Distributionally Robust Federated Learning for Differentially Private Data

Siping Shi[1], Chuang Hu[2], Dan Wang[1], Yifei Zhu[3], Zhu Han[4]

1: The Hong Kong Polytechnic University

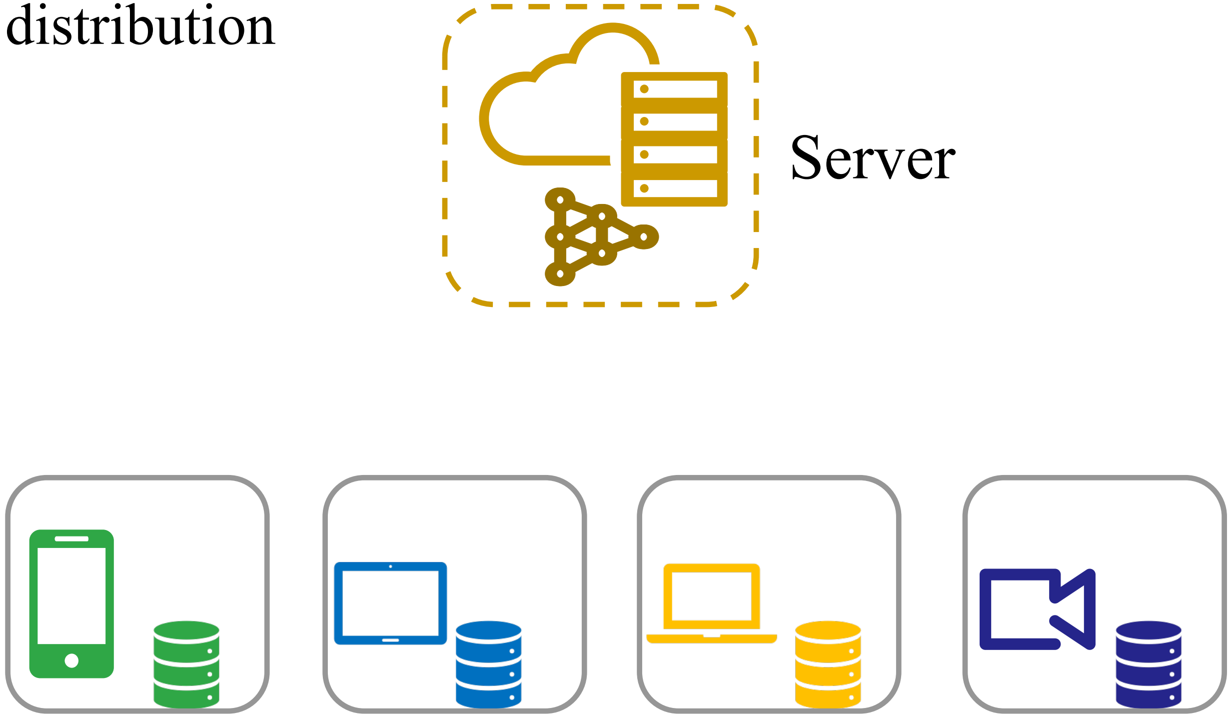
2: Wuhan University

3: Shanghai Jiaotong University

4: University of Houston

Background

- Federated learning
  - A promising learning paradigm proposed to protect user data privacy
  - Collaboratively learn a model while keeping all the data in local
    - Global model distribution

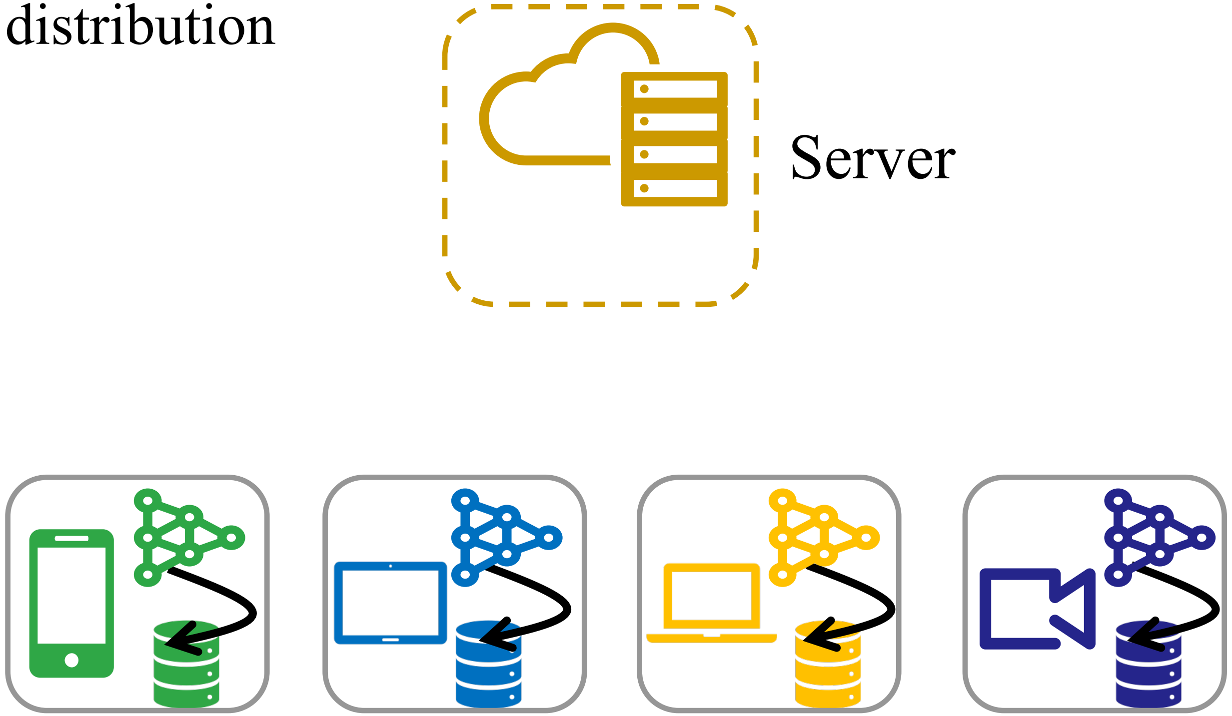


TensorFlow

FATE

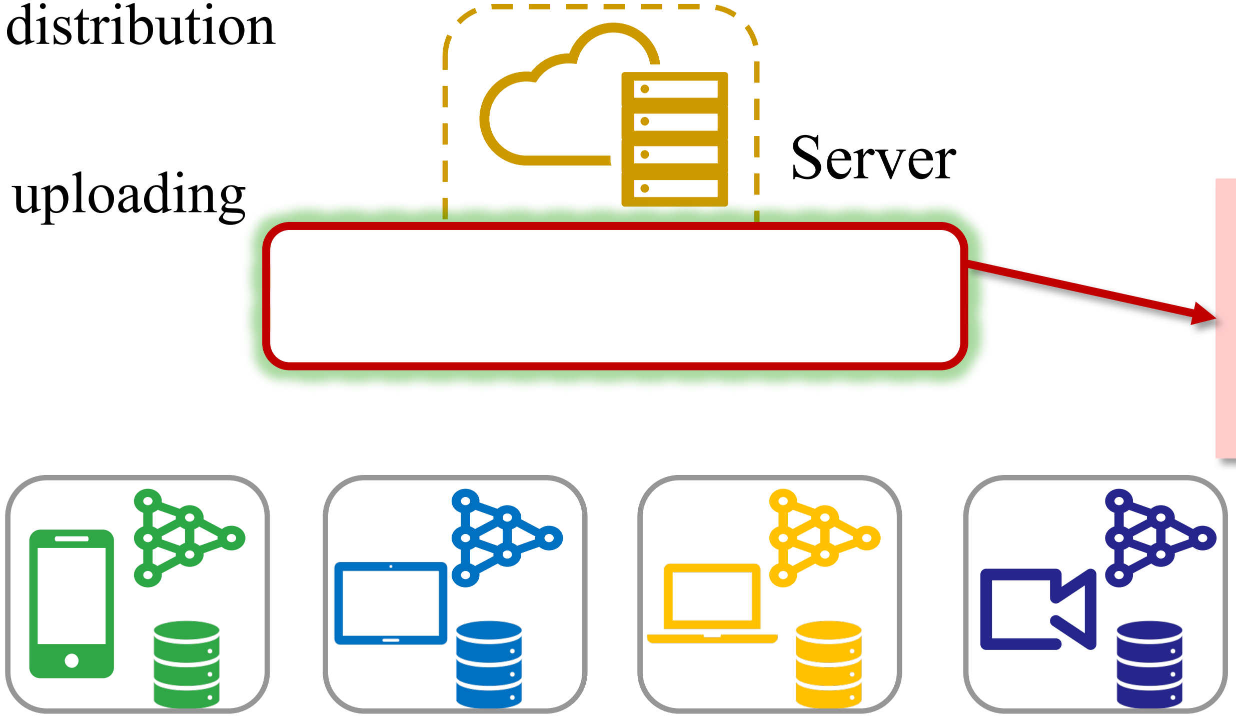
Background

- Federated learning
  - A promising learning paradigm proposed to protect user data privacy
  - Collaboratively learn a model while keeping all the data in local
    - Global model distribution
    - Local training



Background

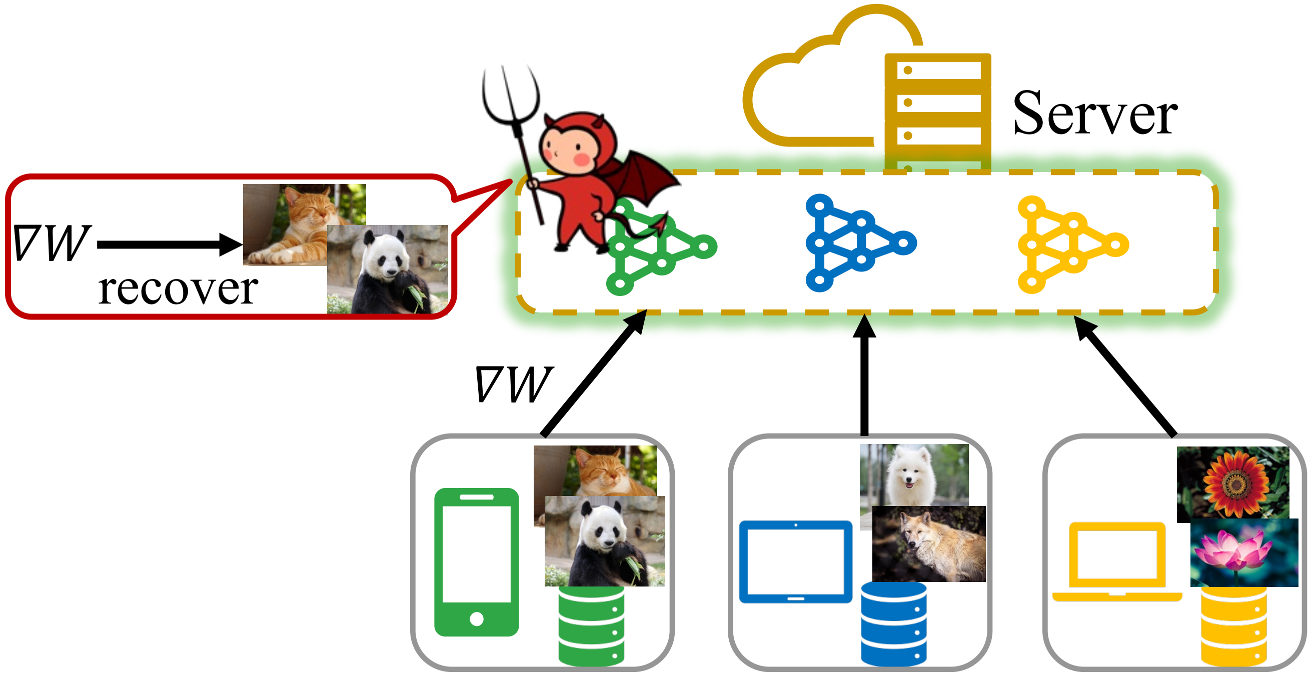
- Federated learning
  - A promising learning paradigm proposed to protect user data privacy
  - Collaboratively learn a model while keeping all the data in local
    - Global model distribution
    - Local training
    - Local updates uploading



Sharing updates rather than local data...
Privacy leakage from model updates

Background

- Privacy leakage from model updates
  - Untrusted server or malicious third-parties has access to the model updates
  - Adversaries attempt to infer the sensitive information from model updates
    - Model inversion attack[1]
    - Property inference attack[2]
    - Membership inference attack[3]



[1] J. Geiping, H. Bauermeister, H. Dr¨oge, and M. Moeller, "Inverting gradients - how easy is it to break privacy in federated learning?" in Proc. of NeurIPS'20, 2020.
[2] L. Melis, C. Song, E. De Cristofaro and V. Shmatikov, "Exploiting Unintended Feature Leakage in Collaborative Learning," in Proc. of IEEE SP'19, 2019.
[3] M. Nasr, R. Shokri and A. Houmansadr, "Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning," in Proc. of IEEE SP'19, 2019.

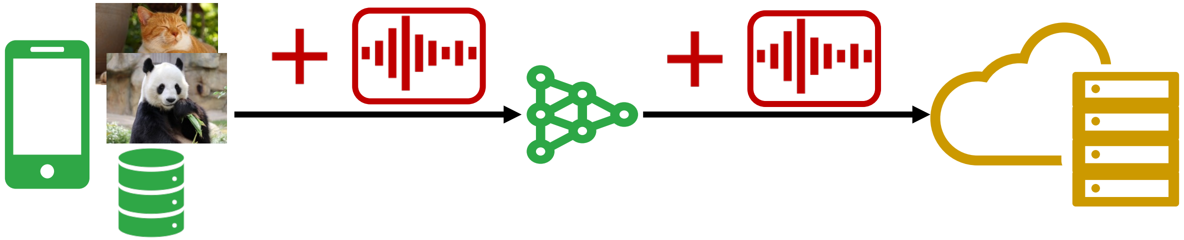
Background

- Privacy protection with local differential privacy (LDP)
  - (ϵ, δ)-LDP

A random mechanism M satisfies (ϵ, δ)-LDP if, for any two input data x and x', and any possible output y,

$$\Pr(M(x) = y) \leq e^{\epsilon} \Pr(x' = y) + \delta$$

  - Add Gaussian noise to the local data[1] or local model updates[2] to achieve LDP



  - Smaller privacy budget ϵ indicates stronger privacy guarantee and leads to higher level of perturbation

[1] K. Fukuchi, Q. K. Tran, and J. Sakuma, "Differentially private empirical risk minimization with input perturbation," in Proc. of DS'17, 2017
[2] Y. Zhao et al., "Local Differential Privacy-Based Federated Learning for Internet of Things," in IEEE Internet of Things Journal, 2021

Background

- Robustness of federated learning is affected by LDP
  - Robustness mainly refers to the capability to defend against adversarial attacks, e.g., backdoor attacks, poisoning attacks, etc.
  - Positive effect: increased robustness with higher level of perturbation
  - Negative effect: decreased robustness with higher level of perturbation after accessing the balance area

How to *allocate privacy budget* to balance the robustness and privacy in federated learning?

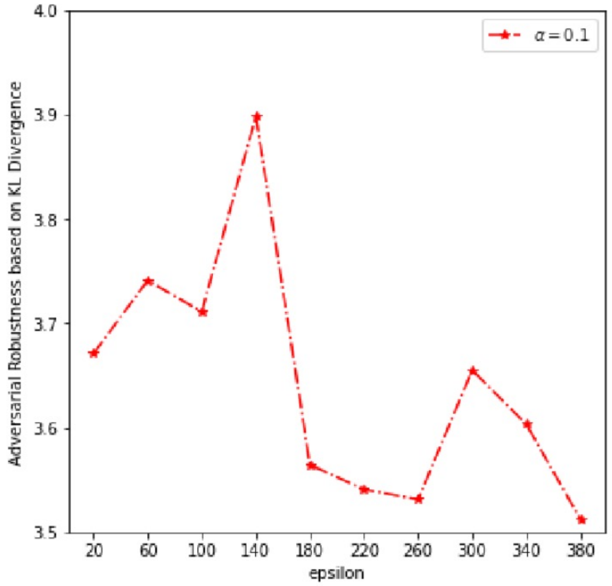


Figure 1. Robustness vs. Privacy[1]

[1] Yaowei Han, Yang Cao, and Masatoshi Yoshikawa. "Understanding the interplay between privacy and robustness in federated learning". in arXiv, 2021.

Motivation

- Jointly consider robustness and privacy in federated learning
  - LDP noise introduces uncertainty to the local training data
  - Adversarial attacks generating noisy data also cause data uncertainty
  - Connecting data uncertainty with privacy budget and adversarial attacks
  - Collaboratively training a model under data uncertainty

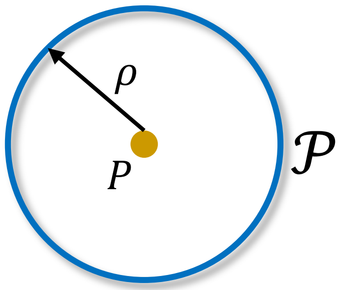
- Challenge
  - How to express data uncertainty explicitly?
  - How to learn a model with data uncertainty?

Motivation

- Our idea
  - Leverage Distributionally Robust Optimization (DRO)
    - DRO enables to model the problem under uncertainty
    - Aiming to find a solution x that minimizes the worst-case cost under all possible distributions in the uncertainty set $\mathcal{P}$:

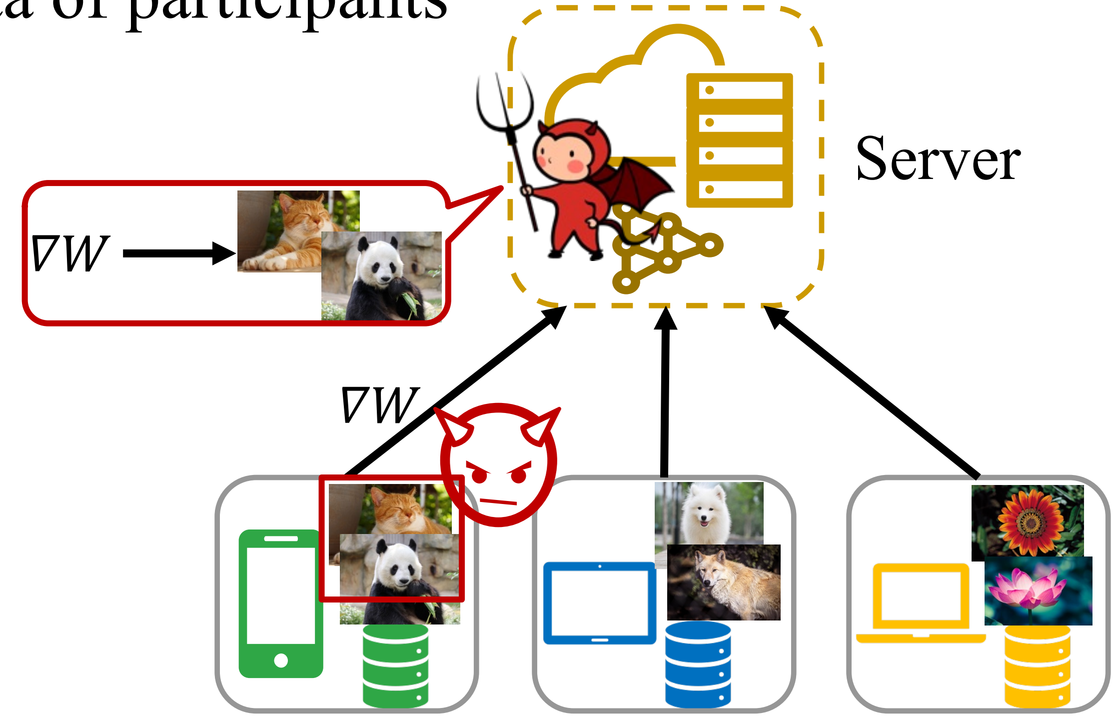
$$\min_{x\in\chi} \max_{P\in\mathcal{P}} E_P[h(x,\xi)]$$

  - Construct the uncertainty set with Wasserstein Distance
    - $\mathcal{P} = \mathbb{B}_\rho(P) = \{Q: \underbrace{W_p(Q,P)}_{\text{Wasserstein distance}} \leq \rho\}$



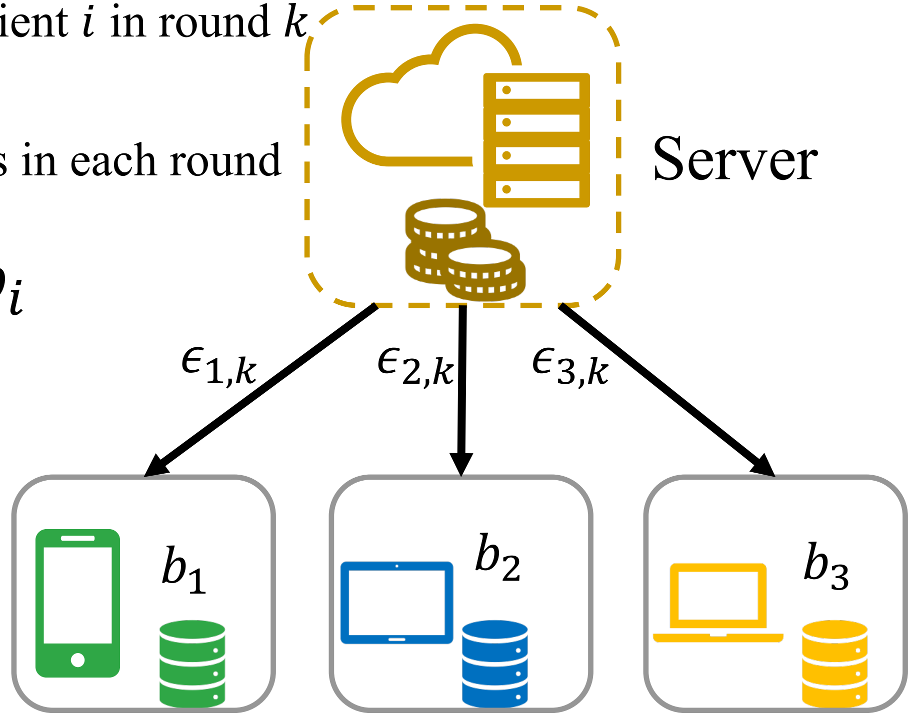
Distributionally Robust and Private FL Problem

- Threat model
  - Server: curious but honest
  - Attacker: manipulate the training data of participants
    - Poison the training sample
    - Manipulate the label of training sample



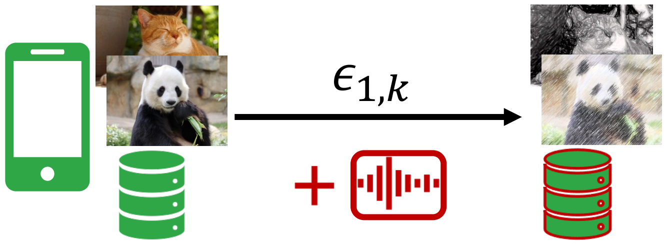
Distributionally Robust and Private FL Problem

- Privacy budget allocation model
  - Monetary reward is required to incentivize local clients to contribute private information of data with some degree of privacy loss
    - $\sum_{i=1}^{N} \epsilon_{i,k} \cdot p \leq V,$
      - $\epsilon_{i,k}$: allocated privacy budget for client i in round k
      - p: unit price of privacy loss
      - V: monetary budget for privacy loss in each round
  - Each client has the baseline of privacy loss b_i
    - $\epsilon_{i,k} \leq b_i$



Distributionally Robust and Private FL Problem

- Local differentially private data model
  - Add Gaussian noise to each data sample locally
    - Original local dataset: $D_i \triangleq \{(x_j, y_j)\}_{j=1}^{d_i}$
    - Differentially private local data set: $\tilde{D}_i \triangleq \{(\tilde{x}_j, y_j)\}_{j=1}^{d_i}$
      - $\tilde{x}_j := x_j + w_{i,k}$, where $w_{i,k} \sim N(0, \sigma_{i,k}^2)$



- Distributionally robust local training model
  - Uncertainty set $\mathcal{P}_i$ consists of probability distributions generated from $\tilde{D}_i$
  - To minimize the worst-case expected loss over $\mathcal{P}_i$
    - $\widehat{J}_i := \min_{\theta_i} \sup_{G_i \in \mathcal{P}_i} \mathrm{E}^{\mathrm{G}_i}\{\ell(x, \theta_i, y)\}$

Distributionally Robust and Private FL Problem

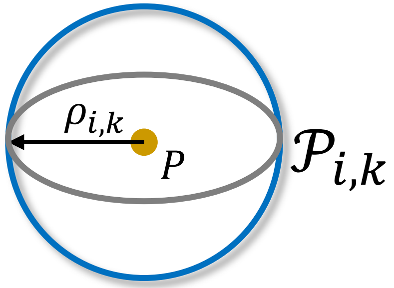
- Problem formulation
  - To maximize the model performance with limited privacy budget and individual privacy requirement constraints.
  - DRPri problem is formulated as:

$$
\begin{aligned}
\min_{\epsilon,\theta} \quad & \sum_{i=1}^{N} \frac{d_i}{D} \sup_{\mathbb{G}_{i,k} \in \boxed{\mathcal{P}_{i,k}}} \mathbb{E}^{\mathbb{G}_{i,k}} \{\ell(x,\theta,y)\} \\
s.t. \quad & \sum_{i=1}^{N} \epsilon_{i,k} \cdot p \leq V, \quad \forall k = 1, \ldots, K, \\
& \epsilon_{i,k} \leq b_i, \quad \forall k = 1, \ldots, K, \\
& \epsilon_{i,k} \geq 0, \quad \forall k = 1, \ldots, K.
\end{aligned}
$$

How to construct the uncertainty set $\mathcal{P}_{i,k}$?

Uncertainty Set Construction for DRPri Problem

- Wasserstein distance based uncertainty set
  - A Wasserstein ball with radius $\rho_{i,k}$ around the empirical distribution $\tilde{\mathbb{D}}_{i,k}$ which represents the local noisy dataset $\tilde{D}_{i,k}$.
    - $\mathcal{P}_{i,k} = \{\mathbb{G}_{i,k}\colon \mathrm{W}_1(\mathbb{G}_{i,k}, \tilde{\mathbb{D}}_i) \le \rho_{i,k}\}$



  - Choose the value of $\rho_{i,k}$
    - Enabling the constructed uncertainty set $\mathcal{P}_{i,k}$ contains the true underlying local data distribution $\mathbb{P}_i$ with $1-\gamma$ confidence level
    - Obtain a proper $\rho_{i,k}$ based on the measure concentration theorem[1]

$$\min_{\epsilon,\theta} \sum_{i=1}^{N} \frac{d_i}{D} \sup_{\boxed{\mathbb{G}_{i,k} \in \mathcal{P}_{i,k}}} \mathbb{E}^{\mathbb{G}_{i,k}} \{\ell(x,\theta,y)\} \rightarrow \min_{\epsilon,\theta} \sum_{i=1}^{N} \frac{d_i}{D} \sup_{\boxed{\mathbb{G}_{i,k}:\mathrm{W}_1(\mathbb{G}_{i,k},\tilde{\mathbb{D}}_i) \le \rho_{i,k}}} \mathbb{E}^{\mathbb{G}_{i,k}} \{\ell(x,\theta,y)\}$$

[1] N. Fournier and A. Guillin, “On the rate of convergence in wasserstein distance of the empirical measure,” in Probability Theory and Related Fields, 2013.

Tractable Reformulation of DRPri Problem

- Reduce the computation overhead of the inner worst-case expectation problem of DRPri
  - Lipschitz continuous loss function assumption

Assumption 1*. The loss function ℓ is $G(\theta)$-Lipschitz continuous: for all ξ_1 and ξ_2,*

$$|\ell(\theta,\xi_1) - \ell(\theta,\xi_1)| \leq G(\theta) \cdot \|\xi_1 - \xi_2\|.$$

  - The upper bound for the worst-case expectation problem

Lemma 1*. Let Assumption 1 hold, then*

$$\sup_{\mathbb{G}:W_1(\mathbb{G},\tilde{D})\leq\rho} \mathbb{E}^{\mathbb{G}}\{\ell(x,\theta,y)\} \leq \mathbb{E}^{\tilde{D}}\{\ell(x,\theta,y)\} + G(\theta) \cdot \rho.$$

Tractable Reformulation of DRPri Problem

- Reformulate DRPri problem with *Lemma 1* to DRPri-W problem

$$\min_{\epsilon,\theta} \quad \sum_{i=1}^{N} \frac{d_i}{D} \boxed{\sup_{\mathbb{G}_{i,k}\in\mathcal{P}_{i,k}} \mathbb{E}^{\mathbb{G}_{i,k}}\{\ell(x,\theta,y)\}}$$
$$\text{s.t.} \quad \sum_{i=1}^{N} \epsilon_{i,k}\cdot p \le V, \quad \forall k = 1,\ldots,K,$$
$$\epsilon_{i,k} \le b_i, \quad \forall k = 1,\ldots,K,$$
$$\epsilon_{i,k} \ge 0, \quad \forall k = 1,\ldots,K.$$

$$\Longrightarrow$$

$$\min_{\epsilon,\theta} \quad \sum_{i=1}^{N} \frac{d_i}{D} \boxed{\mathbb{E}^{\tilde{\mathbb{D}}_{i,k}}\{\ell(x,\theta,y)\} + \rho_{i,k}\cdot G(\theta)}$$
$$\text{s.t.} \quad \sum_{i=1}^{N} \epsilon_{i,k}\cdot p \le V, \quad \forall k = 1,\ldots,K,$$
$$\epsilon_{i,k} \le b_i, \quad \forall k = 1,\ldots,K,$$
$$\epsilon_{i,k} \ge 0, \quad \forall k = 1,\ldots,K.$$

- DRPri-W problem is computationally much easier than the DRPri problem

Robust and Private Federated Learning Algorithm

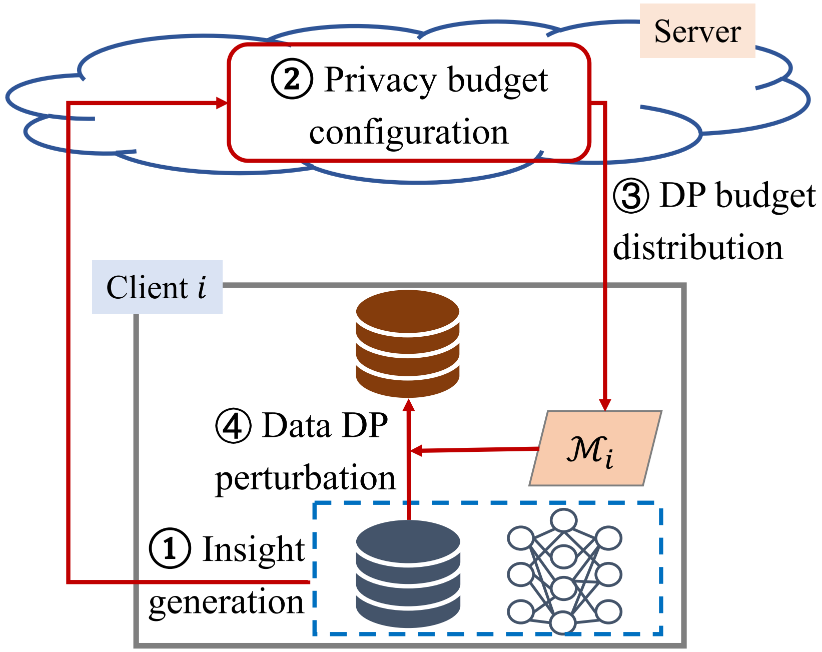
- To solve DRPri-W problem in two steps
  - Determine the privacy budget allocation strategy ϵ
    - Given a determined model $\hat{\theta}$, to derive an optimal ϵ with the DRPri-W problem

$$\min_{\epsilon} \quad \sum_{i=1}^{N} \frac{d_i}{D} \left(\frac{1}{d_i} \sum_{j=1}^{d_i} \ell(x_j, \hat{\theta}, y_j) + (\eta_i + \frac{C_i}{\epsilon_{i,k}}) G(\hat{\theta}) \right)$$

$$s.t. \quad \sum_{i=1}^{N} \epsilon_{i,k} \cdot p \le V, \quad \forall k = 1, \ldots, K,$$

$$\epsilon_{i,k} \le b_i, \quad \forall k = 1, \ldots, K,$$

$$\epsilon_{i,k} \ge 0, \quad \forall k = 1, \ldots, K.$$



    - A typical minimization problem and easy to solve.

Robust and Private Federated Learning Algorithm

- To solve DRPri-W problem in two steps
  - Derive the robust model θ with the differentially private data
    - Each client performs distributionally robust local training with the allocated privacy budget.

Algorithm 1: Robust and private federated learning algorithm (RPFL)

Input: Local dataset $\mathcal{D} = \{D_1, D_2, \ldots, D_N\}$, the objective function $\ell(\cdot)$, total privacy budget V in each round, baseline of privacy loss b_i in each client i, and the confidence level $1-\gamma$;

Output: Learned global model θ^* and the privacy budget allocation strategy ϵ^* ;

$\theta \leftarrow$ `Random(`θ`)`,
for $k = 1, 2, \ldots, K$ **do**
  // Phase I: Privacy budget allocation
  Central Server:
  Broadcast θ to each client i.
  Client i:
  $\theta_i \leftarrow \theta$,
  $s_i \leftarrow \frac{1}{d_i} \sum_{j=1}^{d_i} \ell(x_j, \theta_i, y_j)$,
  Send s_i to central server.
  Central Server:
  Solving (19) with the uploaded s_i to obtain ϵ_k^*,
  Broadcast ϵ_k^* to each client i.
  // Phase II: Global model updating
  Client i:
  for $(x_j, y_j) \in D_i$ **do**
    $\tilde{x}_i \leftarrow x_j + w_{i,k}$,
    $\tilde{y}_i \leftarrow y_j$,
  for $t = 1, 2, \ldots, T$ **do**
    $\theta_i \leftarrow \theta_i - \eta \cdot \frac{1}{d_i} \sum_{j=1}^{d_i} \nabla \ell(\tilde{x}_j, \theta_i, \tilde{y}_j)$,
  Send θ_i to central server.
  Central Server:
  $\theta \leftarrow \sum_{i=1}^{N} \frac{d_i}{D} \cdot \theta_i$
return θ, ϵ^*

Theoretical Analysis

- Privacy analysis
  - Privacy leakage of each training round

Theorem 1. *Let Assumption 2 and 3 hold and* $\epsilon_{i,k}, \delta_{i,k} > 0$ *for all* $i = 1, \ldots, N$ *and all* $k = 1, \ldots, K$. *In Algorithm 1 , if we have*

$$\sigma_{i,k}^2 = c\frac{G^2 T ln\left(1/\delta_{i,k}\right)}{d_i(d_i - 1)\sqrt{\mu}\epsilon_{i,k}^2}$$

then the local model $\theta_{i,k}$ *learned in round* k *satisfies* $\left(\epsilon_{i,k}, \delta_{i,k}\right)$*-local differential privacy for some constant* c.

Theoretical Analysis

- Privacy analysis
  - Privacy leakage of all training rounds

Theorem 2. *Let Assumption 2 and 3 hold and* $\epsilon_{i,k}, \delta_{i,k} > 0$ *for all* $i = 1, \dots, N$ *and all* $k = 1, \dots, K$. *With* $\sigma_{i,k}$ *satisfies Theorem 1, the Algorithm 1 guarantees* $\left(c_0\sqrt{K}\epsilon_{m,\tilde{k}}, \delta_{m,\bar{k}}\right)$ *differential privacy for some constant* c_0, *where* $\epsilon_{m,\tilde{k}} = max\left\{\epsilon_{i,k} \mid \forall i = 1, \dots, N, \forall k = 1, \dots, K\right\}$.

- Utility analysis

Theorem 3. *Let Assumption 2 to 4 hold and the value of the gradient is upper bounded with B (i.e.,* $\nabla \ell(\theta) \leq B$ *), with* $\sigma_{i,k}$ *satisfies Equation (20), we have*

$$\mathbb{E}[\ell(\theta^K)] - \ell^* \leq \frac{LB^2}{\mu^2 K}(1 + p\sigma^2)$$

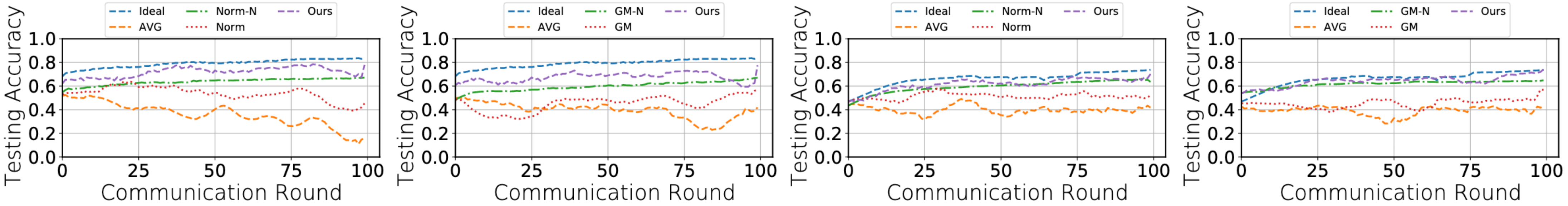
where $\sigma = max\left\{\sigma_{i,k} \mid \forall i = 1, \dots, N, \forall k = 1, \dots, K\right\}$, *p is the dimension of each training data sample.*

Evaluation

- Setup
  - Training dataset and model
    - Adult dataset with the logistics regression model and loan dataset with the multi layer perception model
  - Attack model
    - Backdoor attack
    - Label flipping attack
    - Membership inference attack
  - Benchmarks
    - FedAVG (baseline)
    - Norm bounding (Norm) and adding Gaussian noise (Weak-DP)
    - GeoMed (GM), Trimmed Mean (TM)
    - LDP and CDP

Evaluation

- Results
  - Robustness
    - RPFL (ours) improves the accuracy of the learned model by 3.39 times compared to the baseline FedAvg, and 0.76 times compared to the benchmark norm bounding.



(a) Backdoor attack (LR) (b) Label flipping attack (LR) (c) Backdoor attack (MLP) (d) Label flipping attack (MLP)

Fig. 2: The testing accuracy of the learned global model when defending against attacks with different defense methods.

Evaluation

- Results
  - Robustness
    - RPFL (ours) can still achieve relatively high testing accuracy even when the number of attackers is increased to 0.4.
    - RPFL performs the best compared to the benchmarks, the testing accuracy of the learned model is improved up to 2.75 times compared to benchmarks

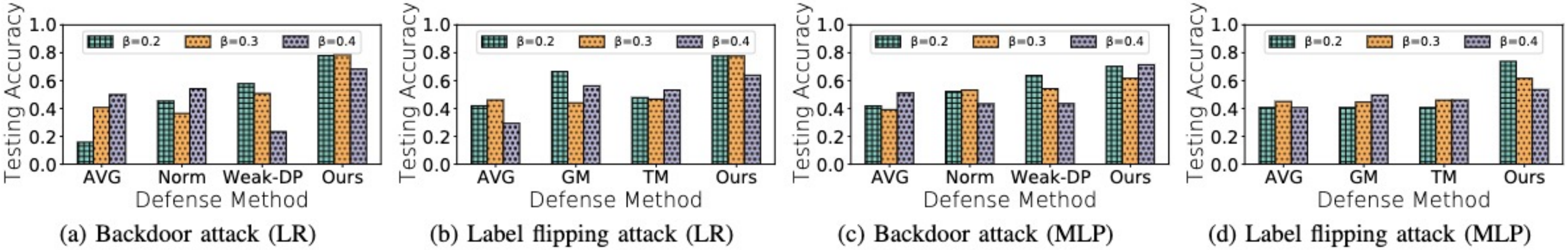


Fig. 3: Top-1 accuracy of different defense methods when defending against attacks with various fraction of malicious clients (from 0.2 to 0.4).

Evaluation

- Results
  - Privacy
    - RPFL (ours) achieves almost the best privacy guarantee compared to the benchmarks, with an attack accuracy decreased up to 0.71 times
    - RPFL (ours) shows better privacy and utility trade-off than other DP-related privacy-preserving methods.

TABLE I: Performance of different private federated learning methods (dataset: Loan)

Methods / Accuracy	CDP	LDP	Ours	AVG
Attack	64%	46%	48%	82%
Classification	73%	58%	80%	86%

Conclusion

- DRPri: Distributionally robust and private FL problem
  - Leverage DRO to jointly consider robustness and privacy problem in federated learning
  - Design an algorithm RPFL to solve the formulated problem with high robustness and privacy guarantee

- Experiment results show RPFL outperforms other defense methods in robustness with privacy guarantee.

Thank you!

Q&A

Email: si-ping.shi@connect.polyu.hk