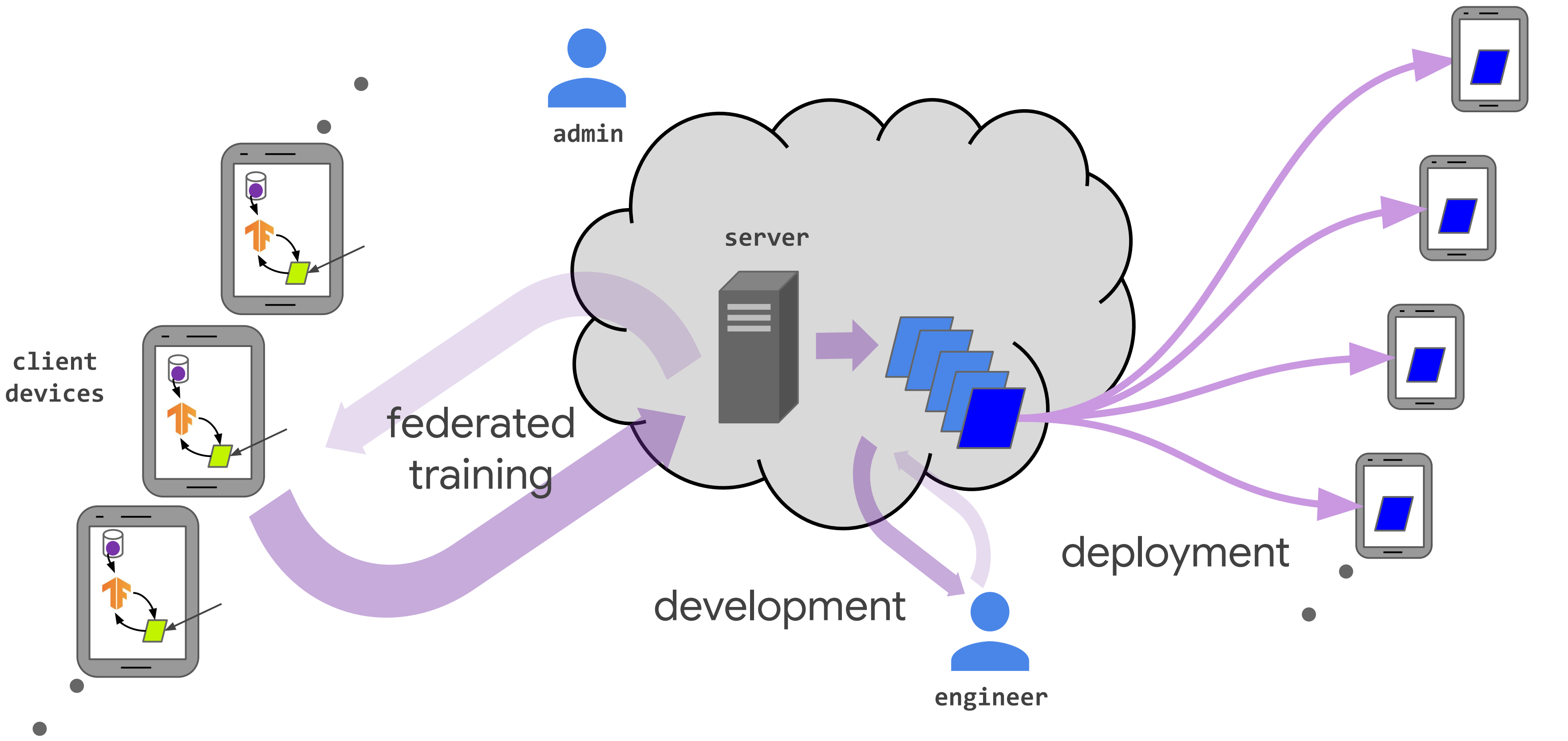
Communication-Privacy-Accuracy Tradeoffs under Distributed DP for FL

University of California Santa Barbara

Peter Kairouz
Research Scientist | Google

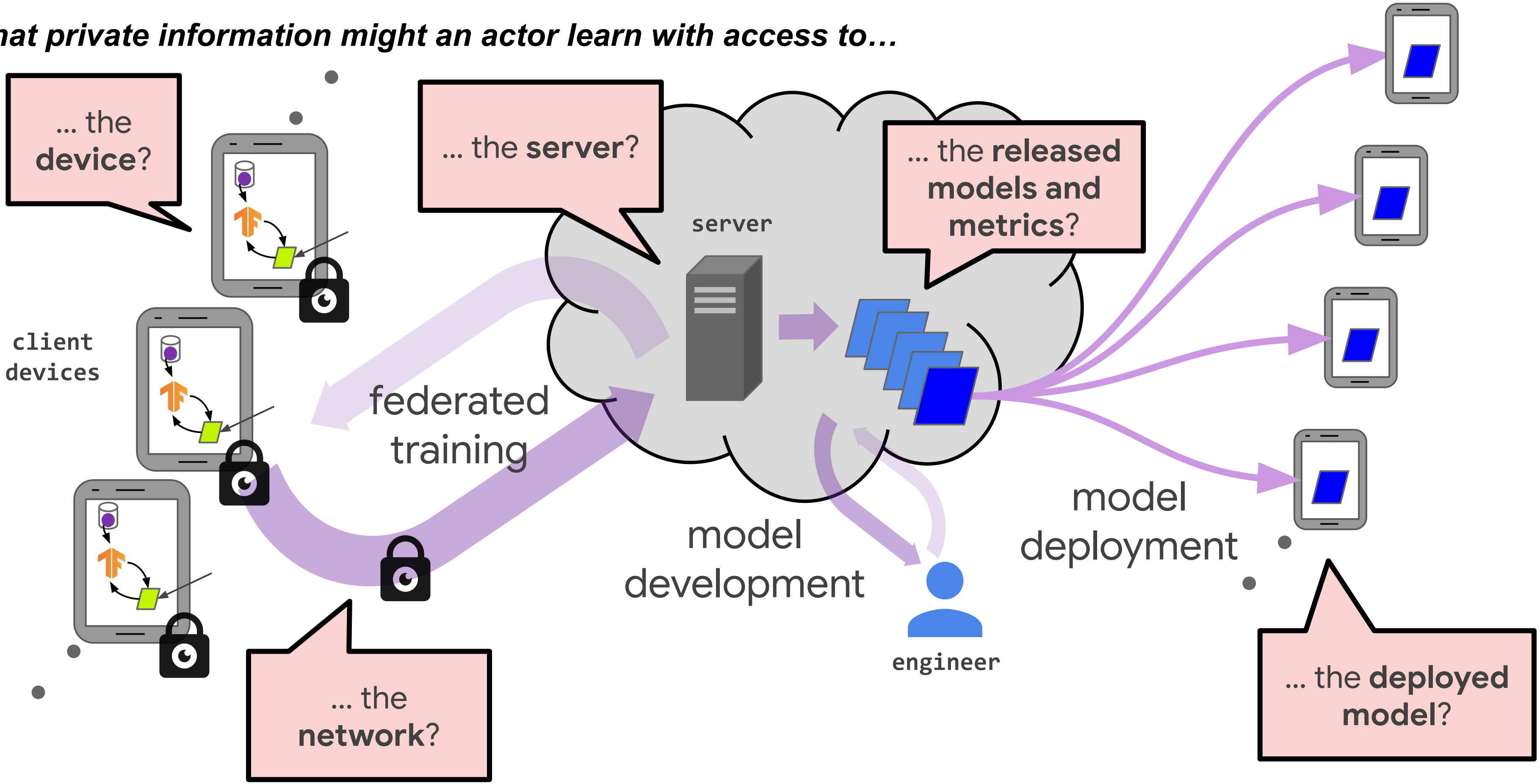
Twitter: @KairouzPeter
Presenting joint work with Ken Liu, Thomas Steinke, Naman Agarwal, and others

Federated Learning

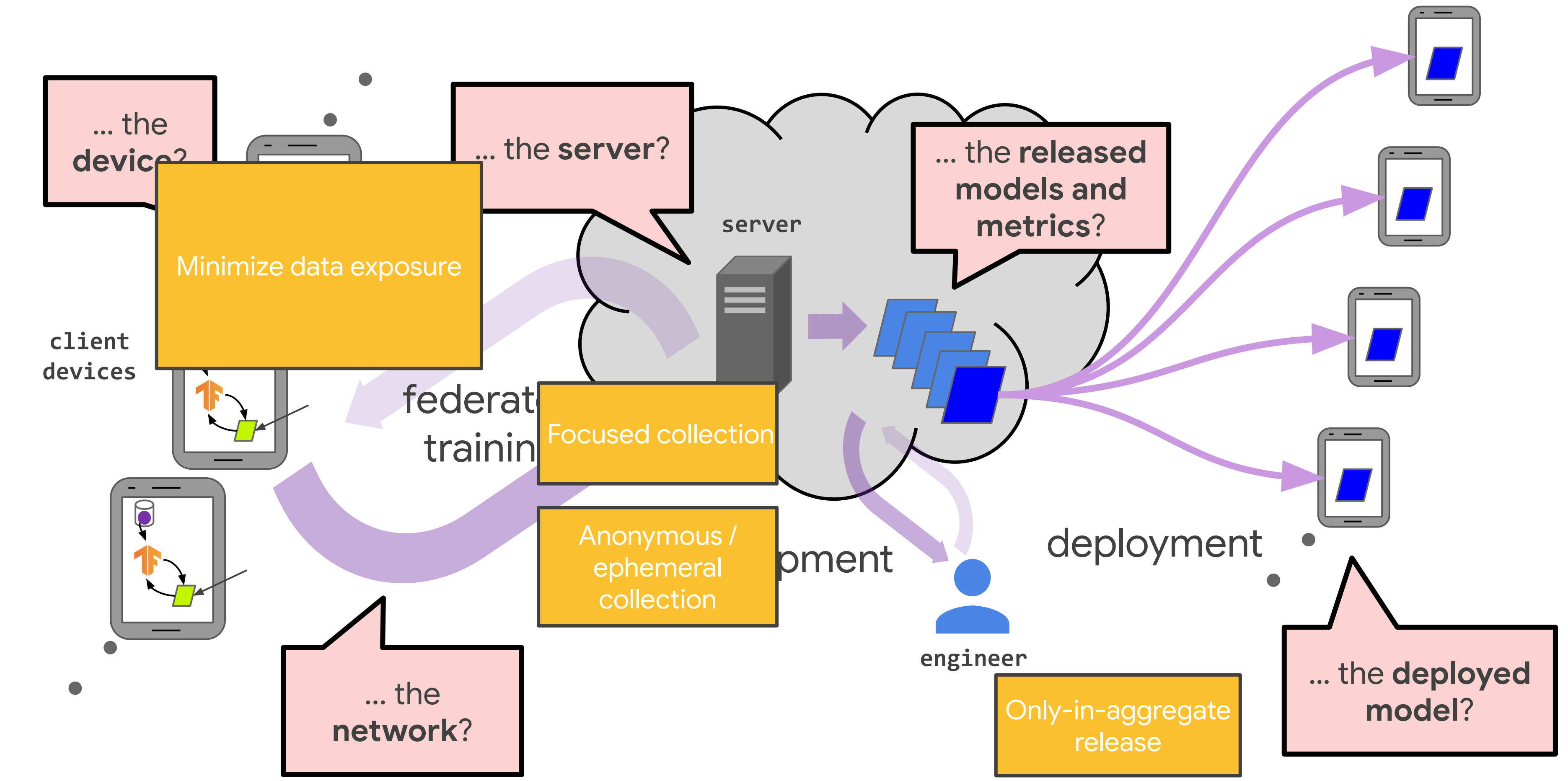


Ensuring privacy of participating users

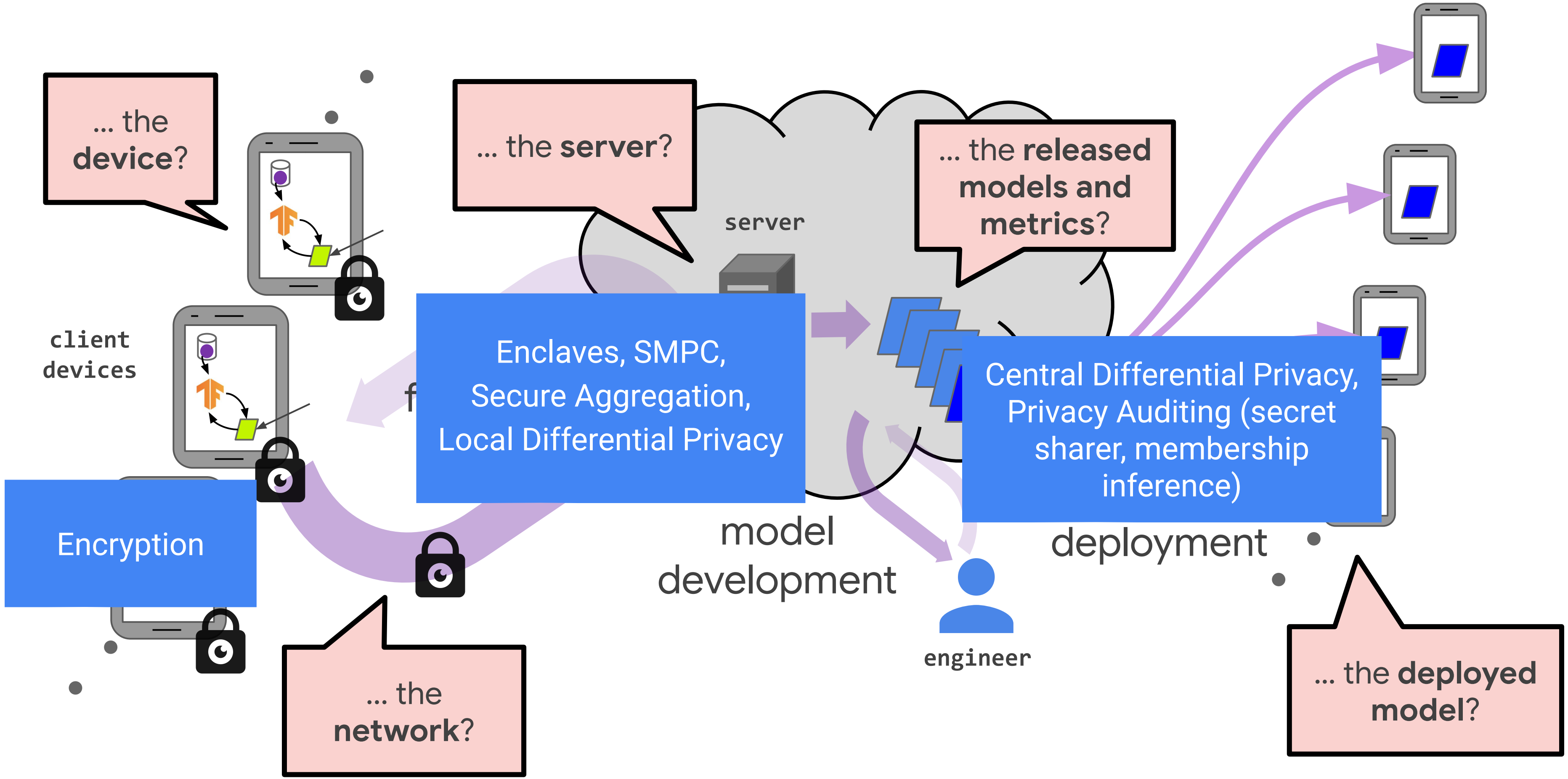
What private information might an actor learn with access to…



Data minimization principles for FL

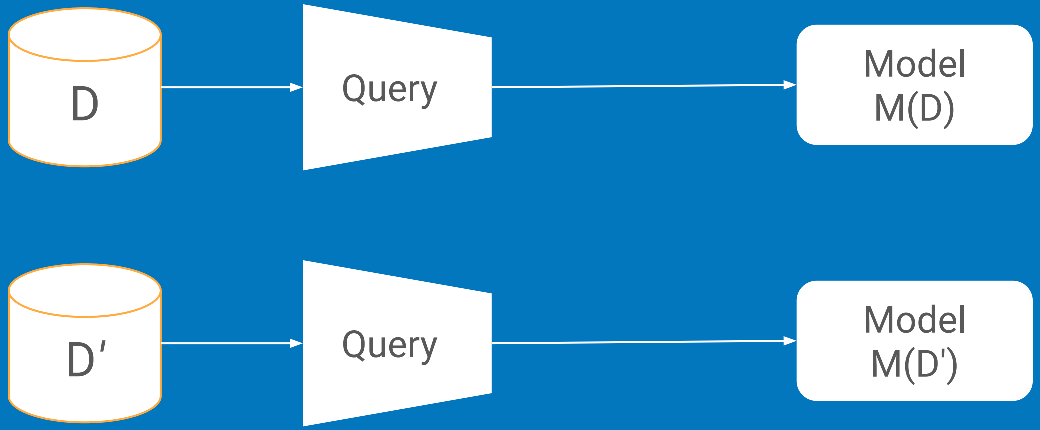


Complementary privacy technologies



Differential Privacy for FL

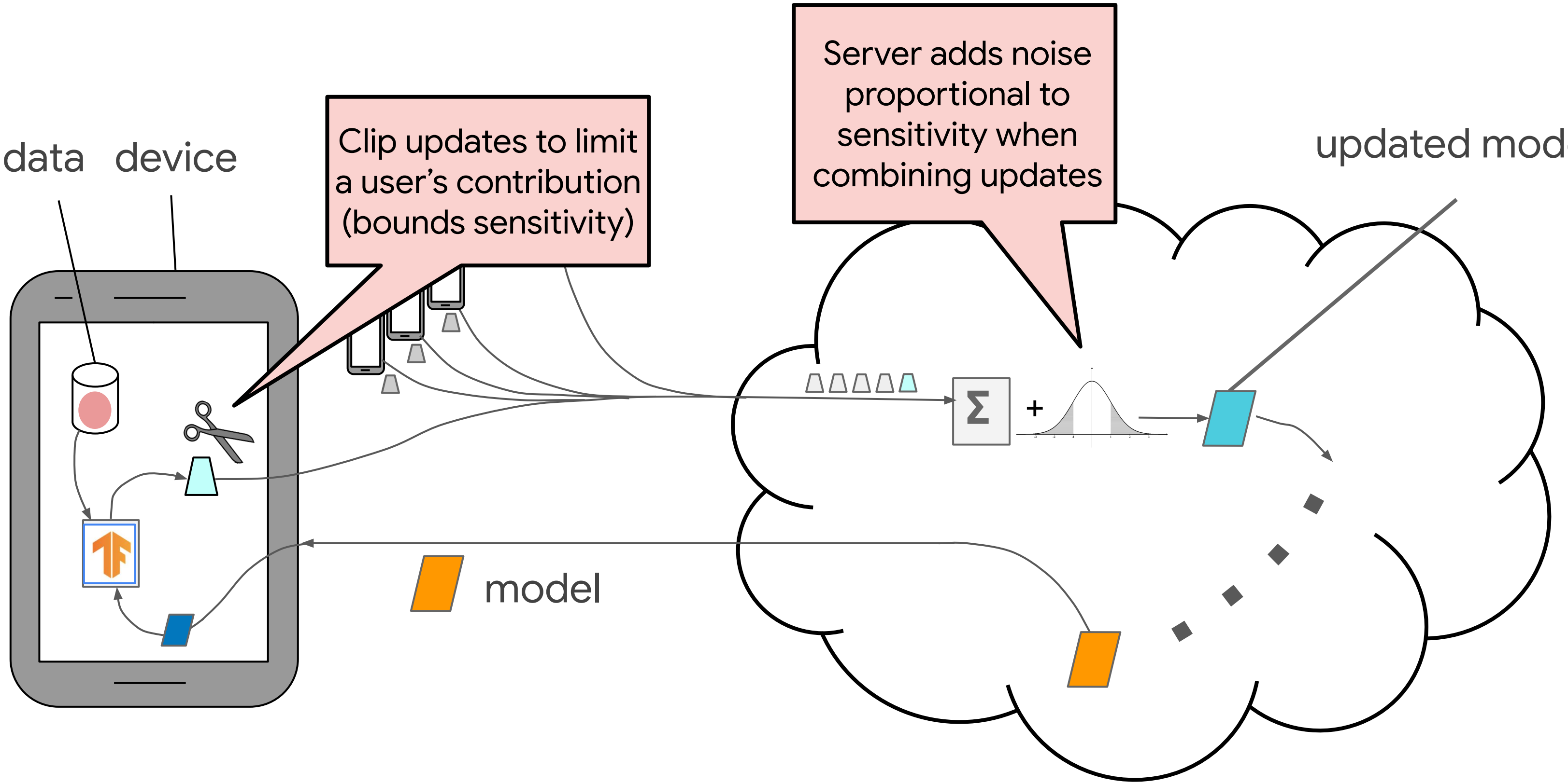
Differential Privacy



(ε, δ)-Differential Privacy: The distribution of the output $M(D)$ (a trained model) on database (training dataset) D is **nearly the same** as $M(D')$ for all adjacent databases D and D'

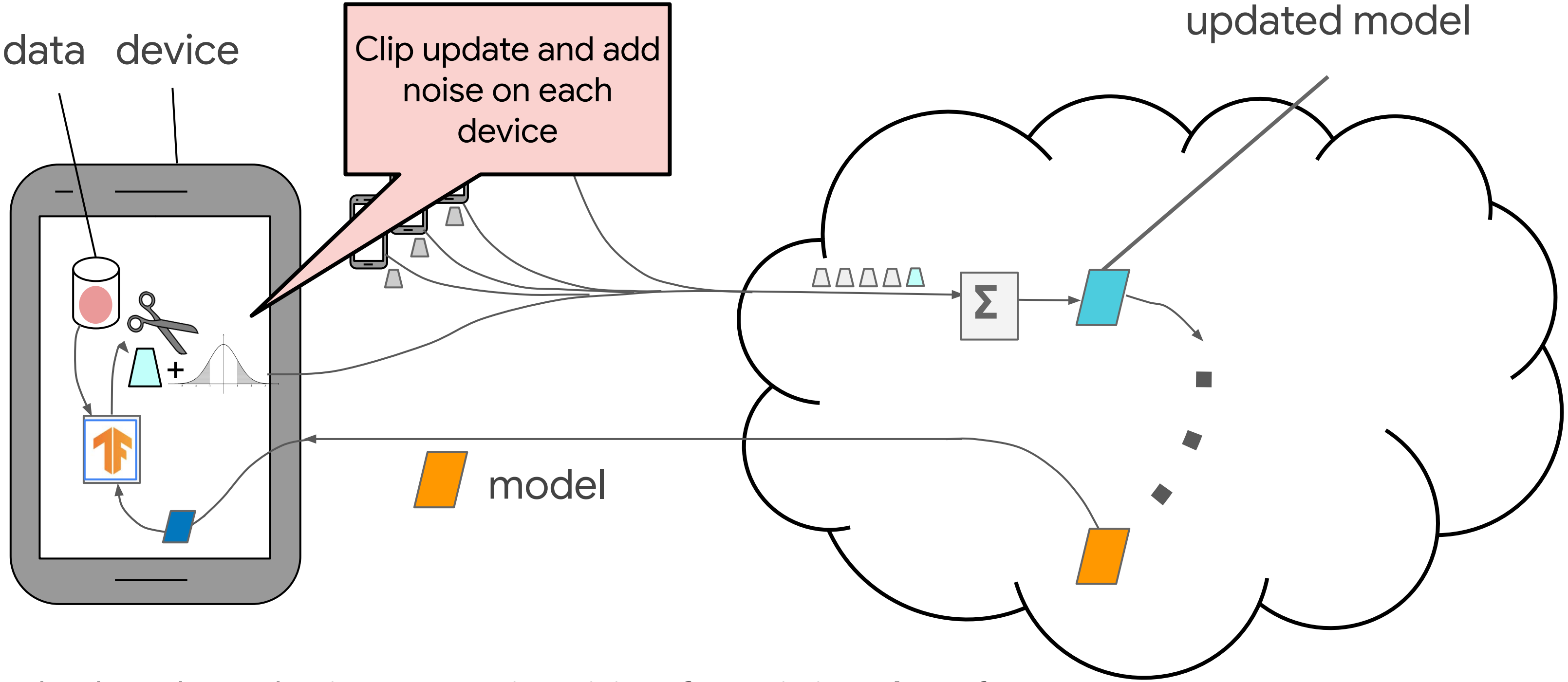
$$\forall S: \quad \Pr[M(D) \in S] \leq \exp(\varepsilon) \cdot \Pr[M(D') \in S] + \delta$$

Centrally differentially private federated learning



H. B. McMahan, et al. ***Learning Differentially Private Recurrent Language Models.*** *ICLR 2018*

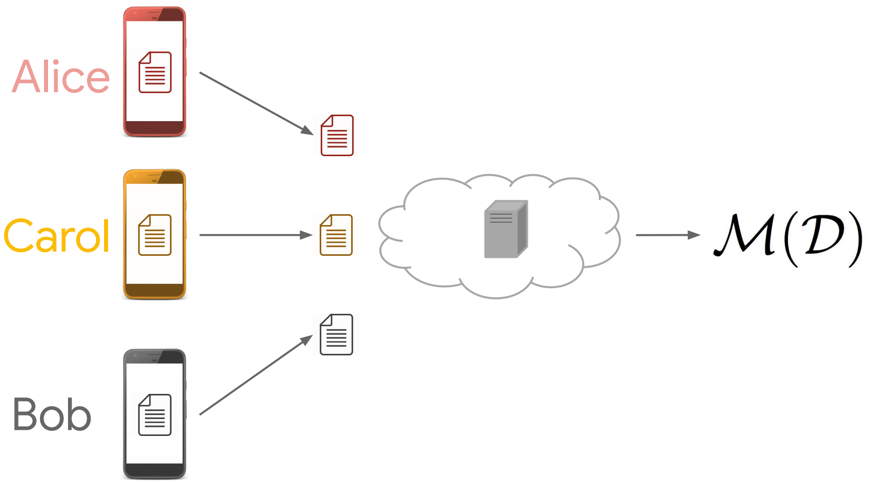
Locally differentially private federated training

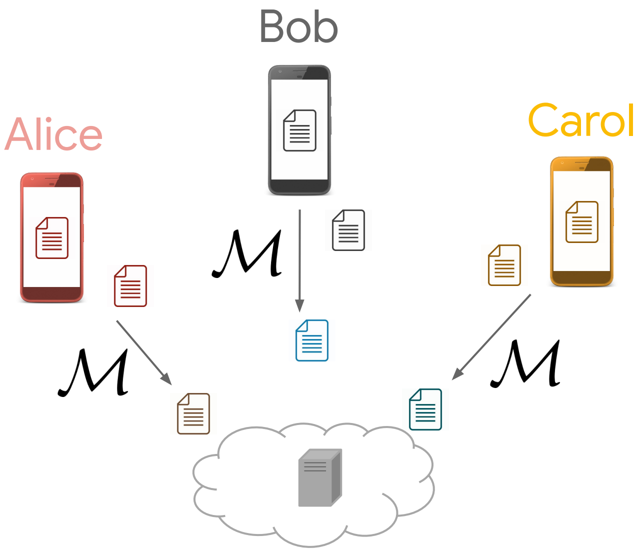


Evfimievski, Alexandre, et al. **Privacy preserving mining of association rules.** Information Systems 2004
Warner, Stanley L. **Randomized response: A survey technique for eliminating evasive answer bias.** JASA 1965

Can we combine the best of both worlds?

Distributed Differential Privacy



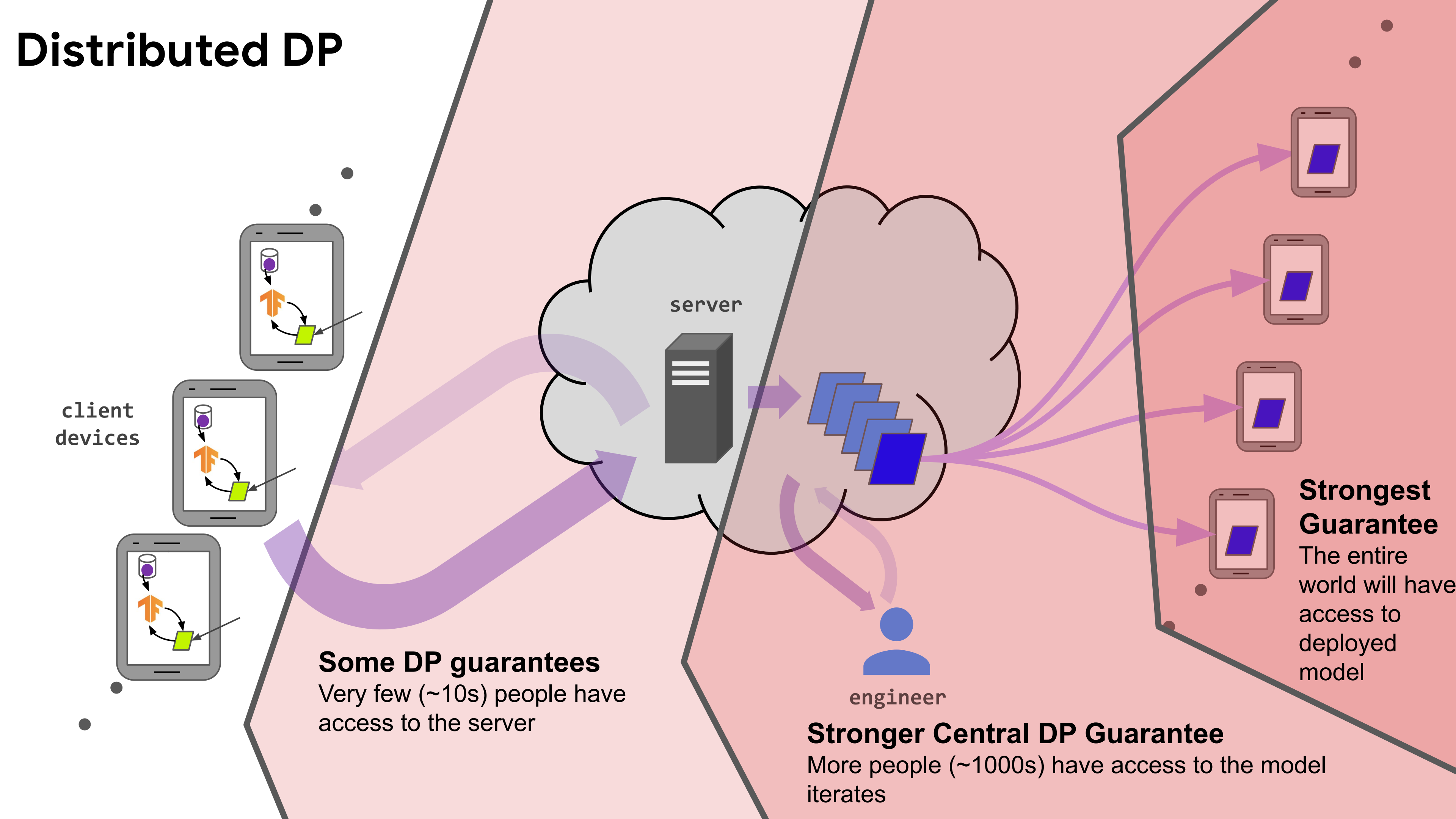


Central DP: full trust in service provider
Higher utility at reasonable privacy levels

Local DP: weaker trust assumptions
Utility often suffers

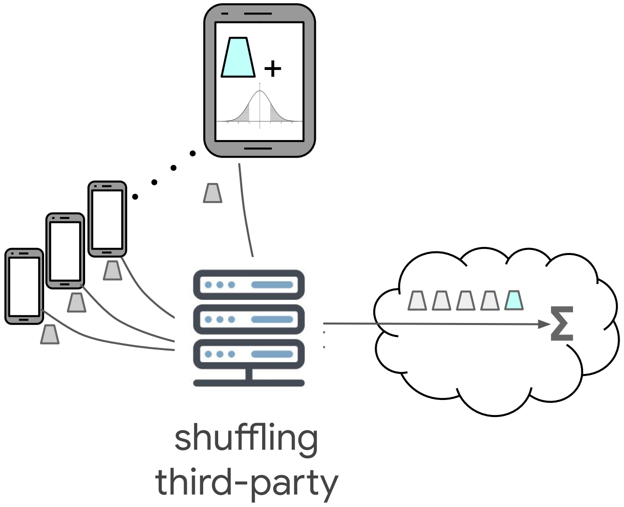
Dwork, et. al. “Our Data, Ourselves: Privacy Via Distributed Noise Generation”. 2006.

Distributed DP



Distributing trust for private aggregation

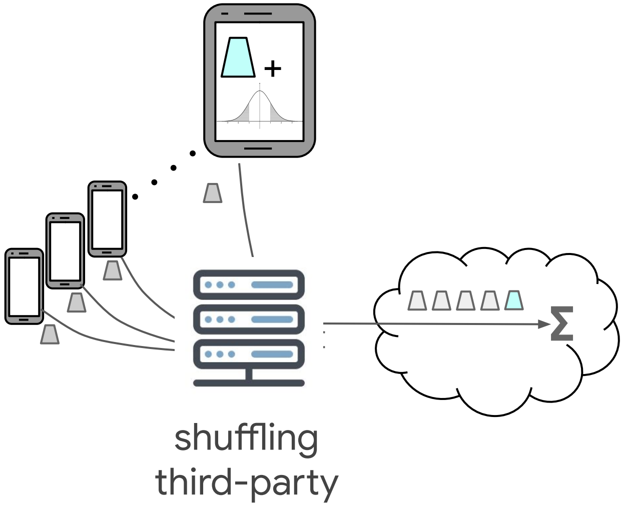
1 Trusted “third party”



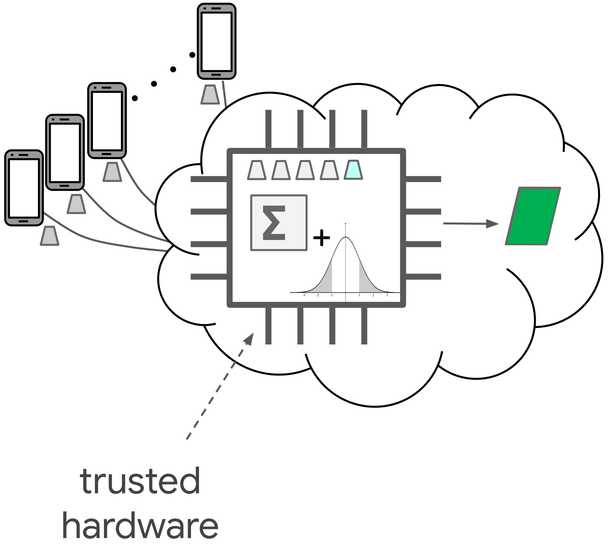
Andrea Bittau, et al. ***Prochlo: Strong Privacy for Analytics in the Crowd****. SOSP 2017*
Úlfar Erlingsson, et al. ***Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity.*** *SODA 2019*

Distributing trust for private aggregation

1 Trusted "third party"



2 Trusted Execution Environments

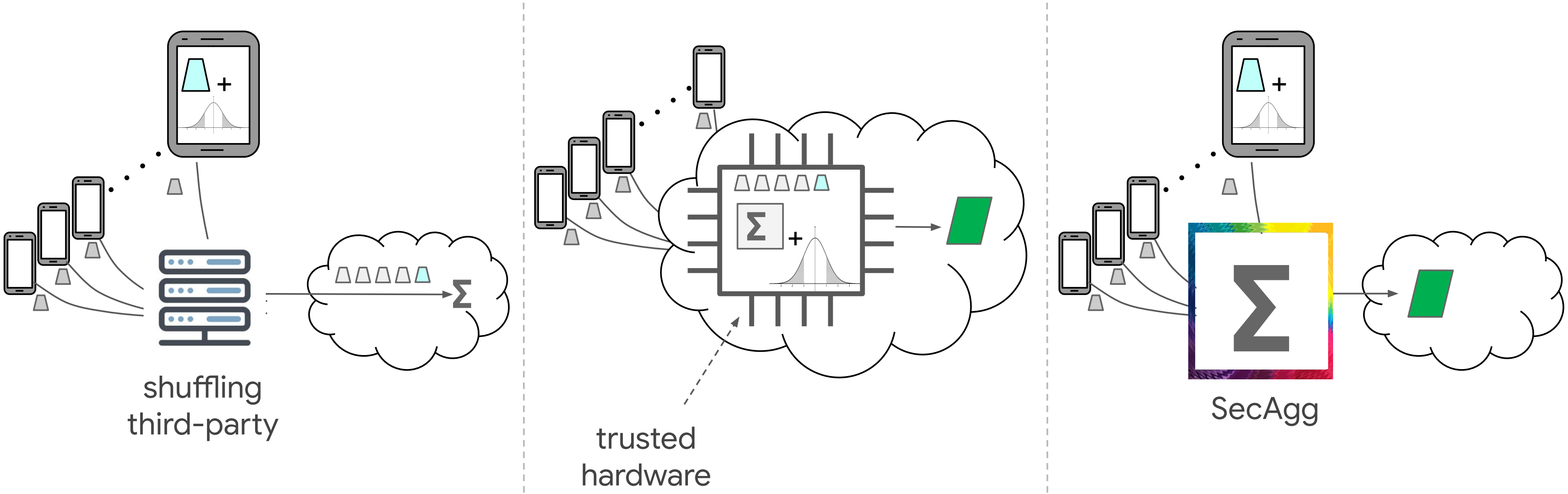


Andrea Bittau, et al. ***Prochlo: Strong Privacy for Analytics in the Crowd***. *SOSP 2017*

Úlfar Erlingsson, et al. ***Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity.*** *SODA 2019*

Distributing trust for private aggregation

1 Trusted “third party” | 2 Trusted Execution Environments | 3 **Trust via Cryptography**

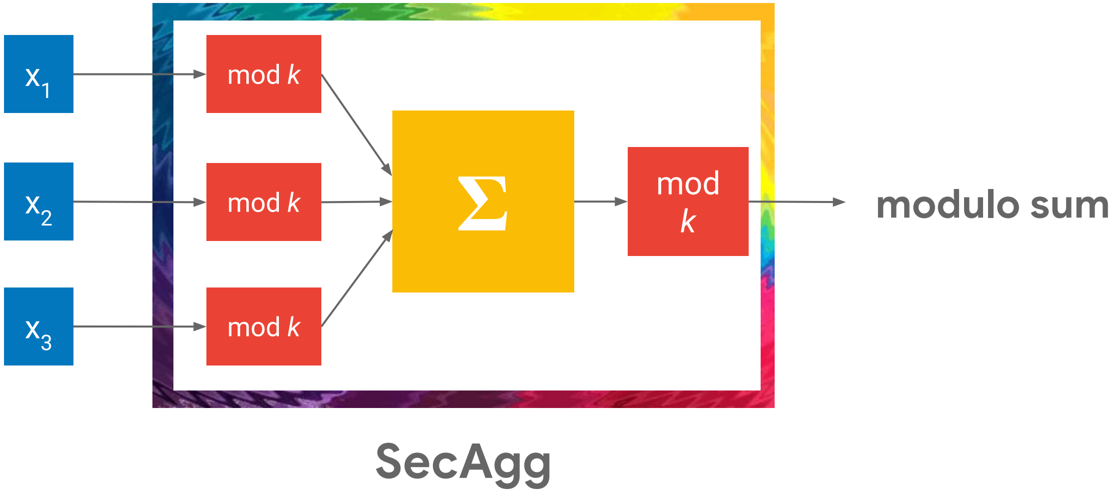


K. A. Bonawitz, et al. ***Practical secure aggregation for privacy-preserving machine learning*** *CCS 2017*
J. Bell, et al. ***Secure Single-Server Vector Aggregation with (Poly) Logarithmic Overhead*** *CCS 2020*

Secure Aggregation allows a server to obtain the sum of high-dimensional vectors of client-held data in a way that ensures (cryptographically) that the server learns *just the sum*, and *no individual data whatsoever* *.

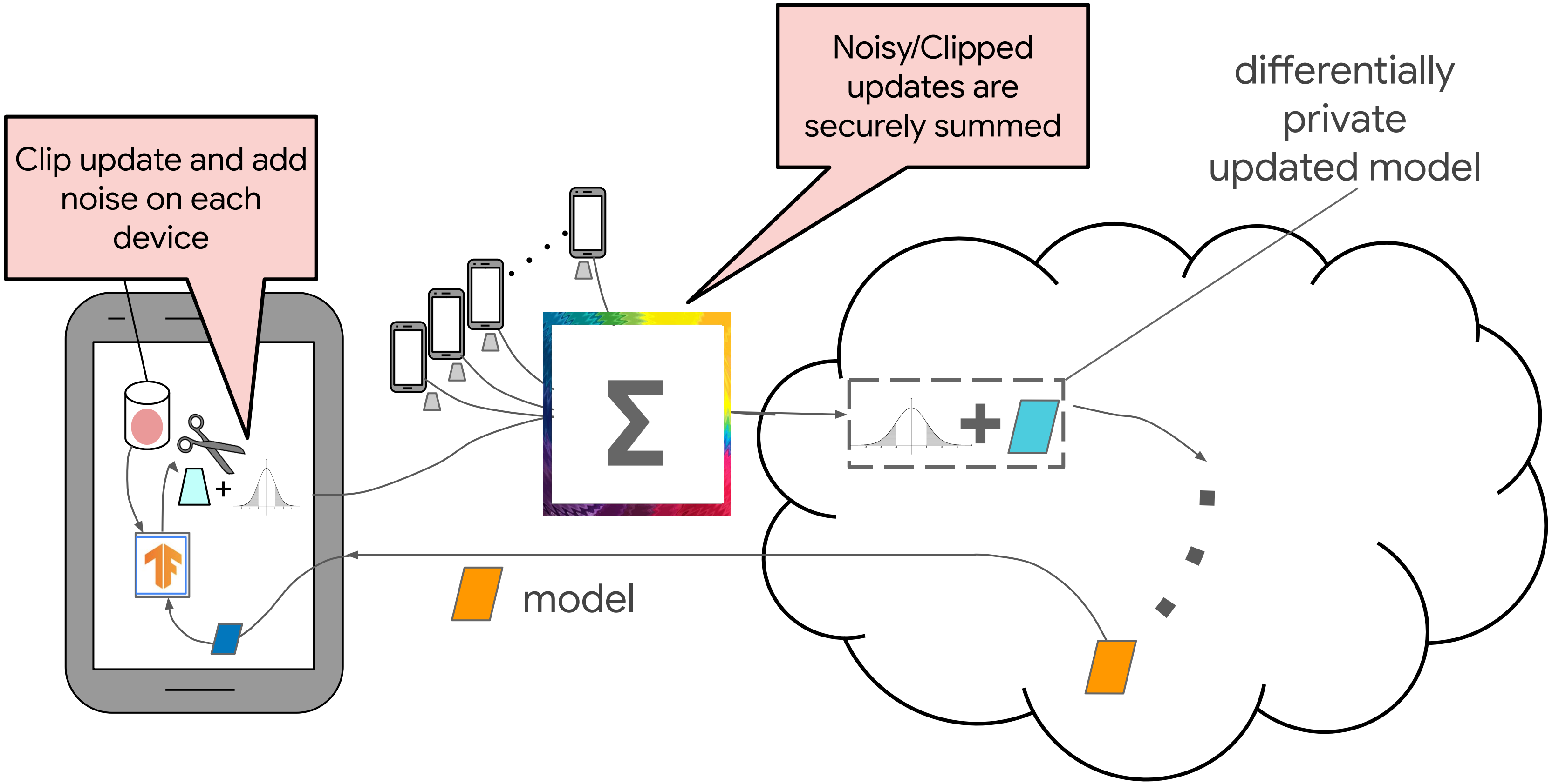
* even if some users are malicious (and collude with the server), and some drop out.

SecAgg: a closed box that performs integer modulo sums



K. A. Bonawitz, et al. ***Practical secure aggregation for privacy-preserving machine learning*** *CCS 2017*
J. Bell, et al. ***Secure Single-Server Vector Aggregation with (Poly) Logarithmic Overhead*** *CCS 2020*

Why not just add continuous Gaussian noise?



Challenges

- SecAgg operates on a finite group with integer modulo arithmetic
  - Need clever data discretization methods that do not inflate the sensitivity
  - Cannot use continuous mechanisms

Challenges

- SecAgg operates on a finite group with integer modulo arithmetic
  - Need clever data discretization methods that do not inflate the sensitivity
  - Cannot use continuous mechanisms
- Communication efficiency is an important bottleneck in FL
  - Do not want privacy to depend on communication cost (SecAgg's group size)

Challenges

- SecAgg operates on a finite group with integer modulo arithmetic
  - Need clever data discretization methods that do not inflate the sensitivity
  - Cannot use continuous mechanisms
- Communication efficiency is an important bottleneck in FL
  - Do not want privacy to depend on communication cost (SecAgg's group size)
- Many discrete mechanisms (e.g. k-RR) are not closed under summation
  - Analyzing sums of these mechanisms is difficult, especially in high dimensions

Challenges

- SecAgg operates on a finite group with integer modulo arithmetic
    - Need clever data discretization methods that do not inflate the sensitivity
    - Cannot use continuous mechanisms
- Communication efficiency is an important bottleneck in FL
    - Do not want privacy to depend on communication cost (SecAgg's group size)
- Many discrete mechanisms (e.g. k-RR) are not closed under summation
    - Analyzing sums of these mechanisms is difficult, especially in high dimensions
- Discrete mechanisms with finite tails* do not satisfy Rényi or concentrated DP
    - Avoids catastrophic privacy failures and allows for tight privacy accounting

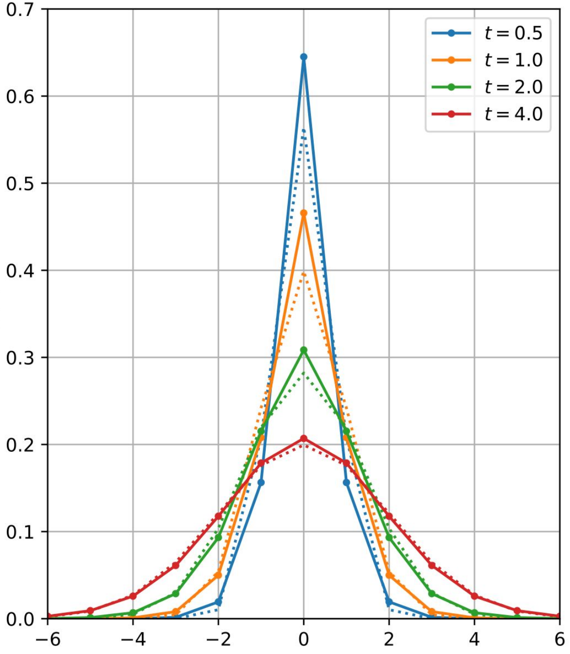
*For example, the multi-dimensional binomial mechanism (Agarwal, et al. **cpSGD: Communication-efficient and differentially-private distributed SGD.** NeurIPS 2018.) does not achieve Rényi DP

Challenges

- SecAgg operates on a finite group with integer modulo arithmetic
  - Need clever data discretization methods that do not inflate the sensitivity
  - Cannot use continuous mechanisms
- Communication efficiency is an important bottleneck in FL
  - Do not want privacy to depend on communication cost (SecAgg's group size)
- Many discrete mechanisms (e.g. k-RR) are not closed under summation
  - Analyzing sums of these mechanisms is difficult, especially in high dimensions
- Discrete mechanisms with finite tails* do not satisfy Rényi or concentrated DP
  - Avoids catastrophic privacy failures and allows for tight privacy accounting
- Need mechanisms that can be sampled from exactly and efficiently

*For example, the multi-dimensional binomial mechanism (Agarwal, et al. **cpSGD: Communication-efficient and differentially-private distributed SGD.** NeurIPS 2018.) does not achieve Rényi DP

The discrete Gaussian mechanism

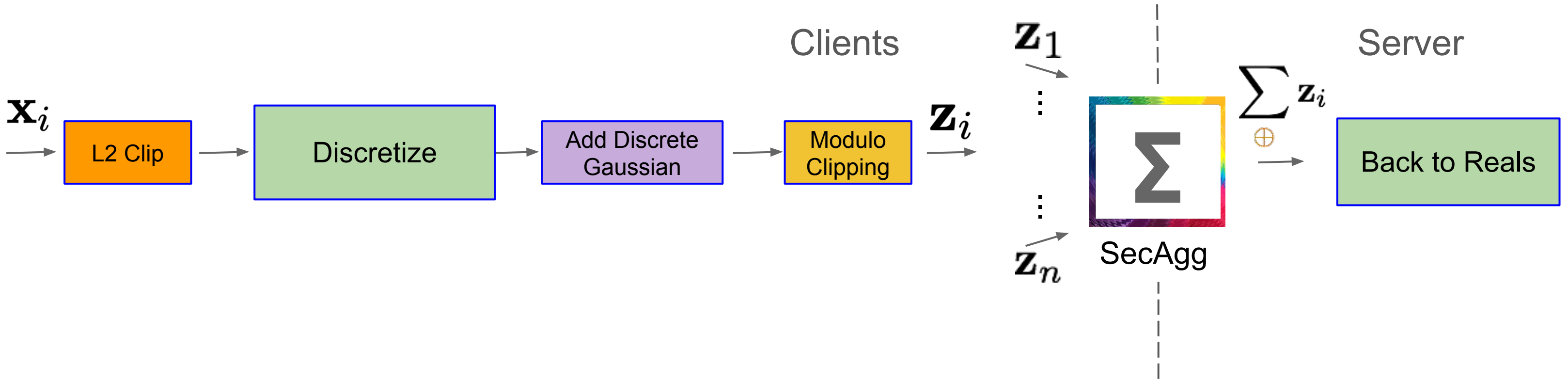


$$\mathbb{P}_{X \leftarrow \mathcal{N}_{\mathbb{Z}}(\mu, t^2)}[X = n] = \frac{e^{-(n-\mu)^2/2t^2}}{\sum_{k \in \mathbb{Z}} e^{-(k-\mu)^2/2t^2}}$$

- **Discrete Gaussian:** discrete analog of continuous Gaussian (≠ rounding Gaussian to nearest ints)
  - Essentially **the same privacy-accuracy trade-off** as continuous Gaussian*
  - zCDP / Rényi DP for tight compositions in learning contexts
- **Problem:** sums of discrete Gaussians ≠ discrete Gaussians

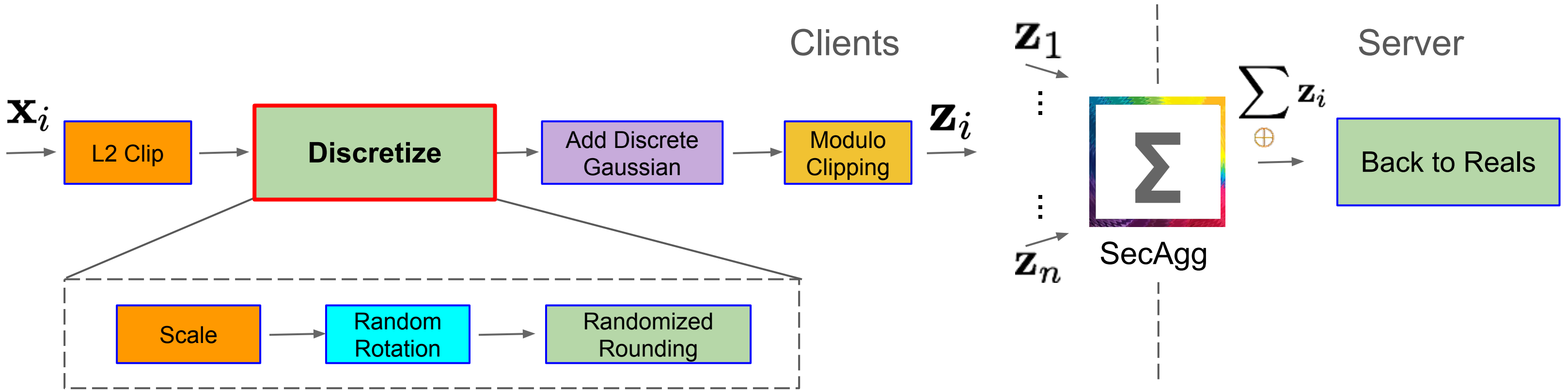
*Clement Canonne, et al. ***The Discrete Gaussian for Differential Privacy***. NeurIPS 2020*

The Distributed Discrete Gaussian Mechanism



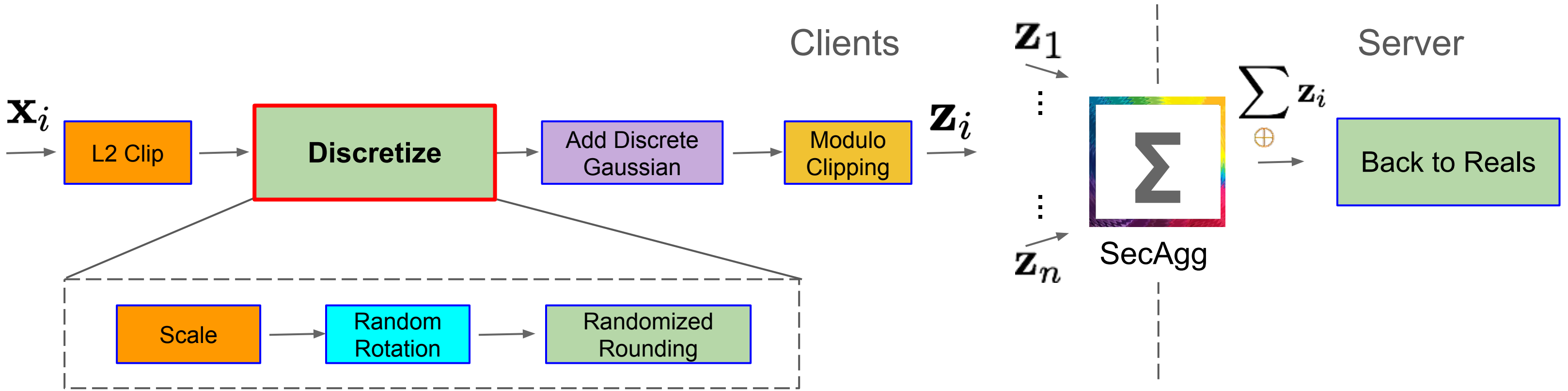
The Distributed Discrete Gaussian Mechanism for Federated Learning with Secure Aggregation. *ICML 2021*

Data Quantization



- **Scaling:** stretching the signal → reduces quantization error
- **Random rotation:** "flatten" concentrated coordinates → controls the L-inf norm
- **Randomized rounding:** values stochastically rounded to integers (unbiased)

Data Quantization



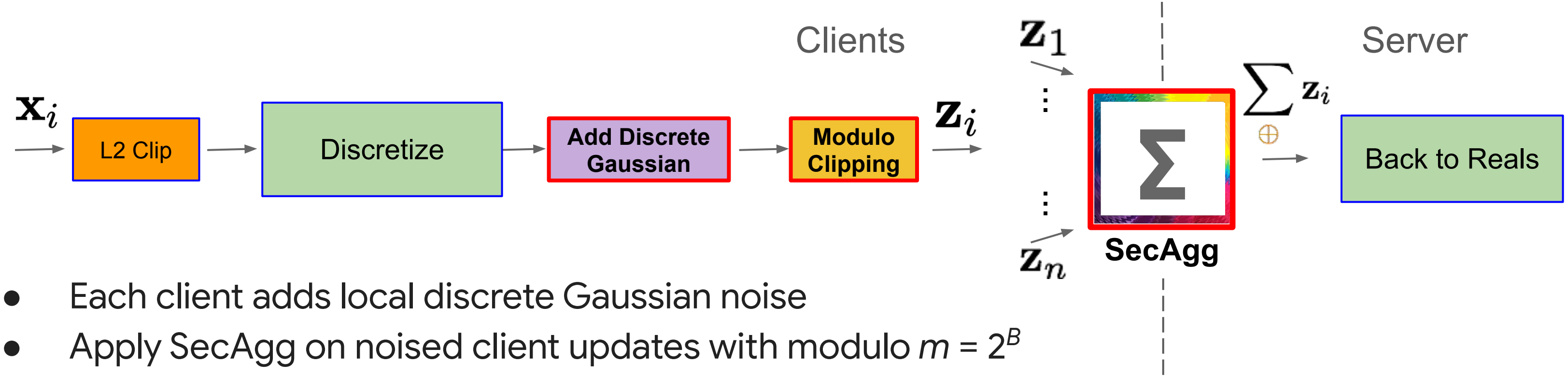
- **Scaling:** stretching the signal ⟶ reduces quantization error
- **Random rotation:** "flatten" concentrated coordinates ⟶ controls the L-inf norm
- **Randomized rounding:** values stochastically rounded to integers (unbiased)
- We can probabilistically bound the L2 norm growth from rounding (**helps reduce DP noise**):

Proposition 22 (Properties of Randomized Rounding). *Let* $\beta \in [0,1)$, $\gamma > 0$, *and* $x \in \mathbb{R}^d$.

$$\Delta_2^2 := \min \left\{ \begin{array}{c} \|x\|_2^2 + \frac{1}{4}\gamma^2 d + \sqrt{2\log(1/\beta)} \cdot \gamma \cdot \left(\|x\|_2 + \frac{1}{2}\gamma\sqrt{d}\right), \\ \left(\|x\|_2 + \gamma\sqrt{d}\right)^2 \end{array} \right\}$$

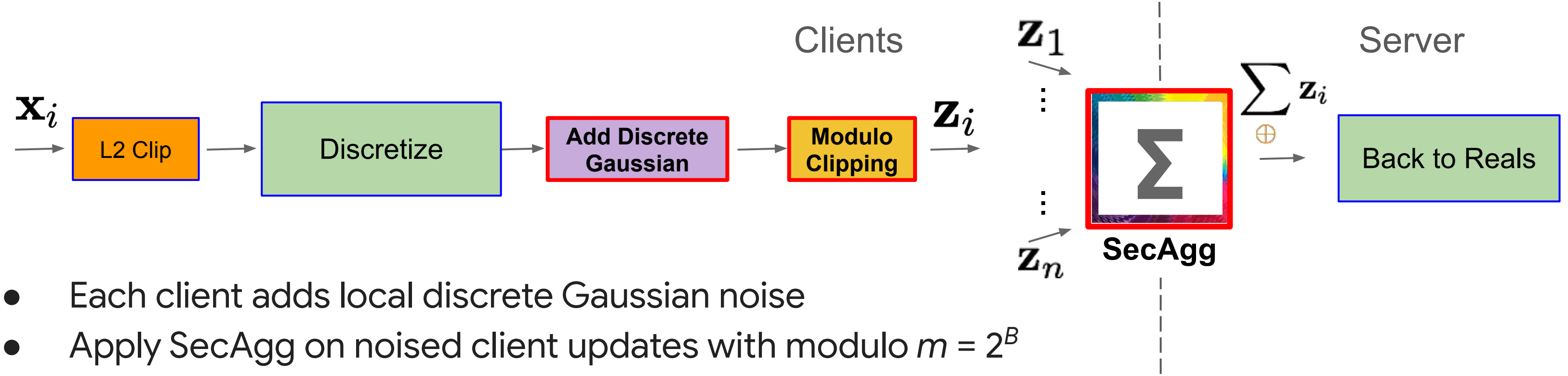
d: client vector dim
γ: rounding granularity; inverse scaling factor
β: rounding bias

Local Noising & (Secure) Sums of Discrete Gaussians



- Each client adds local discrete Gaussian noise
- Apply SecAgg on noised client updates with modulo $m = 2^B$

Local Noising & (Secure) Sums of Discrete Gaussians



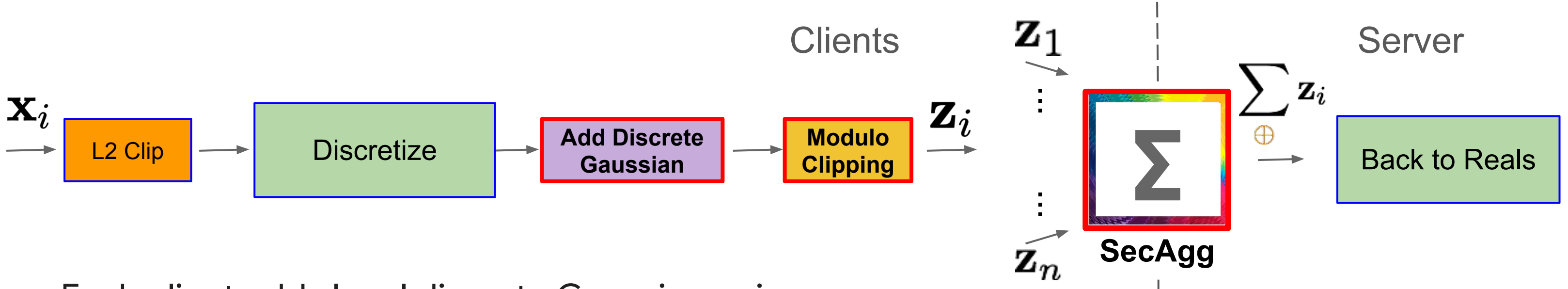
- Each client adds local discrete Gaussian noise
- Apply SecAgg on noised client updates with modulo $m = 2^B$
- While sums of discrete Gaussians ≠ discrete Gaussian, we show that they are **extremely close:**

Theorem 11 (Convolution of two Discrete Gaussians). *Let $\sigma, \tau \geq \frac{1}{2}$. Let $X \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \sigma^2)$ and $Y \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \tau^2)$ be independent. Let $Z = X + Y$. Let $W \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \sigma^2 + \tau^2)$. Then*

$$\mathrm{D}_{\pm\infty}(Z \| W) = \sup_{z \in \mathbb{Z}} \left| \log \left(\frac{\mathbb{P}[Z = z]}{\mathbb{P}[W = z]} \right) \right| \leq 5 \cdot e^{-2\pi^2/(1/\sigma^2 + 1/\tau^2)}.$$

- Exponentially small with larger variance; ≤ 10^{-12} if $\sigma_1^2 = \sigma_2^2 = 3$.
- Noise is added on **quantized** client values, so σ^2 is **scaled** and this is even smaller

Local Noising & (Secure) Sums of Discrete Gaussians



- Each client adds local discrete Gaussian noise
- Apply SecAgg on noised client updates with modulo $m = 2^B$
- While sums of discrete Gaussians ≠ discrete Gaussian, we show that they are **extremely close:**

Theorem 11 (Convolution of two Discrete Gaussians). *Let* $\sigma, \tau \geq \frac{1}{2}$. *Let* $X \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \sigma^2)$ *and* $Y \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \tau^2)$ *be independent. Let* $Z = X + Y$. *Let* $W \leftarrow \mathcal{N}_{\mathbb{Z}}(0, \sigma^2 + \tau^2)$. *Then*

$$\mathrm{D}_{\pm\infty}(Z\|W) = \sup_{z \in \mathbb{Z}} \left| \log\left(\frac{\mathbb{P}[Z = z]}{\mathbb{P}[W = z]} \right) \right| \leq 5 \cdot e^{-2\pi^2/(1/\sigma^2 + 1/\tau^2)}.$$

Main Privacy Guarantees with n clients:

$$\tau := 10 \cdot \sum_{k=1}^{n-1} e^{-2\pi^2 \frac{\sigma^2}{\gamma^2} \cdot \frac{k}{k+1}}$$

$$\varepsilon := \min \left\{ \begin{array}{l} \sqrt{\frac{\Delta_2^2}{n\sigma^2} + \frac{1}{2}\tau d} \\ \frac{\Delta_2}{\sqrt{n}\sigma} + \tau\sqrt{d} \end{array} \right\}$$

- Weak dependence on d (number of model params);
- 1st term same as central Gaussian/DGaussian

Stack Overflow Next Word Prediction

Code

- Next word prediction for question/answer sentences on StackOverflow.com with LSTMs
- ~10^9 sentences grouped by the N = 342477 SO users/clients
- **Fig. 1: DDGauss matches continuous Gaussian as long as the bit-width *B* is sufficient**
- **Fig. 2: DDGauss scales (1000 clients per round) and works in low-noise (utility-first) settings**

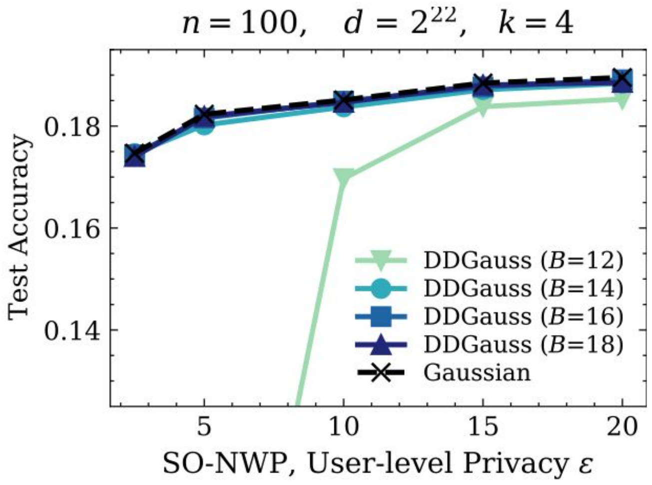


Fig. 1: Test acc with different ε and *B* (n = 100).

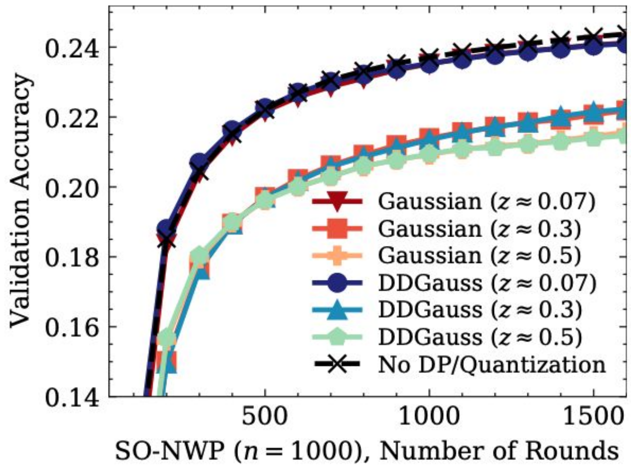
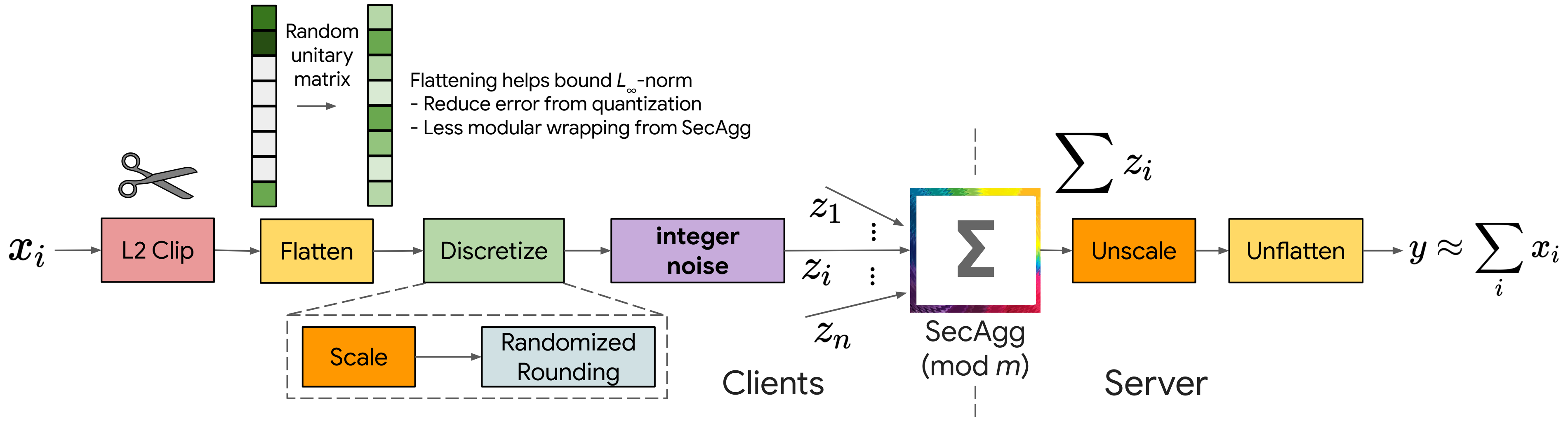


Fig. 2: Val acc with with n = 1000 clients, B = 18.
z: approximate noise multiplier aligned on ε.

Code: https://github.com/google-research/federated/tree/master/distributed_dp

Our end-to-end solution



E.g. 4.23 → 42.3 → {42 w.p 0.7, 43 w.p. 0.3}

(Symmetric) Skellam Distribution

- Difference of two independent Poisson RVs. With mean Δ and variance *μ*,

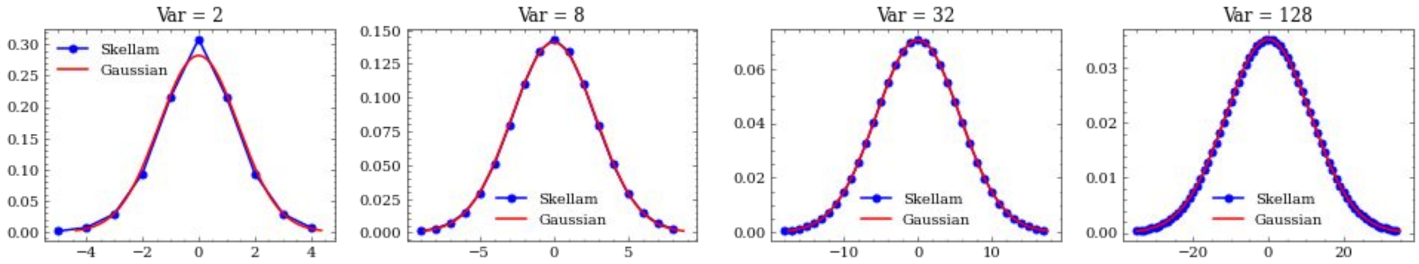
$I_k(z)$: modified Bessel function of the first kind

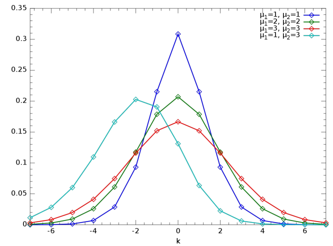
$$X_i \sim \mathrm{Sk}_{\Delta_i,\mu} \quad \text{with} \quad P(X_i = k) = e^{-\mu} I_{k-\Delta_i}(\mu)$$

- **Closed under summation**: easily switch between **central DP** & **distributed DP** (central vs local noise)
- **Easy to sample**: `` `np.random.poisson` ``
- Skellam gets closer to Gaussian as variance increases and we scale the output appropriately

- **Skellam Mechanism**: for an integer-valued query *f(D)*,

$$\mathrm{Sk}_{0,\mu}(f(D)) = f(D) + Z \text{ where } Z \sim \mathrm{Sk}_{0,\mu}$$





(Distributed) Skellam

$$X \sim \mathrm{Sk}_{\Delta,\mu}(X) \triangleq e^{-\mu} I_{k-\Delta}(\mu)$$

- **Main Rényi DP guarantee**

Gaussian RDP

For ℓ_1, ℓ_2 sensitivities Δ_1, Δ_2, central variance μ, and order $\alpha > 1, \alpha \in \mathbb{Z}$,

$$\varepsilon(\alpha) \leq \frac{\alpha \Delta_2^2}{2\mu} + \min\left(\frac{(2\alpha - 1)\Delta_2^2 + 6\Delta_1}{4\mu^2}, \frac{3\Delta_1}{2\mu}\right)$$

L1 bound (after quantization)

$$\Delta_1 \leq \Delta_2 \cdot \min\left(\sqrt{d}, \Delta_2\right)$$

2nd term goes to 0 with larger variance (higher privacy)

(Distributed) Skellam

$$X \sim \mathrm{Sk}_{\Delta,\mu}(X) \triangleq e^{-\mu} I_{k-\Delta}(\mu)$$

- **Main Rényi DP guarantee**

Gaussian RDP

For ℓ_1, ℓ_2 sensitivities Δ_1, Δ_2, central variance μ, and order $\alpha > 1, \alpha \in \mathbb{Z}$,

$$\varepsilon(\alpha) \leq \boxed{\frac{\alpha \Delta_2^2}{2\mu}} + \min\left(\frac{(2\alpha - 1)\Delta_2^2 + 6\Delta_1}{4\mu^2}, \frac{3\Delta_1}{2\mu}\right)$$

- **Effect of scaling** (scale both noise stddev and sensitivity)

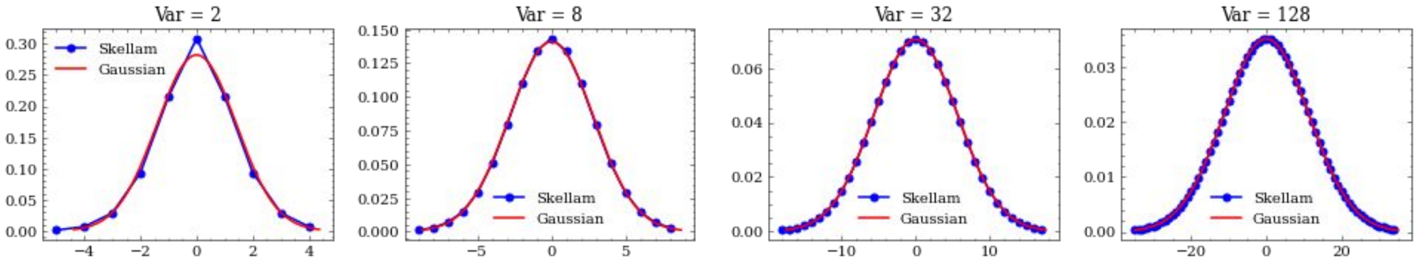
Corollary 4.1 (Scaled Skellam Mechanism). *With a scaling factor $s \in \mathbb{R}$, the multi-dimensional Skellam Mechanism is (α, ε)-RDP with*

$$\varepsilon(\alpha) \leq \frac{\alpha \Delta_2^2}{2\mu} + \min\left(\frac{(2\alpha - 1)\Delta_2^2}{4s^2\mu^2} + \frac{3\Delta_1}{2s^3\mu^2}, \frac{3\Delta_1}{2s\mu}\right)$$

L1 bound (after quantization)

$$\Delta_1 \leq \Delta_2 \cdot \min\left(\sqrt{d}, \Delta_2\right)$$

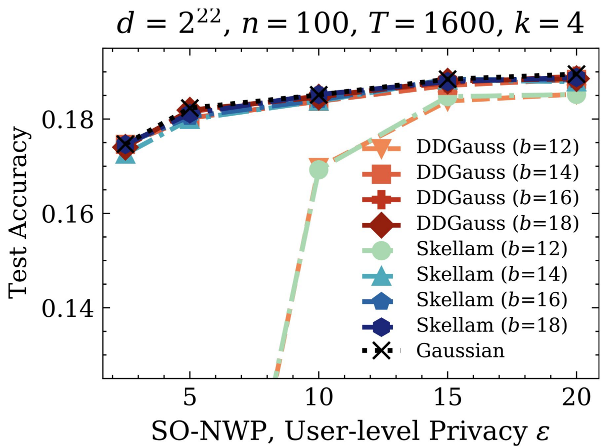
2nd term goes to 0 with larger variance (higher privacy or **large scaling**)

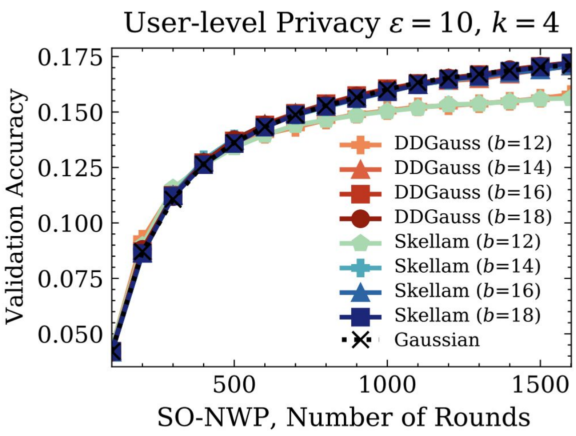


Stack Overflow Next Word Prediction

Code

- Next word prediction for questions/answers sentences on StackOverflow.com with LSTMs
- ~10^9 sentences grouped by N = 342477 users on Stack Overflow
- **Left:** Test acc across various privacy levels ε and bit-widths b
- **Right:** Validation acc across training rounds
- Skellam matches continuous Gaussian and distributed discrete Gaussian





Code: https://github.com/google-research/federated/tree/master/distributed_dp

Better communication efficiency?

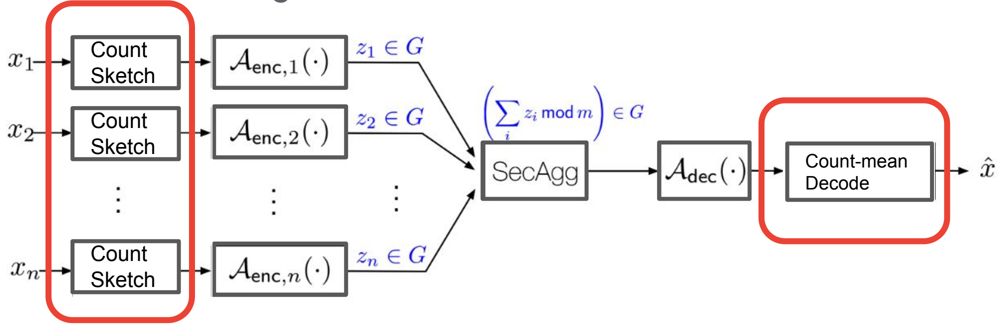
To achieve centralized error of $O\left(\frac{d}{n^2\varepsilon^2}\right)$ each client must transmit $\Omega\left(d\log\left(\frac{n^2\varepsilon^2}{d}\right)\right) = \tilde{\Omega}\left(d\right)$ bits.

Better communication efficiency?

To achieve centralized error of $O\left(\frac{d}{n^2\varepsilon^2}\right)$ each client must transmit $\Omega\left(d\log\left(\frac{n^2\varepsilon^2}{d}\right)\right) = \tilde{\Omega}(d)$ bits.

In the worst-case, each client **cannot** transmit less than the entire gradient!

- But, gradients may be near-sparse! Is their sum?
- We can leverage this structure to compress each x_i !



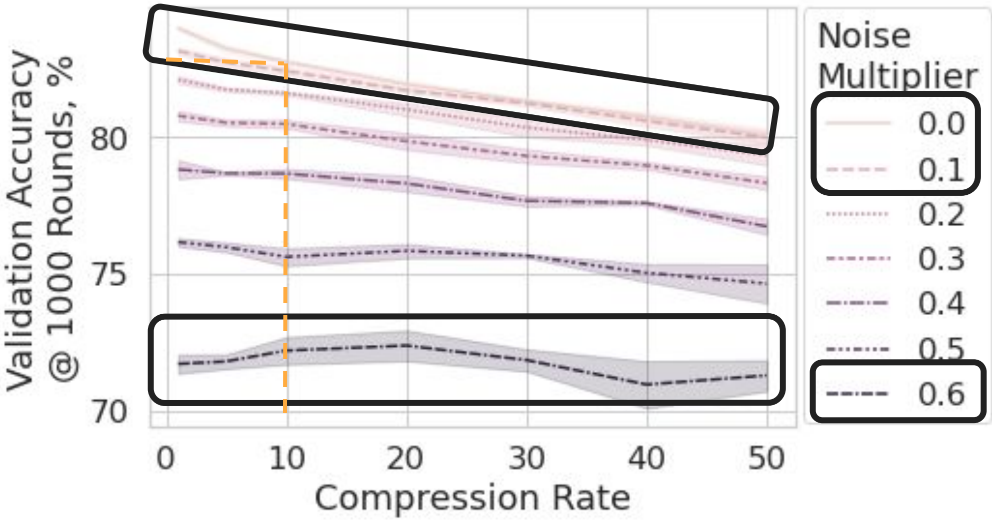
- We will use a count-mean sketch: **efficient** and **linear** dimensionality reduction

Stich, Sebastian U., Jean-Baptiste Cordonnier, and Martin Jaggi. "Sparsified SGD with memory." arXiv preprint arXiv:1809.07599 (2018).
Barnes, Leighton Pate, et al. "rTop-k: A statistical estimation approach to distributed SGD." IEEE Journal on Selected Areas in Information Theory 1.3 (2020): 897-907.
Rothchild, Daniel, et al. "Fetchsgd: Communication-efficient federated learning with sketching." International Conference on Machine Learning. PMLR, 2020.
Haddadpour, Farzin, et al. "Fedsketch: Communication-efficient and private federated learning via sketching." arXiv preprint arXiv:2008.04975 (2020).

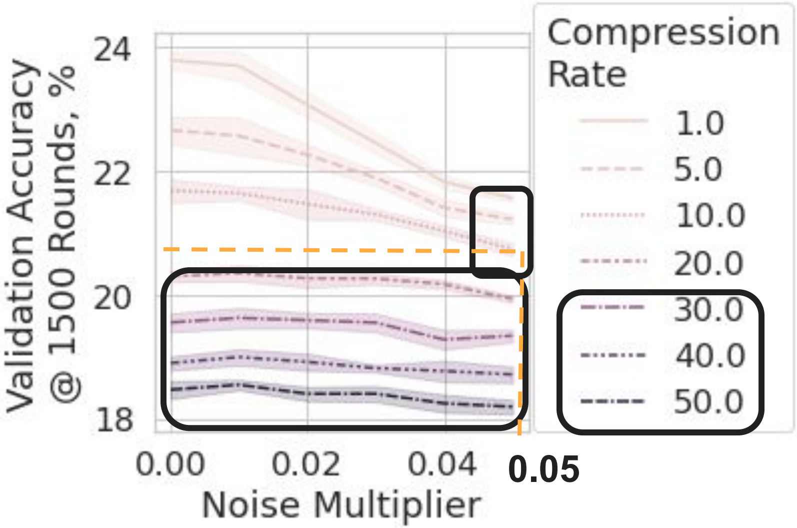
Experiments

Want:
Noise Multiplier ↑ ⟷ ��
Accuracy ↑

Federated EMNIST-62 @ 100 Clients



Stack Overflow Next Word Prediction
@ 100 Clients



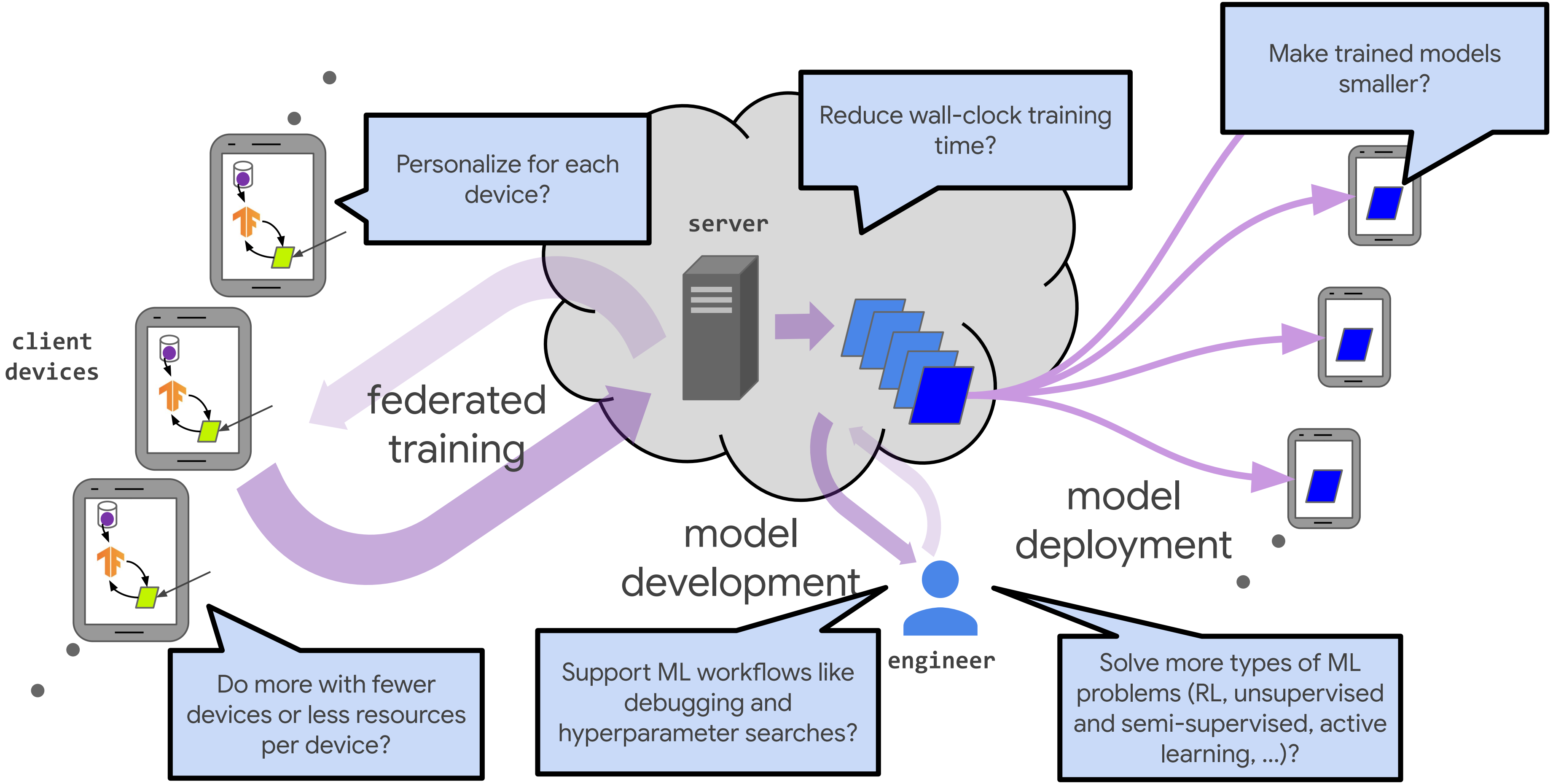
Google Research

Challenges & Opportunities

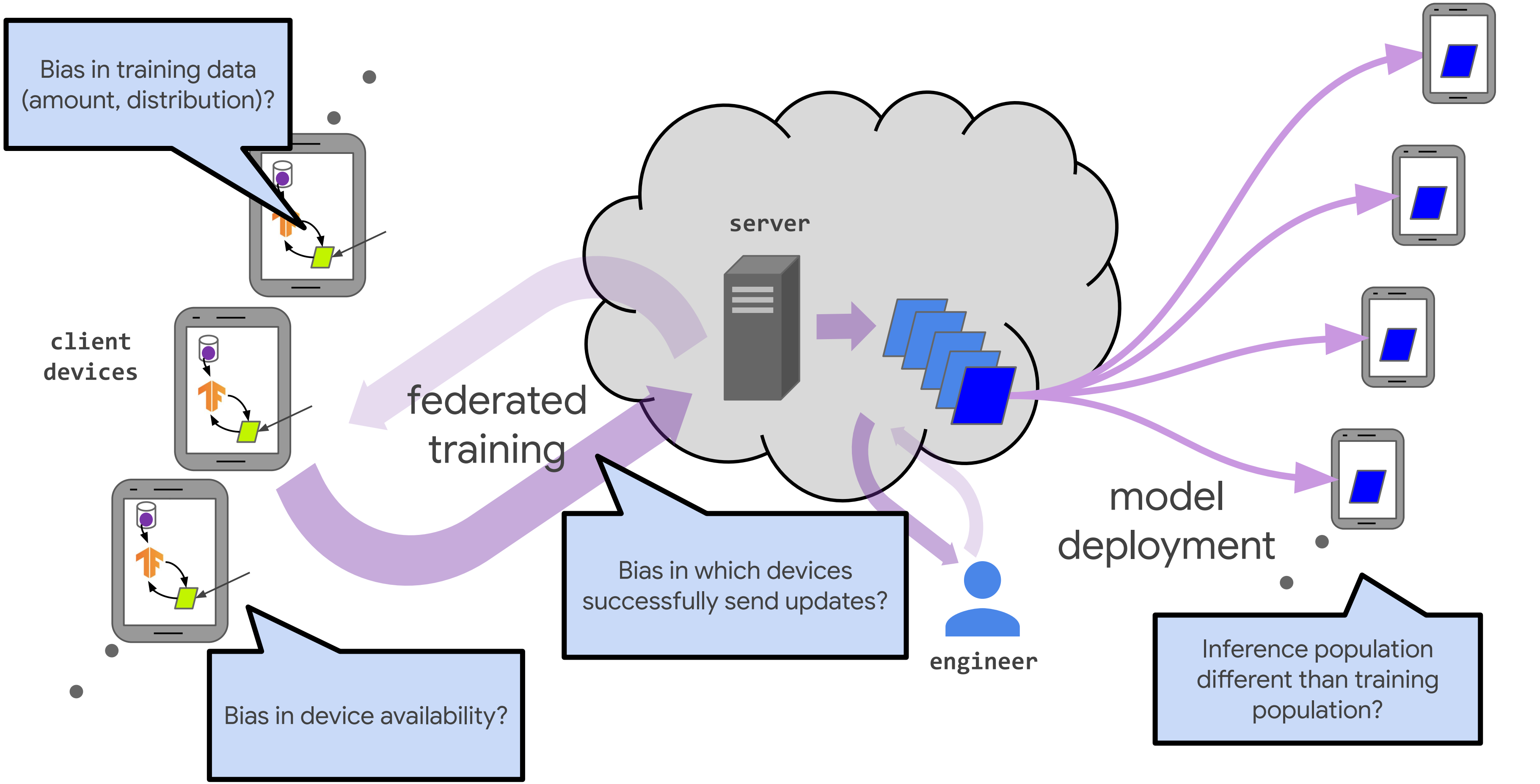
Open technical challenges in privacy

- Privacy is multifaceted
  - Need to better understand *privacy, communication, computation, accuracy, sparsity* tradeoffs
  - Tensions between privacy, robustness, and fairness are very interesting and remain underexplored – personalization may play an important role in easing the tensions
  - Cryptographic techniques will play a critical role in strengthening privacy
- Differential privacy provides an incredibly useful tool
  - But it often comes at a “hit” in accuracy
  - If we have to pay, we'd usually rather pay with more computation (not privacy or accuracy)
  - How to choose epsilon remains (and perhaps will always be) an open question
  - How to make sense of large-ish epsilons?
  - Model auditing techniques for measuring privacy loss (memorization) are complimentary
- Privacy budgeting and management systems are not available
  - Can scientists apply complex and repeated learning tasks on the same or similar datasets?
  - How do we efficiently track and quantify the privacy loss of a complex system?
- Public data is largely underutilized
  - Public data will play a key role in improving privacy-accuracy tradeoffs
  - How do we optimally combine public and private datasets during training?

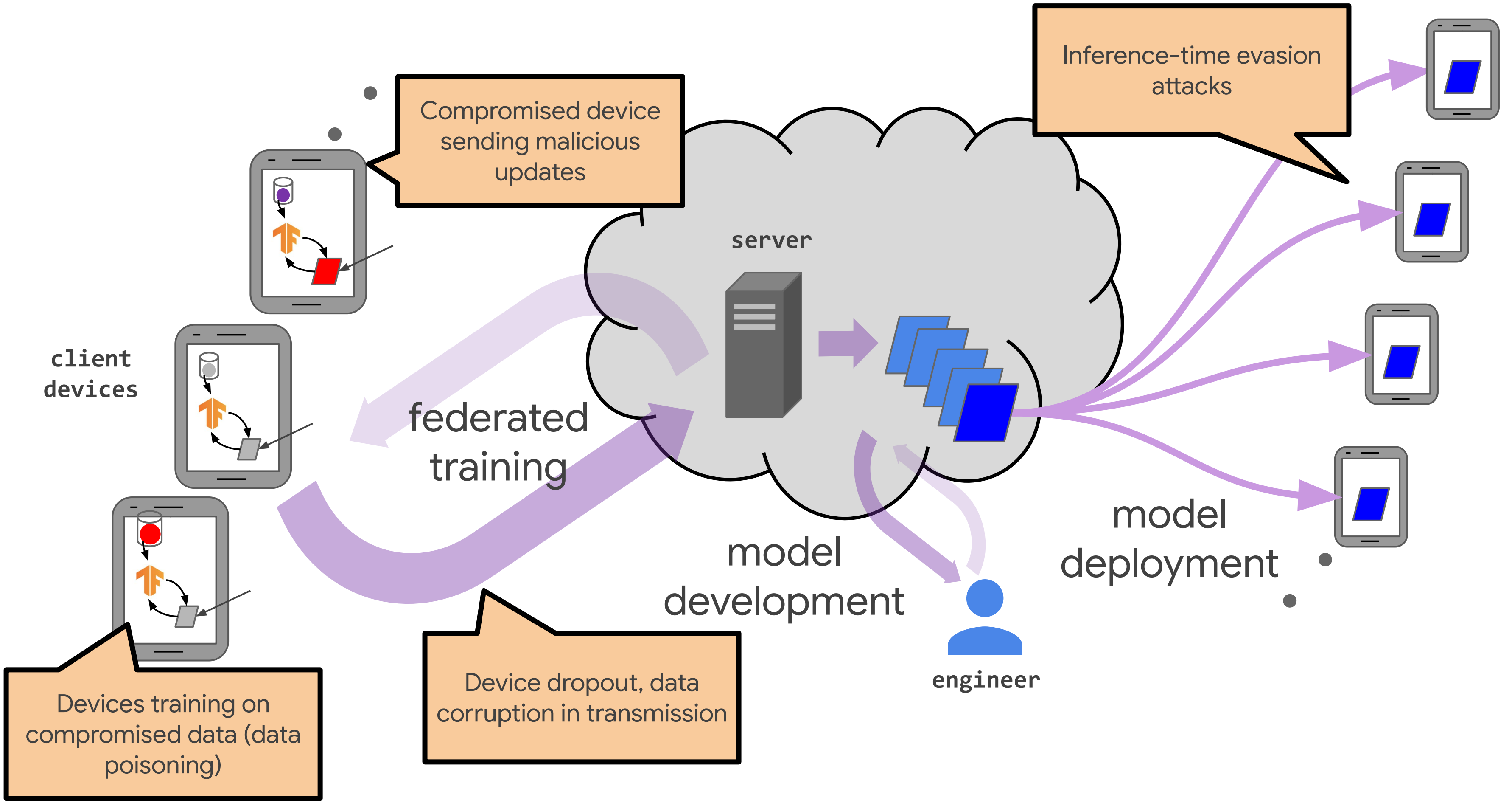
Improving efficiency and effectiveness



Ensuring fairness and addressing sources of bias



Robustness to attacks and failures



Advances and Open Problems in FL

59 authors from 25 top institutions

arxiv.org/abs/1912.04977

Foundations and Trends in Machine Learning

Advances and Open Problems in Federated Learning

Peter Kairouz[7*] H. Brendan McMahan[7*] Brendan Avent[21] Aurélien Bellet[9]
Mehdi Bennis[19] Arjun Nitin Bhagoji[13] Kallista Bonawitz[7] Zachary Charles[7]
Graham Cormode[23] Rachel Cummings[6] Rafael G.L. D'Oliveira[14]
Hubert Eichner[7] Salim El Rouayheb[14] David Evans[22] Josh Gardner[24]
Zachary Garrett[7] Adrià Gascón[7] Badih Ghazi[7] Phillip B. Gibbons[2]
Marco Gruteser[7,14] Zaid Harchaoui[24] Chaoyang He[21] Lie He [4]
Zhouyuan Huo [20] Ben Hutchinson[7] Justin Hsu[25] Martin Jaggi[4] Tara Javidi[17]
Gauri Joshi[2] Mikhail Khodak[2] Jakub Konečný[7] Aleksandra Korolova[21]
Farinaz Koushanfar[17] Sanmi Koyejo[7,18] Tancrède Lepoint[7] Yang Liu[12]
Prateek Mittal[13] Mehryar Mohri[7] Richard Nock[1] Ayfer Özgür[15]
Rasmus Pagh[7,10] Hang Qi[7] Daniel Ramage[7] Ramesh Raskar[11]
Mariana Raykova[7] Dawn Song[16] Weikang Song[7] Sebastian U. Stich[4]
Ziteng Sun[3] Ananda Theertha Suresh[7] Florian Tramèr[15] Praneeth Vepakomma[11]
Jianyu Wang[2] Li Xiong[5] Zheng Xu[7] Qiang Yang[8] Felix X. Yu[7] Han Yu[12]
Sen Zhao[7]

[1]Australian National University, [2]Carnegie Mellon University, [3]Cornell University,
[4]École Polytechnique Fédérale de Lausanne, [5]Emory University, [6]Georgia Institute of Technology,
[7]Google Research, [8]Hong Kong University of Science and Technology, [9]INRIA, [10]IT University of Copenhagen,
[11]Massachusetts Institute of Technology, [12]Nanyang Technological University, [13]Princeton University,
[14]Rutgers University, [15]Stanford University, [16]University of California Berkeley,
[17] University of California San Diego, [18]University of Illinois Urbana-Champaign, [19]University of Oulu,
[20]University of Pittsburgh, [21]University of Southern California, [22]University of Virginia,
[23]University of Warwick, [24]University of Washington, [25]University of Wisconsin–Madison

Abstract

Federated learning (FL) is a machine learning setting where many clients (e.g. mobile devices or whole organizations) collaboratively train a model under the orchestration of a central server (e.g. service provider), while keeping the training data decentralized. FL embodies the principles of focused data collection and minimization, and can mitigate many of the systemic privacy risks and costs resulting from traditional, centralized machine learning and data science approaches. Motivated by the explosive growth in FL research, this paper discusses recent advances and presents an extensive collection of open problems and challenges.

A Field Guide to Federated Optimization

Jianyu Wang[*1], Zachary Charles[*3], Zheng Xu[*3], Gauri Joshi[*1], H. Brendan McMahan[*3], Blaise Agüera y Arcas[3], Maruan Al-Shedivat[1], Galen Andrew[3], Salman Avestimehr[13], Katharine Daly[3], Deepesh Data[9], Suhas Diggavi[9], Hubert Eichner[3], Advait Gadhikar[1], Zachary Garrett[3], Antonious M. Girgis[9], Filip Hanzely[8], Andrew Hard[3], Chaoyang He[13], Samuel Horváth[4], Zhouyuan Huo[3], Alex Ingerman[3], Martin Jaggi[2], Tara Javidi[10], Peter Kairouz[3], Satyen Kale[3], Sai Praneeth Karimireddy[2], Jakub Konečný[3], Sanmi Koyejo[11], Tian Li[1], Luyang Liu[3], Mehryar Mohri[3], Hang Qi[3], Sashank J. Reddi[3], Peter Richtárik[4], Karan Singhal[3], Virginia Smith[1], Mahdi Soltanolkotabi[13], Weikang Song[3], Ananda Theertha Suresh[3], Sebastian U. Stich[2], Ameet Talwalkar[1], Hongyi Wang[14], Blake Woodworth[8], Shanshan Wu[3], Felix X. Yu[3], Honglin Yuan[6], Manzil Zaheer[3], Mi Zhang[5], Tong Zhang[3,7], Chunxiang Zheng[3], Chen Zhu[12], and Wennan Zhu[3]

[1]*Carnegie Mellon University*, [2]*École Polytechnique Fédérale de Lausanne*, [3]*Google Research*, [4]*King Abdullah University of Science and Technology*, [5]*Michigan State Univeristy*, [6]*Stanford University*, [7]*The Hong Kong University of Science and Technology*, [8]*Toyota Technological Institute at Chicago*, [9]*University of California, Los Angeles*, [10]*University of California, San Diego*, [11]*University of Illinois Urbana-Champaign*, [12]*University of Maryland, College Park*, [13]*University of Southern California*, [14]*University of Wisconsin–Madison*

Abstract

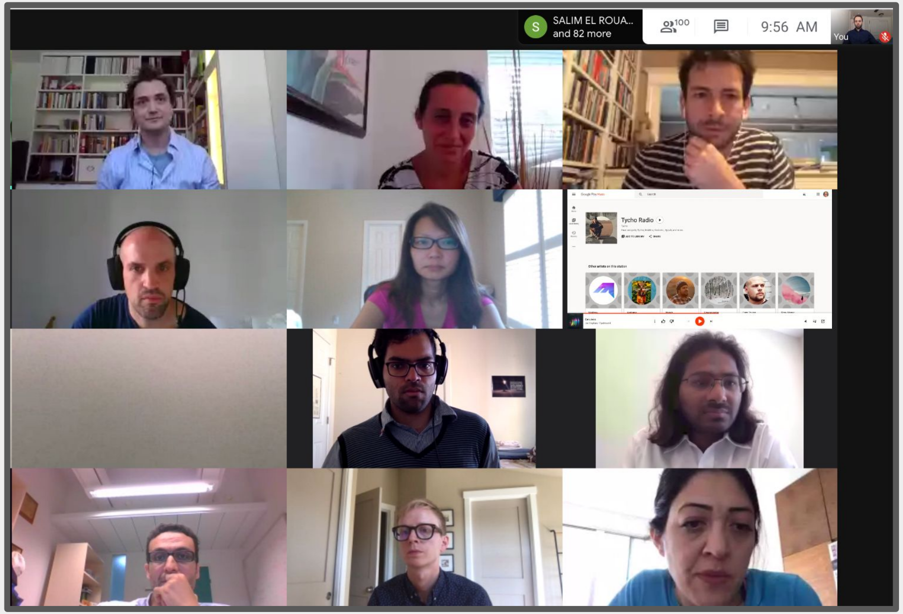
Federated learning and analytics are a distributed approach for collaboratively learning models (or statistics) from decentralized data, motivated by and designed for privacy protection. The distributed learning process can be formulated as solving federated optimization problems, which emphasize communication efficiency, data heterogeneity, compatibility with privacy and system requirements, and other constraints that are not primary considerations in other problem settings. This paper provides recommendations and guidelines on formulating, designing, evaluating and analyzing federated optimization algorithms through concrete examples and practical implementation, with a focus on conducting effective simulations to infer real-world performance. The goal of this work is not to survey the current literature, but to inspire researchers and practitioners to design federated learning algorithms that can be used in various practical applications.

A Field Guide to Federated Optimization

53 authors from 14 top institutions

arxiv.org/abs/2107.06917

Tensorflow Federated Implementation



Thank you for your time!

Twitter: @KairouzPeter