Quantum-Aware Privacy Challenges in Differentially Private Federated Learning

Neha Sharma[1], Prasenjit Chatterjee[2]

[1,2] Lincoln University College, 47301, Petaling Jaya, Selangor Darul Ehsan, Malaysia

[1]nehasharma0110@gmail.com, [2]dr.prasenjitchatterjee6@gmail.com

Abstract: Federated learning (FL) enables multiple parties to collaboratively train models without sharing their raw data, making it suitable for environments where data confidentiality is important. Many FL systems rely on differential privacy (DP) to protect sensitive information, mainly focusing on balancing privacy protection and model performance. However, with the rapid growth of computational capabilities and increasingly sophisticated inference attacks, the long-term effectiveness of these approaches is becoming uncertain. This work reviews the development of privacy-related issues in DP-based FL from 2010 to 2026, focusing on three key areas: privacy protection mechanisms, information leakage caused by attacks, and the role of adversaries. A review of existing studies shows that these aspects have largely been addressed independently, while a comprehensive framework that jointly evaluates privacy, information leakage, and model performance is still lacking.

Keywords: Differential Privacy; Federated Learning; Privacy Leakage; Membership Inference Attacks; Privacy-Preserving Machine Learning; Secure Collaborative Analytics

1. Introduction

The proliferation of data-driven systems has fuelled the need to collaborate and share data between organizations and even countries. For instance, in healthcare, finance, cybersecurity, and climate science, to achieve some insights, data from various institutions has to be combined. Direct data sharing, however, faces several challenges, including privacy and security issues. Federated learning (FL) has been proposed to train models in a decentralized way, employing local data and model updates to aggregate [1]. However, FL has privacy issues, and several studies have demonstrated that attackers can use model updates to carry out several types of attacks, including membership inference, model inversion, and gradient leakage [2]. Hence, privacy in FL is a key area of research. Figure 1 depicts a traditional DP-enabled FL process. The clients train their local models using their private data and share their updates with a server. The private data is protected using gradient clipping and noise addition, but there is a possibility of information leakage from updates and models. Recently, differential privacy (DP) [3] was adopted as a formal approach to secure training data by adding noise to updates and gradients, limiting an individual's impact on the model. DP was combined with FL to prevent information leakage while maintaining performance. The existing approaches [4, 5] have some limitations. For example, some research works have focused on the trade-off between privacy and utility or optimization of noise, but have not taken into consideration the latest advancements in adversarial attacks. In some research works, privacy leakage and attack strategies have been considered in isolation from different DP mechanisms. In this paper, the privacy leakage and defense in differentially private FL have been examined based on a systematic review of the latest research works. It can be concluded that the findings of this paper will help to strengthen the global collaboration in FL in a secure manner.

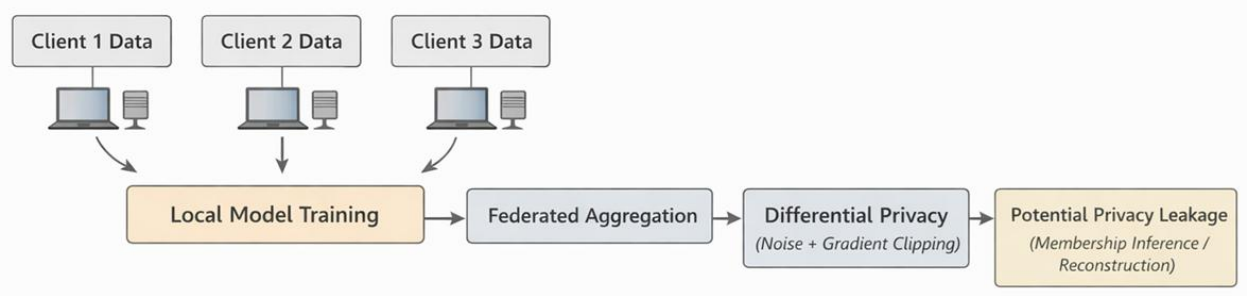

Figure 1. Privacy risks and protection mechanisms in differentially private FL.

2. Related works

Hu et al. [6] discussed privacy-preserving strategies in FL, addressing challenges including leakage, membership inference attacks, DP, and defense mechanisms for machine learning (ML) systems. It highlighted the need to support collaborative model training while safeguarding sensitive data and maintaining model performance. Bai et al. [7] focused on the use of DP within ML frameworks to provide formal privacy guarantees and reduce the risks associated with inference attacks. Differentially Private Stochastic Gradient Descent method relies on gradient clipping and noise addition to restrict the effect of each data point. Later studies have adapted these concepts for FL, focusing on privacy at the client level or aggregation. Alongside the development of privacy mechanisms, several studies have explored the potential privacy risks present in ML systems [8]. In membership inference attacks, it was revealed that an adversary can determine whether a particular piece of information was used in the training process based on the output or gradients of the model. In FL, the decentralized nature of the system provides an expanded attack surface based on shared information or intermediate training results. More recent research works [9] have been conducted to explore defense strategies to ensure the balance between privacy and accuracy in FL systems. In this regard, DP, secure aggregation, and regularization-based defense strategies have been explored to reduce information leakage in FL systems. In the literature, it was observed that most of the research works have been conducted to address specific issues in FL systems without exploring the entire system. In other words, the research works have been conducted in isolation without exploring the entire system. In the following section, a detailed analysis of some representative research works will be conducted to highlight the gaps in the literature. Table 1 summarizes representative works and highlights their key parameters, advantages, and limitations, which reveal the absence of quantum-aware adversary modeling in current approaches.

Table 1. Literature Review related to privacy threat modeling and membership inference attacks

Ref	Approach/ Algorithm	Key parameters	Advantages	Limitations	Gap identified
[1]	PPFL taxonomy and threat classification	Privacy-preserving techniques, threat categories, FL setting	Gives a broad system-level view of privacy-preserving FL	Does not include quantum-aware attacker modeling	Gap 1
[2]	Membership inference attack taxonomy	Attacker access level, confidence scores, logits, shadow models	Comprehensive view of privacy leakage mechanisms in ML	Assumes classical attacker capability	Gap 1

[3]	DP-FL framework with performanc e analysis	ε, δ, clipping norm, noise variance, rounds	Formal DP integration in FL with utility analysis	Assumes classical adversary cost model	Gap 2
[4]	Critical review of DP in ML	ε interpretation, composition, reporting practices	Identifies misuse of DP guarantees and weak assurance practices	Does not provide adversary-scaled recalibration method	Gap 2
[5]	Membershi p inference attack survey	Attacker access level, confidence scores, logits, attack success metrics	Comprehensive taxonomy of privacy leakage evaluation	Does not jointly evaluate utility, FL cost, and DP guarantees	Gap 3
[6]	Defense taxonomy for membershi p inference	Regularization, DP, masking, adversarial defenses	Organizes defense evaluation and tradeoff analysis	Evaluation remains defense-centric rather than system-centric	Gap 3

3. Research Gaps and Analysis

Recent studies on DP in FL have made significant progress in three directions: privacy-preserving training, secure aggregation, and privacy attack. However, most studies have focused on classical adversaries, limited leakage, and lack a direct connection between the growth rate of computation in the future and privacy budgeting. The problem is still open for the evaluation, modeling, and calibration of private FL systems.

3.1 Lack of a quantum-aware adversary model for privacy leakage in differentially private FL

The major gap in the literature is that existing research in privacy for FL considers leakage under classical attacker models. There have been surveys that provide taxonomies of threats and defenses in privacy-preserving FL and membership inference attacks. However, none of them have formally addressed how quantum attackers or attackers who benefit from computational power affect the feasibility of attacks or the amount of leakage in FL privacy. There have been surveys that provide taxonomies of threats in quantum attacks, but they have focused on future attackers in cryptography, not leakage in differentially private FL. Therefore, there is a gap in providing a unified attacker model that relates FL privacy leakage to quantum-enhanced attack capabilities.

3.2 Privacy budgeting and noise calibration in DP-FL are not designed for stronger adversary capability

The second major gap in current differentially private FL approaches is that they are mostly centered around privacy budget and noise in terms of efficiency, convergence, and communication. In other words, current studies have been improving clipping, heterogeneous privacy, and adaptive noise, while assuming

a classical attacker. In summary, privacy-accuracy trade-offs are calibrated with respect to the adversary, not the attacker.

3.3 Evaluation of private FL remains fragmented and lacks unified leakage-aware benchmarking

The major gap identified here is the evaluation. The literature on differentially private FL mostly discusses the privacy budget, accuracy, and communication cost individually. The same applies to the privacy attack literature, where the success of the membership inference attack, the effectiveness of the defense, and the leakage are individually measured but not related to the DP guarantee, system overhead, and FL performance. This becomes more significant when dealing with more powerful adversaries in the future, as the guarantee does not necessarily ensure the absence of leakage.

4. Comparative Analysis

According to the previous literature, the comparative synthesis evaluates the fulfillment of the fundamental quantum-aware privacy requirements of differentially private FL. Instead of focusing on individual algorithms, it considers the field in terms of adversary modeling, privacy budgeting, leakage estimation, incorporation of privacy components, and the system level. The conclusion drawn is that individual components are considered, but the quantum-aware privacy perspective is not.

Table 2. Comparative synthesis of key requirements of quantum-aware privacy in differentially private FL

Key Requirement	Status in Existing Literature	Inference from Review
Formal DP guarantees in FL	Addressed	Mature area
Membership leakage analysis	Partially addressed	Mostly attack-specific
Quantum-aware threat modeling	Not addressed	Core research gap
Privacy budget calibration	Addressed	Utility-focused
Quantum-aware noise calibration	Not addressed	Core research gap
Unified evaluation framework	Not addressed	Core research gap
End-to-end privacy integration	Partially addressed	Fragmented across studies

The findings from the reviewed literature were further synthesized to evaluate how current research addresses the key requirements of privacy protection in differentially private FL. Table 2 summarizes these requirements and highlights the extent to which they are addressed in existing studies.

5. Conclusions

In this research, the challenges in the fusion of DP and FL under dynamic threat models will be examined. A thorough examination of the literature revealed that there have been significant achievements in

privacy-preserving learning and leakage in distributed ML. DP offers strong guarantees, while FL offers collaborative learning without the need to share raw data. There have been some challenges identified in the literature. The majority of the literature assumes classical attacks without considering the impacts of advanced computational attacks. In some literature, the tuning of the privacy budget in DP-FL and the use of noise in DP-FL have been mostly utility-driven without adversary awareness. However, the frameworks used in the literature to evaluate these systems remain fragmented. Differential privacy protections, information leakage, and system performance have mostly been studied separately rather than within an integrated framework. Although DP mechanisms, membership inference attacks, and secure aggregation techniques have been widely investigated, they are often discussed in isolation. However, the fusion of DP, robust adversary modeling, secure aggregation, and privacy evaluations in a unified manner to ensure robust privacy in FL systems remains unexplored. A holistic solution to ensure robust privacy in collaborative analytics in various sectors like healthcare, finance, and critical national infrastructure remains unexplored.

References

1. Yin, X., Zhu, Y., & Hu, J. (2021). A comprehensive survey of privacy-preserving federated learning: A taxonomy, review, and future directions. ACM Computing Surveys (CSUR), 54(6), 1-36.
2. Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P. S., & Zhang, X. (2022). Membership inference attacks on machine learning: A survey. ACM Computing Surveys (CSUR), 54(11s), 1-37.
3. Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., ... & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE transactions on information forensics and security, 15, 3454-3469.
4. Blanco-Justicia, A., Sánchez, D., Domingo-Ferrer, J., & Muralidhar, K. (2022). A critical review on the use (and misuse) of differential privacy in machine learning. ACM Computing Surveys, 55(8), 1-16.
5. Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P. S., & Zhang, X. (2022). Membership inference attacks on machine learning: A survey. ACM Computing Surveys (CSUR), 54(11s), 1-37.
6. Hu, L., Yan, A., Yan, H., Li, J., Huang, T., Zhang, Y., Dong, C., & Yang, C. (2023). Defenses to Membership Inference Attacks: A survey. ACM Computing Surveys, 56(4), 1–34. https://doi.org/10.1145/3620667.
7. Bai, L., Hu, H., Ye, Q., Li, H., Wang, L., & Xu, J. (2024). Membership inference attacks and defenses in federated learning: A survey. ACM Computing Surveys, 57(4), 1-35.
8. Mohammadi, S. (2024). Balancing privacy and performance in federated learning. Journal of Network and Computer Applications.
9. Weng, S., Gou, Y., Zhang, L., & Imran, M. A. (2025). Evaluating privacy loss in differential privacy based federated learning. *Future Generation Computer Systems*, *172*, 107848.