A2FL: Availability-Aware Selection for Machine Learning on Clients with Federated Big Data

IEEE ICC 2023

Rome, Italy - May 2023

Ahmed M. Abdelmoniem , Yomna M. Abdelmoniem , Ahmed Elzanaty

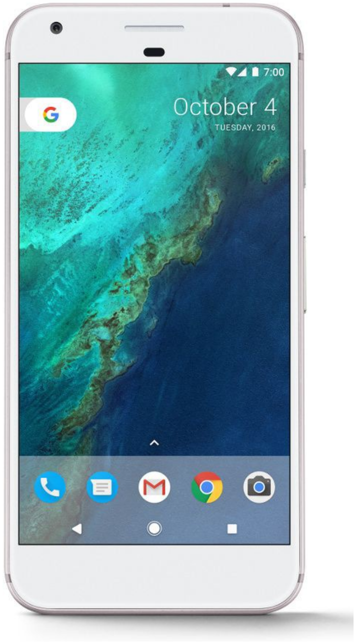
Queen Mary University of London

ASSIUT UNIVERSITY

UNIVERSITY OF SURREY

Data lives at the Edge

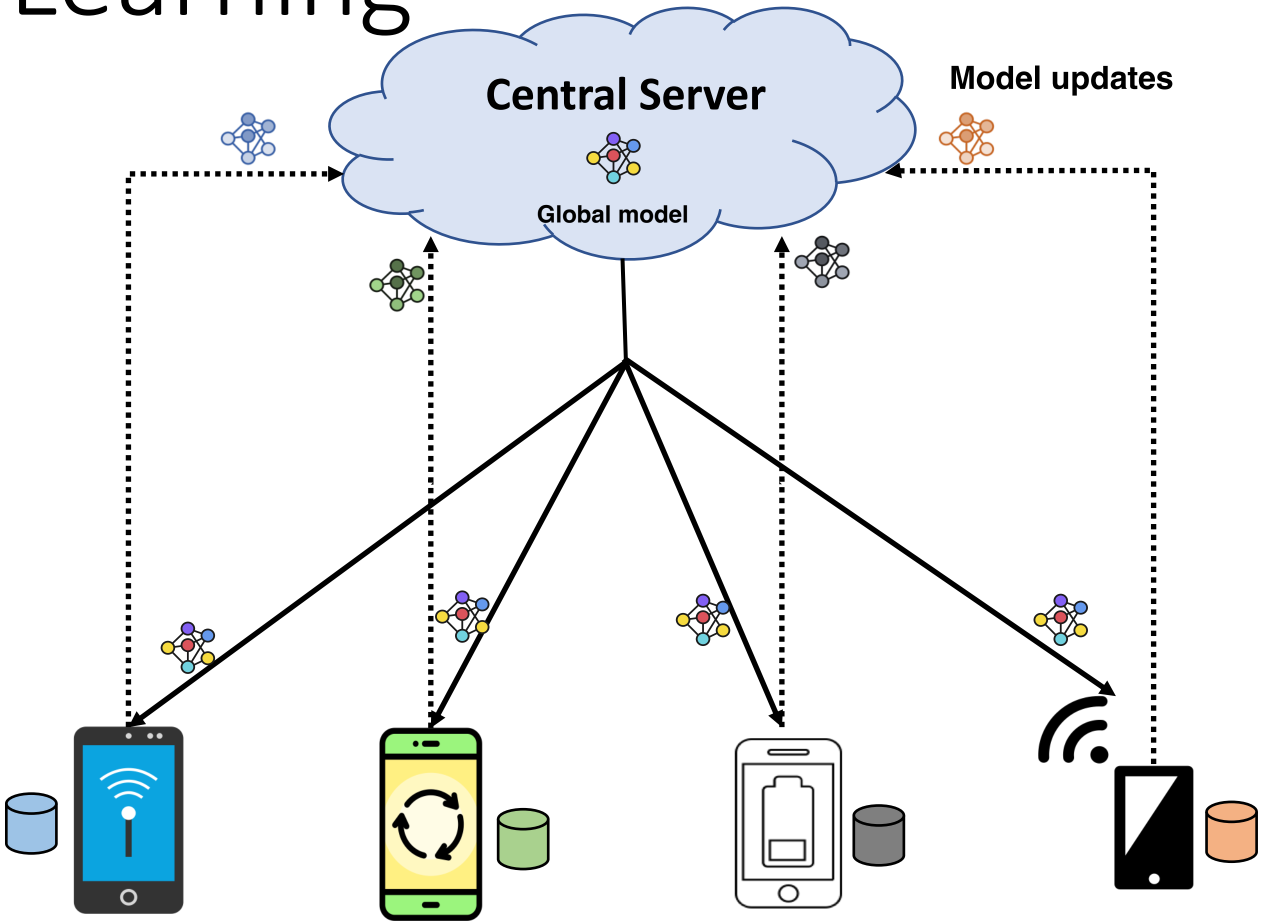
- Billions of phones & IoT devices constantly generate data
- Data enables better products and smarter models
- On-device processing (e.g., inference for mobile keyboards)
  - advanced specialized hardware (e.g., GPU and NPUs on mobile/IoT devices)
- Benefits
  - Improved latency
  - Works offline
  - Better battery life
  - Privacy advantages



What about analytics & learning?

Centralized vs Federated Learning

- Centralized Training:
  - Central (Data) server
  - Expensive data movement
  - **Communication-intensive**
  - **Privacy concerns**
- Federated Learning:
  - Central (Aggregation) server
  - Model exchange
  - **Communication-efficient**
  - **Privacy-preserving (Differential privacy + secure aggregation)**



Practical Use-cases of Federated Learning (FL)

What are good applications for FL?

- On-device data is more relevant than server-side data (or lack of it)
- On-device data is privacy sensitive or large to communicate
- Labels can be inferred naturally from user interaction

Gboard: next-word prediction

Using FL, better next-word prediction accuracy: +24%

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

Apple: Voice recognition

MIT Technology Review
Sign in
Subscribe

Artificial intelligence / Machine learning

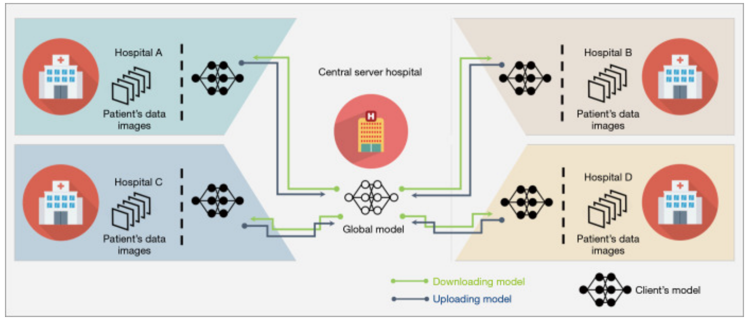
How Apple personalizes Siri without hoovering up your data

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

by Karen Hao

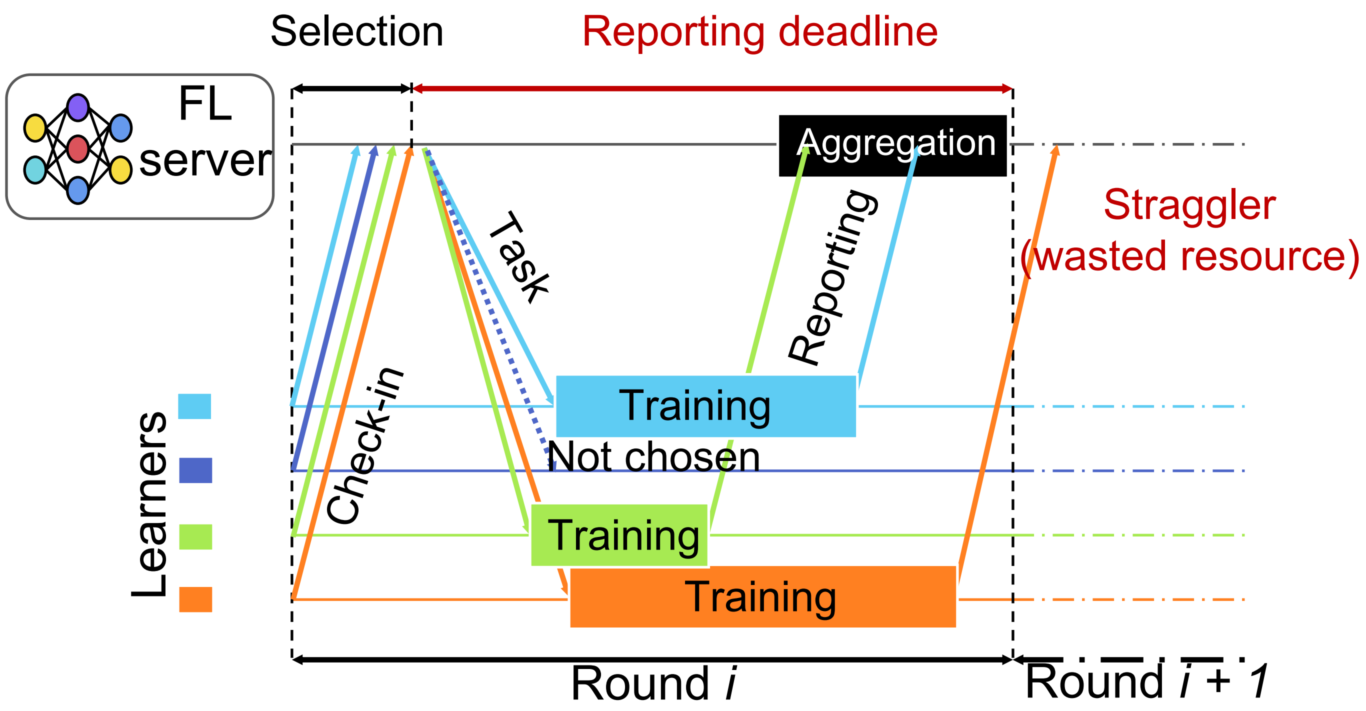
December 11, 2019

Medical Imaging



Ng D, Lan X, Yao MM, Chan WP, Feng M. Federated learning: a collaborative effort to achieve better medical imaging models for individual sites that have small labelled datasets. Quant Imaging Med Surg. 2021

Federated Learning Life-cycle



Heterogeneity impacts quality and time!

- Heterogenous data distributions → non-IID setting (quality)
- Diverse hardware and network capabilities → stragglers (time)
- Clients are not always available → inclusivity is hard (quality)

Non-Practical Selection Methods

- Most existing methods aim to improve the **time-to-quality**

Reduce Time

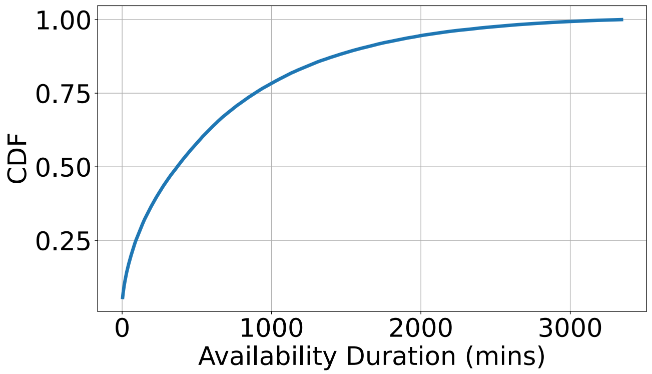
FedCS (ICC'18), Oort (OSDI'21)

- Biases the client selection to **reduce the training time** by exploiting the fast learners

Improve Quality

AdaPow (AISTATS'22)

- Biases the client selection towards ones with high loss to **boost model quality**



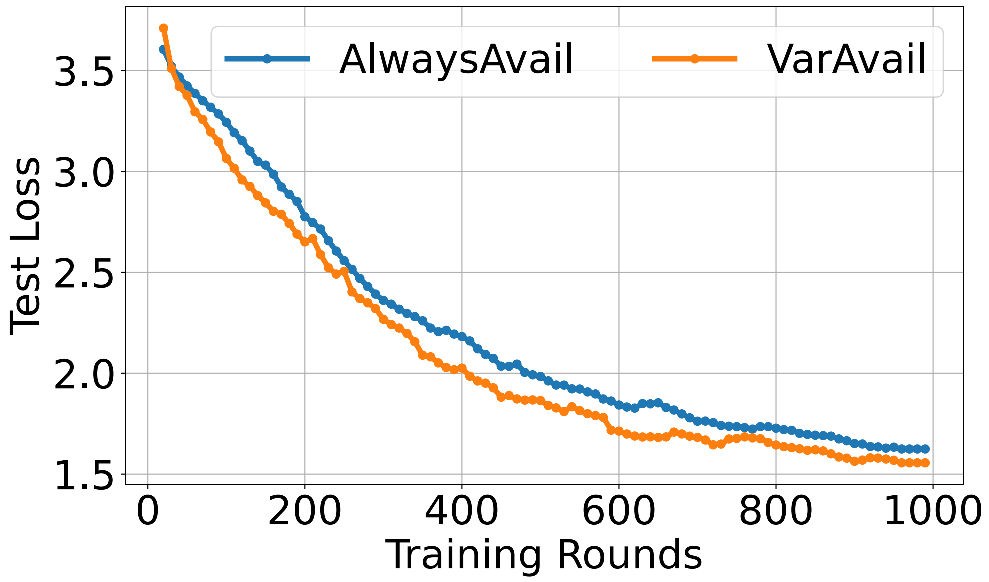
Disregards clients' availability

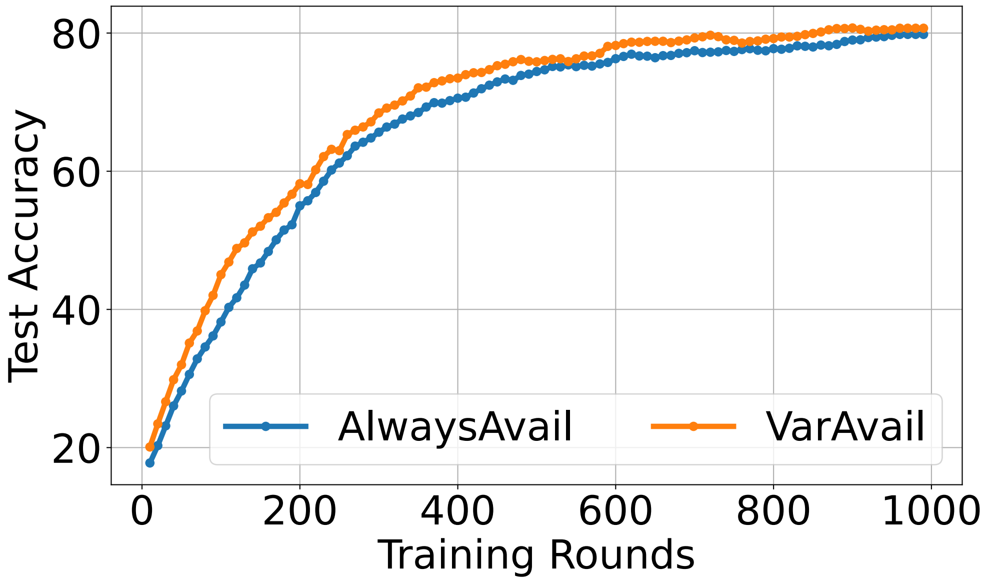
(low inclusivity)

FedCS → T. Nishio, R. Yonetani, Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge, ICC, 2018
Oort → F. Lai, X. Zhu, H. V. Madhyastha, M. Chowdhury, Efficient Federated Learning via Guided Participant Selection, USENIX OSDI, 2021
AdaPow → Yae Jee Cho, Jianyu Wang, Gauri Joshi, Towards Understanding Biased Client Selection in Federated Learning, AISTATS, 2022

Availability does NOT matters in IID case

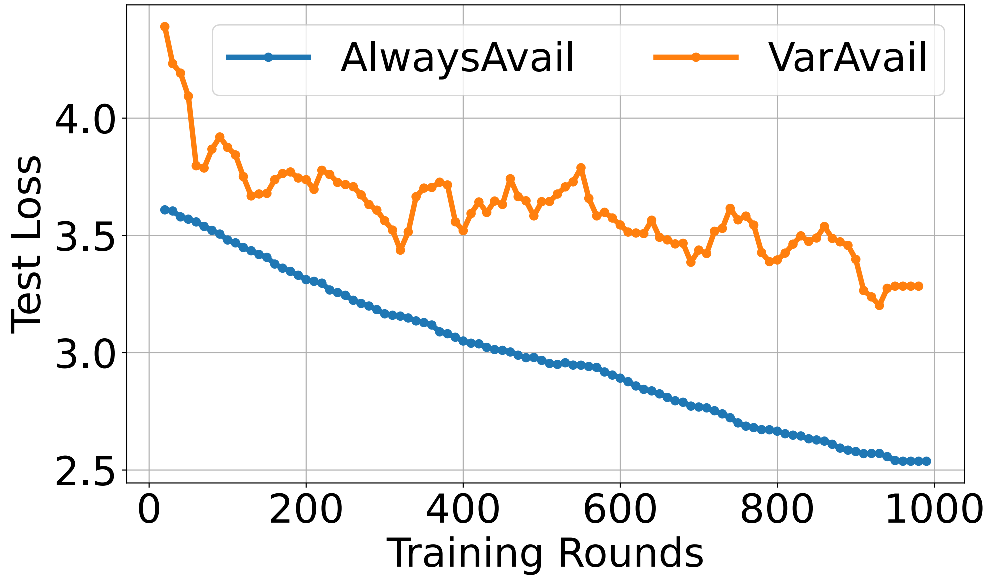
- **Availability does not impact the model quality**
  - Oort in IID data distribution → client's data are uniformly distributed
  - Even biased selection (fast learners) → still can capture the global data distribution

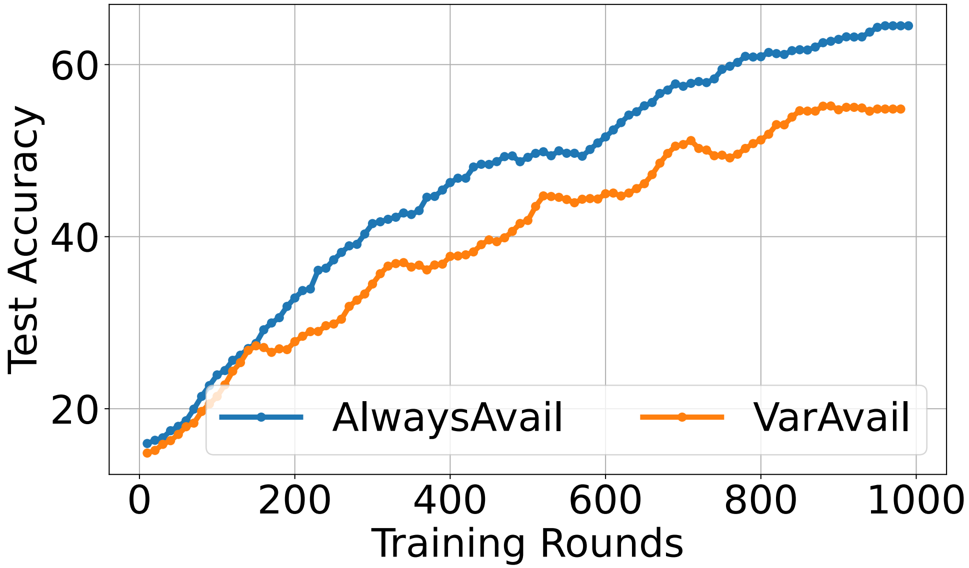




Motivation – Availability matters in non-IID case

- **Availability can impact model quality**
  - In non-IID data distribution → every client's data samples are important
  - Lack of inclusive selection → hard to capture the global data distribution





Takeaway & Proposed Solution

- Existing methods disregard client availability in the selection
- Biased selection can result in low resource diversity

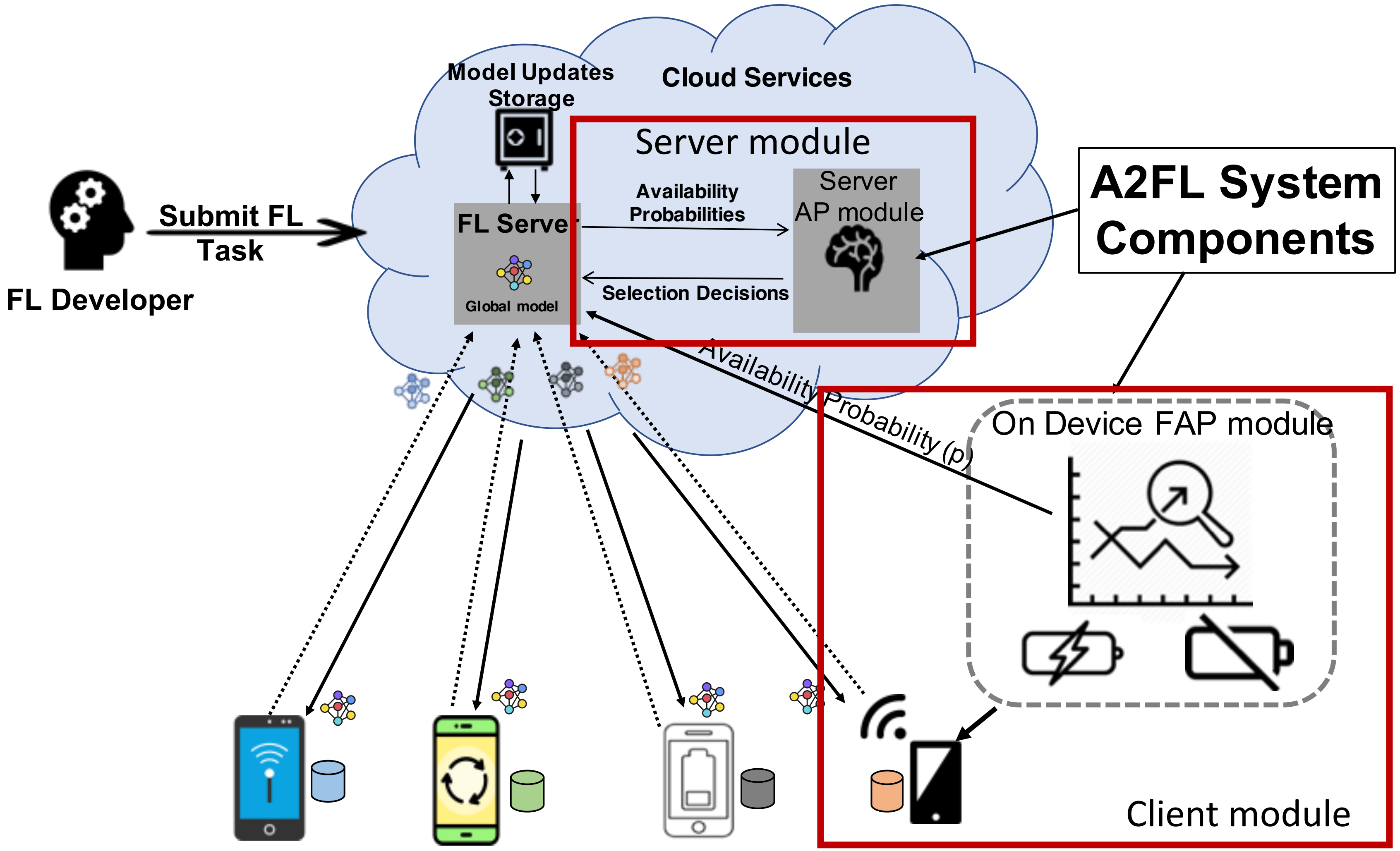
⬇

A2FL: Availability-Aware Federated Learning

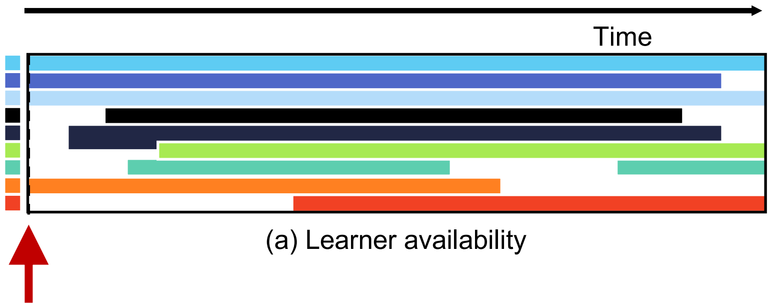
- ❑ **Selection**: prioritize selection of least available learners → Increases **diversity**
  - ❑ **Availability Prediction Module**: on-device prediction models (no privacy violation)
  - ❑ **Hybrid Selection Method:** the selection leverages both availability prioritization and random sampling

In the paper → more detailed description and discussion of the algorithm

System Design



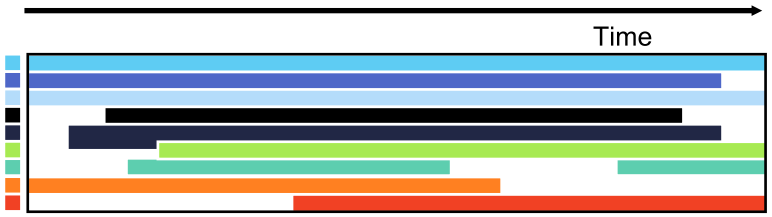
An Illustrative Example



(a) Learner availability

A2FL selects among online clients using availability info and breaks ties by selecting at random

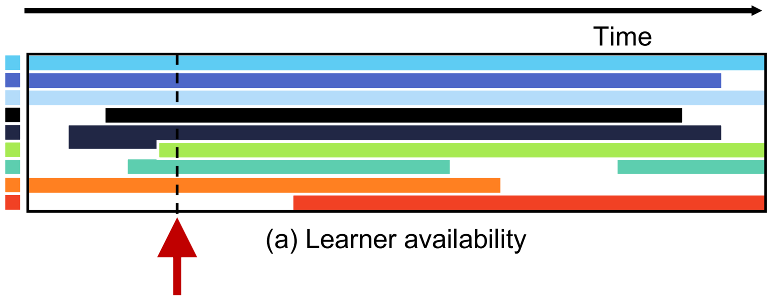
An Illustrative Example



(a) Learner availability

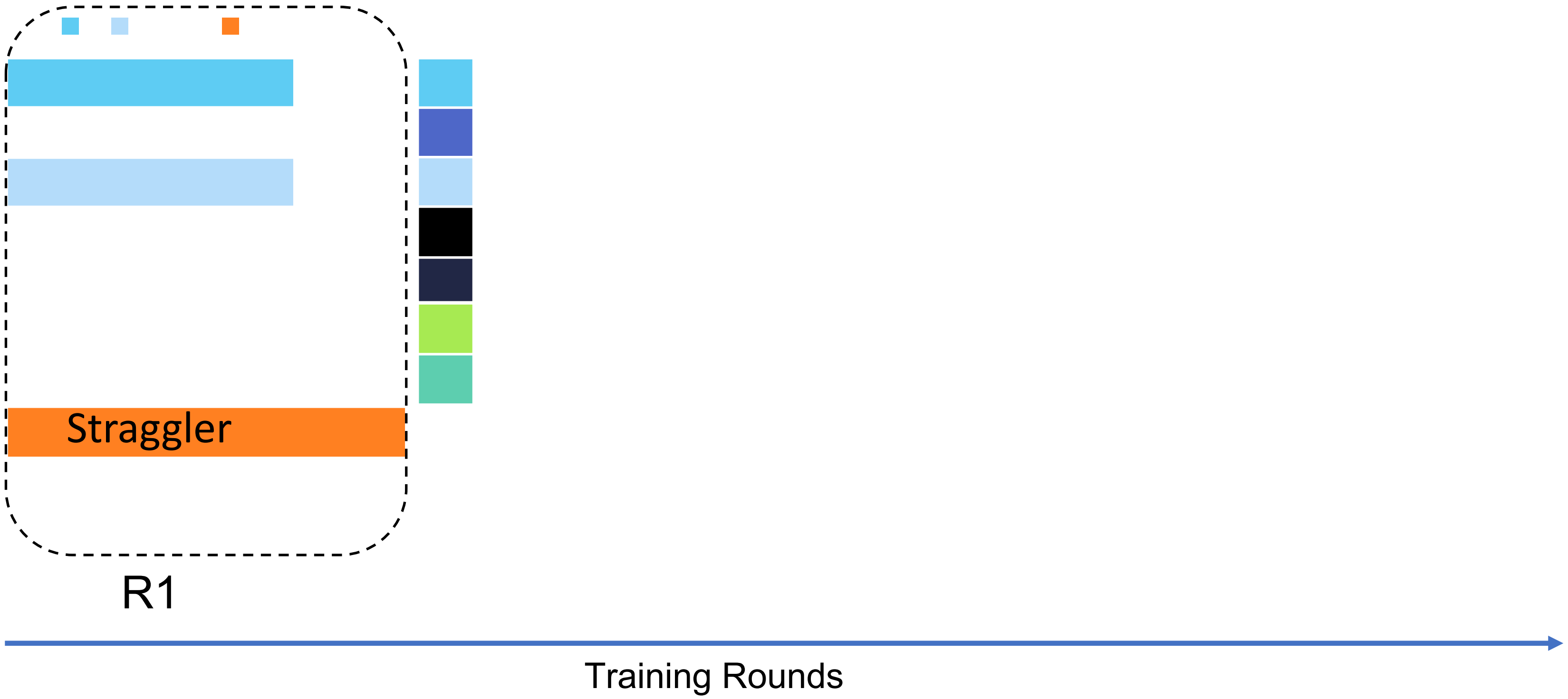
A2FL may select some stragglers but improves diversity, Oort only selects the fast learners

An Illustrative Example

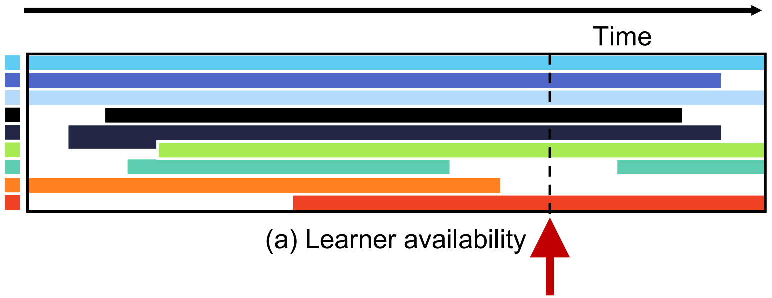


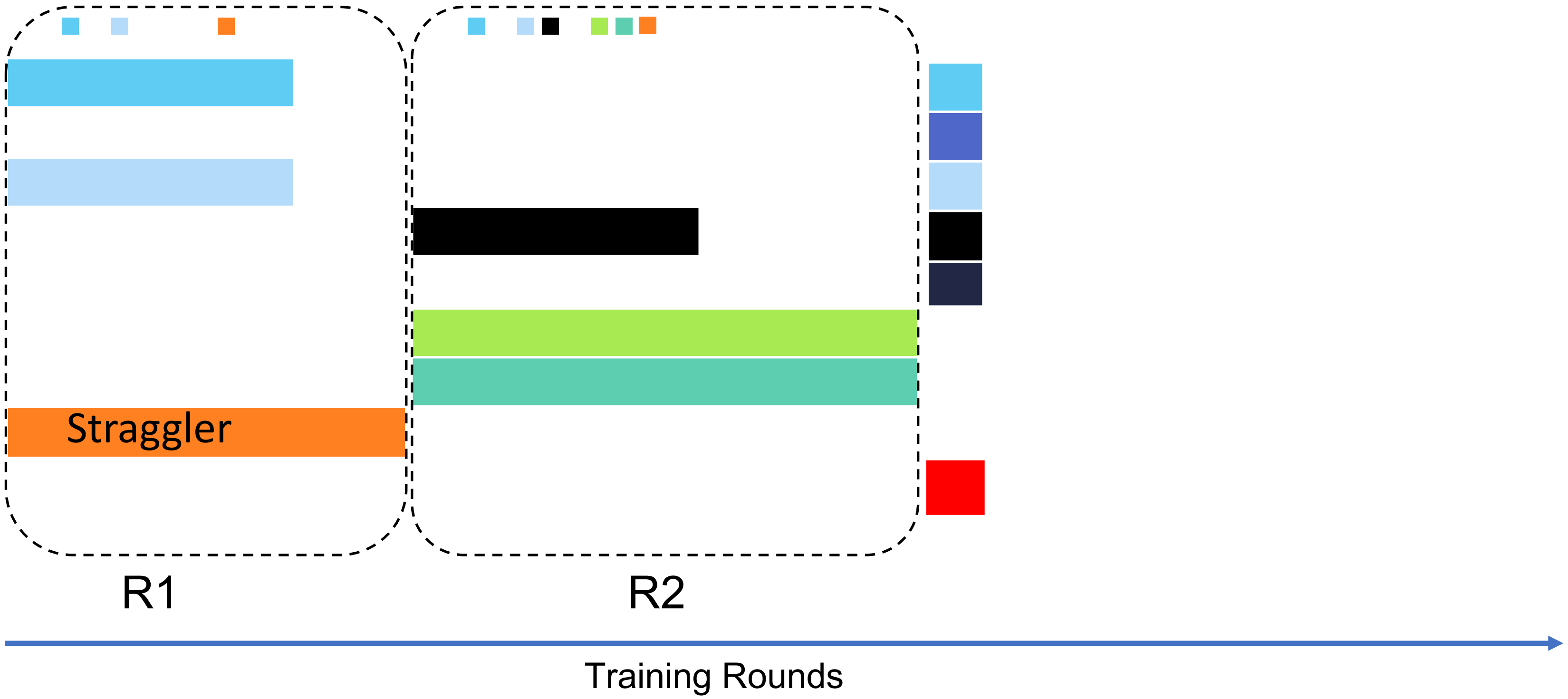
(a) Learner availability

A2FL selects clients with the least availability, Random selects regardless of future availability

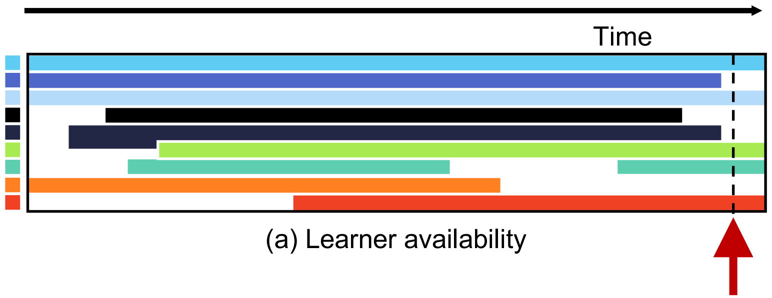


An Illustrative Example



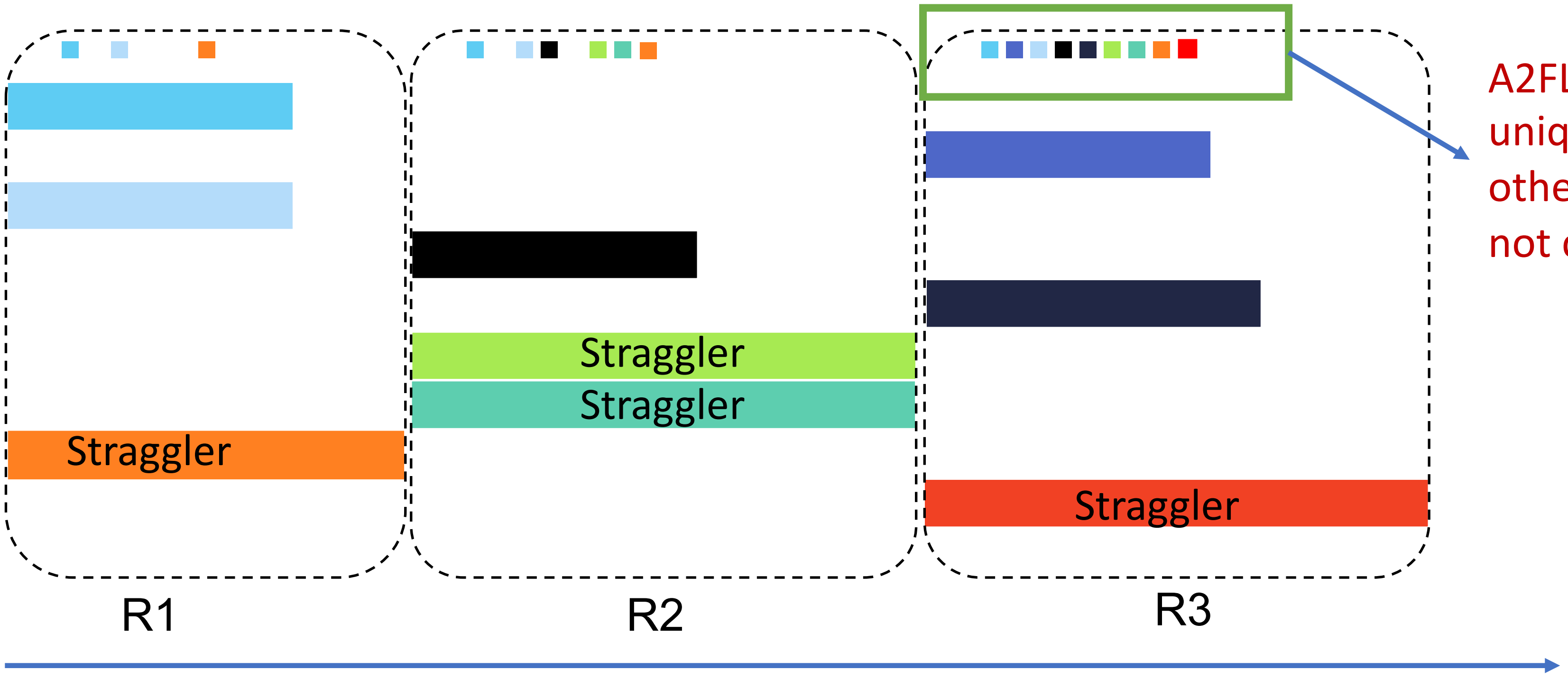


An Illustrative Example



(a) Learner availability

A2FL is able to achieve higher client diversity which improves the statistical efficiency of the model

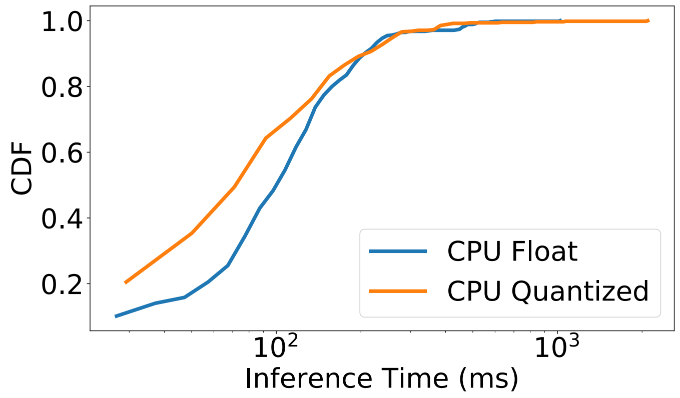


A2FL have high rate of unique clients, while other methods would not cover all the clients

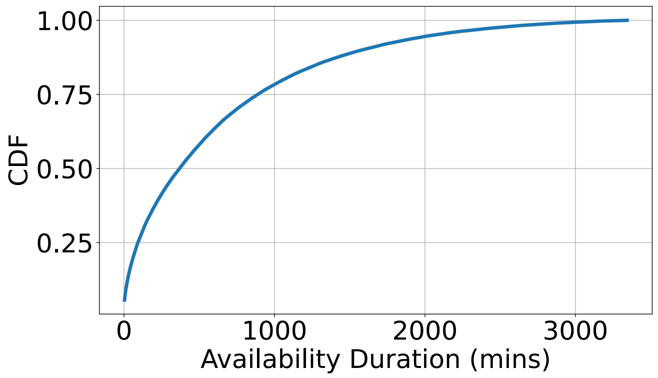
Experimental Evaluation

- FL Benchmarks using Google's Speech Recognition task [1]
- Various data distributions: IID, Label-limited (non-IID)

Heterogenous clients
Device compute profiles (AI benchmark)



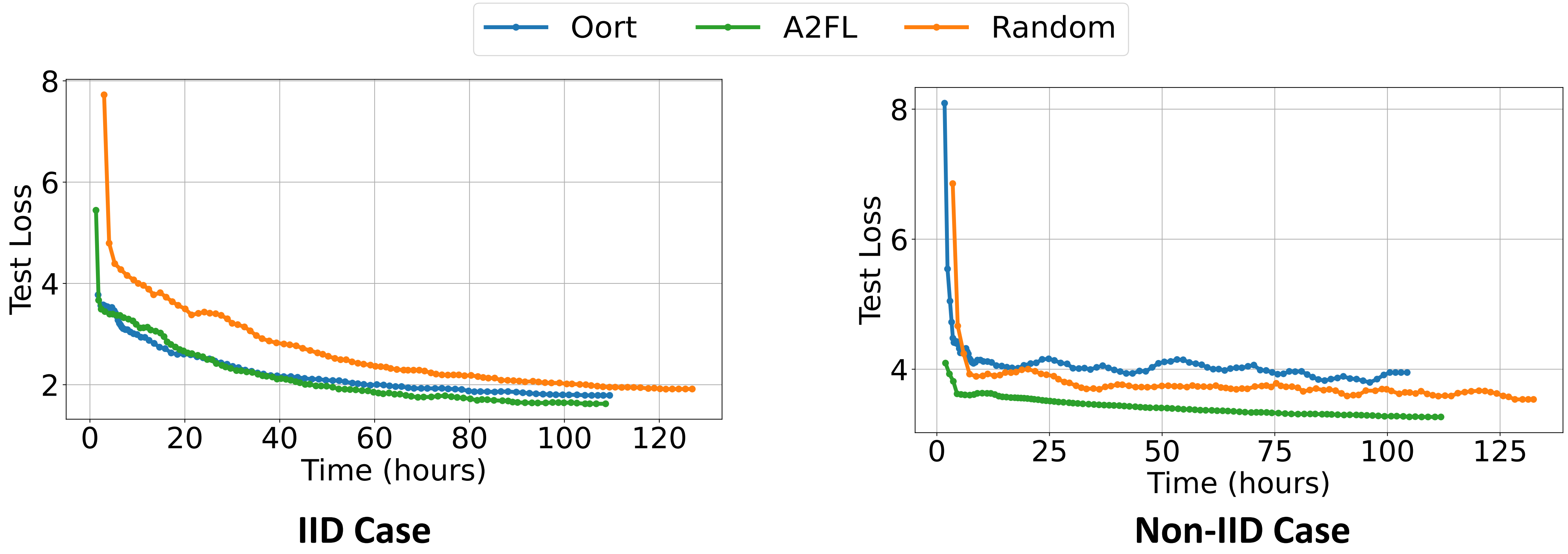
Heterogenous availability
User availability trace of 136K clients [2]



[1] F. Lai et al., "FedScale: Benchmarking Model and System Performance of Federated Learning". In *ICML*, 2022
[2] C. Yang et al., "FLASH: Heterogeneity-Aware Federated Learning at Scale" in IEEE Transactions on Mobile Computing, 2022

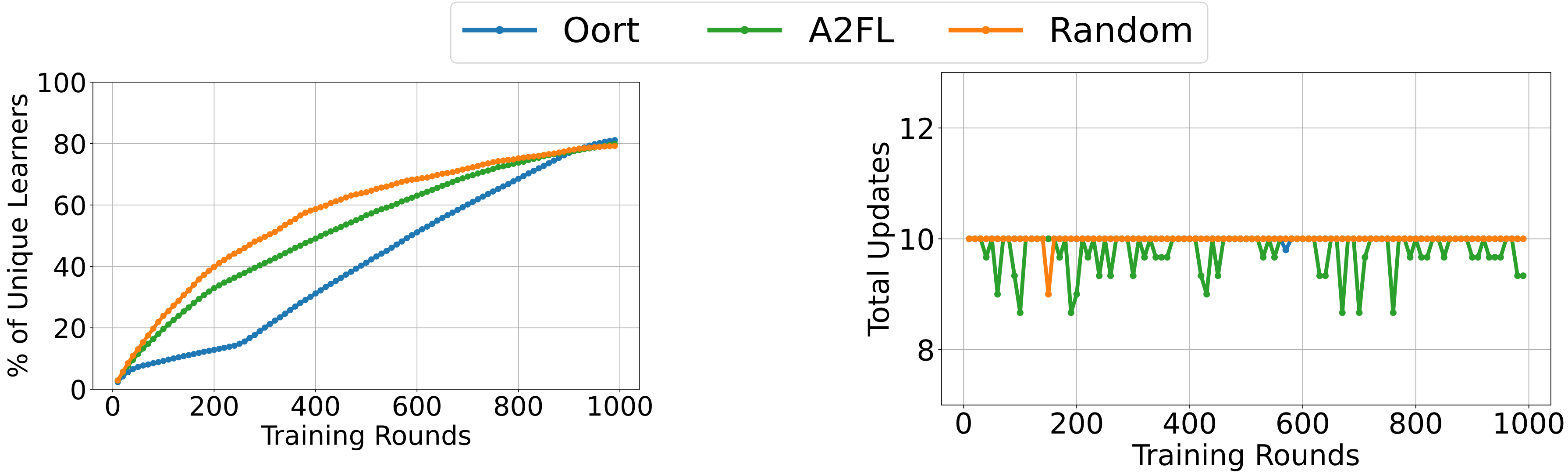
Evaluation of A2FL

- A2FL → best model quality with least amount of resources and time
  - It improves over all the other methods in both IID and non-IID cases



Evaluation of A2FL

- **Availability prioritization** leads to better diversity
  - This is evident by the high rate of unique clients for A2FL (close to random)
  - This is with even lower number of updates (i.e., higher stragglers).



Takeaways

- Heterogeneity is a major challenge for FL:
  - Model quality degradations are not acceptable, esp. in non-IID settings
  - Behavior heterogeneity impacts the quality even more.

- To tackle heterogeneity → adapt to availability dynamics of the clients
  - **A2FL** leverages support of on-device availability prediction module and prioritizes the clients with least availability → gains in model quality.

- Furture Work & Technical Challenges
  - How to deal with mis-information from malicious/non-faithful learners?
  - How to fine-tune knobs to control the trade-off betwen efficiency & diversity?

Thanks

Q & A

For further questions, please reach out to ahmed.sayed@qmul.ac.uk

If interested in solving real-world problems!
Join us at https://sayed-sys-lab.github.io