VIRGINIA TECH

CNSR

ACSAC 2022

December 5-9, 2022 · Austin, Texas, USA

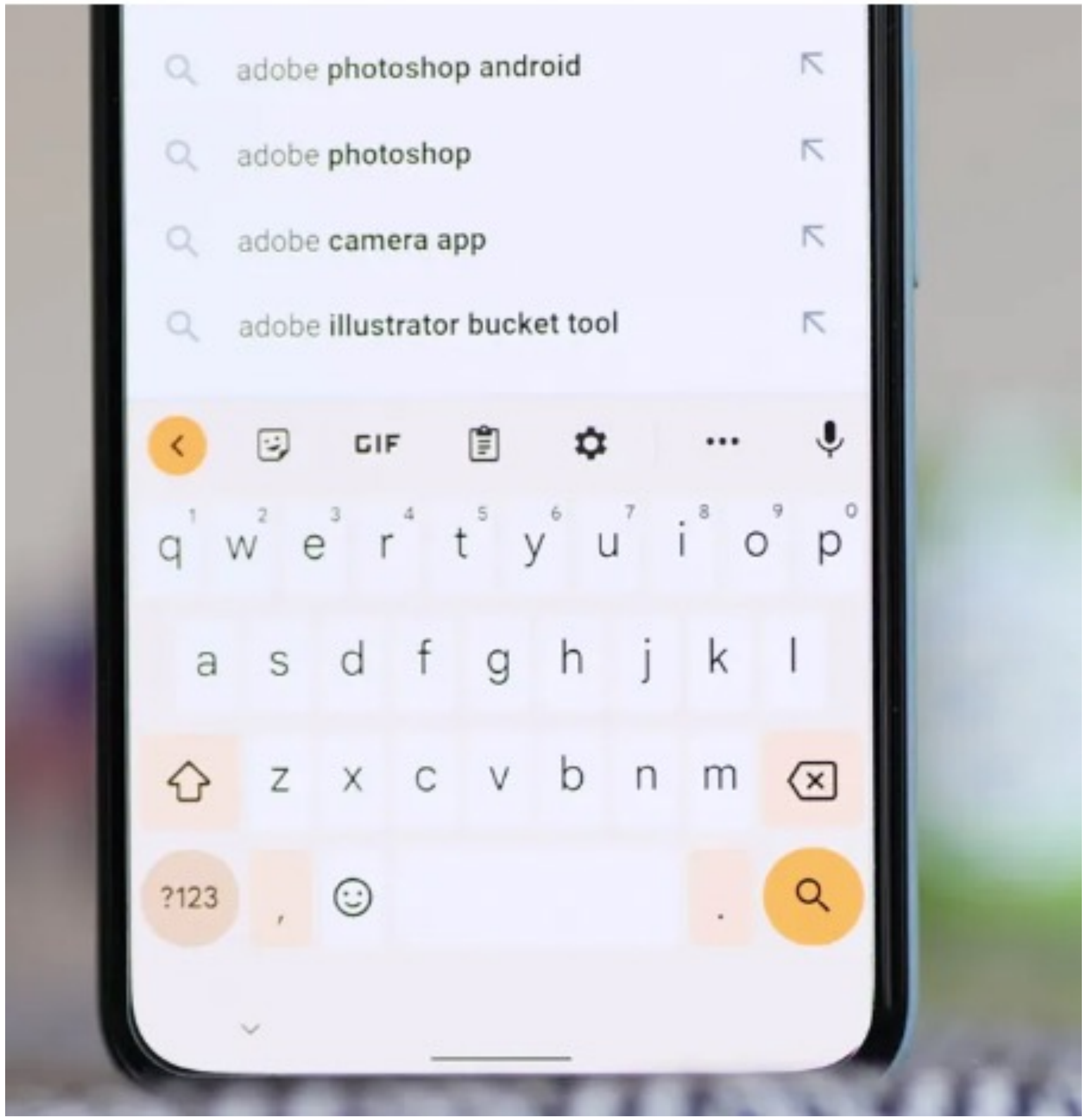
Squeezing More Utility via Adaptive Clipping on Differentially Private Gradients in Federated Meta-Learning

Authors: **Ning Wang**′, Yang Xiao†, Yimin Chen‡, Ning Zhang⋆, Wenjing Lou′, and Y. Thomas Hou′

Virginia Tech ′
University of Kentucky †
University of Massachusetts Lowell ‡
Washington University in St. Louis ⋆

Our data is used by AI applications!

- Next word suggestion of Gboard.



Our data is used by AI applications!

Gboard only uses federated learning while your phone charges, is connected to Wi-Fi, and isn't in u

https://s

Learn

Support

federated learning

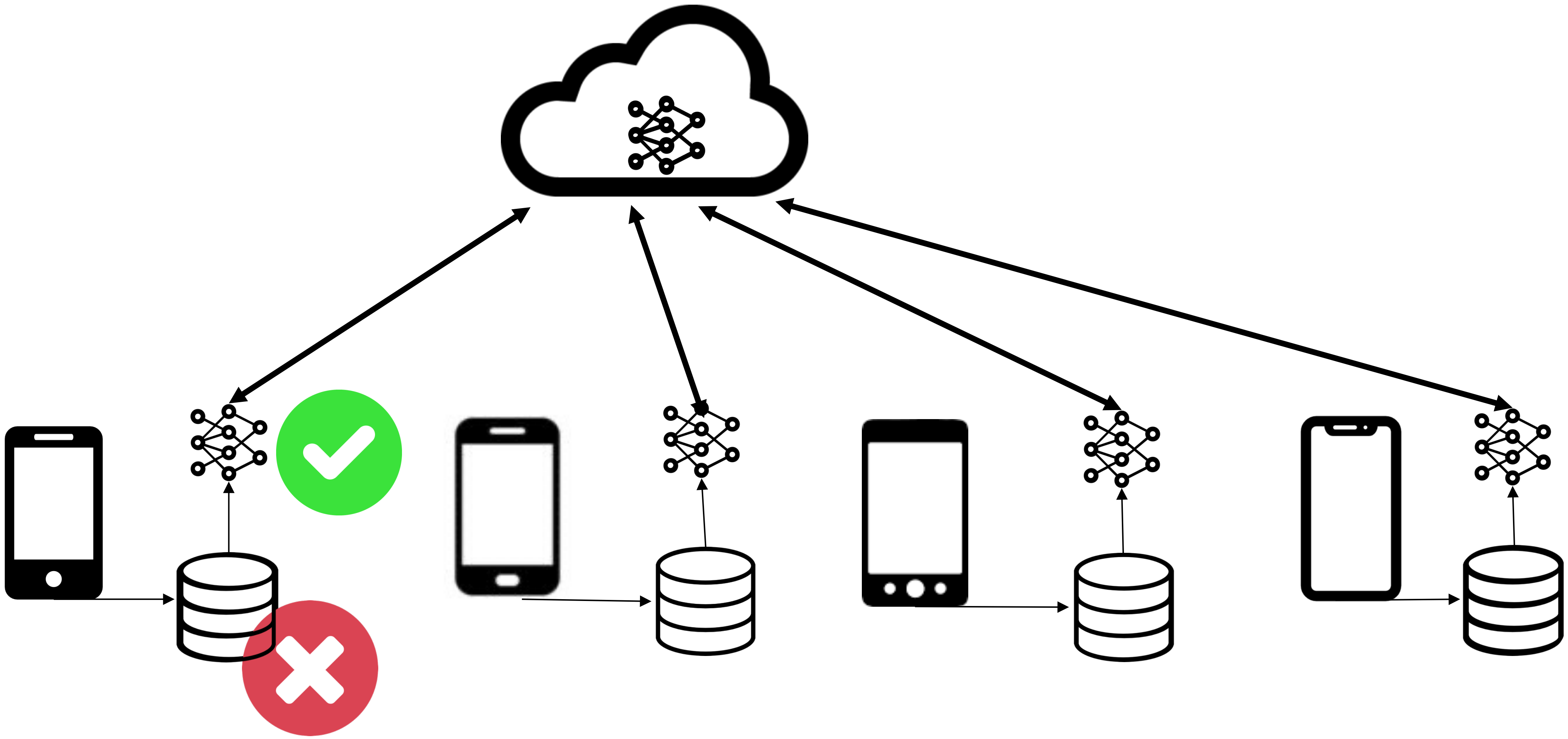
Federated learning (also known as collaborative learning) is a machine learning technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them.

https://en.wikipedia.org › wiki › Federated_learning

Federated learning - Wikipedia

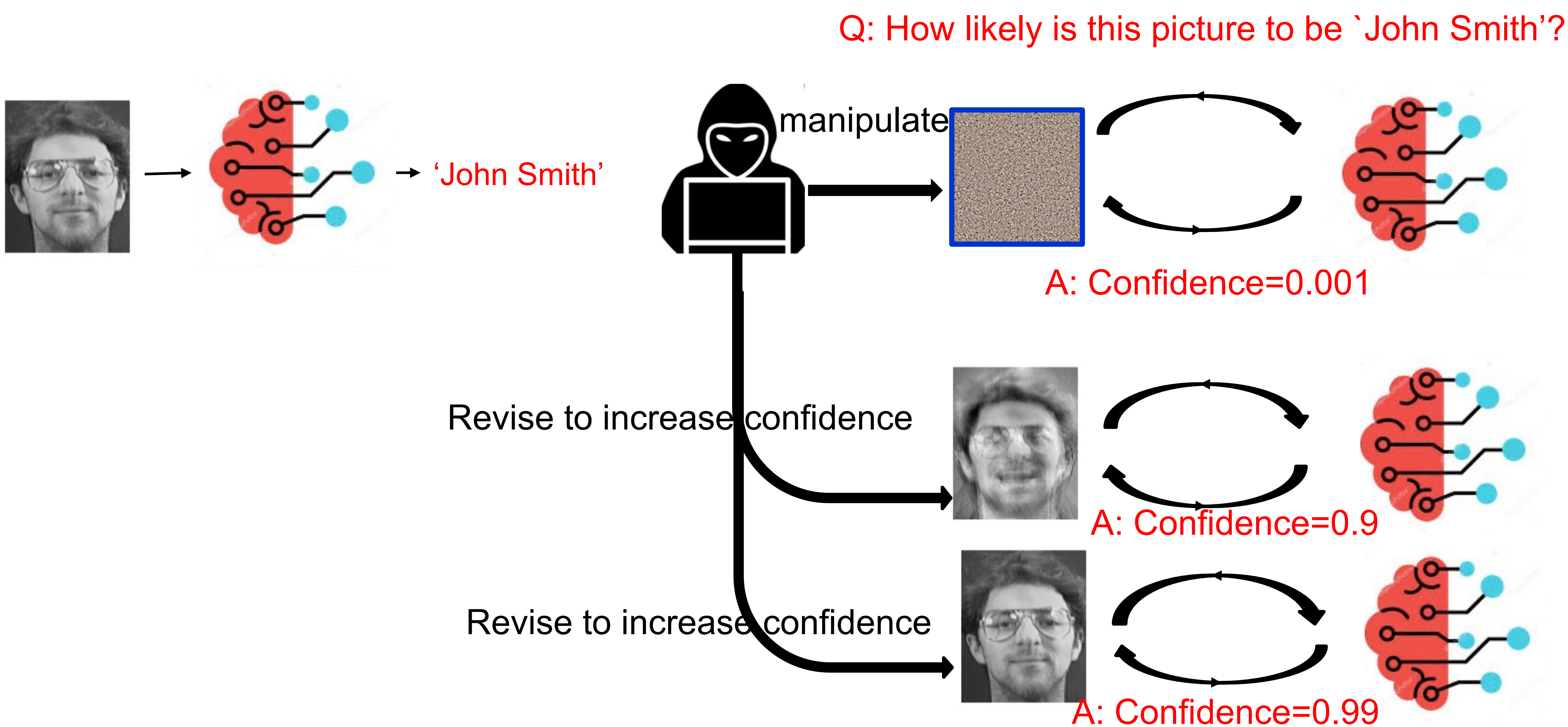
VIRGINIA TECH

What's Federated Learning?



Is Data Privacy fully Protected by FL?

- The model parameters are open to the server directly and to other clients indirectly.
- Can attacker infer data from model?



Training Data is Memorized by the Model

Privacy Attack: Inference Attack

- FL cannot guarantee the training data privacy.
- State-of-the-art Inference Attack
  - Membership inference attack [1,2]
  - Model Inversion Attack [3]
  - Attribute Inference Attack [4]

[1] Milad Nasr et al. 2019. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In 2019 IEEE Symposium on Security and Privacy (SP 19). IEEE, 739–753.
[2] Jingwen Zhang, Jiale Zhang, Junjun Chen, and Shui Yu. 2020. Gan enhanced membership inference: A passive local attack in federated learning. In ICC 20202020 IEEE International Conference on Communications (ICC). IEEE, 1–6.
[3] Zhibo Wang, Mengkai Song, Zhifei Zhang, Yang Song, Qian Wang, and Hairong Qi. 2019. Beyond inferring class representatives: User-level privacy leakage from federated learning. In IEEE Conf. on Computer Communications (INFOCOM). IEEE, 2512–2520.
[4] Rui Wang, Yong Fuga Li, XiaoFeng Wang, Haixu Tang, and Xiaoyong Zhou. 2009. Learning your identity and disease from research papers: information leaks in genome wide association study. In Proceedings of the 16th ACM conference on Computer and communications security (CCS). ACM, 534–544.

Problem and Goal of this Paper

- Problem
  - FL can protect data privacy to some extend.
  - Attackers are still capable to infer training data while knowing the model parameters.
  - Differential Privacy (DP) is a tool for privacy protection, but it harms the accuracy a lot.
  - Mobile used data size can be small.
- Goal
  - Provide rigorous privacy guarantee for users by incorporate DP.
  - Maintain a good trade-off between privacy and accuracy.

DP preliminary: Inference Attack on Databases

- What is the inference attack in a database?
  - Use the statistical/aggregate queries that are authorized to gain information that are not authorized.
- Example: Exam score database.
  - Tuple: (student_id, score)
  - Average score on an exam is a query everyone is allowed to run.
  - Attacker wants to find the exact score of some student.

Inference attack sometimes requires some additional external information.
e.g., Attacker knows Alice took the exam late.

↓

Attacker get
average score **before** some date & average score **after** such date.

↓

It is easy to get the score of Alice.

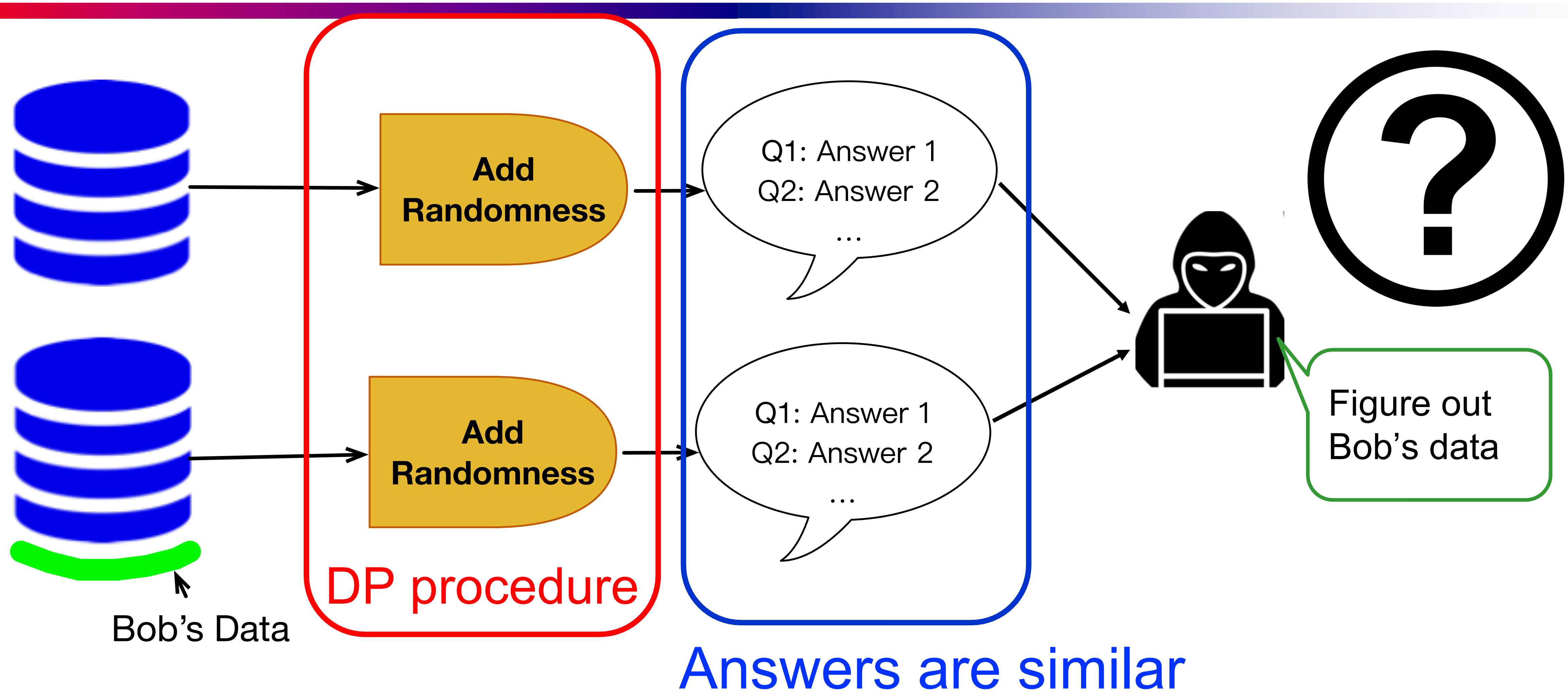
DP preliminary: Adjacent Databases

- What's Adjacent database?
  - Two databases only have one record difference.
  - E.g., a database with Alice data in, a database without Alice Data.
- Once attackers have access to the adjacent database, it can launch inference attack.

How to solve the inference attack?

VIRGINIA TECH.

Differential Privacy (DP)

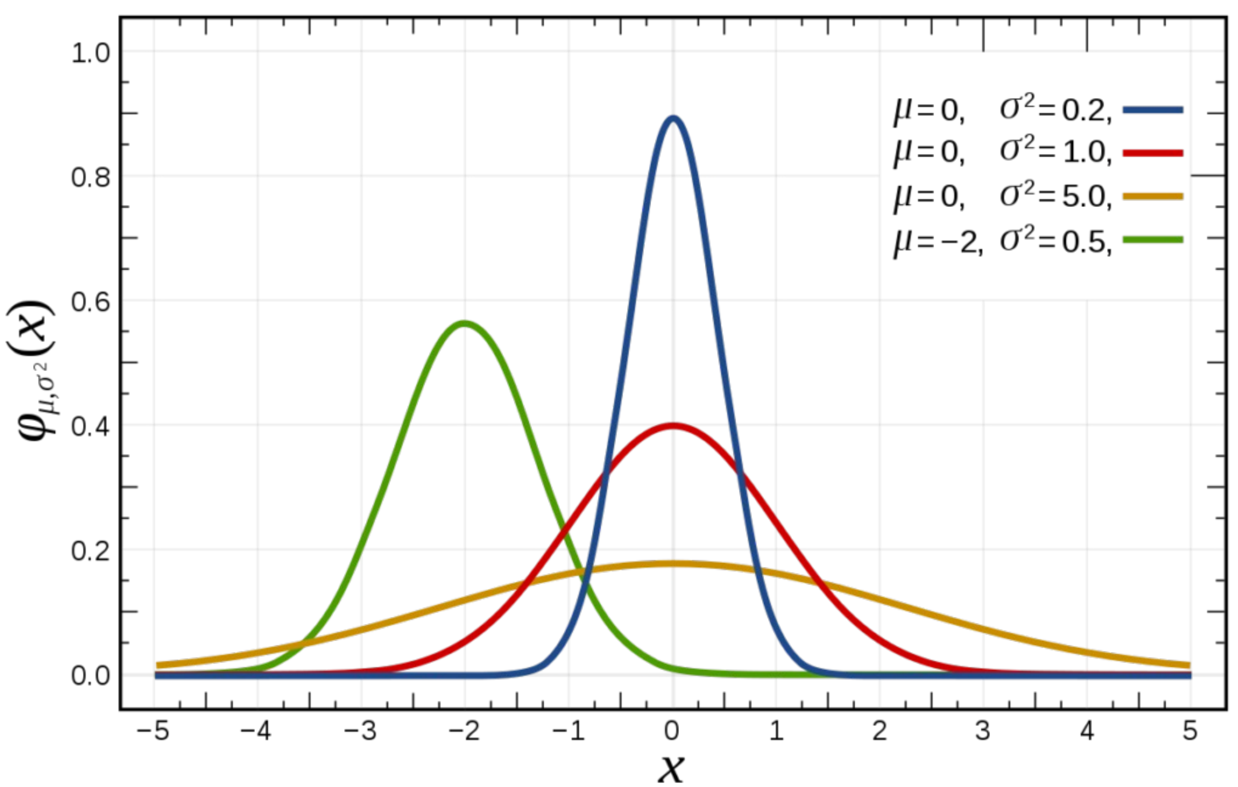


- DP allows to learn useful information of the population without leaking individual privacy.

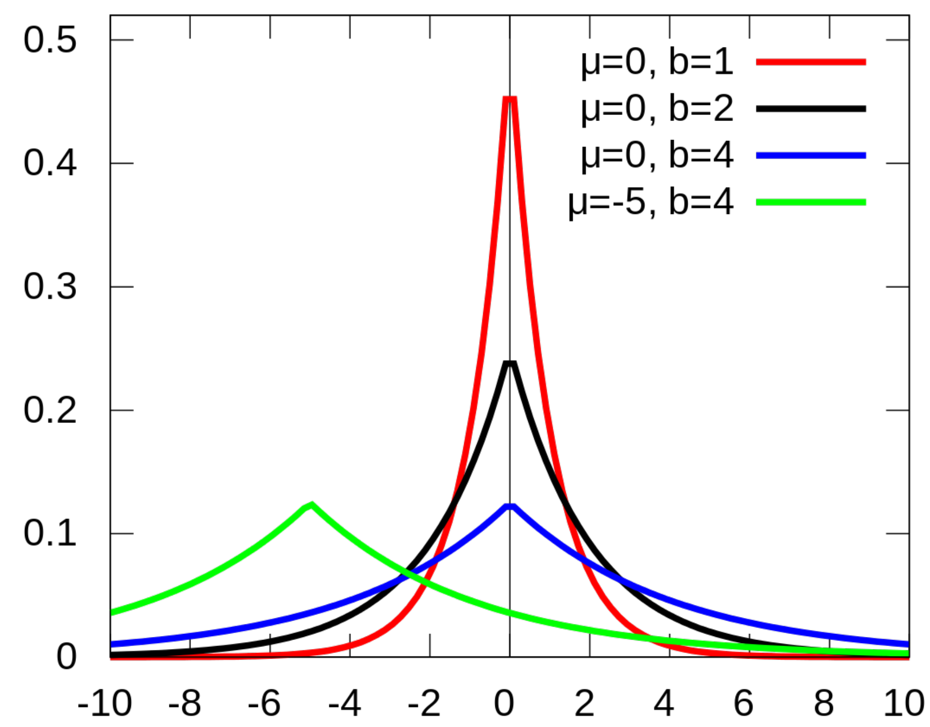
$(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$**−differential privacy**: A random mechanism $\mathcal{M}: \mathcal{D} \rightarrow \mathcal{R}$ satisfies (ε, δ)−differential privacy if for any two adjacent inputs ($d, d' \in \mathcal{D}$) and for any subset of outputs $S \in \mathcal{R}$ it holds that: $\Pr[\mathcal{M}(d) = S] \leq e^{\varepsilon} \Pr[\mathcal{M}(d') = S] + \delta$

Key Step of DP

- Adding Randomness
  - Gaussian Noise
  - Laplace Noise



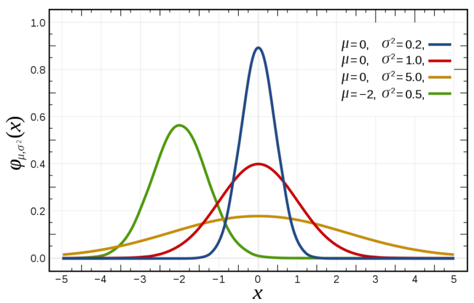
Gaussian Noise

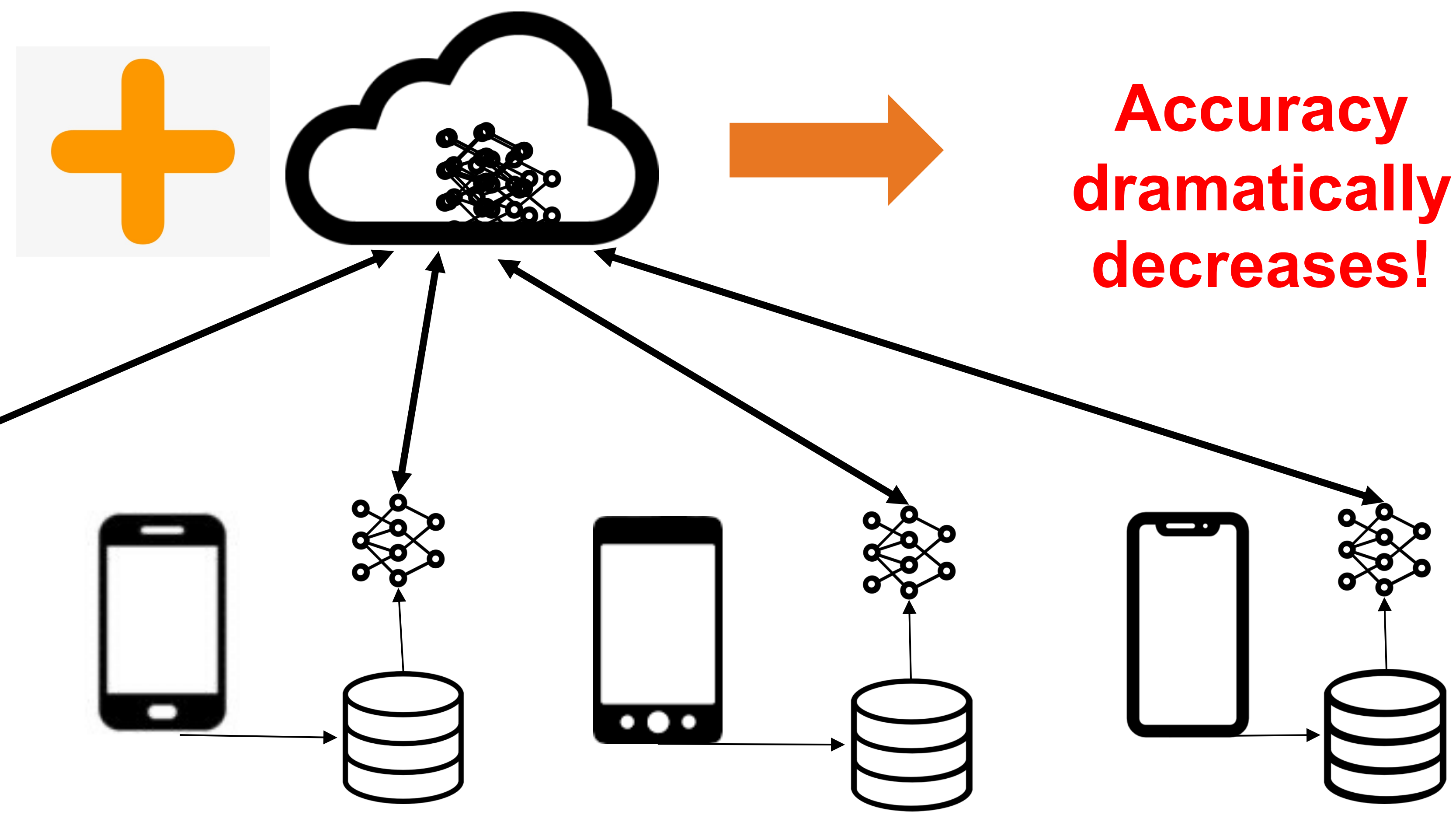


Laplace Noise

Differentially private federated Learning

- Problem: low accuracy





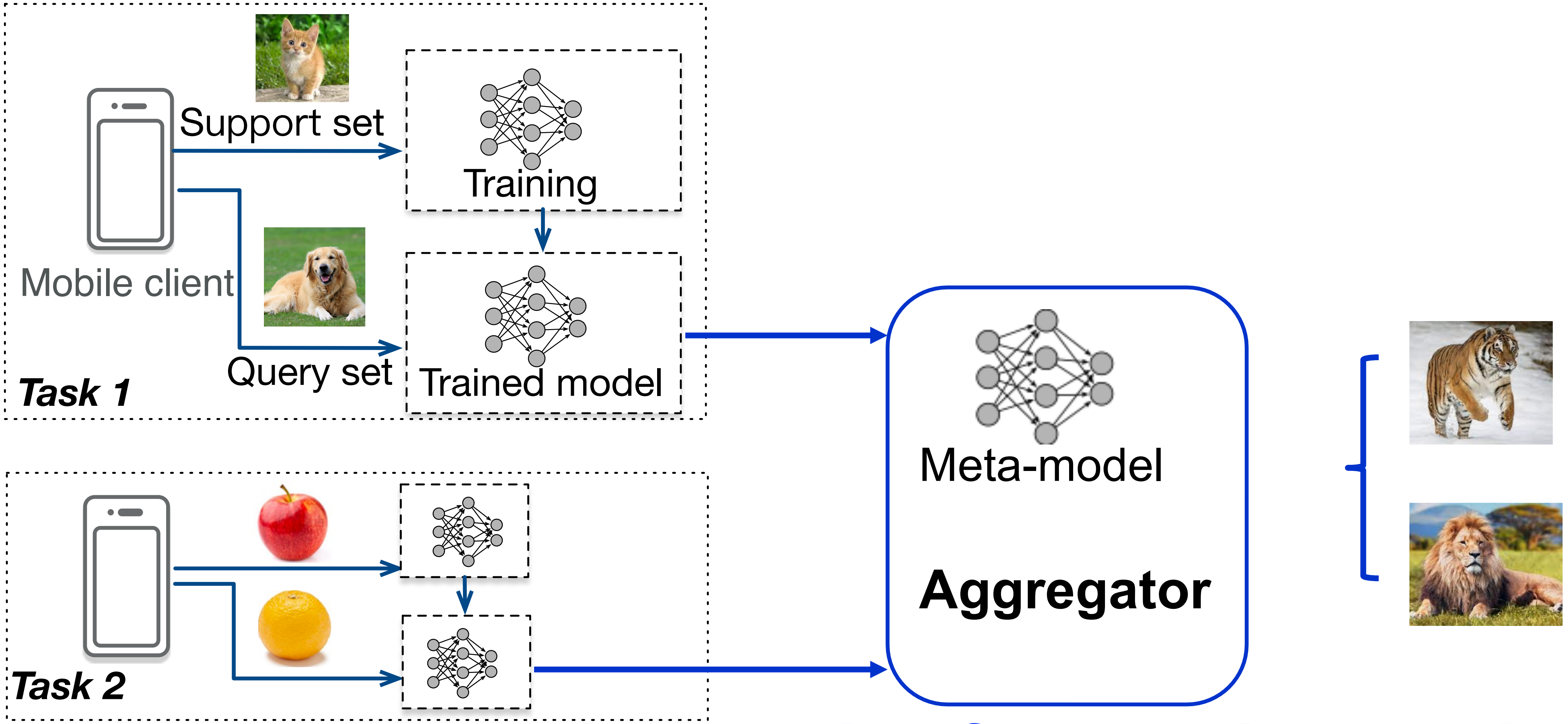
Accuracy dramatically decreases!

Continual Training does help.

But it requires Enough Data & Training Power/Time

To cope with small size of local data

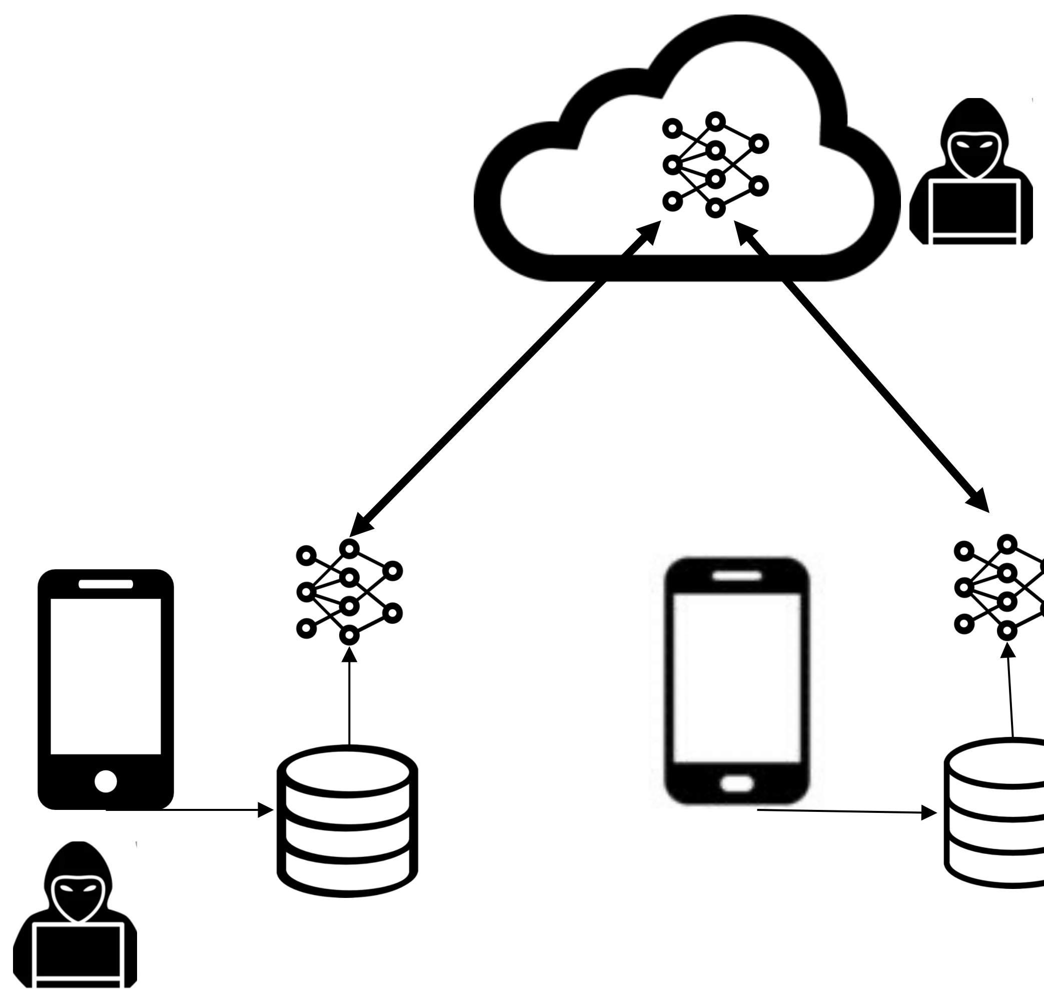
- Federated Learning → **Federated Meta-Learning**
  - Deal with few-shot problem.
  - Fast adaptation/customization.



Learn common knowledge from various tasks to enable quick learning for new/unseen tasks.

Two Threat Models

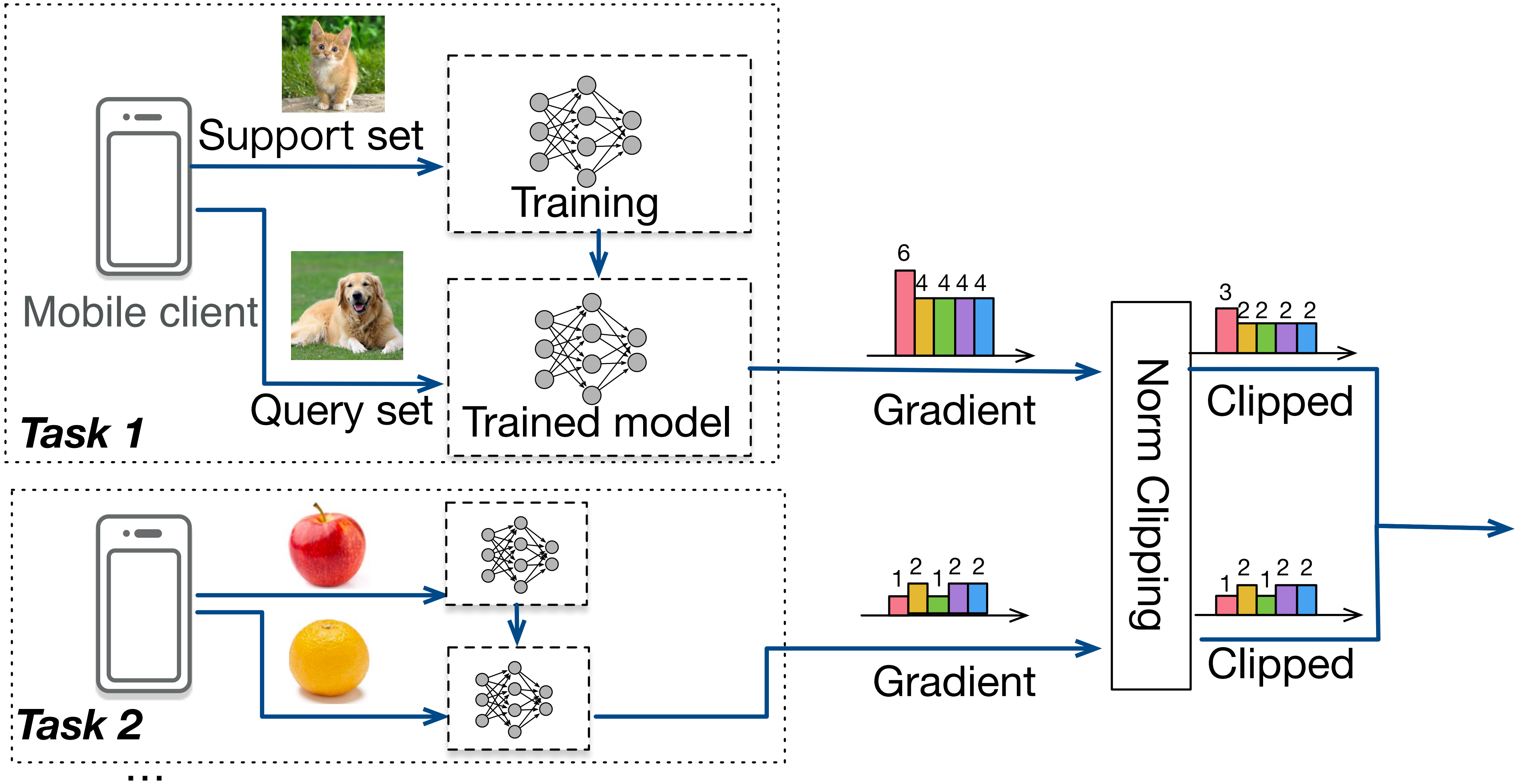
- Central server is trusted, clients are honest-but-curious.
- Both central server and clients are honest-but-curious.



VIRGINIA TECH

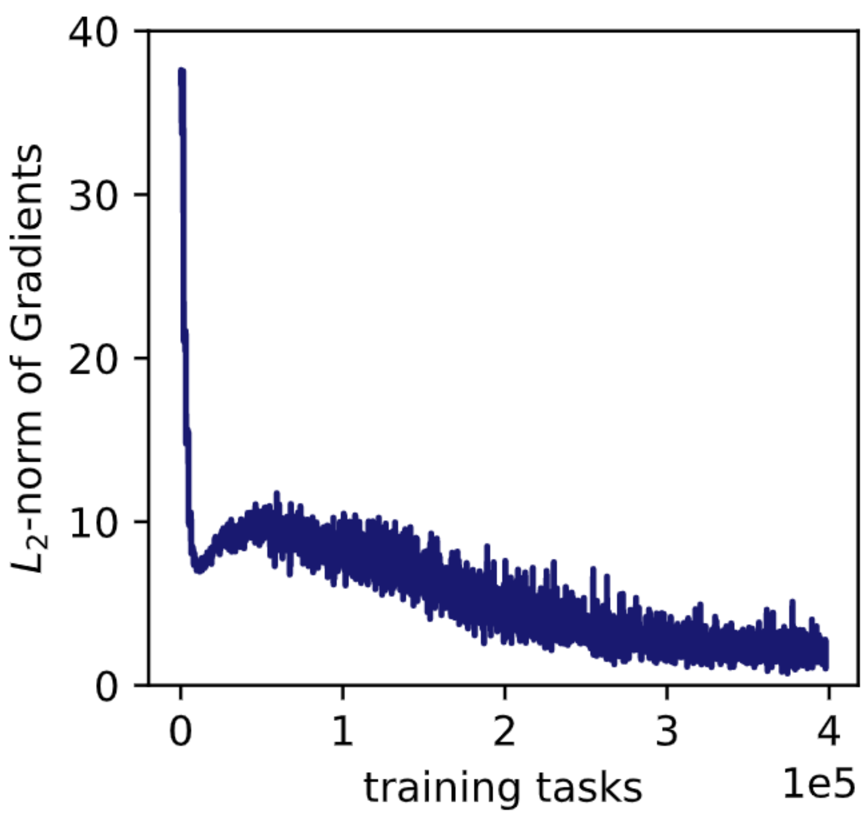
DP in Federated Meta-learning

- Adding noise
  - The noise should be proportional to the largest gradient.
  - To avoid too large noise, we should clip the gradient.



Our Proposal

- **Adaptive** Clipping
  - Naïve constant clipping maintain a fixed clipping threshold C. The noise will be: $k * C$.
  - Adaptive clipping: change the threshold C adaptively.
  - Why adaptive clipping better?



The gradients will decrease during the course of training.

We can change the threshold C according to the gradient change.

Adaptive Clipping

- We ***cannot*** use the true gradient scale to adaptively decide C.
- Our proposal: determine C by using historical ***DP*** gradients in a window of size W:

$$C_{W+1} = f([\tilde{g}_1, \dots, \tilde{g}_W], k)$$

Algorithm 2: Adaptive Cilpping

Input: Clients set $\mathcal{T}$, noise multiplier z, user sample size L, Learning round T, Window size W

Output: Clipping Threshold C

```
Initialize C = C_0;
while and t <= T do
    Randomly Sample L clients T_s ← sample(T, L);
    if t > W then
        C ← f([g̃_{t−W}, ..., g̃_{t−1}], k)
    for i ∈ T_s do
        g_i ← gradient provided by client i;
        Clip gradient: ĝ_i ← g_i * min(1, C/||g_i||) ;
    g̃_t ← (1/L) (Σ_i ĝ_i + N(0, z^2 C^2 I));
```

$C \leftarrow f([\tilde{g}_{t-W}, \dots, \tilde{g}_{t-1}], k)$ — The DP version Gradients

Clip gradient: $\hat{g}_i \leftarrow g_i * \min(1, \frac{C}{\|g_i\|})$; $\quad \tilde{g}_t \leftarrow \frac{1}{L}\left(\sum_i \hat{g}_i + \mathcal{N}(0, z^2C^2\mathbf{I})\right)$;

VIRGINIA TECH.

Will Adaptive Clipping leak any more privacy?

- NO
- Because of the ***Post processing Property of DP***
  - If $F(X)$ satisfies ϵ-differential privacy
  - Then for any (deterministic or randomized) function g, $g(F(X))$ satisfies ϵ-differential privacy

Algorithm 2: Adaptive Cilpping

Input: Clients set $\mathcal{T}$, noise multiplier z, user sample size L, Learning round T, Window size W

Output: Clipping Threshold C

Initialize $\underline{C = C_0}$;

while *and* $t <= T$ **do**

  Randomly Sample L clients $\mathcal{T}_s \leftarrow \text{sample}(\mathcal{T}, L)$;

  if $t > W$ **then**

    $\underline{C \leftarrow f([\tilde{g}_{t-W}, ..., \tilde{g}_{t-1}], k)}$ ← The DP version Gradients

  for $i \in \mathcal{T}_s$ **do**

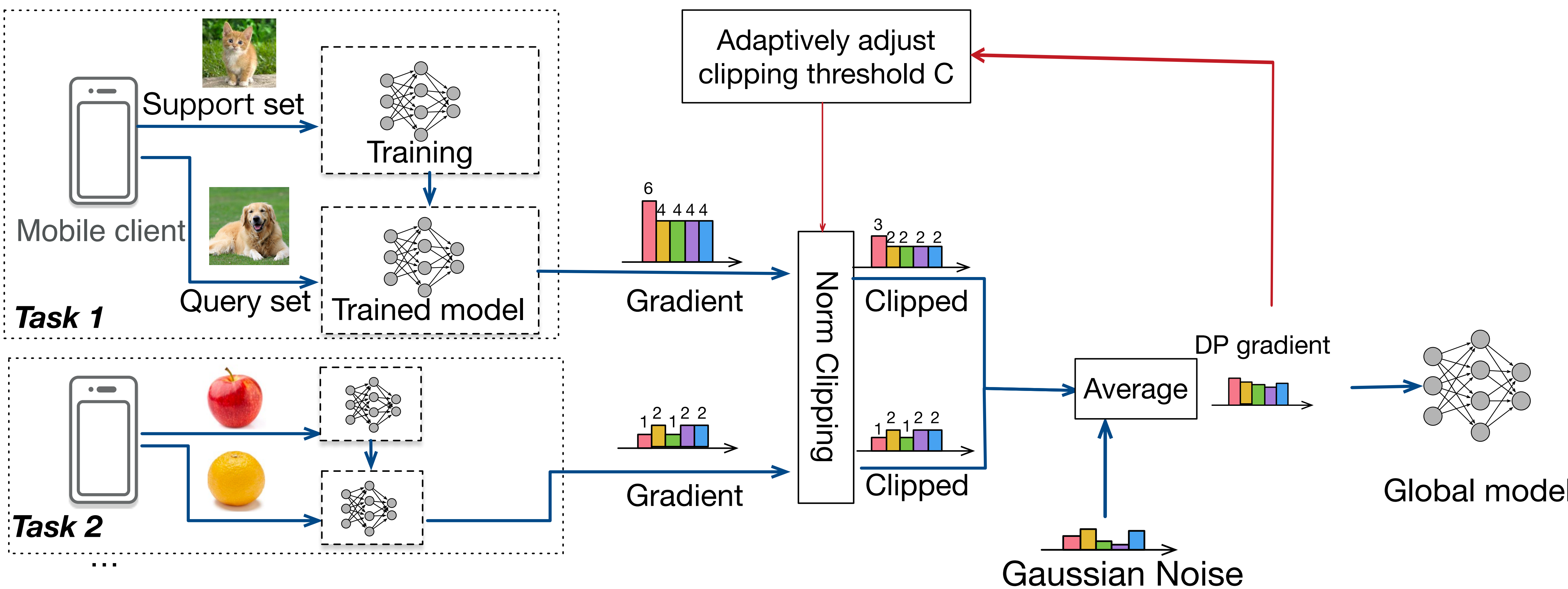
    $g_i \leftarrow$ gradient provided by client i;

    Clip gradient: $\hat{g}_i \leftarrow g_i * \min(1, \frac{C}{\|g_i\|})$;

  $\tilde{g}_t \leftarrow \frac{1}{L} (\sum_i \hat{g}_i + \mathcal{N}(0, z^2 C^2 \mathbf{I}))$;

Differentially private Meta-learning

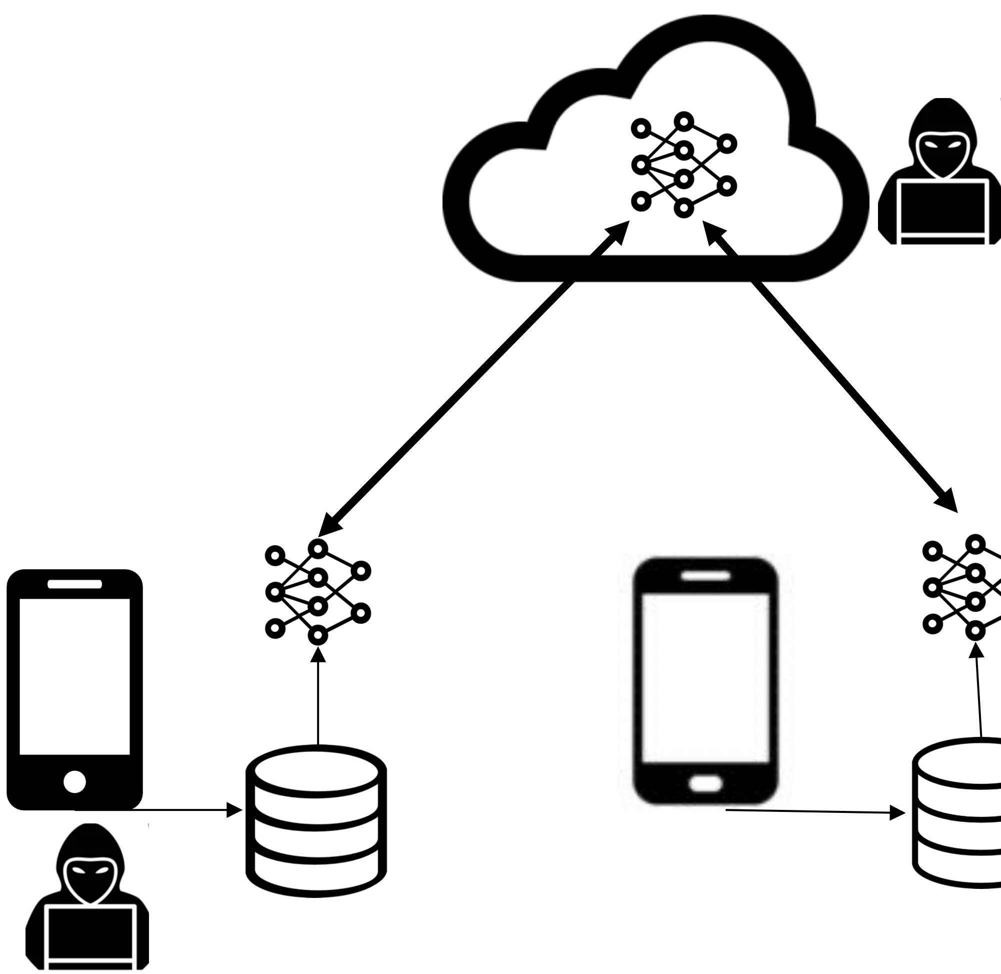
- The history of *Differentially Private* version gradients guides the current clipping.



VIRGINIA TECH.

Two Algorithms

- Two threat models
  - DP-AGR for threat model 1 where server is trusted, clients are honest-but-curious
  - DP-AGRLR for threat model 2 where the server is not trusted, and clients are honest-but-curious



Algorithm 3: DP-AGRLR (Client Side)

Input: Current global model Θ, local data $\mathcal{D}$, DP parameter (ϵ_0, δ_0), C_0, z_0

Output: gradient g

```
Function g = Base-Model-Train(Θ, D^s, D^q):
Initialize base-model: θ ← Θ;
Split local data D^s, D^q ← D;
z_0 ← compute_noise(ε_0, δ_0, *args)
for (x_i, y_i) ∈ D^s do
    record- level gradient: g_i ← ∇_θ L(θ, x_i) ;
    clip gradient: ĝ_i ← g_i * min(1, C_0/||g_i||) ;
g̃ ← 1/|D^s| (Σ_i ĝ_i + N(0, (z_0 C_0)^2 I));
update base-model: θ ← θ − η_1 g̃;
for (x_i, y_i) ∈ D^q do
    record-level gradient: g_i ← ∇_θ L(θ, x_i) ;
    clip gradient: ĝ_i ← g_i * min(1, C_0/||g_i||) ;
g ← 1/|D^q| (Σ_i ĝ_i + N(0, (z_0 C_0)^2 I)).
```

ing

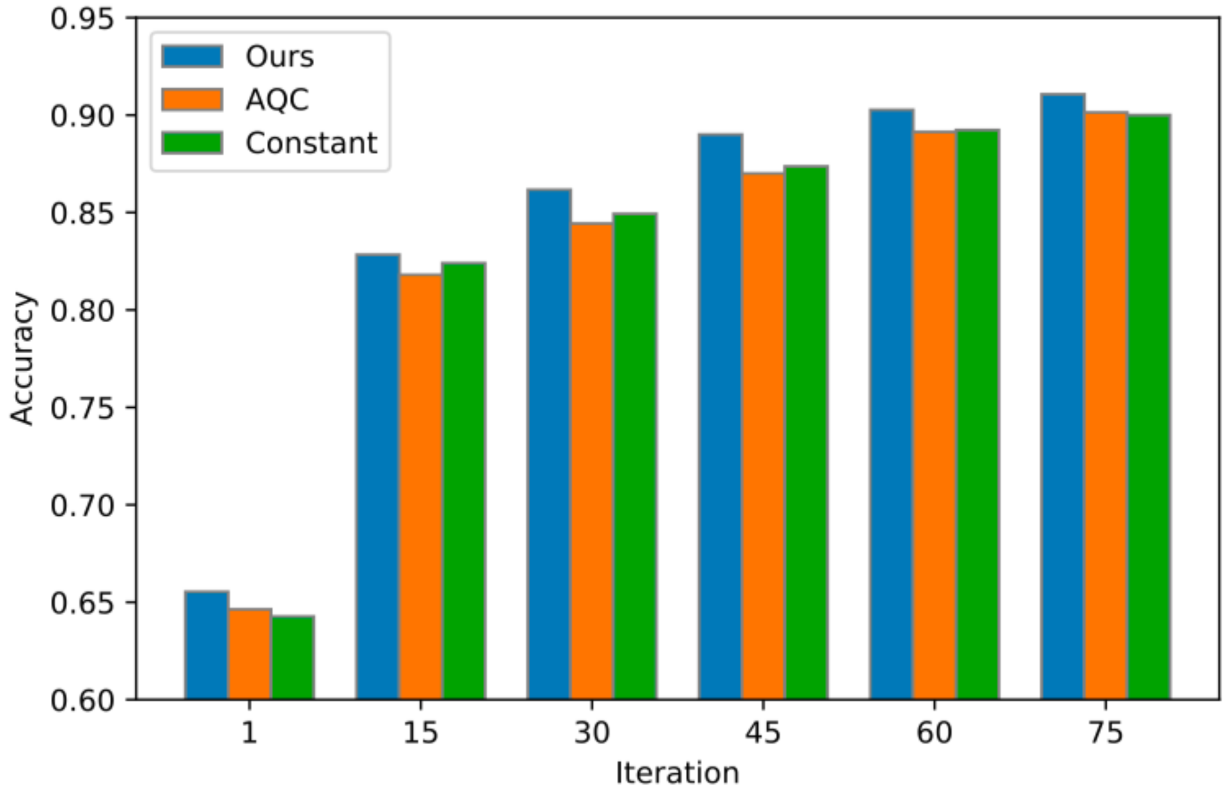
Experimental Setting

- Settings:
  - Image Datasets: Omniglot, CIFAR-FS, Mini-ImageNet
  - Client Number: 400,000
  - Clients in each learning round: 1500
  - Each client has 30 data record.
  - Meta-learning algorithm: MAML.

- Code:
  - Our code is available at https://github.com/ning-wang1/DPFedMeta.
  - Code Evaluated

Evaluation: Adaptive Clipping

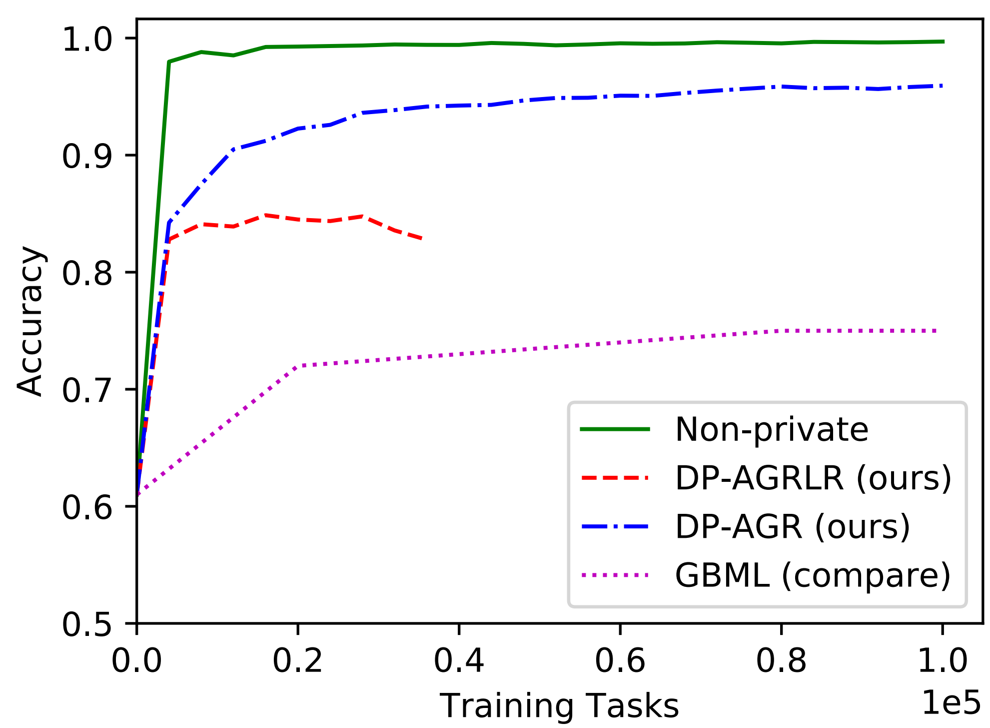
- **Ours** Vs **AQC** Vs **Constant**



All other settings are the same, only change the clipping method.

Evaluations: DP-AGRLR

- More accurate ML model with much lower privacy budget
  - DP-AGR (ours) achieves $(1.5, 10^{-6})$-DP;
  - DP-AGRLR (ours) achieves $(2.5, 10^{-5})$-DP for record-level privacy
  - Baseline achieves $(9.5, 10^{-3})$-DP



Model Accuracy of 5-way 5-shot learning in the Omniglot dataset.

Summary

- Differentially private federated meta-learning architecture.
- Design an adaptive gradient clipping method to conserve the privacy budget and improve accuracy.
- Provide two algorithms, DP-AGR and DP-AGRLR, to deal with different privacy requirements..

VIRGINIA TECH.

Thank You!
Q&A