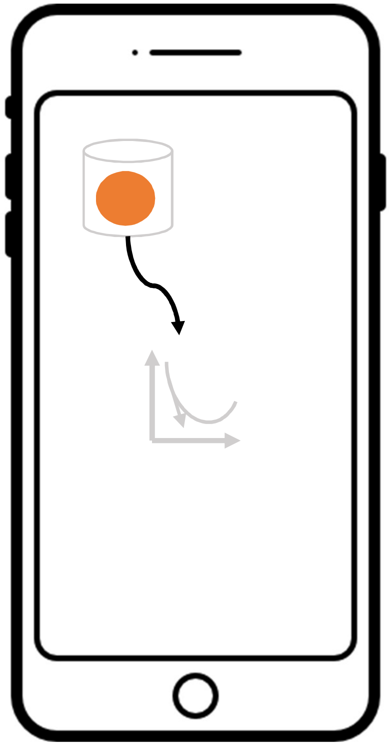
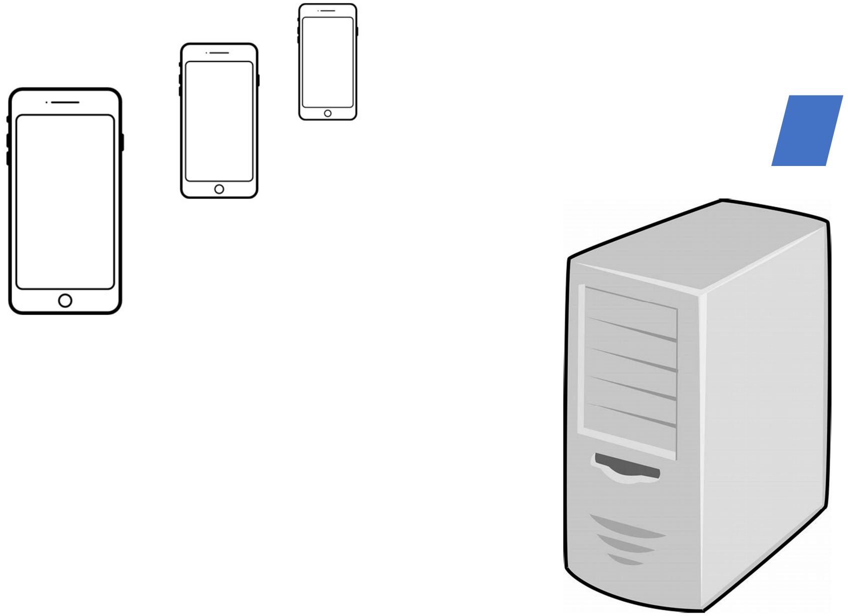
FedSel: Federated SGD under Local Differential Privacy with Top-k Dimension Selection

Ruixuan Liu[1], Yang Cao[2], Masatoshi Yoshikawa[2], Hong Chen[1]
[1]Renmin University of China, [2]Kyoto University
DASFAA, 2020

Federated Learning Overview

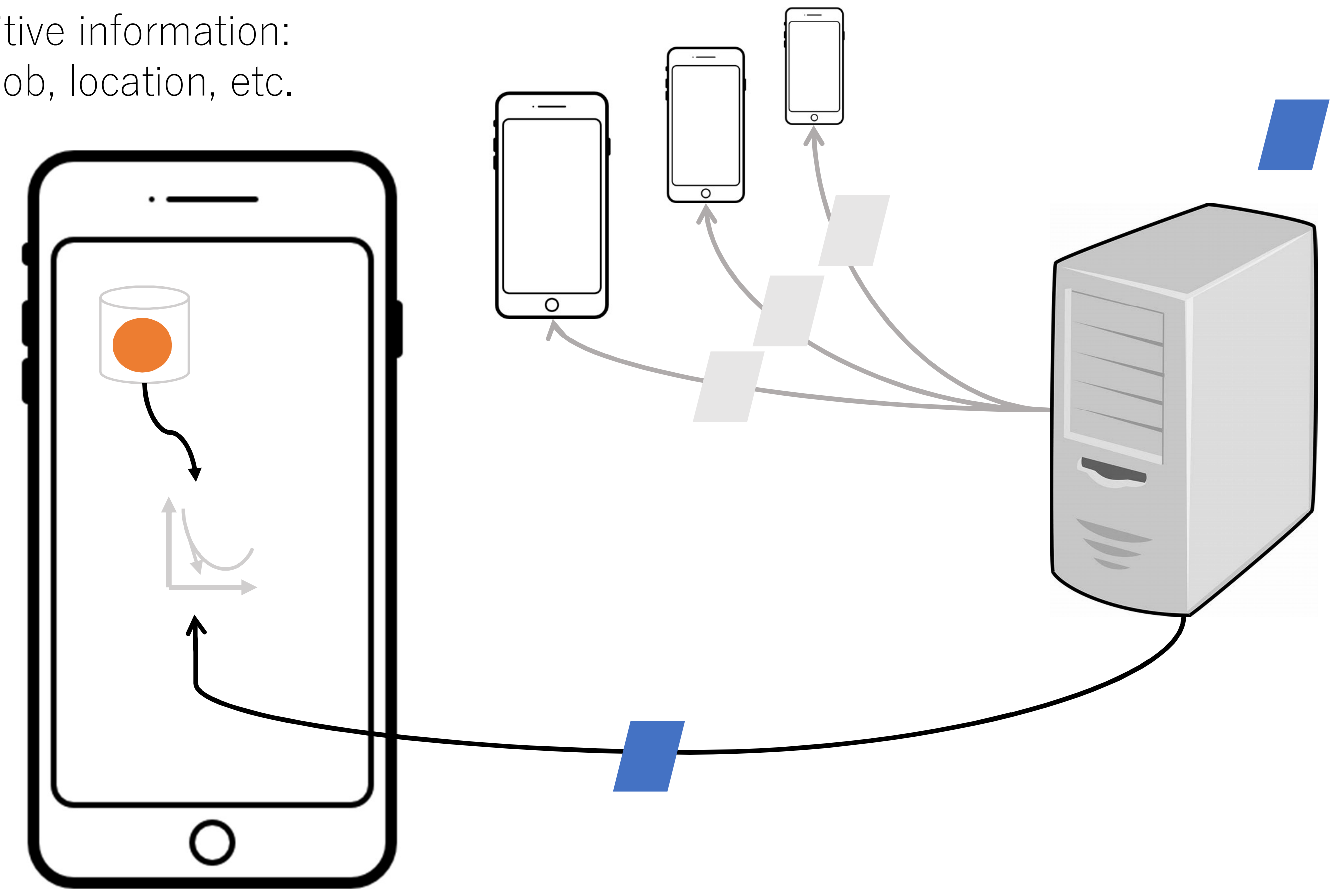
Sensitive information:
age, job, location, etc.





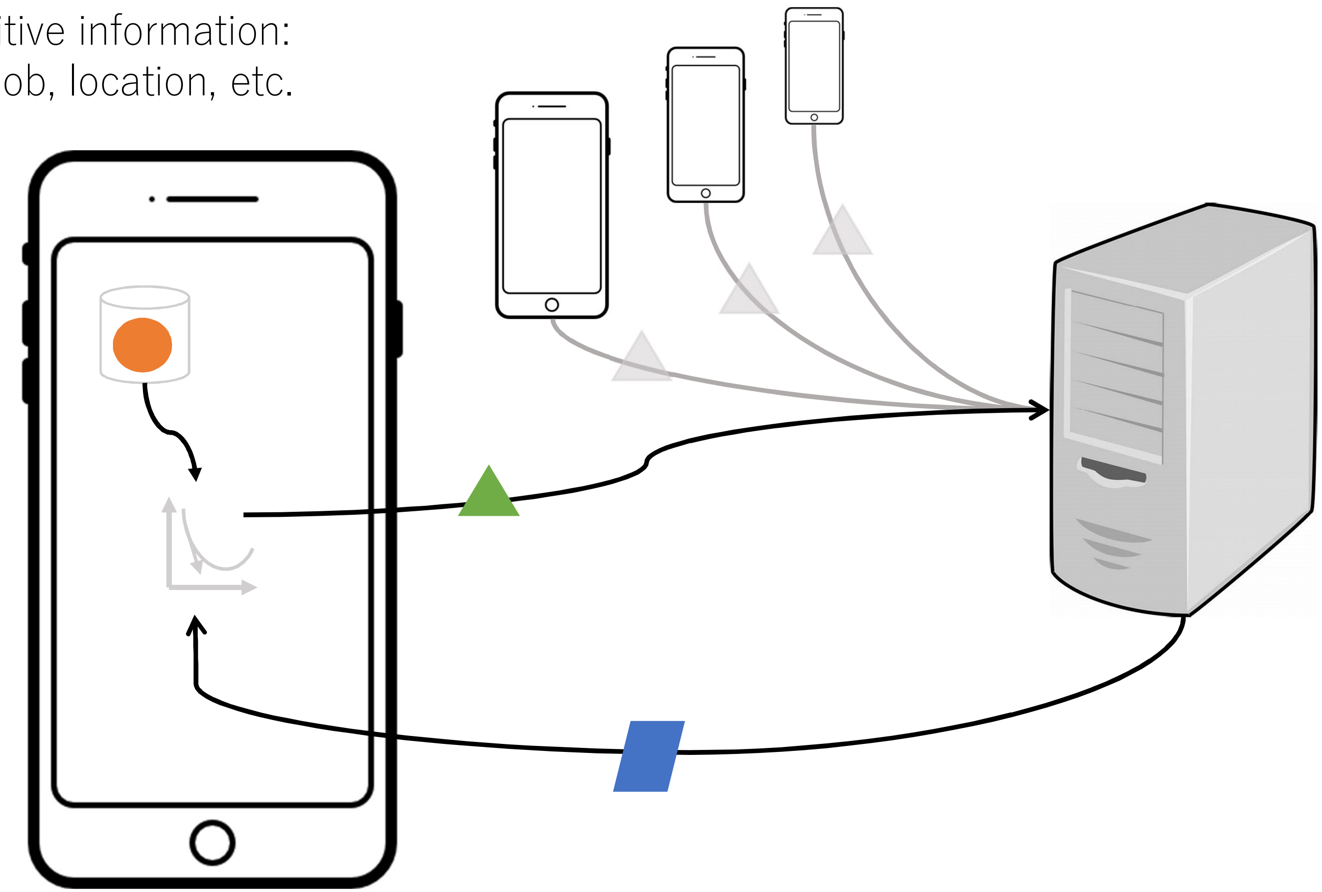
Federated Learning Overview

Sensitive information:
age, job, location, etc.

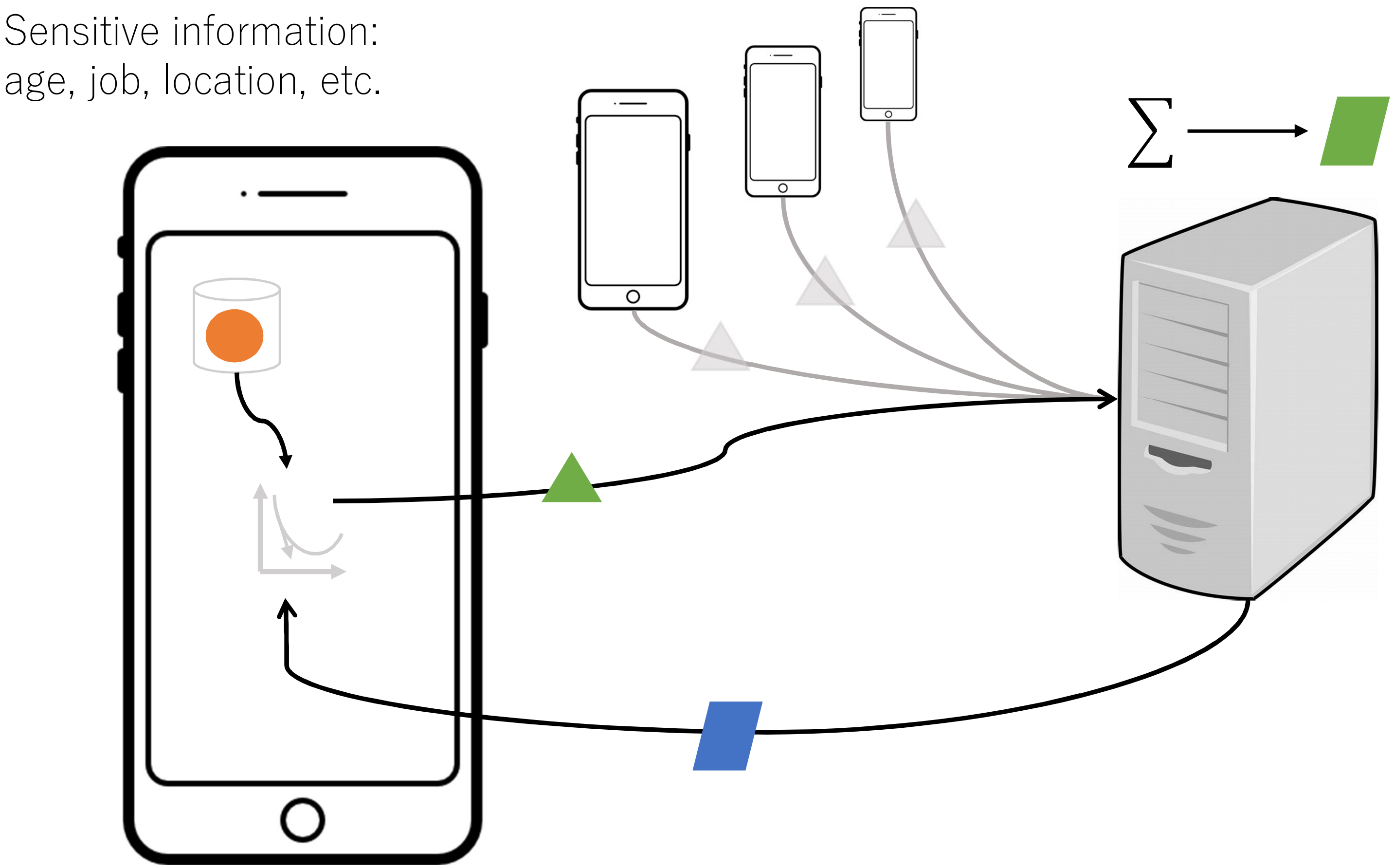


Federated Learning Overview

Sensitive information:
age, job, location, etc.

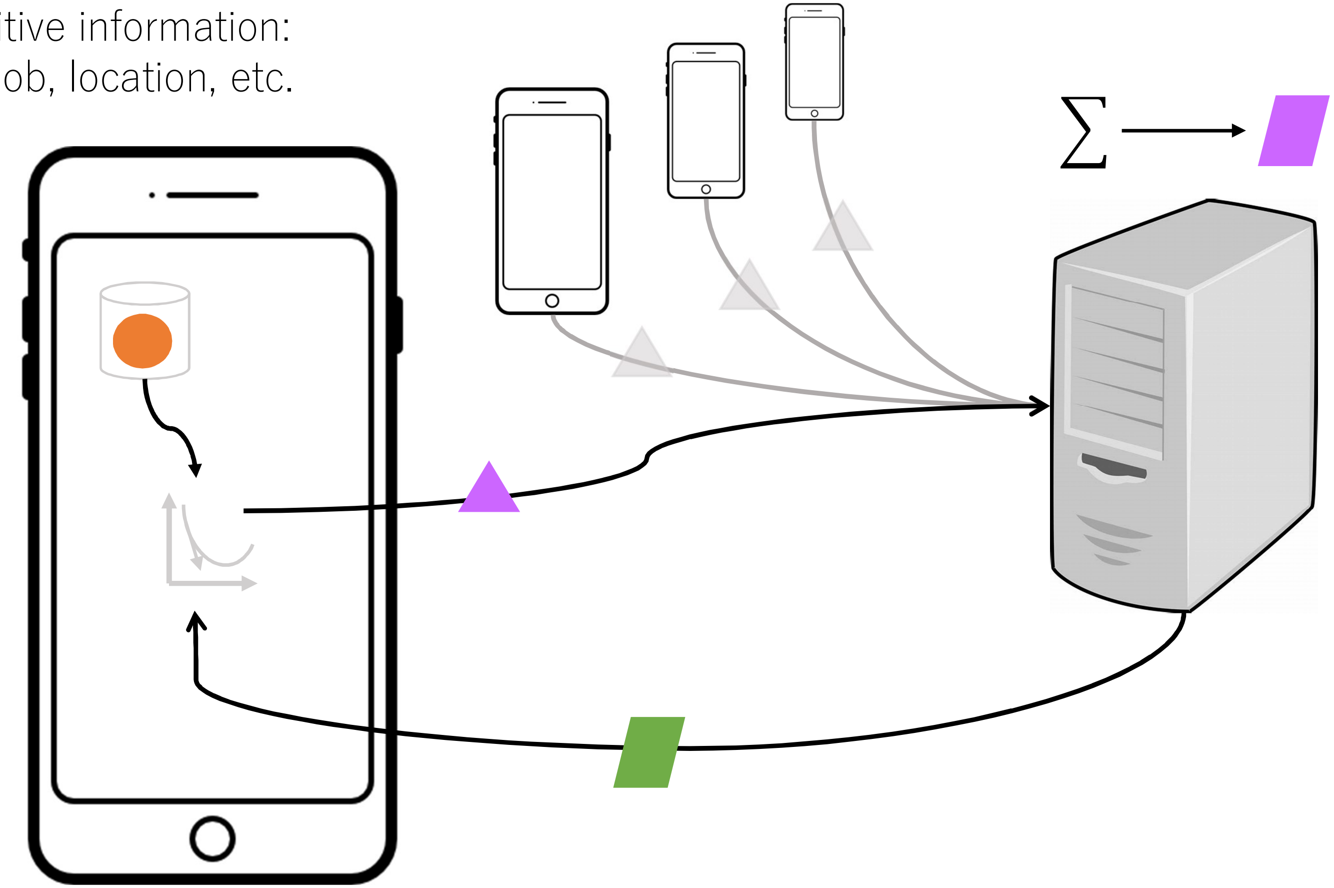


Federated Learning Overview

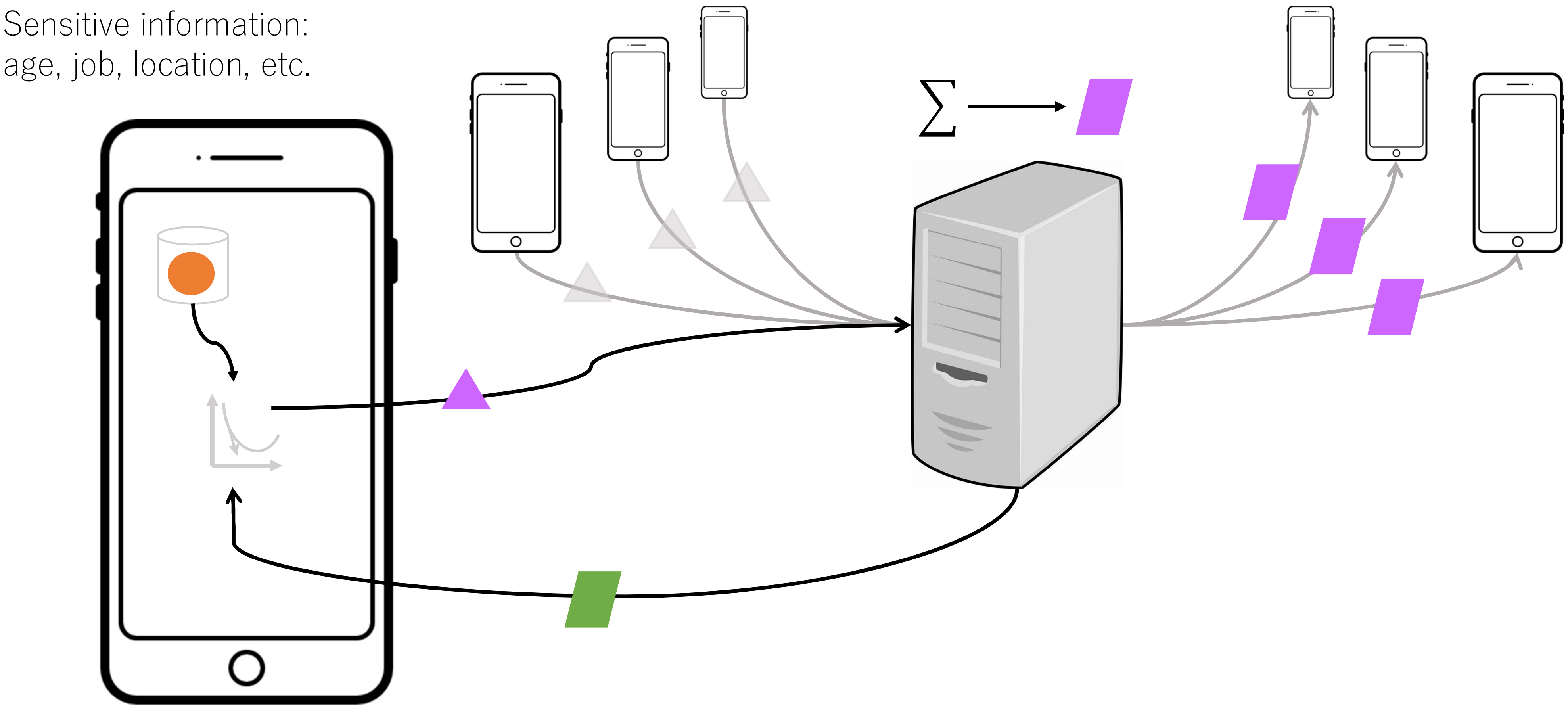


Federated Learning Overview

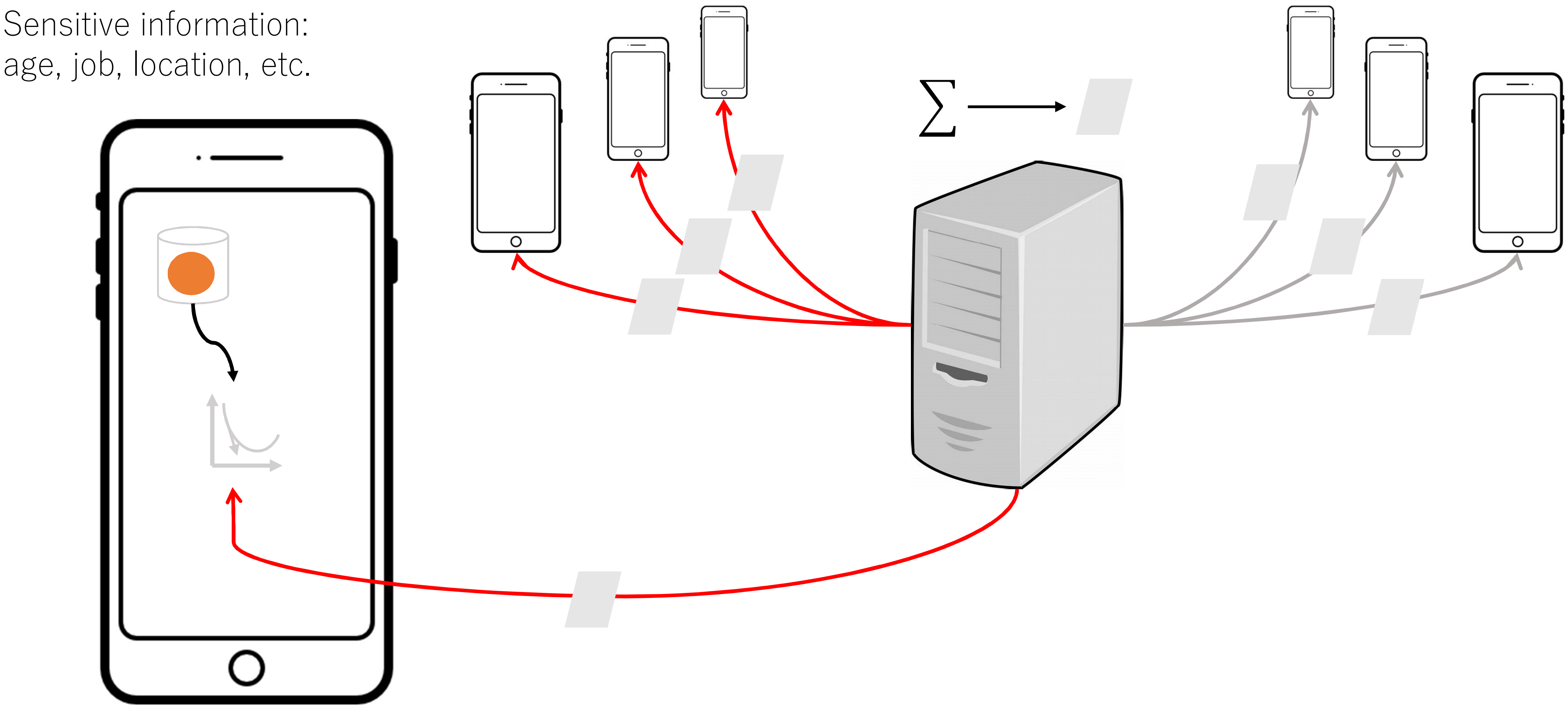
Sensitive information:
age, job, location, etc.



Federated Learning Overview

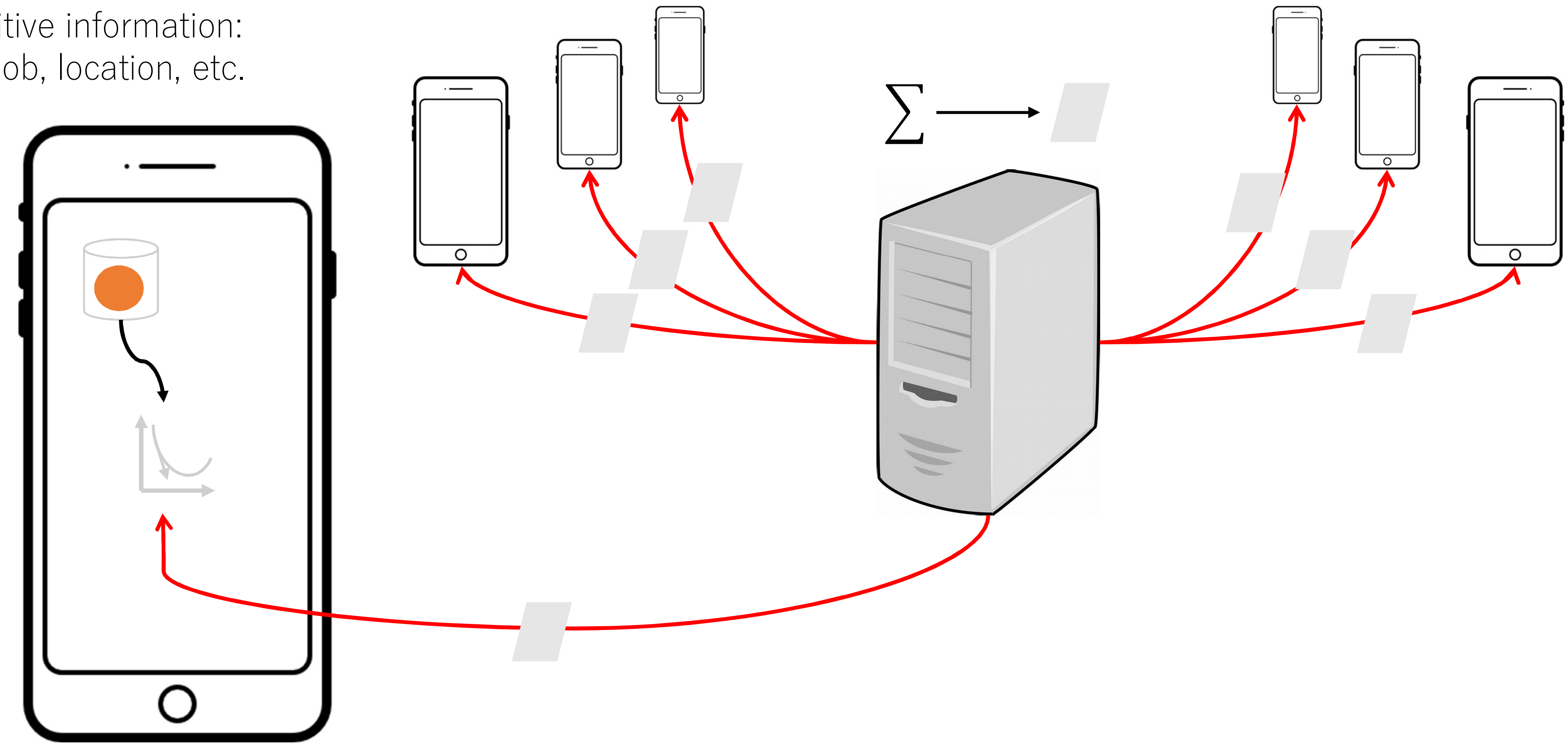


Federated Learning Privacy Vulnerabilities

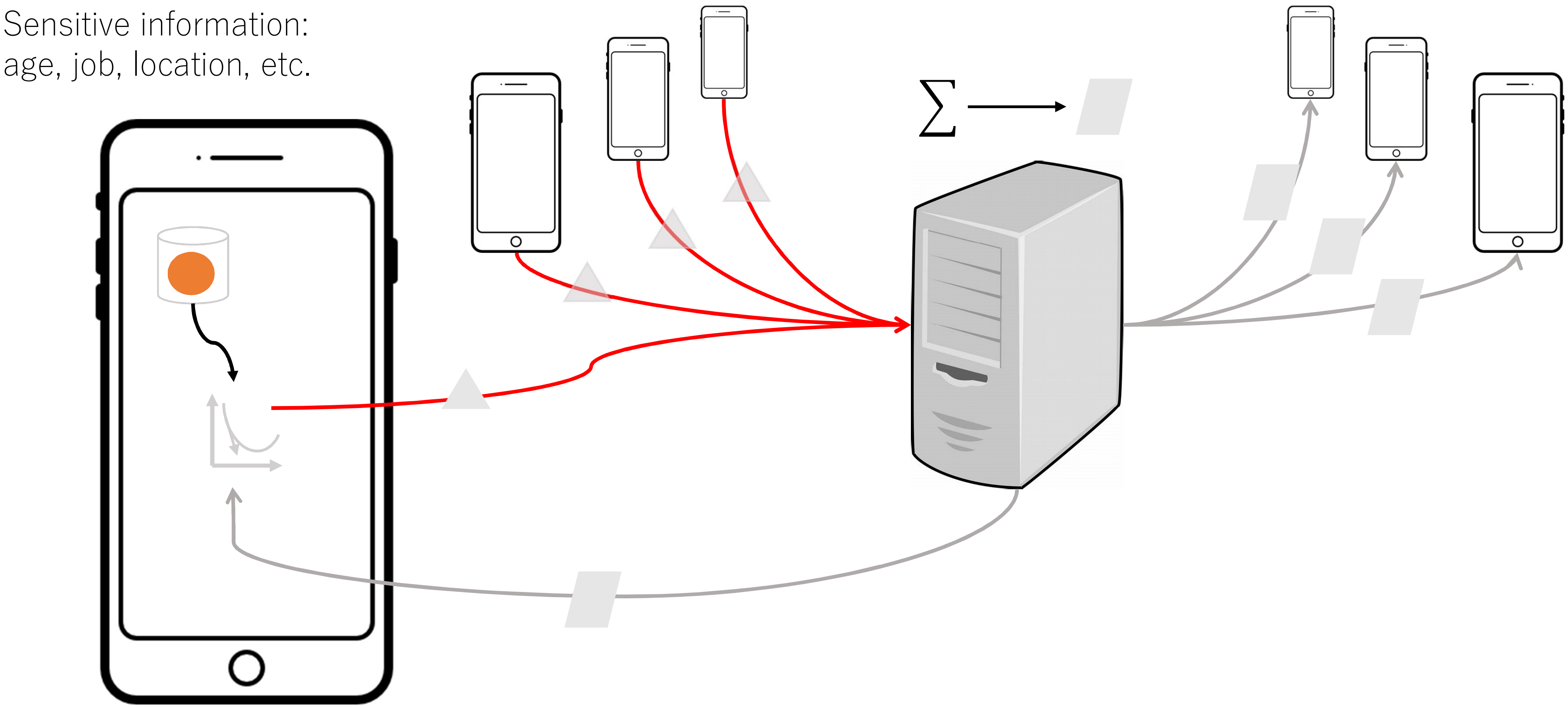


Federated Learning Privacy Vulnerabilities

Sensitive information:
age, job, location, etc.



Federated Learning Privacy Vulnerabilities



Federated Learning Privacy Vulnerabilities

Possible privacy attacks…

- **Membership Inference**

  "Whether data of a target victim has been used to train a model?"

- **Reconstruction attack**

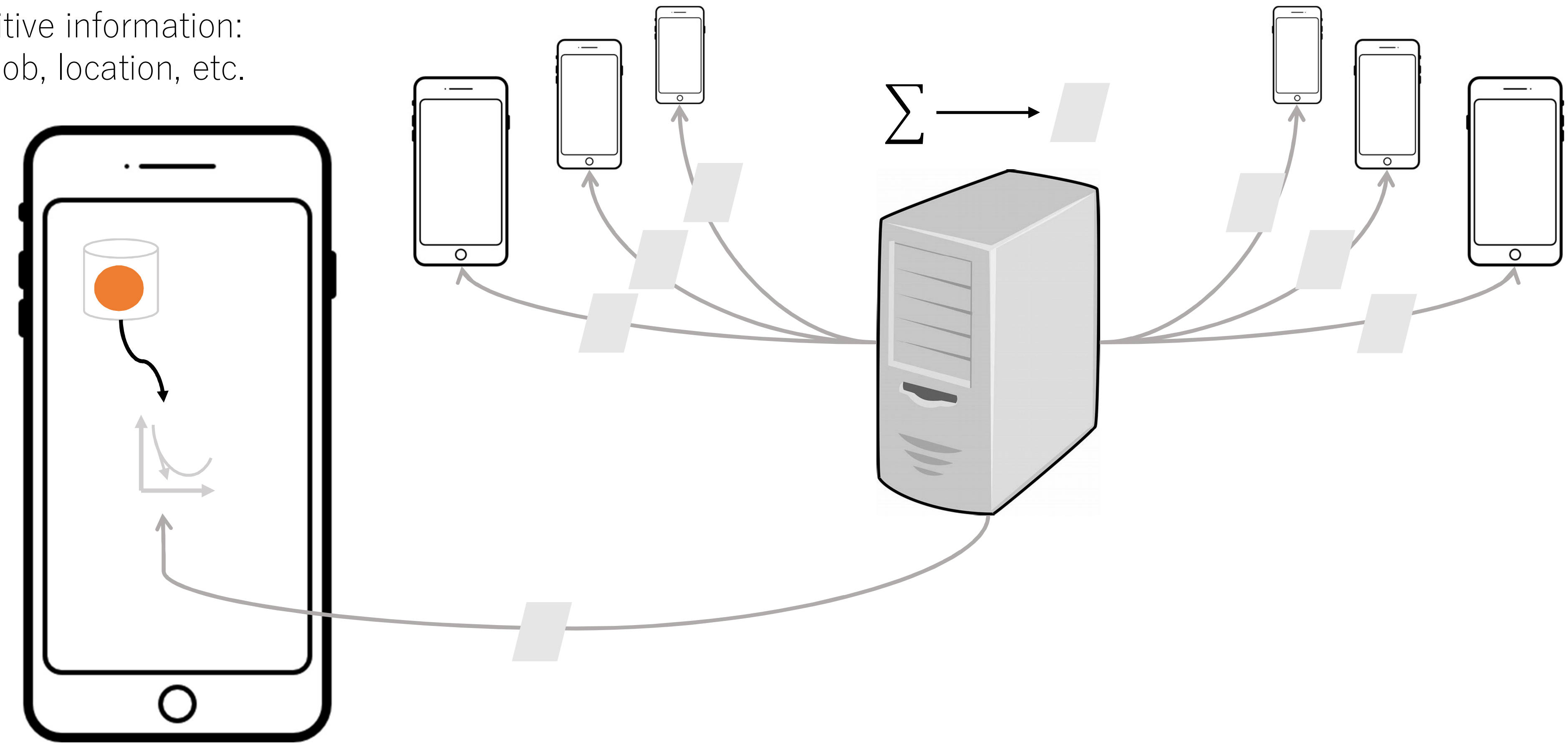
  Given a gender classifier, "What a male looks like?"

- **Unintended inference attack**

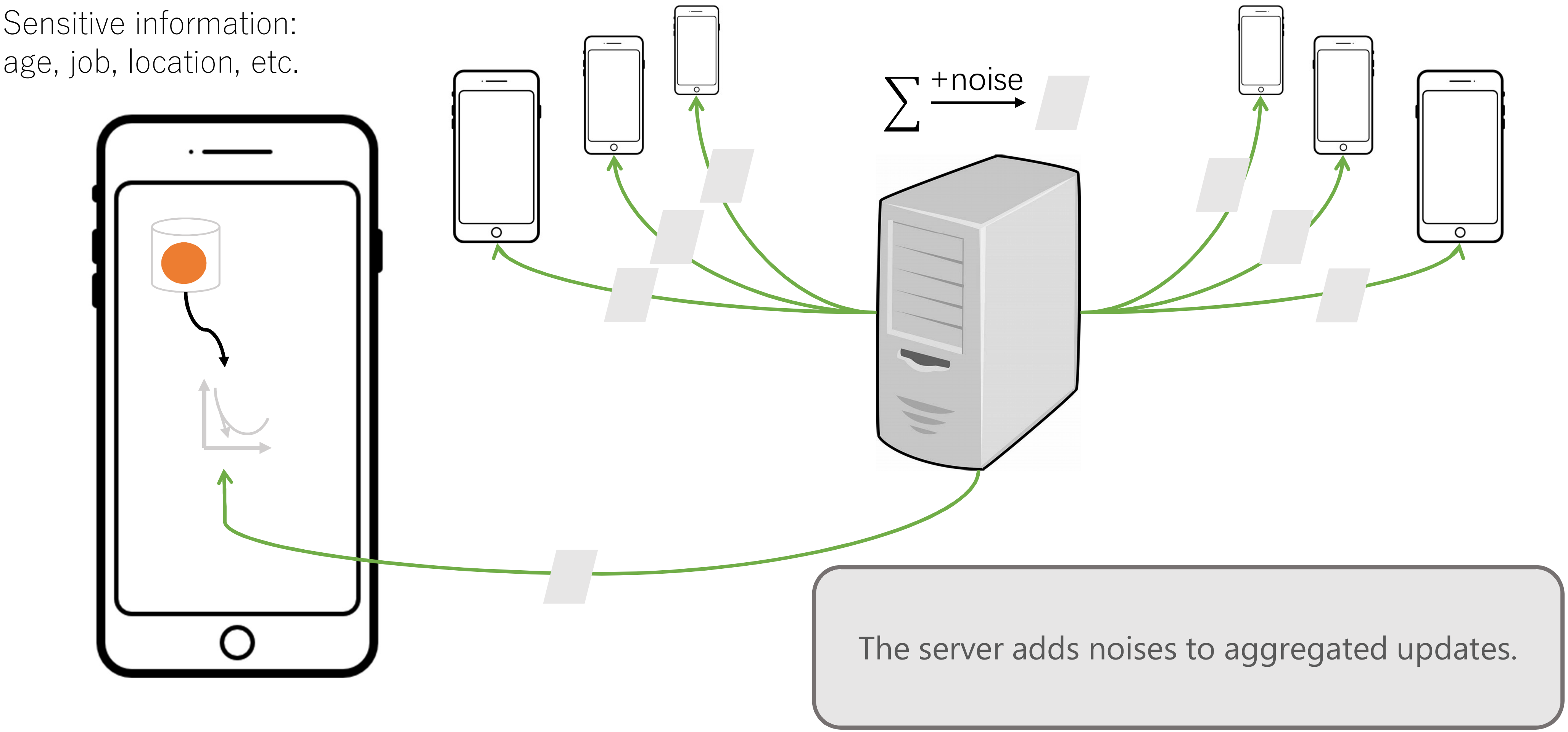
  Given a gender classifier, "What is the race of people in Bob's photos?"

Differential Privacy for Federated Learning

Sensitive information:
age, job, location, etc.

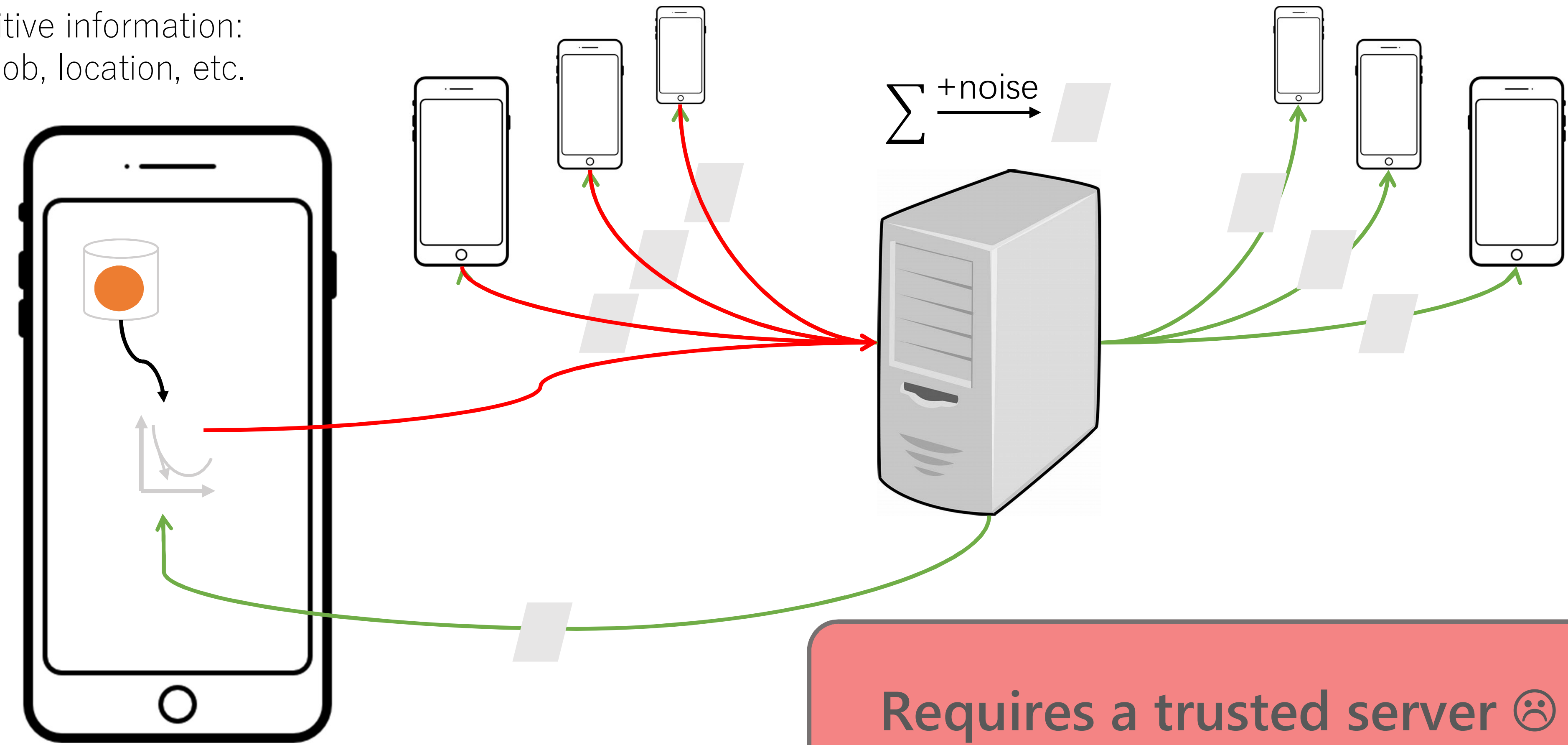


Differential Privacy for Federated Learning



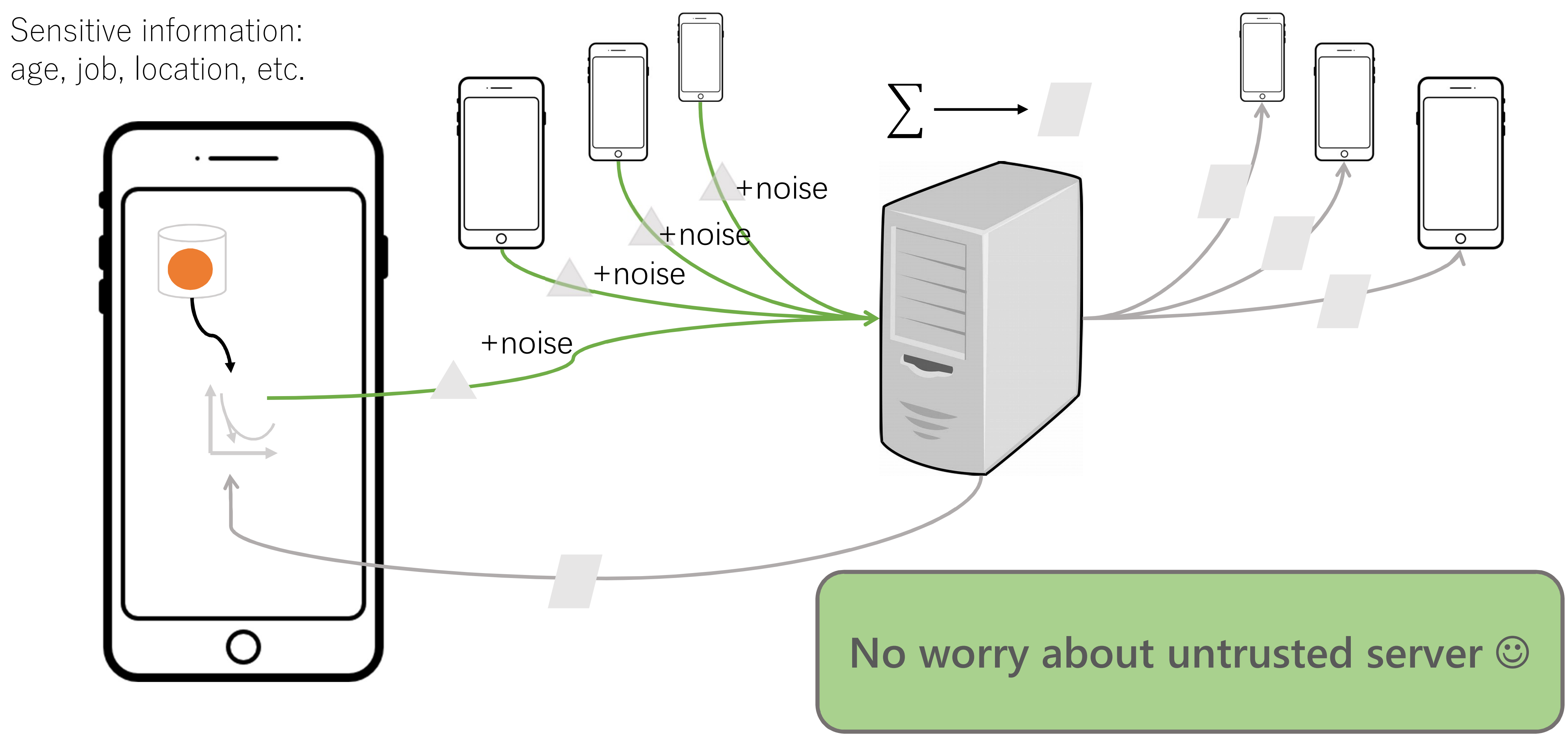
Differential Privacy for Federated Learning

Sensitive information:
age, job, location, etc.

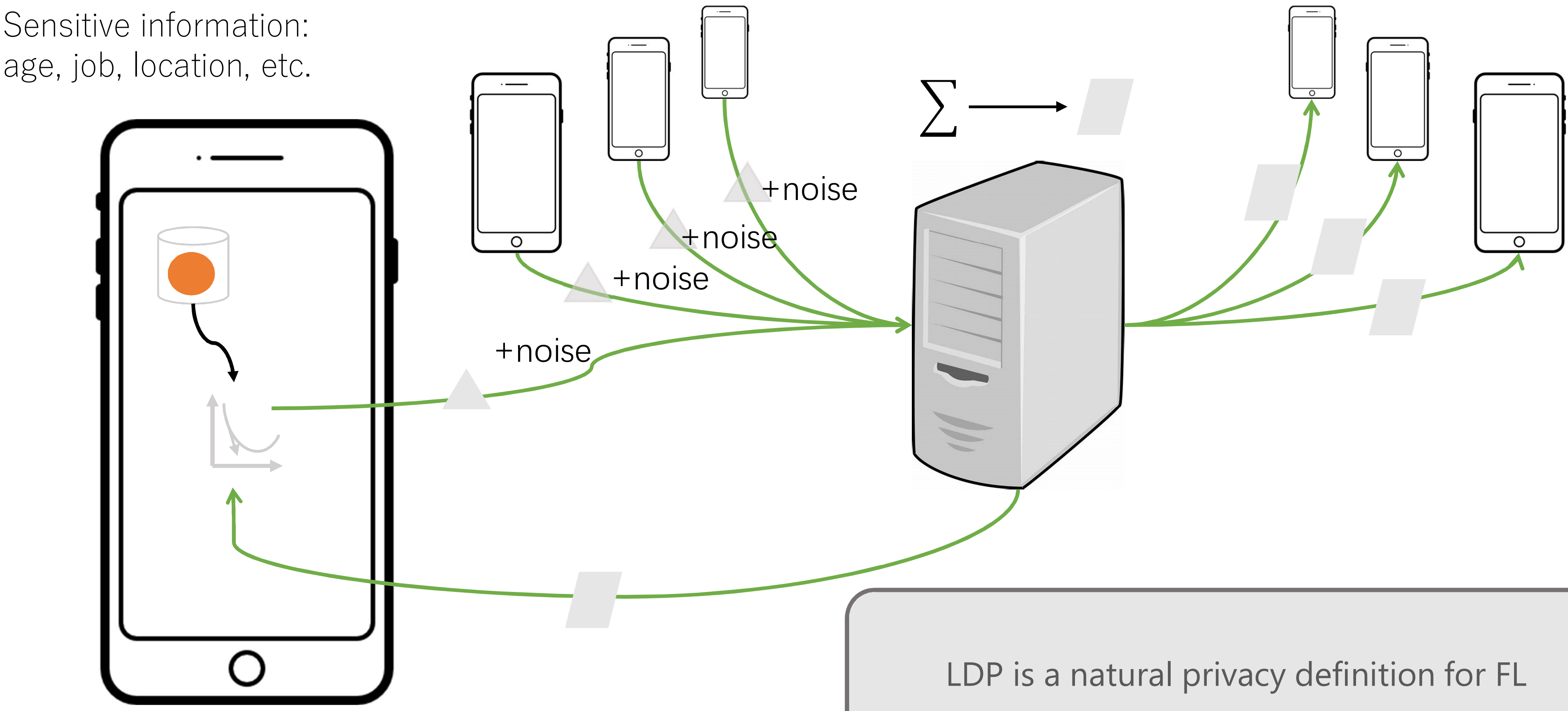


Requires a trusted server ☹

Local Differential Privacy for Federated Learning

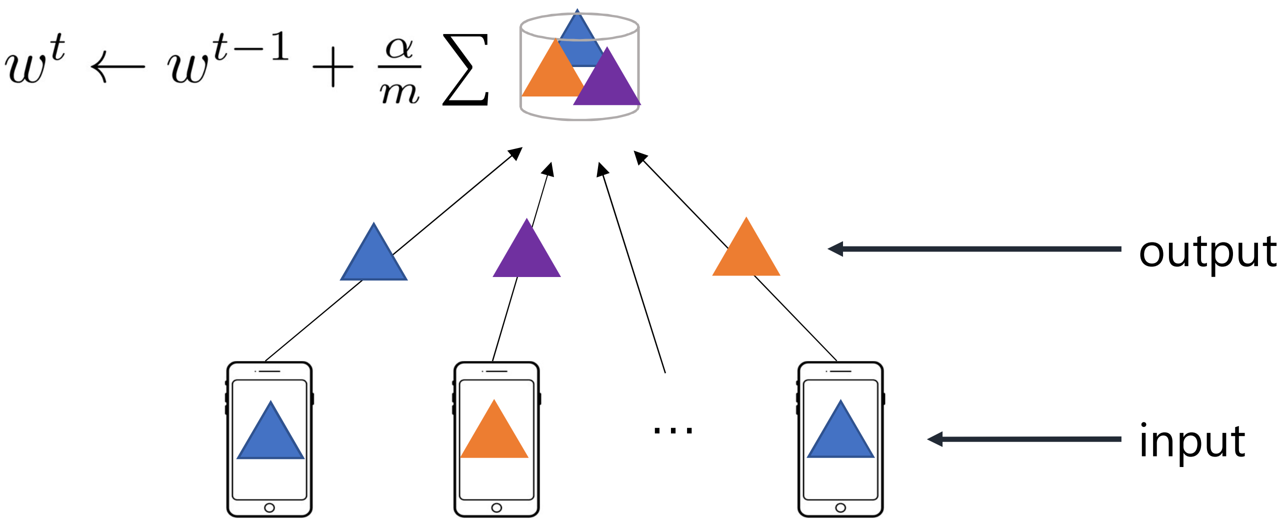


Local Differential Privacy for Federated Learning



LDP is a natural privacy definition for FL

Local Differential Privacy for Federated Learning



A randomized mechanism $\mathcal{M}$ is ϵ-LDP iff. for any two possible inputs v, v' and output v^*: $\frac{Pr[\mathcal{M}(v)=v^*]}{Pr[\mathcal{M}(v')=v^*]} \leq e^{\epsilon}$.

Challenges of LDP in Federated Learning

[1] Wang N, Xiao X, Yang Y, et al. Collecting and analyzing multidimensional data with local differential privacy[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019: 638-649.

For a $\boldsymbol{d}$-dimensional vector, the metric is:

- Given a local privacy budget ϵ for the vector,
- The error in the estimated mean of each dimension

If split local privacy budget to d dimensions[1]:

- The error is super-linear to d, and can be excessive when d is large

$$O\left(\frac{d\sqrt{\log d}}{\epsilon\sqrt{m}}\right)$$

Challenges of LDP in Federated Learning

[1] Wang N, Xiao X, Yang Y, et al. Collecting and analyzing multidimensional data with local differential privacy[C]//2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019: 638-649.

For a d-dimensional vector, the metric is:

- Given a local privacy budget ϵ for the vector,
- The error in the estimated mean of each dimension

If split local privacy budget to d dimensions[1]:

- The error is super-linear to d, and can be excessive when d is large

$$O\left(\frac{d\sqrt{\log d}}{\epsilon\sqrt{m}}\right)$$

An asymptotically optimal conclusion[1]:

1. Random sample k dimensions
   - Increase the privacy budget for each dimension
   - Reduce the noise variance incurred
2. Perturb each sampled dimension with ϵ/k
3. Aggregate and scale up by the factor of $\frac{d}{k}$

$$O\left(\frac{\sqrt{d\log d}}{\epsilon\sqrt{m}}\right)$$

Challenges of LDP in Federated Learning

$$O\left(\frac{\sqrt{d \log d}}{\epsilon \sqrt{m}}\right)$$

Typical orders-of-magnitude

d: 100-1,000,000s dimensions

m: 100-1000s users per round

ϵ: smaller privacy budget = stronger privacy

The dimension curse!

Our Intuition

Common bottleneck of the dimension curse

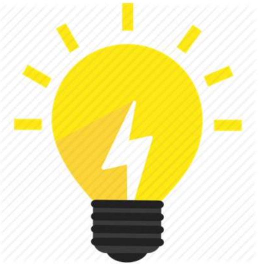
- **Distributed learning**
  - Data are partitioned and distributed for accelerating the training process
  - Gradient vectors are transmitted among separate workers
  - Communication costs = d × bits of representing one real value
- **Gradient sparsification**

  Reduce communication costs by only transmitting important dimensions
- **Intuition**

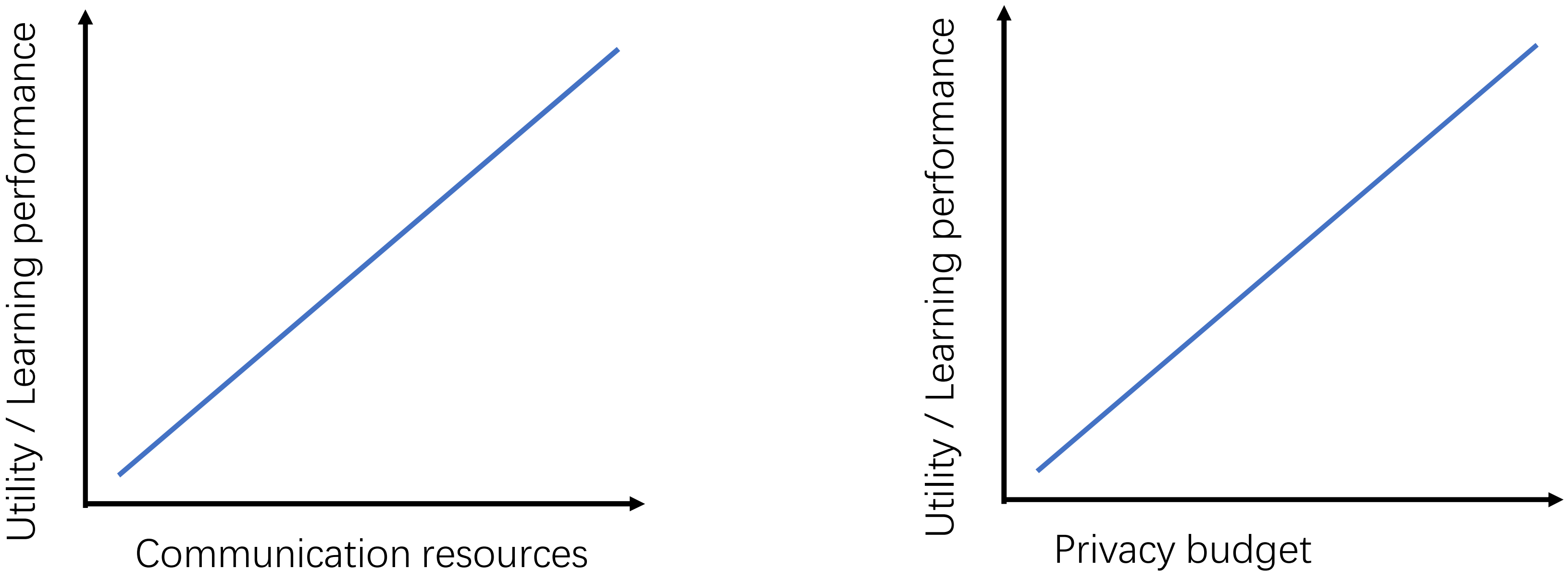
  Dimensions with larger absolute magnitudes are more important

  => Efficient dimension reduction for LDP

Our Intuition

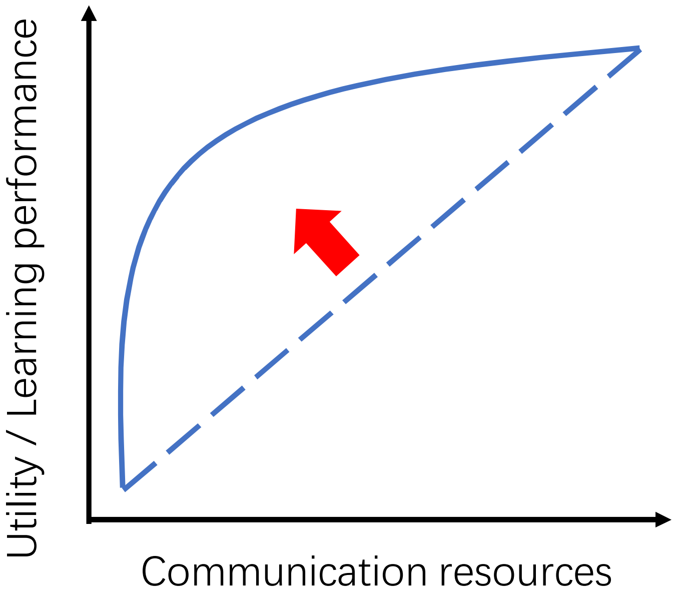


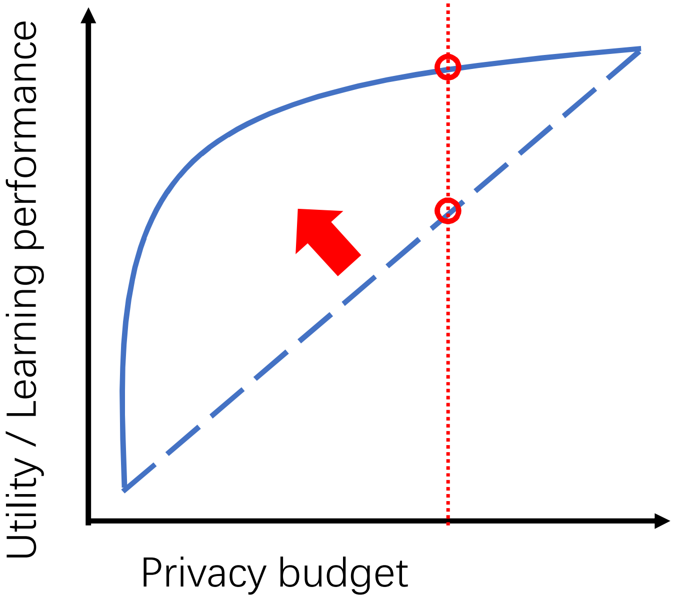
Common focus on selecting Top dimensions



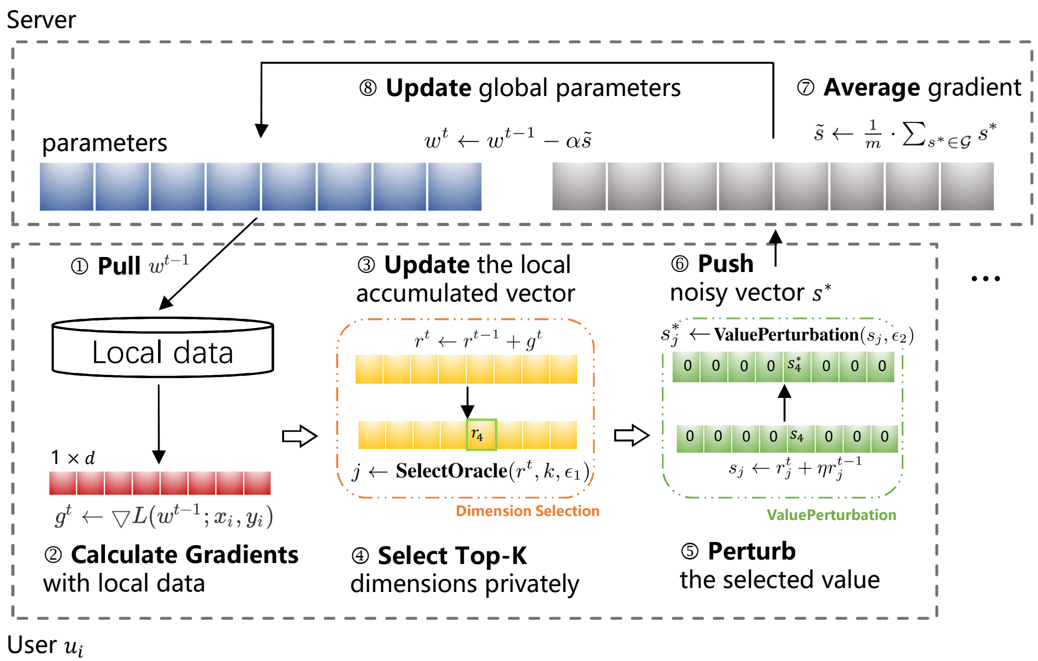
Our Intuition

Common focus on selecting Top dimensions





Two-stage Framework- FedSel



- **Top-k dimension selection is data-dependent**

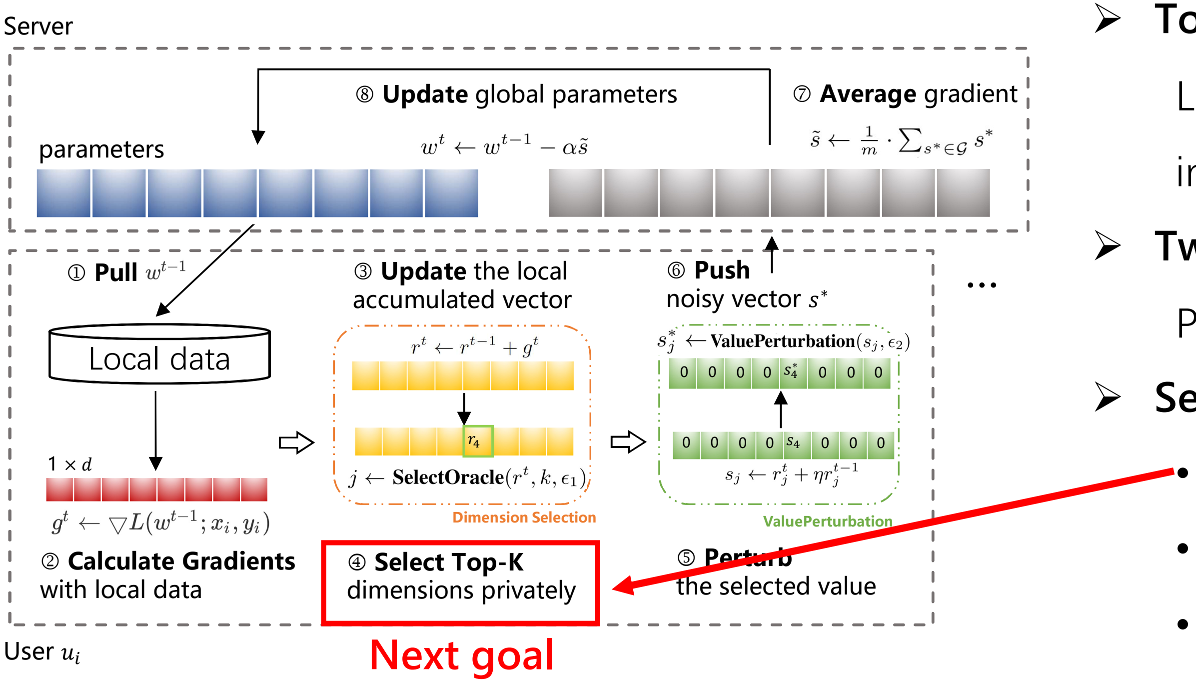
  Local vector = Top-k information + value information

- **Two-stage framework**

  Private selection + Value Perturbation

- **Sequential Composition**
  - The Top-k selection is ϵ_1-LDP
  - The value perturbation is ϵ_2-LDP
  - => The mechanism is ϵ-LDP, $\epsilon = \epsilon_1 + \epsilon_2$

Two-stage Framework- FedSel



Next goal

- **Top-k dimension selection is data-dependent**

  Local vector = Top-k information + value information

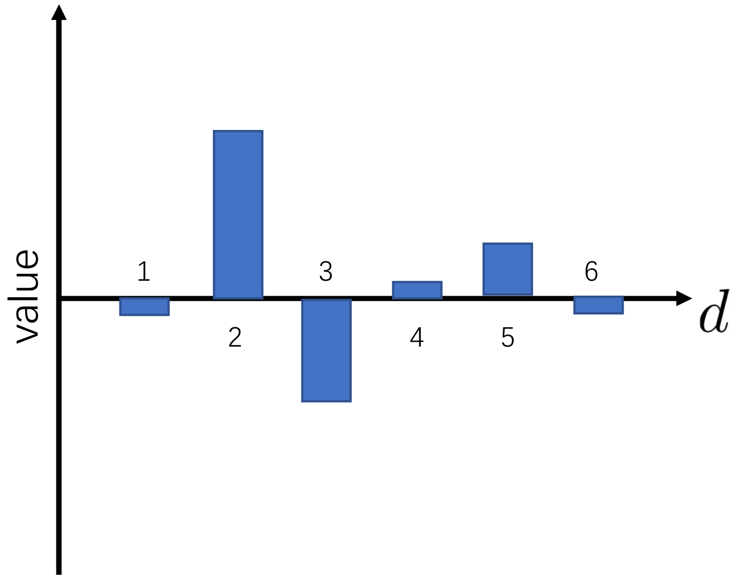
- **Two-stage framework**

  Private selection + Value Perturbation

- **Sequential Composition**
  - The Top-k selection is ϵ_1-LDP
  - The value perturbation is ϵ_2-LDP
  - => The mechanism is ϵ-LDP, $\epsilon = \epsilon_1 + \epsilon_2$

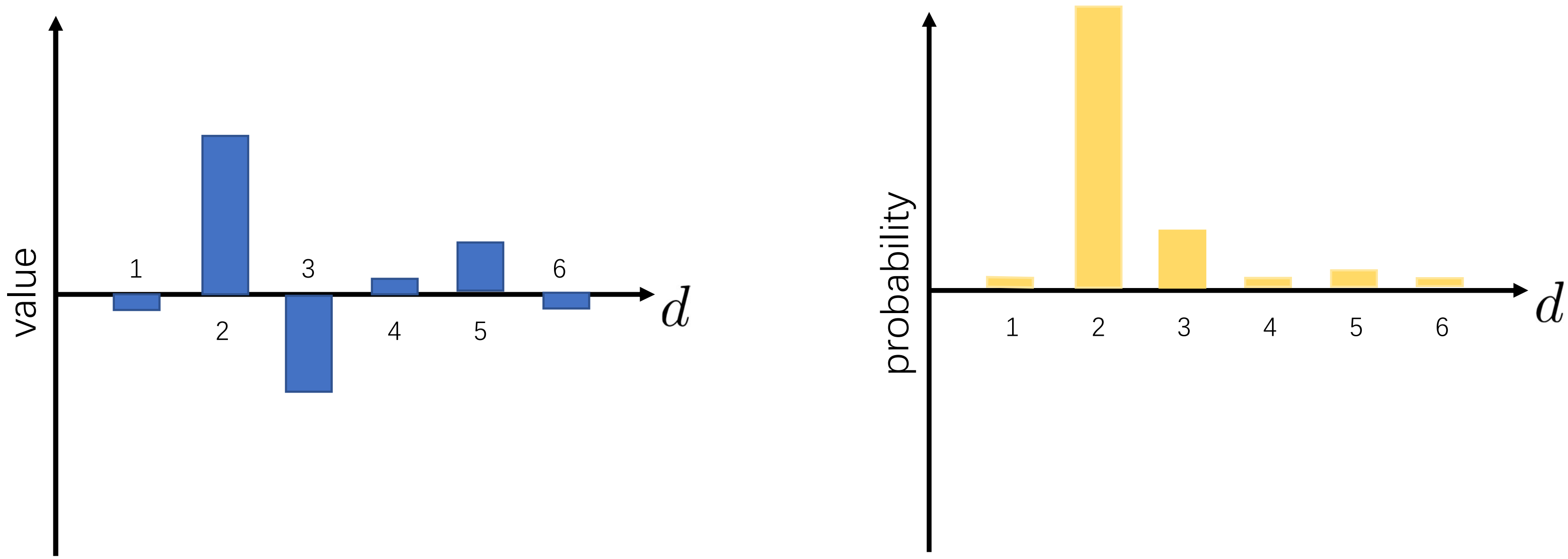
Methods-Exponential Mechanism (EXP)

1. Sorting and the ranking is denoted with $\{z_1, \ldots, z_d\} \in \{1, \ldots, \mathrm{d}\}^d$
2. Sample unevenly with the probability $\frac{\exp(\frac{\epsilon_1 z_j}{d-1})}{\sum_{i=1}^{d} \exp(\frac{\epsilon_1 z_i}{d-1})}$



Methods-Exponential Mechanism (EXP)

1. Sorting and the ranking is denoted with $\{z_1, \ldots, z_d\} \in \{1, \ldots, \mathrm{d}\}^d$
2. Sample unevenly with the probability $\frac{\exp(\frac{\epsilon_1 z_j}{d-1})}{\sum_{i=1}^{d} \exp(\frac{\epsilon_1 z_i}{d-1})}$



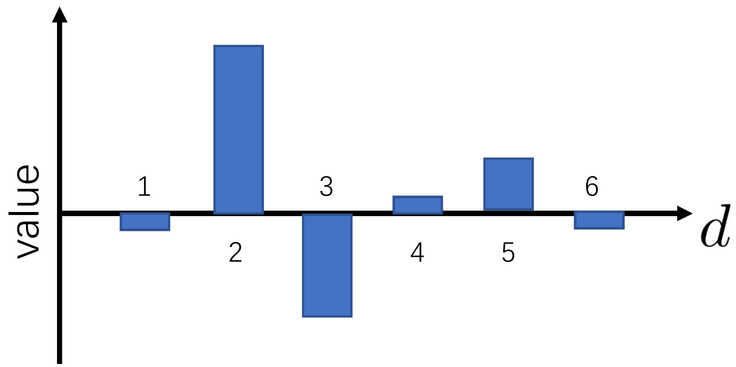
Methods-Perturbed Encoding Mechanism (PE)

1. Sorting and the ranking is denoted the Top-k status with $\{z_1, \ldots, z_d\} \in \{0,1\}^d$
2. For each dimension,

   to retain status z_j with a larger probability p

   to flip z_j has a smaller probability $1 - p$
3. Sample from dimension set $\mathbb{S} = \{j | z_j^* = 1\}$

$$p = \frac{e^{\epsilon_1}}{e^{\epsilon_1} + 1}$$



$$\{z_1, \cdots, z_d\} = \{0, 1, 1, 0, 0, 0\}$$

Methods-Perturbed Encoding Mechanism (PE)

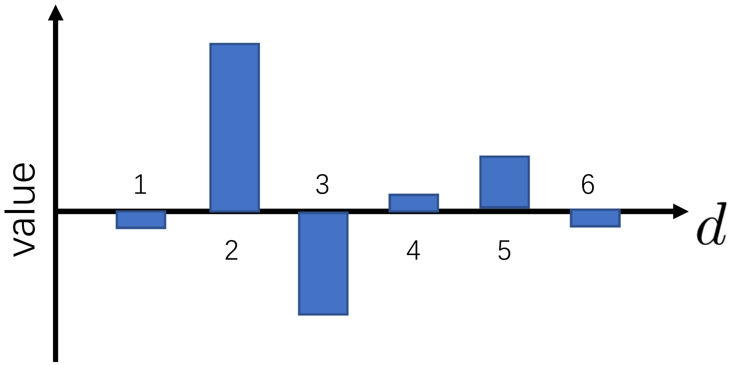
1. Sorting and the ranking is denoted the Top-k status with $\{z_1, \ldots, z_d\} \in \{0,1\}^d$
2. For each dimension,

   to retain status z_j with a larger probability p

   to flip z_j has a smaller probability $1 - p$

3. Sample from dimension set $\mathbb{S} = \{j | z_j^* = 1\}$

$$p = \frac{e^{\epsilon_1}}{e^{\epsilon_1} + 1}$$



$$\{z_1, \cdots, z_d\} = \{0, 1, 1, 0, 0, 0\}$$

$$\{\acute{z}_1, \cdots, \acute{z}_d\} = \{0, 0, 1, 0, 1, 0\}$$

Methods-Perturbed Encoding Mechanism (PE)

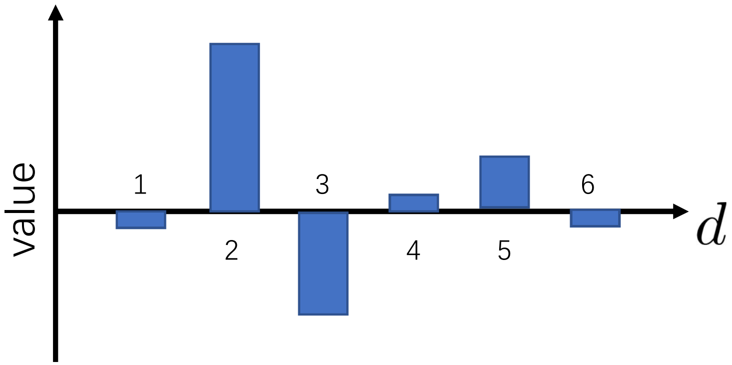
1. Sorting and the ranking is denoted the Top-k status with $\{z_1, \ldots, z_d\} \in \{0,1\}^d$
2. For each dimension,

   to retain status z_j with a larger probability p

   to flip z_j has a smaller probability $1 - p$

$$p = \frac{e^{\epsilon_1}}{e^{\epsilon_1}+1}$$

3. Sample from dimension set $\mathbb{S} = \{j | z_j^* = 1\}$



$$\{z_1, \cdots, z_d\} = \{0, 1, 1, 0, 0, 0\}$$

$$\{\acute{z}_1, \cdots, \acute{z}_d\} = \{0, 0, 1, 0, 1, 0\}$$

$$\mathbb{S} = \{3, 5\}$$

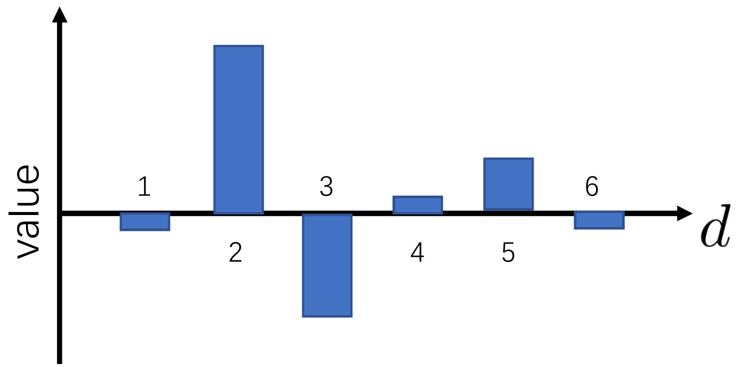
Methods-Perturbed Sampling Mechanism (PS)

1. Sorting and the ranking is denoted the Top-k status with $\{z_1, \ldots, z_d\} \in \{0,1\}^d$
2. Sample a dimension from:

   Top-k dimension set, with a larger probability p

   Non-top dimension set, with a smaller probability $1-p$

$$p = \frac{d^{\epsilon_1} \cdot k}{d-k+e^{\epsilon_1} \cdot k}$$

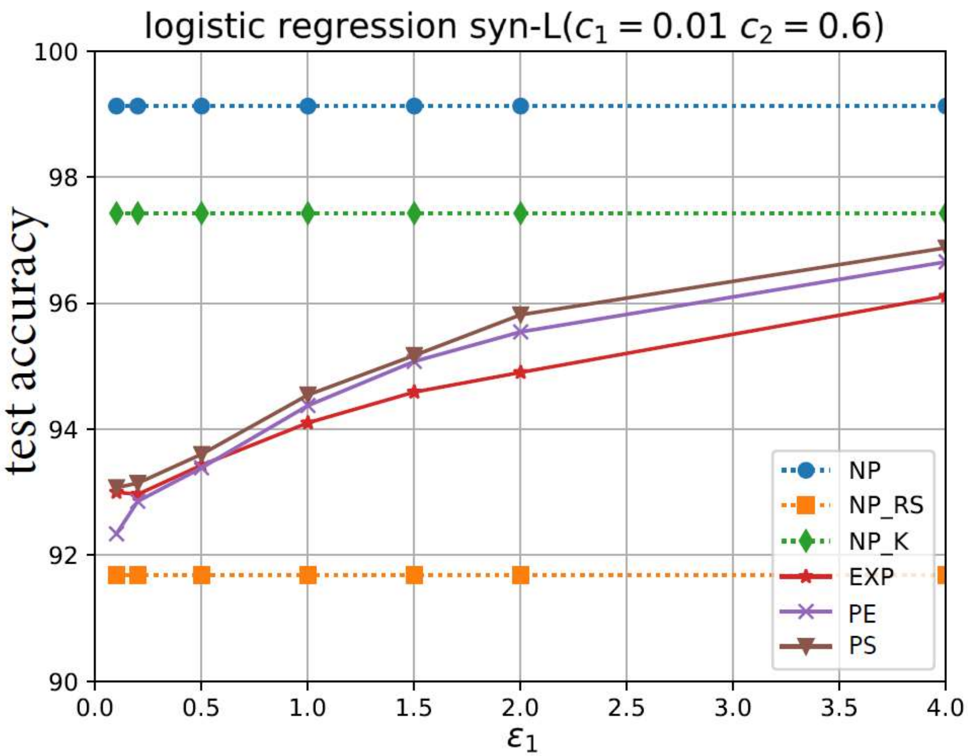


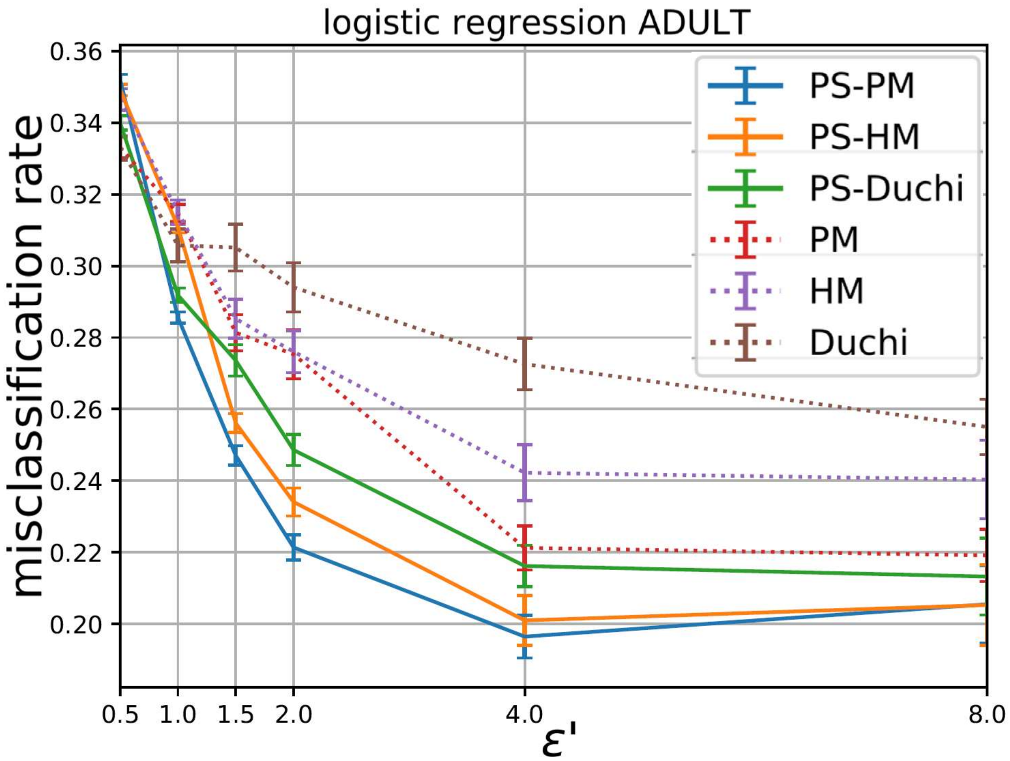
$$\{z_1, \cdots, z_d\} = \{0, 1, 1, 0, 0, 0\}$$

$$\text{Top-k set } \{2, 3\}$$

$$\text{Non-top set } \{1, 4, 5, 6\}$$

Empirical results





- Even a **small** budget in dimension selection helps to increase the learning accuracy
- Private Top-k selection helps to improve the learning utility **independent** of the mechanism for perturbing one dimension.

Empirical results

dataset	model	EXP-gain	EXP-loss	PE-gain	PE-loss	PS-gain	PS-loss
syn-L-0.01-0.9	logistic	**8.6074**	0.3517	**5.410**	1.192	**5.975**	0.4970
syn-L-0.01-0.9	SVM	**7.1950**	2.1593	**3.7704**	0.8533	**5.065**	2.0816
BANK	logistic	**2.4197**	-0.157	**3.2338**	0.0464	**2.5525**	0.1463
BANK	SVM	**4.3823**	0.4436	**3.4369**	0.2530	**4.0244**	0.0164
KDD	logistic	**2.0471**	0.5091	**2.5148**	0.2322	**2.0171**	0.3428
KDD	SVM	**1.85629**	-0.1625	**2.2168**	0.2288	**1.8291**	0.4465
ADULT	logistic	**5.5745**	0.2935	**5.6445**	1.3096	**6.0535**	0.8091
ADULT	SVM	**5.5361**	0.1949	**5.6057**	0.9550	**5.1442**	0.3852

$$\text{gain} = \text{acc}(\text{EXP/PE/PS-PM-C}) - \text{acc}(\text{PM}),$$
$$\text{loss} = \text{acc}(\text{EXP/PE/PS-PM-C}) - \text{acc}(\text{EXP/PE/PS-PM}).$$

What we **gain is much larger** than what we lose
from private and efficient Top-k selection

Summary

Conclusion

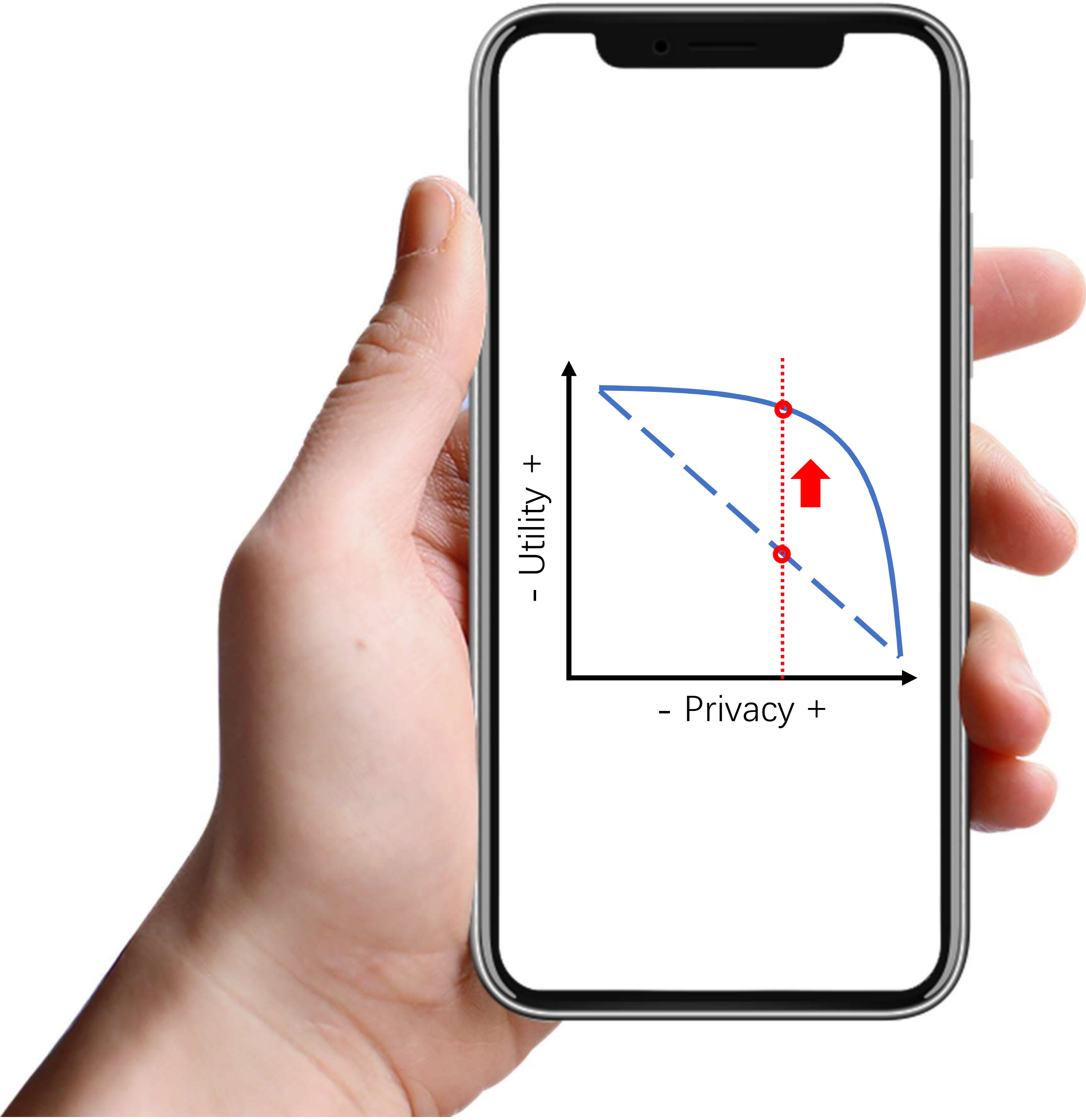
- We propose a two-stage framework for locally differential private federated SGD
- We propose 3 private selection mechanisms for efficient dimension reduction under LDP

Takeaway

- Private mechanism can be specialized for sparse vector
- Private Top-k dimension selection can improve learning utility under a given privacy level

Future work

- Optimal hyper-parameter tuning



Thanks