Securing Secure Aggregation: Mitigating Multi-Round Privacy Leakage in Federated Learning

Ramy E. Ali

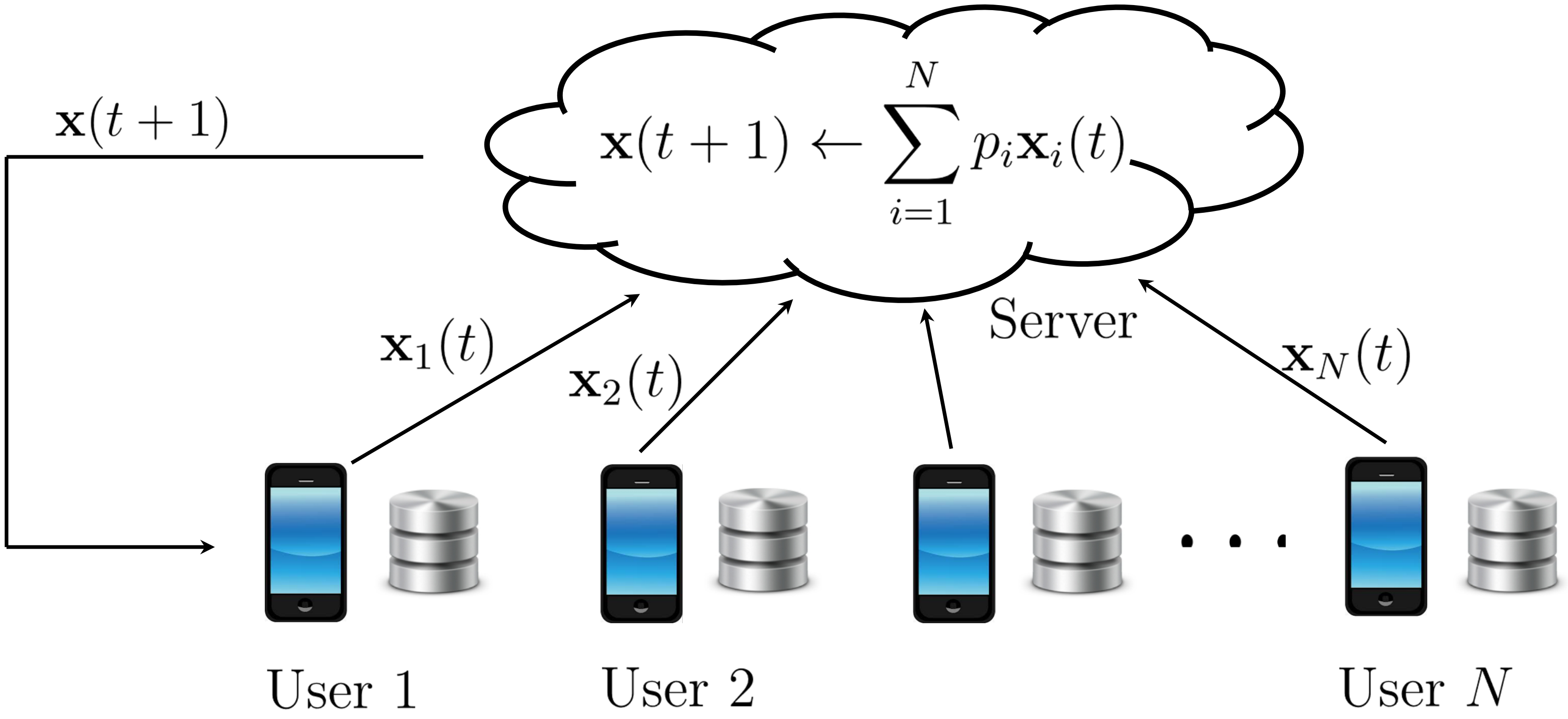
USC University of Southern California

In collaboration with Jinhyun So (USC), Başak Güler (UCR), Salman Avestimehr (USC) and Jiantao Jiao (UC Berkeley)

2021

The Promise of Federated Learning

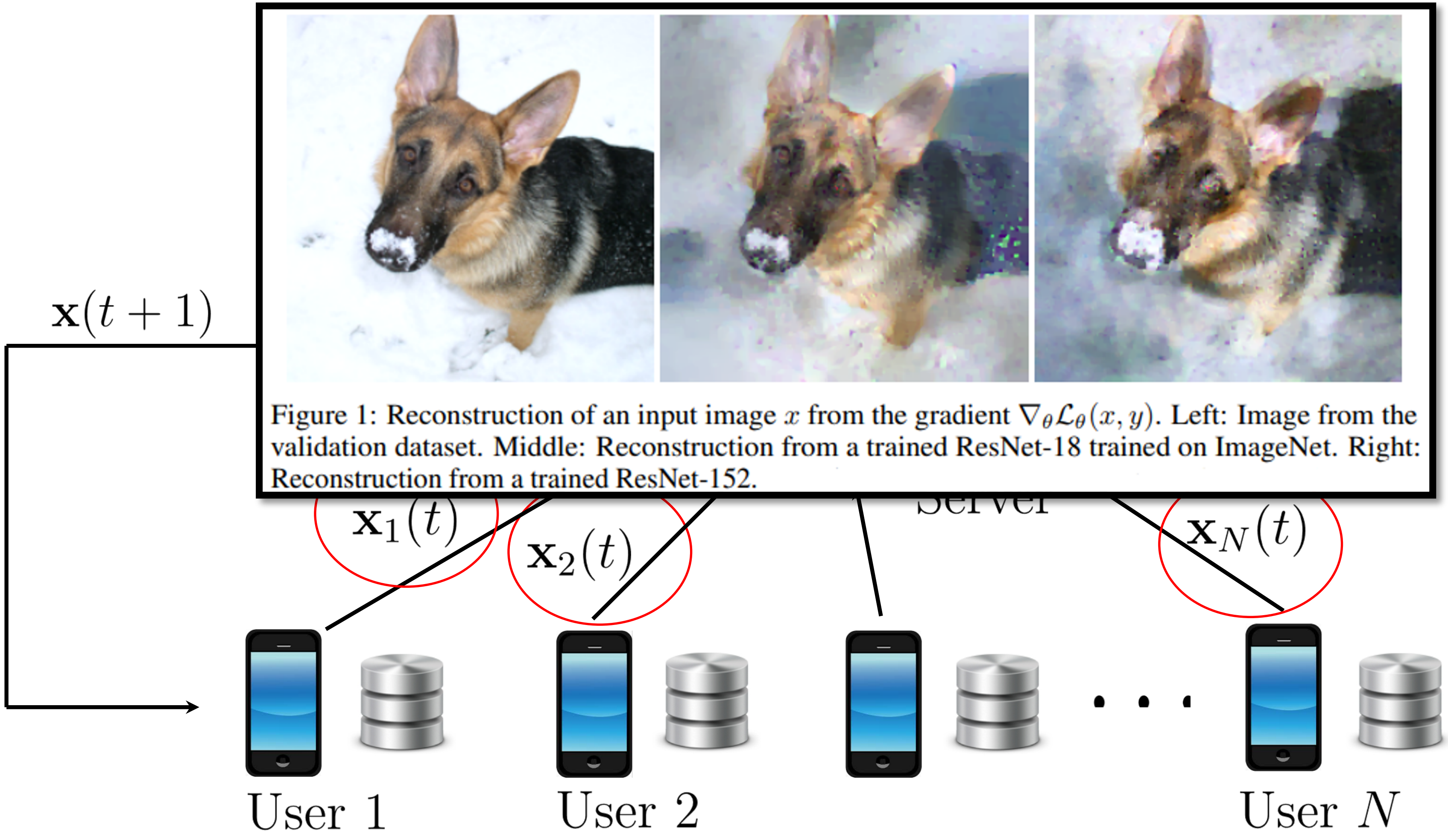
Main principle: train locally - average globally



Ensuring **privacy** by avoiding data sharing?

But ...

Model Inversion Attack

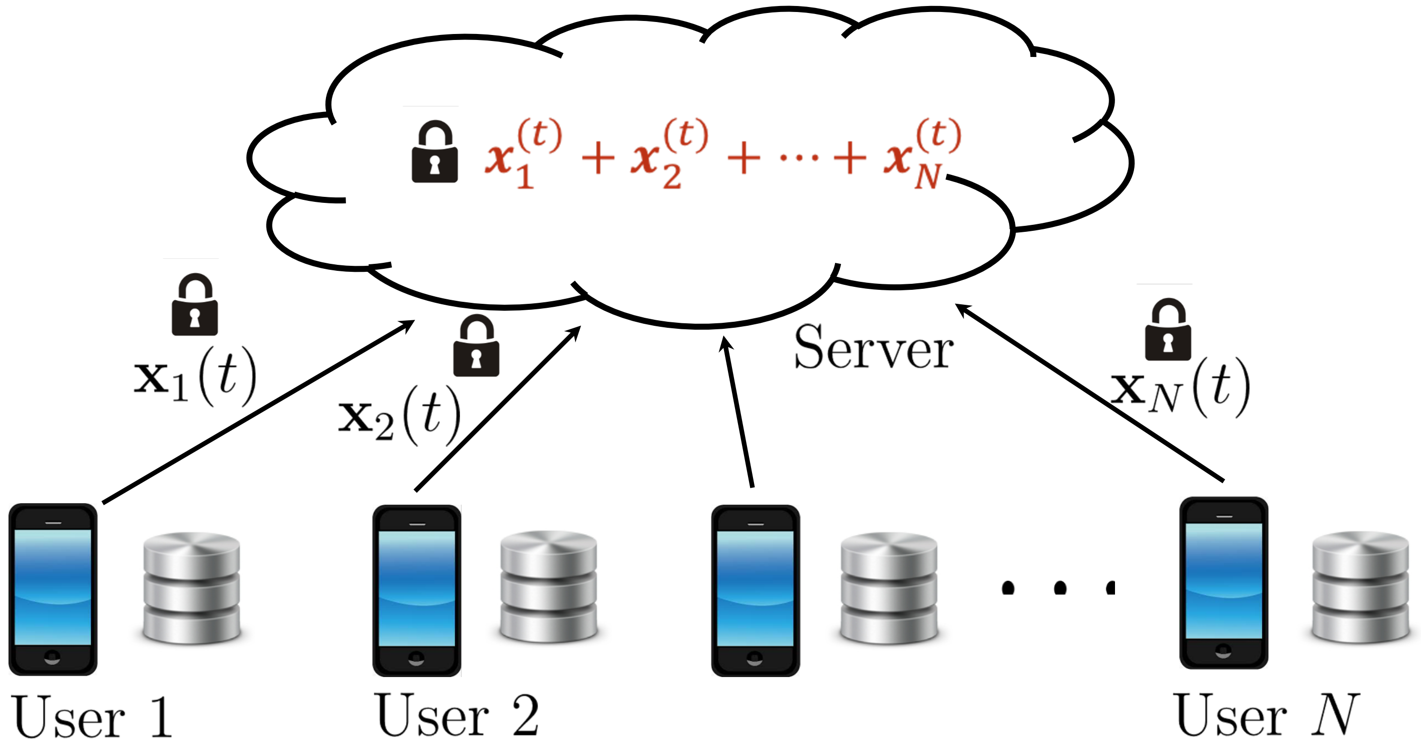


[Geiping et al]

Problem: Individual model updates can leak sensitive data

Remedy: Secure Model Aggregation

- Secure aggregation ensures that the server only learns the global model.



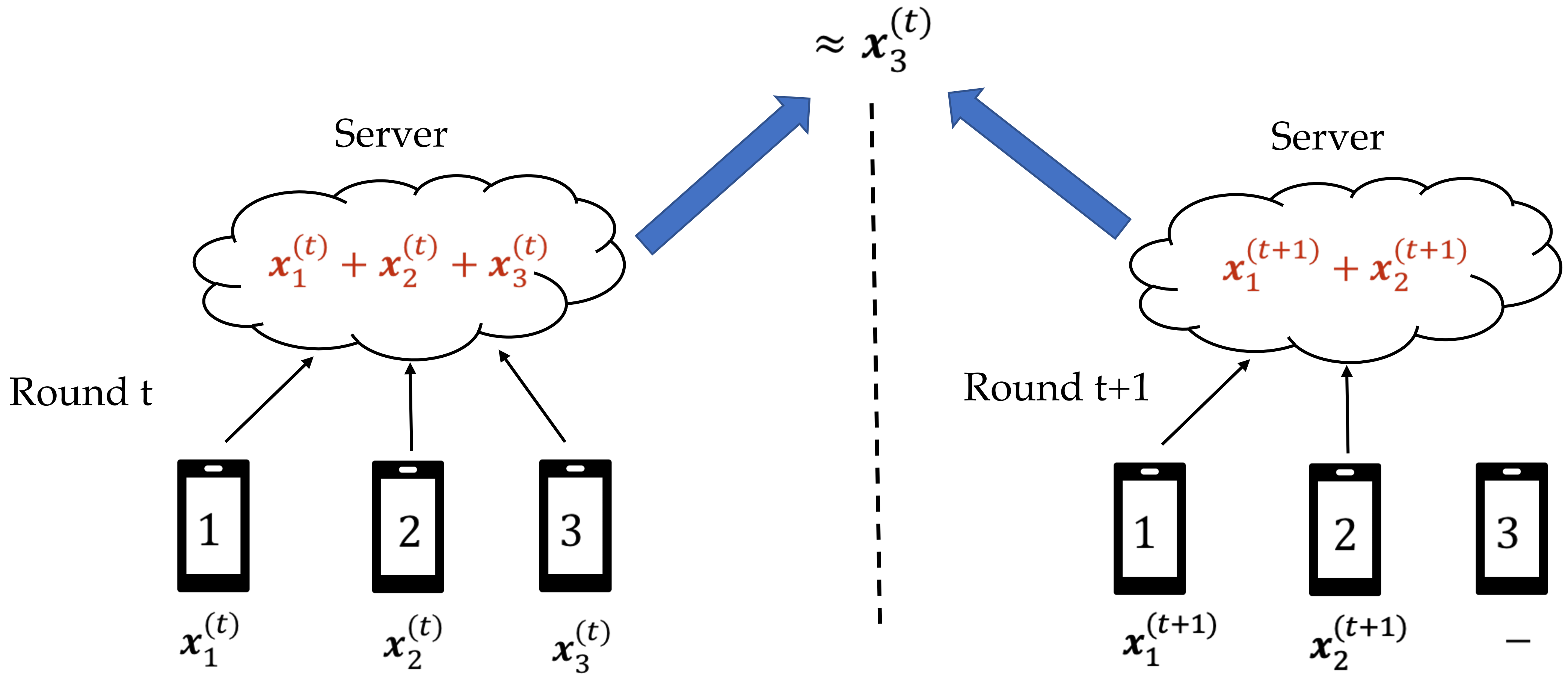
Secure Aggregation is Essentially an MPC problem with User Dropouts

Bad news ...

- Secure aggregation, however, is not secure over multiple rounds.

Bad news ...

- Secure aggregation, however, is not secure over multiple rounds.
- Intuition: partial user participation leads to privacy leakage

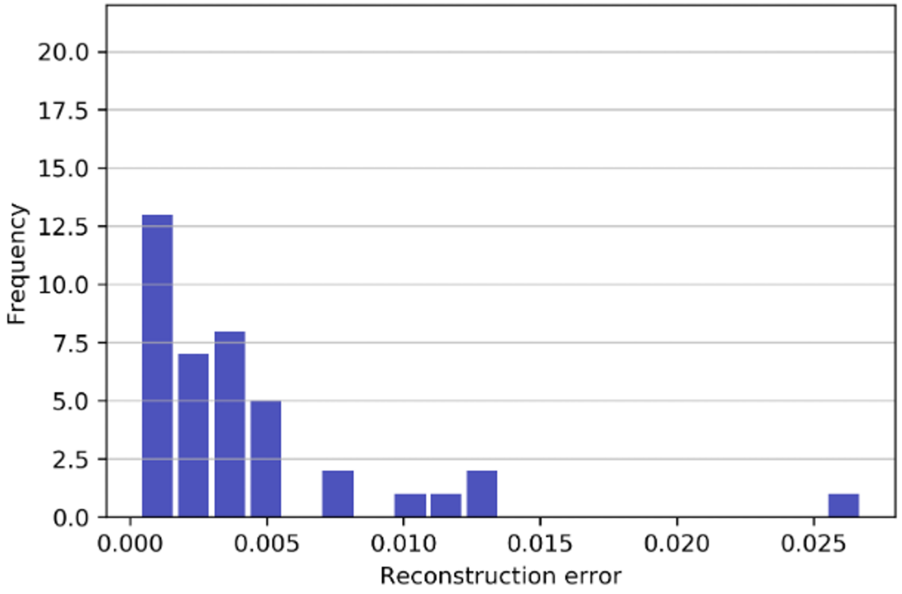


This is a serious issue!

- Random selection may reveal all individual models.
- Experiment
  - N=40 users
  - MNIST dataset with non-IID distribution
  - K=8 users are selected at random at each round
  - The server estimates the individual gradients through least-squares

Reconstruction error of user i at time t

$$\frac{\left\|\boldsymbol{x}_i^{(t)} - \widehat{\boldsymbol{x}}_i^{(t)}\right\|^2}{\left\|\boldsymbol{x}_i^{(t)}\right\|^2}$$



reconstruction error after 600 rounds

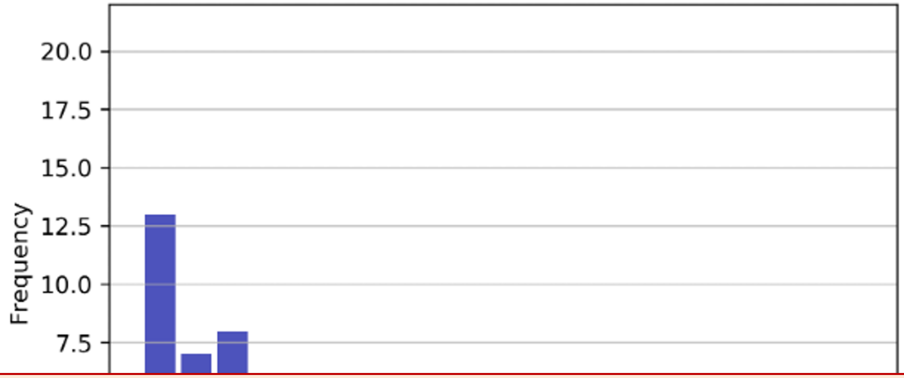
reconstruction error < 0.005 (for many users)

This is a serious issue!

- Random selection may reveal all individual models.
- Experiment
  - N=40 users
  - MNIST dataset with non-IID distribution
  - K=8 users are selected at random at each round
  - The server estimates the individual gradients through least-squares

Reconstruction error of user i at time t

$$\frac{\left\|\boldsymbol{x}_i^{(t)} - \hat{\boldsymbol{x}}_i^{(t)}\right\|^2}{\left\|\boldsymbol{x}_i^{(t)}\right\|^2}$$



reconstruction error < 0.005 (for many users)

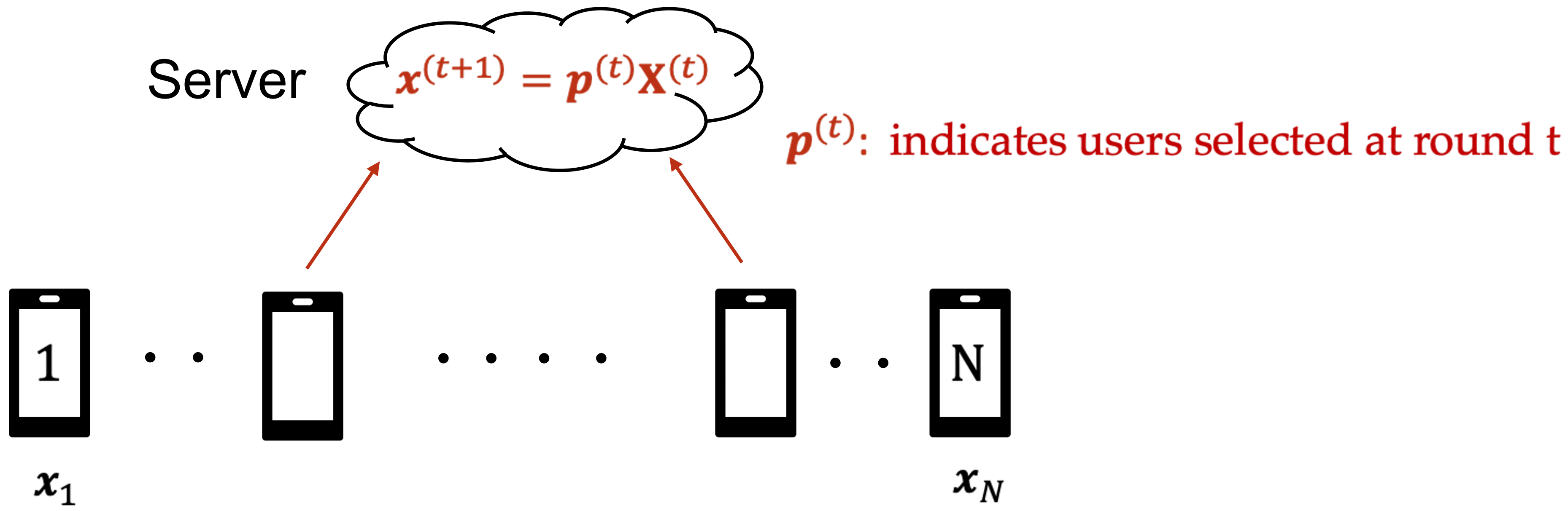
Similar concerns have also been reported in some recent works:

- Pejó et al. "Quality Inference in Federated Learning with Secure Aggregation.", 2020.
- M. Lam et al. "Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix.", ICML 2021.

This talk

- Introduce a notion/metric for multi-round privacy
- Propose Multi-RoundSecAgg, which ensures multi-round privacy
    - It further optimizes the average number of participating users (convergence rate) and fairness in user selection
    - It also introduces a trade-off between “privacy” and “convergence rate”

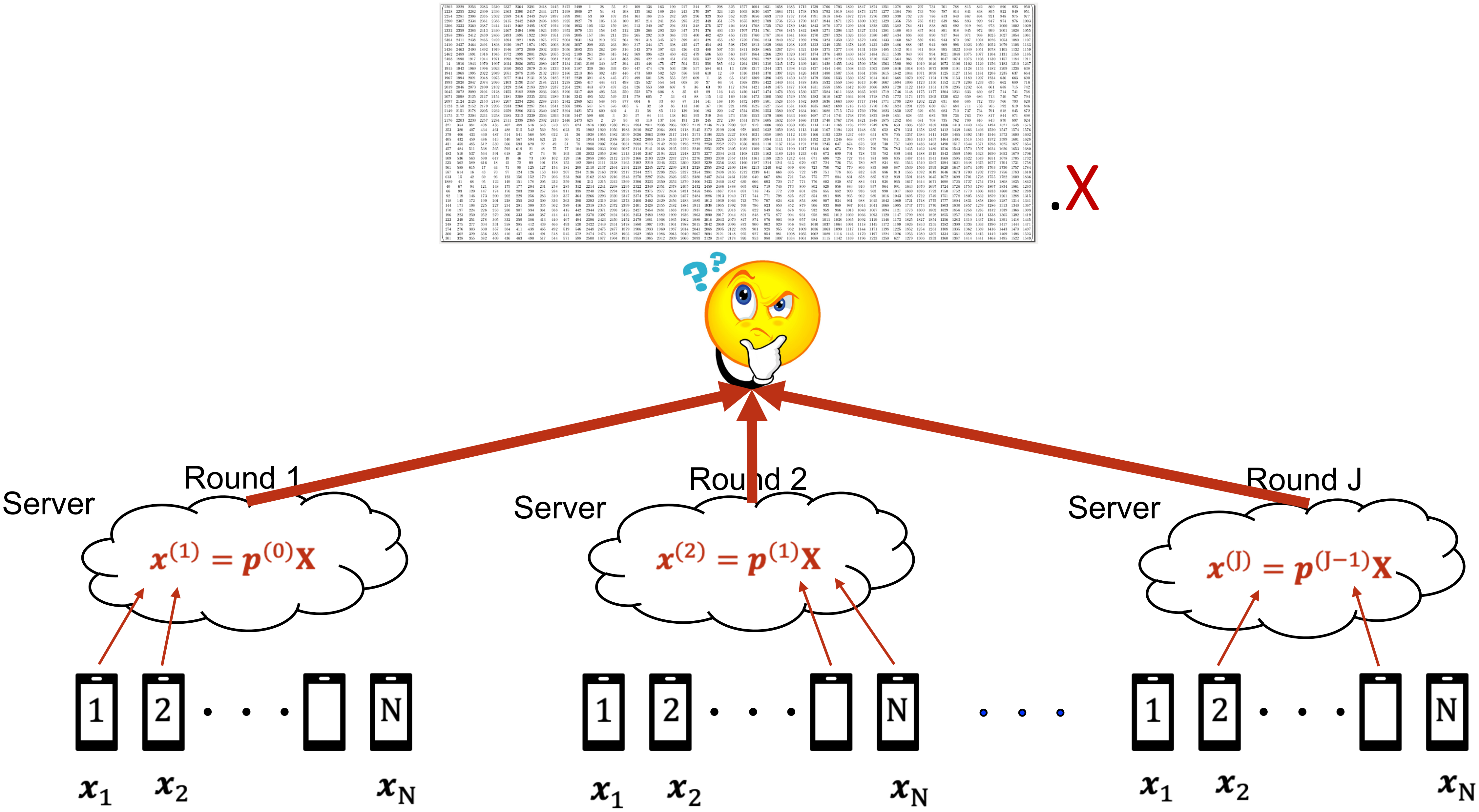
Federated Averaging with Partial User Participation



- Participation matrix $\boldsymbol{P}^{(J)} = \begin{pmatrix} \boldsymbol{p}^{(0)} \\ \vdots \\ \boldsymbol{p}^{(J-1)} \end{pmatrix} \in \{0,1\}^{J \times N}$, J: number of rounds

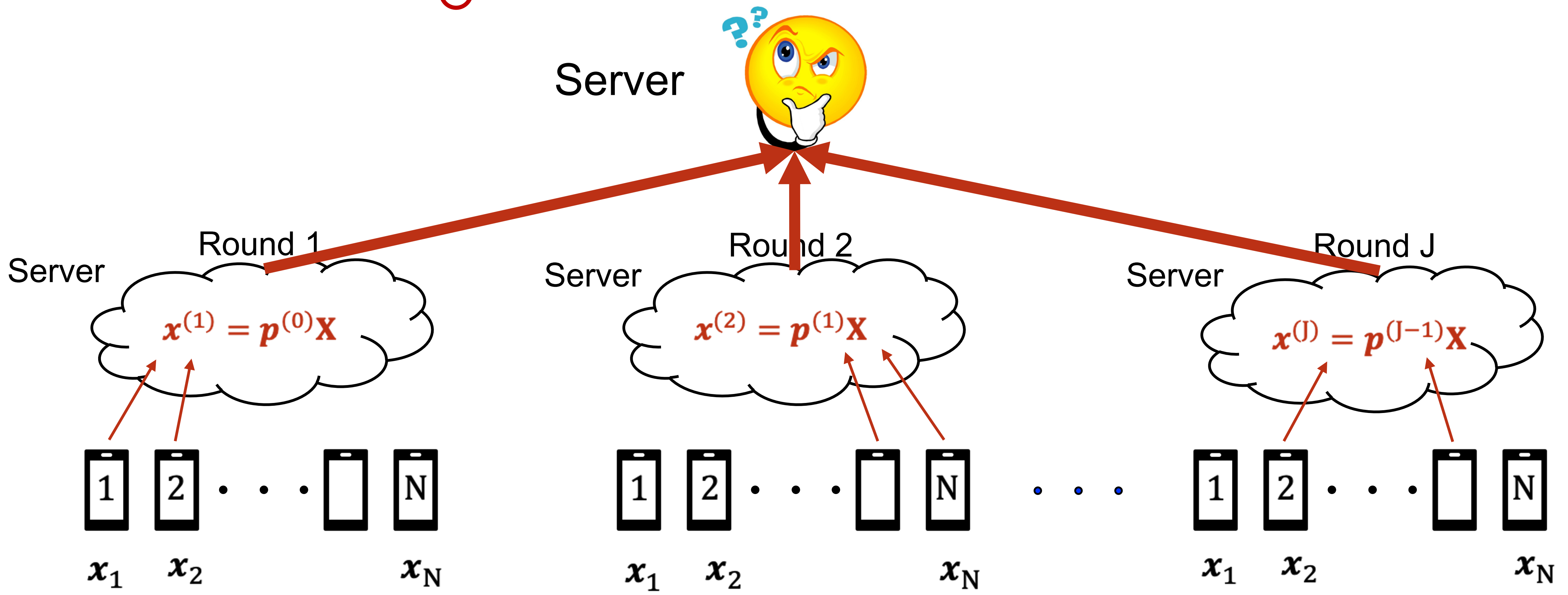
How should we choose $\boldsymbol{P}^{(J)}$ to ensure long-term privacy?

Metric 1: Multi-round Privacy (T)



Metric 1: Multi-round Privacy (T)

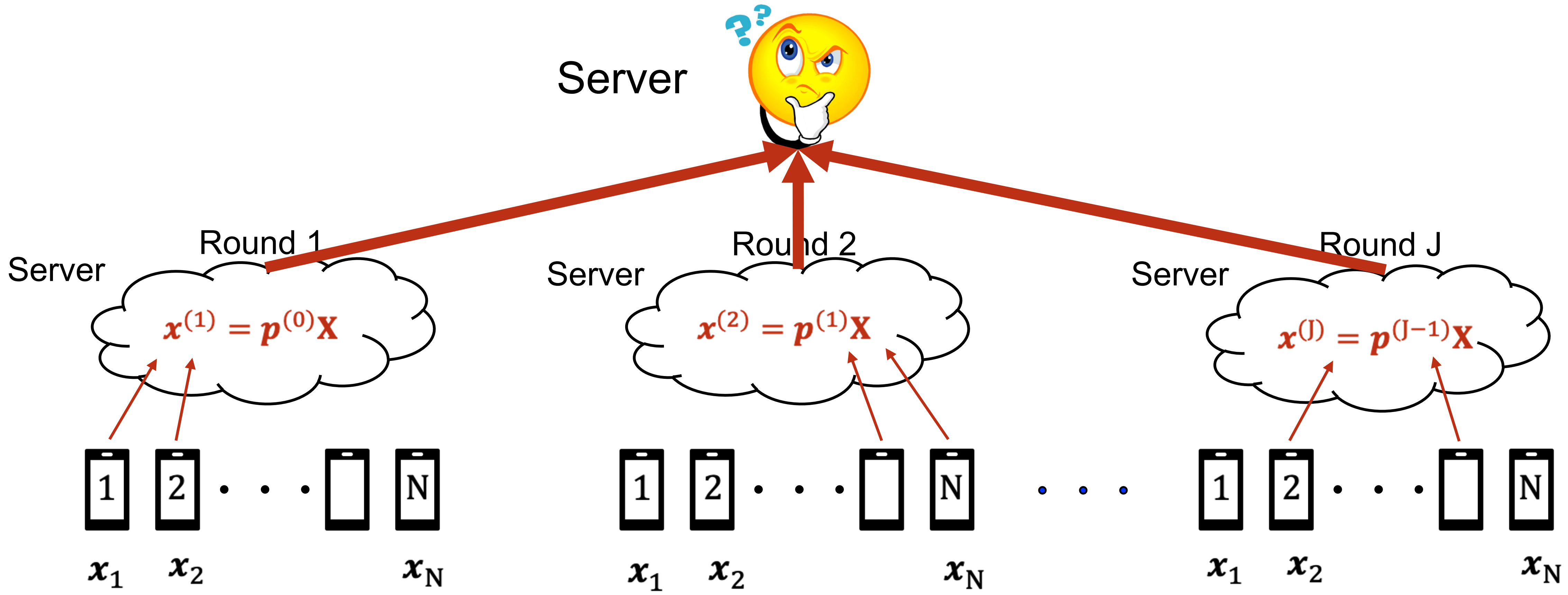
The server must not learn any aggregate model of less than T users.



Metric 1: Multi-round Privacy (T)

- A multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$a_1 \sum_{j\in\mathcal{S}_1} \boldsymbol{x}_j + a_2 \sum_{j\in\mathcal{S}_2} \boldsymbol{x}_j + \cdots + a_n \sum_{j\in\mathcal{S}_n} \boldsymbol{x}_j, \text{ where } |\mathcal{S}_i| \geq T.$$



Metric 1: Multi-round Privacy (T)

- A multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$a_1 \sum_{j \in \mathcal{S}_1} \boldsymbol{x}_j + a_2 \sum_{j \in \mathcal{S}_2} \boldsymbol{x}_j + \cdots + a_n \sum_{j \in \mathcal{S}_n} \boldsymbol{x}_j, \text{ where } |\mathcal{S}_i| \geq T.$$

- **Example (**T=2): the best the server can do is reconstructing x_i+x_j (for some *i* and *j*)

Metric 1: Multi-round Privacy (T)

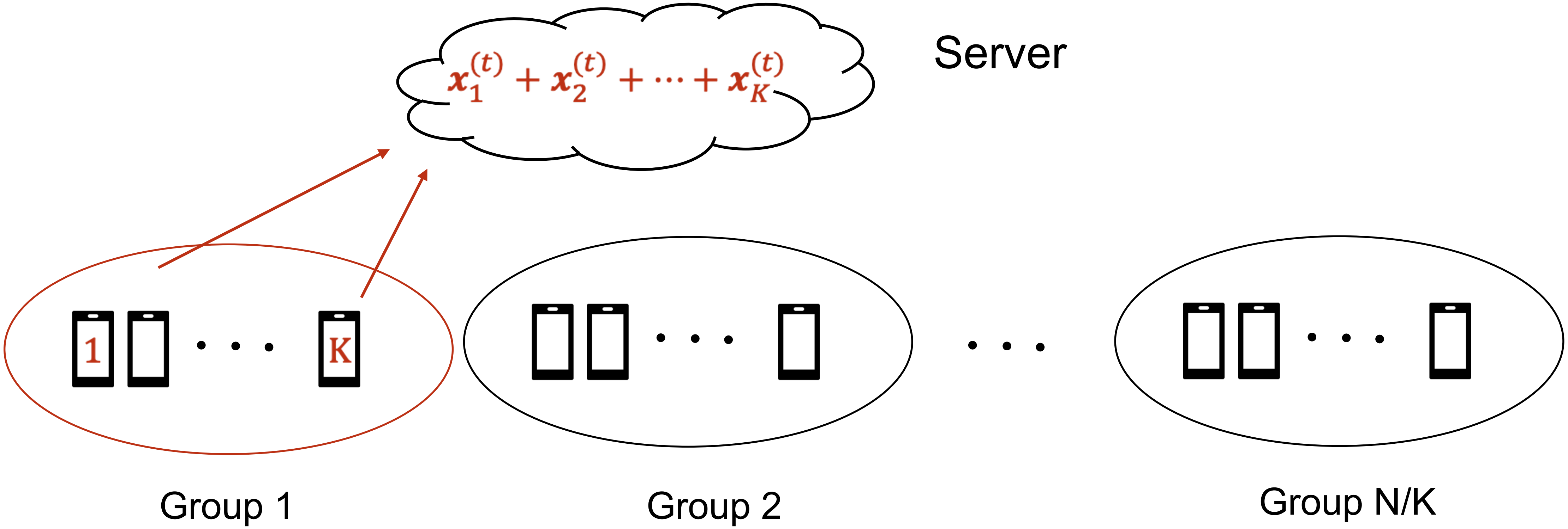
- A multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$a_1 \sum_{j \in \mathcal{S}_1} \boldsymbol{x}_j + a_2 \sum_{j \in \mathcal{S}_2} \boldsymbol{x}_j + \cdots + a_n \sum_{j \in \mathcal{S}_n} \boldsymbol{x}_j, \text{ where } |\mathcal{S}_i| \geq T.$$

- **Example (**T=2): the best the server can do is reconstructing x_i+x_j (for some *i* and *j*)
- Worst-case (strong) assumption 1: the model coefficients in each group are the same
- Worst-case (strong) assumption 2: user's model stays the same across different rounds

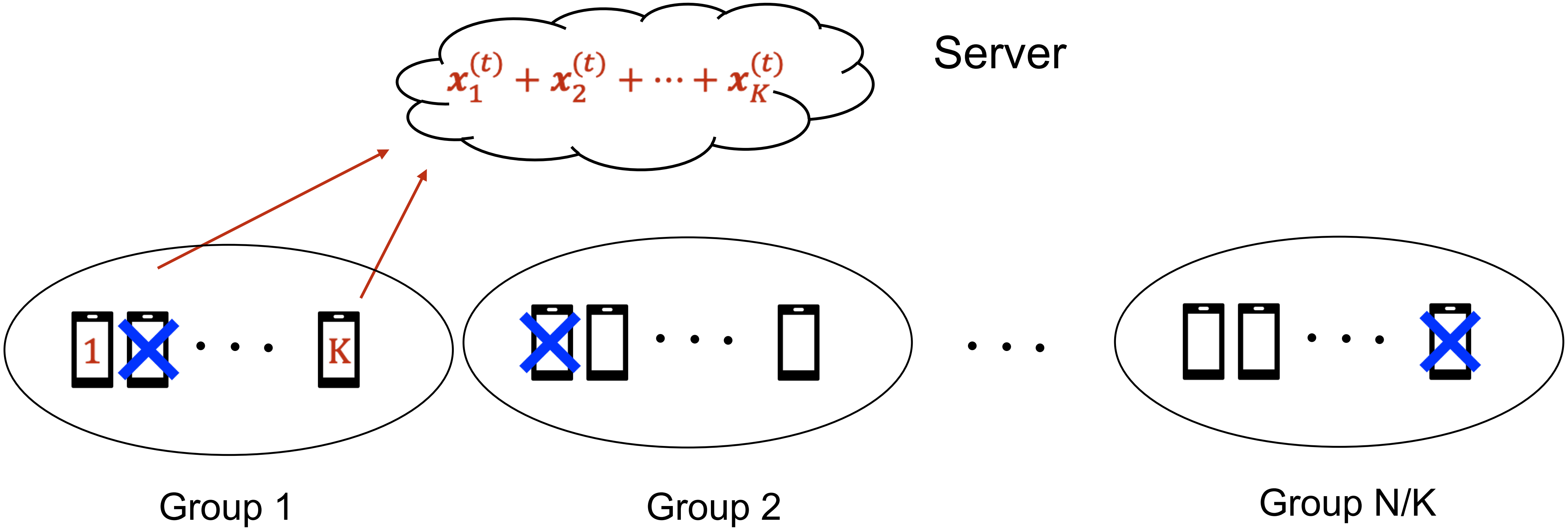
Baselines

1. User Partitioning
   - Large multi-round privacy T = **group size**



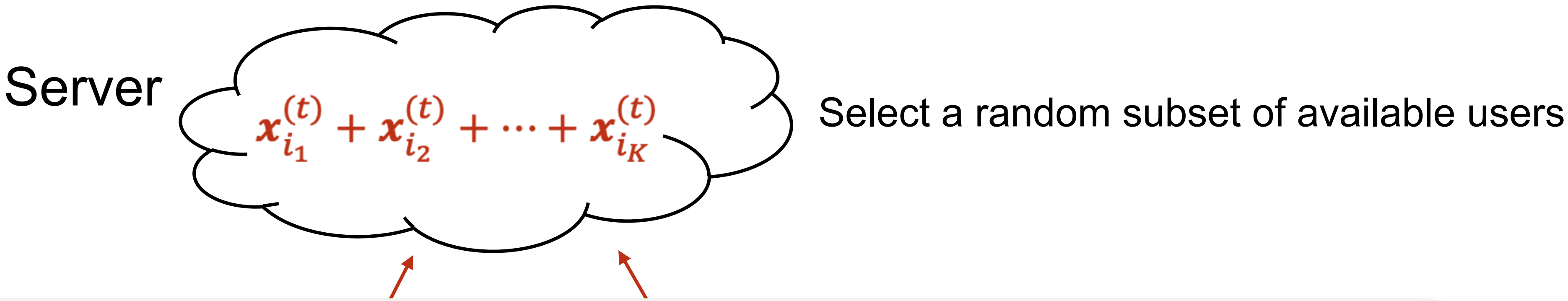
Baselines

1. User Partitioning
   - Large multi-round privacy $T = \textbf{group size}$
   - In many rounds, however, no groups are available



Baselines

2. Random Selection
   - Small multi-round privacy $T = 1$



Theorem: In Random Selection, the server can reconstruct all individual models of the N users after N rounds with probability at least

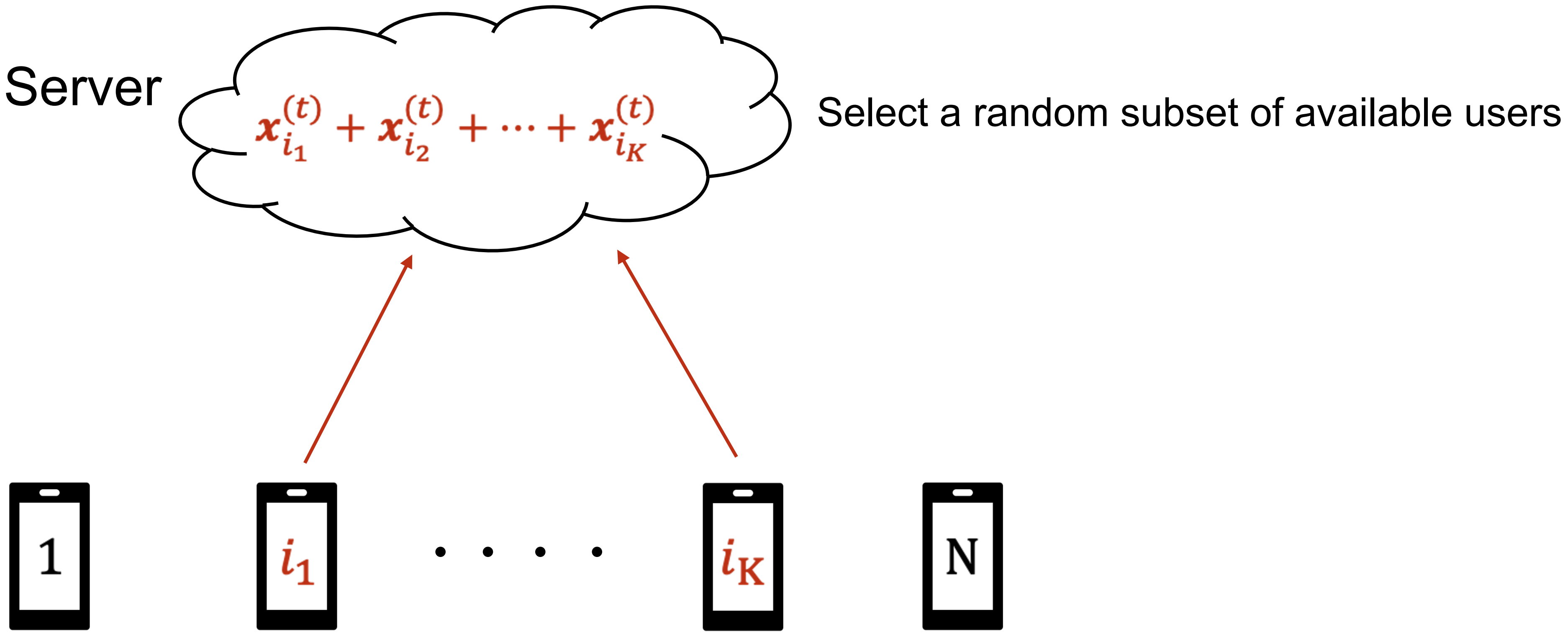
$$1-\exp(-cN),$$

where c is a constant.

Baselines

2. Random Selection

- Small multi-round privacy $T = 1$
- Any subset of available users can be selected in any round



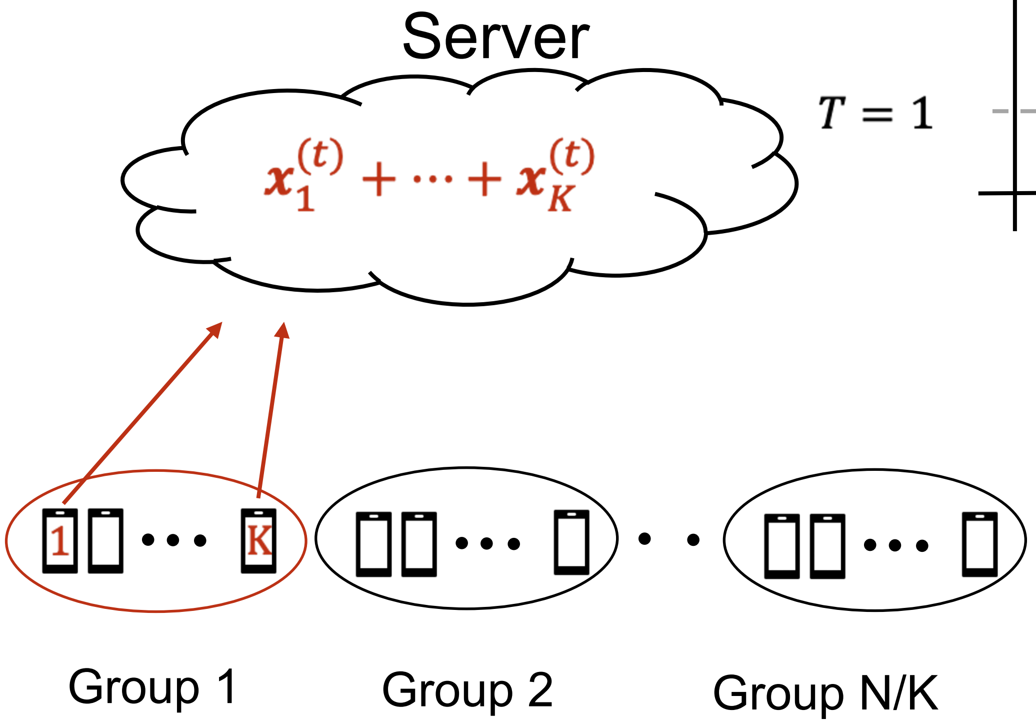
Metric 2: Average Aggregation Cardinality (C)

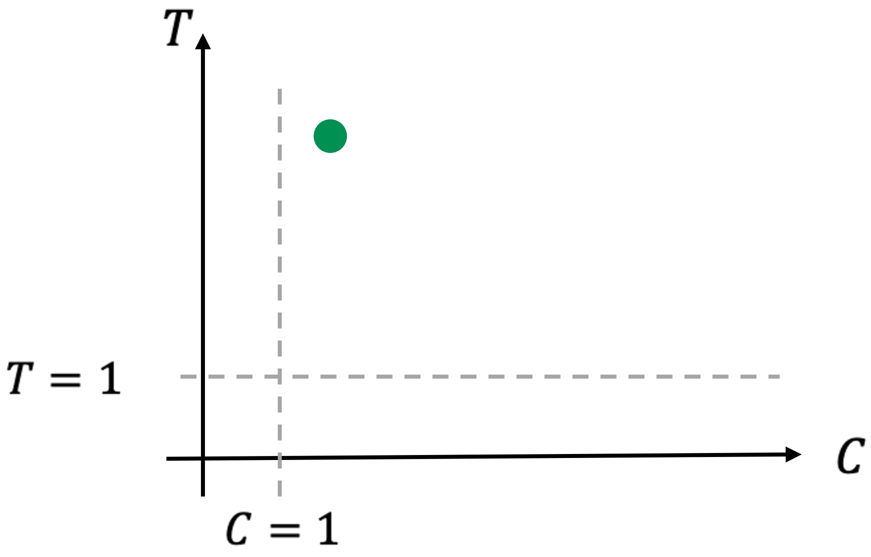
Aggregation Cardinality $\mathbf{C}$ = Average number of participating users over all rounds

Metric 2: Average Aggregation Cardinality (C)

Aggregation Cardinality **C** = Average number of participating users over all rounds

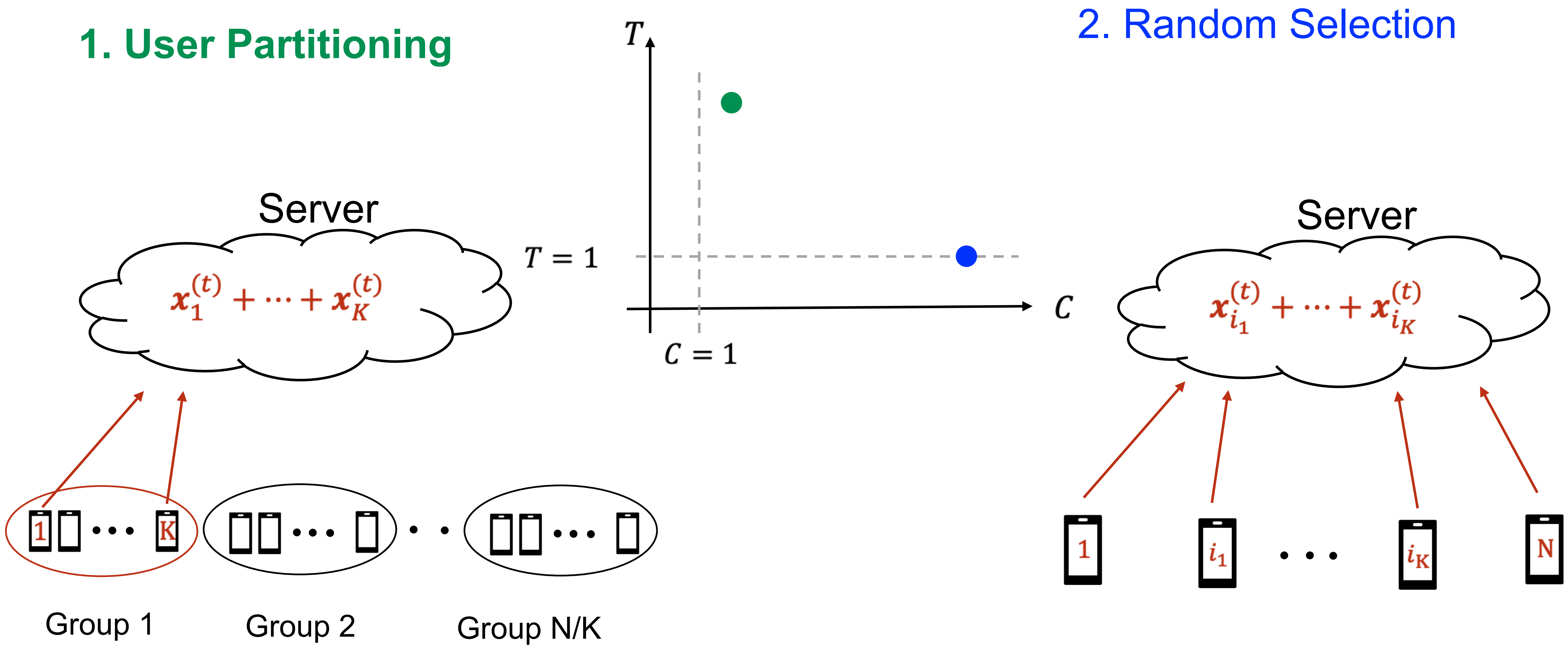
1. User Partitioning





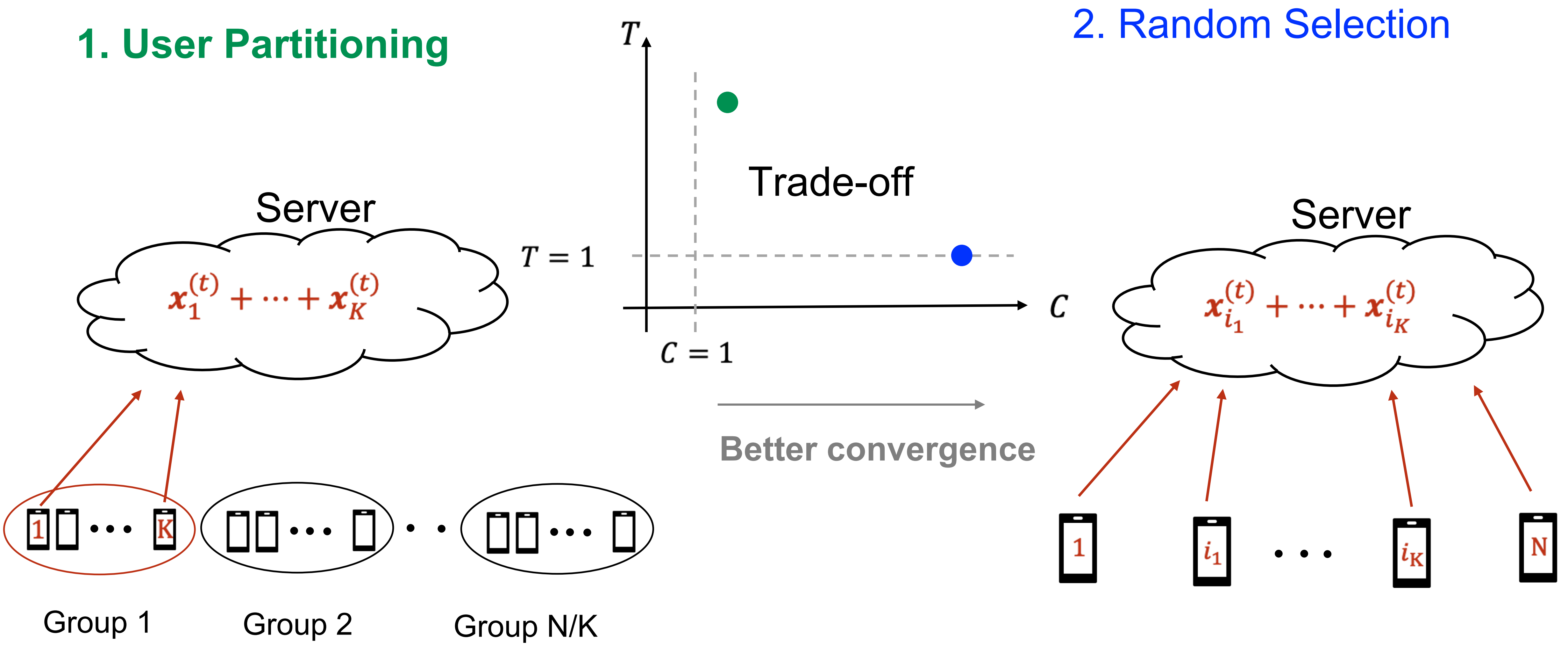
Metric 2: Average Aggregation Cardinality (C)

Aggregation Cardinality **C** = Average number of participating users over all rounds



Metric 2: Average Aggregation Cardinality (C)

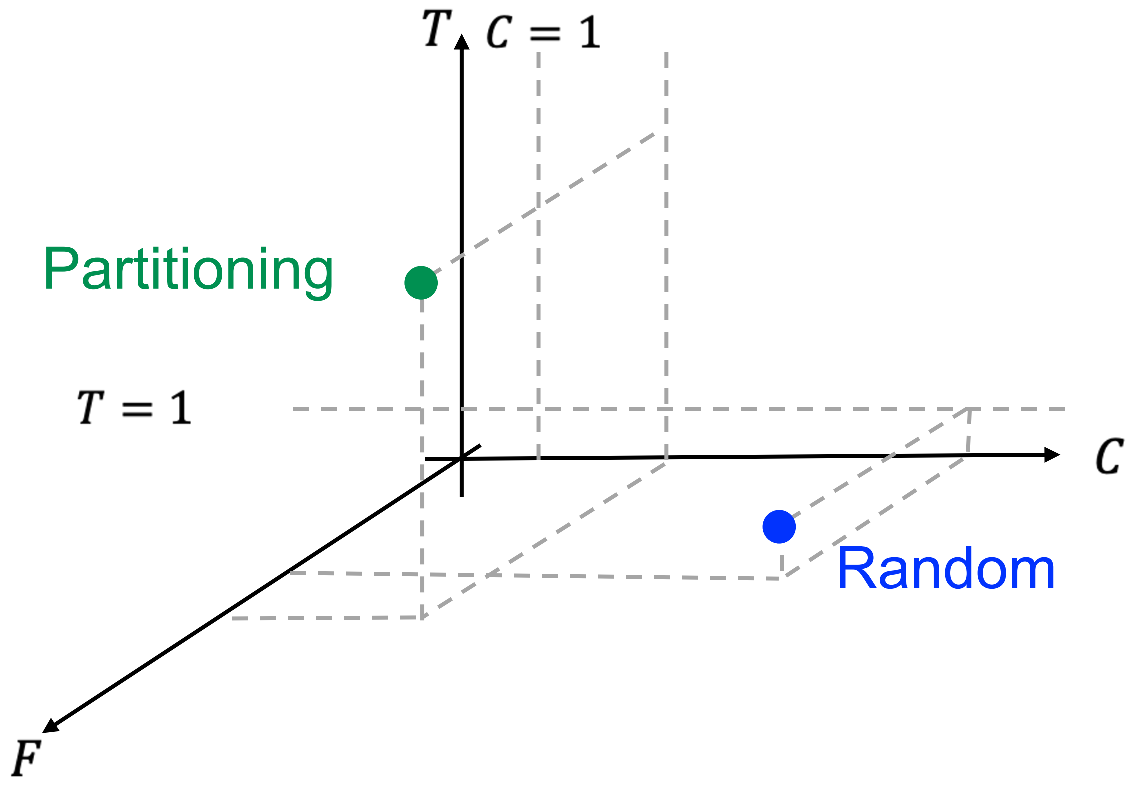
Aggregation Cardinality **C** = Average number of participating users over all rounds



Metric 3: Aggregation Fairness Gap (F)

- Aggregation Fairness Gap $\mathbf{F}$

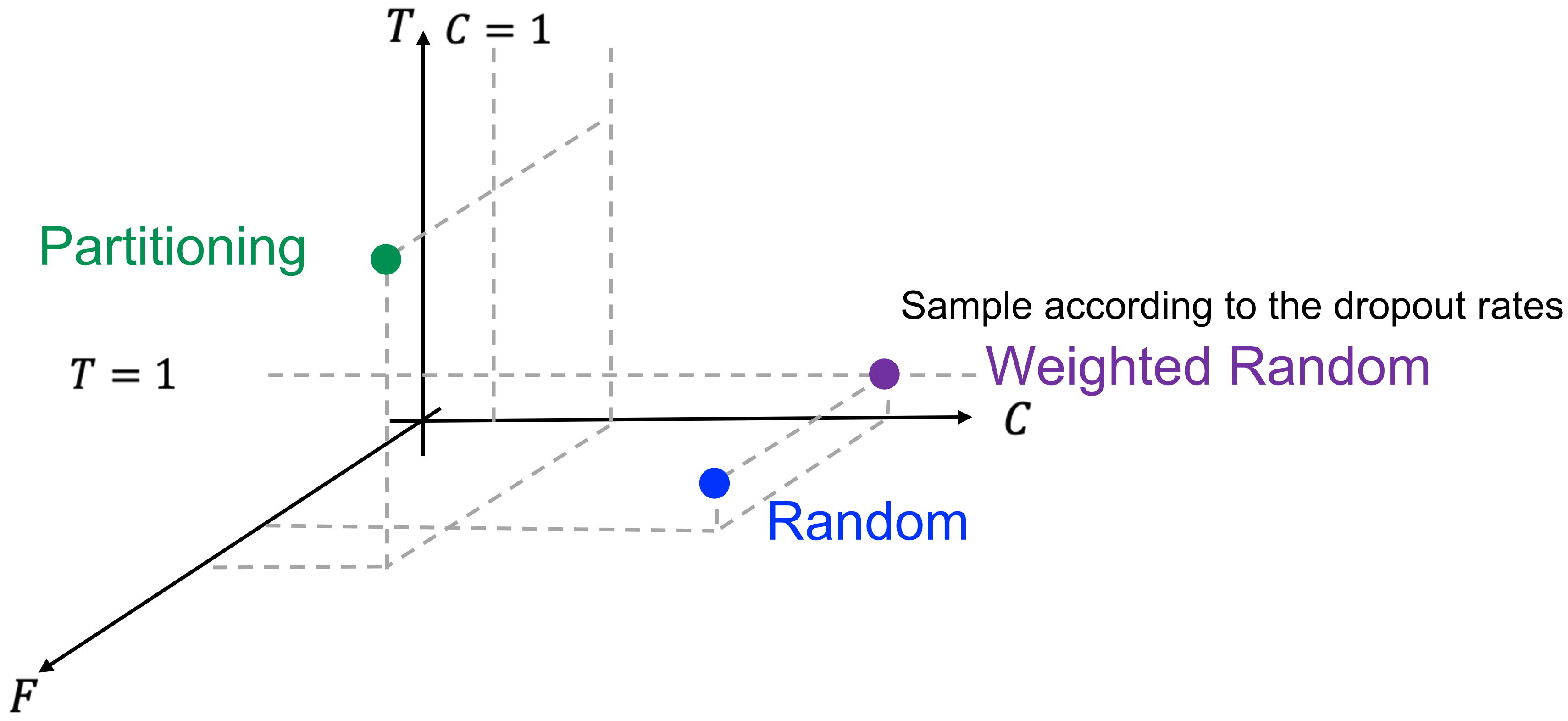
  $\mathbf{F} =$ max. average participation frequency − min. average participation frequency



Metric 3: Aggregation Fairness Gap (F)

- Aggregation Fairness Gap $\mathbf{F}$

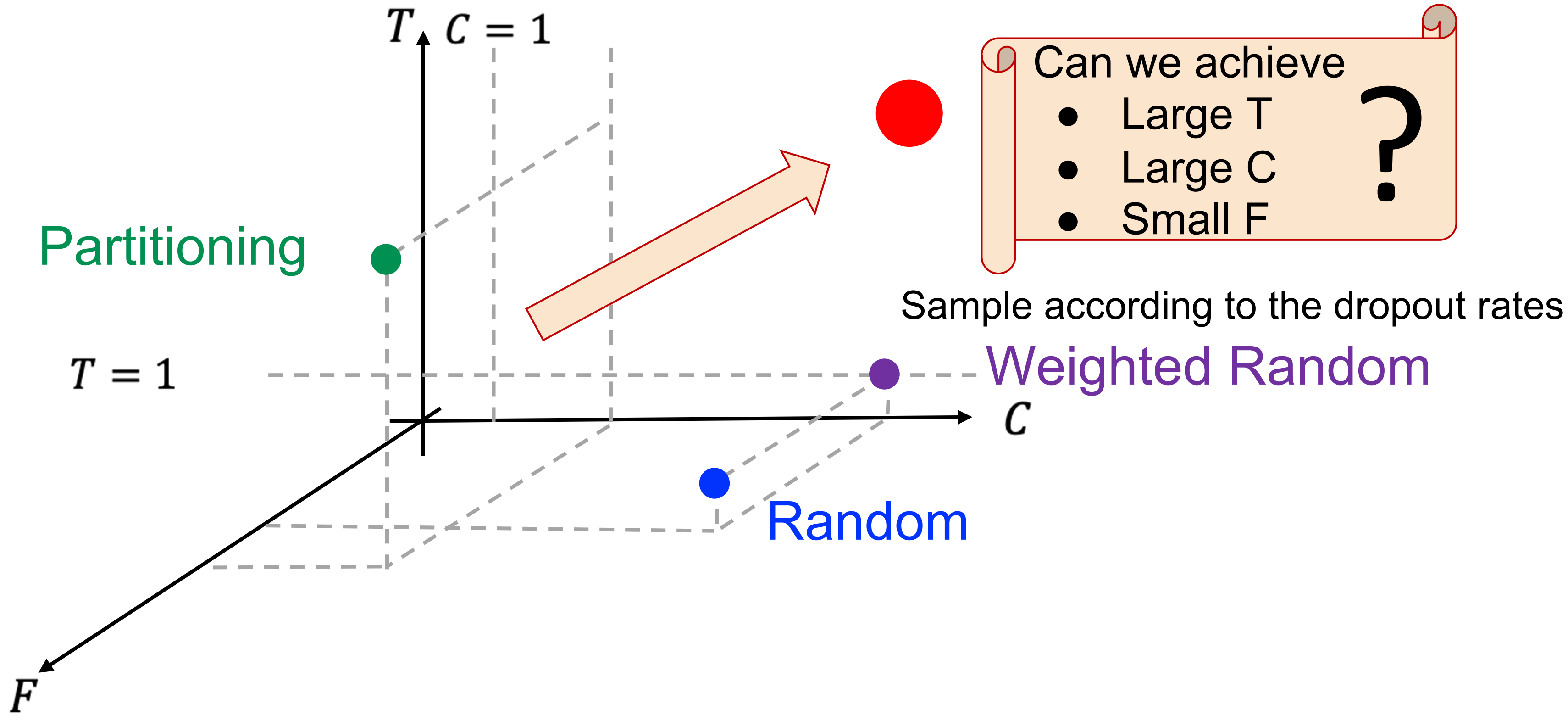
  $\mathbf{F} =$ max. average participation frequency − min. average participation frequency



Metric 3: Aggregation Fairness Gap (F)

- Aggregation Fairness Gap $\mathbf{F}$

  $\mathbf{F} =$ max. average participation frequency − min. average participation frequency



Proposed Approach: Multi-RoundSecAgg

1) Batch Partitioning

- Input: N, $\mathrm{K} \leq \mathrm{N}$, $1 \leq \mathrm{T} \leq \mathrm{K}$
- Output: A family of $\boldsymbol{K}$-user sets satisfying the multi-round privacy $\boldsymbol{T}$.
  This family is represented by a matrix $\boldsymbol{B}$.

2) Available batch selection to guarantee fairness

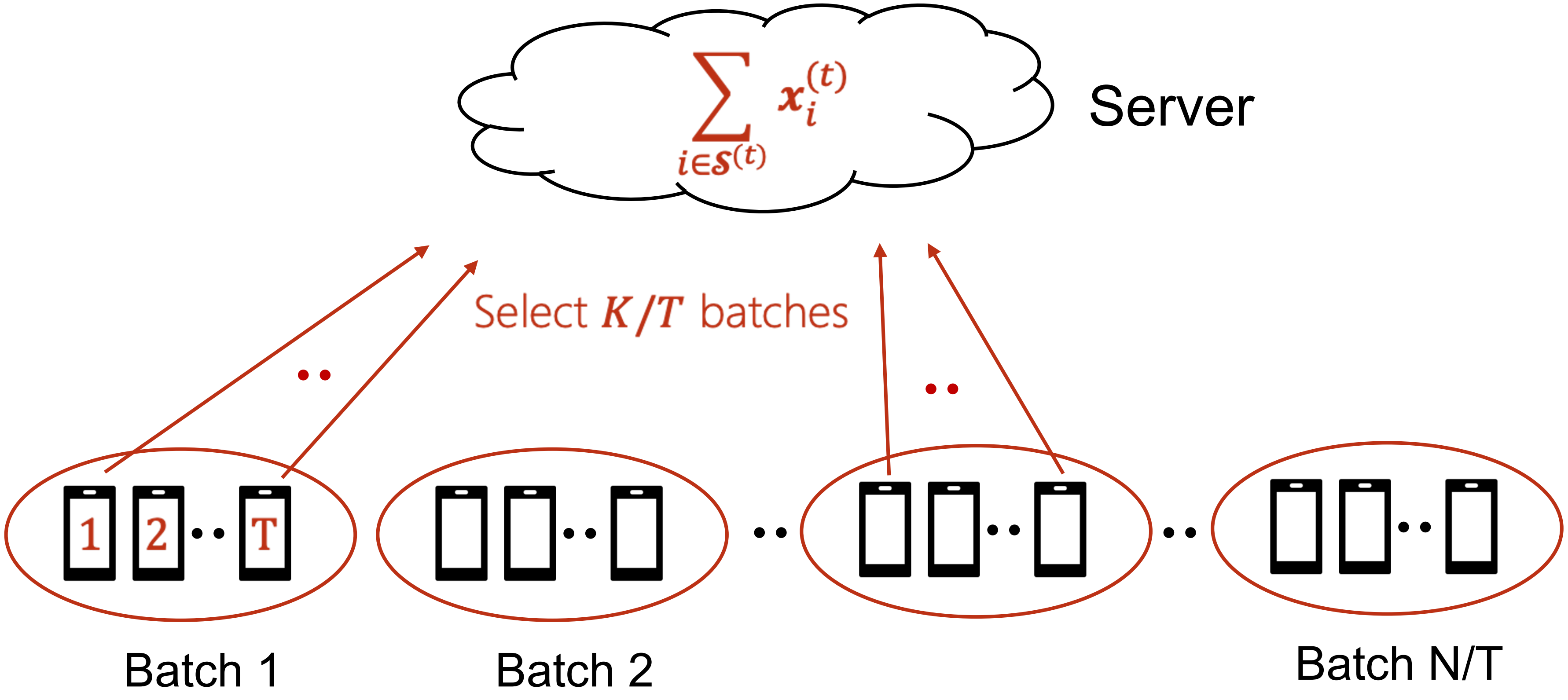
- Input: Set of available users at round $\boldsymbol{t}$ and $\boldsymbol{B}$.
- Output: Set of users that will participate at round $\boldsymbol{t}$.

Multi-RoundSecAgg

1) Batch Partitioning

Idea: Partition users into T-user batches; allow selection of any K/T available batches

This results in a family of $\mathrm{R_{BP}} = \binom{N/T}{K/T}$ sets.



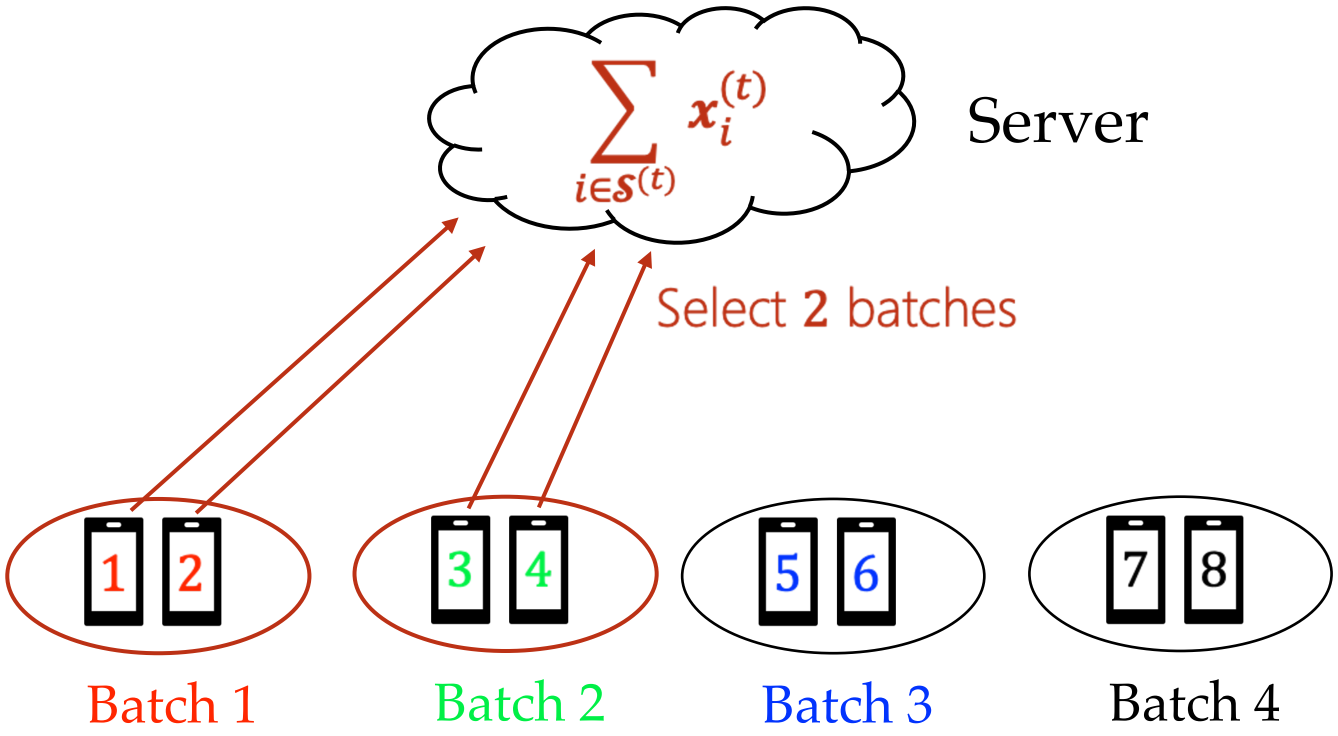
Multi-RoundSecAgg

1) Batch Partitioning

Example ($N = 8, K = 4$ & $T = 2$)

$$R_{BP} = \binom{N/T}{K/T} = 6 \text{ sets}$$

$$\begin{array}{c} \boldsymbol{B} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \\ \text{Batch 1} \quad \text{Batch 2} \quad \text{Batch 3} \quad \text{Batch 4} \end{array}$$



Multi-RoundSecAgg

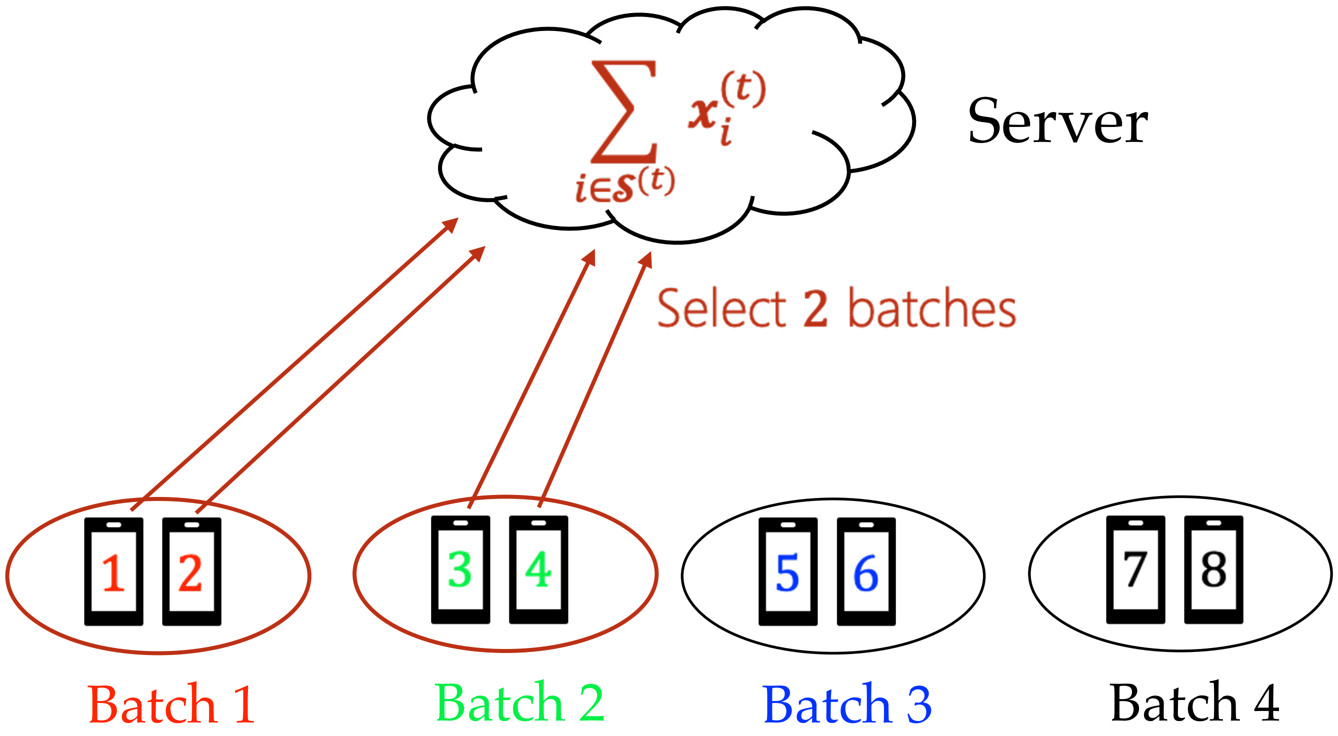
1) Batch Partitioning

Example $(N = 8, K = 4 \,\&\, T = 2)$

$$\mathrm{R_{BP}} = \binom{N/T}{K/T} = 6 \text{ sets}$$

$$\boldsymbol{B} = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array}\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

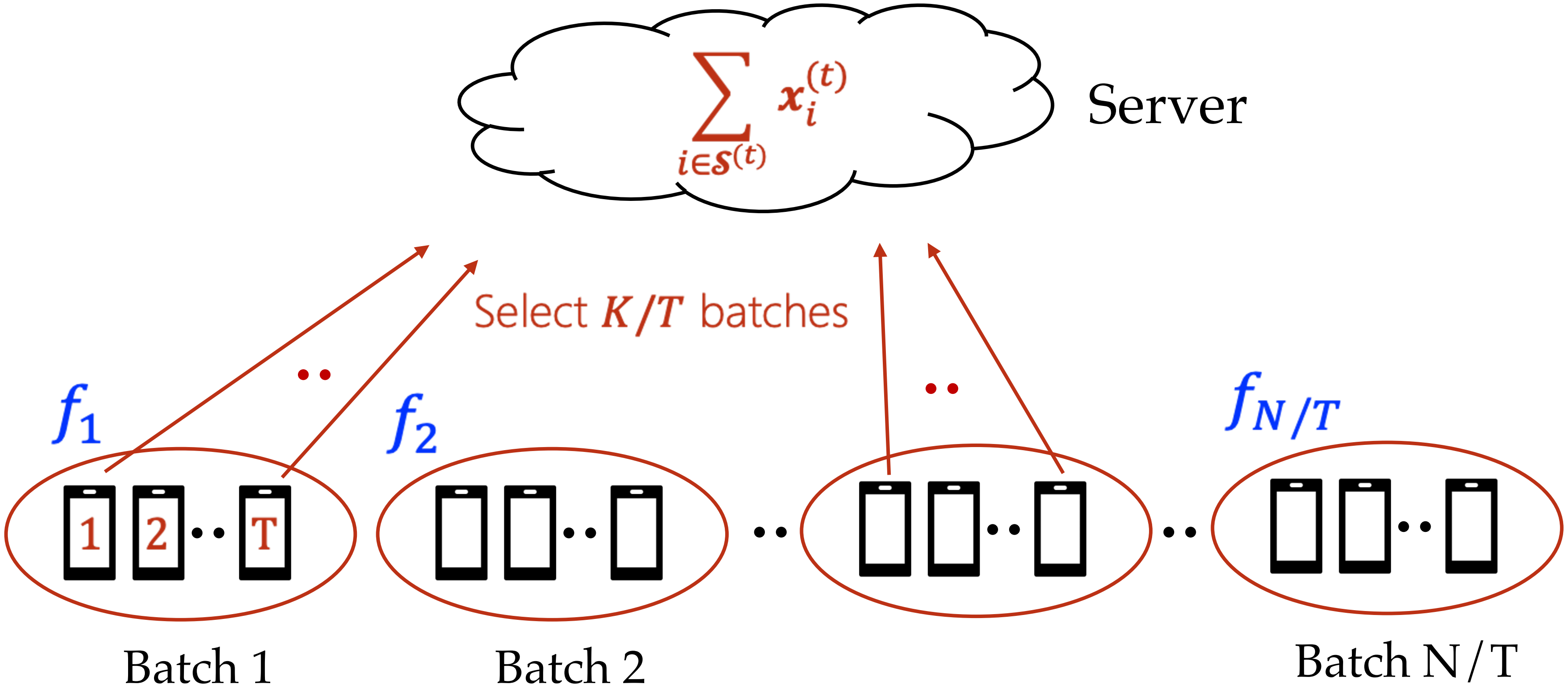
Batch 1 Batch 2 Batch 3 Batch 4



Theorem: aggregated models in the same batch can't be separated across different rounds even through non-linear mixtures of received aggregates.

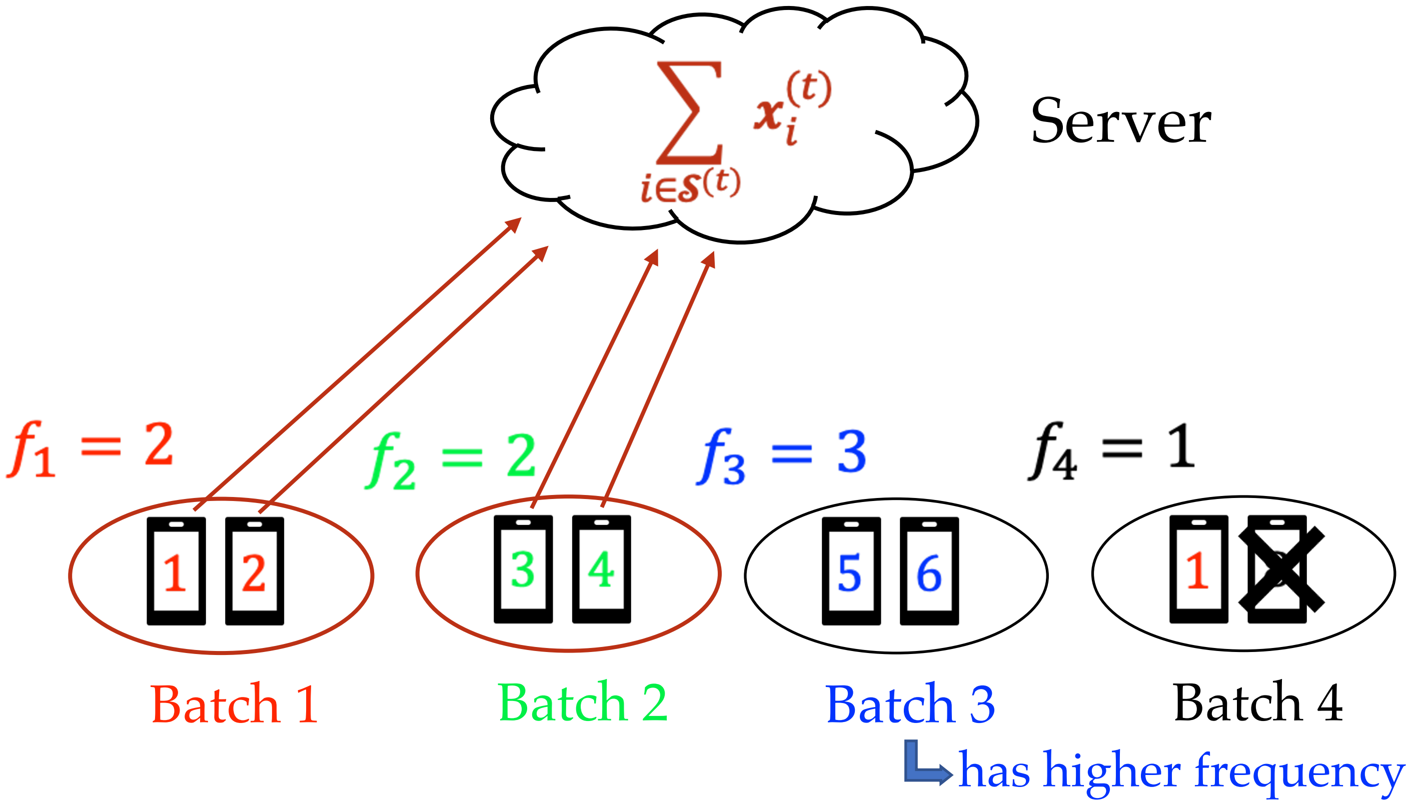
Multi-RoundSecAgg

2) Available batch selection to guarantee fairness
   - Input: Set of available users at round $\boldsymbol{t}$ and $\boldsymbol{B}$.
   - Output: Set of users that will participate at round $\boldsymbol{t}$.
   - Idea: Select based on the minimum frequency of participation.



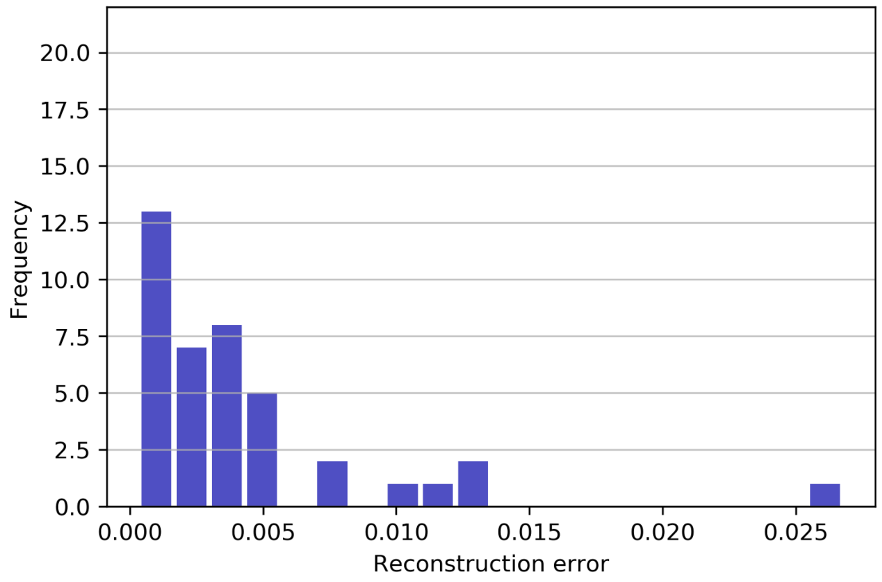
Multi-RoundSecAgg

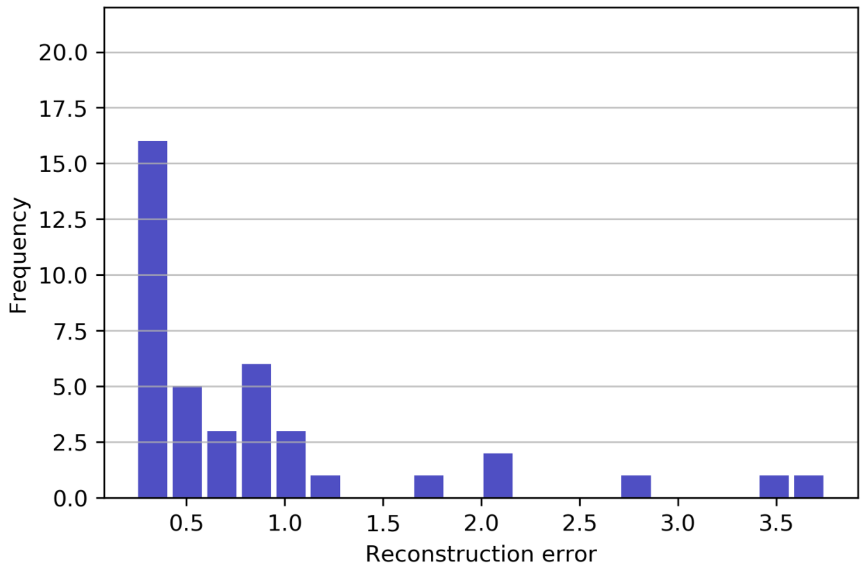
2) Available batch selection to guarantee fairness
   - Input: Set of available users at round $\boldsymbol{t}$ and $\boldsymbol{B}$.
   - Output: Set of users that will participate at round $\boldsymbol{t}$.
   - Idea: Select based on the minimum frequency of participation.



An Illustrative Experiment

- Experiment (N=40 users)
  - MNIST Dataset & Non-IID Setting.
  - K=8 users are selected at random at each round.
  - Dropout probability $p_i \sim \{0.1, 0.2, 0.3, 0.4, 0.5\}$
  - The server estimates the individual gradients through least squares.





- Random Selection
  Reconstruction Error < 0.005 for many users

- Multi-RoundSecAgg (T=2)
  Reconstruction Error > 0.25 for all users

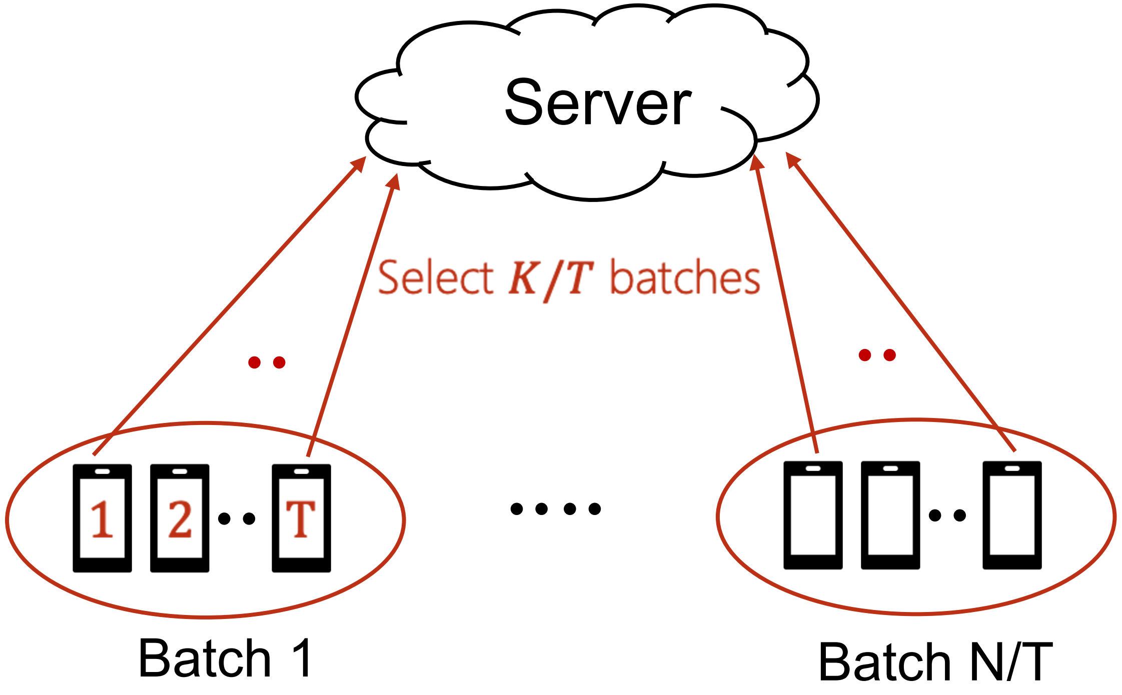
Multi-RoundSecAgg Theoretical Guarantees

Theorem: Multi-RoundSecAgg with parameters $\boldsymbol{N}, \boldsymbol{K} \leq \mathbf{N}$ and $\mathbf{1} \leq \boldsymbol{T} \leq \mathbf{K}$ ensures

1. a multi-round privacy $\mathbf{1} \leq \boldsymbol{T} \leq \mathbf{K}$,
2. an aggregation fairness gap $\boldsymbol{F} = \mathbf{0}$, and
3. an average aggregation cardinality $\boldsymbol{C}$

$$\mathrm{C(T)} = K\left(1 - \sum_{i=N/T-K/T+1}^{N/T} \binom{N/T}{i} q^{i}(1-q)^{N/T-i}\right),$$

$q = 1 - (1-p)^{T}$, p: dropout probability



Each batch is not available with probability q

Multi-RoundSecAgg Theoretical Guarantees

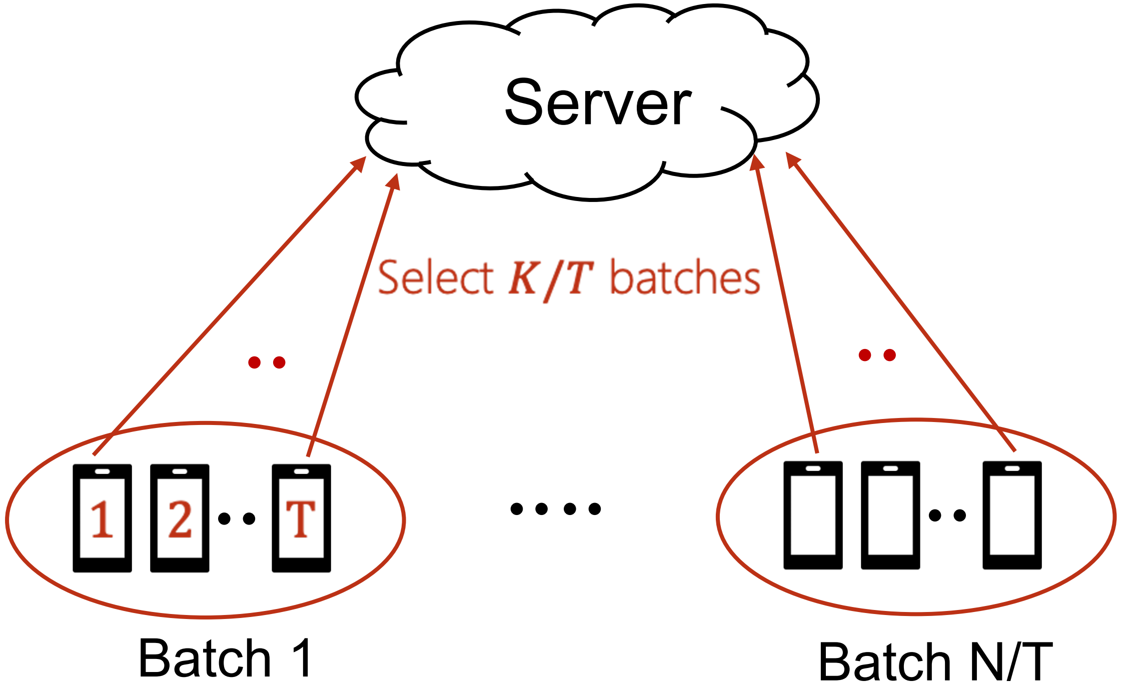
Theorem: Multi-RoundSecAgg with parameters $\boldsymbol{N}, \boldsymbol{K} \leq \mathbf{N}$ **and** $\mathbf{1} \leq \boldsymbol{T} \leq \mathbf{K}$ ensures

1. a multi-round privacy $\mathbf{1} \leq \boldsymbol{T} \leq \mathbf{K}$,
2. an aggregation fairness gap $\boldsymbol{F} = \mathbf{0}$, and
3. an average aggregation cardinality $\boldsymbol{C}$

$$\mathrm{C(T)} = K\left(1 - \sum_{i=N/T-K/T+1}^{N/T} \binom{N/T}{i} q^i (1-q)^{N/T-i}\right),$$

$q = 1 - (1-p)^T$, p: dropout probability

Example $(\mathrm{N} = 120, \mathrm{K} = \mathrm{N}/10 = 12, \mathrm{p} = 0.2)$
for $T = N/20 = 6$, we have C=11.77 $\approx N/10$



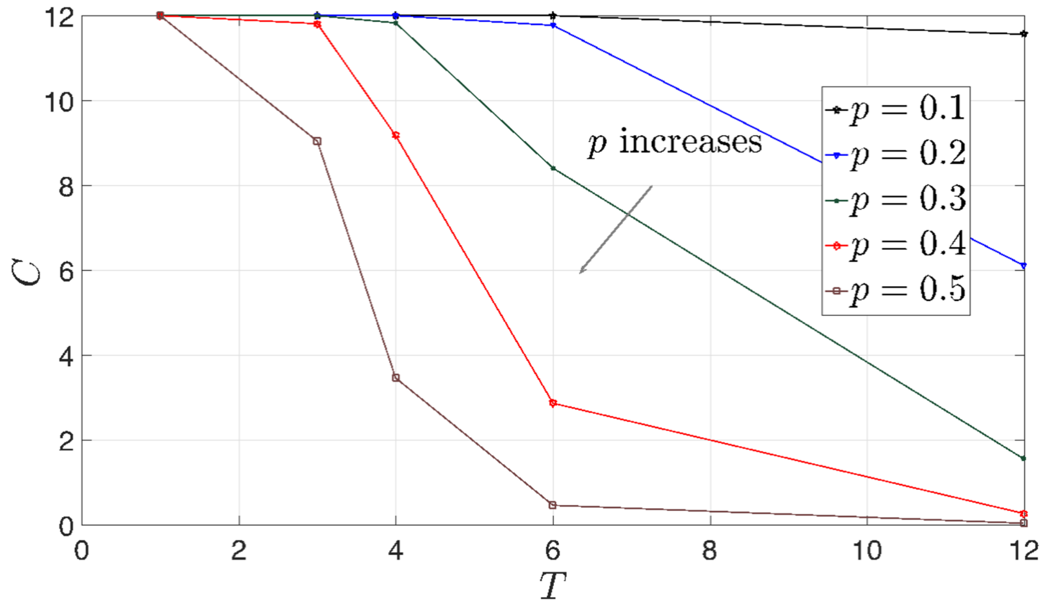
Each batch is not available with probability q

Multi-RoundSecAgg Theoretical Guarantees

Theorem: Multi-RoundSecAgg with parameters $N, K \leq \mathrm{N}$ and $1 \leq T \leq \mathrm{K}$ ensures

1. a multi-round privacy $1 \leq T \leq \mathrm{K}$,
2. an aggregation fairness gap $F = 0$, and
3. an average aggregation cardinality C

Better convergence



$N = 120$ and $K = 12$

Trade-off between "Multi-round Privacy Guarantee" & "Average Aggregation Cardinality"

Multi-RoundSecAgg Convergence Guarantees

Assumptions

1. The loss functions $L_1, L_2, \ldots, L_N$ are ρ-smooth.
2. The loss functions $L_1, L_2, \ldots, L_N$ are μ-strongly convex.
3. The variance of the stochastic gradients at user i is bounded by σ_i^2.
4. The expected squared norm of the stochastic gradients is uniformly bounded by G^2.

$$E\left[L\left(\boldsymbol{x}^{(J)}\right)\right] - L^* \leq \frac{\rho}{\gamma + \frac{C}{K}JE - 1}\left(\frac{2(\alpha+\beta)}{\mu^2} + \frac{\gamma}{2}E\left[\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^*\right\|\right]\right) ,$$

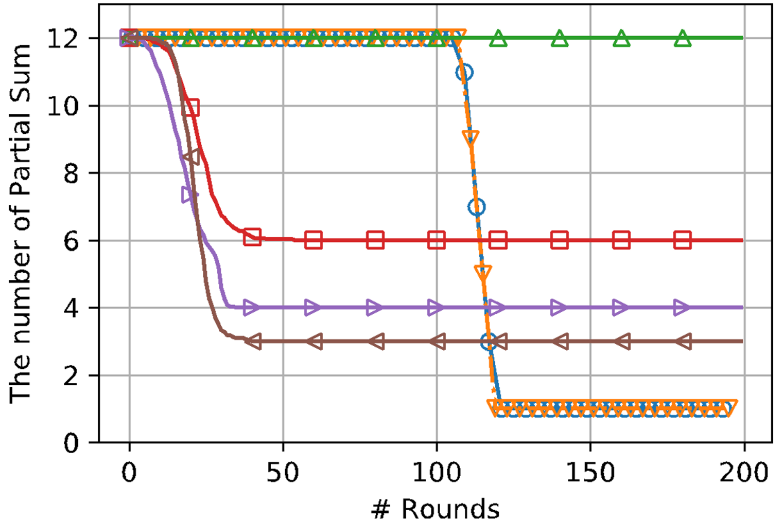
C controls the convergence rate

Experiments

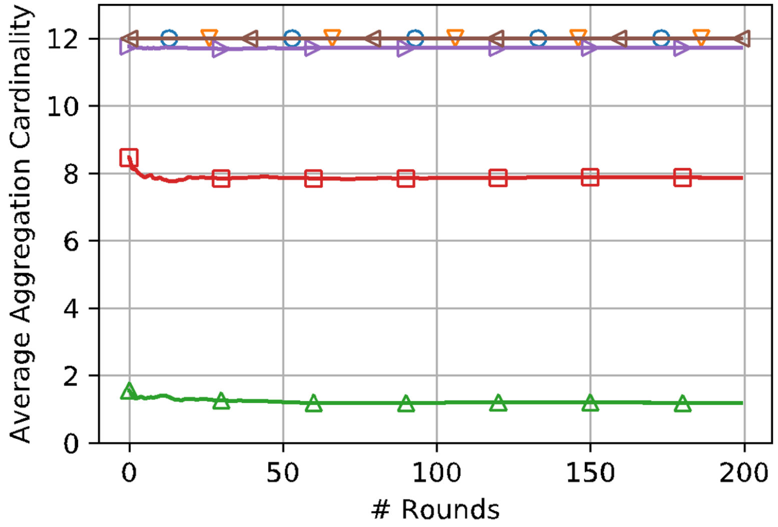
Setup

- $N = 120$ users, $K = 12$ users.
- Dataset: CIFAR-10
- Architecture: LeNet
- Dropout probability, $p_i \sim \{0.1, 0.2, 0.3, 0.4, 0.5\}$

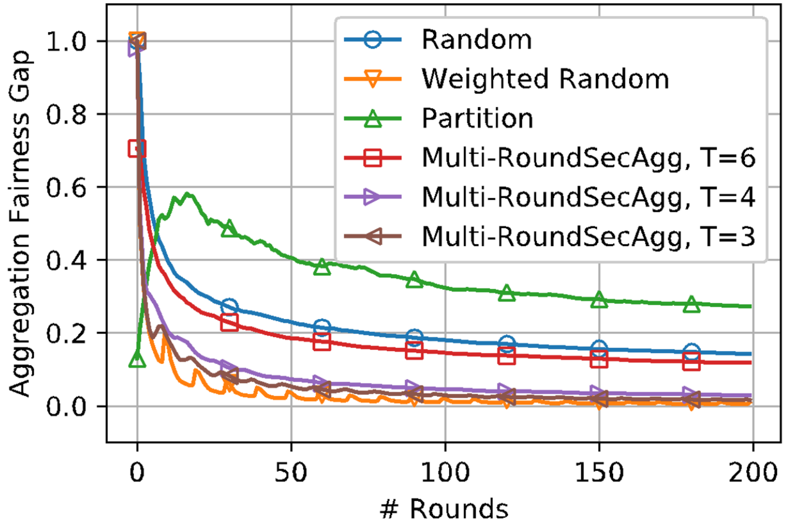
Key Metrics



Multi-round privacy guarantee (T)



Average aggregation cardinality (C)

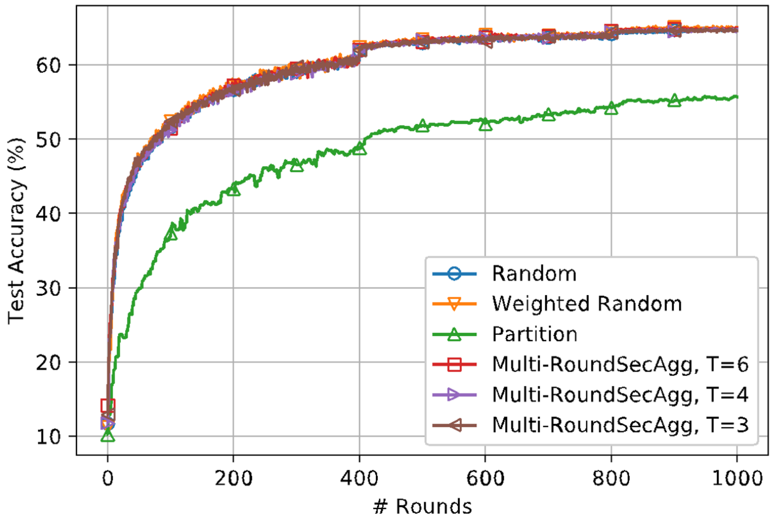


Aggregation fairness gap (F)

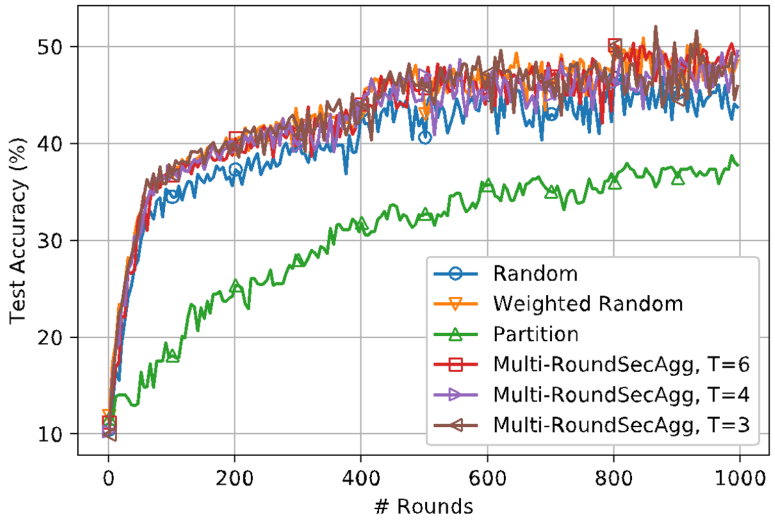
Experiments

Setup

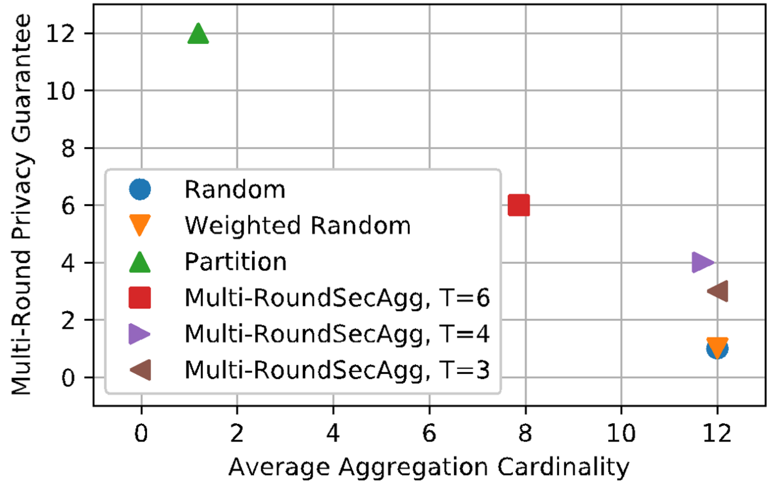
- $N = 120$ users, $K = 12$ users.
- Dataset: CIFAR-10
- Architecture: LeNet
- Dropout probability, $p_i \sim \{0.1, 0.2, 0.3, 0.4, 0.5\}$



(a) I.I.D Data Distribution



(b) Non I.I.D Data Distribution



(c) Trade-off between multi-round privacy guarantee & average aggregation cardinality

Concluding Remarks

- Random user selection in FL can lead to serious privacy leakage
- MultiRoundSecAgg is the first scheme for mitigating this challenge
- MultiRoundSecAgg reveals an interesting tradeoff between “privacy” and “convergence rate” in FL

Concluding Remarks

- Random user selection in FL can lead to serious privacy leakage
- MultiRoundSecAgg is the first scheme for mitigating this challenge
- MultiRoundSecAgg reveals an interesting tradeoff between “privacy” and “convergence rate” in FL
- Potential future directions
  - Our metric for multi-round privacy is very strong. Careful relaxations may lead to substantial improvements (e.g., in aggregation cardinality)
  - Formalizing a fundamental trade-off between privacy and convergence-rate in FL?
  - Our protocol guarantees that only an aggregate of models can be learned. How to bound privacy leakage from aggregate models?

$$\boldsymbol{x}_1^{(t)} + \boldsymbol{x}_2^{(t)} + \cdots + \boldsymbol{x}_N^{(t)}$$

Questions?

Thank you

Additional Slides

Optimality of Multi-RoundSecAgg

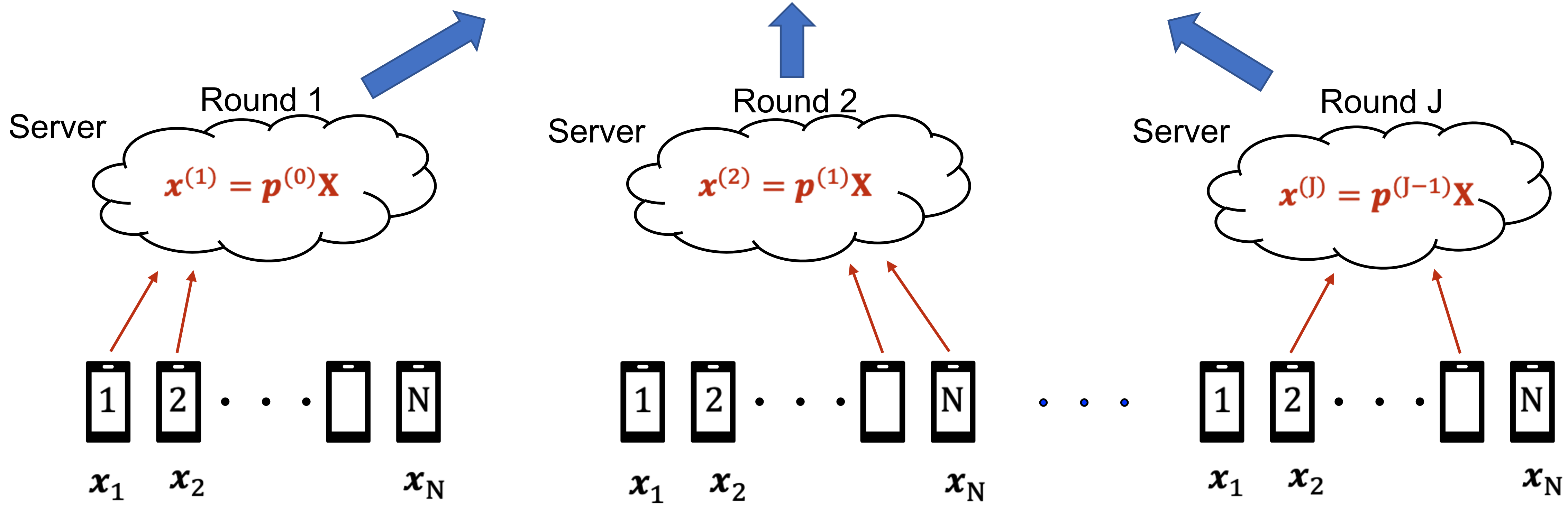
- Any strategy satisfying the multi-round privacy guarantee must have a batch partitioning structure.
- For given $N, K \leq \mathbf{N/2}$ **and** T, any strategy satisfying a multi-round privacy $\mathbf{T}$ can have at most $R_{\mathbf{max}}$ user sets

$$R_{\max} \leq \binom{N/T}{K/T} = R_{\mathrm{BP}}$$ number of sets in BP

Our Multi-round Privacy is Strong

- A multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$a_1 \sum_{j\in\mathcal{S}_1} \boldsymbol{x}_j + a_2 \sum_{j\in\mathcal{S}_2} \boldsymbol{x}_j + \cdots + a_n \sum_{j\in\mathcal{S}_n} \boldsymbol{x}_j, \text{ where } |\mathcal{S}_i| \geq T.$$

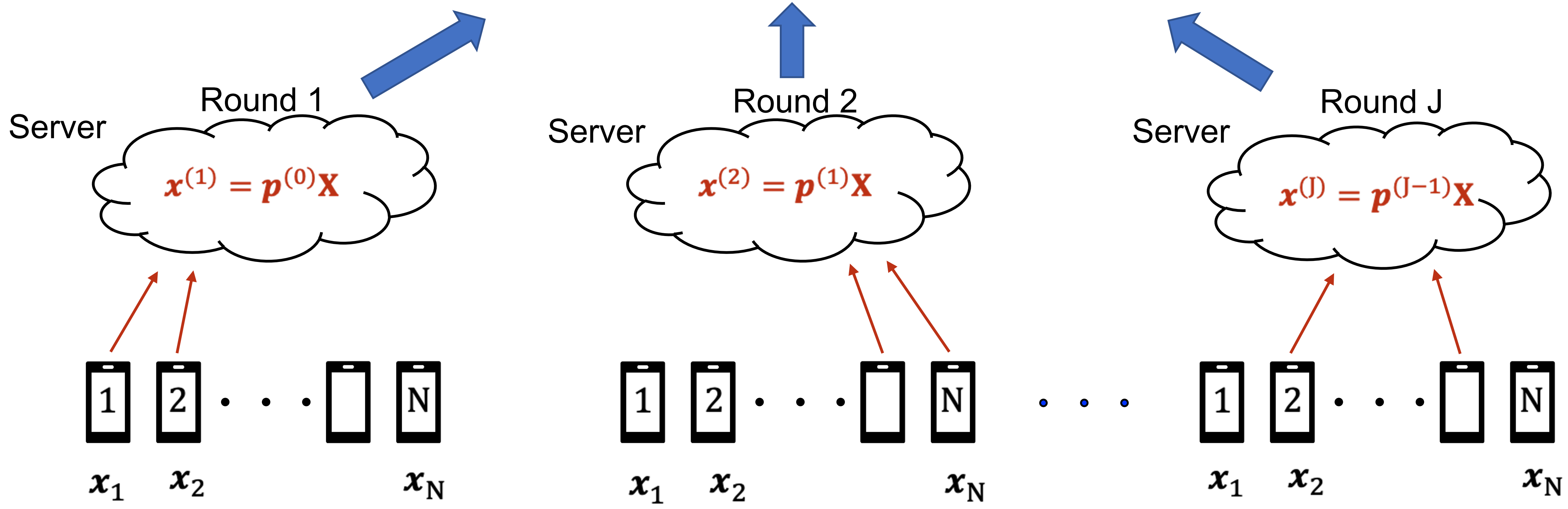


Each group has the same coefficient.

Relaxed Multi-round Privacy

- The relaxed multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

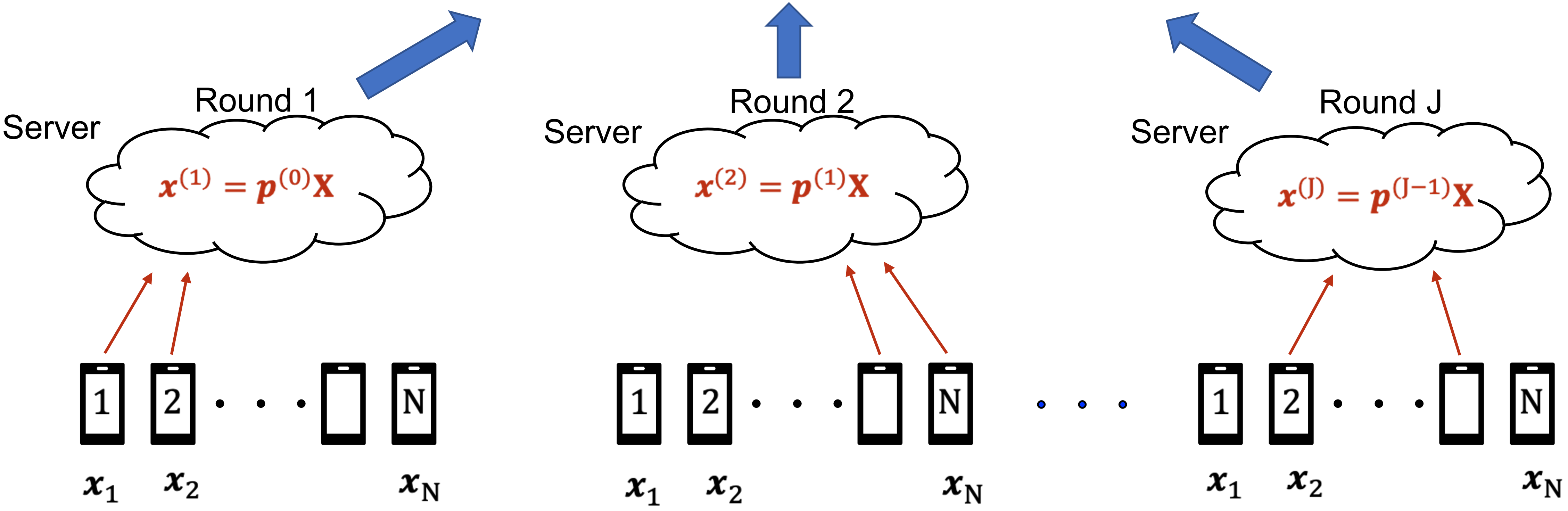
$$\sum_{j\in\mathcal{S}} a_j \boldsymbol{x}_j, \text{ where } |\mathcal{S}| \geq T$$



Relaxed Multi-round Privacy

- The relaxed multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$\sum_{j\in\mathcal{S}} a_j \boldsymbol{x}_j, \text{ where } |\mathcal{S}| \geq T$$

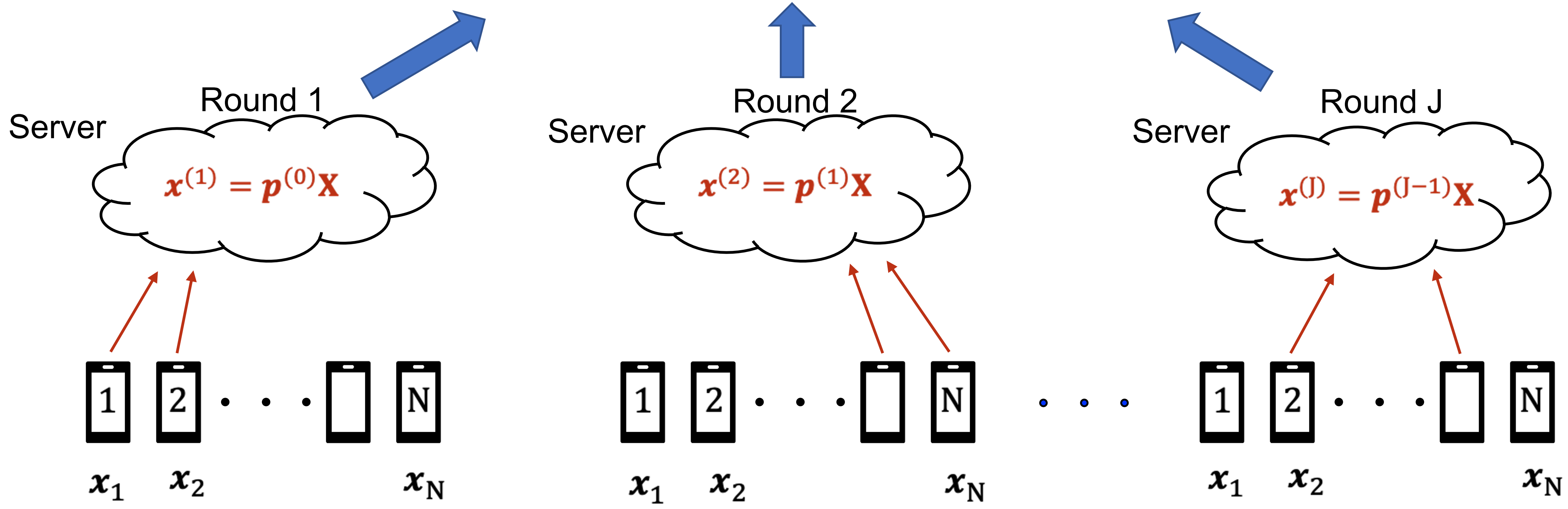


When $T = 2$, this allows reconstructing $a_i\, \boldsymbol{x}_i + a_j\, \boldsymbol{x}_j$ which can reveal $\boldsymbol{x}_i$ if $a_i \gg a_j$.

Relaxed Multi-round Privacy

- The relaxed multi-round privacy T requires that any non-zero partial sum that the server can reconstruct to be of the form

$$\sum_{j \in \mathcal{S}} a_j \boldsymbol{x}_j, \text{ where } |\mathcal{S}| \geq T$$

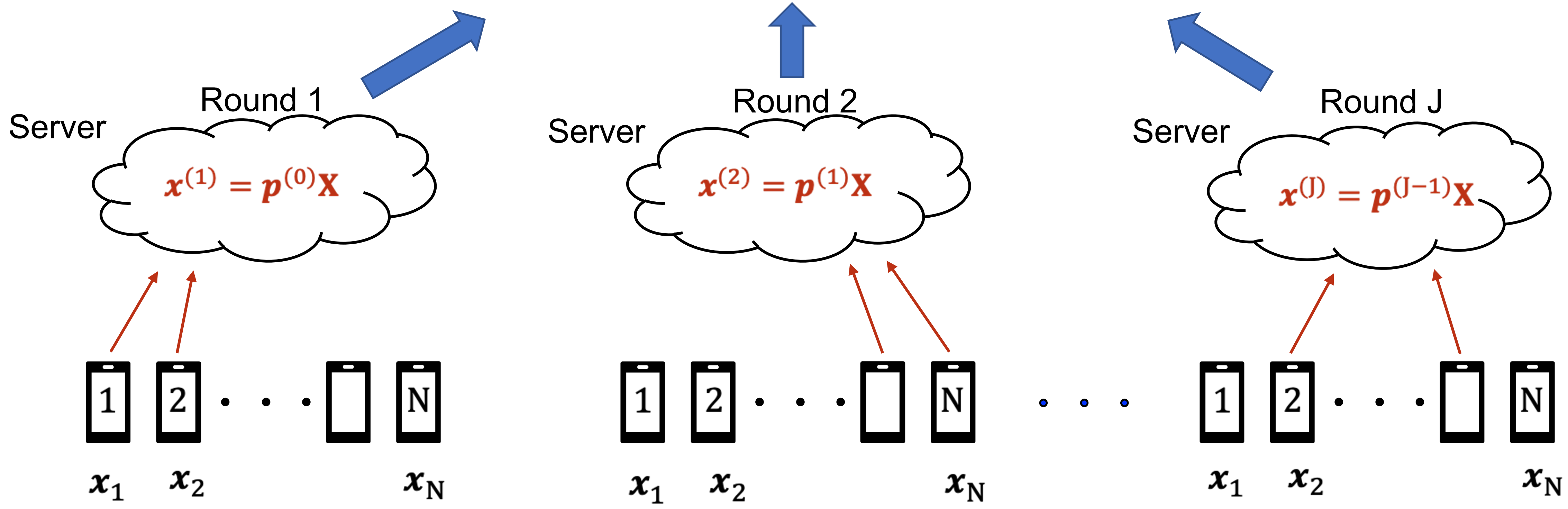


Batch partitioning is not necessary

Relaxed Multi-round Privacy

- The relaxed multi-round privacy T requires that any non-zero partial sum that the server can reconstruct to be of the form

$$\sum_{j\in\mathcal{S}} a_j \boldsymbol{x}_j, \text{ where } |\mathcal{S}| \geq T$$



What's the optimal strategy?

Relaxed Multi-round Privacy

- The relaxed multi-round privacy $\boldsymbol{T}$ requires that any non-zero partial sum that the server can reconstruct to be of the form

$$\sum_{j\in\mathcal{S}} \boldsymbol{a}_j \boldsymbol{x}_j, \text{ where } |\mathcal{S}| \geq T$$

Example (N=6, K=4, T=2)

$$\boldsymbol{B}_{old} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \xrightarrow{2\times} \boldsymbol{B}_{new} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Aggregation Fairness Gap F & Average Aggregation Cardinality C

Aggregation Fairness Gap F

$$\mathrm{F} = \underbrace{\max_{i\in[N]} \limsup_{J\to\infty} \frac{1}{J} \mathrm{E}\left[\sum_{t=0}^{J-1} \{\boldsymbol{p}^{(t)}\}_i\right]}_{\text{maximum average participation frequency}} - \underbrace{\min_{i\in[N]} \liminf_{J\to\infty} \frac{1}{J} \mathrm{E}\left[\sum_{t=0}^{J-1} \{\boldsymbol{p}^{(t)}\}_i\right]}_{\text{minimum average participation frequency}}$$

Average Aggregation Cardinality C (Average number of participating users)

$$\mathrm{C} = \liminf_{J\to\infty} \frac{1}{J} \mathrm{E}\left[\sum_{t=0}^{J-1} \left\|\boldsymbol{p}^{(t)}\right\|_0\right]$$

Multi-RoundSecAgg Convergence Guarantees

Assumptions

1. The loss functions $L_1, L_2, \dots, L_N$ are ρ-smooth.
2. The loss functions $L_1, L_2, \dots, L_N$ are μ-strongly convex.
3. The variance of the stochastic gradients at user i is bounded by σ_i^2.
4. The expected squared norm of the stochastic gradients is uniformly bounded by G^2.

$$E\left[L\left(\boldsymbol{x}^{(J)}\right)\right] - L^* \leq \frac{\rho}{\gamma + \frac{C}{K} JE - 1} \left(\frac{2(\alpha+\beta)}{\mu^2} + \frac{\gamma}{2} E\left[\left\|\boldsymbol{x}^{(0)} - \boldsymbol{x}^*\right\|\right]\right),$$

where E is the number of local SGD steps, $\alpha = \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 + 6\,\rho\Gamma + 8(E-1)^2 G^2$, $\beta = \frac{4(N-K)E^2 G^2}{K(N-1)}, \Gamma = L^* - \sum_{i=1}^{N} L_i$ and $\gamma = \max\{\frac{8\rho}{\mu}, E\}$

Necessity of Batch Partitioning (BP)

- Any strategy satisfying the multi-round privacy guarantee must have a batch partitioning structure.

Necessity of Batch Partitioning (BP)

- Any strategy satisfying the multi-round privacy guarantee must have a batch partitioning structure.

  Proof Idea

  Consider any scheme with a matrix $\boldsymbol{V}_{R\times N}$ and denote the linear combination used by the server by $\boldsymbol{z}_{1\times R}$, $z_i \sim U[0,1], i \in [R]$.

  For the scheme to have privacy $\boldsymbol{T}$, at least $\boldsymbol{T}$ elements of $\boldsymbol{zV}$ have to be equal.

Necessity of Batch Partitioning (BP)

- Any strategy satisfying the multi-round privacy guarantee must have a batch partitioning structure.

  Proof Idea

  Consider any scheme with a matrix $\boldsymbol{V}_{R \times N}$ and denote the linear combination used by the server by $\boldsymbol{z}_{1 \times R}$, $z_i \sim U[0,1], i \in [R]$.

  For the scheme to have privacy $\boldsymbol{T}$, at least $\boldsymbol{T}$ elements of $\boldsymbol{zV}$ have to be equal.

  This implies that at least T columns of $\boldsymbol{V}$ are equal as $z_i \sim U[0,1]$.

  ➡ Batch partitioning is necessary

Necessity of Batch Partitioning (BP)

- Any strategy satisfying the multi-round privacy guarantee must have a batch partitioning structure.

  Proof Idea

  Consider any scheme with a matrix $\boldsymbol{V}_{R\times N}$ and denote the linear combination used by the server by $\boldsymbol{z}_{1\times R}$, $z_i \sim U[0,1], i \in [R]$.

  For the scheme to have privacy $\boldsymbol{T}$, at least $\boldsymbol{T}$ elements of $\boldsymbol{zV}$ have to be equal.

  This implies that at least T columns of $\boldsymbol{V}$ are equal as $\boldsymbol{z_i} \sim \boldsymbol{U[0,1]}$.

  ➡ Batch partitioning is necessary

- For given $\boldsymbol{N, K \leq \mathrm{N/2} \text{ and } T}$, any strategy satisfying a multi-round privacy $\mathbf{T}$ can have at most $\boldsymbol{R_{\max}}$ user sets

$$\boldsymbol{R_{\max} \leq \binom{N/T}{K/T} = R_{\mathrm{BP}}}$$ number of sets in BP

References

[1] H Brendan McMahan et al.. “**Communication-efficient learning of deep networks from decentralized data**.” In Int. Conf. on Artificial Int. and Stat. (AISTATS), pages 1273–1282, 2017.

[2] Balázs Pejó, and Gergely Biczók. "**Quality Inference in Federated Learning with Secure Aggregation.**” arXiv:2007.06236 (2020).

[3] M. Lam et al. "**Gradient Disaggregation: Breaking Privacy in Federated Learning by Reconstructing the User Participant Matrix**." arXiv:2106.06089 (2021).