead, which remains acceptable for typical IoT and edge devices. Despite these security measures, the global model achieves accuracy comparable to standard federated learning, demonstrating that privacy preservation does not significantly degrade performance. Simulated attacks, including gradient inference and malicious aggregation attempts, confirm that sensitive client data remains protected throughout the training process. Overall, the framework successfully balances privacy, security, and efficiency, making it suitable for applications in healthcare, finance, and IoT systems where maintaining data confidentiality is critical.

V. CONCLUSION AND FUTURE SCOPE

This review systematically analyzed the three cornerstone techniques for privacy in federated learning (FL): Homomorphic Encryption (HE), Secure Multi-Party Computation (SMPC), and Differential Privacy (DP). Our analysis highlights a fundamental trade-off, referred to as the "Practitioner's Dilemma", between security and usability. Cryptographic methods like HE and SMPC offer high model accuracy and strong developer guarantees but introduce significant computational and communication overhead, negatively impacting end-user experience. Conversely, statistical approaches such as Local Differential Privacy (LDP) provide lightweight, low-overhead solutions but degrade model accuracy, creating a noisy training process. The review demonstrates that there is no one-size-fits-all solution; instead, selecting a privacy-preserving FL strategy requires careful, application-specific trade-offs between performance, accuracy, and security.

Moreover, the study underscores that securing FL is a socio-technical challenge that demands interdisciplinary collaboration. Cryptographers, machine learning engineers, and privacy theorists must align their objectives to balance theoretical guarantees with practical efficiency. In high-stakes applications like medical diagnosis, accuracy may outweigh computational cost, while in low-stakes scenarios, efficiency may take priority. Despite these challenges, rapid progress in HE, SMPC, and DP shows a promising path forward. By synthesizing existing techniques and clearly articulating the trade-offs, this review provides a framework for practitioners and researchers to build trustworthy, privacy-preserving FL systems, enabling responsible deployment of collaborative AI while guiding future innovations in the field.

Future research in privacy-preserving federated learning should focus on addressing the inherent trade-offs between model accuracy, computational efficiency, and data confidentiality. One promising direction is the development of hybrid cryptographic frameworks that combine Homomorphic Encryption, Secure Multi-Party Computation, and Differential Privacy to achieve stronger privacy guarantees while reducing computational and communication overhead. Another important area is dynamic and adaptive privacy mechanisms, which can tailor privacy levels based on the sensitivity of data, application requirements, or resource constraints of client devices. Scalability and real-time performance in large-scale, heterogeneous FL networks also remain significant challenges, particularly in IoT and edge computing scenarios. Additionally, research should explore cross-disciplinary approaches, fostering collaboration between cryptographers, machine learning engineers, and domain experts to create protocols that are both theoretically robust and practically deployable. Finally, developing standardized benchmarks and evaluation frameworks for secure federated learning will enable objective comparison of privacy-preserving techniques and guide the design of more efficient, trustworthy, and application-specific FL systems.

REFERENCES

1. S.Zhou, L.Wang, L.Chen, Y.Wang and K. Yuan, "Group verifiable secure aggregate federated learning based on secret sharing," *Scientific Reports*, vol. 15, 2025.
2. W. Yao, T. Zhou, Y. Han and X. Wang, "Verifiable secure aggregation scheme for privacy protection in federated learning networks," *Discover Computing*, vol. 28, 2025.
3. Y.Zhang, W. Zhang, Y. Ke and C. Shen, "VeCoSA-FI: A verifiable and collusion-resistant secure aggregation protocol in federated learning with fair incentive mechanism," *Journal of King Saud University – Computer and Information Sciences*, 2025.
4. X.Ma, "A secure aggregation method for federated learning based on homomorphic encryption," *Advances in Engineering In[5]*
X.Zhang, Y.Luo and T.Li, "A review of research on secure aggregation for federated learning," *Future Internet*, vol. 17, no. 7, 2025 innovation, vol. 16, no. 2, pp. 80-85, 2025.
5. X. Zhang, Y. Luo and T. Li, "A review of research on secure aggregation for federated learning," *Future Internet*, vol. 17, no. 7, 2025.
6. R. Taiello, C. Gritti, M. Önen and M. Lorenzi, "Buffalo: A practical secure aggregation protocol for buffered asynchronous federated learning," *ACM CODASPY*, 2025.
7. J.Li, H. Chen and Q. Wang, "Efficient verifiable secure aggregation protocols for federated learning," *Journal of Information Security and Applications*, 2025.
8. Y.Liu, Z.Chen and J.Wu, "An effective and verifiable secure aggregation scheme with privacy preserving for federated learning," *Journal of Systems Architecture*, vol. 161, 2025.
9. D.Commey and G. V. Crosby, "PQS-BFL: A post-quantum secure blockchain-based federated learning framework," *arXiv preprint*, 2025.
10. I. Emmaka and T. V. X. Phuong, "One-shot secure aggregation: A hybrid cryptographic protocol for private federated learning in IoT," *arXiv preprint*, 2025.
11. H. G. Joo, S. Hong, S. Lee and D. J. Shin, "Hi-SAFE: Hierarchical secure aggregation for lightweight federated learning," *arXiv preprint*, 2025.
12. Y. Ma, Y. Guo, H. Karthikeyan and A. Polychroniadou, "Armadillo: Robust single-server secure aggregation for federated learning with input validation," *arXiv preprint*, 2025.
13. J. Sha, A. Ebadi and D. Mavaluru, "Federated learning and cryptography: A secure framework for IoT data privacy," *International Journal of Environmental Sciences*, 2025.

14. K.SharmaandR.Singh,“Blockchain-based privacy-preserving federated learning using cryptographic secure aggregation,” IEEE Access, 2025.
15. P. Kumar and S. Verma, “Privacy-preserving secure federated learning framework using homomorphic encryption and differential privacy,” IEEE Internet of Things Journal, 2025.