A Client-level Conditonal Generative Adversarial Network-based Data Reconstruction Attack and its Defense in Clustered Federated Learning Scenario

Lei Shi *Member, IEEE*, Han Wu*, Xu Ding, Hao Xu, Sinan Pan

***Abstract*—Clustered Federated Learning (CFL) has emerged as an effective solution to address data heterogeneity in traditional Federated Learning (FL). However, the intrinsic cluster-based structure of CFL introduces new privacy risks, making it more vulnerable to client-level inference attacks. In this paper, we propose a novel client-level data reconstruction attack based on Conditional Generative Adversarial Networks (cGANs), which exploits intra-cluster similarities to enhance the quality of reconstructed private data. Unlike prior works, our attack requires only partial access to a victim's model updates through passive eavesdropping, thereby reflecting a more realistic threat model in decentralized and resource-constrained environments such as the Internet of Things (IoT). To mitigate this threat, we develop a lightweight and adaptive defense mechanism grounded in Local Differential Privacy (LDP). Our design incorporates dynamic privacy budget decay, selective layer-wise noise injection, and real-time similarity-guided adaptation. This approach achieves a favorable privacy-utility trade-off while explicitly addressing the computational, communication, and latency constraints inherent in IoT environments. Experimental results demonstrate that our proposed attack improves reconstruction similarity by up to 20% compared with existing baselines, while the defense reduces attack success rate by 27.2% with only a 3.3% accuracy drop. Moreover, it significantly lowers computational cost—reducing FLOPs by 42.7%, memory usage by 23.4%, and DP noise processing time by 45.5%—without introducing additional communication overhead. These findings highlight the underestimated privacy vulnerabilities in CFL and underscore the necessity of efficient, context-aware defense strategies.**

***Index Terms*—Clustered federated learning, Data reconstruction attack, Differential privacy, Generative adversarial nets.**

I. Introduction

FEDERATED Learning (FL) [1] [2] enables collaborative model training across decentralized clients without exchanging raw data, thereby providing a promising privacy-preserving paradigm. However, standard FL frameworks often suffer from data heterogeneity—where clients hold non-independent and identically distributed (non-IID) data—which significantly degrades model performance and convergence stability.

To address this challenge, **Clustered Federated Learning (CFL)** [3] has been proposed. CFL organizes clients into multiple clusters based on data similarity, enabling each cluster to train its own localized global model. This cluster-specific approach improves both accuracy and convergence in heterogeneous environments. Nevertheless, CFL also introduces new security vulnerabilities. Specifically, the enforced homogeneity within clusters reduces the diversity of model updates, thereby lowering the entropy of shared information and making it more susceptible to inference attacks.

Despite the growing attention to privacy risks in FL, the unique threats stemming from CFL's structural characteristics remain unexplored [4]. Most existing reconstruction attacks are designed for standard FL settings, assuming either full access to client gradients (e.g., DLG [5], iDLG [6]) or centralized adversaries capable of observing global model updates (e.g., Hitaj et al.'s GAN-based attack [7]). Such assumptions, however, are often impractical in decentralized real-world deployments, particularly in IoT environments.

In this paper, we propose a novel and practical **client-level data reconstruction attack** tailored for the CFL framework. Our method leverages *Conditional Generative Adversarial Networks (cGANs)* and is executed by a malicious client embedded within a CFL cluster. Unlike prior works, our attacker gains only *partial access* to the victim's uploaded model parameters through passive eavesdropping over insecure communication channels (e.g., Wi-Fi, BLE). Despite this limited visibility, the attacker exploits *intra-cluster similarity* to train a cGAN capable of generating high-fidelity reconstructions of the victim's private training data. This setup reflects a more realistic and severe threat scenario, especially in resource-constrained and decentralized environments such as IoT or edge computing.

To counter this attack, we propose a lightweight and adaptive defense mechanism in *Local Differential Privacy (LDP)*. The method dynamically adjusts privacy budgets and selectively injects noise into model layers based on real-time risk estimation, thereby achieving a favorable privacy-utility trade-off. Importantly, our design explicitly accounts for constraints typical of IoT environments [8]—including **limited computational resources, restricted memory, bandwidth limitations, and strict latency requirements**—which render conventional, heavyweight privacy-preserving methods impractical. By maintaining low overhead and avoiding centralized coordination, the proposed defense is particularly suitable for on-device deployment in real-world IoT settings.

Our work sheds light on a previously underexplored vulnerability in CFL and makes the following key contributions:

- We identify and formulate a practical client-level data re-

L. Shi, H. Wu, X. Ding, H. Xu and S. Pan are all with School of Computer Science and Information Engineering,Hefei University of Technology,Hefei, China

*Corresponding Author: H. Wu (email: 2951817936@qq.com)

This paper was produced by the IEEE Publication Technology Group. They are in Piscataway, NJ.

construction attack in the CFL setting, leveraging cGANs to exploit intra-cluster similarity under an eavesdropping threat model.
- We propose an adaptive hybrid loss function that enhances reconstruction fidelity while improving the stability of GAN training.
- We develop a resource-efficient defense mechanism based on LDP, which integrates privacy budget decay, selective layer-wise noise injection, and real-time adaptive control, making it suitable for IoT constraints.
- We conduct extensive experiments to evaluate the effectiveness and efficiency of the proposed attack and defense framework in realistic CFL and IoT environments.

The remainder of the paper is organized as follows: Section II reviews related work on FL, CFL, data privacy attacks, and GANs. Section III presents the threat model, the details of the proposed attack, and the corresponding experimental results. Section IV introduces the defense methodology and evaluates its performance through comprehensive experiments. Finally, Section V concludes the paper and outlines future research directions.

II. RELATED WORK

A. Data Reconstruction Attacks in FL and CFL

Federated Learning (FL) faces a wide range of security threats, among which data reconstruction attacks [9], [10] are particularly concerning because they attempt to recover clients' original training data, thereby directly violating user privacy. Such attacks undermine the core promise of FL—data confidentiality—and are especially harmful in privacy-sensitive domains such as healthcare and finance.

Data reconstruction attacks in FL can broadly be divided into two main categories: gradient inversion-based attacks and GAN-based attacks.

Gradient inversion-based attacks attempt to recover client data by reversing gradients shared during training. Representative methods include DLG and its improved version iDLG, which optimize synthetic inputs to match observed gradients. More advanced schemes such as GS [11], CPL [12], R-GAP [13], and COPA [14] enhance reconstruction fidelity by estimating gradient sensitivity and leveraging feature priors. FedDC [15], while originally designed to mitigate convergence issues in non-IID settings, also modifies model update dynamics in ways that can inadvertently increase privacy leakage. These methods are typically server-side and assume full access to client gradients and updates, enabling high-fidelity reconstructions but limiting their practicality in decentralized IoT and edge environments.

GAN-based attacks employ Generative Adversarial Networks to learn and replicate the data distribution of target clients. The seminal work by Hitaj et al. [7] demonstrated that a malicious client could train a GAN to generate synthetic data resembling the inputs of other clients. More recently, VagueGAN [16] investigated the use of GANs in poisoning attacks and highlighted their potential for reconstructing structured data in FL. In vertical federated learning (VFL) settings, Active Reconstruction attacks leverage active queries and cross-party feature correlations to enhance attack effectiveness. Recent studies indicate that adversaries can exploit adaptive generative models to achieve improved reconstructions even in resource-constrained scenarios. Zhao et al. [17] provide a comprehensive survey of privacy attacks and defenses in federated learning, offering updated taxonomies of reconstruction threats. Tan et al. [18] empirically reassess the strength of reconstruction attacks under realistic deployments, showing that while some traditional methods weaken, generative and topology-aware strategies remain effective. Moreover, undetectable reconstruction attacks such as URVFL [19] and generative frameworks like GenDRA [20] demonstrate that even with constrained visibility or stealthy interventions, adversaries can still recover sensitive data. These results directly motivate our focus on client-level, cGAN-based attacks in clustered federated learning (CFL).

However, most of these methods were developed within conventional FL frameworks, with limited attention to client-side GAN-based reconstruction attacks in clustered architectures such as CFL. This gap is particularly critical given CFL's unique intra-cluster communication patterns and relatively homogeneous data distributions, which introduce distinct vulnerabilities.

B. Lightweight and Adaptive Defenses for IoT

The rapid proliferation of IoT devices has introduced new challenges for applying privacy-preserving mechanisms within FL, particularly due to the limited computational resources, memory, and energy available on edge devices. Traditional defense strategies, such as global differential privacy and secure aggregation [21], often incur substantial computational and communication overhead, rendering them impractical for many real-world IoT deployments.

To overcome these limitations, recent studies have investigated lightweight and adaptive defense mechanisms. For example, PrivateFL-GAN [22] integrates differential privacy into GAN-based FL systems to generate synthetic data while preserving privacy. However, its reliance on globally added noise makes it less effective in heterogeneous or dynamic environments. Similarly, FedDC [15], although originally proposed to improve training under non-IID conditions, indirectly enhances privacy by decoupling local drift.

In contrast, adaptive privacy-preserving methods dynamically adjust protection levels according to contextual indicators such as model performance or data sensitivity, thereby minimizing unnecessary utility loss. Local Differential Privacy (LDP) has emerged as a particularly promising solution for decentralized and resource-constrained environments, as it allows each client to perturb its own data or update before transmission, removing reliance on a trusted central server. Recent advances have introduced layer-wise LDP schemes that selectively inject noise into sensitive layers [23], improving the privacy-utility balance, as well as privacy budget decay functions [24] that gradually reduce noise during training to preserve convergence and accuracy.

Recent studies increasingly emphasize adaptive and lightweight defenses in federated learning (FL), particularly

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2025.3637061

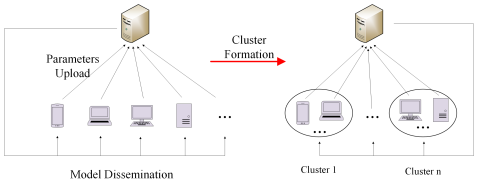

Fig. 1. Cluster process

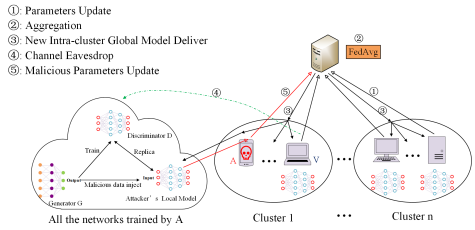

Fig. 2. Threat model

for IoT-oriented environments. A 2024 systematic review of differentially private FL categorizes adaptive-budget and per-layer defenses as promising approaches. Building on this, ALDP-FL [25] introduces adaptive localized differential privacy with bounded perturbations, demonstrating that real-time adjustment of noise can significantly improve the privacy–utility trade-off in IoT scenarios. Similarly, AdapLDP-FL [26] proposes dynamic adaptation of clipping and noise bounds to better resist evolving adversarial strategies, while Zhang et al. [27] explore adaptive differential privacy in asynchronous FL, addressing challenges of resource heterogeneity and communication constraints. Together, these works highlight that effective defenses in IoT-oriented FL must balance privacy, utility, and system efficiency. Our method extends these directions by integrating privacy budget decay, selective layer-wise perturbation, and real-time similarity-guided adjustments, specifically tailored for CFL environments.

Despite these advancements, few existing methods simultaneously address all three aspects—privacy effectiveness, computational efficiency, and adaptiveness—within the CFL setting. Building on this foundation, we propose a client-side, lightweight, and adaptive defense strategy that integrates privacy budget decay, selective parameter protection, and real-time feedback-driven adjustments. This design is specifically tailored for CFL systems deployed in IoT environments, where both resource constraints and adaptive privacy mechanisms are imperative.

III. Attack

In this section, we present the overall threat model of the Clustered Federated Learning (CFL) system and detail the design of our proposed client-level data reconstruction attack.

A. Threat Model

Consider a CFL system that consists of a central parameter server and multiple clients, as illustrated in Fig. 1. Each client holds non-independent and Identically Distributed (Non-IID) data. Assume that clients can be categorized into n groups according to their data distributions. Let D_i $(i = 1 \dots n)$ denote the set of clients sharing the same data distribution. In a typical CFL workflow, two fundamental processes are involved: client clustering and model aggregation. As shown in Algorithm I, during client clustering, the server identifies all D_i through three steps: (1) distributing the initial global model, (2) collecting updated client parameters, and (3) constructing a cosine similarity matrix to guide clustering. Based on the cosine similarity matrix, a clustering algorithm such as K-Means is applied to partition clients into n distinct groups. This procedure is repeated until suitable clusters D_i are established.

Algorithm 1 Clustering process

Input: the number of clients m, the number of clusters n, client k's parameters in round i $p_i^k(k = 1...m)$

1: *Server* calculates cosine similarity according $p_i^k(k = 1...m)$.
2: *Server* constructs the cosine similarity matrix M_c
3: *Server* divides the clients into n clusters using the clustering algorithm(K-Means for example) according to the cosine similarity matrix M_c.
4: **return** Multiple clusters $a_i(i = 1...n)$.

As shown in Fig. 2, we assume an adversary that participates legitimately in CFL and is assigned to one cluster during clustering. The adversary has white-box knowledge of the model architecture and label space and aims to reconstruct a co-clustered victim's local data by exploiting intra-cluster similarity in model updates. Concretely, the adversary obtains partial visibility into the victim's model parameters via passive eavesdropping on unsecured communication channels (e.g., Wi-Fi, BLE, vulnerable edge devices) and accumulates partial parameter snapshots across rounds (a stealthy, passive MITM strategy). In addition, the adversary may be a colluding client within the same cluster. In either scenario, only limited, round-wise parameter access is required, which reflects a realistic constraint in many practical CFL deployments, particularly lightweight IoT systems relying on simplified transmission protocols or local peer-to-peer exchanges. When end-to-end encryption (e.g., TLS or secure aggregation) is enforced, the attack capability is naturally limited. Therefore, this threat model provides an upper-bound analysis of potential data leakage in less protected FL systems.

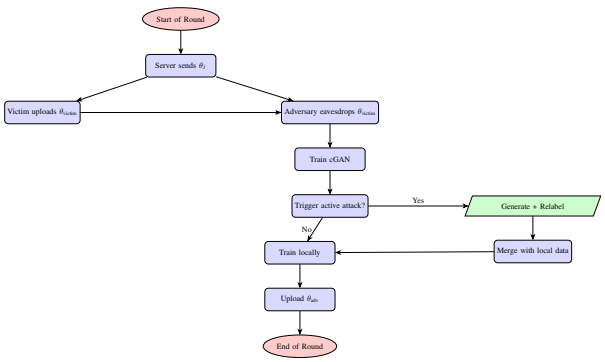

Fig. 3. Compact workflow of the client-level reconstruction attack in CFL.

B. Client-level cGAN-based Data Reconstruction Attack in CFL

We propose a client-level data reconstruction attack tailored for CFL that leverages conditional Generative Adversarial Networks (cGANs). The attack is executed by a malicious client that has been legitimately assigned to the same cluster as its target; it exploits intra-cluster similarity in model updates to recover private training examples from peer clients. As summarized in Algorithm II, the attack proceeds in two phases:

1) Passive Phase (Parameter Eavesdropping and GAN Training): The adversary passively intercepts partial model parameters (e.g., selected layer weights, compressed updates or gradients) uploaded by the target victim during routine CFL rounds. Despite limited visibility, the adversary aggregates these parameter snapshots across training rounds and uses them, together with cluster-level statistics, to supervise the training of a cGAN.

- The generator G synthesizes data conditioned on class labels.
- The discriminator D distinguishes real samples from generated ones.

Concretely, the generator is trained using both the global cluster model and the intercepted victim parameters as supervision. Over multiple communication rounds, G gradually approximates the victim's data distribution and produces increasingly realistic samples.

2) Active Phase (Label Manipulation and Model Poisoning): After the passive phase yields sufficiently plausible samples, the adversary may optionally adopt a more aggressive refinement step. The generator produces samples targeting a specific class (e.g., "5"), deliberately mislabels them (e.g., as "10") and incorporates them into its local training set. This process poisons the cluster model, slows convergence, and extracts more detailed features of the victim class, further improving reconstruction fidelity.

Attack Overview: Fig. 3 illustrates the overall workflow. The adversary alternates between passive learning and active manipulation, enabling progressive and stealthy reconstruction of the victim's private data without remaining compliant with CFL protocols.

Analytical Note on Attack Efficiency. To clarify the theoretical underpinnings of the proposed cGAN-based reconstruction, we briefly analyze the relation between gradient visibility and recovery probability. Let ∇_i denote the proportion of gradients visible to the adversary and S_c the intra-cluster similarity. The empirical success rate P_{succ} can be approximated as:

$$P_{succ} \approx \int P\big(G_\theta(x) \mid \nabla_i, S_c\big)\, dx, \tag{1}$$

which grows sub-linearly with ∇_i when S_c is high. This implies that even partial gradients can yield recognizable reconstructions in clustered settings, consistent with prior gradient inversion analyses. The computational complexity of the attack scales with $O(T{\cdot}|\nabla_i|)$, confirming its feasibility for resource-constrained clients.

Algorithm 2 Client-Level Reconstruction Attack in CFL

Require: Number of clients m, rounds T

for each round $i = 1$ to T **do**
  Server performs clustering and model aggregation
  Server sends cluster-specific model θ_j^i to clients
  for each client k **do**
    if client k is benign **then**
      Train locally and upload update θ_k^i
    else if client k is adversary **then**
      Receive cluster model θ_j^i
      Passive: Eavesdrop partial update θ_{victim}^i
      Update discriminator D with θ_{victim}^i
      Train generator G to match victim distribution
      if the Active phase triggered **then**
        Generate synthetic data and apply label manipulation
        Merge with local training set
      end if
      Upload poisoned update θ_{adv}^i to server
    end if
  end for
end for

C. Adaptive Hybrid Loss Function and Adversarial Attack

Modeling and Derivation of Loss Function

To enhance the stability and fidelity of training in our conditional GAN-based data reconstruction framework, we propose an *adaptive hybrid loss function* that combines L2 loss, a modified L1 loss, and cross-entropy loss. This formulation is specifically designed to address the unique challenges of adversarial training in CFL, where both precision and robustness to outliers are crucial—particularly during the early stages of training.

Let $x \in \mathbb{R}^n$ denote the ground truth data, and $\hat{x} \in \mathbb{R}^n$ denote the output of the generator. The overall loss function is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adaptive}}(x, \hat{x}) + \alpha \cdot \mathcal{L}_{\text{CE}}(y, \hat{y}), \tag{2}$$

where $\mathcal{L}_{\text{CE}}$ is the cross-entropy loss between the true label y and predicted label $\hat{y}$, and α is a balancing coefficient.

The core component, $\mathcal{L}_{\text{adaptive}}$, is defined as a piecewise function that switches between L2 and L1 penalties based on the prediction error magnitude:

$$\mathcal{L}_{\text{adaptive}}(x, \hat{x}) = \begin{cases} \|x - \hat{x}\|_2^2, & \text{if } \|x - \hat{x}\| \le \delta, \\ \lambda \cdot \|x - \hat{x}\|_1, & \text{if } \|x - \hat{x}\| > \delta, \end{cases} \tag{3}$$

where δ is a predefined threshold that determines the switching point between L2 and L1 behavior, and λ is a weighting factor applied to the L1 term.

This formulation allows the model to emphasize quadratic penalties for small errors—promoting fine-grained approximation—while applying linear penalties to large errors, thereby enhancing robustness and training stability. This dynamic mechanism is particularly advantageous during the early stages of GAN training when output deviation is high.

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2025.3637061

To further understand its optimization behavior, we derive the gradient of the adaptive loss with respect to the generator output $\hat{x}$:

$$\frac{\partial \mathcal{L}_{\text{adaptive}}}{\partial \hat{x}} = \begin{cases} 2(\hat{x} - x), & \text{if } \|x - \hat{x}\| \leq \delta, \\ \lambda \cdot \text{sign}(\hat{x} - x), & \text{if } \|x - \hat{x}\| > \delta. \end{cases} \tag{4}$$

This gradient formulation reveals that the L2 region offers smooth and continuous updates, facilitating convergence, while the L1 region provides bounded gradients, improving robustness to large deviations or noisy samples. Although not differentiable at the threshold $\|x - \hat{x}\| = \delta$, the loss function remains continuous and piecewise smooth. This mild discontinuity is common in robust optimization and has been adopted in practical models, such as Huber loss and Smooth L1 loss used in Fast R-CNN and VagueGAN.

Our design is inspired by and extends earlier adaptive loss functions. For example, Smooth L1 provides a continuous transition between L2 and L1 norms but does not incorporate semantic conditioning. Recent studies, such as VagueGAN and Active Reconstruction in VFL [28] apply adaptive loss functions in generative settings but omit classification-aware supervision [29]. In contrast, our method integrates a tunable threshold δ for dynamic adjustment, along with cross-entropy guidance, to enable more effective conditional generation. This design results in improved convergence and higher reconstruction quality.

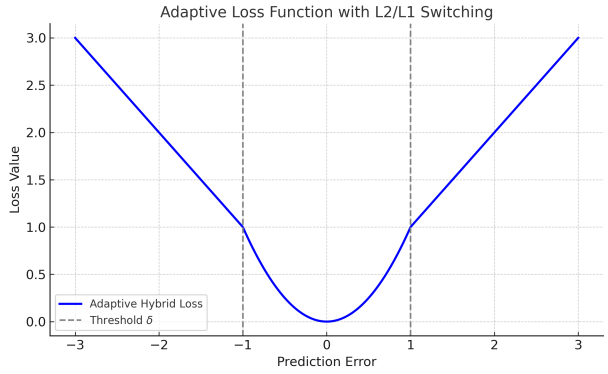

Fig. 4. Visualization of the adaptive hybrid loss function. The function applies L2 loss for small prediction errors ($\|x-\hat{x}\| \leq \delta$) to encourage precision, and switches to L1 loss for large errors ($\|x-\hat{x}\| > \delta$) to improve robustness and stabilize training.

The behavior of the proposed loss function is illustrated in Fig. 4. When the prediction error $\|x-\hat{x}\|$ is below a threshold δ, the loss adopts a quadratic (L2) form, encouraging fine-grained reconstruction and faster convergence—particularly beneficial in later training stages. When the error exceeds δ, the loss transitions to a linear (L1) form, mitigating the impact of large deviations and enhancing training stability. This adaptive structure balances learning between global patterns and local details.

By combining L1 and L2 behaviors in a piecewise manner, the loss enables smooth optimization in low-error regions while maintaining robustness to outliers. As shown in Fig. 4, the curve is continuous but adjusts its curvature based on the error magnitude, supporting stable and progressive learning. This design ultimately improves the fidelity and realism of the reconstructed data.

Our adaptive hybrid loss shares conceptual similarities with the classical Huber loss [30], which also applies a quadratic penalty to small errors and a linear penalty to large ones. As depicted in Fig. 5, both functions provide a principled balance between sensitivity and robustness.

$$\mathcal{L}_{\text{Huber}}(e) = \begin{cases} \frac{1}{2}e^2, & \text{if } |e| \leq \delta, \\ \delta(|e| - \frac{1}{2}\delta), & \text{otherwise.} \end{cases} \tag{5}$$

While we adopt the same switching principle, our formulation simplifies the transition mechanism by directly using L2 and L1 loss without the smooth blending term of Huber loss. This design is better suited for adversarial settings, where gradient stability and computation efficiency are essential.

Moreover, our method integrates cross-entropy loss to guide conditional generation, enabling label-consistent reconstruction—a key requirement in tasks involving semantic conditioning, which is not supported by the traditional Huber loss.

The parameter δ controls the switch between L2 and L1 regimes. A small δ results in a faster transition to the robust L1 region, while a large δ maintains quadratic sensitivity longer, which may amplify outlier effects. In our implementation, δ is empirically set based on the early-stage average reconstruction error, ensuring that the model initially benefits from the L1 robustness and gradually transitions to precision via L2. This threshold-based switching stabilizes adversarial training and improves both convergence and generative fidelity.

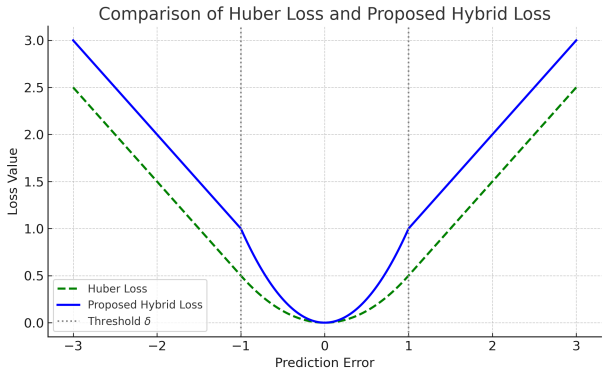

Fig. 5. Comparison between classical Huber loss and our proposed hybrid loss function. The proposed method directly switches between L2 and L1 based on error magnitude, offering simpler gradients and better suitability for GAN-based training.

Adversarial attack As shown in Fig. 6, the adversary may prolong and amplify its reconstruction capability by introducing subtle adversarial interventions. For instance, the adversary can inject synthesized samples conditioned on the victim's target class (e.g., class "5") but intentionally label them as a different class (e.g., "10"), thereby slowing the cluster model's convergence and creating a persistent window for influence. As the victim's local model adapts to these manipulated updates, its internal representations become biased toward features associated with the true target class, inadvertently exposing additional information about that class. The adversary exploits

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2025.3637061

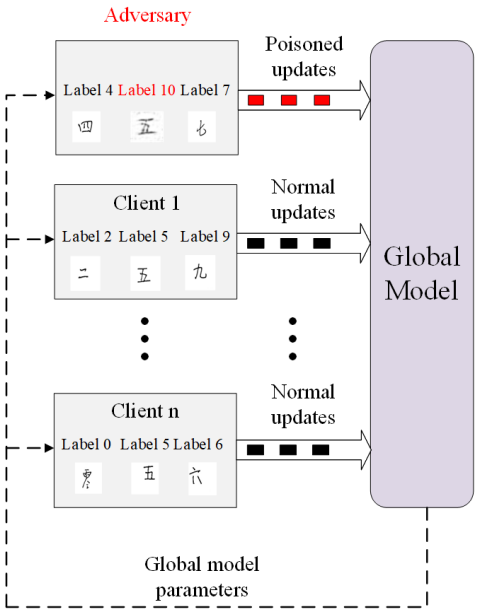

Fig. 6. Introduce adversarial influence

this adaptive response by refining its generator to more closely mimic the revealed characteristics of class "5", progressively improving reconstruction fidelity. Because these interventions conform to the standard training protocol and produce only subtle perturbations to local updates, they are difficult to detect using conventional defenses.

IV. Defense

In this section, we first present the defense mechanism proposed in this paper and then evaluate its performance by comparing it with representative baseline approaches.

A. Defense Method

After the client clustering phase in Clustered Federated Learning (CFL) is completed, where each client is assigned to its designated cluster, the defense mechanism is activated and remains effective throughout the subsequent training process. Integrated into the system architecture, it enables clients to locally apply defense strategies that mitigate potential attack risks, safeguard data privacy, and minimize performance degradation.

The defense mechanism adopts two main strategies:

1) **Differential Privacy in Local Training:** Clients inject noise into model parameters during local training, with the noise scale increasing as training progresses. A randomized mechanism M satisfies (ε, δ)-differential privacy if for any pair of neighboring datasets D and D', and any subset of outputs S, the following condition holds:

$$\Pr[M(D) \in S] \leq e^{\varepsilon} \cdot \Pr[M(D') \in S] + \delta \quad (6)$$

where ε is the privacy budget that controls the degree of information leakage (smaller ε implies stronger privacy), and δ is a relaxation parameter representing the maximum tolerated failure probability.

2) **Dynamic Adjustment Strategy:** Clients dynamically adjust the probability of adding DP noise based on two factors: data reconstruction similarity and local model accuracy. This approach ensures a balance between privacy protection and training effectiveness.

The following sections elaborate on the implementation details, theoretical basis, and practical impact of these mechanisms across different stages of the training process.

B. Adaptive Privacy Mechanism

As shown in Algorithm III, our defense method is lightweight, adaptive, and suitable for resource-constrained environments such as IoT. It comprises three tightly integrated components:

1) **Privacy Budget Decay:** We employ a cosine-based decay function to gradually reduce the privacy budget ϵ across training rounds. The mechanism starts with weaker noise to ensure convergence and gradually increases protection as reconstruction risk accumulates.
2) **Layer-wise Selective Noise Injection:** As shown in Fig. 7, rather than perturbing the entire model uniformly, stronger noise is applied to deep layers that capture sensitive features, while lighter noise is injected into early convolutional layers to preserve general representation quality and reduce computational overhead.
3) **Real-time Similarity-Guided Adaptation:** Clients estimate local reconstruction similarity using a lightweight GAN module and dynamically adjust the probability and strength of noise injection. When both model accuracy and similarity risk are high, stronger protection is activated.

Overview: The defense operates entirely locally on each client, requiring no trusted server or encrypted communication, making it ideal for federated deployments in edge environments.

Decay Function Explanation: The privacy budget ϵ is computed as:

$$\epsilon = \frac{-e^{\text{round}}/\phi_1 + \phi_2}{\phi_3} \quad (7)$$

This cosine-like decay ensures a smooth and monotonic reduction of the privacy budget over training rounds, avoiding abrupt jumps that could destabilize model training. Compared to linear decay, it allocates more noise in later stages—when attacks are more effective—thereby improving the privacy-utility trade-off.

Formal Privacy Budget Composition: Following the Gaussian mechanism and standard composition theorems, the cumulative privacy cost of the proposed layer-wise LDP defense can be approximated as:

$$\epsilon_{\textbf{total}} = \sqrt{2T\log(1/\delta)} \sum_{l=1}^{L} \frac{\Delta_l^2}{\sigma_l^2}, \quad (8)$$

where Δ_l and σ_l denote the sensitivity and noise scale at layer l, respectively. This bound ensures that the composed leakage remains within a controlled $(\epsilon_{\textbf{total}}, \delta)$ privacy budget. The derivation aligns with the moments accountant analysis used in prior adaptive DP-FL studies [31].

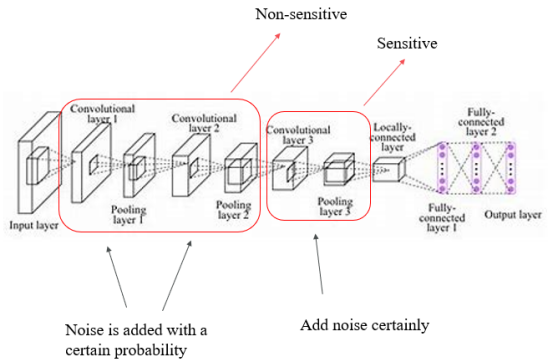

Fig. 7. Selective protection.

C. Lightweight Design for IoT Deployment

To address the constraints of resource-limited edge and IoT environments, we propose a lightweight and adaptive defense mechanism based on Local Differential Privacy (LDP). Clients perturb model parameters locally before transmission, eliminating the need for centralized coordination or computationally expensive cryptographic protocols.

Our method introduces no additional communication overhead, as all operations are performed entirely on-device. By selectively injecting noise into *early convolutional layers*—which primarily encode general features and contain fewer parameters—we reduce computational cost while avoiding excessive privacy noise, thereby achieving layer-aware and resource-efficient protection.

To further adapt to evolving training dynamics and adversarial threats, we incorporate a *feedback-driven noise modulation mechanism* that dynamically adjusts noise levels in response to real-time indicators such as model accuracy and reconstruction risk. This design remains fully local and lightweight, ensuring suitability for deployment in IoT environments, including wearable sensors, smart home hubs, and industrial edge nodes.

Compared to static or uniform-noise defenses, our approach features:

- **Selective layer-wise protection**, minimizing unnecessary overhead.
- **Real-time adaptive noise scheduling**, enhancing robustness.
- **IoT-oriented design**, ensuring low computational and communication costs.

In summary, our method provides a modular, adaptive, and low-overhead privacy solution tailored for CFL in *heterogeneous IoT settings*.

Algorithm 3 Adaptive Privacy Defense in CFL Clients

Require: Total rounds T, initial privacy budget ϵ_0, thresholds θ_1 (accuracy), θ_2 (similarity)

for each round $i = 1$ to T **do**
  for each benign client c **do**
    Receive cluster model θ_j^i from server
    Train local model on private data, obtain update θ_c^{i+1}
    Estimate local accuracy A and reconstruction similarity s
    (1) Privacy Budget Decay:
    Update ϵ using decay function (Eq. 7)
    (2) Adaptive Control:
    if $A > \theta_1$ **and** $s > \theta_2$ **then**
      Set noise probability $p \leftarrow p_{\text{high}}$, sensitivity $s_c \leftarrow s_c + \delta_1$
    else if $A \leq \theta_1$ **and** $s > \theta_2$ **then**
      Set noise probability $p \leftarrow p_{\text{med}}$, sensitivity $s_c \leftarrow s_c - \delta_2$
    else
      Maintain p and s_c
    end if
    (3) Selective Noise Injection:
    if random$(0, 1) < p$ **then**
      Inject Gaussian noise $\mathcal{N}(0, s_c/\epsilon)$ to sensitive layers
    end if
    Upload perturbed update $\theta_{c,\text{DP}}^{i+1}$ to server
  end for
end for

V. Experiments

A. Datasets and Experiment settings

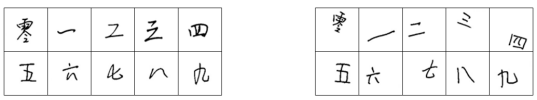

Fig. 8. Handwriting written by adults (left) and children (right) respectively.

1) Datasets: We evaluate our reconstruction attack and defense mechanisms using two datasets under a controlled yet representative experimental setup.

The first dataset is the standard MNIST dataset, containing 60,000 training and 10,000 test grayscale images of handwritten digits (0–9), resized to 32×32 pixels. To simulate a non-IID setting, we apply random rotations to a subset of images, introducing feature skew and intra-class variability, thereby mimicking realistic distribution shifts in federated learning.

The second dataset, *CMNIST*, is a custom collection of 5,400 grayscale images of handwritten Chinese numerals (0–9), sourced from two demographic groups: adults and children. This natural division forms semantically meaningful clusters, reflecting the distributional heterogeneity central to Clustered Federated Learning (CFL). Fig. 8 illustrates this distinction, with samples from adults on the left and children on the right.

We focus on handwritten digit datasets due to their visual interpretability and established use in evaluating generative models. CMNIST further enables cluster-level heterogeneity, aligning with the CFL framework and allowing controlled evaluation of privacy risks both across and within clusters.

Justification of Dataset Selection. While MNIST and CMNIST are canonical benchmarks, they allow controlled evaluation of reconstruction quality and privacy–utility trade-offs without external biases from heterogeneous real-world datasets. Such controlled settings are widely adopted in recent IoT-FL works [32], ensuring fair comparison and reproducibil-

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2025.3637061

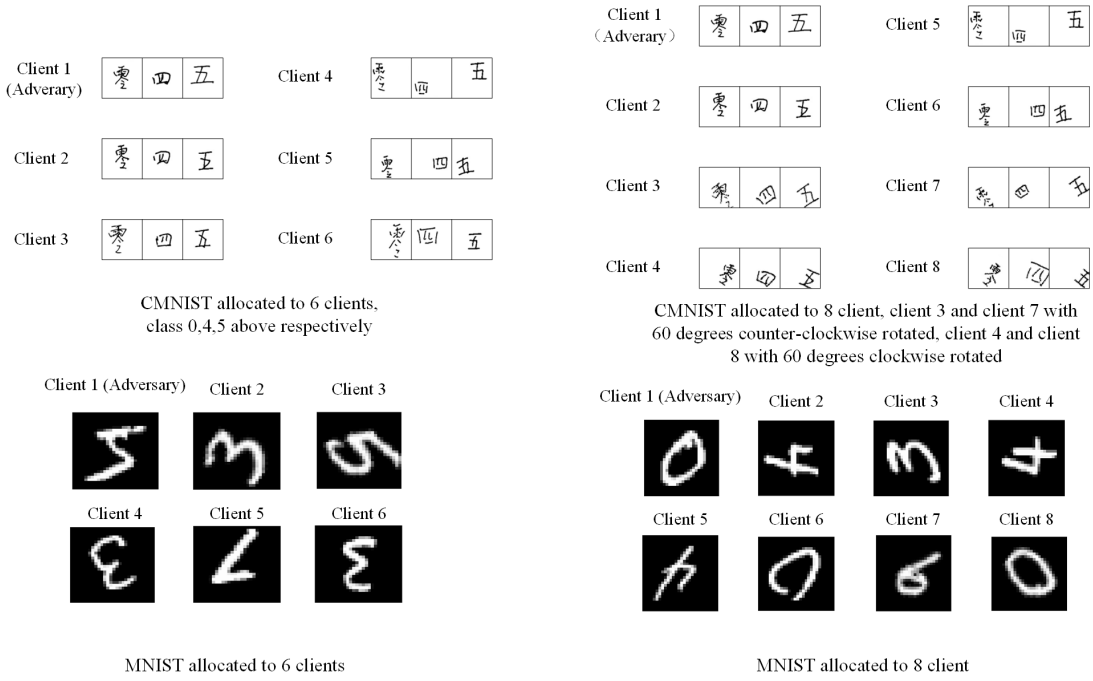

Fig. 9. Allocation of datasets.

ity. Future work will extend the evaluation to more complex IoT datasets (e.g., CIC-IDS2017, IoT23) once large-scale clustered data become available.

2) Model architecture and hyper-parameters: In our experiments, we used a convolutional neural network (CNN) for classification tasks. For the CMNIST dataset, the input size is 64×64, and for MNIST, it is 32×32. The network consists of three convolutional layers with 32, 64, and 128 filters, each using 3×3 kernels, strides of 2, and Leaky ReLU activations, followed by max pooling and two fully connected layers with 256 neurons each, ending with a 10-neuron softmax layer. The discriminator has an 11-neuron output layer and the generator uses three deconvolutional layers, generating 32x32 images for MNIST and 64x64 images for CMNIST.

For MNIST, each client trains for two local epochs per round with an initial learning rate of 10^{-4}, decaying by 10^{-7} per round, over a total of 100 rounds. The adversary's GAN is trained for 10 epochs per round, with initial learning rates of 8×10^{-5} for the discriminator and 8×10^{-4} for the generator, both decaying by 10×10^{-7} per round.

For CMNIST, the learning rate are $8 * 10^{-4}$ for the local model and discriminator, and $8 * 10^{-5}$ for the generator, all decaying by 10^{-7}. The remaining hyperparameters are similar to those used for MNIST.

3) Experiment setting: We conducted two sets of experiments. In the CFL setting, the adversary can only target clients within its own cluster; therefore, we evaluate attack effectiveness by varying the dataset and the number of clients per cluster.

Experiment 1 (6 clients, 2 clusters): As shown in Fig. 9, for the CMNIST dataset, client 1-3 are assigned data featuring adults, whereas the remaining three clients receive child-centric data. Each client possesses data covering all 10 labels, and client 1 is designated as adversary.

For the MNIST dataset, we evenly distribute the entire training set among six clients, introducing feature skew via rotations. Specifically, the data of clients 2 and 3 are rotated 90 degrees clockwise and counterclockwise, respectively, while the data of clients 4, 5, and 6 are rotated 180 degrees.

Experiment 2 (8 clients, 2 clusters — 4 clients per cluster): For the CMNIST dataset, we applied similar data augmentation techniques: client 3 and client 4 were rotated 60 degrees clockwise and counterclockwise, respectively. For MNIST, additional rotation angles were applied to simulate further feature skew. Client 1 is again set as the adversary.

B. Client-level Attack

In this section, we evaluate the performance of our client-level data reconstruction attack. Within the CFL setting, the adversary systematically targets clients inside its own cluster, one victim at a time. As shown in Fig. 10, the reconstructed samples closely resemble the characteristics of the targeted client's data.

For CMNIST, attacks on specific clients, such as client 2 and client 3, reveal distinct features, such as skewed versus straight lines, highlighting the adversary's ability to link attributes to individual clients. Similar results are observed for MNIST, where rotated data produce characteristic reconstruction patterns.

Further analysis demonstrates that the attack remains effective even with limited parameter eavesdropping. As illustrated in Fig. 11, access to only a subset of the victim's parameters is sufficient for achieving notable reconstruction quality. Table I shows that as the eavesdropping level increases to approximately 60%, the similarity of the reconstructed data surpasses

This article has been accepted for publication in IEEE Internet of Things Journal. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2025.3637061

TABLE I
SIMILARITY OF DIFFERENT RATIO OF EAVESDROPPING.

Scenarios / Datasets	6 clients						
	FL	CFL(0%)	CFL-20%	CFL-40%	CFL-60%	CFL-80%	CFL-100%
CMNIST	0.5	0.727	0.704	0.702	0.719	0.738	0.758
MNIST	0.413	0.553	0.545	0.553	0.562	0.571	0.575
Scenario / Datasets	8 clients						
	FL	CFL(0%)	CFL-20%	CFL-40%	CFL-60%	CFL-80%	CFL-100%
CMNIST	0.618	0.693	0.677	0.697	0.705	0.728	0.722
MNIST	0.386	0.523	0.532	0.536	0.544	0.56	0.572

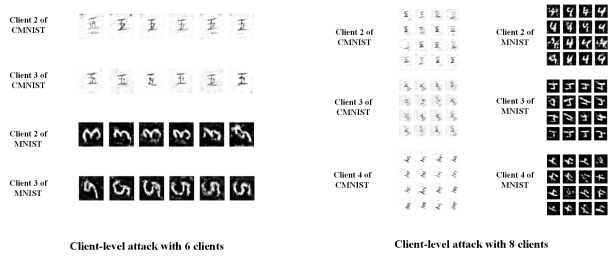

Fig. 10. Client-level attack.

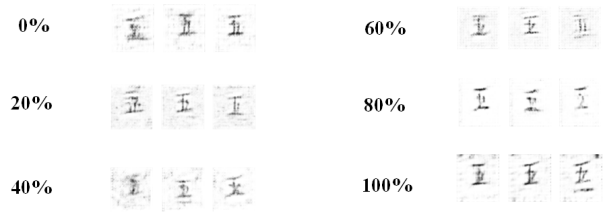

Fig. 11. Effects of different ratio of eavesdropping.

that observed in the baseline CFL scenario. Importantly, even at lower levels of parameter access, the attack's effectiveness exceeds that in conventional FL settings, although it remains slightly below the performance achieved in full CFL cases.

These findings underscore the urgent need for robust defenses in CFL environments.

C. Evaluation of Defense Effectiveness and Resource Trade-offs

To comprehensively evaluate the effectiveness and practicality of our proposed defense mechanism, we compare three strategies in terms of accuracy, privacy protection, and resource efficiency:

- **No Defense** (baseline),
- **Plain Differential Privacy** (fixed noise with $\epsilon = 1.0$),
- **Our Adaptive Defense** (budget decay + selective layer-wise noise injection).

1) Privacy-Utility-Efficiency Trade-off: Fig. 12 presents a radar chart summarizing the performance of each method across four key metrics:

- **Accuracy**: Post-defense model classification performance.
- **Privacy Protection**: Measured as $1 -$ cosine similarity between reconstructed and true samples.
- **Memory Efficiency**: Inverse of memory usage (MB).
- **FLOPs Efficiency**: Inverse of floating-point operation cost.

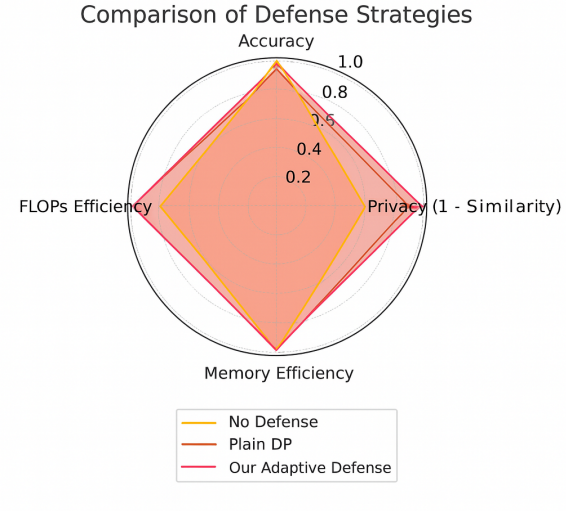

Fig. 12. Radar chart comparing three defense strategies in terms of accuracy, privacy, and resource efficiency.

Compared with Plain DP, our adaptive defense exhibits a superior overall balance:

- It reduces cosine similarity by **28.7%** compared to no defense (from 0.689 to 0.491),
- Maintains only a **3.3%** drop in accuracy relative to the unprotected model (84.6% $\rightarrow$ 81.3%),
- Achieves **23.4%** lower memory usage and **43%** fewer FLOPs than Plain DP,
- Requires no additional communication cost.

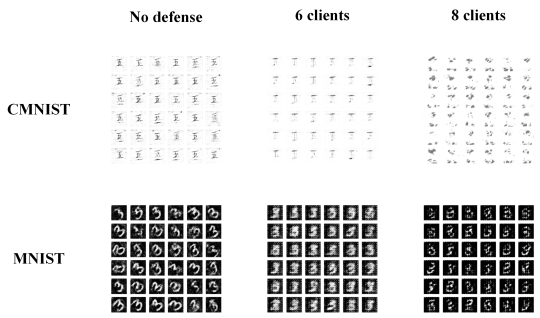

Fig. 13. Defense effect. Take class 5 in CMNIST and class 3 in MNIST as examples, respectively.

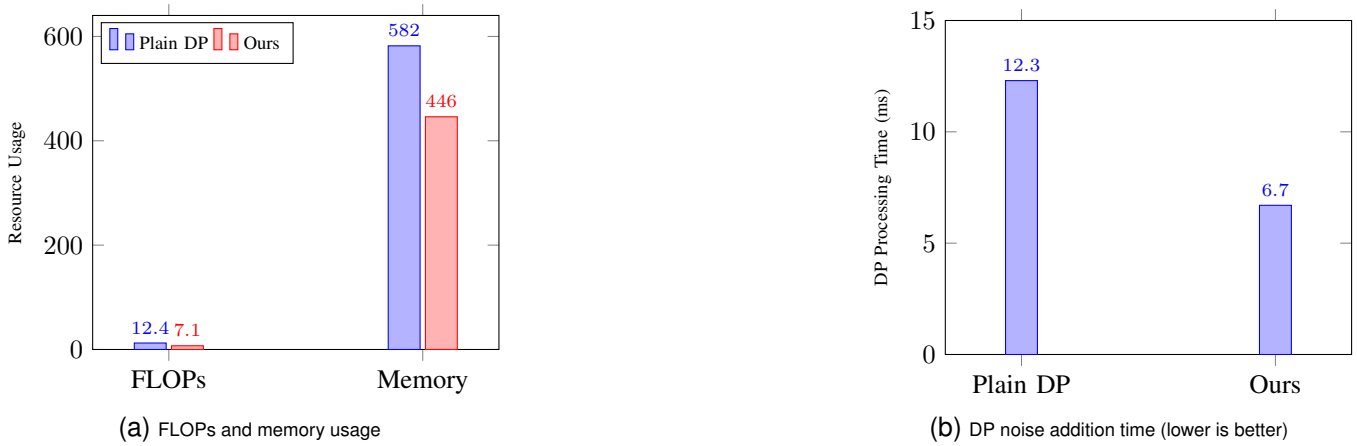

Fig. 14. Comparison between baseline and our method

2) Visual Evaluation of Reconstructed Samples: As illustrated in Fig. 13, without defense, reconstructed samples retain strong personal characteristics. Plain DP reduces fidelity but introduces visible noise. In contrast, our method effectively obfuscates user-specific traits while preserving semantic structure, thereby balancing utility and privacy.

3) Lightweight Performance in IoT Context: Fig. 14 summarizes defense overhead. Our selective perturbation of sensitive layers, guided by dynamic similarity estimation, leads to a substantial reduction in computational load and DP noise processing time—by **23.3%** and **45.5%**, respectively. All operations are performed locally on the client, making the method ideal for low-power federated deployments.

These findings validate our method as a robust and efficient privacy-preserving mechanism, particularly well suited for CFL in resource-constrained environments such as IoT.

VI. Conclusion

In this study, we identify and empirically validate a critical vulnerability in Clustered Federated Learning (CFL) systems. We propose a client-level data reconstruction attack that leverages passive eavesdropping and conditional GANs to exploit intra-cluster similarities, thereby reconstructing private training data. Unlike conventional server-centric or full-gradient attacks, our approach operates under realistic constraints—requiring only partial parameter access from within the victim's cluster—yet achieves notably higher reconstruction fidelity.

This threat model is particularly relevant for practical CFL deployments, where clients often communicate over insecure channels and cluster-based training is increasingly adopted. Our findings underscore the urgency of revisiting privacy assumptions in CFL and of developing more robust, context-aware defenses.

To this end, we propose a lightweight and adaptive defense mechanism grounded in Local Differential Privacy (LDP). By dynamically adjusting privacy budgets and applying layer-wise protection, our method effectively balances privacy preservation with model utility, while remaining feasible for deployment in resource-constrained environments such as IoT. Moreover, The proposed layer-wise LDP mechanism is compatible with secure aggregation, homomorphic encryption, or blockchain-based client verification schemes. For example, recent works in IoT-FL contexts integrate federated learning with blockchain architectures to ensure auditability and tamper-resistance [33], [34]. Such hybrid integration can enhance robustness without modifying the local privacy module. We plan to explore such integration in future deployments.

Overall, the proposed attack and defense framework exposes previously underestimated risks in CFL systems and provides a foundation for developing more nuanced, scalable, and practical privacy-preserving strategies in future federated learning research.

Acknowledgments

This study was funded by Anhui Provincial Natural Science Foundation (No.2308085MF212) and Anhui Provincial Key Research and Development Project (Grant No.202304a05020059).

References

[1] Y. Zhou, M. Shi, Y. Tian, Y. Li, Q. Ye, and J. Lv, "Federated cinn clustering for accurate clustered federated learning," in *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2024, pp. 5590–5594.

[2] R. Zeng, B. Mi, and D. Huang, "A federated learning framework based on csp homomorphic encryption," in *2023 IEEE 12th Data Driven Control and Learning Systems Conference (DDCLS)*. IEEE, 2023, pp. 196–201.

[3] F. Sattler, K.-R. Müller, and W. Samek, "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints," *IEEE transactions on neural networks and learning systems*, vol. 32, no. 8, pp. 3710–3722, 2020.

[4] M. Kamal, I. Rashid, W. Iqbal, M. H. Siddiqui, S. Khan, and I. Ahmad, "Privacy and security federated reference architecture for internet of things," *Frontiers of Information Technology & Electronic Engineering*, vol. 24, no. 4, pp. 481–508, 2023.

[5] L. Zhu, Z. Liu, and S. Han, "Deep leakage from gradients," *Advances in neural information processing systems*, vol. 32, 2019.

[6] B. Zhao, K. R. Mopuri, and H. Bilen, "idlg: Improved deep leakage from gradients," *arXiv preprint arXiv:2001.02610*, 2020.

[7] B. Hitaj, G. Ateniese, and F. Perez-Cruz, "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, pp. 603–618.

[8] E. Dritsas and M. Trigka, "Federated learning for iot: A survey of techniques, challenges, and applications," *Journal of Sensor and Actuator Networks*, vol. 14, no. 1, p. 9, 2025.

[9] Z. Wang, M. Song, Z. Zhang, Y. Song, Q. Wang, and H. Qi, "Beyond inferring class representatives: User-level privacy leakage from federated learning," in *IEEE INFOCOM 2019-IEEE conference on computer communications*. IEEE, 2019, pp. 2512–2520.

[10] J. C. Zhao, A. Sharma, A. R. Elkordy, Y. H. Ezzeldin, S. Avestimehr, and S. Bagchi, "Loki: Large-scale data reconstruction attack against federated learning through model manipulation," in *2024 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2024, pp. 1287–1305.

[11] J. Geiping, H. Bauermeister, H. Dröge, and M. Moeller, "Inverting gradients-how easy is it to break privacy in federated learning?" *Advances in neural information processing systems*, vol. 33, pp. 16 937–16 947, 2020.

[12] W. Wei, L. Liu, M. Loper, K.-H. Chow, M. E. Gursoy, S. Truex, and Y. Wu, "A framework for evaluating gradient leakage attacks in federated learning," *arXiv preprint arXiv:2004.10397*, 2020.

[13] J. Zhu and M. Blaschko, "R-gap: Recursive gradient attack on privacy," *arXiv preprint arXiv:2010.07733*, 2020.

[14] C. Chen and N. D. Campbell, "Understanding training-data leakage from gradients in neural networks for image classification," *arXiv preprint arXiv:2111.10178*, 2021.

[15] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, and C.-Z. Xu, "Feddc: Federated learning with non-iid data via local drift decoupling and correction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 112–10 121.

[16] W. Sun, B. Gao, K. Xiong, Y. Lu, and Y. Wang, "Vaguegan: a gan-based data poisoning attack against federated learning systems," in *2023 20th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2023, pp. 321–329.

[17] J. C. Zhao, S. Bagchi, S. Avestimehr, K. S. Chan, S. Chaterji, D. Dimitriadis, J. Li, N. Li, A. Nourian, and H. R. Roth, "Federated learning privacy: Attacks, defenses, applications, and policy landscape-a survey," *CoRR*, 2024.

[18] Q. Tan, Q. Li, Y. Zhao, Z. Liu, X. Guo, and K. Xu, "Defending against data reconstruction attacks in federated learning: An information theory approach," in *33rd USENIX Security Symposium (USENIX Security 24)*, 2024, pp. 325–342.

[19] D. Yao, S. Li, X. Gong, S. Hou, and G. Pan, "Urvfl: Undetectable data reconstruction attack on vertical federated learning," *arXiv preprint arXiv:2404.19582*, 2024.

[20] C. Zhao, S. Li, Y. He, W. Huang, G. Li, L. Ding, and J. Li, "Gendra: Generative data reconstruction attacks on federated edge learning and countermeasures," *Electronics*, vol. 14, no. 11, p. 2263, 2025.

[21] F. Mosaiyebzadeh, S. Pouriyeh, R. M. Parizi, Q. Z. Sheng, M. Han, L. Zhao, G. Sannino, C. M. Ranieri, J. Ueyama, and D. M. Batista, "Privacy-enhancing technologies in federated learning for the internet of healthcare things: a survey," *Electronics*, vol. 12, no. 12, p. 2703, 2023.

[22] Y. Yang, B. Hui, H. Yuan, N. Gong, and Y. Cao, "{PrivateFL}: Accurate, differentially private federated learning via personalized data transformation," in *32nd USENIX Security Symposium (USENIX Security 23)*, 2023, pp. 1595–1612.

[23] Q. Chen, H. Wang, Z. Wang, J. Chen, H. Yan, and X. Lin, "Lldp: A layer-wise local differential privacy in federated learning," in *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2022, pp. 631–637.

[24] J. Hong, Z. Wang, and J. Zhou, "Dynamic privacy budget allocation improves data efficiency of differentially private gradient descent," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, 2022, pp. 11–35.

[25] L. Cui and X. Wu, "Aldp-fl for adaptive local differential privacy in federated learning," *Scientific Reports*, vol. 15, no. 1, p. 26679, 2025.

[26] G. Yue, L. Yan, L. Kang, and C. Shen, "Adapldp-fl: An adaptive local differential privacy for federated learning," *IEEE Transactions on Mobile Computing*, 2025.

[27] Y. Zhang, H. Zhang, Y. Yang, W. Sun, H. Zhang, and Y. Fu, "Adaptive differential privacy in asynchronous federated learning for aerial-aided edge computing," *Journal of Network and Computer Applications*, vol. 235, p. 104087, 2025.

[28] P. Ye, Z. Jiang, W. Wang, B. Li, and B. Li, "Feature reconstruction attacks and countermeasures of dnn training in vertical federated learning," *IEEE Transactions on Dependable and Secure Computing*, 2024.

[29] Y. Song, Z. Wang, and E. Zuazua, "Approximate and weighted data reconstruction attack in federated learning," *arXiv preprint arXiv:2308.06822*, 2023.

[30] P. J. Huber, "Robust estimation of a location parameter," in *Breakthroughs in statistics: Methodology and distribution*. Springer, 1992, pp. 492–518.

[31] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 2016, pp. 308–318.

[32] J. Li, M. Lu, J. Zhang, and J. Wu, "Aldp-fl: an adaptive local differential privacy-based federated learning mechanism for iot," *International Journal of Information Security*, vol. 24, no. 1, p. 23, 2025.

[33] N. Waheed, A. U. Rehman, A. Nehra, M. Farooq, N. Tariq, M. A. Jan, F. Khan, A. Z. Alalmaie, and P. Nanda, "Fedblockhealth: A synergistic approach to privacy and security in iot-enabled healthcare through federated learning and blockchain," in *GLOBECOM 2023-2023 IEEE Global Communications Conference*. IEEE, 2023, pp. 3855–3860.

[34] M. Shawkat, A. El-desoky, Z. H. Ali, and M. Salem, "Blockchain and federated learning based on aggregation techniques for industrial iot: A contemporary survey," *Peer-to-Peer Networking and Applications*, vol. 18, no. 4, p. 192, 2025.

VII. Biography

Lei Shi received his B.S. degree in 2002, M.S. degree in 2005, and Ph.D. degree in 2012, all from Hefei University of Technology, Hefei, Anhui, China. He is currently an associate professor in School of Computer Science and Information Engineering, Hefei University of Technology. His main research area lies in federated learning, edge computing and wireless network optimization.

Han Wu is currently pursuing the M.S. degree in Computer Science and Technology at Hefei University of Technology, located in Hefei, China. His research focuses on cutting-edge areas such as edge computing and federated learning, aiming to contribute to the development of more efficient and privacy-preserving computational models.

Xu Ding received B.S. and Ph.D. degrees from the School of Computer and Information, Hefei University of Technology in 2006 and 2015.Now he is an associate research fellow with the Institute of Industry and Equipment Technology, Hefei University of Technology. His researchfield mainly lies in wireless communications and wireless sensor networks.

Hao Xu is currently pursuing the M.S. degree in Computer Science and Technology at Hefei University of Technology, located in Hefei, China. His research focuses on cutting-edge areas such as edge computing and federated learning, aiming to contribute to the development of more efficient and privacy-preserving computational models.

Sinan Pan is currently pursuing the M.S. degree in Computer Science and Technology at Hefei University of Technology, located in Hefei, China. Her research focuses on cutting-edge areas such as edge computing and federated learning, aiming to contribute to the development of more efficient and privacy-preserving computational models.