Harmonizing Differential Privacy Mechanisms for Federated Learning: Boosting Accuracy and Convergence

Shuya Feng
University of Connecticut
Storrs, CT, USA

Meisam Mohammady
Iowa State University
Ames, IA, USA

Hanbin Hong
Shenao Yan
University of Connecticut
Storrs, CT, USA

Ashish Kundu
Cisco Research
San Jose, CA, USA

Binghui Wang
Illinois Institute of Technology
Chicago, IL, USA

Yuan Hong
University of Connecticut
Storrs, CT, USA

Abstract

Differentially private federated learning (DP-FL) offers a compelling approach to collaborative model training by ensuring robust privacy for clients. Despite its potential, current methods face challenges in effectively balancing privacy, utility, and performance across diverse federated learning scenarios. Addressing these challenges, we introduce UDP-FL, to our knowledge the first DP-FL framework that universally harmonizes any randomization mechanism, including those considered optimal, by employing the Gaussian Moments Accountant (viz. DP-SGD). Central to UDP-FL is the 'Harmonizer,' a dynamic module engineered to intelligently select and apply the most suitable DP mechanism tailored to each client's specific privacy requirements, data sensitivities, and computational capacities. This selection process is driven by the principle of Rényi Differential Privacy, which serves as a crucial mediator for aligning privacy budgets effectively. Our comprehensive evaluation of UDP-FL, benchmarked against established baseline methods, demonstrates superior performance in upholding privacy guarantees and enhancing model functionality. The framework's robustness has been rigorously tested against a broad spectrum of privacy attacks, making it one of the most thorough validations of a DP-FL framework to date. [1]

CCS Concepts

• **Security and privacy** → *Privacy-preserving protocols*; • **Computing methodologies** → *Machine learning algorithms.*

Keywords

Differential Privacy, Federated Learning, Privacy Attacks

ACM Reference Format:
Shuya Feng, Meisam Mohammady, Hanbin Hong, Shenao Yan, Ashish Kundu, Binghui Wang, and Yuan Hong. 2025. Harmonizing Differential Privacy Mechanisms for Federated Learning: Boosting Accuracy and Convergence. In *Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy (CODASPY '25), June 4–6, 2025, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3714393.3726517

1 Introduction

As the volume of data generated by emerging applications continues to grow exponentially, federated learning (FL) [56] has emerged as a promising solution for collaborative model training without sharing raw data. Despite the absence of local data exposure, sensitive information about the data can still be leaked through the exchanged model parameters via, e.g., membership inference attacks [8, 33, 61, 65, 83], data reconstruction attacks [26, 32, 39, 49], and attribute inference attacks [4, 10, 25, 52, 54].

Differential privacy (DP) has been proposed to provide rigorous privacy guarantees, ensuring that any data sample or user's data at any client cannot influence the output of a function (e.g., the gradient or model parameters in FL) [1, 2, 48, 79]. However, directly applying existing DP mechanisms and accounting approaches to FL can result in excessive noise addition and loose privacy guarantees.

Recent techniques, such as the advanced accounting of privacy [1, 71, 74, 89] and Rényi Differential Privacy (RDP) [69], have shown remarkable results in optimizing accounting for privacy loss in machine learning. These techniques significantly improved the tradeoff between data privacy and model utility. Nevertheless, applying the Moments Accountant (MA) as a budget economic solution to other DP mechanisms can be challenging. One reason is that deriving the moments accountant for each DP mechanism often requires a heavy analysis of the tails in their probability density function (PDF), which can be difficult and may differ from one mechanism to another. Accountants rely heavily on the Gaussian mechanism to ensure DP, but this often results in excessive perturbation of gradients. This can make it hard to achieve a satisfactory balance between privacy and utility in existing DP-FL methods [3, 24, 29, 38, 64, 68, 75, 85], which are dominantly based on the Gaussian mechanism and the DP-SGD variants. The need to harmonize different DP mechanisms in FL arises from the varying privacy requirements and data characteristics in different applications. Due to the different privacy-utility trade-offs and the changing privacy regulations, no single DP mechanism can be universally optimal.

[1] Source code, and the full version of this paper are available at https://github.com/datasec-lab/UDP-FL.

CODASPY '25, June 4–6, 2025, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-1476-4/2025/06
https://doi.org/10.1145/3714393.3726517

Table 1: An example of Harmonizer's output for choosing the DP mechanism. This is the ideal "after-the-fact" recommendation after observing experiments. MIA, DRA and AIA stand for membership inference attacks, data reconstruction attacks and attribute inference attacks, respectively.

DP Mechanism	**Convergence**	**Accuracy**	MIA Resilience	DRA Resilience	AIA Resilience	FL Arch	**Data Pattern**	Computational Efficiency
Gaussian	◐	●	●	◐	◐	FedAvg, FedSGD	Any size, Dense	●
Laplace	○	◐	◐	●	●	FedProx, SCAFFOLD	Any size, Sparse	◐
Staircase	●	●	●	●	●	FedAvg, FedOpt, q-FedAvg	Any size, Complex	◐

Harmonizing DP mechanisms allows for personalized privacy for clients and stricter privacy controls. This approach enables improved privacy guarantees while maintaining high utility in diverse FL scenarios, adapting to specific client needs and data characteristics, and selecting the most suitable mechanism for each use case.

Building upon the limitations of existing DP-FL methods [3, 29, 38, 64, 68, 75, 85], we propose a novel universal solution for DP-FL called UDP-FL, which offers a comprehensive approach for achieving DP in FL that extends beyond the popular DP-SGD algorithm (Gaussian). It universally adopts different DP mechanisms (e.g., Staircase [27] *which greatly outperforms Gaussian in terms of noise magnitude*[2]) and harmonizes privacy guarantees under a unified framework. It allows for tighter budget accounting and comparison of privacy guarantees between different DP-FL techniques, e.g., the Gaussian, Laplace and Staircase noise additive mechanisms, using the Rényi DP notion as a mediator variable [57, 74]. This approach provides greater flexibility and generalizability for real-world scenarios that suit specific client and data characteristics.

DP-noise Harmonizer. The Harmonizer component is central to UDP-FL, addressing the challenges of distributed learning systems. It automatically selects and harmonizes different DP mechanisms based on specific client requirements, dataset characteristics, and privacy concerns. We also rely on the Harmonizer component to enhance the budget accountant and management and improve the applicability and convergence of DP-FL across diverse scenarios. It harmonizes different DP mechanisms with the Rényi divergence, providing a generalized, flexible, and universal approach to DP-FL that adapts to various requirements, applications, and scenarios while ensuring the best privacy-utility trade-off for each case. Moreover, the Harmonizer maps the Gaussian Moments Accountant [1] to the corresponding Rényi DP of other DP mechanisms for FL, allowing it to measure privacy loss using Rényi divergence [1, 57, 74]. This algorithm calculates privacy leakage for each training round and ensures that the leakage of the adopted DP mechanism (e.g., Staircase mechanism [27]) does not exceed the Gaussian version (viz. DP-SGD and other variants).

Robustness against Privacy Attacks. UDP-FL demonstrates significant resilience against a spectrum of privacy attacks, aligning with advanced theories presented in recent work [63]. Through extensive empirical studies, our results reveal a substantial reduction in the success rate of Membership Inference Attacks (MIA) (i.e., LiRA [8]) compared with non-private FL, highlighting the framework's effectiveness in preserving privacy. While DP is not inherently designed to combat Attribute Inference Attacks (AIA), we observed a reduced correlation in feature learning. Against Data Reconstruction Attacks (DRA), UDP-FL has proven to be adept at preventing the reconstruction of original training data, thereby reinforcing its robustness in protecting data privacy within DP-FL environments.

Thus, the key contributions of this paper are summarized below:

(1) To our best knowledge, we propose the first DP-FL framework (UDP-FL) that harmonizes different differential privacy mechanisms in federated learning, achieving tighter privacy bounds and higher model accuracy compared to baseline methods while reducing computational and communication overheads.
(2) We propose a scoring-based approach for DP mechanism selection that considers privacy strength, utility preservation, and computational efficiency, providing insights into how different mechanisms can be optimally chosen for varying FL scenarios.
(3) We conduct comprehensive privacy evaluations of UDP-FL against membership inference [33, 65], data reconstruction [23, 88], and attribute inference attacks [25, 52]. Our results demonstrate superior defense capabilities, particularly with the Staircase mechanism showing consistent improvement across all attack scenarios while preserving model utility.

2 Preliminaries

2.1 System and Adversaries

In this work, we follow the standard semi-honest adversarial setting for differentially private federated learning (DP-FL) where the adversary can possess arbitrary background knowledge. The server is honest-but-curious by following the protocol but attempting to derive private information about the client's data from the exchanged messages during the training process. Clients are also categorized as "honest-but-curious", by strictly adhering to the protocol without deviating from established procedures [36]. Key responsibilities include refraining from manipulating local model updates and avoiding the use of poisoned or false data in the training. Upholding these guidelines is essential for maintaining the integrity and security of the global model, ensuring its reliability and robustness.

In terms of privacy, both UDP-FL (across all mechanisms) and the DP-SGD are adding noise into local gradients during the training process. Despite this, the disclosure of trained model parameters is proven to preserve (ϵ, δ)-DP [1]. The model parameters, viewed as post-processed results of the DP guaranteed noisy gradients, do not affect the privacy leakage.

Despite the distributed nature of federated learning offering collaborative model training, the challenge of preserving data privacy persists. The privacy concerns stem from the potential leakage of sensitive information through clients' local model updates. We also empirically evaluate the performance of UDP-FL against privacy

[2] The Staircase mechanism [27] has been proven to be optimal for ℓ_1 and ℓ_2 metrics for a wide range of privacy budget ϵ [27, 59].

attacks, including membership inference attacks (MIAs) [33, 65], data reconstruction attacks (DRAs) [23, 88], and attribute inference attacks (AIAs) [25, 52]. Their settings (which are different from DP-FL) will be discussed in Section 5.

2.2 Federated Learning

FL is an emerging distributed learning approach that enables a central server to coordinate multiple clients to jointly train a model without accessing to the raw data. Assuming the FL system has N clients $C = \{C_1, C_2, \cdots, C_N\}$ and each client C_k owns a private training dataset $\mathcal{D}_k = \{(\mathbf{x}_j^k, y_j^k)\}$ with $|\mathcal{D}_k|$ samples and each sample $\mathbf{x}_j^k$ has a label y_j^k. Then, FL considers the following distributed optimization problem:

$$\min_w F(w) = \sum_{k=1}^{N} p_k F_k(w), \tag{1}$$

where $p_k \geq 0$ is the client C_k's weight and $\sum_{k=1}^{N} p_k = 1$; Each client C_k's local objective is defined by $F_k(w) = \frac{1}{|\mathcal{D}_k|} \sum_{j=1}^{|\mathcal{D}_k|} \ell(w; (\mathbf{x}_j^k, y_j^k))$, with $\ell(\cdot;\cdot)$ a user-specified loss function, e.g., cross-entropy loss.

FedAvg [55] is the *de facto* FL algorithm to solve Equation (1) in an iterative way. It has the following steps:

(1) **Global Model Initialization.** The server initializes a global model w^0, selects a random subset $\mathcal{S}_n$ of n clients from C, and broadcasts w^0 to all clients in $\mathcal{S}_n$.
(2) **Local Model Update.** In each global epoch t, each client C_k receives the global model w^t, initializes its local model w_k^t as w^t, and updates the local model by minimizing $F_k(w^t)$ on the local dataset D_k. E.g., when running SGD, we have: $w_k^t \leftarrow w_k^t - \eta_t \nabla_{w_k^t} F_k(w^t)$, where η_t is the learning rate in the t-th epoch.
(3) **Global Model Update.** The server collects the updated client models $\{w_k^t\}$ and updates the global model w^{t+1} for the next round via an aggregation algorithm. For instance, when using FedAvg [55], the updated global model is: $w^{t+1} \leftarrow \frac{N}{n} \sum_{C_k \in \mathcal{S}_n} p_k w_k^t$, which is then broadcasted to clients for the next round.
(4) Repeat Steps 2 and 3 until the global model converges.

2.3 Differential Privacy and Rényi Accountant

The use of DP in FL enhances the benefits of collaborative model training with the need for protecting data privacy. It ensures that each data sample or user's contribution to the model training process is indistinguishable from others, and it can be implemented by adding noise to the gradients or parameters of the model or by using secure aggregation techniques. The notion of DP can be defined as below.

Definition 1 ((ϵ, δ)-Differential Privacy [16, 17]). *A randomization algorithm $\mathcal{A}$ is (ϵ, δ)-differentially private if for any adjacent databases d, d' that differ on a single element, and for any output set $\Omega \subseteq range(\mathcal{A})$, we have $Pr[\mathcal{A}(d) \in \Omega] \leq e^{\epsilon} Pr[\mathcal{A}(d') \in \Omega] + \delta$, and vice versa.*

FL with (ϵ, δ)-DP generally requires hundreds of training rounds to obtain a satisfactory model. Rényi accountant [71, 74] via Rényi Differential Privacy (RDP) [57] has been proposed to provide tighter privacy bounds on the privacy loss than the standard DP. RDP is defined over the Rényi divergence [69]. Recall that for two probability distributions P and Q, their Rényi divergence is defined as $\mathcal{D}_\alpha(P||Q) = \frac{1}{\alpha-1} \log \mathbf{E}_{x \sim Q}(\frac{P(x)}{Q(x)})^\alpha$ where x denotes a random variable and $\alpha > 1$ is the Rényi divergence order. Thus, the RDP can be defined as below.

Definition 2 ((α, γ)-Rényi Differential Privacy [57]). *A randomized mechanism $\mathcal{A}$ is said to have γ-Rényi differential privacy of order α, if for any adjacent datasets d, d' that differ on a single element, and for any output set $\Omega \subseteq range(\mathcal{A})$, the Rényi divergence $\mathcal{D}_\alpha[\mathcal{A}(d) = \Omega || \mathcal{A}(d') = \Omega] \leq \gamma$ holds.*

Rényi accountant [1, 58, 74] is a method for managing and assessing the cumulative privacy loss in a sequence of DP operations. It operates by tracking the cumulant generating function (CGF) of the privacy loss random variable over the sequence of operations. Specifically, the Rényi accountant evaluates the CGF at a series of fixed points corresponding to different orders of Rényi divergence, thus enabling the calculation of an overall privacy guarantee for the sequence. This overall guarantee is expressed in terms of Rényi Differential Privacy, providing a more nuanced and tighter estimation of privacy loss compared to traditional methods. The Rényi accountant is particularly effective in complex scenarios, such as those encountered in machine learning algorithms, where multiple DP operations are composed over time.

Although the efficacy of the Rényi accountant in providing a refined estimation of privacy loss becomes increasingly significant when addressing privacy loss in FL, several challenges still exist. These include non-optimal noise mechanisms like Gaussian or Laplace that degrade accuracy, loose DP guarantees in complex systems, difficulty in tracking privacy loss across diverse clients, and reduced convergence speed. These limitations underscore the need for developing an enhanced DP-FL framework that *universally ensures tighter privacy out of diverse DP mechanisms* while maintaining fast convergence and accuracy. The Rényi accountant [58] adopted in UDP-FL helps to address these challenges by providing tighter accounting of privacy loss across numerous training rounds. Moreover, other recent accountants [1, 71, 74, 89] can also act as viable alternatives, offering flexibility in the choice of DP composition. Our framework's design is orthogonal to the specific choice of accountant, meaning it is adaptable and could incorporate even tighter accounting methods as they become available in the future.

3 UDP-FL Framework

In this section, we propose a comprehensive framework, called universal DP-FL (UDP-FL), for achieving superior privacy-utility tradeoff and faster convergence in FL.

3.1 Building Blocks of UDP-FL

DP Mechanisms. UDP-FL integrates diverse DP mechanisms, including Gaussian, Laplace, and Staircase [27], offering versatility for various FL applications. This approach provides stronger privacy guarantees while maintaining training quality, surpassing Gaussian-only methods in adaptability. We primarily utilize the Staircase mechanism for its optimality in ℓ_1 and ℓ_2 norms [27, 59], making it particularly effective in scenarios requiring robust privacy guarantees without significant accuracy loss. UDP-FL is designed to accommodate other advanced DP mechanisms [9, 59], enabling flexibility to adapt to specific FL task requirements, such as data sensitivity, privacy-utility trade-offs, and computational constraints.

Harmonizer. The Harmonizer dynamically selects the optimal DP mechanism (e.g., Gaussian, Laplace, Staircase) for each client in FL, based on their privacy requirements, data sensitivity, and computational resources. It ensures privacy-preserving gradient updates by clipping gradients, adding noise, and tracking privacy budgets using Rényi Differential Privacy, balancing privacy and utility across clients.

3.2 UDP-FL Framework

In this section, we present the main steps in UDP-FL.

(1) **Local Data Preparation**. The clients collect and store their data locally. Once their data is ready, clients specify privacy parameters (ϵ_k, δ_k) and send them to the server.

(2) **Global Model Initialization**. The server initializes a global model w^0, selects a subset of clients $\mathcal{S}_n \subset C$, and sends the current global model parameters w^0 to the selected clients.

(3) **Local Model Update**. Clients perform local training over their data. The Harmonizer manages all aspects of differential privacy during this process, including gradient processing, noise addition, and privacy accounting. Finally, clients send their updated local models w_k^t to the server.

(4) **Global Model Update**. The server aggregates the received local models: (1) Initialize intermediate models by averaging pairs of client models. (2) Optimize intermediate models to find a low-loss path between client models. The Harmonizer ensures privacy-preserving computations for any client-side operations. (3) Update the global model w^{t+1} based on the optimized intermediate models.

The detailed procedures of UDP-FL are illustrated in Algorithm 1.

Algorithm 1: Harmonizer

```
Input: Clients $C = \{C_1, ..., C_N\}$, datasets $D = \{D_1, ..., D_N\}$,
       computational resources $R$, privacy requirements
       $E = \{\epsilon_1, ..., \epsilon_N\}$, $\Delta = \{\delta_1, ..., \delta_N\}$
Output: Selected DP mechanism $M^*$, harmonized privacy
        parameters $\epsilon^*$, $\delta^*$
mechanisms ← ['Gaussian', 'Laplace', 'Staircase'];
best_score ← $-\infty$;
$M^*$ ← null;
for each $M$ in mechanisms do
    score ← calculate_score($M, D, R$);
    if score > best_score then
        best_score ← score;
        $M^*$ ← $M$;
$\epsilon^*, \delta^*$ ← initialize_privacy_params($E, \Delta$);
for each $C_k \in C$ do
    $g_k$ ← compute_gradient($C_k$);
    $g_k^c$ ← clip_gradient($g_k, S_k$);
    $g_k^n$ ← add_noise($g_k^c, M^*, \epsilon^*, \delta^*$);
    send_noisy_gradient_to_server($C_k, g_k^n$);
RD ← calculate_renyi_divergence($M^*, \epsilon^*, \delta^*$);
adjust_privacy_params($\epsilon^*, \delta^*, RD$);
while privacy_budget_not_exceeded do
    check_privacy_budget($\epsilon^*, \delta^*$);
return $M^*, \epsilon^*, \delta^*$;
```

3.3 UDP-FL Harmonizer

The Harmonizer serves as the central privacy management component within UDP-FL, orchestrating the selection and application of differential privacy mechanisms across the federated learning system. Its functionality can be divided into three primary aspects: data characteristic analysis, privacy requirement assessment, and adaptive noise management.

First, the Harmonizer performs comprehensive data characteristic analysis for each client. This analysis begins by examining the statistical properties of client data, including distribution patterns, gradient sparsity, and feature sensitivity levels. For gradient distribution analysis, the Harmonizer computes both the sparsity ratio and gradient magnitude distribution. Dense gradients with normal-like distributions typically benefit from Gaussian mechanisms, while sparse gradients with heavy-tailed distributions are better suited for Laplace mechanisms. The Harmonizer also analyzes data sensitivity by computing the maximum change possible in the gradient when a single training example is modified, which is crucial for calibrating noise addition.

Following data analysis, the Harmonizer conducts a thorough privacy requirement assessment using a specialized scoring system. This system evaluates each DP mechanism's suitability through a weighted combination of privacy strength, utility preservation, and computational efficiency. The scoring function is defined as:

$$\text{Score}(M) = wp * \text{Privacy}(M) + wu * \text{Utility}(M) + we * \text{Efficiency}(M)$$

The Privacy(M) score quantifies the mechanism's theoretical privacy guarantees and empirical resistance to known attacks. For example, when handling medical data, this score reflects how well the mechanism resists membership inference and reconstruction attacks. The Utility(M) score measures the mechanism's ability to preserve model accuracy by examining historical performance data and theoretical bounds on noise addition. The Efficiency(M) score evaluates computational overhead, memory requirements, and communication costs associated with implementing the mechanism. The scoring-based mechanism selection process incurs minimal computational overhead while maintaining robust performance, making UDP-FL highly practical for real-world deployments. The selection logic employs efficient lookup tables, as presented in Table 2, along with simple gradient statistics, thereby avoiding computationally intensive operations during training.

The Harmonizer dynamically adjusts scoring weights based on client-specific requirements. For medical institutions with strict privacy regulations, wp might be set to 0.5 or higher, while resource-constrained IoT devices might be set to 0.4 or higher to prioritize computational efficiency. These weights are continuously refined based on observed performance and changing requirements.

A key function of the Harmonizer is calculating and managing the noise multiplier, which controls the noise to be added to the model updates while balancing privacy and utility. It adjusts this multiplier throughout training by computing Rényi Divergence at each iteration to ensure privacy loss remains within the set bounds.

Furthermore, the Harmonizer incorporates a penalty term to address heterogeneity in individual client privacy guarantees, adjusting the gradient updates based on the client's privacy settings:

$$w_k^t \leftarrow w_k^t - \eta \left(g_k^c + \lambda_k (w_k^t - w_{\max}^t) \right)$$

Where $\lambda_k = \frac{\epsilon_{\max} - \epsilon_k}{\epsilon_{\max}}$ adjusts based on each client's privacy needs, ensuring clients with weaker DP guarantees are balanced against those with stronger guarantees.

Ultimately, the Harmonizer enables a highly customizable and adaptable DP-FL process, allowing clients to define their own privacy parameters while ensuring the system harmonizes these diverse preferences using Rényi Divergence. This ensures efficient management of the privacy budget across all participants, without sacrificing model performance. By selecting optimal DP mechanisms and tracking privacy budgets, the Harmonizer promotes convergence to a unified global model while maintaining robust privacy guarantees.

Through this comprehensive approach, the Harmonizer ensures optimal privacy-utility tradeoffs while adapting to diverse client requirements and data characteristics. Its modular design allows for the incorporation of new DP mechanisms as they are developed, making UDP-FL extensible and future-proof. The system continuously monitors and adjusts its decisions based on observed performance metrics, ensuring robust privacy protection throughout the federated learning process.

Table 2: Harmonizer's Mechanism Selection Criteria. A practical interpretation of the scoring function Score (M), showing how different characteristics influence mechanism selection in UDP-FL.

Client Data	Privacy Requirements	Resource	Mechanism
Dense gradients	High ($\epsilon < 3$)	Standard compute	Gaussian
Sparse gradients	Moderate (ϵ = 3-8)	Limited compute	Laplace
Complex distributions	Low ($\epsilon > 8$)	High compute	Staircase
Heterogeneous data	High ($\epsilon < 3$)	Limited compute	Gaussian
Time-series data	Moderate (ϵ = 3-8)	Standard compute	Staircase
High-dimensional data	High ($\epsilon < 3$)	High compute	Gaussian
Small datasets	Moderate (ϵ = 3-8)	Limited compute	Laplace
Large client pool	Low ($\epsilon > 8$)	Standard compute	Staircase

4 Theoretical Analyses

In this section, we provide a theoretical analysis of the privacy and utility of UDP-FL and examine the influence of its parameters on the overall privacy guarantees. The proofs of the theorems are provided in the full version of the appendix.

4.1 Error Bounds Analysis of UDP-FL

The Staircase mechanism can be viewed as a geometric mixture of uniform probability distributions, ensuring an optimal privacy-utility tradeoff, particularly for medium to large ϵ values. This mechanism generates noise by carefully mixing uniform distributions, adjusting for the privacy budget and other requirements, and adding the noise to query responses in a way that preserves privacy without significantly compromising accuracy. The Staircase mechanism applied to a function f is defined as follows:

$$f = \begin{cases} e^{-\rho\lambda} \cdot y, & ||x||_1 \in [\rho\Delta, (\rho + \nu)\Delta] \\ e^{-(\rho+1)\lambda} \cdot y, & ||x||_1 \in [(\rho + \nu)\Delta, (\rho + 1)\Delta] \end{cases}$$

for $\rho \in \mathbf{N}$, where λ and ν are the parameters controlling the noise distribution, Δ is the sensitivity of the query, and ρ defines the intervals for the ℓ_1-norm of x. Furthermore, y is given by:

$$y \triangleq \frac{1 - e^{-1}}{2\Delta(\nu + e^{-\lambda}(1 - \nu))}$$

Theorem 1 formalizes the privacy guarantees of the Staircase mechanism using Rényi differential privacy (RDP). This enables the Harmonizer in UDP-FL to track the privacy loss of the Staircase mechanism across different rounds of federated learning.

Theorem 1 (Proof in full version Appendix A.4). *For any order $\alpha > 1$ and privacy budget $\epsilon_\alpha > 0$, the Staircase mechanism satisfies $(\alpha, \epsilon_\alpha)$-Rényi differential privacy (RDP), where ϵ_α is given by:*

$$\epsilon_\alpha = \frac{1}{2} e^{(\alpha-1)\lambda} + \frac{1}{2} e^{-\alpha\lambda} + \frac{1 - e^{-1}}{2(\nu + e^{-\lambda}(1 - \nu))}$$
$$\times \left(\left(e^{(\alpha-1)\lambda} + e^{-\alpha\lambda} \right) (1 - \nu) + |2\nu - 1| e^{-sgn(\frac{1}{2} - \nu)\lambda} \right)$$

We now provide the error bounds for the DP mechanism (i.e., noise applied to the model parameters).

Lemma 1 (Proof in Geng et al. [27]). *Given the Staircase mechanism $f(\lambda, \Delta, \nu)$, when $\nu = \frac{1}{1+e^{\lambda/2}}$, the minimum expectation of noise amplitude is $\Delta \frac{e^{\lambda/2}}{e^{\lambda} - 1}$.*

As demonstrated by Geng et al. [27], when the privacy budget ϵ is sufficiently small, the Staircase mechanism saves at least $\Delta^2 \left(\frac{1}{12} - \frac{\epsilon^2}{720} + O(\epsilon^4) \right)$ perturbation in variance compared to Gaussian and Laplacian mechanisms. Since our privacy budget spent in each round is negligible ($\epsilon \longrightarrow 0$), we can capitalize on this improved accuracy payoff per round, leading to enhanced model performance while maintaining robust privacy guarantees.

Theorem 2 (Proof in full version A.1). *For any $\alpha > 1$, $\gamma > 0$, Staircase mechanism $f(\lambda, \Delta, \nu)$ satisfies (α, γ)-Rényi differential privacy, where $\nu = \frac{\log(\gamma - 1)}{\alpha - 1}$*

Theorem 3 (Proof in full version Appendix A.5). *The expectation of the ℓ_1 distance for the output model parameters preserved by UDP-FL with the Staircase mechanism after T training rounds is:*

$$\frac{mT}{1 - e^{-\lambda}} \left(\nu^2\Delta^2 + e^{-\lambda}\Delta^2 - e^{-\lambda}\nu^2\Delta^2 + \Delta e^{-\lambda} \right)$$

where m is the length of the loss function, and ν, ρ, λ are the noise multipliers computed by UDP-FL.

Furthermore, the utilization of Staircase noise has been demonstrated to significantly accelerate convergence compared to the baseline, as empirically validated in Figure 3, Figure 4, and Table 1. The enhanced convergence speed is a byproduct of applying the optimal noise for ℓ_1 and ℓ_2 distance metrics (for a wide range of ϵ). In essence, when DP is fixed to guarantee ϵ, this noise has been proven to minimize both ℓ_1 and ℓ_2 distances. This implies that all noise-additive operations, including gradient perturbation and client model averaging, yield more accurate results, closer to the non-private scenario.

5 Experiments

In this section, we will evaluate the performance of UDP-FL on privacy, accuracy and efficiency. The key objectives of our evaluations are: (1) Assessing the accuracy of UDP-FL in comparison with SOTA mechanisms. (2) Investigating the convergence behavior of UDP-FL to understand how its hyperparameters influence the training performance. (3) Demonstrating UDP-FL's computational and communication efficiency against baseline methods. (4) Rigorously testing UDP-FL's resilience against common privacy attacks, including membership inference, data reconstruction, and attribute inference attacks, to validate its defense performance.

5.1 Implementation

Our implementation leverages PyTorch 1.9.0, allowing easy integration with existing ML pipelines. The core of UDP-FL, our Harmonizer component, is designed as a flexible Python module that can seamlessly switch between different DP mechanisms without requiring changes to the overall federated learning setup. To use UDP-FL, practitioners only need to specify their desired privacy budget ϵ and choose a DP mechanism, with the framework automatically handling the rest of the privacy-preserving process. This simplicity and flexibility enable UDP-FL to adapt to a wide range of datasets and use cases beyond those presented in our experiments, addressing potential concerns about overfitting specific scenarios. We validate this adaptability by testing UDP-FL on diverse datasets, including MNIST [45], Medical MNIST [44], UTKface [84] and CIFAR-10 [43], demonstrating its effectiveness across various data types and distributions.

5.2 Experiment Setup

ML Models. For the MNIST dataset, we utilize a two-layer CNN with ReLU activation, max pooling, and two fully connected layers. In contrast, the Medical MNIST dataset employs a four-layer CNN with additional fully connected layers designed for 3-channel image classification. For CIFAR-10, we use the ResNet-18 architecture [34] other than pre-trained models. Focusing on initial training may lead to lower accuracy, but is crucial for assessing UDP-FL's influence on early-stage learning and privacy in federated learning.

Parameters Setting. Although the Harmonizer supports dynamic DP mechanism selection, we fix the mechanisms to Laplace, Gaussian, and Staircase in our experiments to independently evaluate their robustness and accuracy. Thus, the mechanisms used in the experiments are not chosen by the Harmonizer but are manually set to ensure consistency in evaluating each mechanism's performance across different datasets and privacy conditions. We have used a learning rate of 0.01 in all experiments. For the MNIST and Medical datasets, we have set the clipped gradient (ℓ_2) as 1 and 0.1, respectively. For the CIFAR-10, and UTKFace datasets (to be used in the defense evaluation against privacy attacks in Section 5.6), the clipped gradient is set as 0.01. The default number of clients is set at 10, and the sampling rate is set at 0.05. We set the local communication round as 150 and the local training epoch as 2.

Harmonizer Validation. While the Harmonizer is designed to dynamically select the optimal DP mechanism based on client requirements, data characteristics, and privacy constraints, for experimental validation purposes, we first evaluate each mechanism independently to verify the Harmonizer's selection criteria. This systematic evaluation helps validate the scoring function used by the Harmonizer and demonstrates why certain mechanisms are preferred in specific scenarios. We evaluate three primary mechanisms: Laplace, Gaussian, and Staircase. For each dataset (MNIST, Medical MNIST, and CIFAR-10), we first run experiments with fixed mechanisms to establish baseline performance across different privacy budgets and data characteristics. These results inform the Harmonizer's scoring function and validate its mechanism selection logic.

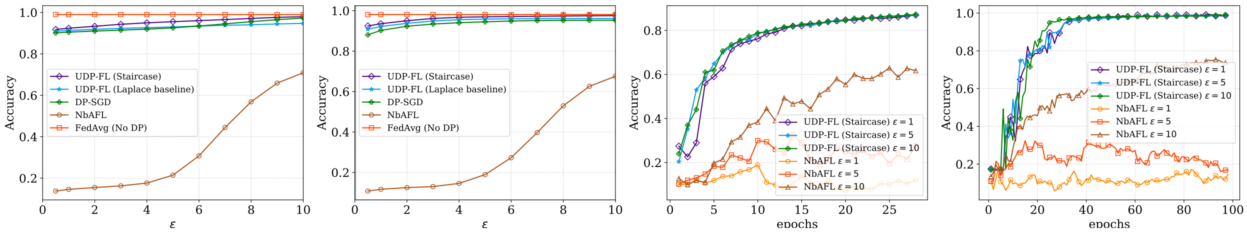
5.3 Accuracy Comparison vs Baseline Methods

Due to the diverse settings for DP guarantee, trust model, noise injection, and model architecture in DP-FL, there is no universally accepted benchmark for evaluating DP-FL methods. Therefore, in this work, we will compare UDP-FL with NbAFL [75] and DP-SGD applied to FedAvg (equivalent to UDP-FL with Gaussian noise), as they share similar settings (sample-level DP within each client and local noise injection before aggregation). We also apply the classic FedAvg [56] without DP guarantees as the baseline in the experiments. Figure 1 shows the comparison results. We can observe from Figure 1(a) and 1(b) that UDP-FL, when using the Staircase mechanism, obtains higher accuracy and faster convergence rates compared to other methods. For instance, on the MNIST dataset, UDP-FL achieves 90% accuracy in about 25 epochs, while other methods struggle to reach this level even after 100 epochs. This faster convergence is particularly evident in the Medical MNIST dataset, where UDP-FL converges in approximately 30 epochs, while other methods fail to converge even after 100 epochs. Furthermore, UDP-FL achieves nearly the same accuracy as FedAvg without DP guarantee. The primary reason is that we evenly distribute the datasets to each client so that their local datasets are unique.

Moreover, each time we randomly select some clients to update their models, this may cause the global model to take more time to converge because the data distribution from clients is more heterogeneous. Thus, the noise generated from UDP-FL helps to balance the unbiased distribution and contributes to faster convergence. UDP-FL requires fewer training epochs to reach the optimal performance (as shown in Figure 1(c) and 1(d)) whereas other methods take more training epochs but still result in lower accuracy.

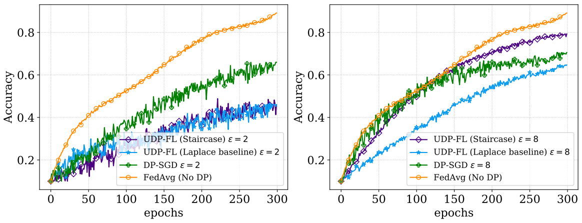
5.4 Training Performance Analysis

Figure 1(c) and 1(d) further prove that even when the data distributions from different clients are heterogeneous, UDP-FL can still preserve good performance and converge faster than other baselines. For instance, in Figure 1(c), it takes about 25 epochs for UDP-FL to achieve an accuracy of 0.9 and the accuracy tends to be stable since then. However, the accuracy of other baselines is low and fluctuates. Similar results can be seen in Figure 1(d), where it only takes about 30 epochs for UDP-FL to converge on the medical dataset. In contrast, other baselines did not converge even with 100 epochs. Thus, UDP-FL converges faster and has better performance.



(a) MNIST: Accuracy vs ϵ (b) Medical: Accuracy vs ϵ (c) MNIST: Accuracy vs epochs (d) Medical: Accuracy vs epochs

Figure 1: Accuracy and convergence results of UDP-FL and the baselines. 1) among the three mechanisms, the Staircase always performs the best with the same privacy budget; 2) UDP-FL obtains significantly better privacy-utility tradeoff and faster convergence than the baseline; and 3) UDP-FL (Staircase) even has a comparable accuracy with FedAvg (No DP).



(a) Accuracy vs epochs (b) Accuracy vs epochs

Figure 2: UDP-FL on CIFAR-10 when (a) $\epsilon = 2$ and (b) $\epsilon = 8$. For a small privacy budget ($\epsilon = 2$), DP-SGD yields better performance, while at a larger privacy budget ($\epsilon = 8$), UDP-FL with Staircase mechanism outperforms DP-SGD and Laplace.

Figure 2 for UDP-FL on CIFAR-10 shows that with a small privacy budget, UDP-FL with Staircase and Laplace mechanism converge more smoothly than DP-SGD. At a larger privacy budget, UDP-FL achieves rapid convergence comparable to the non-private baseline, matching the theoretical results on faster convergence.

5.5 UDP-FL Evaluation

The Number of Clients. As discussed in Section 4, the enhanced privacy is related to a number of clients N and sampling rate q. Analyzing the impact of these hyperparameters allows us to gain insights into the scalability and adaptability of UDP-FL, ensuring optimal performance across different settings. We use UDP-FL with the Staircase mechanism to represent the optimal randomization (w.r.t. a range of ϵ) and Laplace mechanism as another baseline besides DP-SGD. We will present the performance of UDP-FL with other mechanisms in Section 5.3.

The number of clients significantly impacts the performance, communication overhead, model convergence, and privacy-utility tradeoff in FL frameworks. We experimented with 50, 100, 150, 200 clients (as shown in Figure 3), selecting 10% of all the clients randomly per training round. Each chosen client trains locally on 5% of their data for 2 epochs. As shown in Figures 3(a) and 3(c), the accuracy slightly decreases with more clients, likely due to the increased complexity in aggregating diverse model updates. This variance affects convergence and generalization but, notably, the performance drop is not substantial, showing UDP-FL's effectiveness in handling scalable, heterogeneous client scenarios in FL.

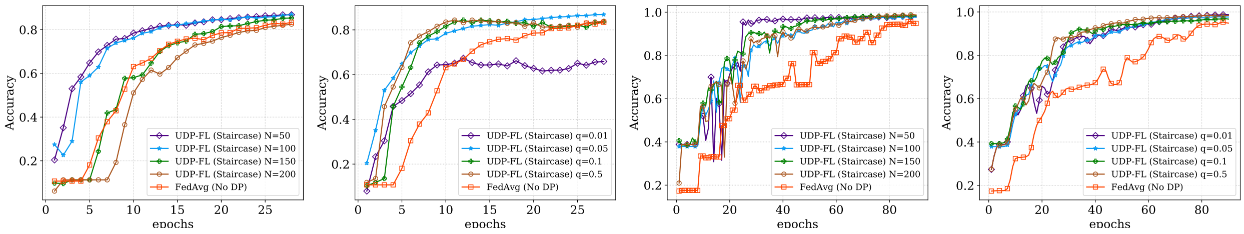
Sampling Rate. The sampling rate significantly impacts convergence speed, model performance, and the privacy-utility tradeoff in UDP-FL. A higher rate means more data samples are used per training epoch, leading to faster convergence and better model performance, but also higher privacy loss. Thus, more noise will be used to preserve privacy. Figure 3(b) and 3(d) confirm the results in the theoretical analysis (as discussed in Section 4). Experiments demonstrate that higher sampling rates (e.g., 0.5) lead to improved accuracy and faster convergence, whereas lower rates (e.g., 0.01) result in slower convergence and reduced accuracy. These findings suggest that utilizing a larger number of data samples per training round enhances overall model performance.

Table 3: UDP-FL accuracy vs. privacy guarantees.

Datasets	Mechanisms	$\epsilon = 2$	$\epsilon = 4$	$\epsilon = 6$	$\epsilon = 8$	$\epsilon = \infty$
CIFAR-10	Gaussian	**0.704**	**0.714**	0.719	0.742	0.871
	Laplace	0.475	0.461	0.506	0.638	0.871
	Staircase	0.633	0.691	**0.733**	**0.780**	0.871

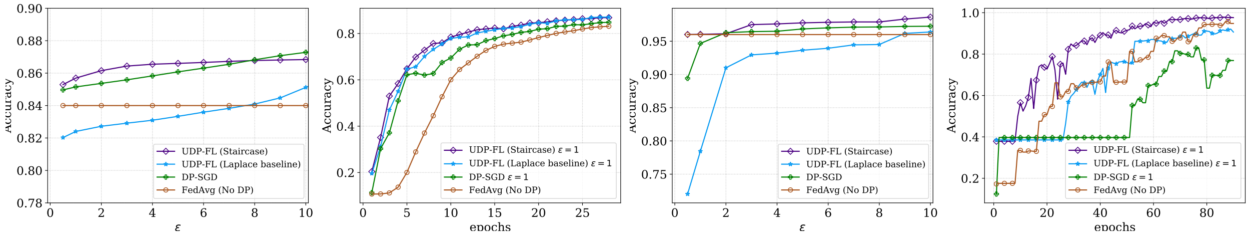
Performance across Privacy Settings. In our comprehensive evaluation, we first delve into the privacy-utility trade-off on the MNIST and Medical datasets, as depicted in Figure 4. UDP-FL's performance with the Staircase mechanism exhibits a steady increase in accuracy as the privacy budget increases and outperforms other baseline methods. The accuracy versus epochs on both datasets reveals UDP-FL's capability for consistent learning over time, even outperforming the non-private baseline (FedAvg) during early epochs on the Medical dataset. These results have validated the practicality of UDP-FL in scenarios where stringent privacy is required without substantially compromising model performance.

Subsequently, we extend our evaluation to the CIFAR-10 dataset, which offers a more complex challenge due to its higher dimensionality and diverse image representations. This further examination on CIFAR-10 aims to validate UDP-FL's robustness and scalability in more intricate visual data scenarios. From the experimental results on CIFAR-10 dataset in Table 3, we observed that DP-SGD has shown consistent moderate accuracy across varying privacy levels, peaking when no privacy constraint was applied ($\epsilon = \infty$). UDP-FL with the Laplace mechanism, while less accurate at stricter privacy settings, improved as privacy constraints were relaxed. Notably, UDP-FL with the Staircase mechanism initially underperformed at lower ϵ values but significantly improved with relaxed privacy, surpassing Gaussian. This observation demonstrates that when



(a) MNIST: Acc vs client # (b) MNIST: Acc vs sampling rate (c) Medical: Acc vs client # (d) Medical: Acc vs sampling rate

Figure 3: Impact of the number of clients and sampling rate on UDP-FL. We observe that: 1) when the number of clients increases, shown in Figures (a) and (c), UDP-FL needs more epochs to converge; 2) with the increase of data sampling rate shown in Figures (b) and (d), UDP-FL converges faster.



(a) MNIST: Accuracy vs ϵ (b) MNIST: Accuracy vs epochs (c) Medical: Accuracy vs ϵ (d) Medical: Accuracy vs epochs

Figure 4: Performance evaluation of UDP-FL on MNIST and Medical datasets. (a) and (c) present UDP-FL's accuracy under various differential privacy noise mechanisms, compared to a non-private baseline, with varying ϵ values. (b) and (d) illustrate the learning curves over training epochs for the MNIST and Medical datasets without privacy and with DP guarantees.

the privacy protection is medium and relatively weaker, UDP-FL with the Staircase mechanism can effectively improve the trade-off between privacy and utility, even in the FL setting.

Computation and Communication Overheads. The efficiency of a DP-FL framework is a critical aspect to be considered, where the computation and communication overheads can accurately reflect the overall system performance. In this section, we evaluate the computation and communication overheads of UDP-FL. Specifically, we will present the total local training time and noise multiplier computation time in UDP-FL, which are executed on the Flower platform [5]. All the results are shown in Table 4 (can greatly reduce the training time due to faster convergence). It implies that UDP-FL does not require heavy computational resources, particularly when compared to the training times with larger datasets and clients, which reinforces the idea of efficient or quick parameter handling.

Table 4: Runtime of UDP-FL (sec) vs # training iterations.

Iterations	50	100	500	1000
MNIST (50 clients)	1049.71	2056.34	10500.80	20803.10
Medical (50 clients)	644.97	1328.23	6482.50	13177.98
CIFAR-10 (50 clients)	1808.67	3515.08	16443.23	31956.76
MNIST (100 clients)	1108.91	2196.13	10934.62	21579.99
Medical (100 clients)	714.02	1445.75	7064.78	14353.64
CIFAR-10 (100 clients)	1873.93	3653.36	17215.17	33562.24
Computing Parameters	50.50	50.60	51.60	54.30

Moreover, since each client sends a local model to the server with a size of ~2MB in each communication round, with a faster convergence by UDP-FL, the total bandwidth consumption can be reduced by more than 50%, e.g., 40GB bandwidth reduction for 10 clients involved in the FL.

5.6 Defense against Privacy Attacks

In the following, we will evaluate the performance of our UDP-FL against several common privacy attacks in the domain of federated learning, specifically Membership Inference Attack (MIA) [65] and Data Reconstruction Attack (DRA) [23, 88]. Notably, we underscore the significance of DP's core advantage lies in its ability to offer plausible deniability [7], maintaining the privacy defenses of UDP-FL against these attacks with strong indistinguishability. This fundamental attribute ensures that, regardless of an attack's sophistication, the indistinguishability introduced by DP mechanisms significantly complicates the accurate reconstruction or direct association of any data with individual participants. Since these privacy attacks are primarily developed based on deterministic results, the evaluation against these attacks can only be based on several sampled random results instead of the entire output space. DP inherently produces randomized outputs, which introduces uncertainty into any inference made from the released data. As a result, even if a privacy attack achieves high accuracy on a subset of the perturbed results, clients and data owners retain plausible deniability—they can legitimately dispute the validity of such inferences, as formalized in [7].

Membership Inference Attacks. We assess the resilience of UDP-FL against three advanced Membership Inference Attacks (MIAs) using the CIFAR-10 dataset. The first attack, Shokri et al.[65], examines how models reveal information about their training data. The second, Likelihood Ratio Attack (LiRA) [8], uses shadow models to statistically ascertain if a data point was used in training. The third, the Canary attack, employs synthetic images, refined through iteration, to probe the model's disclosure of training data characteristics. Our goal is to identify and mitigate potential privacy risks in the model's outputs. Following the evaluation setting from Canary [77], we maintain a strict True Positive Rate (TPR) of 0.01 to measure False Positive Rate (FPR), AUC, and Accuracy (ACC). This conservative approach minimizes false positives, addressing the significant legal and ethical concerns associated with erroneous membership inferences and highlighting the need for a robust defense mechanism that effectively prevents unauthorized inferences while minimizing errors.

Table 5: Evaluation of the Shokri et al. [65], SOTA LiRA [8] and Canary [77] MIAs on CIFAR-10. TPR* denotes the TPR when FPR=0.01. The TPR*, ACC and AUC for Shokri et al. are 0.053, 0.710, and 0.757. The TPR*, ACC and AUC LiRA are 0.126, 0.651, and 0.716. The TPR*, ACC and AUC LiRA are 0.137, 0.649, and 0.719.

Attack	ϵ	DP-SGD			UDP-FL (Laplace baseline)			UDP-FL (Staircase)		
		TPR*	ACC	AUC	TPR*	ACC	AUC	TPR*	ACC	AUC
Shokri	2	**0.009**	**0.502**	**0.493**	0.011	0.509	0.508	**0.009**	0.506	0.501
	4	**0.009**	**0.505**	**0.499**	0.013	0.512	0.509	**0.009**	**0.505**	0.502
	6	0.013	0.510	0.509	0.014	0.505	0.501	**0.009**	**0.504**	**0.499**
	8	0.011	0.503	0.499	**0.007**	0.505	0.500	0.008	**0.504**	**0.496**
LiRA	2	0.013	0.506	0.497	0.009	0.506	0.501	**0.008**	**0.504**	**0.491**
	4	0.014	0.513	0.505	0.013	**0.505**	**0.495**	**0.008**	0.510	0.501
	6	0.017	0.514	0.511	0.011	**0.505**	**0.493**	**0.009**	0.510	0.504
	8	0.011	**0.504**	0.497	**0.009**	0.513	0.505	**0.009**	0.507	**0.492**
Canary	2	0.016	0.519	0.513	0.011	0.504	0.497	**0.009**	0.506	**0.495**
	4	0.016	**0.510**	0.504	0.020	0.512	**0.498**	**0.013**	0.512	0.509
	6	0.015	0.523	0.528	0.011	0.510	0.507	**0.009**	**0.507**	**0.500**
	8	**0.010**	0.521	0.518	0.015	0.516	0.507	0.012	**0.511**	**0.509**

Table 5 highlights the effectiveness of various noise mechanisms in mitigating membership inference attacks. Both attacks show that the Staircase noise addition under UDP-FL consistently yields the lowest False Positive Rate (FPR), indicating superior privacy preservation. The Laplace baseline also effectively reduces FPR, although not as consistently low as the Staircase, suggesting good but variable privacy protection. In terms of Accuracy and AUC, the Staircase and Laplace mechanisms demonstrate moderate success in maintaining model utility while ensuring privacy.

Data Reconstruction Attacks. We evaluate the data reconstruction attacks on UDP-FL on CIFAR-10. These attacks aim to reconstruct training data points from a target model. Instead of training a separate reconstruction model, we directly optimize synthetic inputs to match the gradient information obtained from the target model, following [26]. To evaluate multi-image attacks, we average the gradients from batches of up to 100 images before running the reconstruction. We assess the attack's efficacy by measuring the mean squared error (MSE) and Structural Similarity (SSIM) between the reconstructed and original images.

The evaluation of data reconstruction attacks on CIFAR-10 demonstrates the efficacy of UDP-FL against DRA threats. Notably, across varying privacy budgets (ϵ values), Staircase noise consistently results in higher MSE and lower SSIM, substantially reducing the attackers' ability to reconstruct original images accurately. While both DP-SGD and UDP-FL with Laplace baseline offer comparable levels of protection, evidenced by similar MSE and SSIM values.

Further experiments reveal that UDP-FL's protection against reconstruction attacks [23] remains robust even under varying attack conditions. The adversary first trains a separate reconstruction model on a dataset from a similar distribution as the target model's training data. The reconstruction model learns to generate synthetic inputs that closely match real samples, using the predictions from the target model as feedback. By optimizing the synthetic inputs to minimally change the target model's outputs, the reconstruction attack extracts information about the original training data. We assess the attack's efficacy by measuring the MSE and MAE between the reconstructed and original images. The result in Table 7 emphasizes the effectiveness of the UDP-FL framework, particularly with its Staircase mechanism, in mitigating data reconstruction attacks. This configuration consistently exhibits slightly higher MSE and MAE compared to both DP-SGD and UDP-FL with Laplace baseline, suggesting a more robust defense against reconstruction attacks.

Table 6: Evaluation on the SOTA InvGrad DRA [26] on CIFAR-10. The MSE, PSNR and SSIM for the Non-private method are 1.7104, 9.79, and 0.0751, respectively.

ϵ	DP-SGD			UDP-FL (Laplace baseline)			UDP-FL (Staircase)		
	MSE	PSNR	SSIM	MSE	PSNR	SSIM	MSE	PSNR	SSIM
2	2.2646	8.51	0.0195	2.2686	8.63	0.0573	**2.3399**	**8.37**	**0.0096**
4	2.2058	8.69	0.0414	2.1840	8.68	0.0629	**2.2405**	**8.39**	**0.0204**
6	2.1532	8.76	0.0417	2.1532	8.83	0.0692	**2.1910**	**8.54**	**0.0207**
8	2.1463	8.78	0.0519	2.1290	8.95	0.0746	**2.1832**	**8.73**	**0.0225**

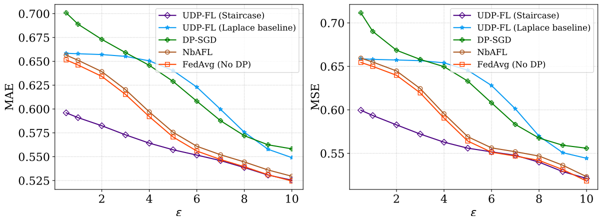
Table 7: Evaluation for data reconstruction attacks [23] on CIFAR-10.

ϵ	DP-SGD		UDP-FL (Laplace baseline)		UDP-FL (Staircase)	
	MSE	MAE	MSE	MAE	MSE	MAE
2	2.40	1.26	2.42	1.27	**2.44**	**1.28**
4	2.35	1.24	2.38	1.25	**2.40**	**1.26**
6	2.30	1.22	2.33	1.23	**2.35**	**1.24**
8	2.25	1.20	2.28	1.21	**2.30**	**1.22**

6 Discussion

Diverse DP Mechanisms. Our experiments mainly evaluate the Staircase, Laplace and Gaussian mechanisms with UDP-FL. As a universal DP-FL work, UDP-FL is flexible and can be extended to incorporate more advanced DP mechanisms, such as the Matrix Variate Gaussian (MVG) mechanism [9], R^{2}DP mechanism [59], DP Boosting [18, 20], and more. Moreover, UDP-FL is designed to be flexible and adaptable to various FL settings. One of the key aspects of this flexibility is the compatibility of our framework with other aggregation functions commonly used in FL, beyond the widely-used FedAvg and FedSGD algorithm [55, 56].

Support Diverse Aggregation Functions. Several aggregation functions have been proposed to address the limitations of FedAvg, such as dealing with non-IID data, mitigating the effects of stragglers, or improving convergence rates. Some of these alternative aggregation functions include Scaffold [40], FedMed [78], FedProx [82]. To show the possibility of integrating alternative aggregation functions into UDP-FL, we have conducted another experiments to evaluate it. We use a wind forecasting dataset [35], and train a simple CNN to predict hourly power generation up to 48 hours ahead at 7 wind farms. The baselines use NON-DP Scaffold aggregation functions in the FL frameworks. The sampling rate is 0.05, and the client number is 10, and the clipped gradient value is 10.



(a) MAE vs ϵ

(b) MSE vs ϵ

Figure 5: Accuracy results on FL with the Scaffold aggregation [40]. On small training epochs, UDP-FL with the Staircase mechanism can achieve better accuracy more quickly.

From Figure 5, we can see UDP-FL with the Staircase mechanism still outperforms the STOA DP-FL-based NbAFL [75]. Moreover, when ϵ is smaller than 5, the performance of UDP-FL is better than FedAvg without DP. One possible reason is that each class of noise is optimal for one specific metric and Staircase is designed for ℓ_1 and ℓ_2 metrics, and the noise may serve as a regularization to mitigate the unbiased distribution of clients' data.

DP Accounting Extensions. UDP-FL can also be extended for other accounting of differential privacy. Recent work has proposed using characteristic functions of the privacy loss random variable as an alternative approach for optimal privacy accounting [71, 74, 89]. This technique provides a natural composition similar to RDP while avoiding RDP's limitations. UDP-FL's architecture is flexible enough that it could potentially be extended to incorporate characteristic functions instead of just Rényi divergence. Specifically, the Harmonizer could be adapted to compute and track characteristic functions for each mechanism. The analytic Fourier accountant method could also replace the MA for conversion to (ϵ, δ)-DP. With these modifications, UDP-FL could achieve tighter accounting and flexibility by leveraging characteristic functions. The modularity of UDP-FL makes these extensions possible without changing the overall existing framework.

Future Works. A crucial direction for future work is the development of a comprehensive theoretical convergence analysis framework for UDP-FL. The primary challenge lies in developing a unified theoretical framework that can handle the diverse nature of different DP mechanisms within the same analysis. This would require novel mathematical tools to bridge the gap between privacy accounting methods and optimization theory. The framework would need to establish formal convergence bounds for UDP-FL across different DP mechanisms while analyzing the impact of mechanism switching on convergence behavior. Of particular interest is understanding convergence rates under varying privacy budgets and client distributions, especially in heterogeneous data scenarios. Such theoretical foundations would strengthen UDP-FL's applicability in practice by providing guarantees on performance across diverse FL environments.

7 Related Works

Differential Privacy and Rényi Differential Privacy. Differential privacy [16, 17] has been widely studied and applied to ensure privacy protection in data analysis and machine learning tasks. Since the introduction of DP-SGD [1], significant research has focused on tightly tracking privacy loss during training. The formalization of Rényi Differential Privacy (RDP) [57] has facilitated easier quantification of privacy loss, leading to further advancements in mechanisms to preserve RDP [28, 58, 59, 72, 74].

In FL, RDP has been employed to enhance privacy protection. Research [3, 68] explored RDP and shufflers [19, 30, 31] in FL systems for improved privacy and utility tradeoffs. Geyer et. al. [29] propose a client-sided DP approach for federated optimization. Bhowmick et al. [6] demonstrate scalable, locally private model training with minimal utility loss in large-scale image and language tasks. Li et al. [47] investigate the feasibility of applying differential privacy techniques to protect patient data in an FL setup.

Differentially Private Federated Learning. DP-FL is evolving rapidly. Li et al. [46] introduced FedMask, which protects both data and model privacy using gradient masking and perturbation. Wei et al. [76] tackled heterogeneity in FL with a personalized DP-FL framework that adapts to client characteristics while ensuring strong privacy. Xu et al. [80] proposed FedCORP, a communication-efficient personalized FL framework incorporating DP. Zhu et al. [87] developed a DP-FL algorithm with optimal sample complexity and theoretical guarantees.

Integrating DP-FL with other privacy-preserving technologies has also been explored. Chen et al. [13] combined DP with secure multi-party computation, and Fort et al. [22] examined privacy amplification by iteration in FL. Theoretical advances have emerged, with Liu et al. [50] improving privacy accounting methods for subsampled mechanisms, and Ding et al. [15] introducing an LDP-based approach enhancing both privacy and communication efficiency. Ding et al. [15] and Varun et al. [70] further enhanced DP-FL with local differential privacy approaches that improve communication, model accuracy, and robustness against attacks.

Recent efforts have also addressed data heterogeneity. Luo et al. [53] combined DP with multi-task learning for personalized models across diverse clients. Zhou et al. [86] refined privacy composition bounds, enabling better privacy loss tracking over multiple rounds. In the realm of large language models, Dagan et al. [14] proposed a DP-FL framework tailored to high-dimensional models. Sun et al. [67] introduced an adaptive DP mechanism that dynamically adjusts privacy levels, optimizing the privacy-utility trade-off. Zheng et al. [85] introduced federated f-DP, specifically designed for the federated setting, while Khanna et al. [42] presented a simple FL algorithm implementing DP for privacy across different institutions.

Privacy Attacks in FL. Chen et al. [12] proposed generative models for private data generation, focusing on the effectiveness of DP in defending against model inversion and GAN-based attacks. In the context of 5G networks, Liu et al. [51] developed a blockchain-based secure FL framework, enhancing privacy preservation for participants. Naseri et al. [60] conducted a comprehensive evaluation of Local and Central DP in FL, assessing their impact on privacy and robustness. Yang et al. [81] propose a robust distributed backdoor attack in federated learning inspired by secret sharing to evade detection and maintain attack efficacy. Sun and Lyu [66] proposed FEDMD-NFDP, a federated model distillation framework incorporating a Noise-Free DP mechanism, effectively eliminating the risk of white-box inference attacks. Kerkouche et al. [41] introduced a new FL scheme, offering a balance between robustness, privacy, bandwidth efficiency, and model accuracy. Chen et al. [11] developed a decentralized, privacy-preserving global model training protocol for FL in P2P networks. Hossain et al. [37] demonstrated how DP could be exploited for stealthy and persistent model poisoning attacks in FL. Feng et al. [21] evaluated user-level DP in FL, specifically in the context of speech-emotion recognition systems. Lastly, Wang et al. [73] proposed a platform-free proof of FL consensus mechanism, focusing on sustainable blockchains and privacy protection in FL models. Salem et al. [63] proposed a game-based framework to unify definitions and analysis of privacy inference risks. They use reductions between games to relate notions like MIA and RIA. Nie et al. [62] develop an efficient federated learning algorithm that is provably privacy-preserving and resilient to Byzantine adversaries.

8 Conclusion

In this paper, we introduced UDP-FL, a novel framework for DP-FL that addresses the critical challenge of optimizing the tradeoff between privacy and accuracy. A key innovation in UDP-FL is the integration of the Harmonizer, which dynamically selects the most appropriate DP mechanism for each client, considering their privacy requirements, etc. By harmonizing various DP mechanisms with Harmonizor, UDP-FL achieves tighter privacy bounds and faster convergence compared to SOTA methods. Our experimental results demonstrate the superior performance of UDP-FL in terms of both privacy guarantees and model accuracy. Furthermore, we proposed a mode connectivity-based method for analyzing the convergence of DP-FL models, providing valuable insights into the faster convergence. Through extensive evaluations, we also showed that UDP-FL exhibits substantial resilience against advanced privacy attacks, further validating the significant advancement in data protection in FL environments.

Acknowledgments

We sincerely thank all the anonymous reviewers for their valuable feedback and constructive suggestions. This work is supported in part by the National Science Foundation (NSF) under Grants No. CNS-2308730, CNS-2302689, CNS-2319277, CMMI-2326341, ECCS-2216926, CNS-2241713, CNS-2331302 and CNS-2339686. This work is also partially supported by the Cisco Research Award and the Synchrony Fellowship.

References

[1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. 2016. Deep learning with differential privacy. In *CCS*.

[2] Mohammed Adnan, Shivam Kalra, Jesse C Cresswell, Graham W Taylor, and Hamid R Tizhoosh. 2022. Federated learning and differential privacy for medical image analysis. *Scientific reports* (2022).

[3] Naman Agarwal, Peter Kairouz, and Ziyu Liu. 2021. The skellam mechanism for differentially private federated learning. *NIPS* (2021).

[4] Caridad Arroyo Arevalo, Sayedeh Leila Noorbakhsh, Yun Dong, Yuan Hong, and Binghui Wang. 2024. Task-Agnostic Privacy-Preserving Representation Learning for Federated Learning against Attribute Inference Attacks. In *AAAI*.

[5] Daniel J Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Javier Fernandez-Marques, Yan Gao, et al. 2020. Flower: A Friendly Federated Learning Research Framework. *arXiv preprint arXiv:2007.14390* (2020).

[6] Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers. 2018. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984* (2018).

[7] Vincent Bindschaedler, Reza Shokri, and Carl A. Gunter. 2017. Plausible Deniability for Privacy-Preserving Data Synthesis. *Proc. VLDB Endow.* (2017).

[8] Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramer. 2022. Membership inference attacks from first principles. In *S&P*. IEEE, 1897–1914.

[9] Thee Chanyaswad, Alex Dytso, H Vincent Poor, and Prateek Mittal. 2018. Mvg mechanism: Differential privacy under matrix-valued query. In *CCS*.

[10] Chen Chen, Lingjuan Lyu, Han Yu, and Gang Chen. 2022. Practical attribute reconstruction attack against federated learning. *IEEE Transactions on Big Data.* (2022).

[11] Qian Chen, Zilong Wang, Wenjing Zhang, and Xiaodong Lin. 2021. PPT: A Privacy-Preserving Global Model Training Protocol for Federated Learning in P2P Networks. *arXiv preprint arXiv:2101.02281* (2021).

[12] Qingrong Chen, Chong Xiang, Minhui Xue, Bo Li, Nikita Borisov, Dali Kaarfar, and Haojin Zhu. 2018. Differentially Private Data Generative Models. *arXiv preprint arXiv:1812.02274* (2018).

[13] Xiaochen Chen, Yongqiang Li, Yuncheng Wang, Jiacheng Liu, and Kim-Kwang Raymond Choo. 2023. Federated Learning with Differential Privacy and Secure Multiparty Computation. *IEEE TIFS* (2023).

[14] Idan Dagan, Tomer Gafni, Oleksii Romanenko, and Idit Keidar. 2023. Federated Learning of Large Language Models with Differential Privacy. *arXiv preprint arXiv:2306.10635* (2023).

[15] Hao Ding, Xiangyu Gao, Ming Li, and Qian Tang. 2023. Differentially Private Federated Learning with Local Randomization and Adaptive Optimization. *IEEE Transactions on Information Forensics and Security* (2023).

[16] Cynthia Dwork. 2006. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33.* Springer.

[17] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *TCC 2006.* Springer.

[18] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. 2010. Boosting and Differential Privacy. In *FOCS*. 51–60.

[19] Vitaly Feldman, Audra McMillan, and Kunal Talwar. 2023. Stronger privacy amplification by shuffling for Rényi and approximate differential privacy. In *SODA*. SIAM.

[20] Shuya Feng, Meisam Mohammady, Han Wang, Xiaochen Li, Zhan Qin, and Yuan Hong. 2024. DPI: Ensuring Strict Differential Privacy for Infinite Data Streaming. In *IEEE S&P*.

[21] Tiantian Feng, Raghuveer Peri, and Shrikanth Narayanan. 2022. User-Level Differential Privacy Against Attribute Inference Attack of Speech Emotion Recognition in Federated Learning. *arXiv preprint arXiv:2202.01684* (2022).

[22] Stanislav Fort, Stefanie Günther, Zoltán Szabó, Andrew McMillan, and Vitaly Feldman. 2023. Privacy Amplification by Iteration in Federated Learning. *arXiv preprint arXiv:2305.15046* (2023).

[23] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. 2015. Model inversion attacks that exploit confidence information and basic countermeasures. In *CCS*. 1322–1333.

[24] Jie Fu, Yuan Hong, Xinpeng Ling, Leixia Wang, Xun Ran, Zhiyu Sun, Wendy Hui Wang, Zhili Chen, and Yang Cao. 2024. Differentially Private Federated Learning: A Systematic Review. *CoRR* abs/2405.08299 (2024).

[25] Karan Ganju, Qi Wang, Wei Yang, Carl A Gunter, and Nikita Borisov. 2018. Property Inference Attacks on Fully Connected Neural Networks using Permutation Invariant Representations. In *CCS*.

[26] Jonas Geiping, Hartmut Bauermeister, Hannah Dröge, and Michael Moeller. 2020. Inverting gradients-how easy is it to break privacy in federated learning? *NeurIPS* 33 (2020), 16937–16947.

[27] Quan Geng and Pramod Viswanath. 2014. The optimal mechanism in differential privacy. In *IEEE ISIT*. IEEE, 2371–2375.

[28] Joseph Geumlek, Shuang Song, and Kamalika Chaudhuri. 2017. Renyi differential privacy mechanisms for posterior sampling. *NeurIPS* 30 (2017).

[29] Robin C Geyer, Tassilo Klein, and Moin Nabi. 2017. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557* (2017).

[30] Antonious M Girgis, Deepesh Data, Suhas Diggavi, Peter Kairouz, and Ananda Theertha Suresh. 2021. Shuffled model of federated learning: Privacy, accuracy and communication trade-offs. *IEEE journal on selected areas in information theory* 2, 1 (2021), 464–478.

[31] Antonious M Girgis, Deepesh Data, Suhas Diggavi, Ananda Theertha Suresh, and Peter Kairouz. 2021. On the rényi differential privacy of the shuffle model. In *ACM SIGSAC Conference on Computer and Communications Security*.

[32] Haimei Gong, Liangjun Jiang, Xiaoyang Liu, Yuanqi Wang, Omary Gastro, Lei Wang, Ke Zhang, and Zhen Guo. 2023. Gradient leakage attacks in federated learning. *Artificial Intelligence Review* 56, Suppl 1 (2023), 1337–1374.

[33] Jamie Hayes, Luca Melis, George Danezis, and Emiliano De Cristofaro. 2019. LOGAN: Membership Inference Attacks Against Generative Models. In *Proceedings on Privacy Enhancing Technologies*, Vol. 2019. 133–152.

[34] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *CVPR*. 770–778.

[35] Tao Hong, Pierre Pinson, Shu Fan, Hamidreza Zareipour, Alberto Troccoli, and Rob Hyndman. 2016. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. *International Journal of Forecasting* (2016).

[36] Yuan Hong, Jaideep Vaidya, Haibing Lu, Panagiotis Karras, and Sanjay Goel. 2015. Collaborative Search Log Sanitization: Toward Differential Privacy and Boosted Utility. *TDSC* (2015).

[37] Md Tamjid Hossain, Shafkat Islam, Shahriar Badsha, and Haoting Shen. 2021. DeSMP: Differential Privacy-exploited Stealthy Model Poisoning Attacks in Federated Learning. *arXiv preprint arXiv:2102.03070* (2021).

[38] Rui Hu, Yuanxiong Guo, Hongning Li, Qingqi Pei, and Yanmin Gong. 2020. Personalized federated learning with differential privacy. *IEEE Internet of Things Journal* (2020).

[39] Xiao Jin, Pin-Yu Chen, Chia-Yi Hsu, Chia-Mu Yu, and Tianyi Chen. 2021. Cafe: Catastrophic data leakage in vertical federated learning. *NeurIPS* (2021).

[40] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. 2020. Scaffold: Stochastic controlled averaging for federated learning. In *ICML*. PMLR.

[41] Raouf Kerkouche, Gergely Ács, and Claude Castelluccia. 2020. Federated Learning In Adversarial Settings. *arXiv preprint arXiv:2001.05641* (2020).

[42] Amol Khanna, Vincent Schaffer, Gamze Gürsoy, and Mark Gerstein. 2022. Privacy-preserving Model Training for Disease Prediction Using Federated Learning with Differential Privacy. In *EMBC*.

[43] Alex Krizhevsky, Vinod Nair, and Geoffrey Hinton. 2009. *CIFAR-10 (Canadian Institute for Advanced Research)*. Technical Report. University of Toronto.

[44] Sammy Kus. 2022. Medical MNIST. *Mendeley Data* (2022).

[45] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* (1998).

[46] Jing Li, Daoxin Lin, Huaxin Shu, Yuan Wang, and Xindong Liu. 2023. Federated Learning with Data and Model Privacy Preservation. *arXiv preprint arXiv:2306.08005* (2023).

[47] Qinbin Li, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, and Bingsheng He. 2021. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *IEEE Transactions on Knowledge and Data Engineering* (2021).

[48] Qinbin Li, Zhaomin Wu, Zeyi Wen, and Bingsheng He. 2020. Privacy-preserving gradient boosting decision trees. In *AAAI*.

[49] Zhuohang Li, Jiaxin Zhang, Luyang Liu, and Jian Liu. 2022. Auditing privacy defenses in federated learning via generative gradient leakage. In *CVPR*.

[50] Xinwei Liu, Penghui Zhao, Chao Li, and Jiawen Zhang. 2023. Tighter Privacy Bounds for Subsampled Mechanisms in Federated Learning. *arXiv preprint arXiv:2304.02140* (2023).

[51] Yi Liu, Jialiang Peng, Jiawen Kang, Abdullah M. Iliyasu, Dusit Niyato, and Ahmed A. Abd El-Latif. 2020. A Secure Federated Learning Framework For 5G Networks. *arXiv preprint arXiv:2001.05637* (2020).

[52] Yugeng Liu, Rui Wen, Xinlei He, Ahmed Salem, Zhikun Zhang, Michael Backes, Emiliano De Cristofaro, Mario Fritz, and Yang Zhang. 2022. ML-Doctor: Holistic Risk Assessment of Inference Attacks Against Machine Learning Models. In *USENIX Security 22*. USENIX Association.

[53] Yitong Luo, Jiaheng Zhang, Qiaoyu Tan, Yu-Xiang Wang, and Qiang Yang. 2023. Federated Multi-Task Learning with Differential Privacy. *arXiv preprint arXiv:2307.03449* (2023).

[54] Lingjuan Lyu and Chen Chen. 2021. A novel attribute reconstruction attack in federated learning. *arXiv preprint arXiv:2108.06910* (2021).

[55] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. 1273–1282.

[56] H. Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. 2018. Learning Differentially Private Recurrent Language Models. In *International Conference on Learning Representations*.

[57] Ilya Mironov. 2017. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*. IEEE.

[58] Ilya Mironov, Kunal Talwar, and Li Zhang. 2019. R\'enyi differential privacy of the sampled gaussian mechanism. *arXiv preprint arXiv:1908.10530* (2019).

[59] Meisam Mohammady, Shangyu Xie, Yuan Hong, Mengyuan Zhang, Lingyu Wang, Makan Pourzandi, and Mourad Debbabi. 2020. R2dp: A universal and automated approach to optimizing the randomization mechanisms of differential privacy for utility metrics with no known optimal distributions. In *CCS'20*.

[60] Mohammad Naseri, Jamie Hayes, and Emiliano De Cristofaro. 2020. Local and central differential privacy for robustness and privacy in federated learning. *arXiv preprint arXiv:2009.03561* (2020).

[61] Milad Nasr, Reza Shokri, and Amir Houmansadr. 2019. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *IEEE S&P*. IEEE.

[62] Chenfei Nie, Qiang Li, Yuxin Yang, Yuede Ji, and Binghui Wang. 2024. Efficient Byzantine-Robust and Provably Privacy-Preserving Federated Learning. *arXiv preprint arXiv:2407.19703* (2024).

[63] Ahmed Salem, Giovanni Cherubin, David Evans, Boris Köpf, Andrew Paverd, Anshuman Suri, Shruti Tople, and Santiago Zanella-Béguelin. 2023. SoK: Let the privacy games begin! A unified treatment of data inference privacy in machine learning. In *IEEE S&P*. IEEE.

[64] Lu Shi, Jiangang Shu, Weizhe Zhang, and Yang Liu. 2021. HFL-DP: Hierarchical Federated Learning with Differential Privacy. In *GLOBECOM*.

[65] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. 2017. Membership inference attacks against machine learning models. In *IEEE S&P*.

[66] Lichao Sun and Lingjuan Lyu. 2020. Federated Model Distillation with Noise-Free Differential Privacy. *arXiv preprint arXiv:2012.04187* (2020).

[67] Lichao Sun, Jianwei Qian, and Xiang Chen. 2023. Federated Learning with Adaptive Differential Privacy. *IEEE TDSC* (2023).

[68] Aleksei Triastcyn and Boi Faltings. 2019. Federated learning with bayesian differential privacy. In *2019 IEEE International Conference on Big Data*. IEEE.

[69] Tim van Erven and Peter Harremos. 2014. Rényi Divergence and Kullback-Leibler Divergence. *IEEE Transactions on Information Theory* (2014).

[70] Matta Varun, Shuya Feng, Han Wang, Shamik Sural, and Yuan Hong. 2024. Towards Accurate and Stronger Local Differential Privacy for Federated Learning with Staircase Randomized Response. In *CODASPY*. 307–318.

[71] Hua Wang, Sheng Gao, Huanyu Zhang, Milan Shen, and Weijie J Su. 2022. Analytical composition of differential privacy via the edgeworth accountant. *arXiv preprint arXiv:2206.04236* (2022).

[72] Han Wang, Jayashree Sharma, Shuya Feng, Kai Shu, and Yuan Hong. 2022. A Model-Agnostic Approach to Differentially Private Topic Mining. In *KDD'22*.

[73] Yuntao Wang, Haixia Peng, Zhou Su, Tom H Luan, Abderrahim Benslimane, and Yuan Wu. 2022. A Platform-Free Proof of Federated Learning Consensus Mechanism for Sustainable Blockchains. *arXiv preprint arXiv:2202.01884* (2022).

[74] Yu-Xiang Wang, Borja Balle, and Shiva Prasad Kasiviswanathan. 2019. Subsampled Rényi differential privacy and analytical moments accountant. In *AISTATS*.

[75] Kang Wei, Jun Li, Ming Ding, Chuan Ma, Howard H. Yang, Farhad Farokhi, Shi Jin, Tony Q. S. Quek, and H. Vincent Poor. 2020. Federated Learning With Differential Privacy: Algorithms and Performance Analysis. *TIFS* (2020).

[76] Kang Wei, Jun Li, Chuan Ma, Ming Ding, Wen Chen, Jun Wu, Meixia Tao, and H. Vincent Poor. 2023. Personalized Federated Learning With Differential Privacy and Convergence Guarantee. *IEEE TIFS* (2023).

[77] Yuxin Wen, Arpit Bansal, Hamid Kazemi, Eitan Borgnia, Micah Goldblum, Jonas Geiping, and Tom Goldstein. 2022. Canary in a Coalmine: Better Membership Inference with Ensembled Adversarial Queries. *arXiv preprint arXiv:2210.10750* (2022).

[78] Xing Wu, Zhaowang Liang, and Jianjia Wang. 2020. Fedmed: A federated learning framework for language modeling. *Sensors* (2020).

[79] Xiaokui Xiao, Guozhang Wang, and Johannes Gehrke. 2010. Differential privacy via wavelet transforms. *IEEE Transactions on knowledge and data engineering* 23, 8 (2010), 1200–1214.

[80] Mingzhe Xu, Gen Li, Jiaojiao Zheng, Qian Xiao, and Jian Liu. 2022. Federated learning with privacy-preserving and model protection. *Inf. Sci.* (2022).

[81] Yuxin Yang, Qiang Li, Yuede Ji, and Binghui Wang. 2025. A Secret Sharing-Inspired Robust Distributed Backdoor Attack to Federated Learning. *ACM Transactions on Privacy and Security* (2025).

[82] Xiaotong Yuan and Ping Li. 2022. On convergence of FedProx: Local dissimilarity invariant bounds, non-smoothness and beyond. *NeurIPS* (2022).

[83] Jingwen Zhang, Jiale Zhang, Junjun Chen, and Shui Yu. 2020. Gan enhanced membership inference: A passive local attack in federated learning. In *ICC*. IEEE.

[84] Zhifei Zhang, Yang Song, and Hairong Qi. 2017. Age progression/regression by conditional adversarial autoencoder. In *CVPR*. 5810–5818.

[85] Qinqing Zheng, Shuxiao Chen, Qi Long, and Weijie Su. 2021. Federated f-differential privacy. In *International Conference on Artificial Intelligence and Statistics*.

[86] Wei Zhou, Jiacheng Mao, Ankai Xu, Shutian Wan, Qianqian Gong, Xin Li, Yi Wen, and Jian Li. 2023. Tight Privacy Composition Analysis for Federated Learning. In *ICML*.

[87] Haonan Zhu and Ruoxi Ding. 2023. Efficient Differentially Private Federated Learning for Heterogeneous Data. *arXiv preprint arXiv:2306.08750* (2023).

[88] Ligeng Zhu, Zhijian Liu, and Song Han. 2019. Deep leakage from gradients. *NeurIPS* 32 (2019).

[89] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. 2022. Optimal Accounting of Differential Privacy via Characteristic Function. In *AISTATS*.