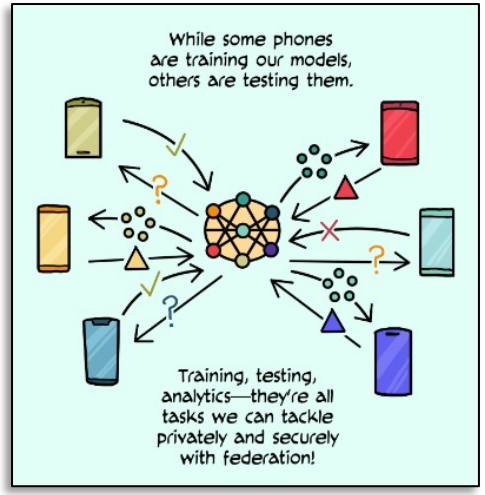
Trustworthy AI for Business and Society

Ilker Birbil

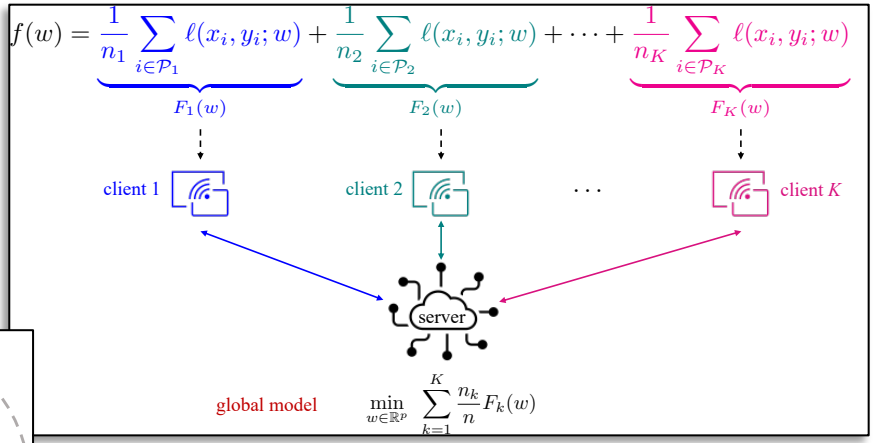
Federated Learning

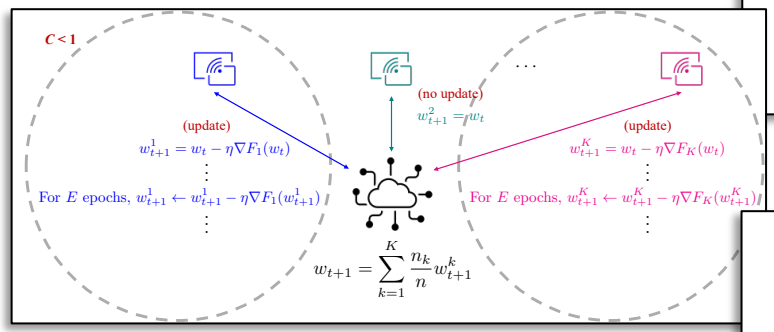
Big Picture

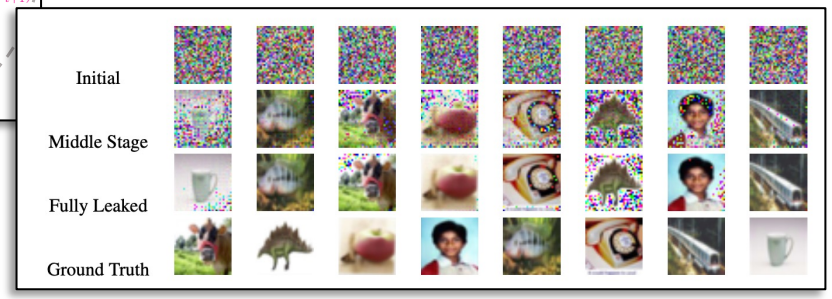
- Motivation and Setup
- Federated Averaging
- Challenges: Privacy
- Secure Sum - A Game

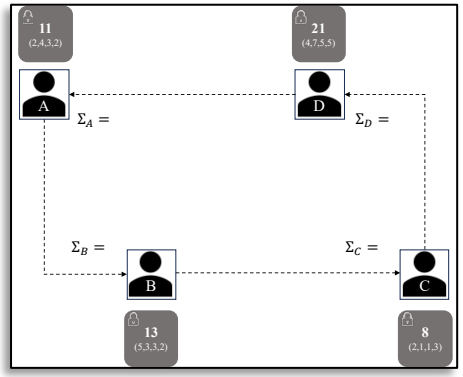


(link)



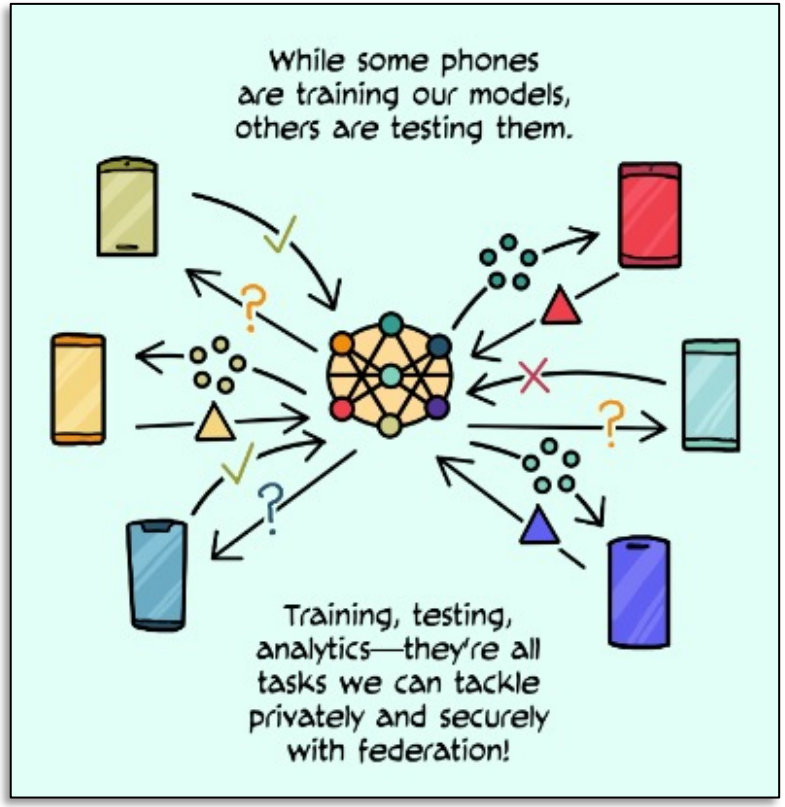






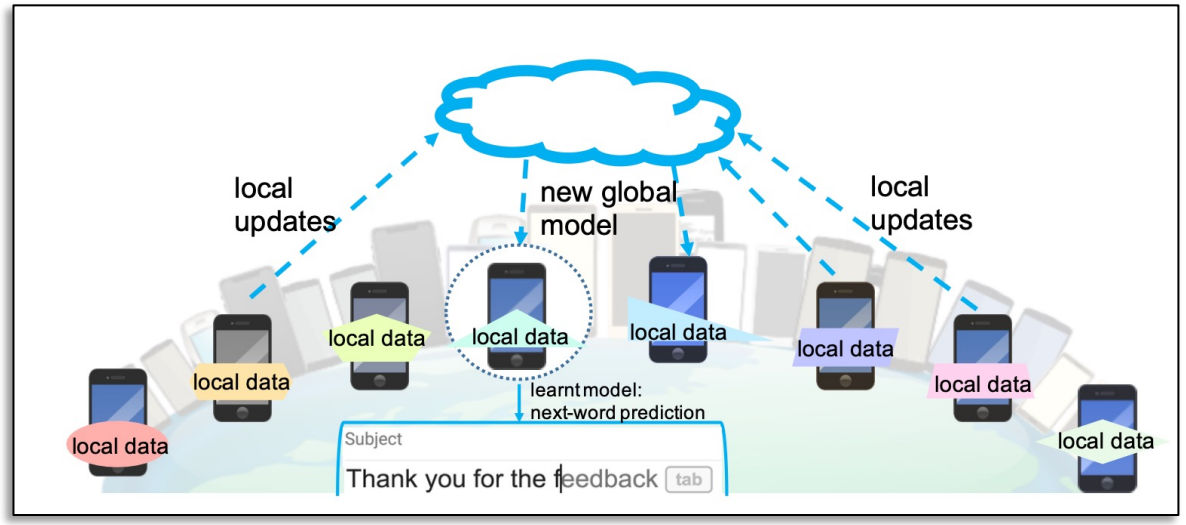
(link)

Motivation



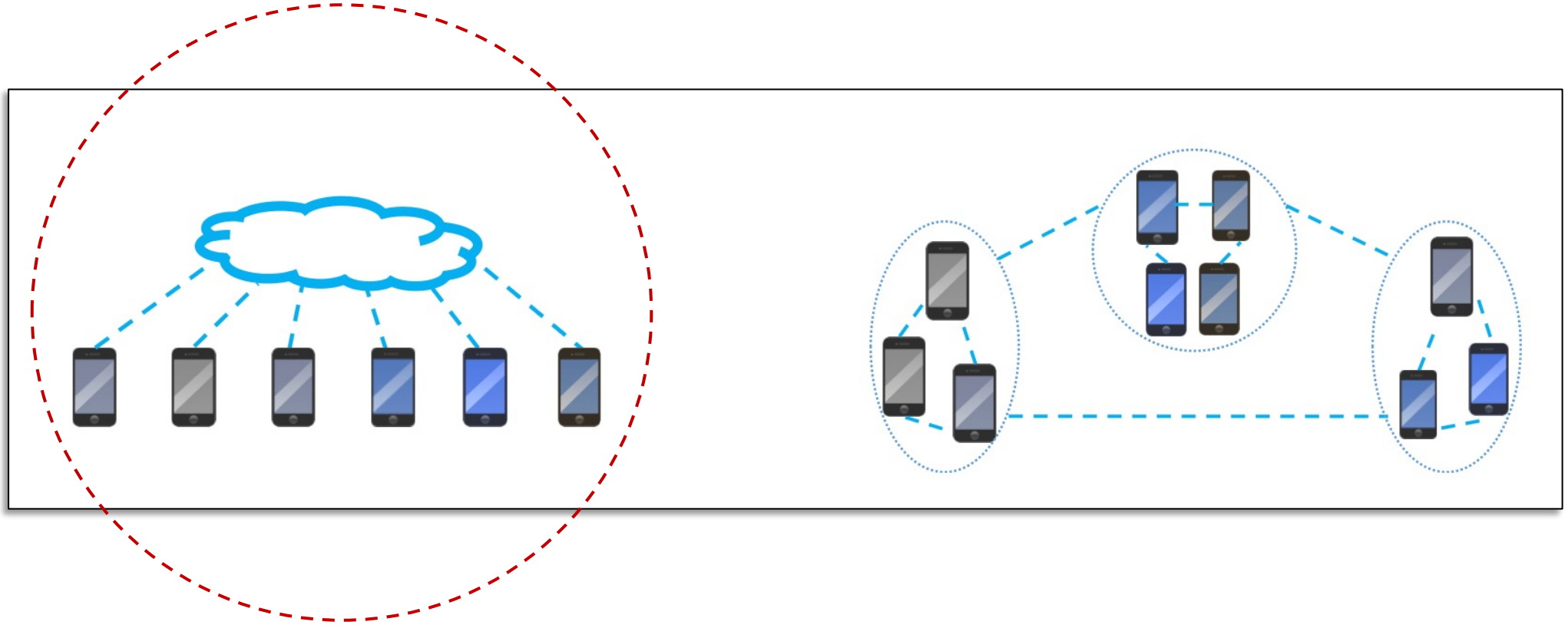
(link)

(link)



(link)

Centralized vs. Decentralized



Li, T., Sahu, A.K. Talwalker, A. & Smith, V. "Federated learning: challenges, methods, and future directions," arXiv:1908.07873, 2019.

Setup

$$\{(x_i, y_i) : \underbrace{i = 1, \ldots, n}_{i \in \mathcal{I}}\}$$

$$x_i \in \mathbb{R}^p$$

$$\min_{w \in \mathbb{R}^p} \underbrace{\frac{1}{n} \sum_{i=1}^{n} \ell(x_i, y_i; w)}_{f(w)}$$

loss function

Data Partitioning

$$\bigcup_{k=1}^{K} \mathcal{P}_k = \mathcal{I} \quad \text{and} \quad \mathcal{P}_s \cap \mathcal{P}_t = \emptyset, s \neq t$$

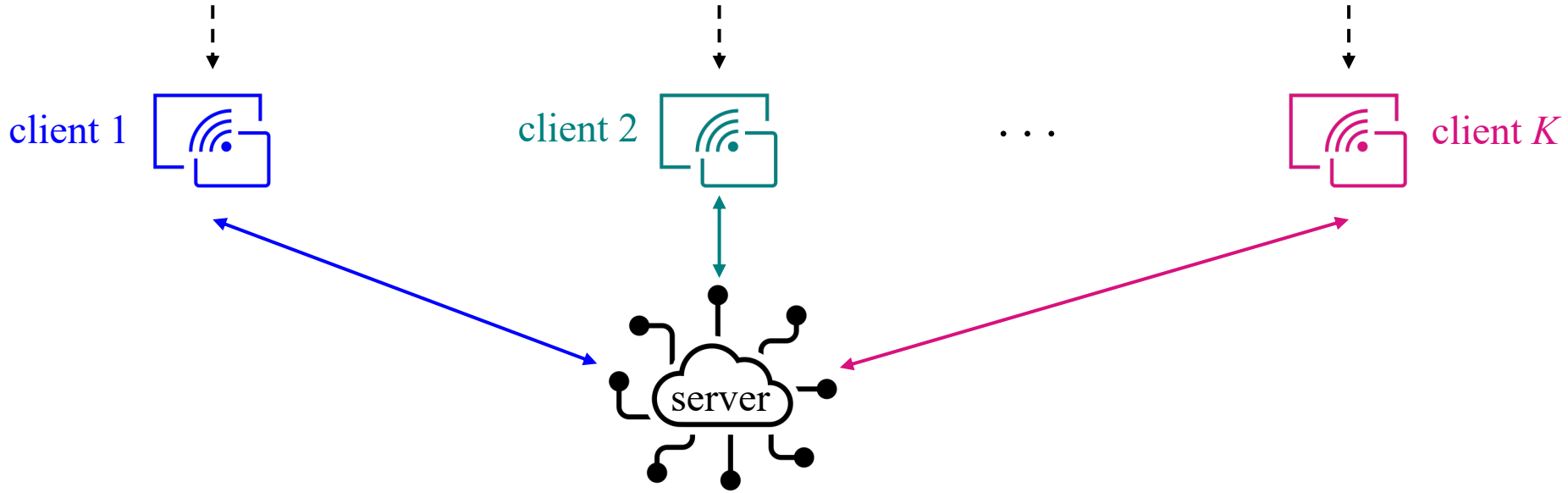
$$n_k = |\mathcal{P}_k|, \quad k = 1, \ldots, K$$

$$n = n_1 + n_2 + \cdots + n_K$$

$$f(w) = \underbrace{\frac{1}{n_1} \sum_{i \in \mathcal{P}_1} \ell(x_i, y_i; w)}_{F_1(w)} + \underbrace{\frac{1}{n_2} \sum_{i \in \mathcal{P}_2} \ell(x_i, y_i; w)}_{F_2(w)} + \cdots + \underbrace{\frac{1}{n_K} \sum_{i \in \mathcal{P}_K} \ell(x_i, y_i; w)}_{F_K(w)}$$

Setup

$$f(w) = \underbrace{\frac{1}{n_1}\sum_{i\in\mathcal{P}_1}\ell(x_i, y_i; w)}_{F_1(w)} + \underbrace{\frac{1}{n_2}\sum_{i\in\mathcal{P}_2}\ell(x_i, y_i; w)}_{F_2(w)} + \cdots + \underbrace{\frac{1}{n_K}\sum_{i\in\mathcal{P}_K}\ell(x_i, y_i; w)}_{F_K(w)}$$

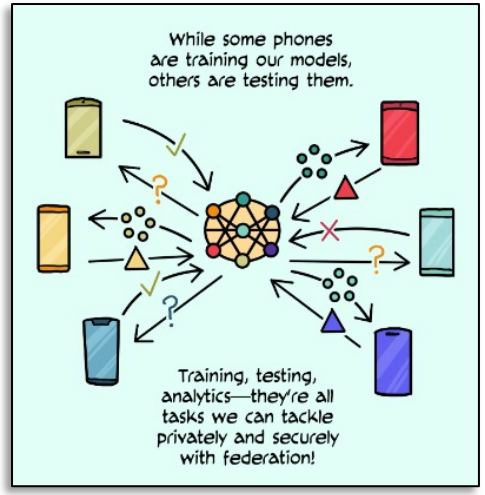


global model $\quad \min_{w\in\mathbb{R}^p} \sum_{k=1}^{K} \frac{n_k}{n} F_k(w)$

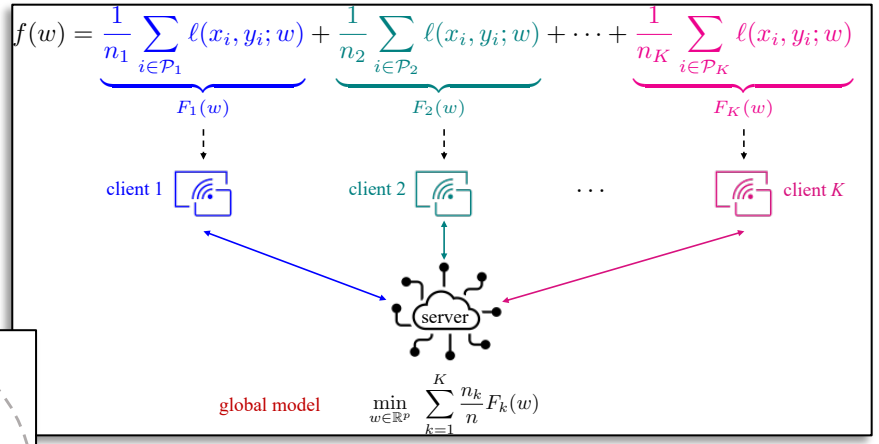
How to train the global model at the server without sending clients' raw data?

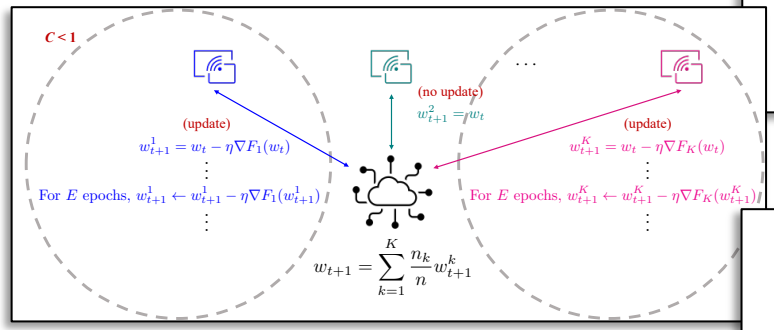
Big Picture

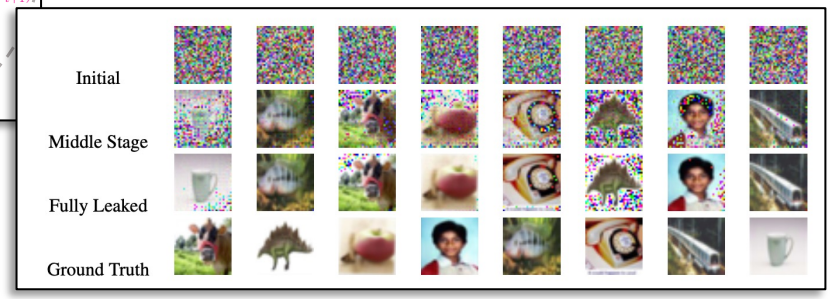
- Motivation and Setup
- Federated Averaging
- Challenges: Privacy
- Secure Sum - A Game

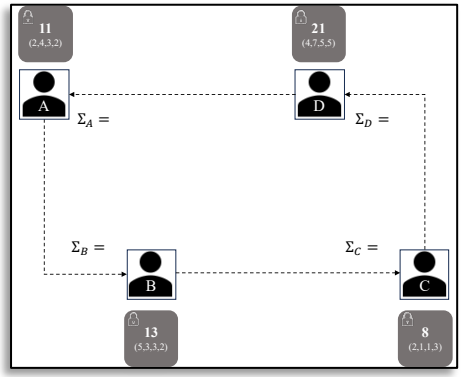


(link)







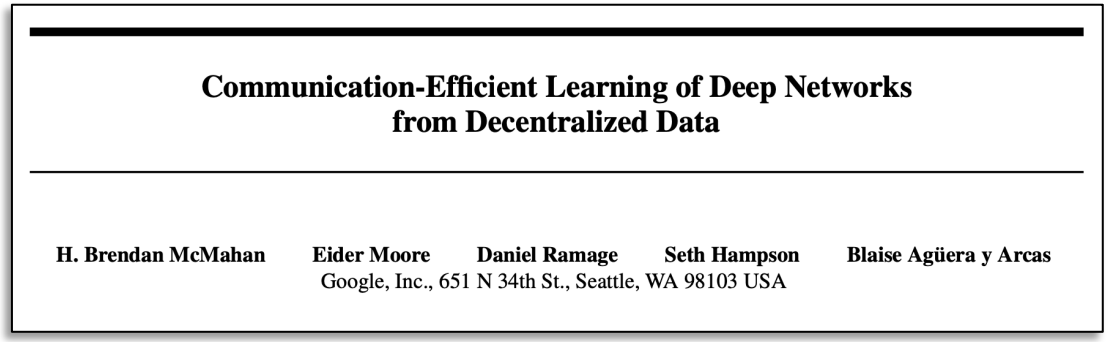


(link)

Federated Averaging

"A principal advantage of this approach is the decoupling of model training from the need for direct access to the raw training data. Clearly, some trust of the server coordinating the training is still required."

"More concretely, we introduce the *Federated Averaging* algorithm, which combines local stochastic gradient descent (SGD) on each client with a server that performs model averaging."



Communication-Efficient Learning of Deep Networks from Decentralized Data

H. Brendan McMahan **Eider Moore** **Daniel Ramage** **Seth Hampson** **Blaise Agüera y Arcas**
Google, Inc., 651 N 34th St., Seattle, WA 98103 USA

(link)

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w)$$

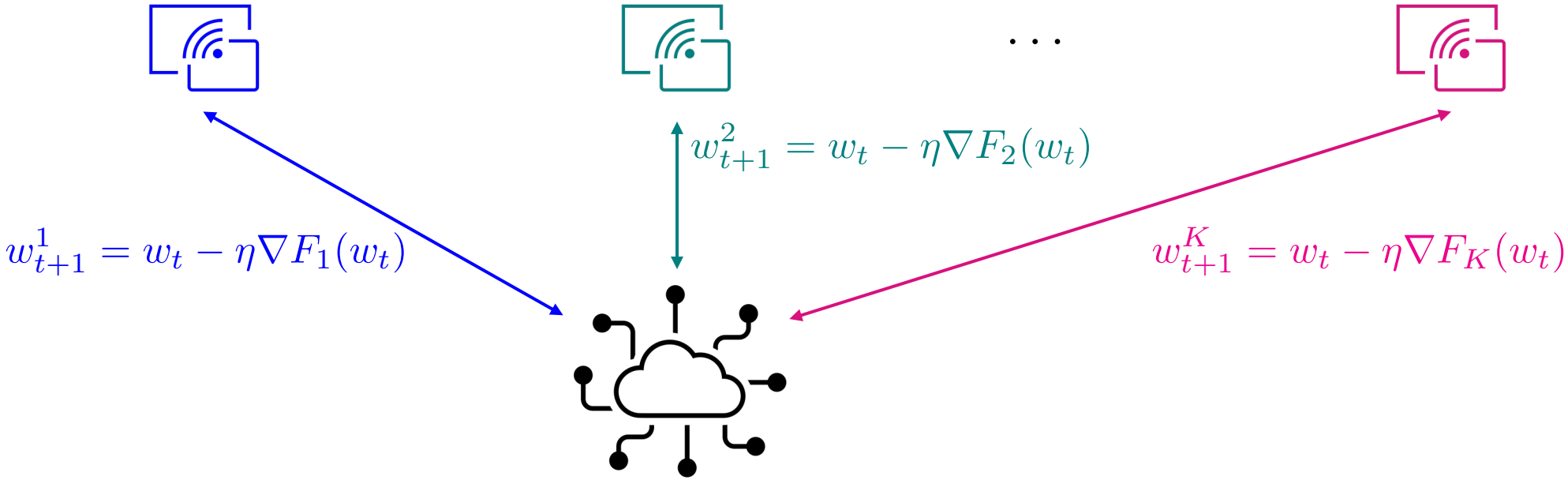
$$\nabla f(w) = \sum_{k=1}^{K} \frac{n_k}{n} \nabla F_k(w)$$

$$w_{t+1} = w_t - \eta \nabla f(w) = w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} \nabla F_k(w) \qquad \text{(full batch at iteration } t\text{)}$$

Federated Averaging

C : fraction of clients updating (**global batch size**)

***C* = 1**

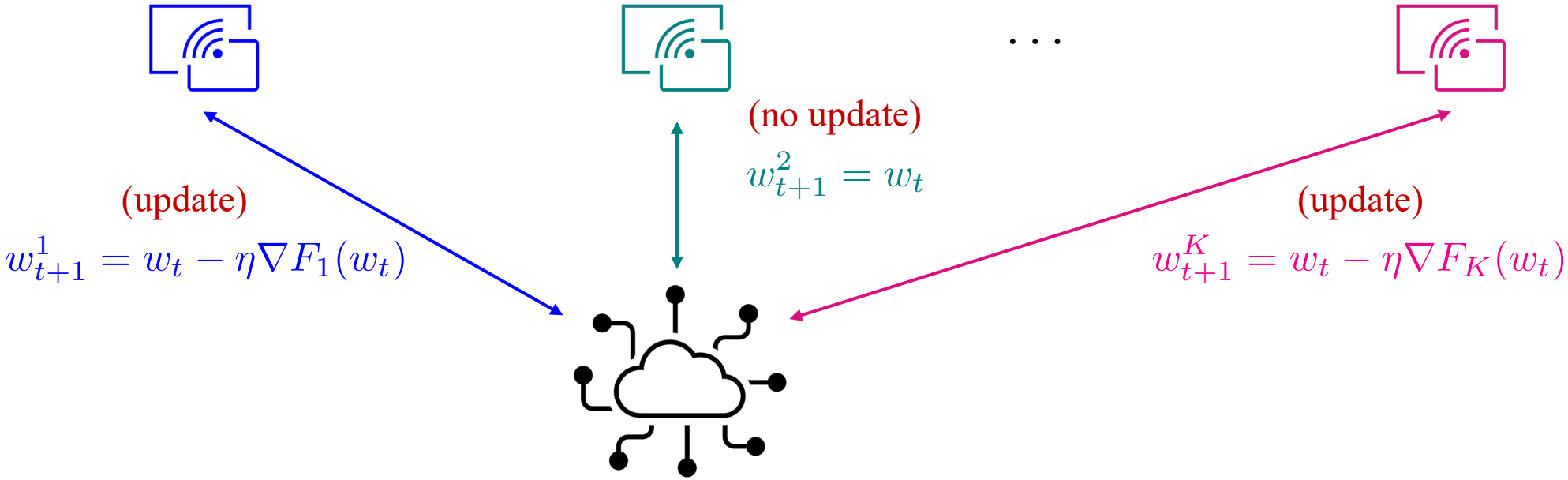


$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

Federated Averaging

C : fraction of clients updating (**global batch size**)

***C* < 1**



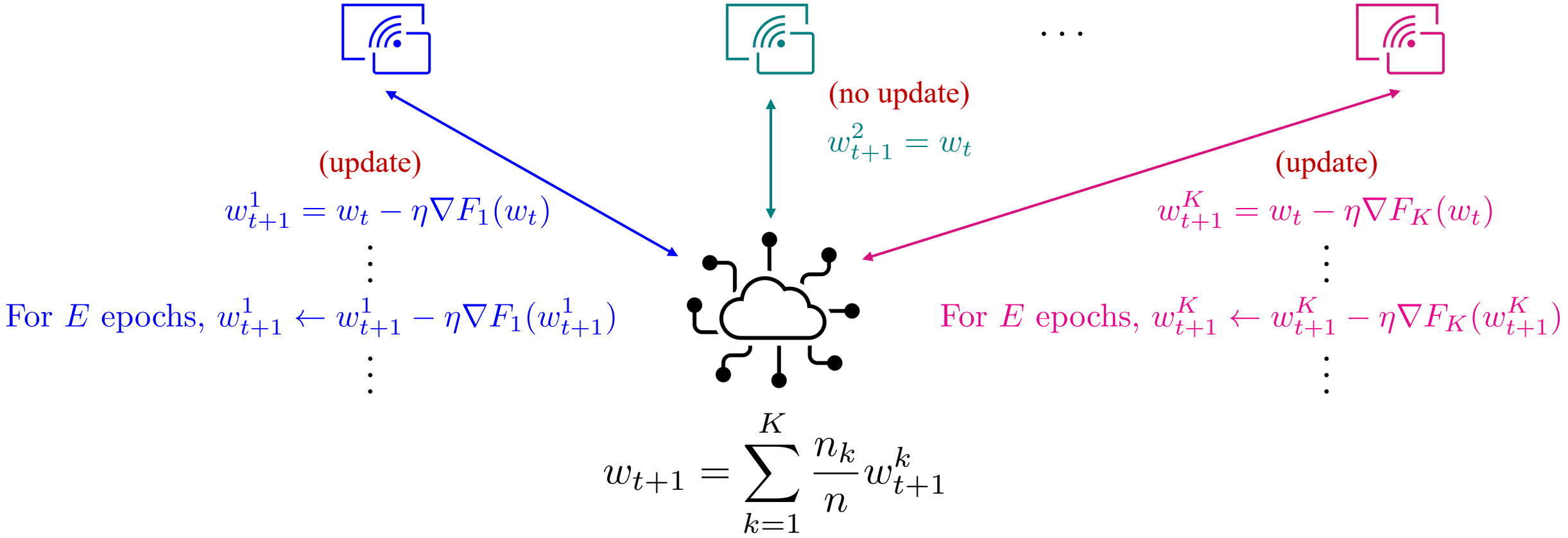
$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

Federated Averaging

C : fraction of clients updating (**global batch size**)

E : number of passes over the local data (**local epochs**)

***C* < 1**



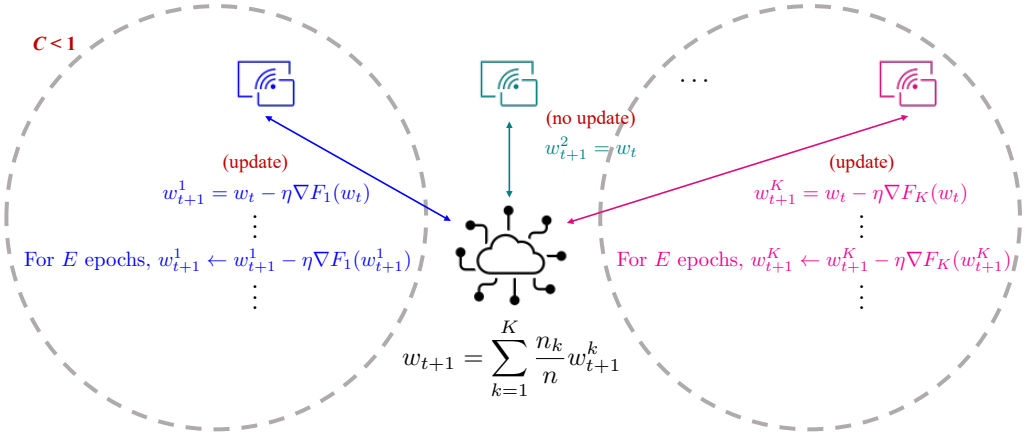
$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

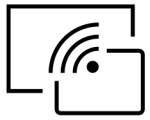
Federated Averaging

C : fraction of clients updating (**global batch size**)

E : number of passes over the local data (**local epochs**)

$\mathcal{B}$: splitting of $\mathcal{P}_k$ into batches of size B **(local batch size)**





(update)

$$w^k_{t+1} = w_t - \eta \nabla F_k(w_t)$$

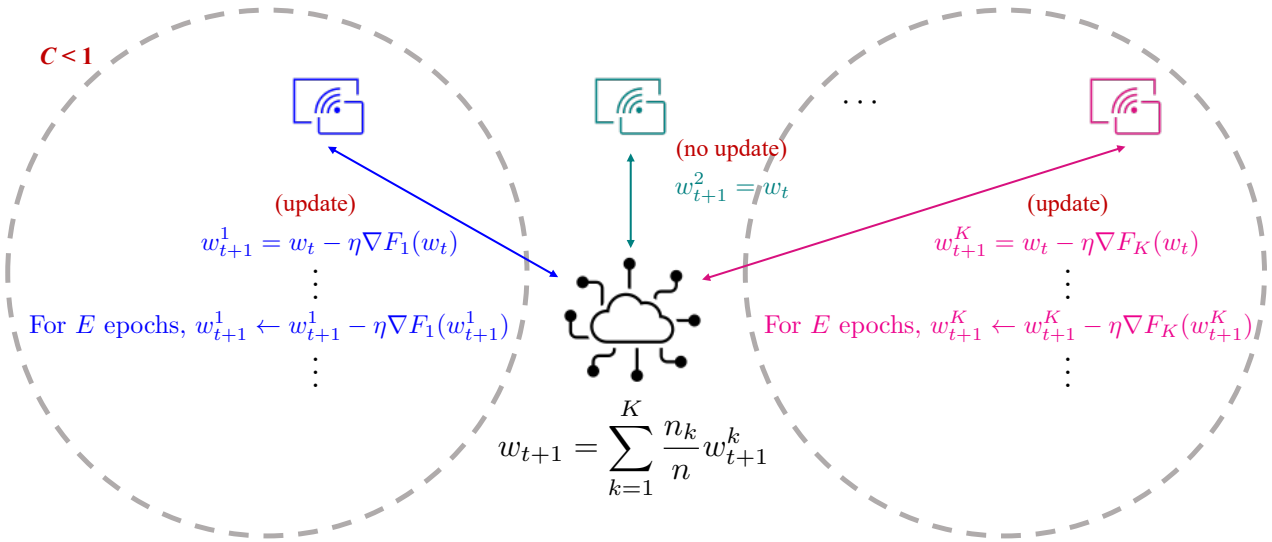
$$\vdots$$

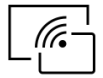
For E epochs,

For $\mathcal{S} \subseteq \mathcal{B}$, $w^k_{t+1} \leftarrow w^k_{t+1} - \eta \sum_{i \in \mathcal{S}} \nabla \ell(x_i, y_i; w^k_{t+1})$

$$\vdots$$

Federated Averaging





(update)

$$w_{t+1}^k = w_t - \eta \nabla F_k(w_t)$$

$$\vdots$$

For E epochs,

$$\text{For } \mathcal{S} \subseteq \mathcal{B},\ w_{t+1}^k \leftarrow w_{t+1}^k - \eta \sum_{i \in \mathcal{S}} \nabla \ell(x_i, y_i; w_{t+1}^k)$$

$$\vdots$$

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:

initialize w_0

for each round $t = 1, 2, \ldots$ **do**

  $m \leftarrow \max(C \cdot K, 1)$

  $S_t \leftarrow$ (random set of m clients)

  for each client $k \in S_t$ **in parallel do**

    $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$

  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w): *// Run on client k*

$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)

for each local epoch i from 1 to E **do**

  for batch $b \in \mathcal{B}$ **do**

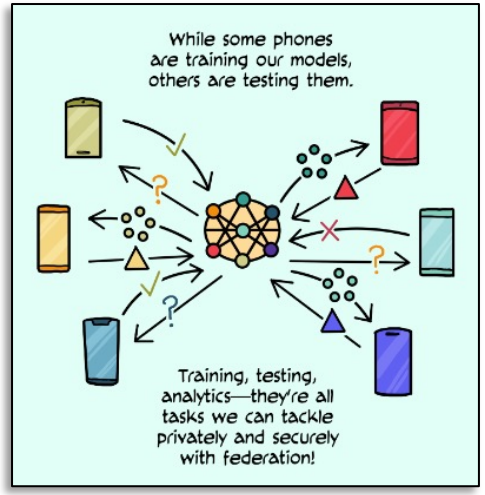
    $w \leftarrow w - \eta \nabla \ell(w; b)$

return w to server

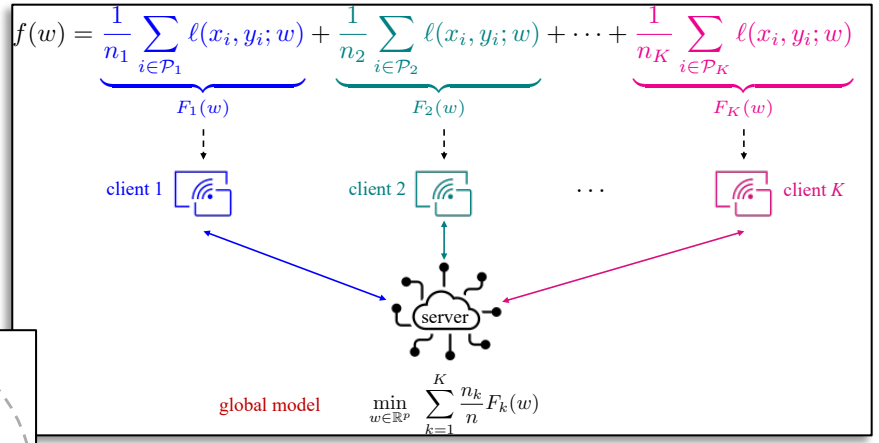
(link)

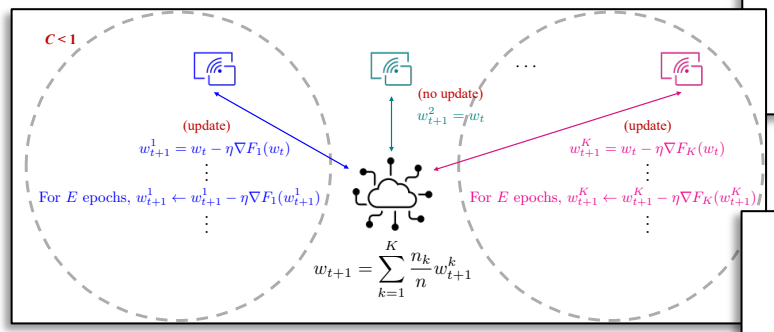
Big Picture

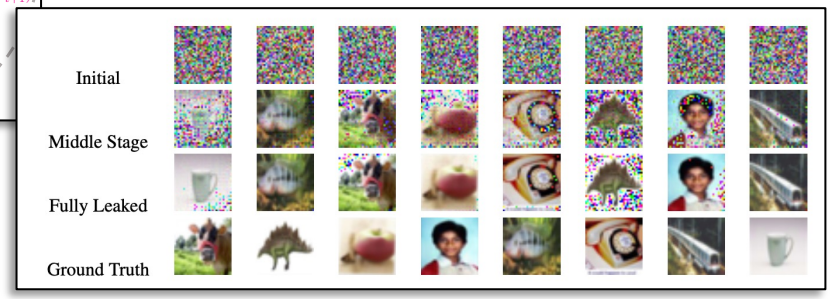
- Motivation and Setup
- Federated Averaging
- Challenges: Privacy
- Secure Sum - A Game

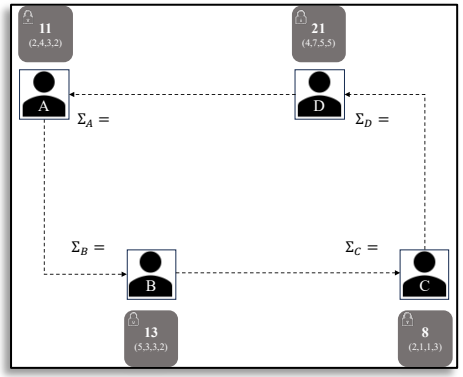


(link)





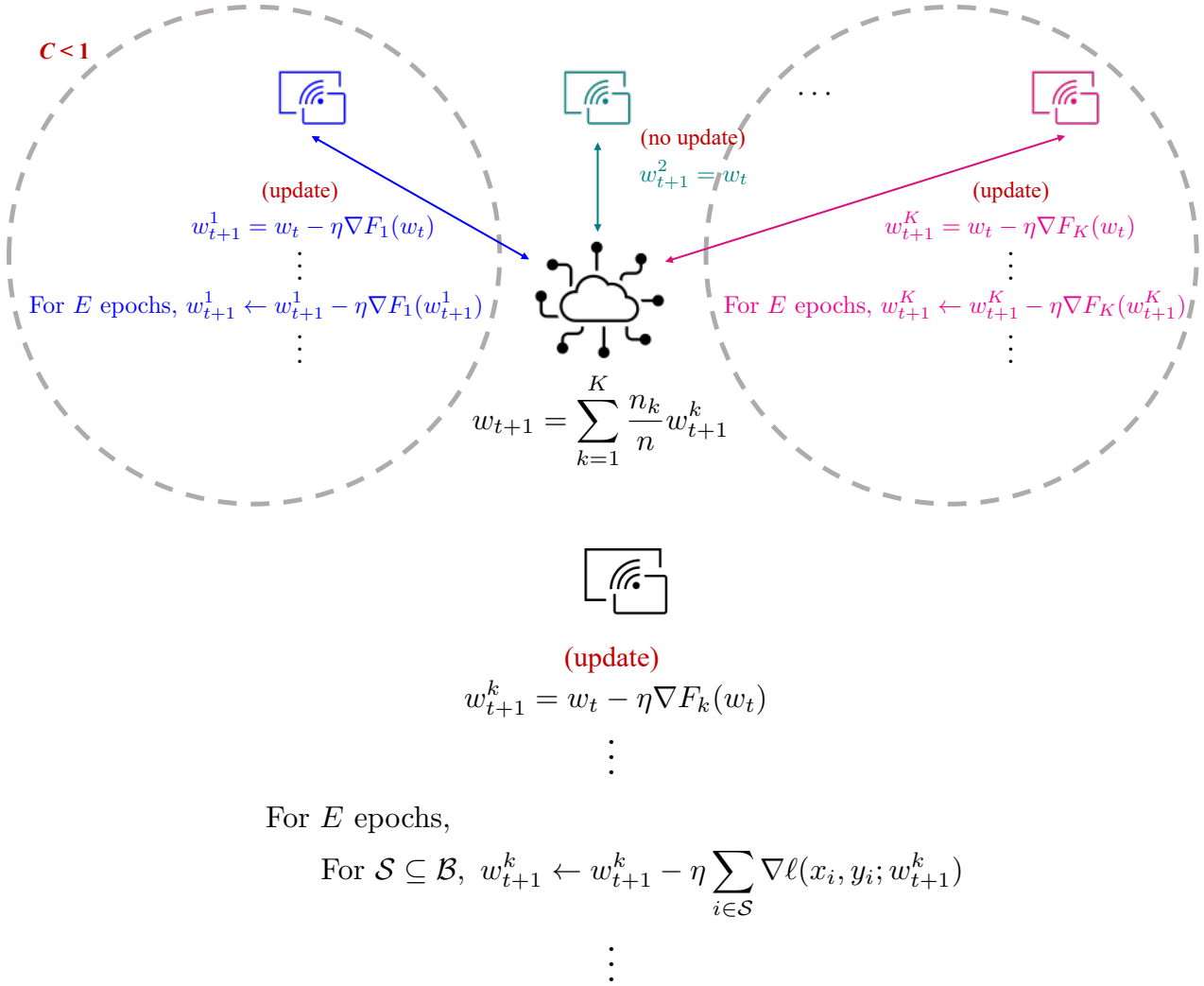




(link)

Challenges

- Communication cost
- Heterogenous network
- Data privacy
- Fairness
- Personalization



Privacy

WHEN YOU TRAIN PREDICTIVE MODELS ON INPUT FROM YOUR USERS, IT CAN LEAK INFORMATION IN UNEXPECTED WAYS.

(link)

The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks

Nicholas Carlini[1,2] Chang Liu[2] Úlfar Erlingsson[1] Jernej Kos[3] Dawn Song[2]

[1]*Google Brain* [2]*University of California, Berkeley* [3]*National University of Singapore*

(link)

"Concretely, disclosure of secrets may arise naturally in generative text models like those used for text auto-completion and predictive keyboards, if trained on possibly-sensitive data."

"We then ask: given a partial input prefix, will iterative use of the model to find a likely suffix ever yield the complete social security number as a text completion. We find the answer to our question to be an emphatic "Yes!" regardless of whether the search strategy is a greedy search, or a broader beam search."

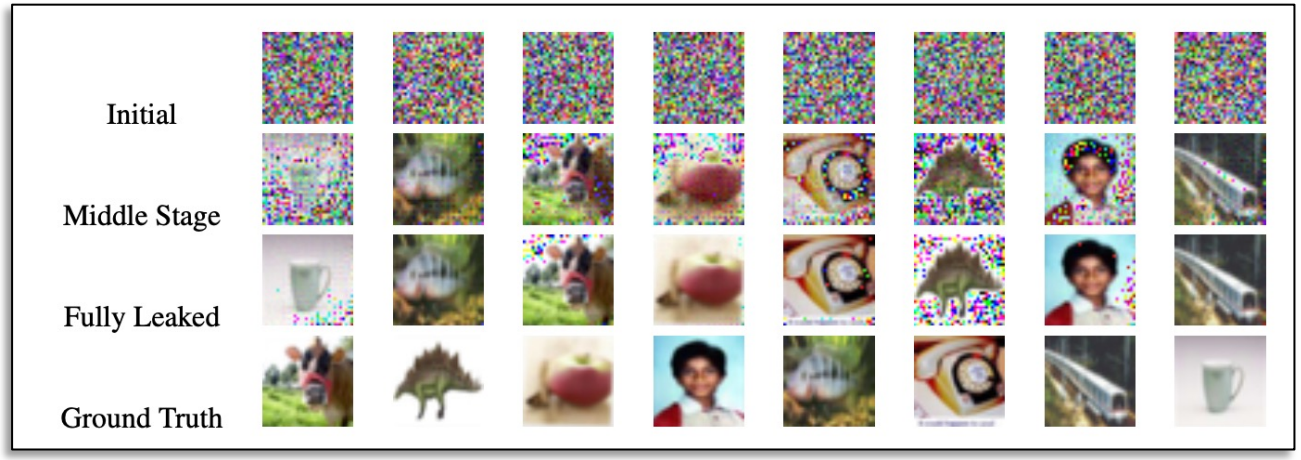
Privacy

"The deep leakage puts a severe challenge to the multi-node machine learning system. The fundamental gradient sharing scheme, as shown in our work, is not always reliable to protect the privacy of the training data."

Deep Leakage from Gradients

Ligeng Zhu Zhijian Liu Song Han
Massachusetts Institute of Technology
{ligeng, zhijian, songhan}@mit.edu

(link)

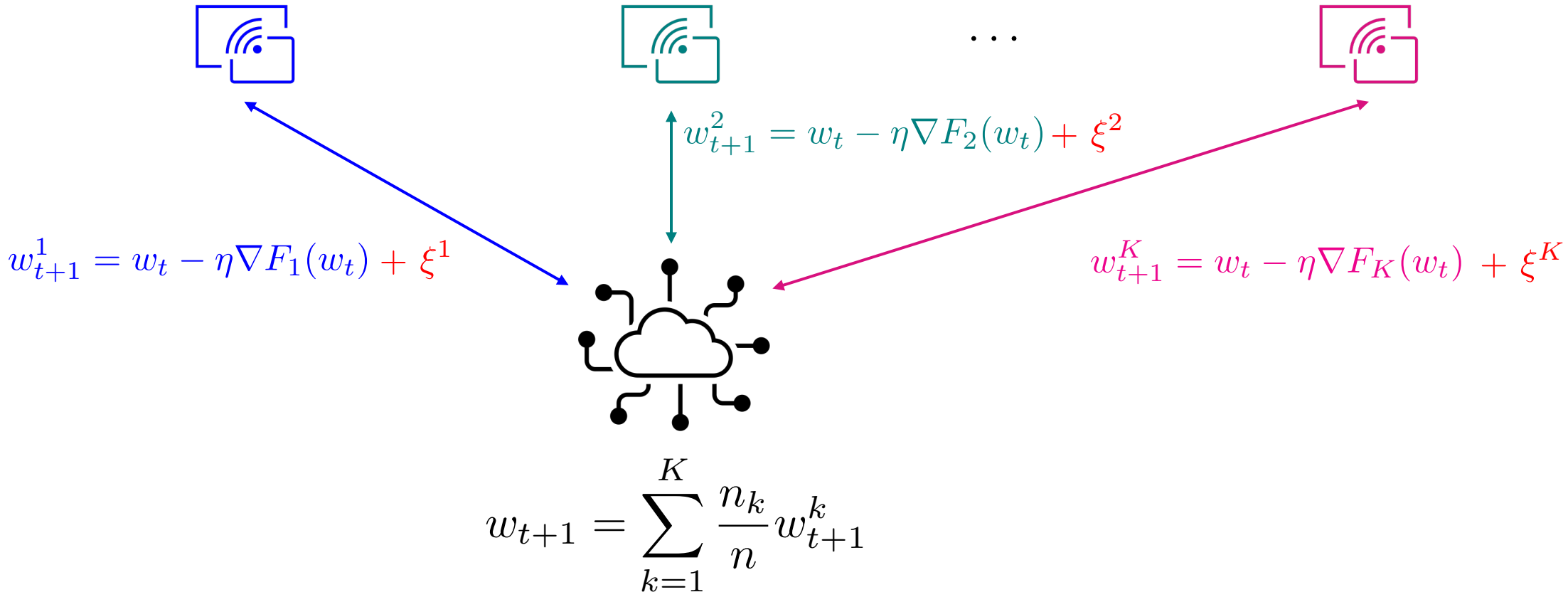


(link)

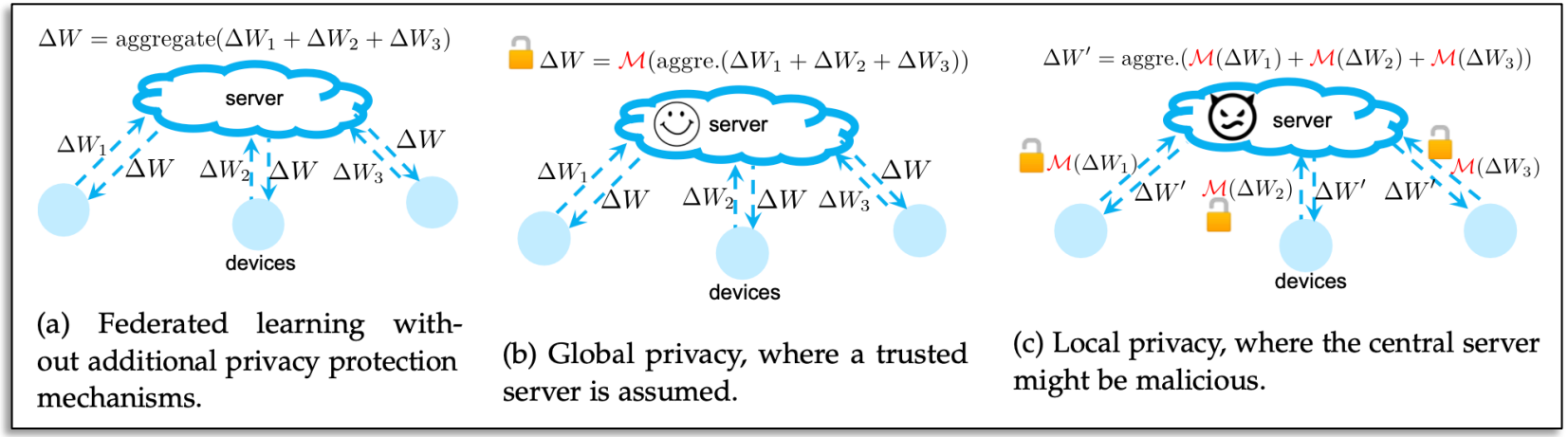
"To prevent the deep leakage, we demonstrate three defense strategies: gradient perturbation, low precision, and gradient compression. For gradient perturbation, we find both Gaussian and Laplacian noise with a scale higher than 10^{-2} would be a good defense. While half precision fails to protect, gradient compression successfully defends the attack with the pruned gradient is more than 20%."

Differential Privacy - Teaser

(more in the next lecture)



Local vs. Global Privacy



(a) Federated learning without additional privacy protection mechanisms.

(b) Global privacy, where a trusted server is assumed.

(c) Local privacy, where the central server might be malicious.

([link])

Differential Privacy

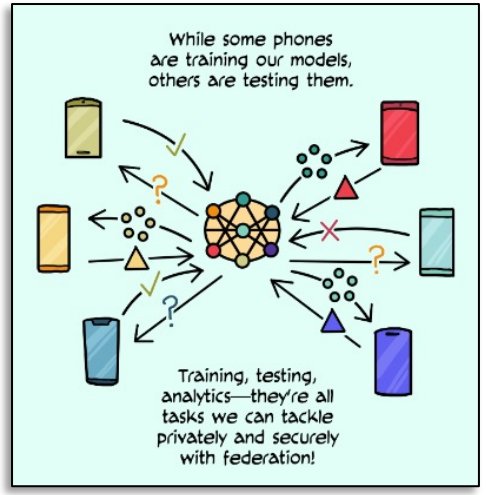
Secure Multiparty Computation

Encryption Methods

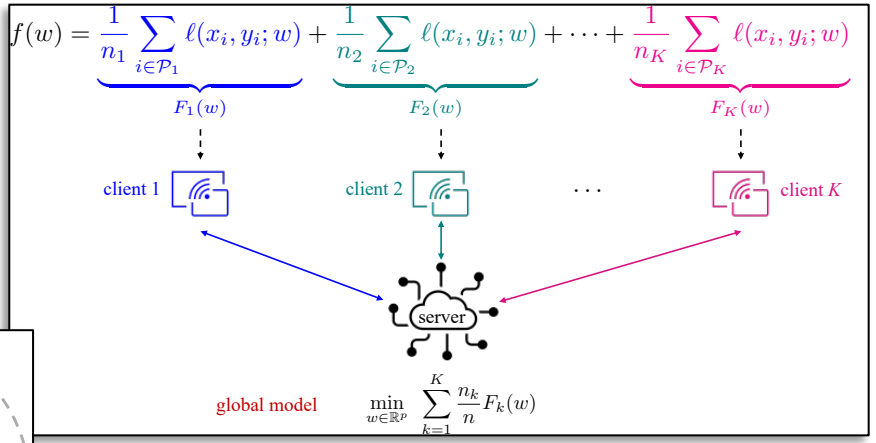
...

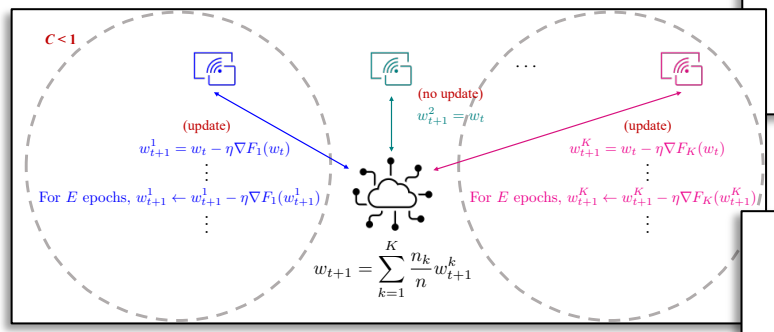
Big Picture

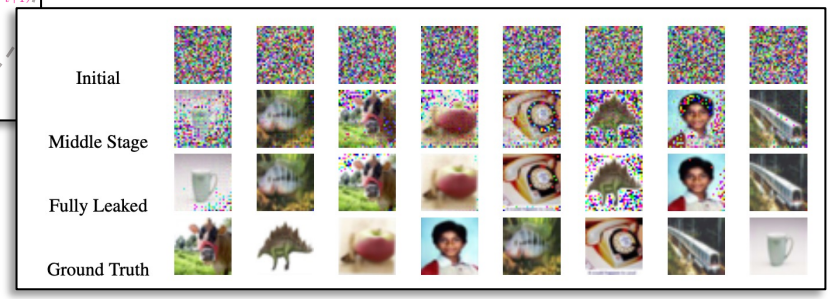
- Motivation and Setup
- Federated Averaging
- Challenges: Privacy
- Secure Sum - A Game

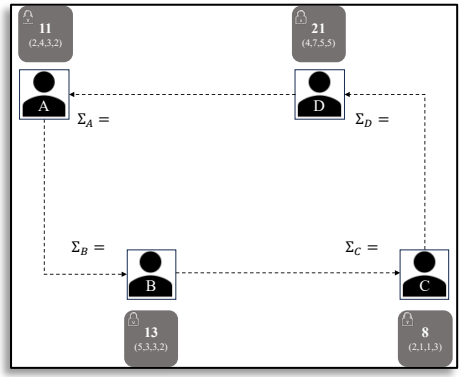


(link)



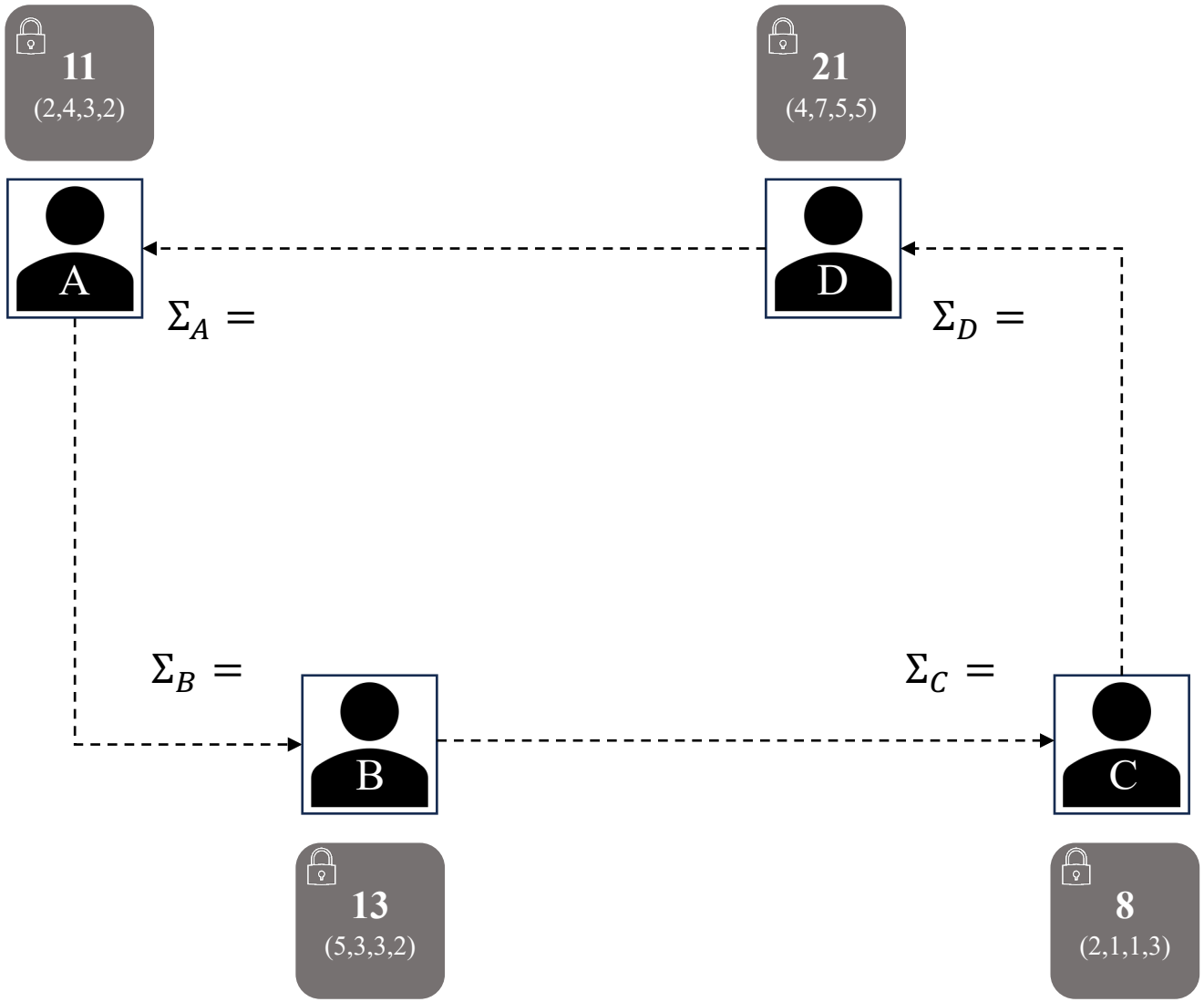




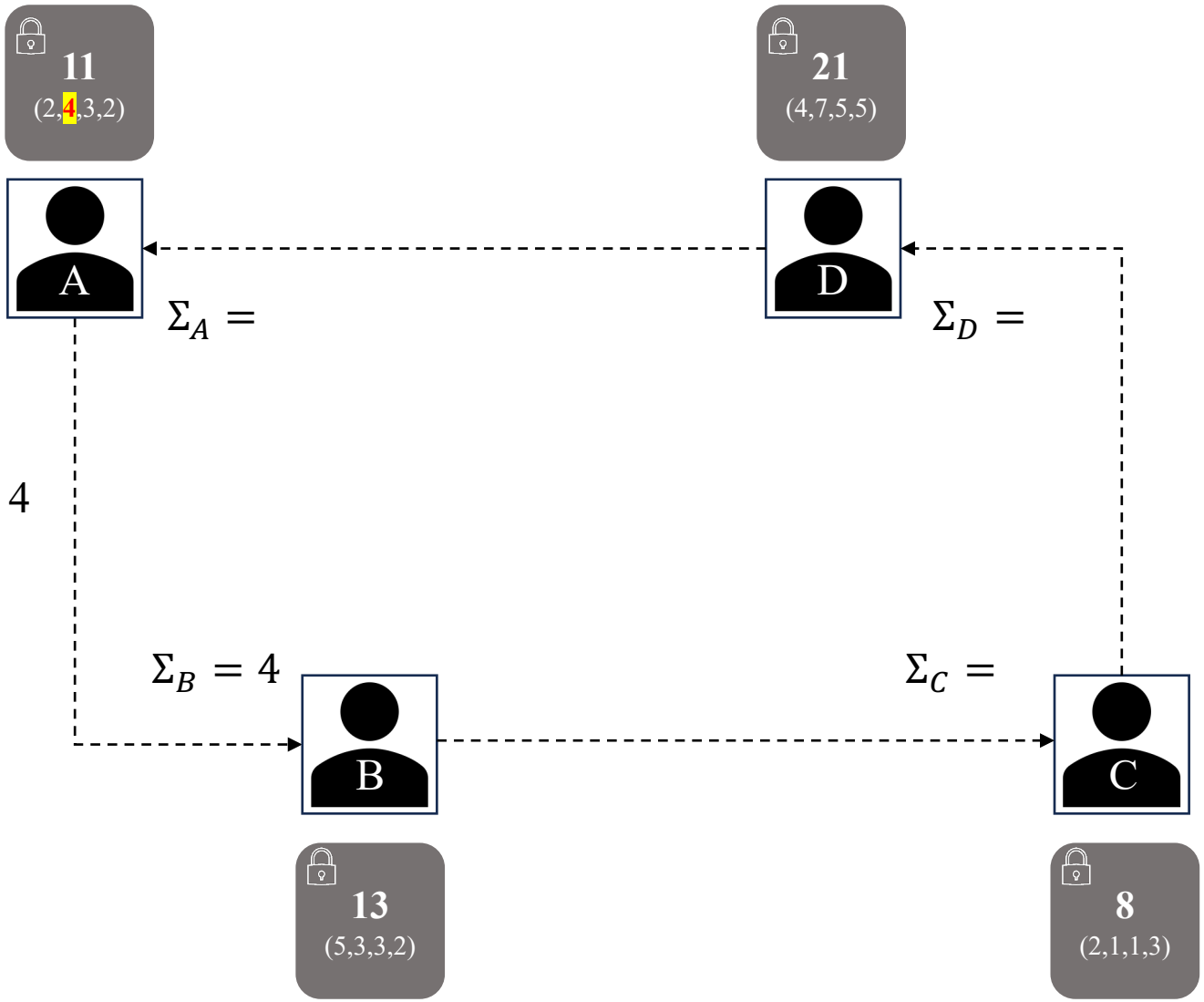


(link)

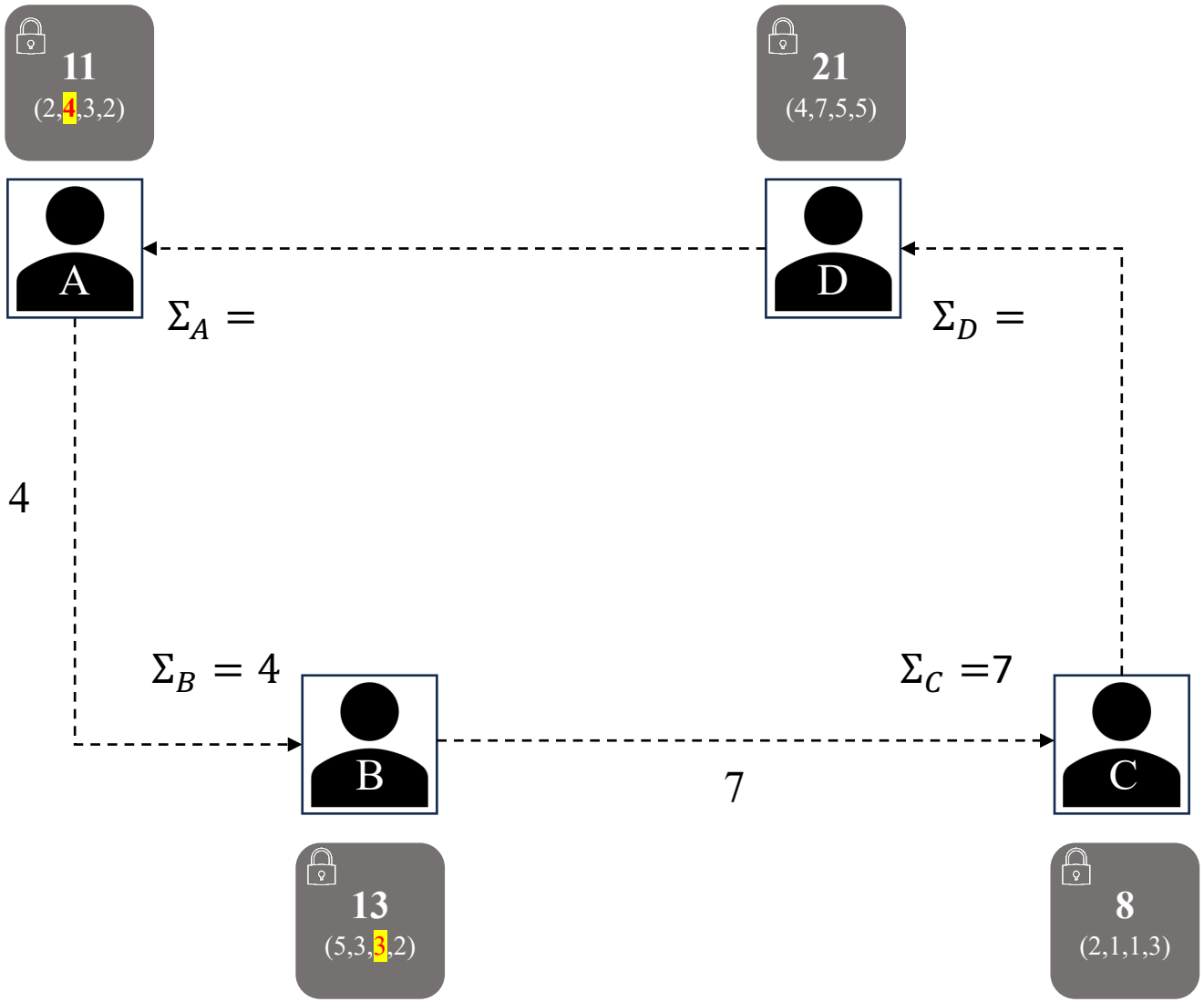
Secure Sum – A Game



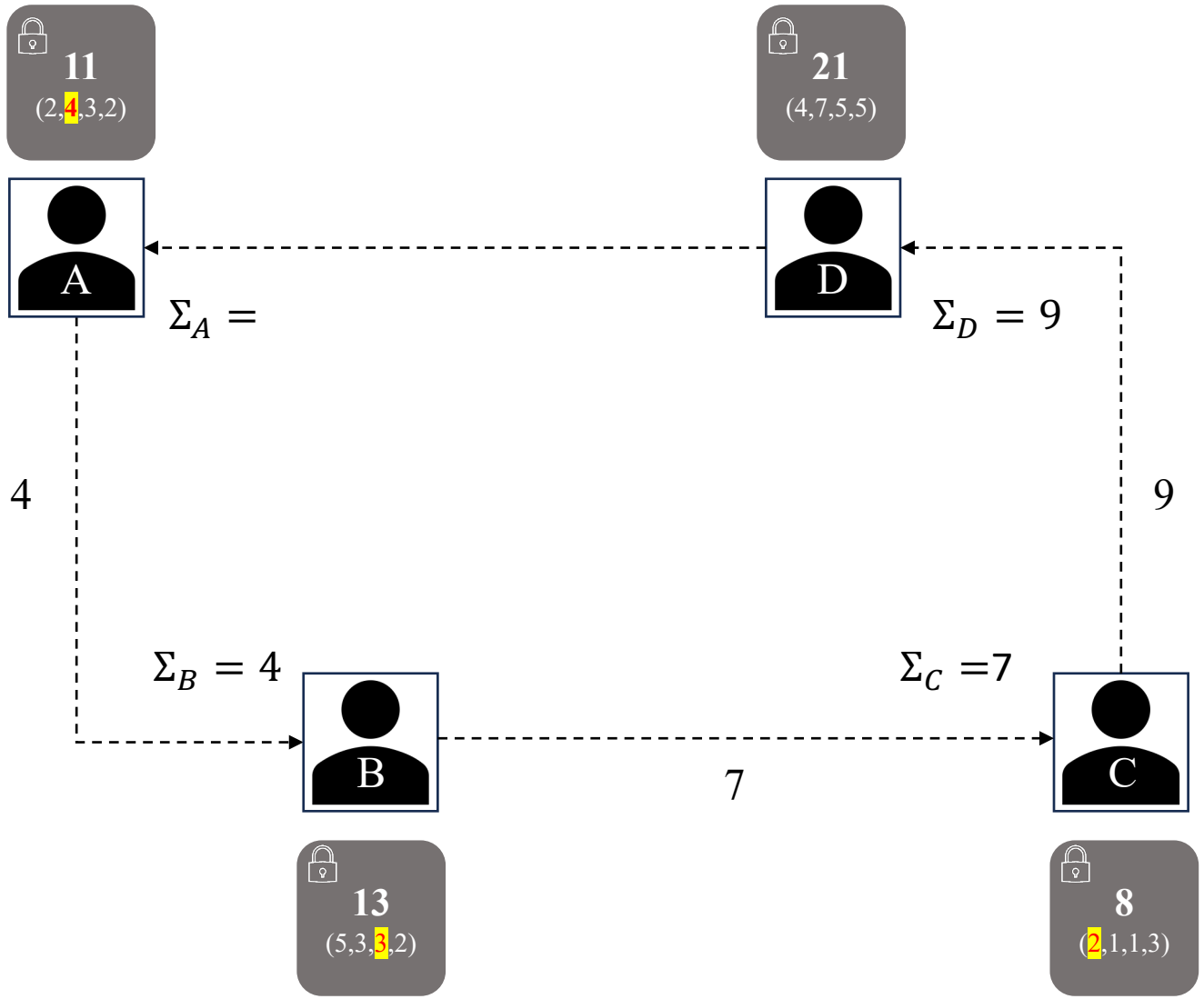
Secure Sum – A Game



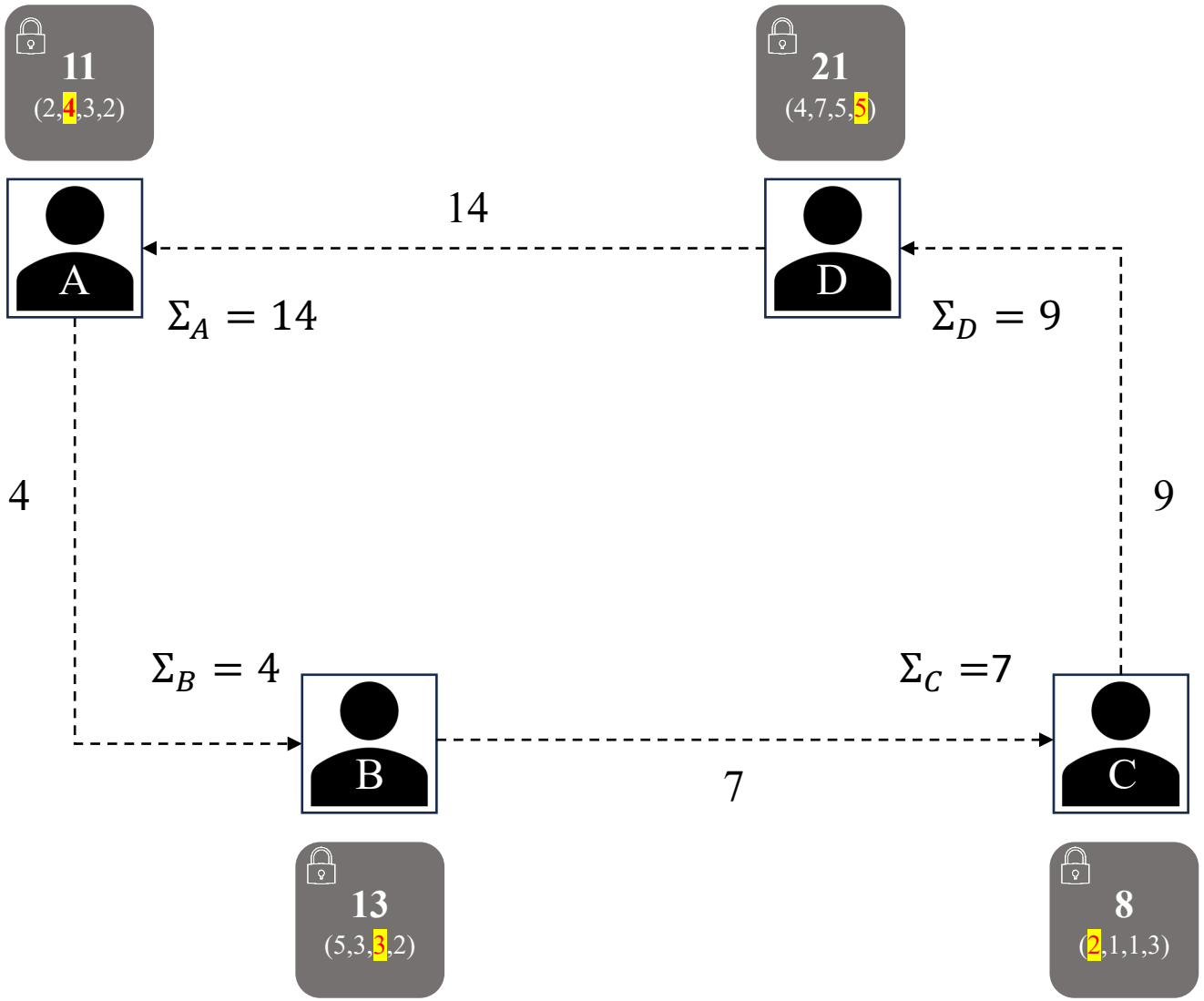
Secure Sum – A Game



Secure Sum – A Game

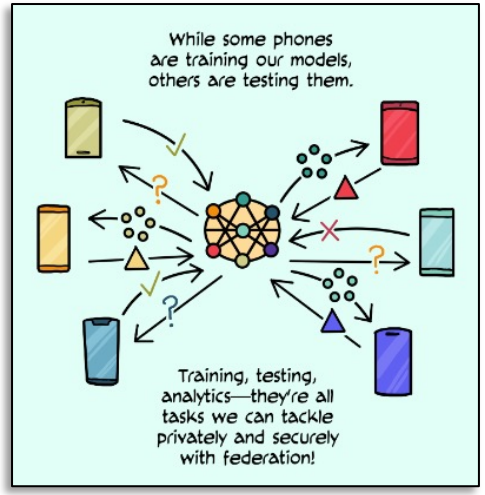


Secure Sum – A Game

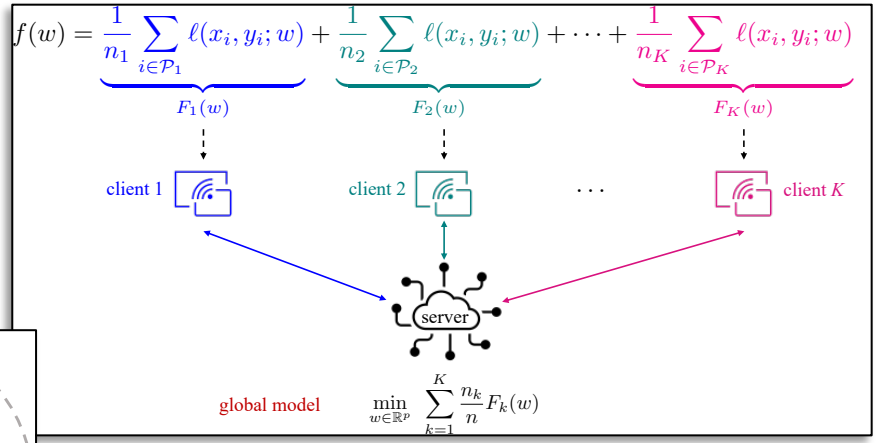


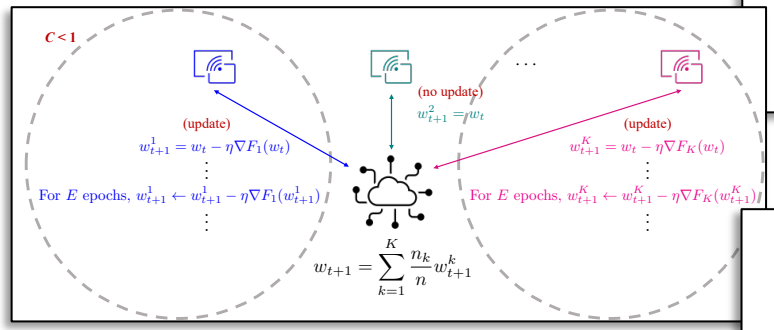
Big Picture

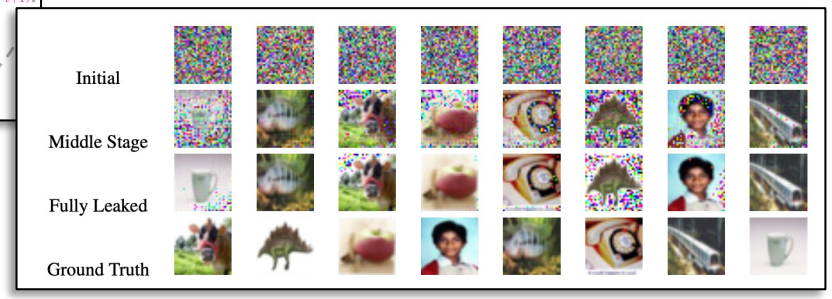
- Motivation and Setup
- Federated Averaging
- Challenges: Privacy
- Secure Sum - A Game

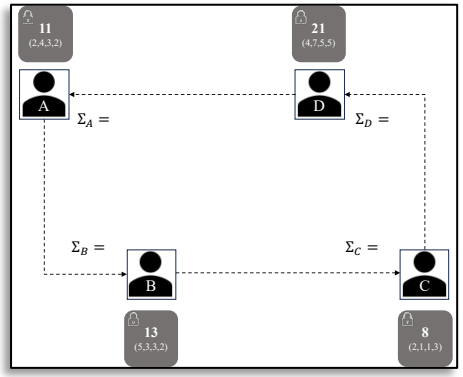


(link)









(link)

Reading Material

Communication-Efficient Learning of Deep Networks from Decentralized Data

H. Brendan McMahan **Eider Moore** **Daniel Ramage** **Seth Hampson** **Blaise Agüera y Arcas**
Google, Inc., 651 N 34th St., Seattle, WA 98103 USA

Federated Learning:
Challenges, Methods, and Future Directions

Tian Li
Carnegie Mellon University
tianli@cmu.edu

Anit Kumar Sahu
Bosch Center for Artificial Intelligenc
anit.sahu@gmail.com

Ameet Talwalkar
Carnegie Mellon University & Determined AI
talwalkar@cmu.edu

Virginia Smith
Carnegie Mellon University
smithv@cmu.edu

Deep Leakage from Gradients

Ligeng Zhu **Zhijian Liu** **Song Han**
Massachusetts Institute of Technology
{ligeng, zhijian, songhan}@mit.edu

The Secret Sharer: Evaluating and Testing Unintended Memorization in Neural Networks

Nicholas Carlini[1,2] Chang Liu[2] Úlfar Erlingsson[1] Jernej Kos[3] Dawn Song[2]
[1]*Google Brain* [2]*University of California, Berkeley* [3]*National University of Singapore*

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*