10-605/10-805: Machine Learning with Large Datasets

Fall 2022

Federated Learning

Announcements

- **No class on Thursday**
- Mini-projects due **Thursday December 1st**
- Recitation this week: Exam 2 Office Hours
- Exam II (December 8) will be in-person

Outline

1. Federated learning: motivation
2. Heterogeneity: optimization, fairness, and personalization

MOTIVATION

Where is ML data coming from?

SUMMARY

LOTS OF DATA

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion - fuelled by internet of things and the use of connected devcies – are hard to comprehend, particularly when looked at in the context of one day

500m

tweets are sent every day

Twitter

4PB

of data created by Facebook, including

350m photos

100m hours of video watch time

Facebook Research

294bn

billion emails are sent

Radicati Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made

Facebook

3.9bn

people use emails

4TB

of data produced by a connected car

Intel

DEMYSTIFIYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit		Value	Size
b	bit	0 or 1	1/8 of a byte
B	byte	8 bits	1 byte
KB	kilobyte	1,000 bytes	1,000 bytes
MB	megabyte	$1,000^2$ bytes	1,000,000 bytes
GB	gigabyte	$1,000^3$ bytes	1,000,000,000 bytes
TB	terabyte	$1,000^4$ bytes	1,000,000,000,000 bytes
PB	petabyte	$1,000^5$ bytes	1,000,000,000,000,000 bytes
EB	exabyte	$1,000^6$ bytes	1,000,000,000,000,000,000 bytes
ZB	zettabyte	$1,000^7$ bytes	1,000,000,000,000,000,000,000 bytes
YB	yottabyte	$1,000^8$ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

463EB

of data will be created every day by 2025

IDC

95m

photos and videos are shared on Instagram

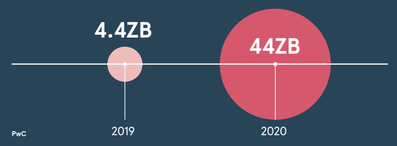
Instagram Business

Searches made a day → **5bn**

Searches made a day from Google → **3.5bn**

Smart insights

ACCUMULATED DIGITAL UNIVERSE OF DATA

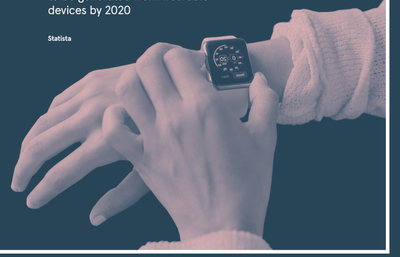


PwC

28PB

to be generated from wearable devices by 2020

Statista



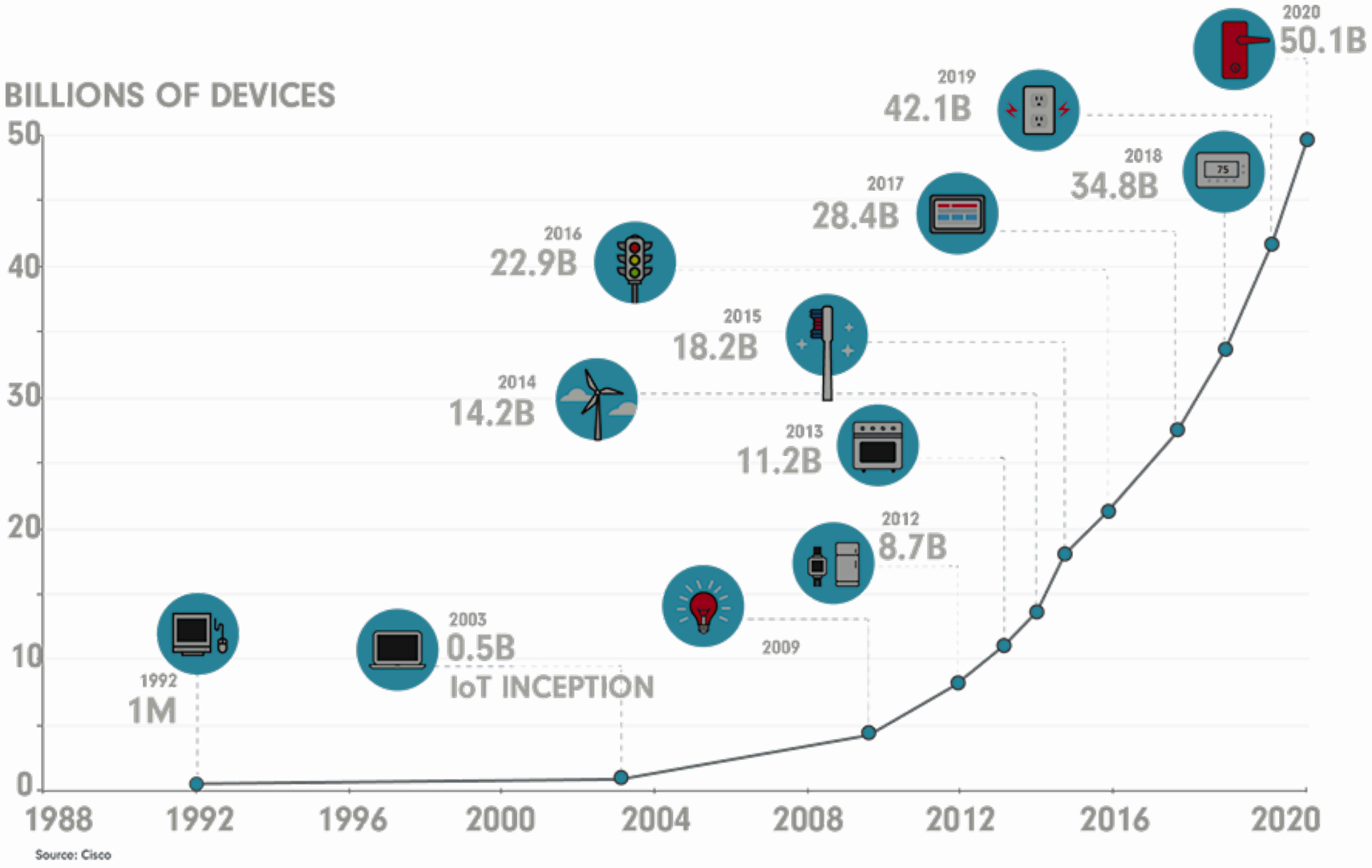
RACONTEUR

SOURCES OF BIG DATA

Internet of Things

SOURCES OF BIG DATA

Internet of Things



MOTIVATION

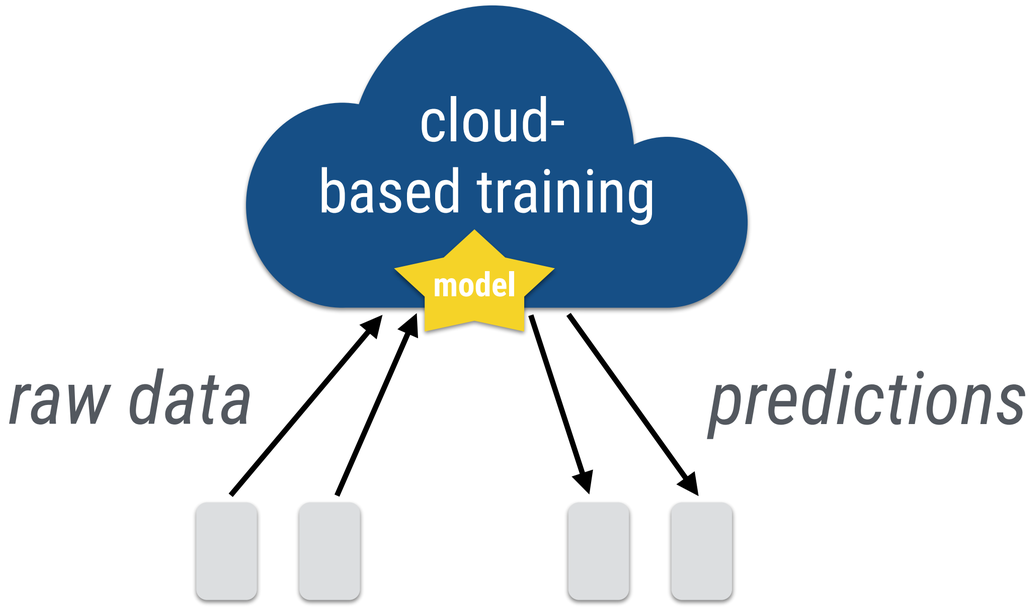
Where is ML data coming from?

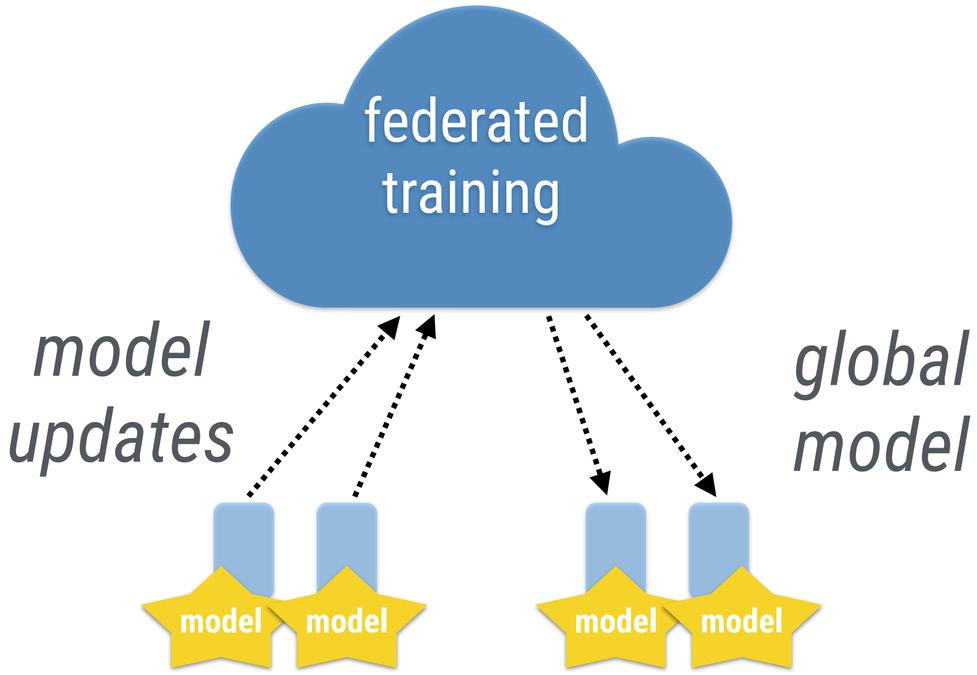
KEY SOURCE: MOBILE PHONES, WEARABLE DEVICES, IOT DEVICES

💡 ***Can we also use these devices for distributed training?***

federated learning

training machine learning models at the edge





why? ✓*quickly incorporate new data* ✓*reduce strain on network* ✓ *privacy*

federated learning

example applications

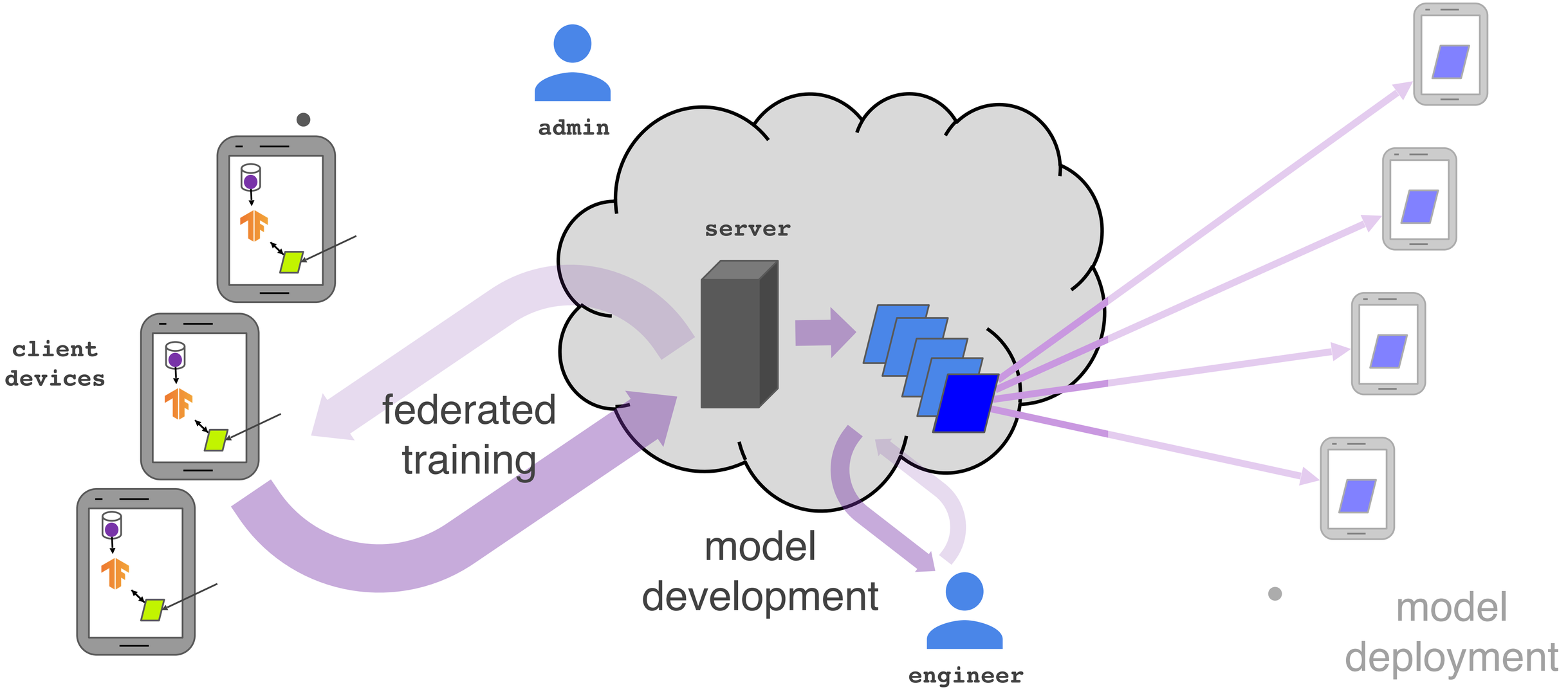
language modeling for voice recognition on mobile phones

adapting to pedestrian behavior in autonomous vehicles

predicting low blood sugar via wearable devices

assumptions? ✓*local data is important* ✓*labels available* ✓*privacy is a concern*

Cross-device federated learning



[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Federated Learning

Federated learning is a machine learning setting where multiple entities (clients) collaborate in solving a machine learning problem, under the coordination of a central server or service provider. Each client's raw data is stored locally and not exchanged or transferred; instead, focused updates intended for immediate aggregation are used to achieve the learning objective.

definition proposed in
Advances and Open Problems in Federated Learning (arxiv/1912.04977)

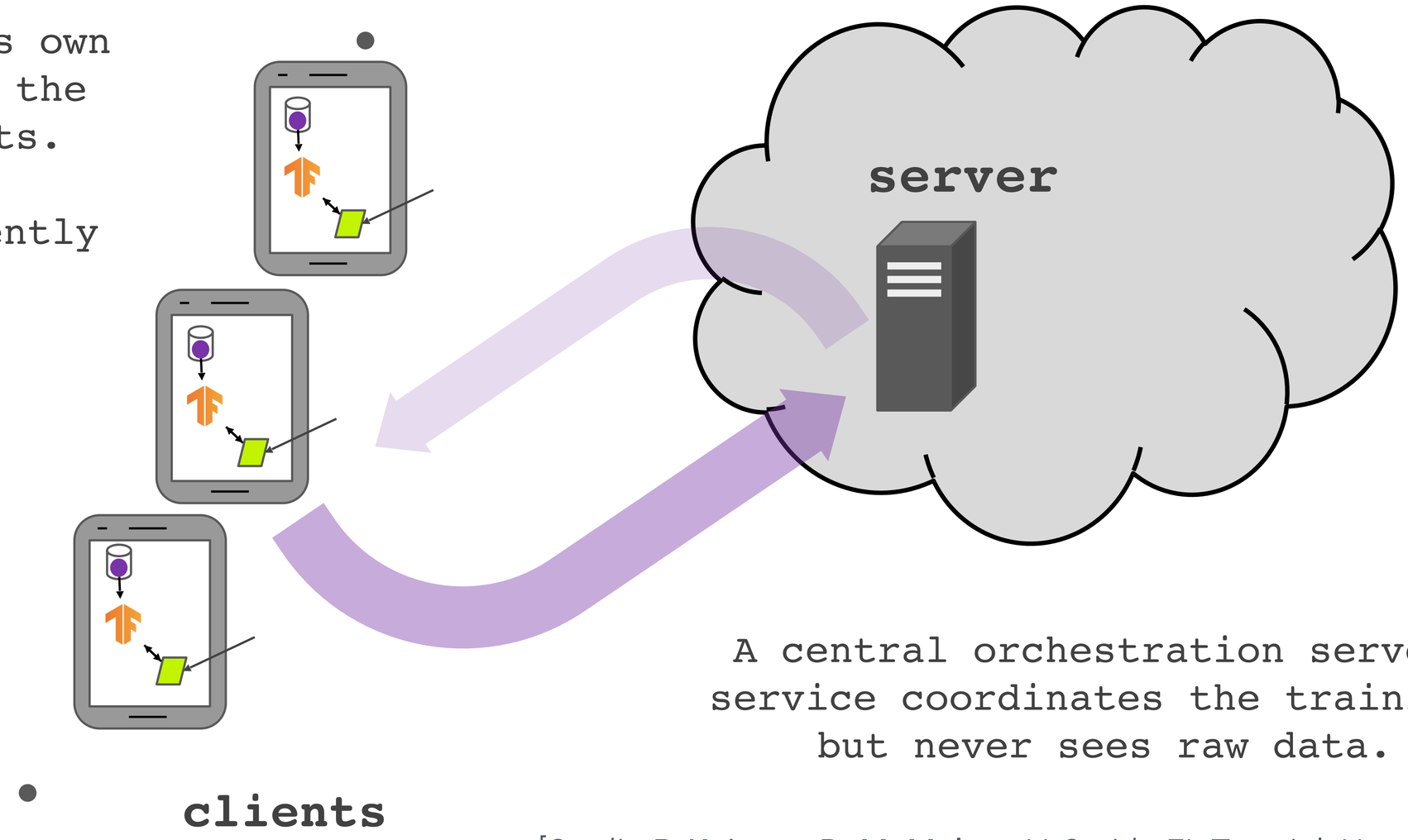
[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Federated learning - defining characteristics

Data is generated locally and remains decentralized.

Each client stores its own data and cannot read the data of other clients.

Data is not independently or identically distributed.



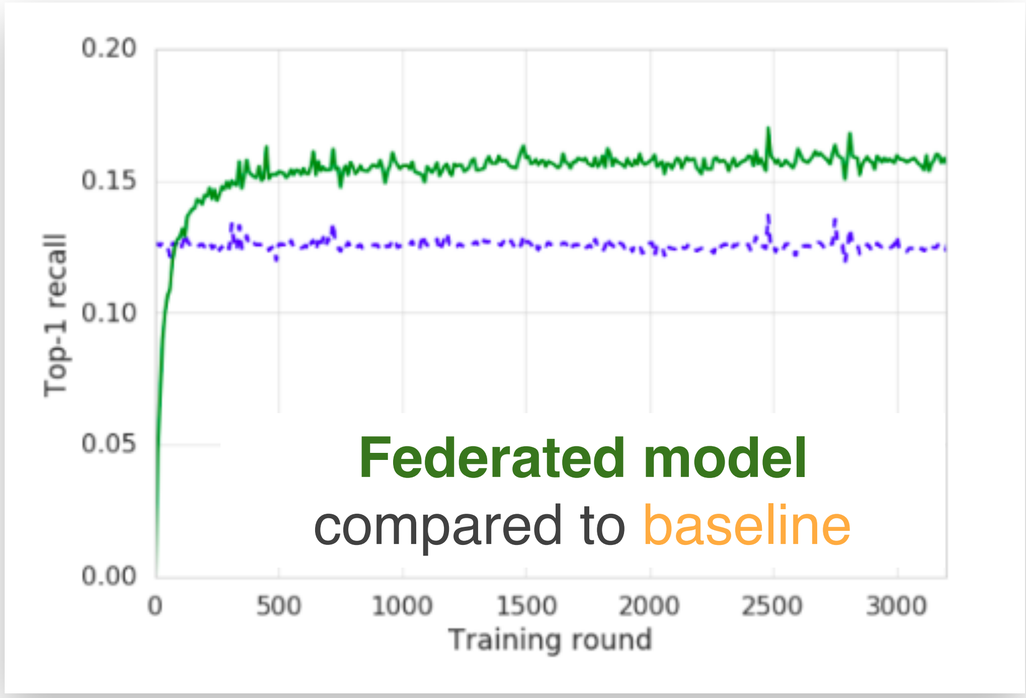
A central orchestration server/ service coordinates the training, but never sees raw data.

[Credit: P. Kairouz, B. McMahan, V. Smith, FL Tutorial, NeurIPS 2020]

Gboard: next-word prediction

Federated RNN (compared to prior n-gram model):

- Better next-word prediction accuracy: +24%
- More useful prediction strip: +10% more clicks



A. Hard, *et al*. **Federated Learning for Mobile Keyboard Prediction**. *arXiv:1811.03604*

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Other federated models in Gboard

Emoji prediction

- 7% more accurate emoji predictions
- prediction strip clicks +4% more
- 11% more users share emojis!

Ramaswamy, *et al*. **Federated Learning for Emoji Prediction in a Mobile Keyboard**. arXiv:1906.04329.

Action prediction

When is it useful to suggest a gif, sticker, or search query?

- 47% reduction in unhelpful suggestions
- increasing overall emoji, gif, and sticker shares

T. Yang, *et al*. **Applied Federated Learning: Improving Google Keyboard Query Suggestions**. *arXiv:1812.02903*

Discovering new words

Federated discovery of what words people are typing that Gboard doesn't know.

M. Chen, *et al*. **Federated Learning Of Out-Of-Vocabulary Words**. *arXiv:1903.10635*

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-device federated learning at Apple

MIT Technology Review

Sign in Subscribe

Artificial intelligence / Machine learning

How Apple personalizes Siri without hoovering up your data

The tech giant is using privacy-preserving machine learning to improve its voice assistant while keeping your data on your phone.

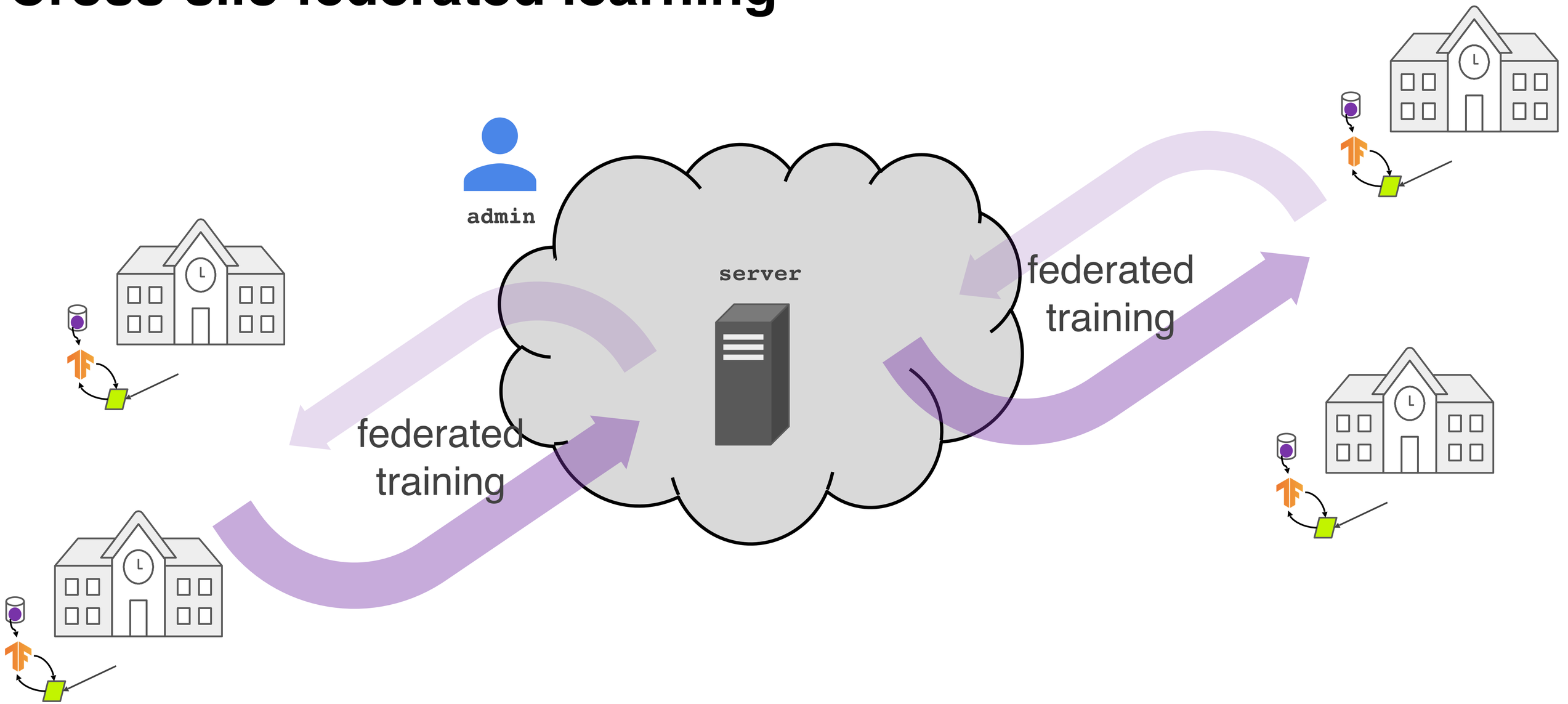
by **Karen Hao** December 11, 2019

*"Instead, **it relies primarily on a technique called federated learning**, Apple's head of privacy, Julien Freudiger, told an audience at the Neural Processing Information Systems conference on December 8. Federated learning is a privacy-preserving machine-learning method that was first introduced by Google in 2017. It allows Apple to train different copies of a speaker recognition model across all its users' devices, using only the audio data available locally. It then sends just the updated models back to a central server to be combined into a master model. In this way, raw audio of users' Siri requests never leaves their iPhones and iPads, but the assistant continuously gets better at identifying the right speaker."*

https://www.technologyreview.com/2019/12/11/131629/apple-ai-personalizes-siri-federated-learning/

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-silo federated learning



[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-silo federated learning from Intel

ARTIFICIAL INTELLIGENCE, DIAGNOSTICS

UPenn, Intel partner to use federated learning AI for early brain tumor detection

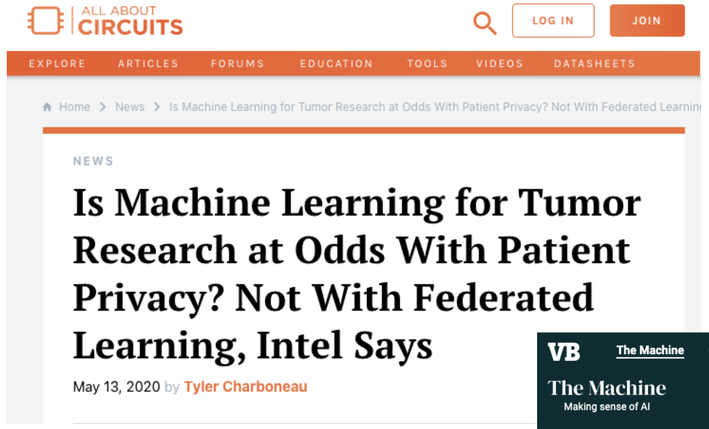
The project will bring in 29 institutions from North America, Europe and India and will use privacy-preserved data to train AI models. Federated learning has been described as being born at the intersection of AI, blockchain, edge computing and the Internet of Things.

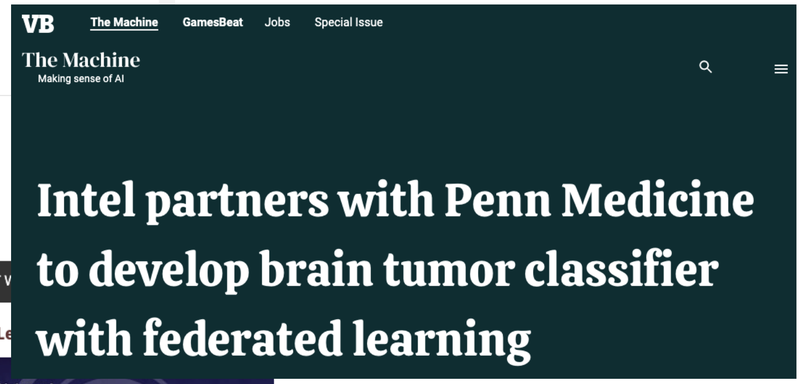
By ALARIC DEARMENT

Post a comment / May 11, 2020 at 10:03 AM

"The University of Pennsylvania and chipmaker Intel are forming a partnership to enable 29 heatlhcare and medical research institutions around the world to train artificial intelligence models to detect brain tumors early."

"The program will rely on a technique known as federated learning, which enables institutions to collaborate on deep learning projects without sharing patient data. The partnership will bring in institutions in the U.S., Canada, U.K., Germany, Switzerland and India. The centers – which include Washington University of St. Louis; Queen's University in Kingston, Ontario; University of Munich; Tata Memorial Hospital in Mumbai and others – will use Intel's federated learning hardware and software."

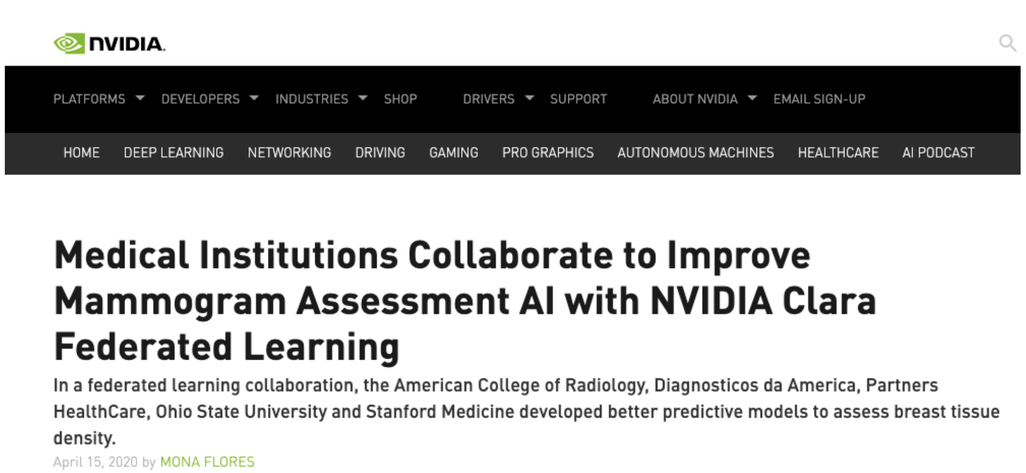




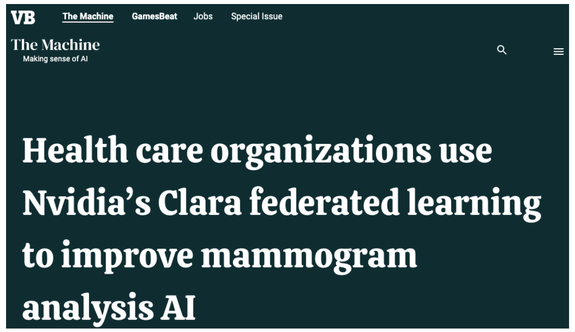
[1] https://medcitynews.com/2020/05/upenn-intel-partner-to-use-federated-learning-ai-for-early-brain-tumor-detection/
[2] https://www.allaboutcircuits.com/news/can-machine-learning-keep-patient-privacy-for-tumor-research-intel-says-yes-with-federated-learning/
[3] https://venturebeat.com/2020/05/11/intel-partners-with-penn-medicine-to-develop-brain-tumor-classifier-with-federated-learning/
[4] http://www.bio-itworld.com/2020/05/28/intel-penn-medicine-launch-federated-learning-model-for-brain-tumors.aspx

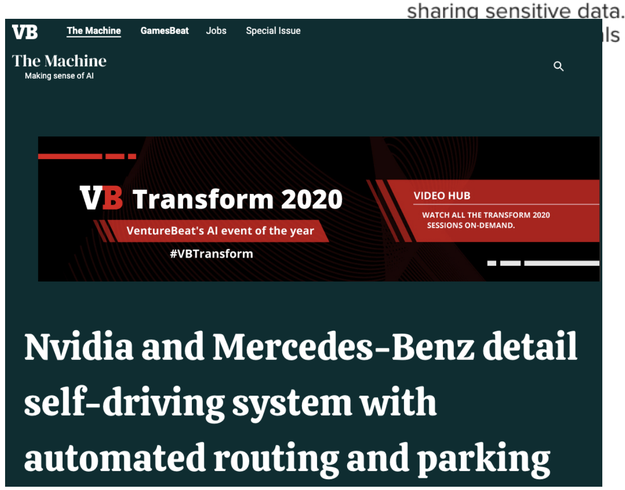
[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-silo federated learning from NVIDIA



"Federated learning addresses this challenge, enabling different institutions to collaborate on AI model development without sharing sensitive clinical data with each other. The goal is to end up with more generalizable models that perform well on any dataset, instead of an AI biased by the patient demographics or imaging equipment of one specific radiology department."



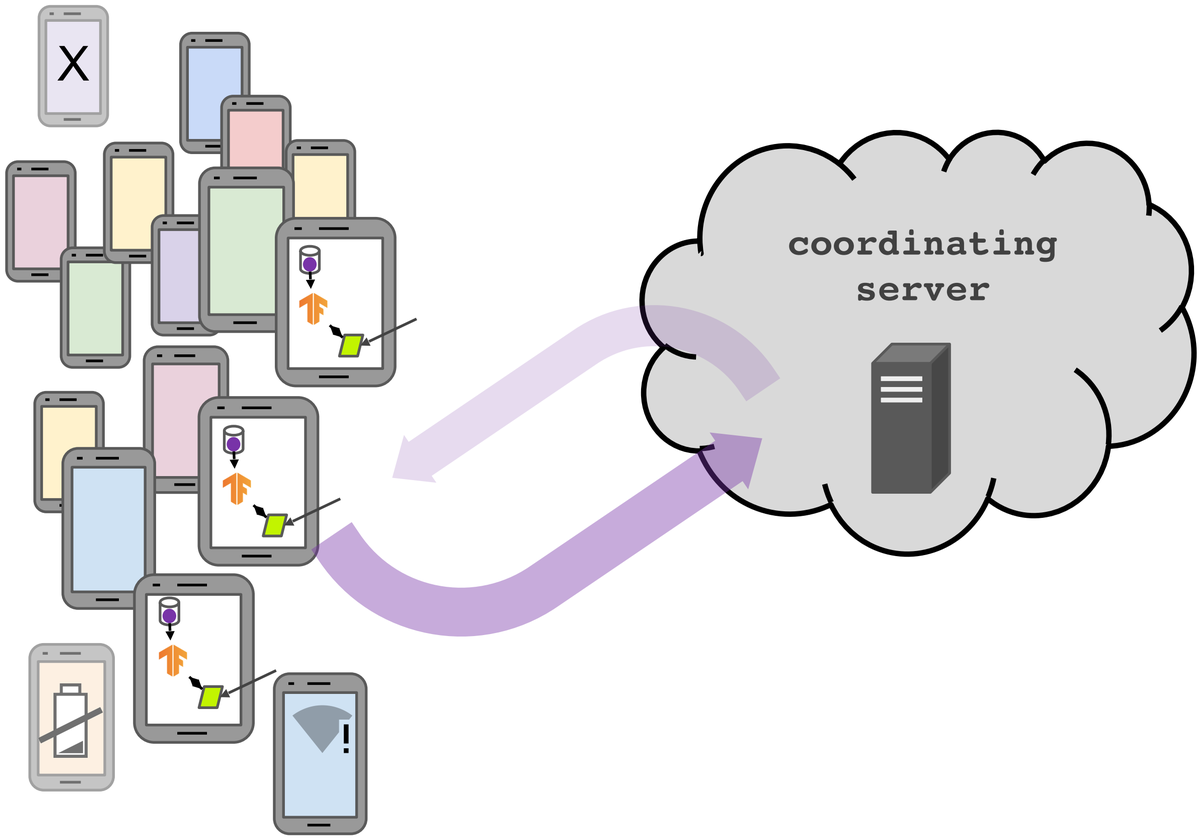


[1] https://blogs.nvidia.com/blog/2020/04/15/federated-learning-mammogram-assessment/
[2] https://venturebeat.com/2020/04/15/healthcare-organizations-use-nvidias-clara-federated-learning-to-improve-mammogram-analysis-ai/
[3] https://medcitynews.com/2020/01/nvidia-says-it-has-a-solution-for-healthcares-data-problems/
[4] https://venturebeat.com/2020/06/23/nvidia-and-mercedes-benz-detail-self-driving-system-with-automated-routing-and-parking/

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

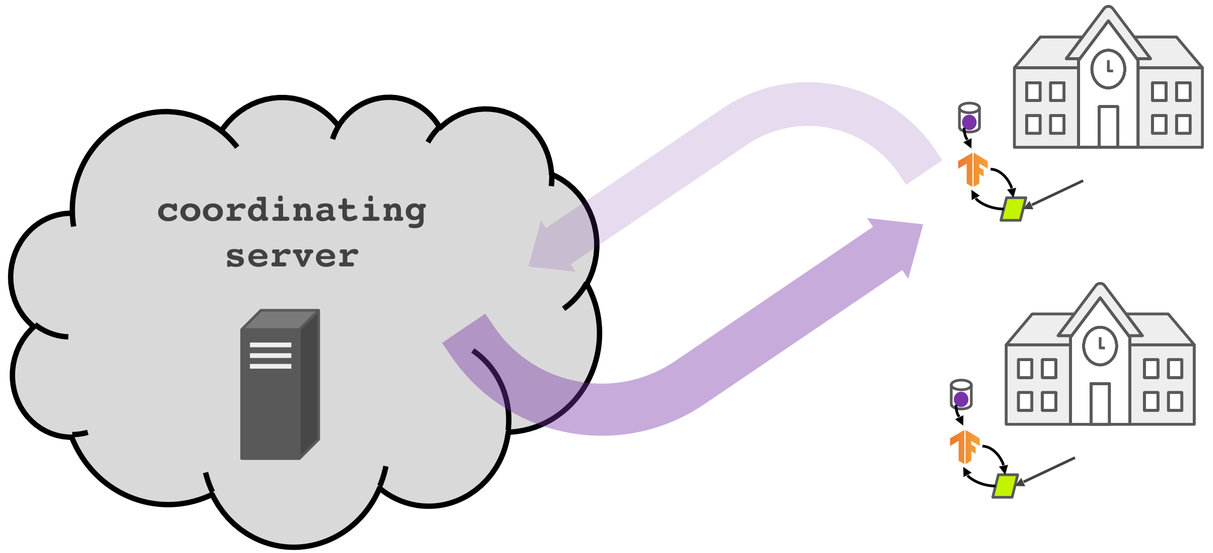
Cross-device federated learning

millions of intermittently
available client devices



Cross-silo federated learning

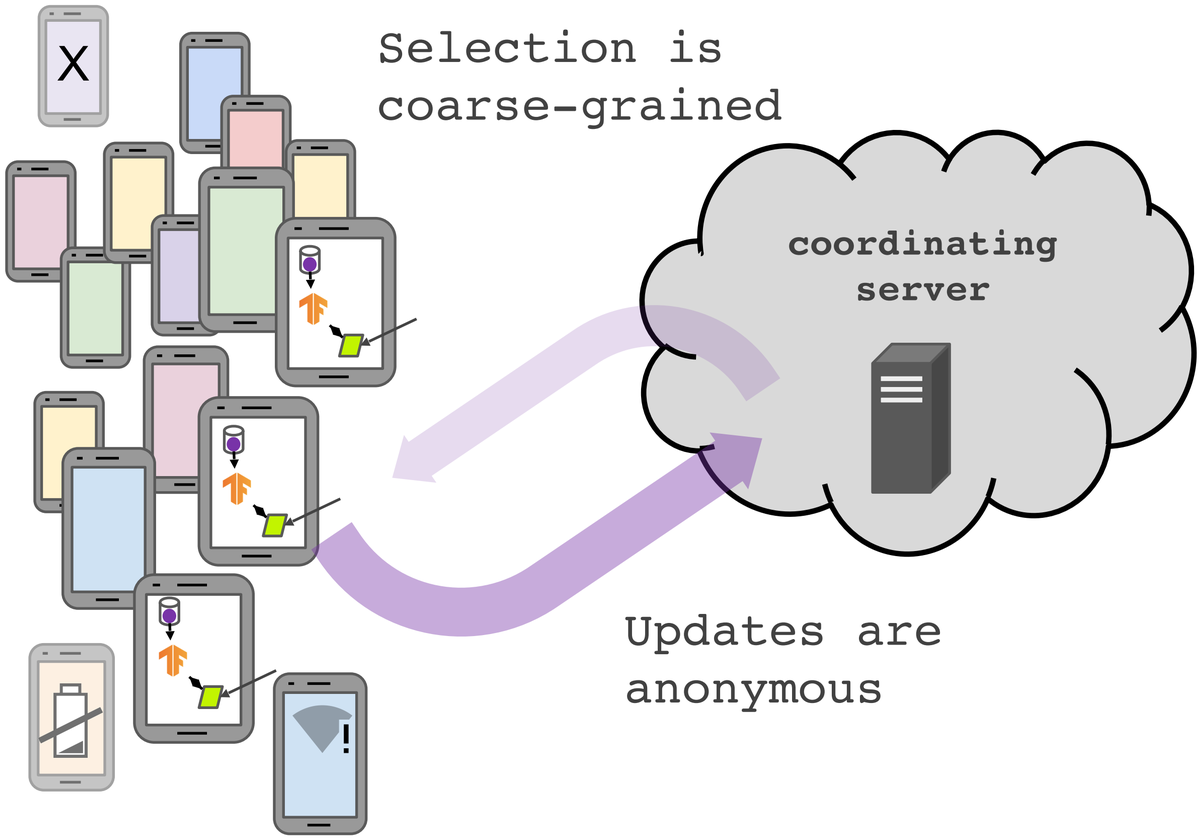
small number of clients
(institutions, data silos),
high availability



[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

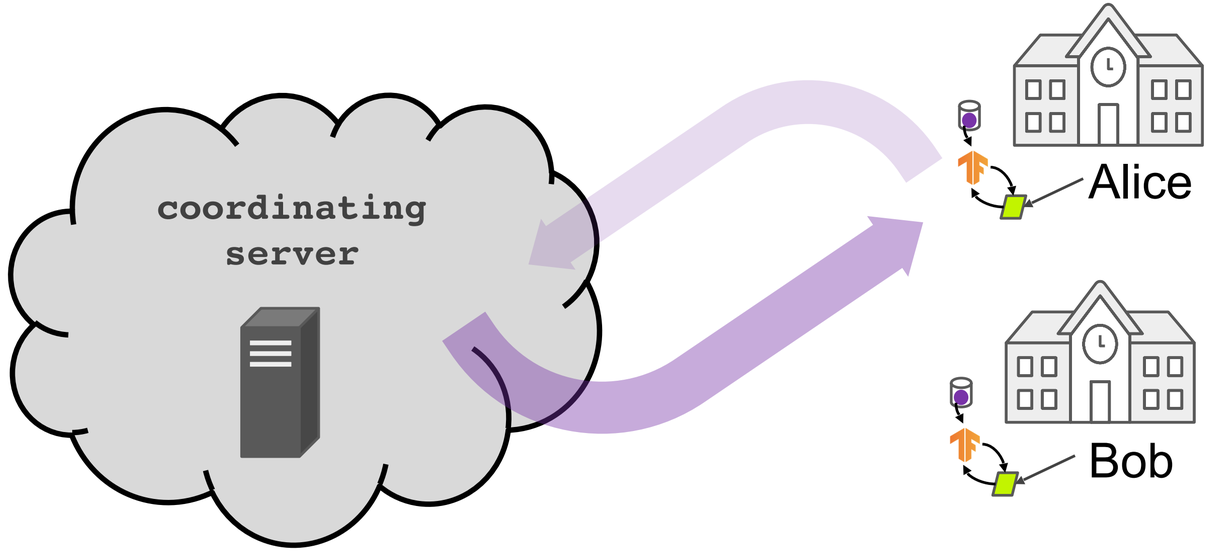
Cross-device federated learning

clients cannot be indexed directly (i.e., no use of client identifiers)



Cross-silo federated learning

each client has an identity or name that allows the system to access it specifically

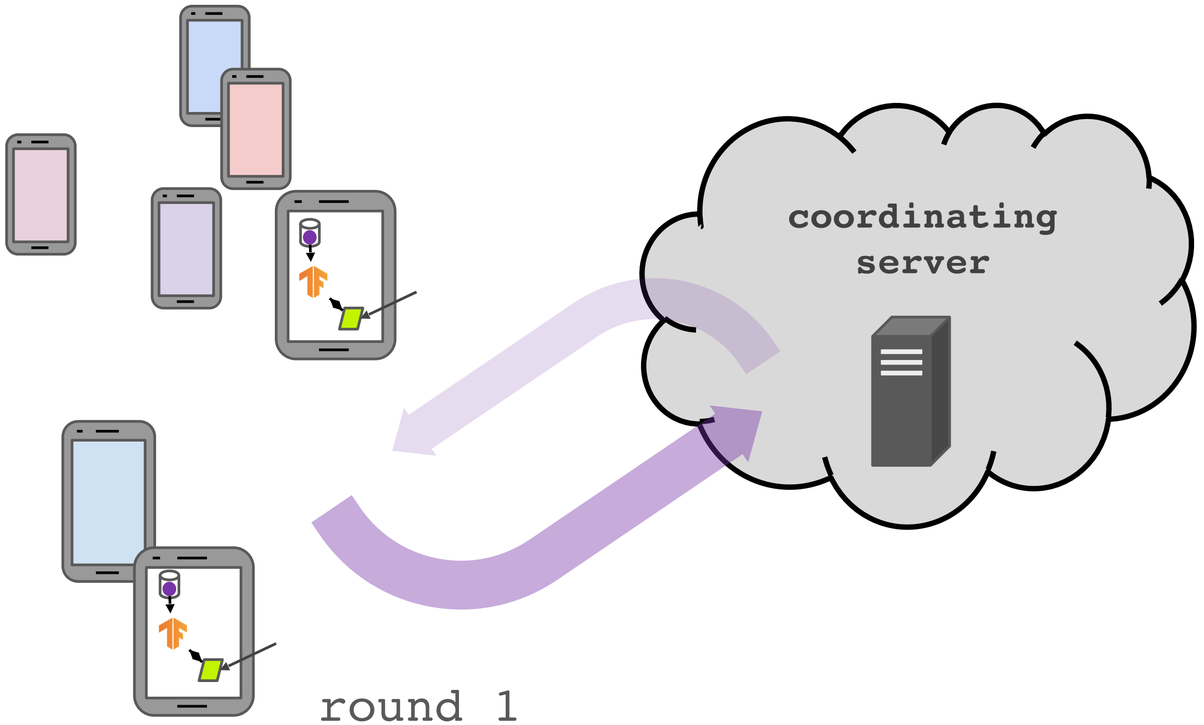


[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-device federated learning

Server can only access a (possibly biased) random sample of clients on each round.

Large population => most clients only participate once.

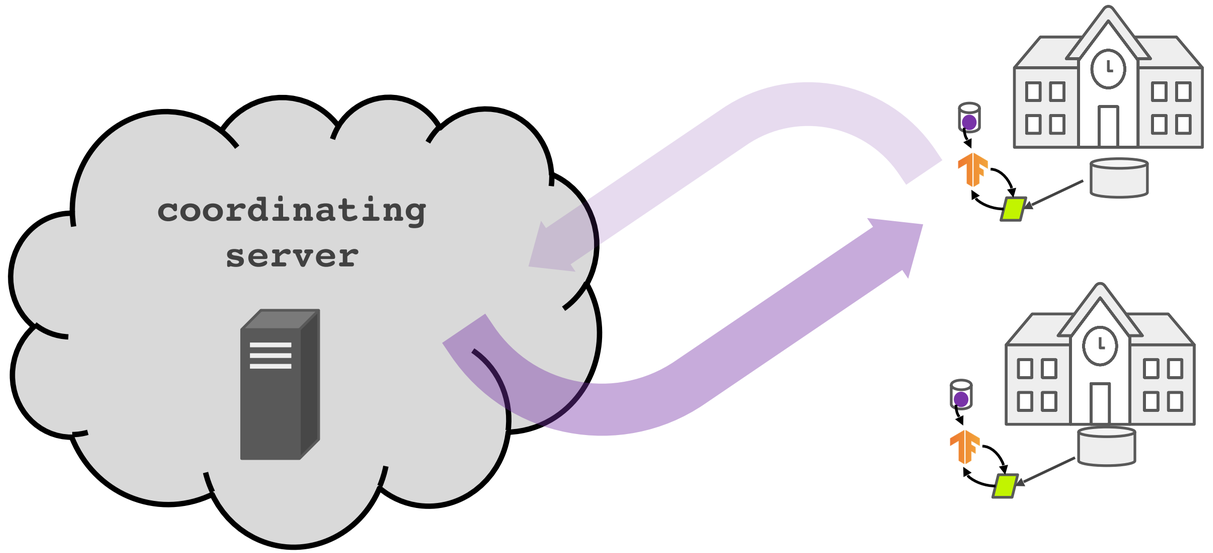


round 1

Cross-silo federated learning

Most clients participate in every round.

Clients can run algorithms that maintain local state across rounds.



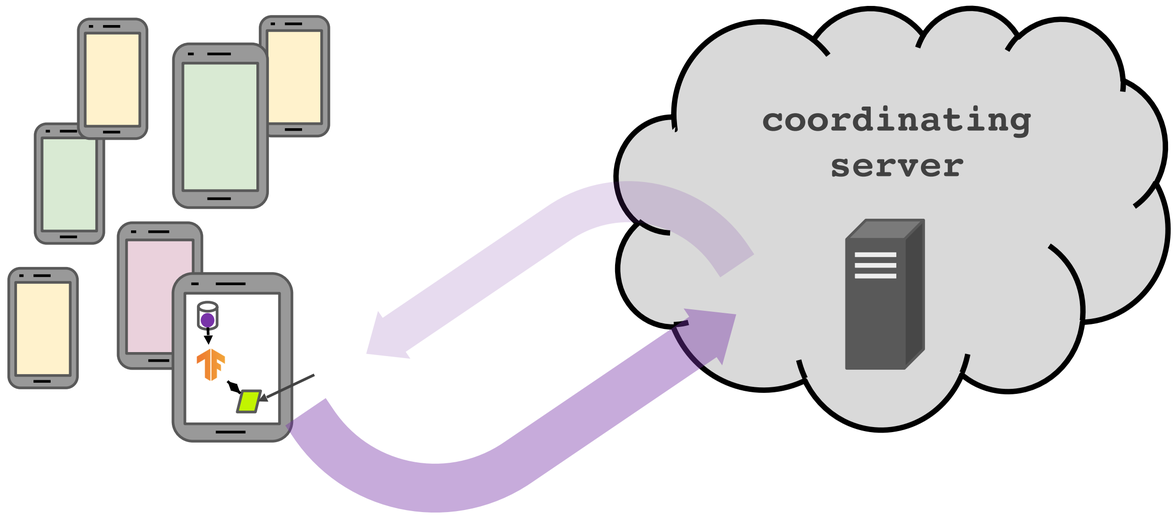
round 1

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

Cross-device federated learning

Server can only access a (possibly biased) random sample of clients on each round.

Large population => most clients only participate once.

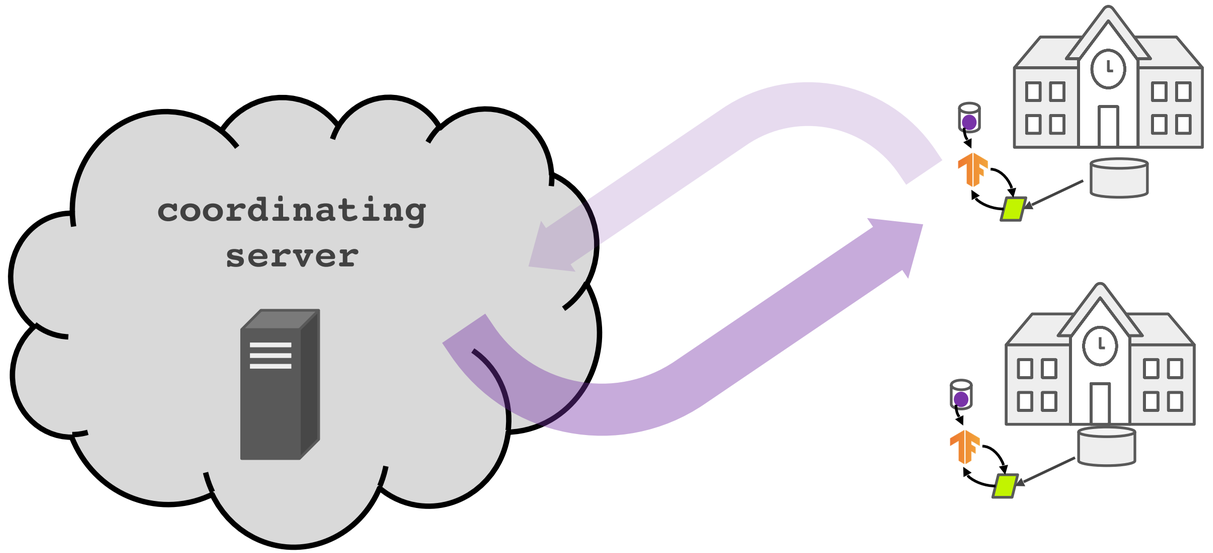


round 2
(completely new set of devices participate)

Cross-silo federated learning

Most clients participate in every round.

Clients can run algorithms that maintain local state across rounds.

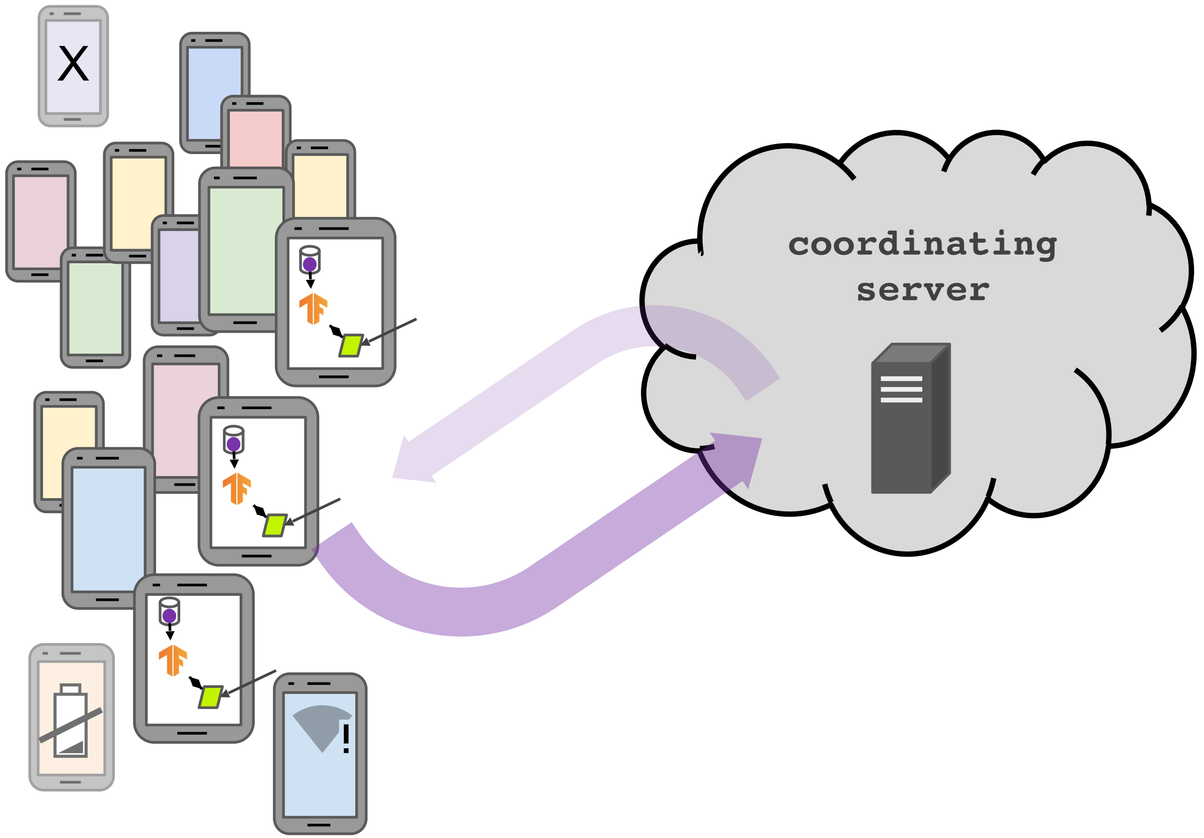


round 2
(same clients)

[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

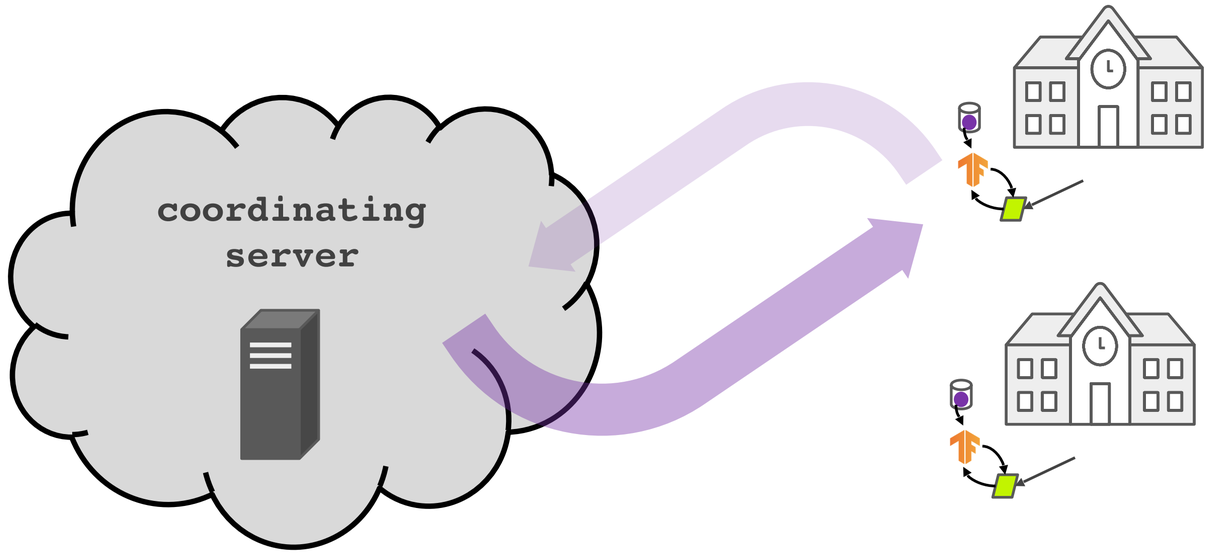
Cross-device federated learning

communication is often the primary bottleneck



Cross-silo federated learning

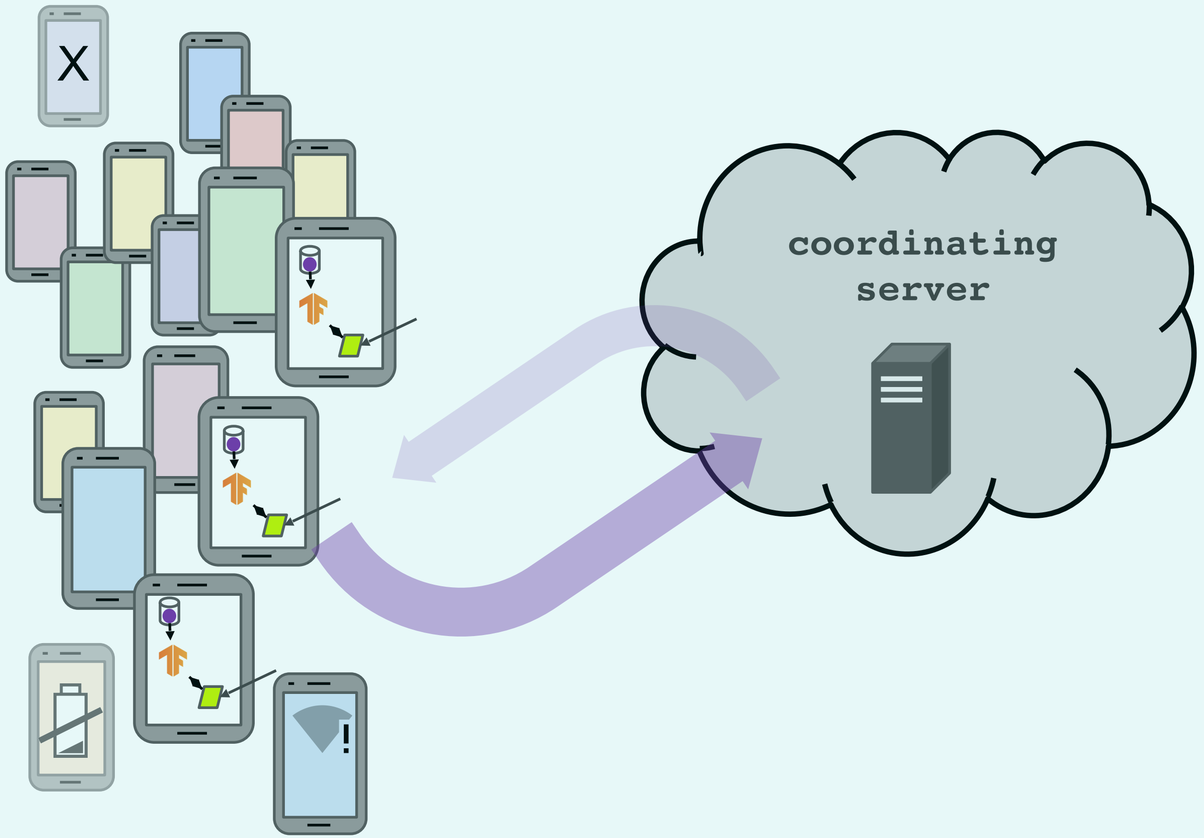
communication or computation might be the primary bottleneck



[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

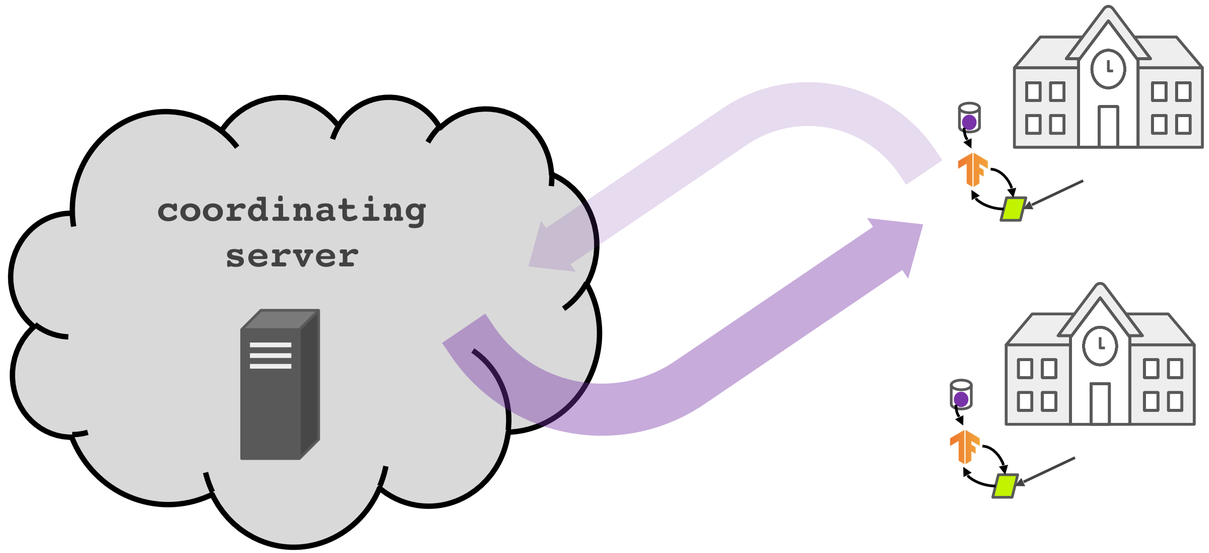
Cross-device federated learning

communication is often the primary bottleneck



Cross-silo federated learning

communication or computation might be the primary bottleneck



[Credit: **P. Kairouz, B. McMahan**, V. Smith, FL Tutorial, NeurIPS 2020]

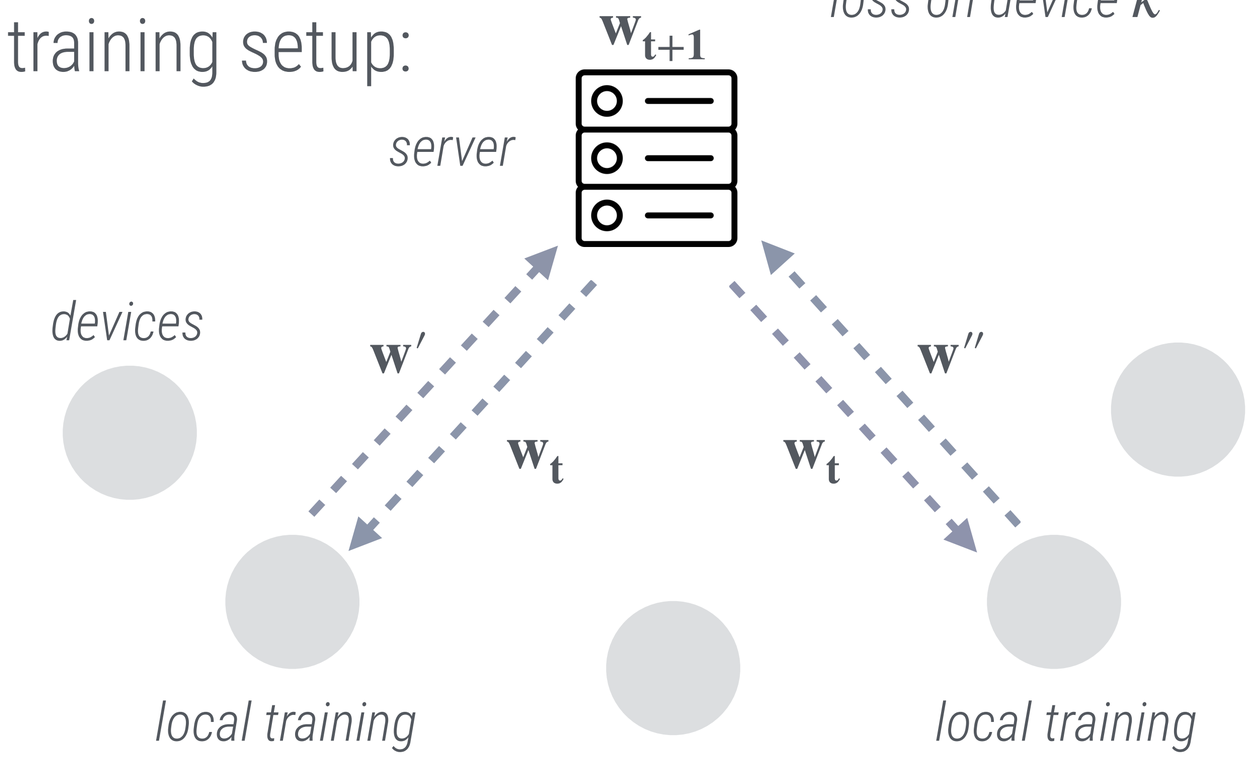

last 20+ years:
distributed optimization
in the data center

federated learning: workflow & challenges

objective: $\min_w f(w) = \sum_{k=1}^{N} p_k F_k(w)$

F_k: *loss on device k*

training setup:



expensive communication
- massive, slow, unreliable networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data

systems heterogeneity
- variable hardware, connectivity, etc

federated learning: workflow & challenges

Federated Learning: Challenges, Methods, and Future Directions,
T. Li, A. K. Sahu, A. Talwalkar, V. Smith,
IEEE Signal Processing Magazine 2020

Federated Learning and Analytics: Industry Meets Academia
P. Kairouz, B. McMahan, V. Smith
NeurIPS Tutorial, https://slideslive.com/38935813/federated-learning-tutorial

expensive communication
- massive, slow, unreliable networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data

systems heterogeneity
- variable hardware, connectivity, etc

federated learning: workflow & challenges

Federated Learning: Challenges, Methods, and Future Directions,
T. Li, A. K. Sahu, A. Talwalkar, V. Smith,
IEEE Signal Processing Magazine 2020

Federated Learning and Analytics: Industry Meets Academia
P. Kairouz, B. McMahan, V. Smith
NeurIPS Tutorial, https://slideslive.com/38935813/federated-learning-tutorial

expensive communication
- massive, slow networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data

systems heterogeneity
- variable hardware, connectivity, etc

on heterogeneity in federated settings

- *how does heterogeneity affect federated optimization methods?*
- *can we equalize performance across diverse networks?*
- *how can we personalize models?*

on heterogeneity in federated settings

optimization

- *how does heterogeneity affect federated optimization methods?*

fairness

- *can we equalize performance across diverse networks?*

modeling

- *how can we personalize models?*

on heterogeneity in federated settings

optimization

- *how does heterogeneity affect federated optimization methods?*

fairness

- *can we equalize performance across diverse networks?*

modeling

- *how can we personalize models?*

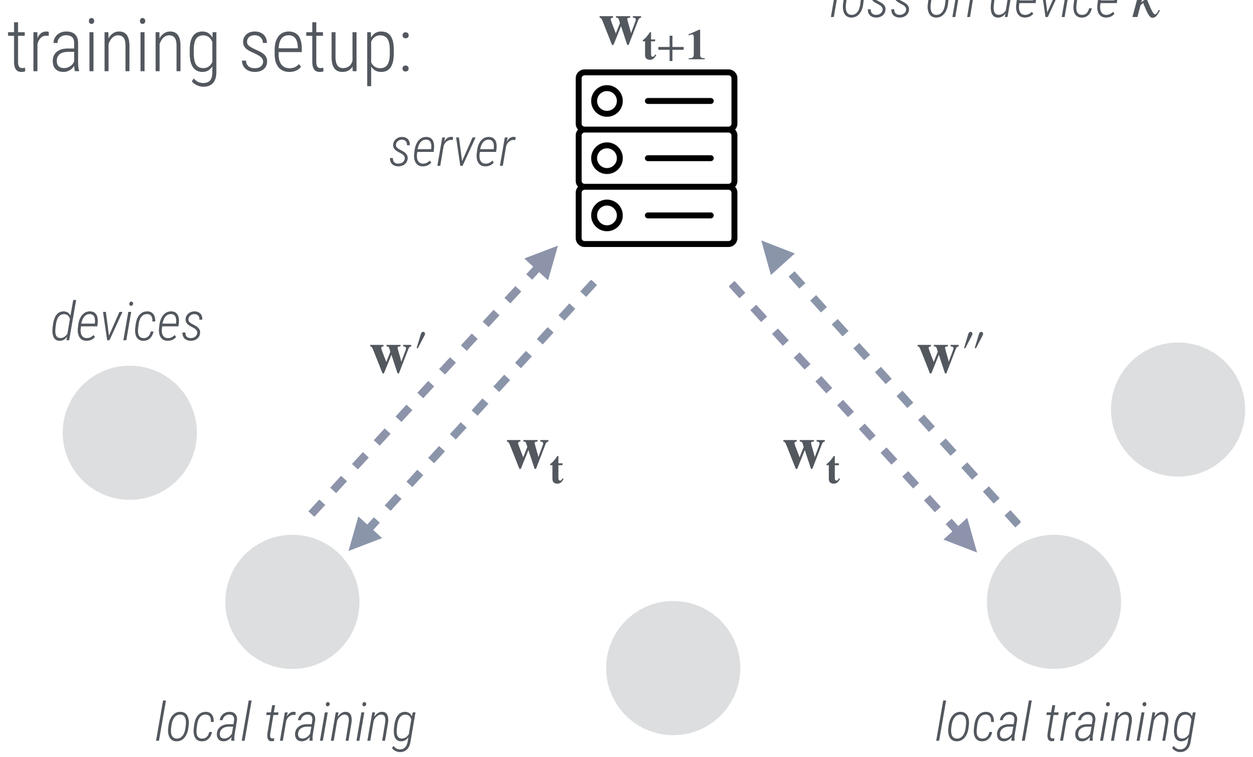
Federated Optimization in Heterogeneous Networks
Li, Sahu, Sanjabi, Zaheer, Talwalkar, Smith, MLSys 2020

Federated learning: workflow & challenges

objective: $\min_w f(w) = \sum_{k=1}^{N} p_k F_k(w)$

loss on device k

training setup:



expensive communication
- massive, slow networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data

systems heterogeneity
- variable hardware, connectivity, etc

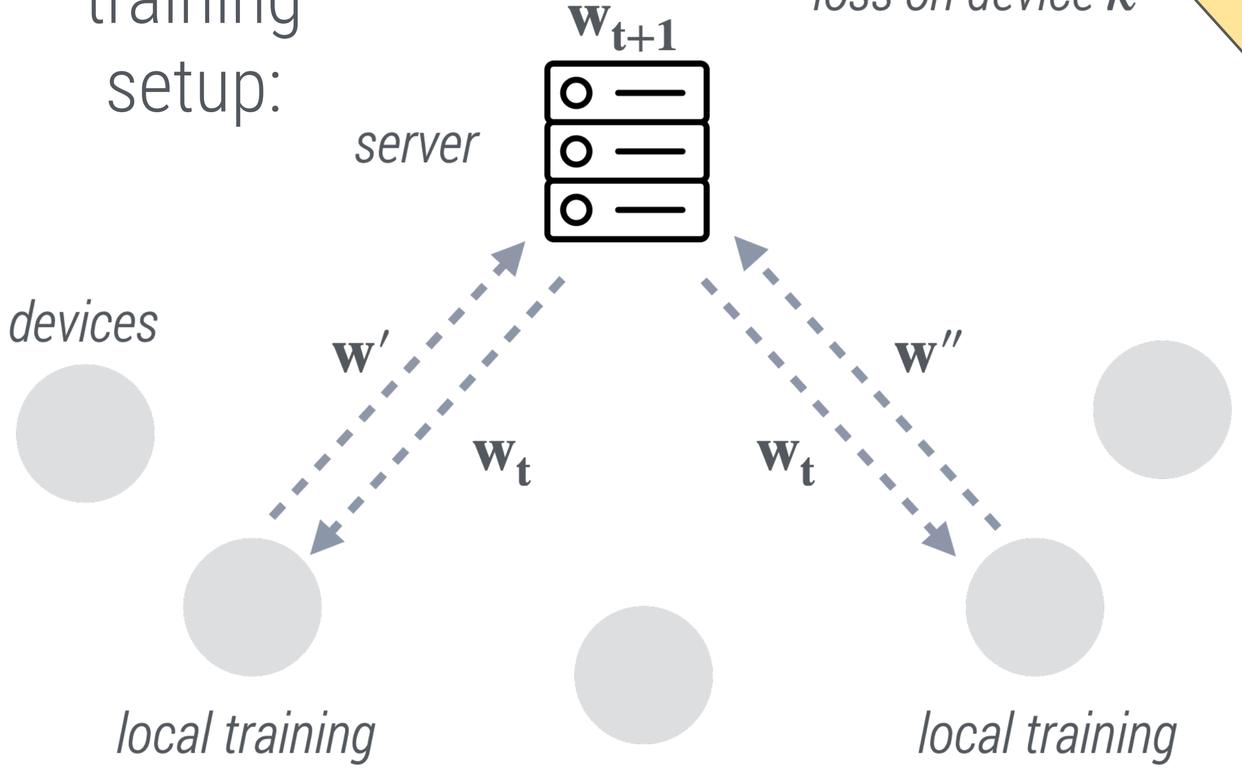
A CLOSER LOOK

Federated learning: workflow & challenges

objective: $\min_w f(w) = \sum_{k=1}^{m} p_k F_k(w)$

F_k: loss on device k

training setup:



Typically consider solving an ERM objective, which is a (possibly) weighted average of losses across the *m* devices and their local data, i.e.,

$$\min_w \sum_{k=1}^{m} p_k \sum_{i=1}^{n_k} \ell(h(x_k^{(i)}; w), y_k^{(i)})$$

However, challenges discussed translate to other common ML objectives as well

ERM objective

Risk

$$R(h) = \mathbb{E}_{(x,y)\sim P}[\ell(h(x; w), y)]$$

Assume we have access to a sample of data $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$

Goal is to estimate the expected risk using this sample, i.e.:

Empirical Risk

$$R_{emp}(h) = \frac{1}{N}\sum_{i=1}^{N} \ell(h(x^{(i)}; w), y^{(i)})$$

Federated ERM objective

Risk

$$R(h) = \mathbb{E}_{k \sim Q} \mathbb{E}_{(x,y) \sim P_k} [\ell(h(x; w), y)]$$

We have a sample of data $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(N)}, y^{(N)})\}$ **generated from *m* devices**

Each device may generate data according to its own distribution, P_k

Goal is to estimate the true risk over this sample, i.e.:

Empirical Risk

$$R_{emp}(h) = \sum_{k=1}^{m} p_k \sum_{i=1}^{n_k} \ell(h(x_k^{(i)}; w), y_k^{(i)})$$

Terminology

m	number of devices
N	total number of data points
n_k	number of data points on device k
$(x_k^{(i)}, y_k^{(i)})$	i-th data point on device k
w	model parameters

Empirical Risk

$$R_{emp}(h) = \sum_{k=1}^{m} p_k \sum_{i=1}^{n_k} \ell(h(x_k^{(i)}; w), y_k^{(i)})$$
$$= \sum_{k=1}^{m} p_k F_k(w)$$

A CLOSER LOOK

Federated learning: workflow & challenges

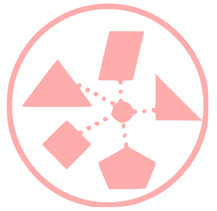
Can **reduce communication** in federated optimization by:

1. Limiting *number of devices* involved in communication
2. Reducing number of *communication rounds*
3. Reducing *size of messages* sent over network

expensive communication
- massive, slow networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data



systems heterogeneity
- variable hardware, connectivity, etc

A CLOSER LOOK

Federated learning: workflow & challenges

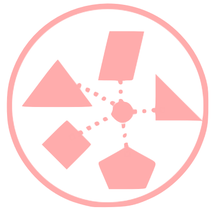
Keeping **raw data local** to each device is a first step

Often also use privacy mechanisms in conjunction with federated optimization

expensive communication
- massive, slow networks

privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data



systems heterogeneity
- variable hardware, connectivity, etc

A CLOSER LOOK

Federated learning: workflow & challenges

Heterogeneous (i.e., non-identically distributed) data and systems can bias optimization procedures

expensive communication

- massive, slow networks

privacy concerns

- user privacy constraints

statistical heterogeneity

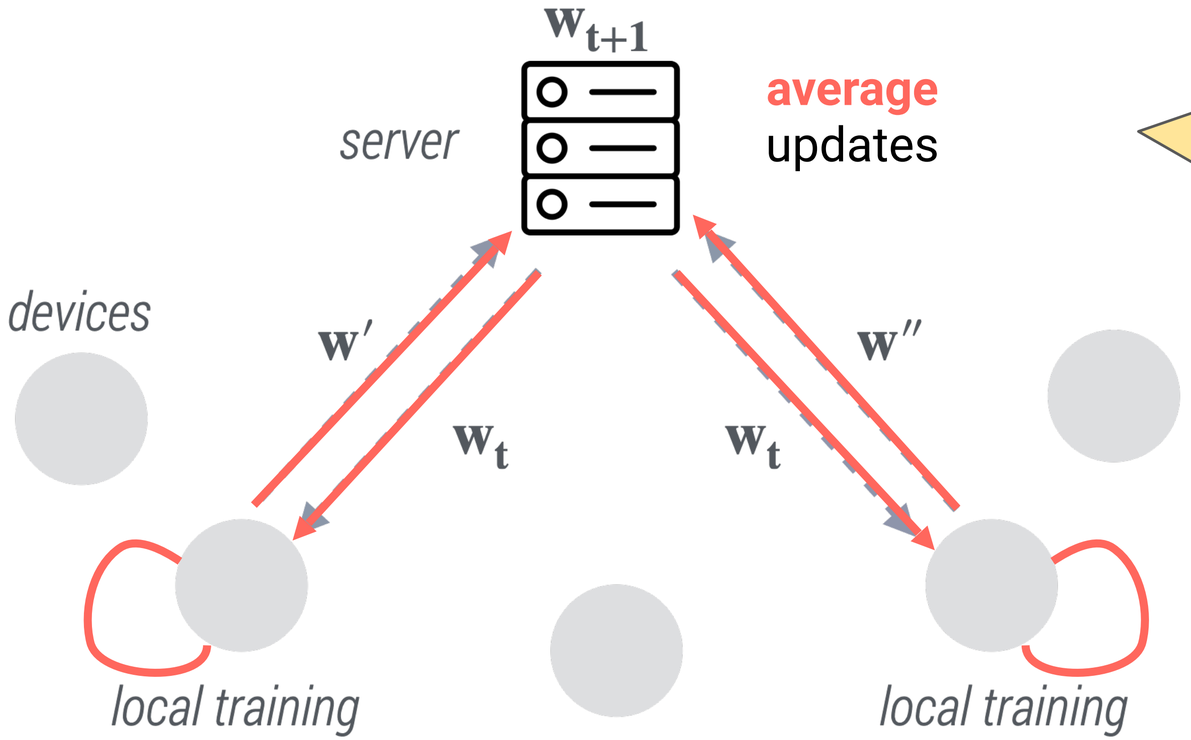
- unbalanced, non-IID data

systems heterogeneity

- variable hardware, connectivity, etc

A STANDARD BASELINE

Federated Averaging (FedAvg)



(e.g. **E** epochs of SGD)

(e.g. **E** epochs of SGD)

- At each communication round:
  (i) run SGD locally, then
  (ii) average the model updates
- Can add privacy mechanisms to procedure
- Reduces communication by:
  (i) performing local updating,
  (ii) communicating with a subset of devices

How does FedAvg differ from distributed SGD?

Distributed SGD: computation on device k

for $i \in$ *mini-batch* B
  $\Delta \mathbf{w} \leftarrow \Delta \mathbf{w} - \alpha \nabla f_i(\mathbf{w})$
end
$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$

FedAvg: computation on device k

for $t = 1, 2, \ldots,$ *local iterations* T
  $\Delta \mathbf{w} \leftarrow \Delta \mathbf{w} - \alpha \nabla f_{i_t}(\mathbf{w})$
  $\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$
end

Why is it useful to perform `local-updating`?

1. Can perform more local computation (i.e., more than just one mini-batch)
2. Incorporate updates more quickly (immediately apply gradient information)

✓ **Can lead to method converging in many fewer communication rounds**

How does FedAvg differ from distributed SGD?

MNIST CNN, 99% ACCURACY							
CNN	E	B	u	IID		NON-IID	
FEDSGD	1	∞	1	626		483	
FEDAVG	5	∞	5	179	(3.5×)	1000	(0.5×)
FEDAVG	1	50	12	65	(9.6×)	600	(0.8×)
FEDAVG	20	∞	20	234	(2.7×)	672	(0.7×)
FEDAVG	1	10	60	34	(18.4×)	350	(1.4×)
FEDAVG	5	50	60	29	(21.6×)	334	(1.4×)
FEDAVG	20	50	240	32	(19.6×)	426	(1.1×)
FEDAVG	5	10	300	20	(31.3×)	229	(2.1×)
FEDAVG	20	10	1200	18	(34.8×)	173	(2.8×)

SHAKESPEARE LSTM, 54% ACCURACY							
LSTM	E	B	u	IID		NON-IID	
FEDSGD	1	∞	1.0	2488		3906	
FEDAVG	1	50	1.5	1635	(1.5×)	549	(7.1×)
FEDAVG	5	∞	5.0	613	(4.1×)	597	(6.5×)
FEDAVG	1	10	7.4	460	(5.4×)	164	(23.8×)
FEDAVG	5	50	7.4	401	(6.2×)	152	(25.7×)
FEDAVG	5	10	37.1	192	(13.0×)	41	(95.3×)

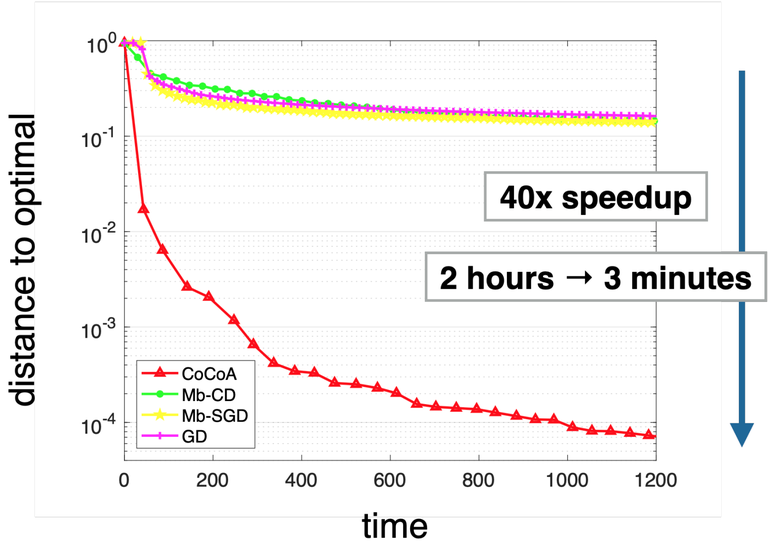
Local-updating (FedAvg) can reduce communication rounds by ~100x relative to SGD

[McMahan et al, Communication-efficient learning of deep networks from decentralized data, AISTATS 2017]

ASIDE

Local-updating is not new

- Extreme setting: one-shot averaging
  (e.g., [Zhang, Duchi, Wainwright, Communication-Efficient Algorithms for Statistical Optimization, JMLR 2013])
- Consensus-based optimization
  - ADMM
    [Boyd et al, Distributed Optimization and Statistical Learning via ADMM, FnT in ML, 2010]
  - CoCoA
    [Jaggi & Smith et al, Communication-Efficient Distributed Dual Coordinate Ascent, NeurIPS 2014]



ASIDE

Local-updating is not new

- Extreme setting: one-shot averaging
  (e.g., [Zhang, Duchi, Wainwright, Communication-Efficient Algorithms for Statistical Optimization, JMLR 2013])
- Consensus-based optimization
  - ADMM
    [Boyd et al, Distributed Optimization and Statistical Learning via ADMM, FnT in ML, 2010]
  - CoCoA
    [Jaggi & Smith et al, Communication-Efficient Distributed Dual Coordinate Ascent, NeurIPS 2014]
- Local-SGD
  (e.g., [MacDonald et al, Efficient large-scale distributed training of conditional maxent models, NeurIPS 2009])
- Decentralized optimization

Federated setting is distinct in considering local-updating with heterogeneous data and partial device participation, often for non-convex objectives

How does FedAvg differ from distributed SGD?

Distributed SGD: computation on device k

$$\begin{aligned}&\textbf{for } \; i \in \; \textit{mini-batch } B \\ &\quad \Delta\mathbf{w} \leftarrow \Delta\mathbf{w} - \alpha \nabla f_i(\mathbf{w}) \\ &\textbf{end} \\ &\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}\end{aligned}$$

FedAvg: computation on device k

$$\begin{aligned}&\textbf{for } \; t = 1, 2, \ldots, \; \textit{local iterations } T \\ &\quad \Delta\mathbf{w} \leftarrow \Delta\mathbf{w} - \alpha \nabla f_{i_t}(\mathbf{w}) \\ &\quad \mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w} \\ &\textbf{end}\end{aligned}$$

Why is it useful to perform `local-updating`?

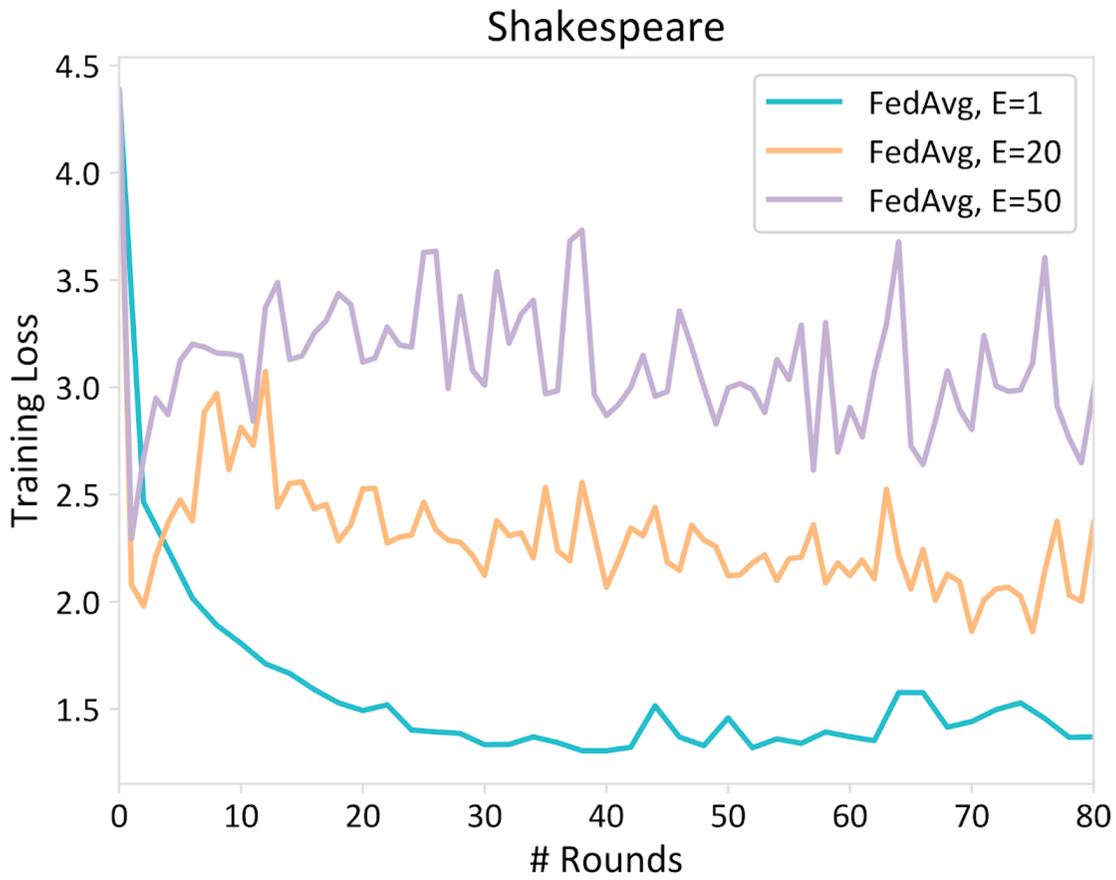
1. Can perform more local computation (i.e., more than just one mini-batch)
2. Incorporate updates more quickly (immediately apply gradient information)

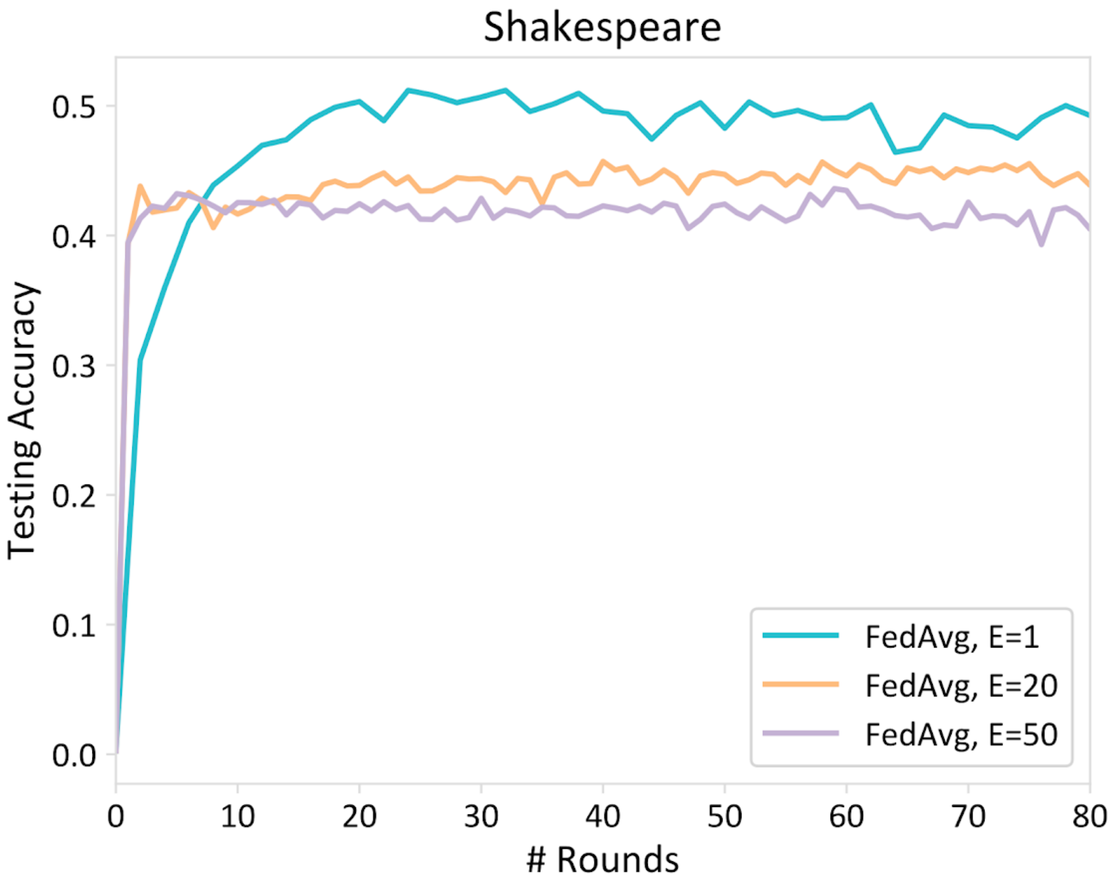
✓ **Can lead to method converging in many fewer communication rounds**

✗ **But, can potentially hurt convergence if not properly tuned …**

WILL THIS CONVERGE?

Challenge: heterogeneity



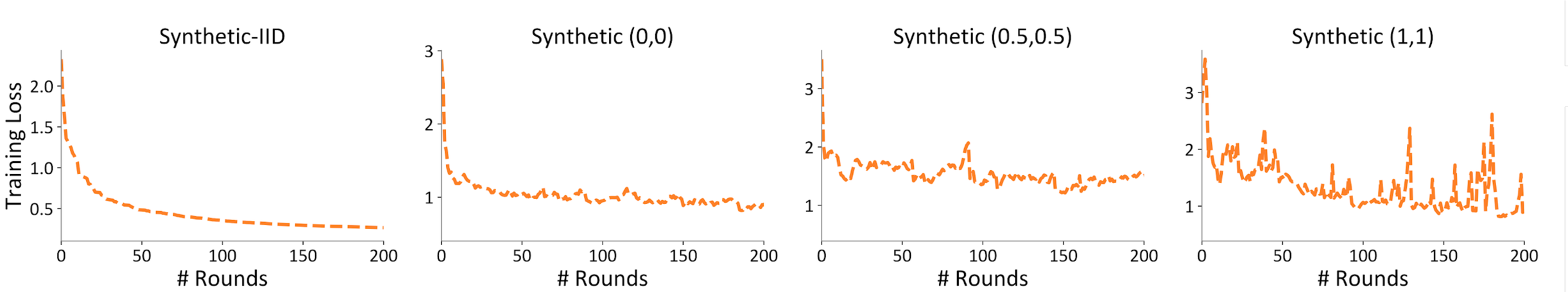


[Li et al, Federated optimization in heterogeneous networks, MLSys 2020]

WILL THIS CONVERGE?

Challenge: heterogeneity

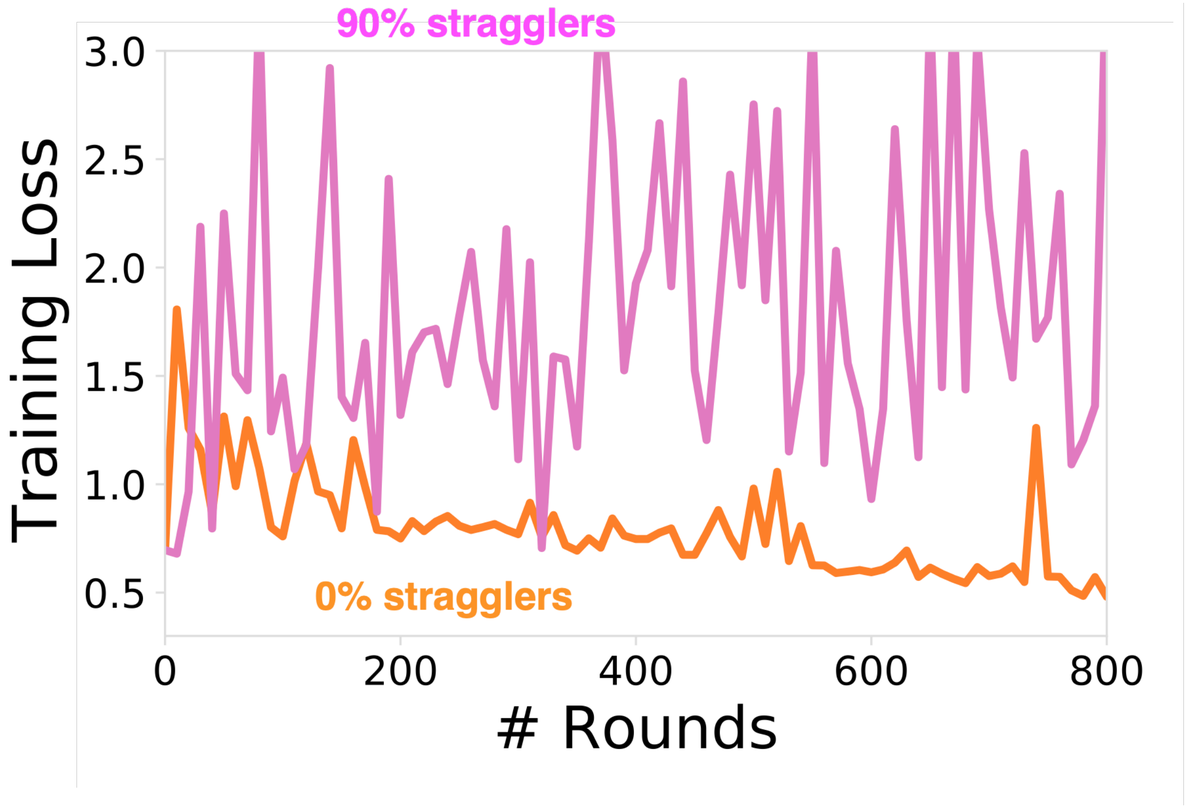
effect of heterogeneity on FedAvg convergence, assuming all other hyperparameters fixed



[Li et al, Federated optimization in heterogeneous networks, MLSys 2020]

WILL THIS CONVERGE?

Challenge: heterogeneity

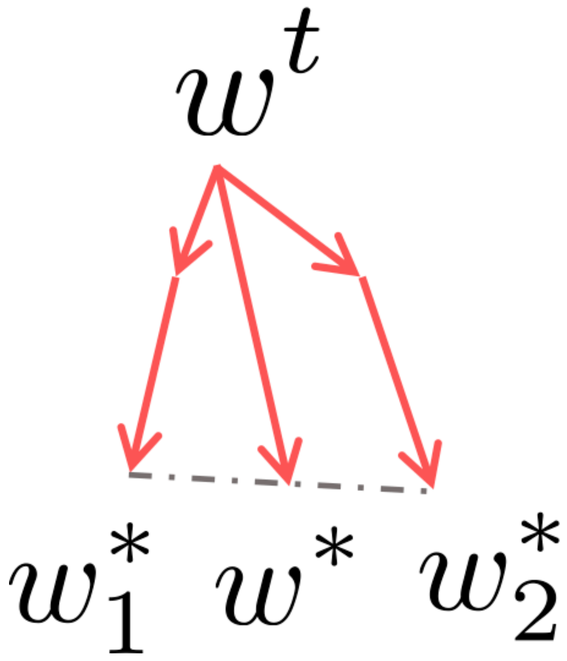


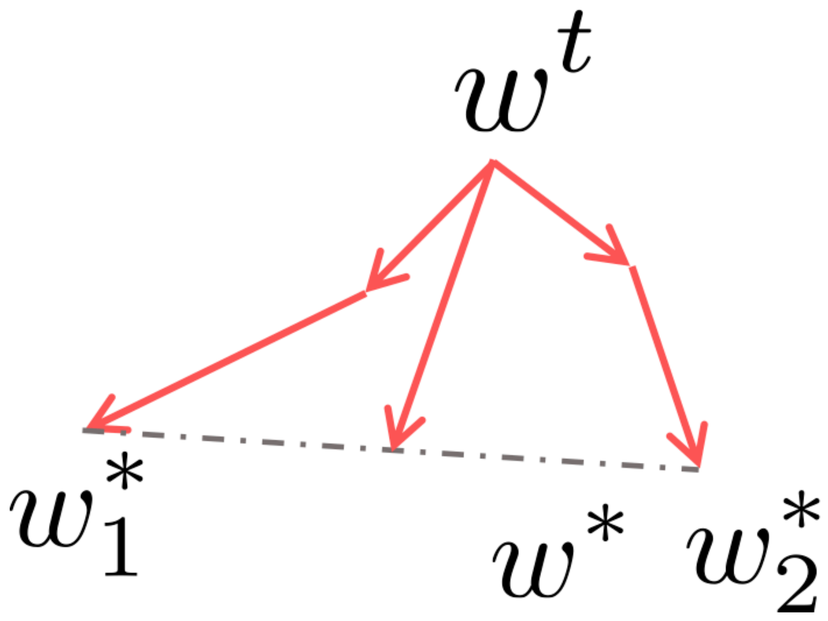
systems heterogeneity (e.g., dropping devices*) can exacerbate convergence issues

*[Bonawitz, et al. Towards Federated Learning at Scale: System Design, MLSys, 2019]
[Li et al, Federated optimization in heterogeneous networks, MLSys 2020]

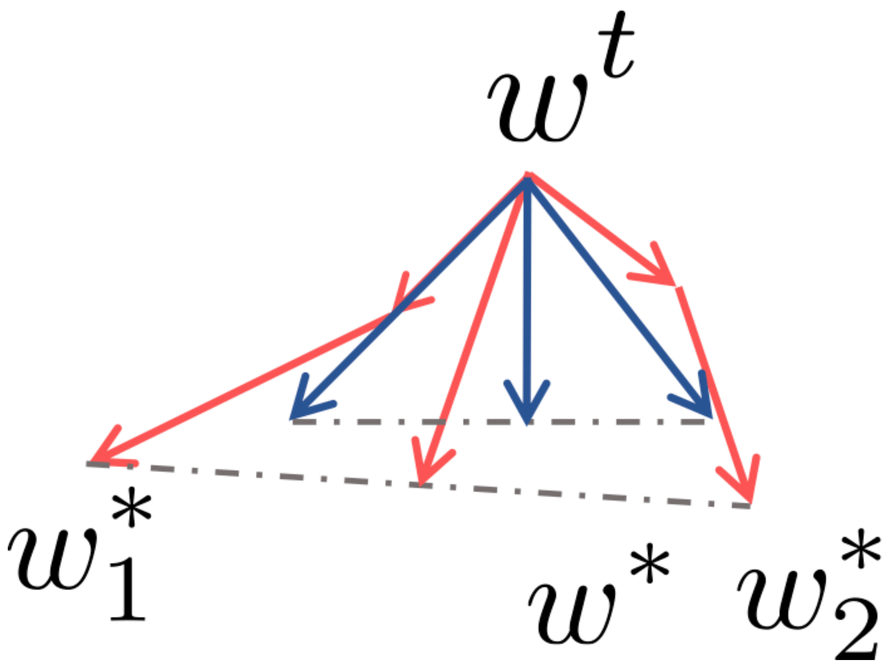
WILL THIS CONVERGE?

Challenge: heterogeneity





Simple modification: FedProx



$$\min_{w_k} F_k(w_k) + \frac{\mu}{2} \left\| w_k - w^t \right\|^2$$

proximal term

- proximal term *limits the impact of heterogeneous local updates*
- don't drop devices: instead [safely] incorporate partial work
- theoretical convergence guarantees (intuitively, to guarantee convergence you need to set mu appropriately to account for degree of heterogeneity)

[Li et al, Federated optimization in heterogeneous networks, MLSys 2020]

optimization

- *how does heterogeneity affect federated optimization methods?*

- heterogeneity can lead to:
  - slower convergence, reduced stability, divergence
- critical to analyze and evaluate federated methods with:
  - non-IID data, partial / variable participation

on heterogeneity in federated settings

optimization

- *how does heterogeneity affect federated optimization methods?*

fairness

- *can we equalize performance across diverse networks?*

modeling

- *how can we personalize models?*

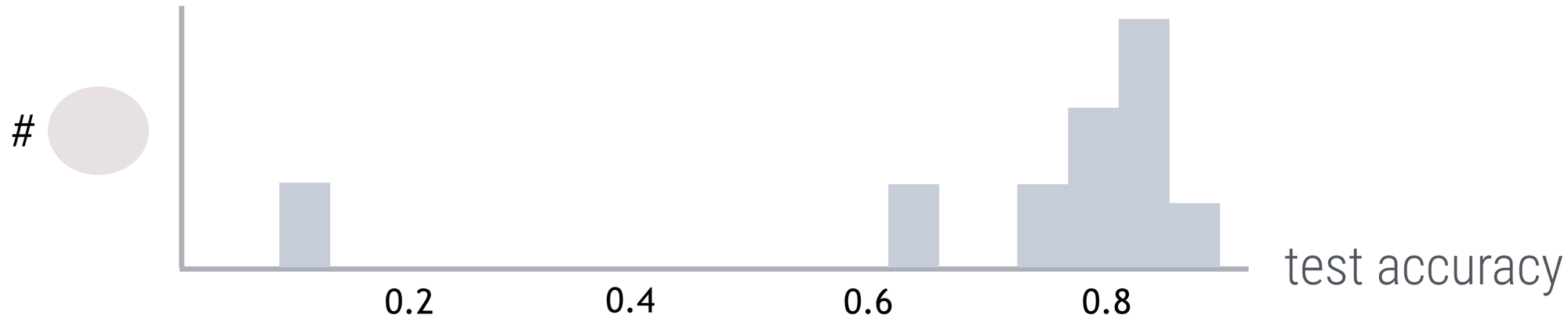
Fair Resource Allocation in Federated Learning
Li, Sanjabi, Beirami, Smith, ICLR 2020

FL: traditional empirical risk minimization

ERM: $\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$

potential issues:

- no accuracy guarantees for individual devices
- performance can vary widely across network



FL: traditional empirical risk minimization

ERM: $\min_{w} \left(p_1 F_1 + p_2 F_2 + \cdots + p_N F_N \right)$

potential issues:

- no accuracy guarantees for individual devices
- performance can vary widely across network

Can we encourage a more fair (i.e., uniform) distribution of the model performance across devices?

fair resource allocation objective

q-FFL: $$\min_w \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

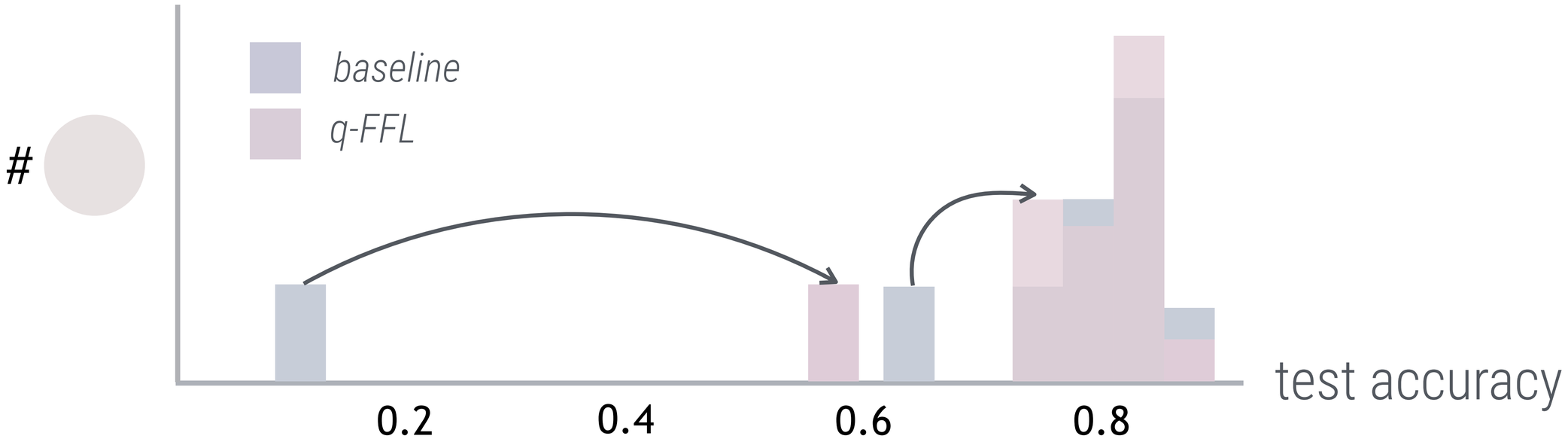
- inspired by α-fairness for fair resource allocation in wireless networks
- a tunable framework ($q \to 0$: previous objective; $q \to \infty$: minimax fairness*)
- theory: increasing q results in more uniform accuracy distributions (e.g., reduced variance)

*[*Fairness without Demographics in Repeated Loss Minimization*, Hashimoto et al, ICML 2018]

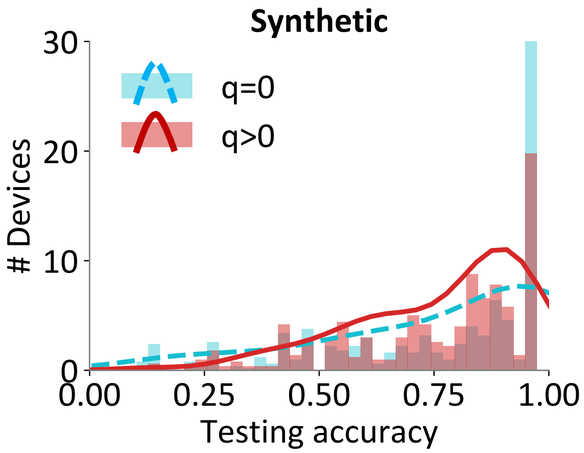
*[*Agnostic Federated Learning*, Mohri, Sivek, Suresh, ICML 2019]

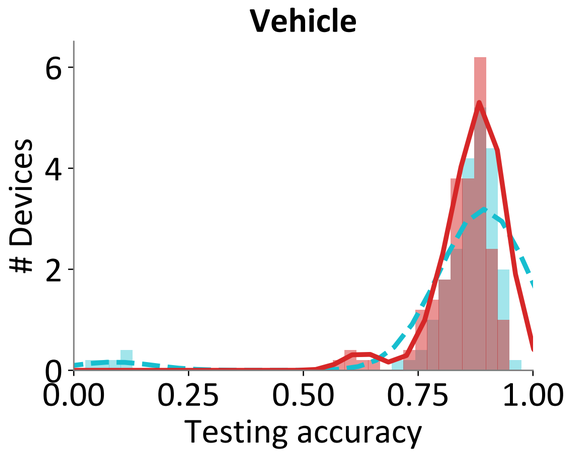
fair resource allocation objective

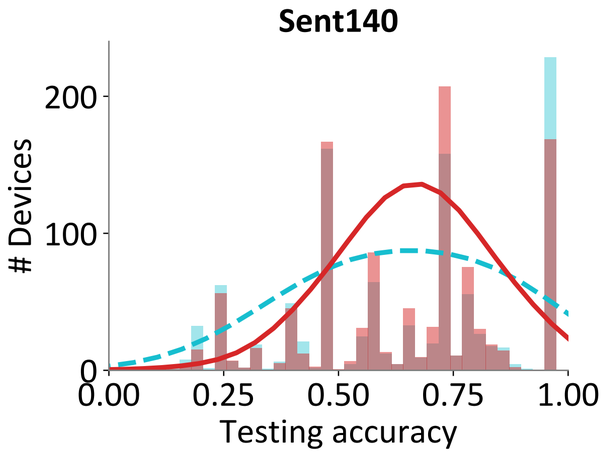
q-FFL: $$\min_{w} \frac{1}{q+1}\left(p_1 F_1^{q+1} + p_2 F_2^{q+1} + \cdots + p_N F_N^{q+1}\right)$$

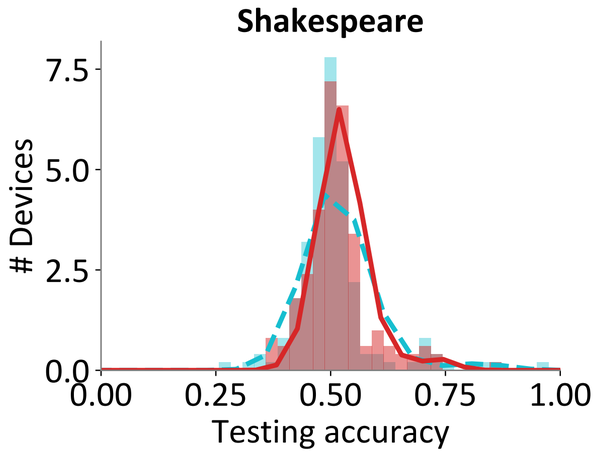


empirical results









on average,
can cut variance in half
while maintaining mean accuracy

fairness

- *can we equalize performance across diverse networks?*

- vanilla ERM may deliver poor quality of service to heterogeneous networks
- q-FFL allows for flexible trade-off between fairness and accuracy

on heterogeneity in federated settings

optimization

- *how does heterogeneity affect federated optimization methods?*

fairness

- *can we equalize performance across diverse networks?*

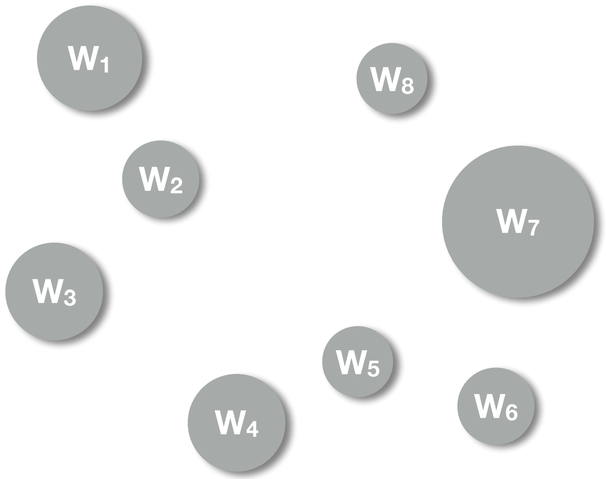
modeling

- *how can we personalize models?*

Federated Multi-Task Learning
Smith, Chiang, Sanjabi, Talwalkar, NeurIPS 2017

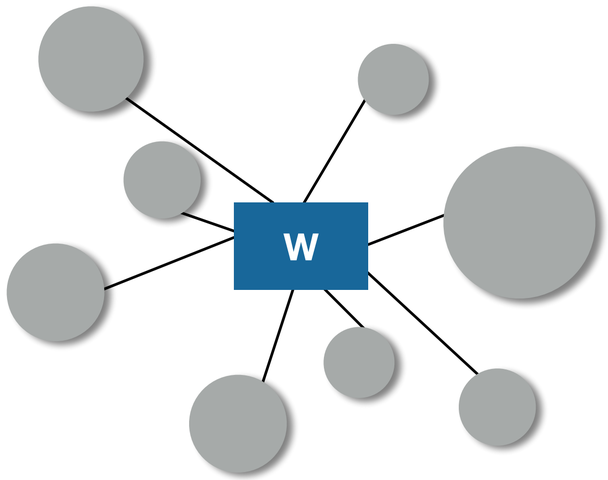
how to model federated data?

local



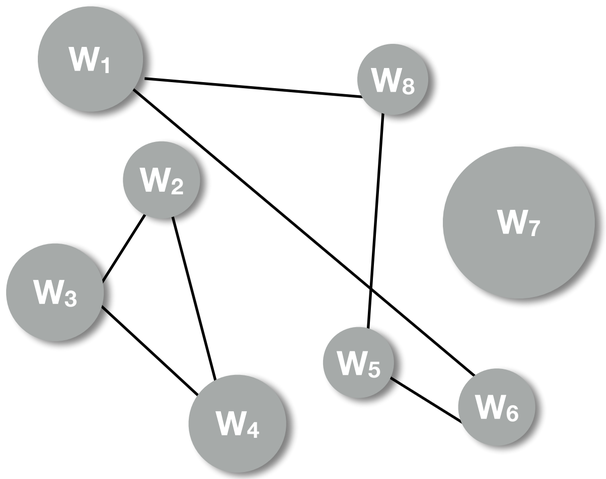
✓ *personalized models*
✗ *don't learn from peers*

global



✗ *non-personalized models*
✓ *learn from peers*

??



✓ *personalized models*
✓ *learn from peers*

multi-task learning

$$\min_{\mathbf{W},\boldsymbol{\Omega}} \sum_{t=1}^{m} \sum_{i=1}^{n_t} \ell_t(\mathbf{w}_t, \mathbf{x}_t^i) + \mathcal{R}(\mathbf{W}, \boldsymbol{\Omega})$$

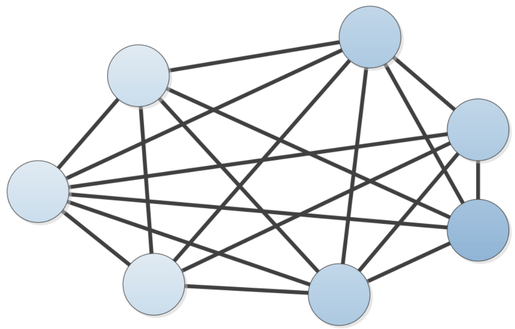
models

task relationship

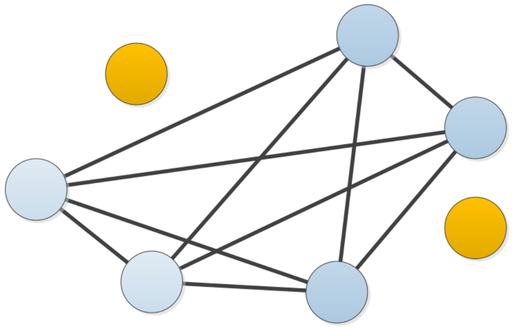
losses

regularizer

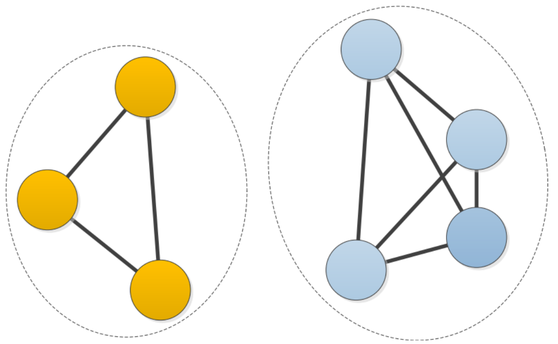
all tasks related



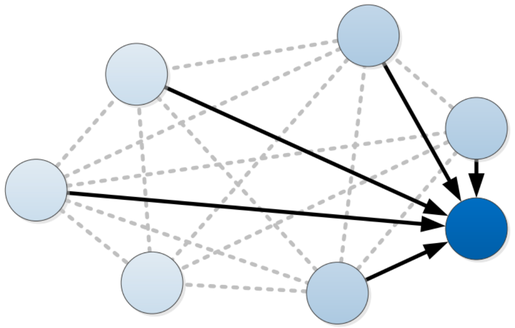
outlier tasks



clusters / groups



asymmetry



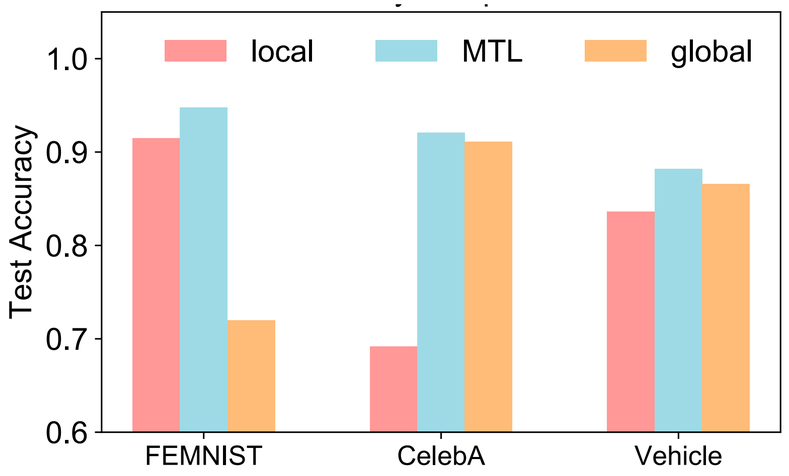
[EP, KDD 2004; JVB, NIPS 2009; ZCY, NIPS 2011]

example: prediction error on federated data

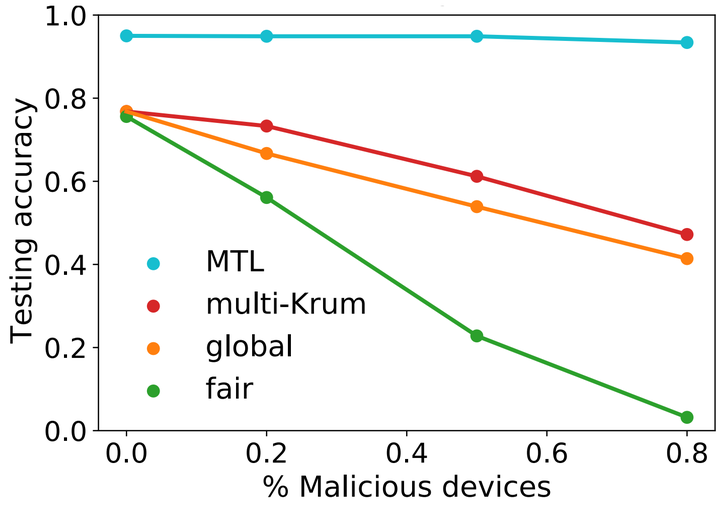
	Global	Local	MTL
Human Activity 	2.23 (0.30)	1.34 (0.21)	**0.46 (0.11)**
Google Glass 	5.34 (0.26)	4.92 (0.26)	**2.02 (0.15)**
Land Mine 	27.72 (1.08)	23.43 (0.77)	**20.09 (1.04)**
Vehicle Sensor 	13.4 (0.26)	7.81 (0.13)	**6.59 (0.21)**

benefits of personalization

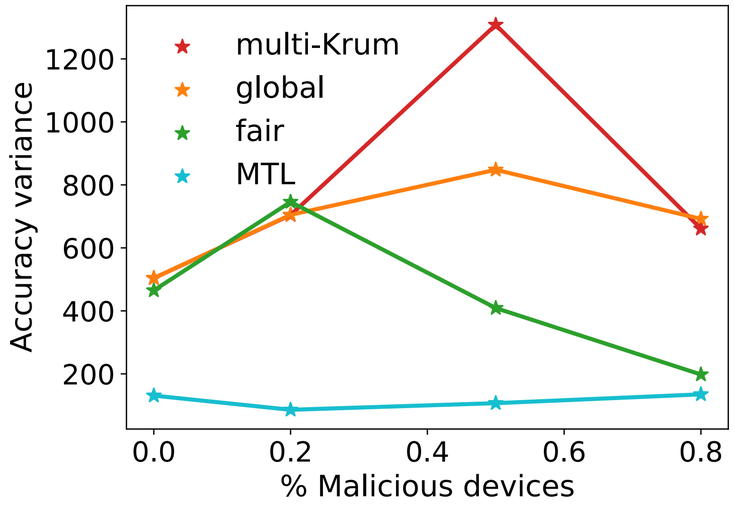
✓ accuracy



✓ robustness



✓ fairness



Ditto: Fair and Robust Federated Learning Through Personalization
Li, Hu, Beirami, Smith, ICML 2021

modeling

- *how can we personalize models?*

- multi-task learning is a principled way to learn personalized, shared models
- personalization not only improves accuracy, but also has benefits in terms of **robustness** and **fairness**

federated learning: workflow & challenges

Federated Learning: Challenges, Methods, and Future Directions,
T. Li, A. K. Sahu, A. Talwalkar, V. Smith,
IEEE Signal Processing Magazine 2020

Federated Learning and Analytics: Industry Meets Academia
P. Kairouz, B. McMahan, V. Smith
NeurIPS Tutorial, https://slideslive.com/38935813/federated-learning-tutorial

expensive communication
- massive, slow, unreliable networks

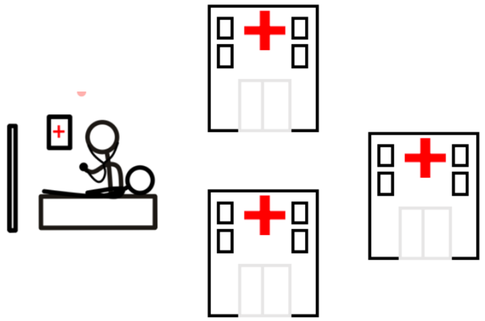
privacy concerns
- user privacy constraints

statistical heterogeneity
- unbalanced, non-IID data

systems heterogeneity
- variable hardware, connectivity, etc

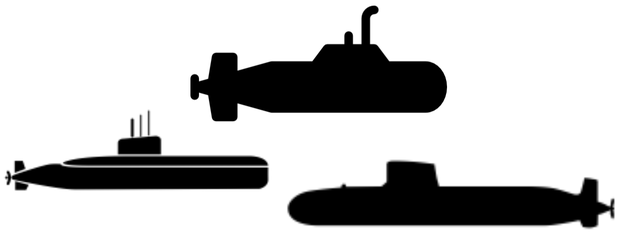
non-mobile FL applications

smart home
behavior modeling

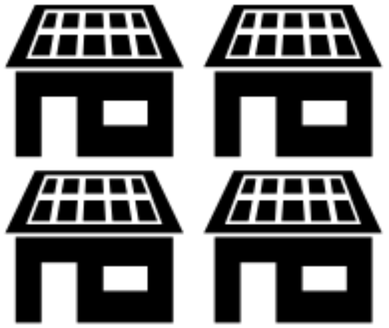


medical imaging
detection

anomaly
detection in
submersibles



power grid
usage prediction



Carnegie Mellon University

Symantec.

UNIVERSITY OF PITTSBURGH

UPMC
LIFE CHANGING MEDICINE

ASSETT

UNITED STATES NAVY

MIT

borsetta