FedDyna: Privacy-Preserving Federated Learning with Dynamic Noise Adaptation and Structural Bias Alignment for Non-IID Environments

Guijuan Wang[1,2,†], Zhiyu Zuo[1,2,†], Anming Dong[1,2,*],
Jiguo Yu[3], Yanqi Zhao[4]

{ guijuan_wang@126.com, zzhy20010925@163.com, donganming@gmail.com,
jiguoyu@sina.com, zhaoyanqi2019@163.com }

[1] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

[2] Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan 250353, China

[3] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

[4] School of Cyberspace Security, Xi'an University of Posts and Telecommunications, Xi'an, China

[†] These authors contributed equally. [*] Corresponding author.

Abstract. In Internet of Things (IoT) scenarios, federated learning (FL) facilitates data value mining through distributed collaborative learning while preserving user privacy. However, the performance of existing privacy-preserving methods in federated learning systems is greatly reduced due to the problem of non-independent and identically distributed (non-IID) data. In this paper, we study the privacy preservation problem of FL in non-IID data environments and propose FedDyna, a novel framework that uses an adaptive noise injection mechanism to enhance privacy. A theoretical analysis reveals an upper bound on the convergence of additive noise under non-IID and proves the limitations of the traditional static noise strategy. FedDyna uses a decaying noise injection method that dynamically correlates with model biases to synchronize structured privacy preservation of the local parameter space with global consistency maintenance in the client training phase. Meanwhile, through a locally drifting tracking mechanism constrained by matrix trace, we explicitly decouple model parameter deviations and perform dynamic correction, combined with gradient variance regularization to suppress divergence in local updates. Finally, through simulation experiments conducted on two benchmark datasets, MNIST and CIFAR-10, we effectively demonstrate the significant divergence between dynamic and static noise under non-IID settings.

Keywords: Federated Learning, Non-IID, IoT, Privacy Protection

IIKI 2025, December 18-21, Chengdu, People's Republic of China

DOI 10.4108/eai.18-12-2025.2365254

1 Introduction

With the advancement of AI, machine learning has been applied in IoT[1]. Servers train high-accuracy models via IoT device data, but this violates data privacy [2][3]. To address this, Google proposed FL—a distributed learning technique where the server aggregates model updates from IoT devices (without raw data sharing) to train a global model, mitigating privacy threats. However, FL's distributed nature inevitably causes non-IID problems, which severely degrade system performance[4].

In FL, data distribution across clients no longer meets the IID assumption, posing dual challenges of convergence degradation and privacy preservation. Under conventional IID, static noise injection strategies based on DP [5, 6] balance privacy and accuracy effectively. But non-IID data heterogeneity leads to significant distributional divergence, resulting in two critical issues: (1) enhanced deviation between local model updates and the global model, where static noise fails to adapt to dynamic model biases; (2) attackers exploiting data heterogeneity to infer sensitive information, amplifying privacy leakage [7]. Theoretical analysis shows that non-IID data increases DP budget consumption by over 32% [8], reducing the effectiveness of existing privacy mechanisms.

Current FL research suffers from a disconnect between optimization and privacy protection. For example, FedProx [9] introduces a proximal regularization term to align local and global parameters but limits global optimum exploration; Scaffold [10] corrects bias via gradient-controlled variables but ignores privacy needs. Conversely, privacy-preserving FL mechanisms incorporating mobility awareness, dynamic sparsification, random projection, or client collaboration [11, 12, 13] operate beyond the IID data assumption, effectively mitigating dynamic privacy leakage risks in heterogeneous data scenarios. Thus, designing dynamic privacy-preserving mechanisms for non-IID has become an urgent FL challenge. In response to this problem, this paper presents FedDyna—an innovative FL framework. Our primary contributions are outlined below:

- ***Dynamic Bias Correction and Gradient Optimization:*** Proposes a matrix trace-constrained dynamic bias tracking mechanism, decoupling model parameter deviations via local drift variables. Integrated with gradient variance regularization, it mitigates non-IID-induced gradient divergence through high-dimensional parameter structured alignment.

- ***Adaptive Privacy-Preservation:*** Designs an exponentially decaying noise injection strategy dynamically linked to model bias, coupling noise intensity with local deviations for self-adjusting privacy protection. Under Rényi differential privacy, structured Gaussian noise ensures data privacy and effective parameter updates.

- ***Convergence Analysis and Experiments:*** Conducts non-convex convergence analysis via augmented Lagrangian method, proving dynamic noise's controllable impact on convergence and revealing privacy-optimization trade-off. Experiments on MNIST and CIFAR-10 verify the approach's effectiveness.

2 Related Work

2.1 Non-IID Problems in FL

In recent years, FL has garnered significant attention as a distributed machine learning paradigm [14, 15, 16]. The foundational algorithms FedAvg [17] employs a weighted parameter averaging strategy for multi-client parameter aggregation, which guarantees approaching convergence in IID settings. However, Woodworth et al. [18] theoretically demonstrated the significant impact of data heterogeneity on convergence performance, revealing that client update bias induced by non-IID constitutes a key factor degrading convergence rates [19]. Such distributional can trigger gradient dispersion, optimization direction drift, and even lead to failure in convergence guarantees. To ease client update variance, existing studies have attempted to constrain local models through regularization techniques. For instance, FedProx introduces a proximal regularization term for the global model, suppressing gradient divergence by minimizing the distance between local and global parameters. However, this approach overlooks the fundamental discrepancy between local and global models, potentially limiting model performance. Subsequent research has further explored personalized local objective design. For instance, Scaffold customizes client-specific gradient corrections to compensate for model bias, while FedDyn [20] proposes dynamic regularization strategies to align global and local solutions.

Another type of research focuses on server aggregation optimization. The work [21] proposed a reinforcement learning-based adaptive aggregation framework to balance heterogeneity and communication efficiency while enhancing fairness and robustness. The work [22] achieved provable convergence and bilateral acceleration via a dual hypergradient descent-based online learning rate adaptation mechanism, improving performance in both global and local updates. Yang et al. achieved linear speedup through a bilateral learning rate mechanism, improving performance in both local and global updates. Although these methods have advanced convergence and communication efficiency, accumulated parameter bias may still lead to suboptimal solutions. The proposed FedDyna method addresses data heterogeneity by tracking and bridging local deviations, effectively decoupling local models from the global model through dynamic bias correction. This provides a novel method to resolving non-IID challenges in federated learning.

2.2 Privacy Protection

With the widespread adoption of FL systems, research on security attacks against this framework has emerged as a critical area of focus. Currently identified attack vectors primarily include inference attacks and poisoning attacks [23, 24, 25]. Regarding inference attacks, Melis et al. put forward a membership inference attack approach relying on non-zero gradients within the embedding layer of deep learning-based natural language processing models, while Hitaj et al. designed an active inference attack framework using generative adversarial networks (GANs) that can reconstruct private samples from target clients. Based on a systematic analysis of FL privacy leakage mechanisms, work [26] constructed a multi-layer inference attack model utilizing gradient information. Regarding poisoning attacks, data poisoning attacks are implemented by modifying training data labels or features, while model poisoning attacks insert hidden backdoors into partial clients to

trigger anomalies in the global model. Work pioneered modeling the interplay between training loss and attack efficacy as an adversarial game, proposing an innovative poisoning attack that can evade detection mechanisms. Regarding privacy protection, secure multi-party computation (SMC) [27] and differential privacy (DP) represent two major mainstream techniques. SMC provides strict security guarantees through complex computational protocols, but its excessive computational overhead limits applications on mobile devices. McMahan et al. were the first to apply DP to privacy protection in recurrent language models. Agarwal et al. optimized communication costs under specific DP constraints , while work enhanced privacy protection through a two-phase noise injection mechanism. Notably, all existing solutions assume IID data, which significantly differs from real-world application scenarios.

3 Preliminaries

3.1 Local Drift in FL

FedAvg, a foundational federated learning (FL) paradigm, enables collaborative training of a global model across multiple participating clients under the coordination of a central server, with inherent data privacy preservation [6]. In the FL framework, we consider a total of N clients, each equipped with a private local data repository D_i that is not shared with other parties. Our primary goal is to learn an optimal global model parameter w^* by leveraging the aggregated global dataset D—formed as the union of all local datasets D_i (i.e., $D = \bigcup_{i\in[N]} D_i$)—and the corresponding optimization objective is formulated as follows:

$$w^* = \text{argmin}_\theta \mathscr{L}(\theta) = \sum_{i=1}^{N} \frac{|D_i|}{|D|} \mathscr{E}_i(\theta) \tag{1}$$

Where, $\mathscr{L}(\theta)$ denotes the empirical risk over the aggregated global dataset, $|D_i|$ stands for the sample size of the local dataset at client i, and $\mathscr{E}_i(\theta) = \mathbb{E}_{(x,y)\in D_i}\left[l\left(\theta;(x,y)\right)\right]$ refers to the local empirical loss for client i, defined as the expected value of the loss function $l(\theta;(x,y))$ over its local data repository D_i. To safeguard data privacy throughout the collaborative training process, no client is permitted to disclose raw data to other clients or the central server. Each client trains its local model independently, but the non-IID nature of data induces local drift [28] between client models and the global model. If this drift is ignored, the server will aggregate a deviated global model, thereby intensifying the challenges of data heterogeneity in FL. Under highly non-IID data distributions, FedAvg exhibits significant performance degradation, demonstrating that its failure to address local drift systematically biases the global model.

3.2 Differential Privacy

DP is a reliable privacy protection method intended to guarantee that personal data information remains uncompromised in the course of data analysis. Let $\mathscr{D}$ be the set consisting of all possible datasets. For two neighboring datasets D and D'(where the datasets differ by only one sample),

a randomized algorithm $\mathscr{A}$ is deemed to be ε-differential privacy if, for any set of outputs $S \subseteq \text{Range}(\mathscr{A})$, the following condition is satisfied:

$$\Pr[\mathscr{A}(D) \in S] \leq e^{\varepsilon} \cdot \Pr[\mathscr{A}(D') \in S] \tag{2}$$

ε controls privacy protection strength: smaller ε brings higher privacy but lower data usability, while larger ε weakens privacy but improves usability. In FL, traditional DP mechanisms use static noise injection (e.g., adding Gaussian noise $\mathscr{N}(0, \sigma^2 I)$ to gradient $\nabla\theta$, resulting in $\nabla\theta' = \nabla\theta + \mathscr{N}(0, \sigma^2 I)$, where σ is noise standard deviation and I is identity matrix). However, under non-IID data, significant client-side data distribution differences make static noise unable to dynamically adapt to actual data and model conditions—leading to excessive noise (impairing model convergence/performance) or insufficient noise (failing to ensure privacy).

Algorithm 1 FedDyna

Input: Randomly initialize the global model parameters$w$, set the number of training rounds T, the number of clients N, initialize the local drift variables, the learning rate η, the number of local training batches K, the initial noise intensity σ_0, and the decay factor α.
Output: The trained global model w
for client t = 1,2,...,T **do**;
    Sample the set of active clients $C_t \subseteq [N]$
    for each client $i \in C_t$ in parallel **do**;
        Set the local model parameters $\theta_i = w$
        for $k = 1, 2, \ldots, K$ **do**;
            // Calculate the structural deviation tracking term
            Calculate the deviation tensor Δ_i and optimize it according to the matrix trace constraint: $\mathscr{R}_i = \alpha \cdot Tr((\Delta_i + \theta_i - w)^T(\Delta_i + \theta_i - w))$
            // Calculate the gradient consistency optimization term
            Calculate the gradient regularization term $\mathscr{G}_i = \mu \cdot \left(1 - \frac{\nabla \varepsilon_i \cdot \nabla \varepsilon_g}{\|\nabla \varepsilon_i\|_2 \cdot \|\nabla \varepsilon_g\|_2}\right)$
            // Calculate the privacy enhancement term
            Calculate the noise standard deviation $\sigma(\|\Delta_i\|_F) = \frac{\sigma_0}{1+\alpha \cdot \|\Delta_i\|_F}$
            Calculate the noise injection term $\mathscr{N}_i(\Delta_i) = \sigma(\|\Delta_i\|_F) \cdot \Delta_i \odot \mathscr{N}(0, I)$
            // Update the local model parameters
            $\theta_i = \theta_i - \eta(\nabla \mathscr{E}_i + \nabla \mathscr{R}_i + \nabla \mathscr{G}_i + \nabla \mathscr{N}_i)$
            Set the local gradient drift $\Delta_i = w - \theta_i$
            Update the local drift $h_i = h_i + \Delta_i$
        end
        // Aggregate and update the global model
        $w = \sum_{i=1}^{N}(\theta_i + h_i)$
    end
Return w

4 Suggested Methods

4.1 Incorporation of dynamic noise

In the context of traditional federated learning facing numerous challenges, we propose a federated learning algorithm called FedDyna, which integrates a deep fusion structural bias tracking

mechanism with a privacy enhancement mechanism. This method aims to achieve precise alignment of the parameter space and efficient privacy protection during the client training phase.

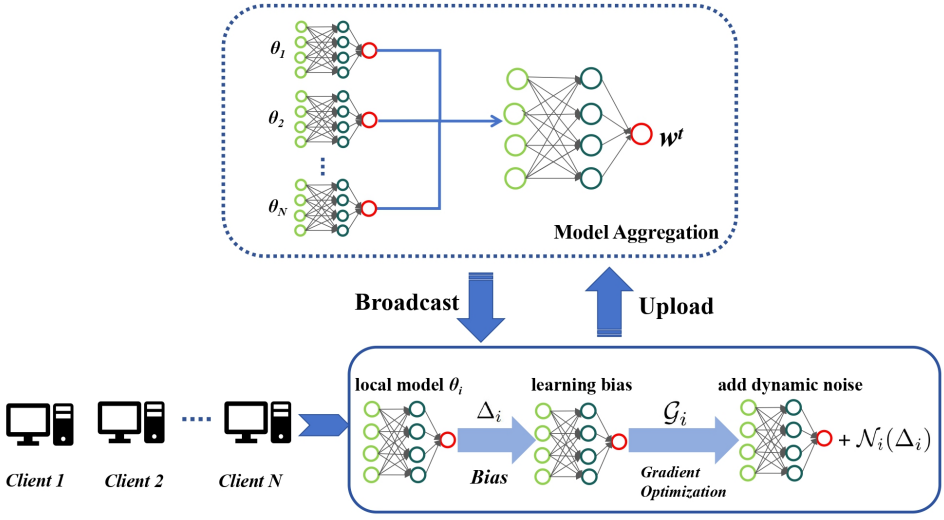

Fig. 1. FedDyna algorithm. It includes several key processes: gradient bias, consistency optimization, and the addition of dynamic noise.

4.1.1 Structure Bias Tracking Mechanism

To address the parameter space mismatch issue caused by non-IID data, we propose a bias tracking mechanism based on tensor decomposition. We introduce a learnable bias tensor $\Delta_i \in \mathbb{R}^{d\times m}$, whose dimensions are consistent with the model parameter matrix. The tensor is optimized through the matrix trace constraint:

$$\mathscr{R}_i = \alpha \cdot \mathrm{Tr}\left((\Delta_i + \theta_i - w)^T (\Delta_i + \theta_i - w)\right) \tag{3}$$

where *Tr(.)* represents the matrix trace operation. By minimizing the sum of the diagonal elements of the matrix, it effectively captures the low-rank characteristics of the parameter matrix. Specifically, the trace constraint allows the bias to be decomposed into a coordinated adjustment of the row space and column space, enabling the local model to adapt to local data characteristics while preserving the global structure. For example, in convolutional layers, Δ_i can capture the response differences between different channels, achieving dynamic balance among the channels through the trace constraint.

4.1.2 Gradient Consistency Optimization

At the same time, we designed a gradient regularization strategy that enforces local gradients to align with the global gradient direction using cosine similarity loss, denoted as:

$$\mathscr{G}_i = \mu \cdot \left(1 - \frac{\nabla \mathscr{E}_i \cdot \nabla \mathscr{E}_g}{\|\nabla \mathscr{E}_i\|_2 \cdot \|\nabla \mathscr{E}_g\|_2}\right) \tag{4}$$

This loss term transforms the directional difference of the gradients into a penalty value within the range of 0 to 1. When the directions are perfectly aligned, $\mathscr{G}_i$=0; when the directions are perpendicular, $\mathscr{G}_i = 2\mu$.

4.1.3 Privacy-Enhancing Mechanisms

In order to strengthen privacy protection for clients whose local models exhibit substantial deviations from the global model, we put forward a deviation-aware dynamic noise injection mechanism. The mathematical formulation for noise injection is defined as follows:

$$\mathscr{N}_i(\Delta_i) = \sigma(\|\Delta_i\|_F) \cdot \Delta_i \odot \mathscr{N}(0, I) \tag{5}$$

where $\sigma(\|\Delta_i\|_F) = \frac{\sigma_0}{1+\alpha \cdot \|\Delta_i\|_F}$, σ_0 represents the initial noise intensity, and α is the decay factor.

4.1.4 The Client Optimization Objective

Combining the above mechanisms, the optimization objective function for the client is:

$$\mathscr{L}_i = \mathscr{E}_i(\theta) + \mathscr{R}_i(\Delta, \theta; w) + \mathscr{G}_i(\mathscr{E}) + \mathscr{N}_i(\Delta_i) \tag{6}$$

where $\varepsilon_i(\theta)$ represents the experience loss term; $\mathscr{R}_i$ denotes the structural alignment term, which achieves low-rank alignment in the parameter space via matrix trace constraints; $\mathscr{G}_i$ is the gradient regularization term; and $\mathscr{N}_i(\Delta_i)$ is the privacy enhancement term. The objective function is optimized alternately with respect to θ_i and Δ_i, enabling adaptive adjustment of the local model in the feature space. The overall framework diagram is shown in Figure 1. And we present the proposed FedDyna algorithm in Algorithm 1.

5 Proof

5.1 Proof of Privacy Budget

5.1.1 One-step Rényi Differential Privacy

Assume that in each round of training, the noise injection mechanism on the client side satisfies α-Rényi Differential Privacy (α-RDP). For adjacent datasets D and D', the one-step RDP is defined as:

$$\mathscr{R}_\alpha(\sigma_t, D, D') = \frac{1}{\alpha - 1} \ln \mathbb{E}\left[\left(\frac{d\mathscr{A}_t(D')}{d\mathscr{A}_t(D)}\right)^{\alpha-1}\right] \leq \varepsilon_t \tag{7}$$

where, ε_t represents the RDP privacy budget for the t-th round, and $\mathscr{A}_t$ denotes the random algorithm for the t-th round (i.e., the dynamic noise injection process).

5.1.2 The Cumulative RDP for Multiple Rounds of Training

If the noise injection process in each round of training is independent, then the cumulative RDP for T rounds of training is:

$$\mathscr{R}_\alpha^{\text{total}} = \sum_{t=1}^{T} \varepsilon_t \tag{8}$$

This property arises from the sub-additivity of RDP, where the total RDP for multiple independent steps is the sum of the RDPs of each step.

5.1.3 The Conversion from RDP to (ε, δ)-DP.

Based on the properties of RDP, we can convert the cumulative RDP to standard (ε, δ)-DP by choosing an appropriate α. The conversion formula is:

$$\varepsilon \leq \mathscr{R}_\alpha^{\text{total}} + \frac{\ln(1/\delta)}{2(\alpha-1)} \tag{9}$$

This formula is derived based on exponential tilting and Markov's inequality. We define the likelihood ratio $\mathrm{L} = \frac{\mathrm{d}\mathscr{A}(D')}{\mathrm{d}\mathscr{A}(D)}$, and according to equation (9), we obtain:

$$\mathbb{E}[L^{\alpha-1}] \leq e^{(\alpha-1)\mathscr{R}_\alpha^{\text{total}}}$$

According to Markov's inequality, we obtain:

$$\Pr[L \geq e^{\varepsilon}] \leq \frac{\mathbb{E}[L^{\alpha-1}]}{e^{\varepsilon(\alpha-1)}} \leq e^{(\alpha-1)\mathscr{R}_\alpha^{\text{total}} - \varepsilon(\alpha-1)} \tag{10}$$

let $\delta = e^{(\alpha-1)\mathscr{R}_\alpha^{\text{total}} - \varepsilon(\alpha-1)}$, we obtain:

$$\varepsilon = \mathscr{R}_\alpha^{\text{total}} + \frac{\ln(1/\delta)}{2(\alpha-1)} \tag{11}$$

5.1.4 The RDP Calculation for Dynamic Noise Mechanisms

For the updated dynamic noise mechanism $\sigma_t = \frac{\sigma_0}{1+\alpha\cdot\|\Delta_i\|_F}$, assuming that the noise in each round follows a Gaussian distribution $\mathscr{N}(0, \sigma_t^2 I)$, the one-step RDP is given by:

$$\mathscr{R}_\alpha(\sigma_t) = \frac{\sigma_t^2}{2} \cdot \frac{\alpha}{\alpha-1} \cdot \frac{1}{(1+\alpha\|\Delta_t\|_F)^2} \tag{12}$$

According to the definition of the likelihood ratio, the outputs for the neighboring datasets D and D' are $x+\varepsilon$ and $x'+\varepsilon$, respectively, and the likelihood ratio is denoted as $L = \exp\left(\frac{\|x+\varepsilon\|^2 - \|x'+\varepsilon\|^2}{2\sigma_t^2}\right)$. By using the moment-generating function of Gaussian noise, we compute the expectation $\mathbb{E}[L^{\alpha-1}]$ and simplify $||x'-x|| \leq 1$ through worst-case analysis to obtain

$$\mathbb{E}[L^{\alpha-1}] \leq \exp\left(\frac{(\alpha-1)^2\sigma_t^2}{2}\right) \tag{13}$$

Based on the definition of Rényi differential privacy, we ultimately arrive at $\mathscr{R}_\alpha(\sigma_t) = \frac{1}{\alpha-1}\ln\mathbb{E}[L^{\alpha-1}] = \frac{\sigma_t^2}{2}\cdot\frac{\alpha}{\alpha-1}$. Substitute dynamic noise $\sigma_t = \frac{\sigma_0}{1+\alpha\cdot\|\Delta_i\|_F}$ to obtain:

$$\mathscr{R}_\alpha(\sigma_t) = \frac{\sigma_0^2}{2}\cdot\frac{\alpha}{\alpha-1}\cdot\frac{1}{(1+\alpha\|\Delta_t\|_F)^2} \tag{14}$$

This formula indicates that as the deviation $\|\Delta_t\|_F$ between the local model and the global model increases, the noise intensity σ_t decreases, resulting in a reduction in the one-step RDP budget $\mathscr{R}_\alpha(\sigma_t)$. This allows for a dynamic balance between privacy protection and model performance.

5.1.5 Final Privacy Budget

Substitute the one-step RDP into the cumulative formula (10), and combine with the transformation formula (11), yielding:

$$\varepsilon(\alpha) = \sum_{t=1}^{T}\frac{\sigma_0^2}{2}\cdot\frac{\alpha}{\alpha-1}\cdot\frac{1}{(1+\alpha\cdot\|\Delta_t\|_F)^2} + \frac{\ln(1/\delta)}{2(\alpha-1)} \tag{15}$$

By optimizing α to minimize $\varepsilon(\alpha)$, we ultimately obtain:

$$\varepsilon(\alpha) = \sum_{t=1}^{T}\mathscr{R}_\alpha(\mathscr{A}_t, D, D') + \frac{\ln(1/\delta)}{2(\alpha-1)} \tag{16}$$

5.2 Convergence Analysis Proof

We have four established assumptions.

A1. Lipschitz Continuity: The objective function $\mathscr{L}_i(\theta_i)$ satisfies Lipschitz continuity and has a constant β. The conditions for Lipschitz continuity are:

$$\|\nabla\mathscr{L}_i(\theta) - \nabla\mathscr{L}_i(\theta')\| \leq \beta\|\theta-\theta'\|, \forall\theta,\theta' \tag{17}$$

where β is a positive constant, controlling the smoothness of the objective function.

A2. Strong Convexity: There exists a constant $\mu > 0$, such that for the objective function $\mathscr{L}_i(\theta_i)$, the following condition holds:

$$\mathscr{L}_i(\theta') \geq \mathscr{L}_i(\theta) + \langle\nabla\mathscr{L}_i(\theta), \theta'-\theta\rangle + \frac{\mu}{2}\|\theta'-\theta\|^2 \tag{18}$$

This condition ensures that the objective function $\mathscr{L}_i(\theta)$ is strongly convex, thus guaranteeing a unique global minimum.

A3. Bounded Gradient: The gradient of the objective function is bounded, and satisfies the following condition:

$$\mathbb{E}_i\|\nabla\mathscr{E}_i(\theta)-\nabla\mathscr{E}(w)\|_2^2 \leq B^2\|\nabla\mathscr{E}(w)\|_2^2 \tag{19}$$

This assumption controls the fluctuations of the gradient across different data samples, ensuring that the gradient does not change too drastically.

A4. Dynamic Noise Condition: The noise intensity satisfies

$$\sigma(\|\Delta_i\|_F) \leq \frac{\sigma_0}{1+\alpha\|\Delta_i\|_F}, \alpha > 0 \tag{20}$$

where α is a constant that controls the decay of the noise over time.

Proof (convergence in non-convex scenarios) Under assumptions ***A1-A4***, the global objective function of FedDyna satisfies the following:

$$\mathbb{E}[\mathscr{L}(\theta^t)] \leq \mathscr{L}(\theta^{t-1}) - \frac{\mu}{2}\|\theta^t-\theta^{t-1}\|^2 + \frac{\beta\eta^2}{2}\left(\frac{2B^2}{\eta^2\kappa^2}\|\nabla\mathscr{E}(\theta)\|^2 + \frac{\sigma_0^2}{(1+\alpha\|\Delta_t\|_{\mathrm{F}})^2}\cdot\mathbb{E}[\|\nabla\theta_i+\Delta\theta_i\|^2]\right) \tag{21}$$

where, $\mathscr{L}(\theta)=\sum_i \frac{|D_i|}{|D|}\mathscr{E}_i(\theta_i)$ is the global loss function.

Proof According to the Taylor expansion and ***A1***, for any θ^t and θ^{t-1}:

$$\mathscr{L}(\theta^t) \leq \mathscr{L}(\theta^{t-1}) + \langle\nabla\mathscr{L}(\theta^{t-1}), \theta^t-\theta^{t-1}\rangle + \frac{\beta}{2}\|\theta^t-\theta^{t-1}\|^2 \tag{22}$$

The client-side local updates satisfy:

$$\theta_i^t = \theta_i^{t-1} - \eta(\nabla\mathscr{E}_i(\theta_i^{t-1}) + \alpha(\theta_i^{t-1}+\Delta_i^{t-1}-\theta^{t-1}) + \nabla\mathscr{G}_i + \nabla\mathscr{N}_i) \tag{23}$$

where the gradient of the gradient variance suppression term $\mathscr{G}_i$ is $\nabla\mathscr{G}_i = \frac{1}{\eta K}(\nabla\mathscr{E}_i - \nabla\mathscr{E}_g)$. According to ***A3***, its variance satisfies: $\mathbb{E}\|\nabla\mathscr{G}_i\|^2 \leq \frac{2B^2}{\eta^2K^2}\|\nabla\mathscr{E}(\theta)\|^2$. Expand the local update equation $\theta^t = \mathbb{E}_i(\theta_i^t+\Delta_i^t)$ to obtain $\theta^t-\theta^{t-1} = \mathbb{E}_i(\theta_i^t+\Delta_i^t-\theta_i^{t-1}-\Delta_i^{t-1})$. Combining the local update equations (25). Substitute $\theta^t-\theta^{t-1}$ into the gradient term $\langle\nabla\mathscr{L}(\theta^{t-1}), \theta^t-\theta^{t-1}\rangle$ in the Taylor expansion and take the expectation:

$$\mathbb{E}[\langle\nabla\mathscr{L}(\theta^{t-1}), \theta^t-\theta^{t-1}\rangle] = -\eta\mathbb{E}[\langle\nabla\mathscr{L}(\theta^{t-1}), \nabla\mathscr{E}_i + \alpha(\theta_i^{t-1}+\Delta_i^{t-1}-\theta^{t-1}) + \nabla\mathscr{G}_i + \nabla\mathscr{N}_i\rangle] \tag{24}$$

The server aggregates updates to refine the global model under the strong convexity assumption ***A2*** and the optimization of gradient consistency, resulting in a further simplification to

$$\mathbb{E}[\langle\nabla\mathscr{L}(\theta^{t-1}), \theta^t-\theta^{t-1}\rangle] \leq -\frac{\mu}{2}\mathbb{E}[\|\theta^t-\theta^{t-1}\|^2] + O \tag{25}$$

where O denotes the higher order term. The expectation of the second-order term $\frac{\beta}{2}\|\theta^t-\theta^{t-1}\|^2$in the Taylor expansion is:

$$\mathbb{E}\left[\frac{\beta}{2}\|\theta^t-\theta^{t-1}\|^2\right] = \frac{\beta\eta^2}{2}\left(\frac{2B^2}{\eta^2\kappa^2}\|\nabla\mathscr{E}(\theta)\|^2 + \frac{\sigma_0^2}{(1+\alpha\|\Delta_t\|_{\mathrm{F}})^2}\cdot\mathbb{E}[\|\nabla\theta_i+\Delta\theta_i\|^2]\right) \tag{26}$$

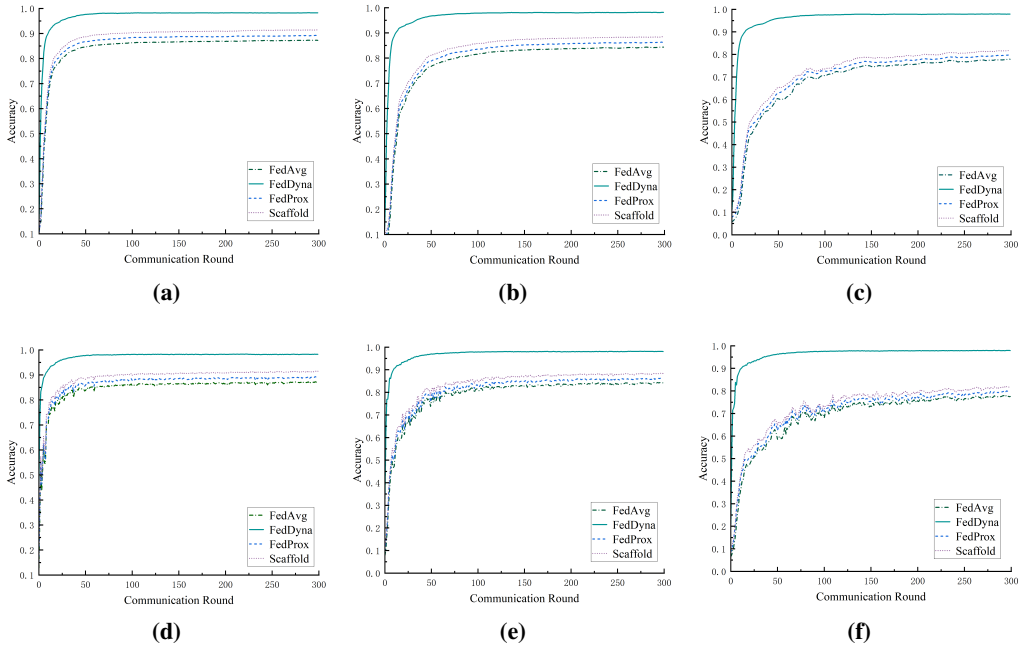

Fig. 2. (a) ε=6 MNIST all client participation, (b) ε=4 MNIST all client participation, (c) ε=2 MNIST all client participation, (d) ε=6 MNIST partial client participation, (e) ε=4 MNIST partial client participation, (f) ε=2 MNIST partial client participation.

Method	FedAvg	FedProx	Scaffold	FedDyna	FedAvg	FedProx	Scaffold	FedDyna
Setting	100 clients full participation				100 clients partial participation			
CIFAR10-0.3	0.4267	0.4591	0.4806	**0.8614**	0.3549	0.4213	0.4516	**0.803**
CIFAR10-0.6	0.4535	0.4808	0.5023	**0.8238**	0.4338	0.4654	0.4821	**0.8168**
CIFAR10-0.9	0.5458	0.5523	0.5697	**0.8818**	0.5312	0.5396	0.561	**0.874**
MNIST-0.3	0.8806	0.8816	0.8827	**0.9818**	0.8769	0.8783	0.8819	**0.9810**
MNIST-0.6	0.8834	0.8825	0.8845	**0.8917**	0.8820	0.8815	0.8831	**0.9812**
MNIST-0.9	0.8439	0.8634	0.8836	**0.9809**	0.8827	0.8820	0.8844	**0.9813**

Table 1: In the context of Non-IID data, the model's performance in terms of accuracy was assessed under two client engagement modes: complete client involvement and partial client participation. With the total number of clients configured as 100, three distinct Non-IID distribution coefficients were adopted, specifically 0.3, 0.6, and 0.9.

Taking the expectation of the global objective function yields equation (23).

6 Experiments

6.1 Datasets and baseline settings

This study conducts experimental verification on MNIST [29] and CIFAR-10 [30] two standard test datasets. The training and testing set partitioning methods for each dataset follow those used in the literature [31]. Under IID conditions, the training samples are randomly divided equally among the clients, ensuring that each terminal device holds an equal and class-balanced number of training samples. In the non-IID configuration, the labeling proportion per client conforms to the Dirichlet distribution, and we performed comparative experiments under diverse parameter settings. If not specified the Dirichlet distribution setting parameter is 0.6. Every client obtains samples via non-replacement sampling from the entire training dataset, based on the label distribution that conforms to a Dirichlet distribution. The experimental setup sets the maximum number of global training rounds to 400, as preliminary verification has shown that the model's performance stabilizes after this number of rounds. To validate the universality of the algorithm, we construct two typical network architectures: a three-layer fully connected network for the MNIST task (consistent with reference), and a CNN architecture that includes convolutional and pooling layers for the CIFAR-10 task.

6.2 Analysis

Privacy budget analysis. To verify the dynamic balance between privacy protection strength and model accuracy in federated learning, privacy budget sensitivity experiments were conducted on the MNIST dataset. The privacy budget ε was set to 2, 4, and 6 (smaller ε indicates higher privacy protection strength), and model performance was compared under full and partial client participation scenarios (Figure 2).

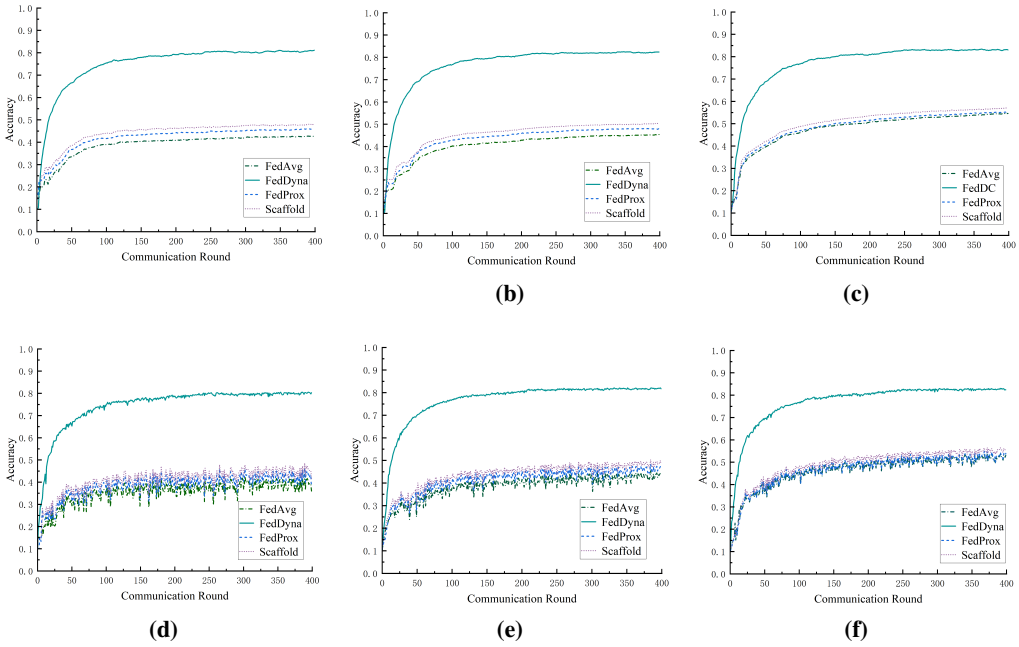

Fig. 3. (a)$D1$=0.3 All clients participate, (b) $D1$=0.6 All clients participate, (c)$D1$=0.9 All clients participate, (d)$D1$=0.3 partial client participation, (e)$D1$=0.6 partial client participation, (f)$D1$=0.9 partial client participation.

As ε decreases, model accuracy declines in both scenarios, but FedDyna consistently outperforms FedAvg, FedProx, and Scaffold across different privacy budgets and participation scales. Specifically, at $\varepsilon = 6$ (weak privacy protection), FedDyna achieves over 95% accuracy with full participation and 94% with partial participation, 3%-5% higher than other algorithms; at $\varepsilon = 4$, it maintains 93% accuracy in full participation (while FedAvg, etc., fall below 92%); at $\varepsilon = 2$ (strong privacy constraint), it reaches 85% in partial participation, 7% higher than FedAvg (78%) and 3% higher than Scaffold (82%), demonstrating low sensitivity and strong robustness to privacy budget changes.

FedDyna's advantage comes from the synergistic optimization of its dynamic noise injection strategy and bias correction mechanism. Traditional methods use static noise injection (noise intensity unchanged with model bias), leading to excessive noise interference and disrupted local-global model consistency under strong privacy budgets. In partial participation scenarios, increased data heterogeneity disperses local model biases and reduces effective aggregation samples, causing significant performance degradation in traditional methods due to blind static noise perturbation and insufficient bias correction. In contrast, FedDyna's "precise noise reduction" (dynamic noise) and "directional bias correction" limit accuracy degradation—at $\varepsilon = 2$, its accuracy drop in partial participation (4%) is only half of FedAvg's (7%), showing strong adaptability to data heterogeneity and participation randomness.

Impact of the degree of non-IID on accuracy. To verify the effect of Non-IID in FL on model accuracy, this experiment set three data distribution inhomogeneities on the CIFAR-10 dataset via Dirichlet distribution parameter $D1$ ($D1 = 0.3$ for high Non-IID, $D1 = 0.6$ for medium Non-IID, and $D1 = 0.9$ for low Non-IID). As shown in Figure 3, under low Non-IID ($D1 = 0.9$), all algorithms achieve high accuracy, with FedDyna stabilizing above 82% after 200 rounds. Under high Non-IID, FedAvg, FedProx, and Scaffold converge slowly and see significant precision drops, while FedDyna separates model parameter deviations via a matrix trace-constrained local drift tracking mechanism and gradient variance regularization, stabilizing accuracy at 81% and completing the key performance node 100 rounds earlier than FedAvg. When some clients participate, FedAvg, FedProx, and Scaffold never converge regardless of data distribution, whereas FedDyna still gradually tends to converge.

Experiments show that as Non-IID degree deepens, traditional FL algorithms suffer significant performance degradation due to static noise strategies and insufficient bias correction. In contrast, FedDyna achieves synergistic optimization between dynamic noise adjustment and structured bias correction, sustaining a steady convergence rate while retaining high model accuracy in the context of highly heterogeneous data distributions. This not only validates its robust adaptability to Non-IID data within practical IoT application scenarios but also offers a viable approach to addressing the trade-off between privacy protection and model accuracy amid complex data distributions. Table 1 presents the accuracy of CIFAR-10 and MNIST datasets across different models and settings.

7 Conclusion

Federated learning has become an important paradigm for privacy-aware intelligence in distributed IoT environments, because it enables multiple devices to collaboratively train models with-

out directly sharing raw data. Nevertheless, practical deployments are still constrained by highly heterogeneous local data, unstable participation, and the persistent tension between privacy preservation and model utility. This paper addresses the performance degradation of privacy protection methods in federated learning under non-IID data and proposes the FedDyna framework based on a dynamic noise mechanism. Specifically, FedDyna constructs a dynamic bias correction and gradient optimization system: it explicitly separates and corrects model parameter biases via a matrix trace-constrained local drift tracking mechanism, and suppresses local update divergence with gradient variance regularization. In addition, it designs a decaying noise injection method dynamically correlated with model biases, thereby satisfying Rényi differential privacy while avoiding the excessive disturbance often caused by static noise strategies. Experimental results on MNIST and CIFAR-10 demonstrate that FedDyna achieves faster convergence, higher accuracy, and stronger robustness under both full and partial client participation. Overall, the proposed framework provides an effective and practical solution for privacy protection and performance optimization in non-IID federated learning scenarios, and it offers useful guidance for future privacy-preserving model design in complex real-world systems.

Acknowledgments

This work was supported in part by the Colleges and Universities 20 Terms Foundation of Jinan City under Grant 202228093, in part by the Major Program of Shandong Provincial Natural Science Foundation for the Fundamental Research under Grant ZR2022ZD03, in part by the National Science Foundation of China under Grants 62272256 and 62202250, in part by the Shandong Province Youth Innovation Team Project under Grant 2024KJH032, in part by the National Natural Science Foundation of China under Grant 62402254, and in part by the Natural Science Foundation of Shandong Province of China under Grants ZR2022QF094 and ZR2022QF010.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] Deng D, Wu X, Zhang T, Xiang C, Zhao W, Xu M, et al. pFedCal: Lightweight Personalized Federated Learning with Adaptive Calibration Strategy. IEEE Transactions on Services Computing. 2025.

[2] He Z, Wang L, Cai Z. Clustered federated learning with adaptive local differential privacy on heterogeneous iot data. IEEE Internet of Things Journal. 2023;11(1):137-46.

[3] Myakala PK, Kamatala S, Bura C. Privacy-Preserving Federated Learning for IoT Botnet Detection: A Federated Averaging Approach. Preprint. 2025.

[4] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. PMLR; 2017. p. 1273-82.

[5] Geyer RC, Klein T, Nabi M. Differentially private federated learning: A client level perspective. arXiv preprint arXiv:171207557. 2017.

[6] McMahan HB, Ramage D, Talwar K, Zhang L. Learning differentially private recurrent language models. arXiv preprint arXiv:171006963. 2017.

[7] Melis L, Song C, De Cristofaro E, Shmatikov V. Exploiting unintended feature leakage in collaborative learning. In: 2019 IEEE symposium on security and privacy (SP). IEEE; 2019. p. 691-706.

[8] Hitaj B, Ateniese G, Perez-Cruz F. Deep models under the GAN: Information leakage from collaborative deep learning. In: Proceedings of the 2017 ACM SIGSAC conference on computer and communications security; 2017. p. 603-18.

[9] Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems. 2020;2:429-50.

[10] Karimireddy SP, Kale S, Mohri M, Reddi S, Stich S, Suresh AT. Scaffold: Stochastic controlled averaging for federated learning. In: International conference on machine learning. PMLR; 2020. p. 5132-43.

[11] Yan J, Chen T, Sun Y, Nan Z, Zhou S, Niu Z. Mobility-Aware Asynchronous Federated Learning with Dynamic Sparsification. arXiv preprint arXiv:250607328. 2025.

[12] Narimani MH, Tavassolipour M. FedRP: A Communication-Efficient Approach for Differentially Private Federated Learning Using Random Projection. arXiv preprint arXiv:250910041. 2025.

[13] Li Y, Fu L, Wang T, Lou J, Chen B, Yang L, et al. Clients collaborate: Flexible differentially private federated learning with guaranteed improvement of utility-privacy trade-off. arXiv preprint arXiv:240207002. 2024.

[14] Collins E, Wang M. Federated Learning: A Survey on Privacy-Preserving Collaborative Intelligence. arXiv preprint arXiv:250417703. 2025.

[15] Kairouz P, McMahan HB, Avent B, et al. Advances and Open Problems in Federated Learning. Foundations and Trends in Machine Learning. 2021;14(1–2):1-210.

[16] Hsu TMH, Qi H, Brown M. Federated Learning on Non-IID Data Silos: An Experimental Study. arXiv preprint arXiv:190906335. 2019.

[17] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. PMLR; 2017. p. 1273-82.

[18] Woodworth BE, Wang J, Smith A, McMahan B, Srebro N. Graph oracle models, lower bounds, and gaps for parallel stochastic optimization. Advances in neural information processing systems. 2018;31.

[19] Jian L, Liu D. Widening the Network Mitigates the Impact of Data Heterogeneity on FedAvg. arXiv preprint arXiv:250812576. 2025.

[20] Acar DAE, Zhao Y, Navarro RM, Mattina M, Whatmough PN, Saligrama V. Federated learning based on dynamic regularization. arXiv preprint arXiv:211104263. 2021.

[21] He J, Chen W, Zhang X. FedAA: A Reinforcement Learning Perspective on Adaptive Aggregation for Fair and Robust Federated Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39; 2025. p. 17085-93.

[22] Chu YC, Gao W, Ye Y, Udell M. Provable and practical online learning rate adaptation with hypergradient descent. arXiv preprint arXiv:250211229. 2025.

[23] Xie C, Huang K, Chen PY, Li B. Dba: Distributed backdoor attacks against federated learning. In: International conference on learning representations; 2019. .

[24] Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V. How to backdoor federated learning. In: International conference on artificial intelligence and statistics. PMLR; 2020. p. 2938-48.

[25] Bhagoji AN, Chakraborty S, Mittal P, Calo S. Analyzing federated learning through an adversarial lens. In: International conference on machine learning. PMLR; 2019. p. 634-43.

[26] Liu G, Xu T, Yang Y, Abdelmoniem AM, Wang C, Liu J. Poisoning as a Post-Protection: Mitigating Membership Privacy Leakage From Gradient and Prediction of Federated Models. IEEE Transactions on Dependable and Secure Computing. 2025.

[27] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security; 2017. p. 1175-91.

[28] Weng S, Ren C, Xiao M, Skoglund M. Heterogeneity-Aware Client Sampling: A Unified Solution for Consistent Federated Learning. arXiv preprint arXiv:250511304. 2025.

[29] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 2002;86(11):2278-324.

[30] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. University of Toronto; 2009.

[31] Bensiah OA, Benaboud R. FedDPA: Dynamic Prototypical Alignment for Federated Learning with Non-IID Data. Electronics. 2025;14:3286.