Towards Asynchronous Client Collaboration in Personalized Federated Learning

Boyi Liu[1,2], Zimu Zhou[1],

Pengfei Gao[2], Shuo Kang[2], Yongxin Tong[2]

[1]City University of Hong Kong

[2]Beihang University

香港城市大學
City University of Hong Kong

北京航空航天大學
BEIHANG UNIVERSITY

Outline

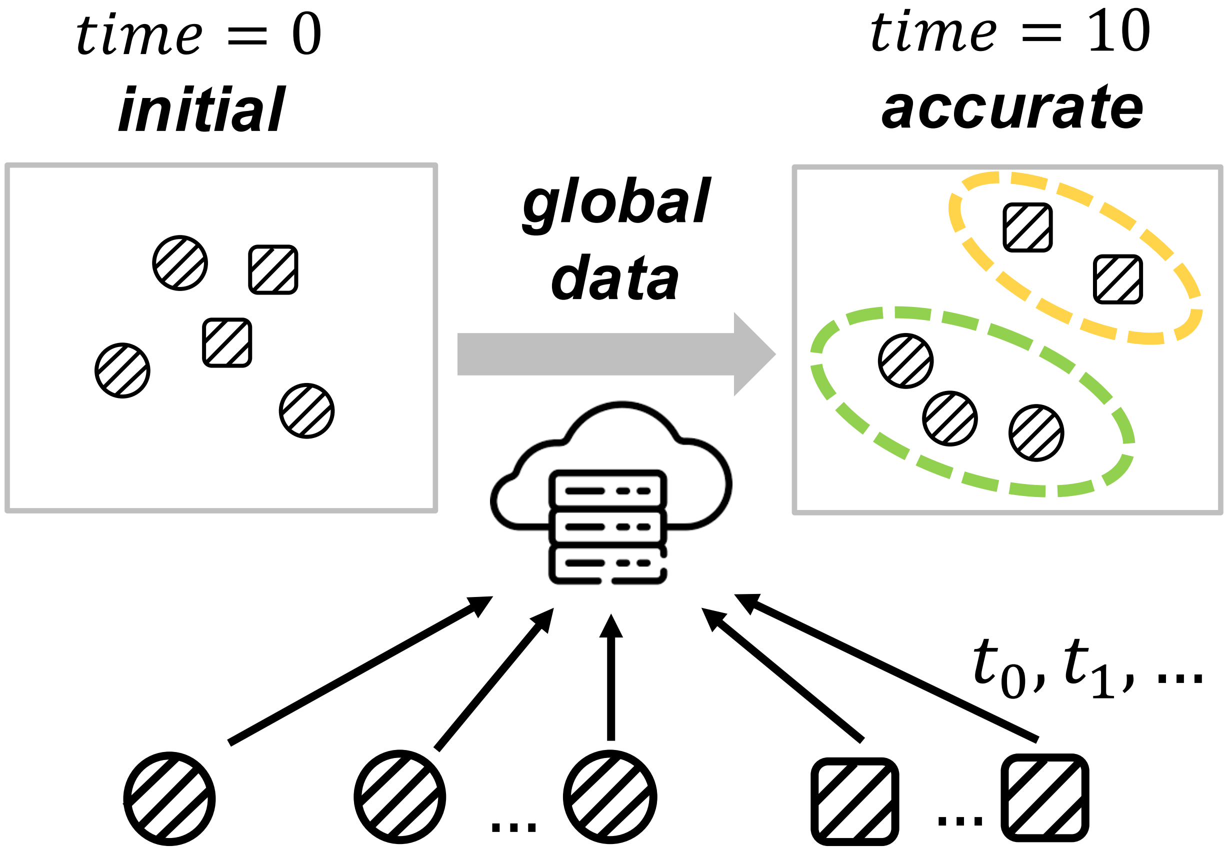
- Background & Motivation
- Problem Statement
- Our Solutions
- Experiments
- Conclusion

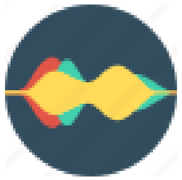
Outline

- Background & Motivation
- Problem Statement
- Our Solutions
- Experiments
- Conclusion

Background & Motivation

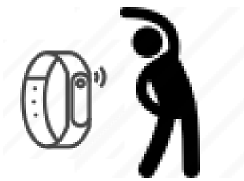
- **Collaboration-based Personalized Federated Learning (Co-PFL)**





Personal Voice Assistant

Smart Keyboards

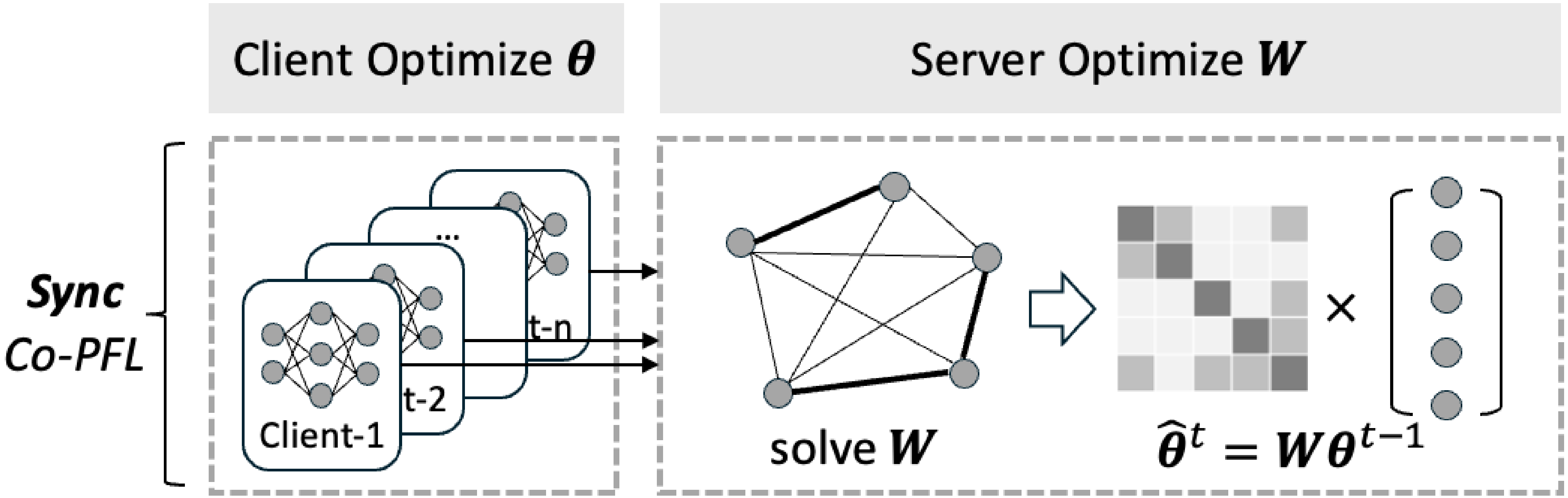


Human Activity Recognition

Data is often *heterogeneous* yet exhibits *natural similarity*

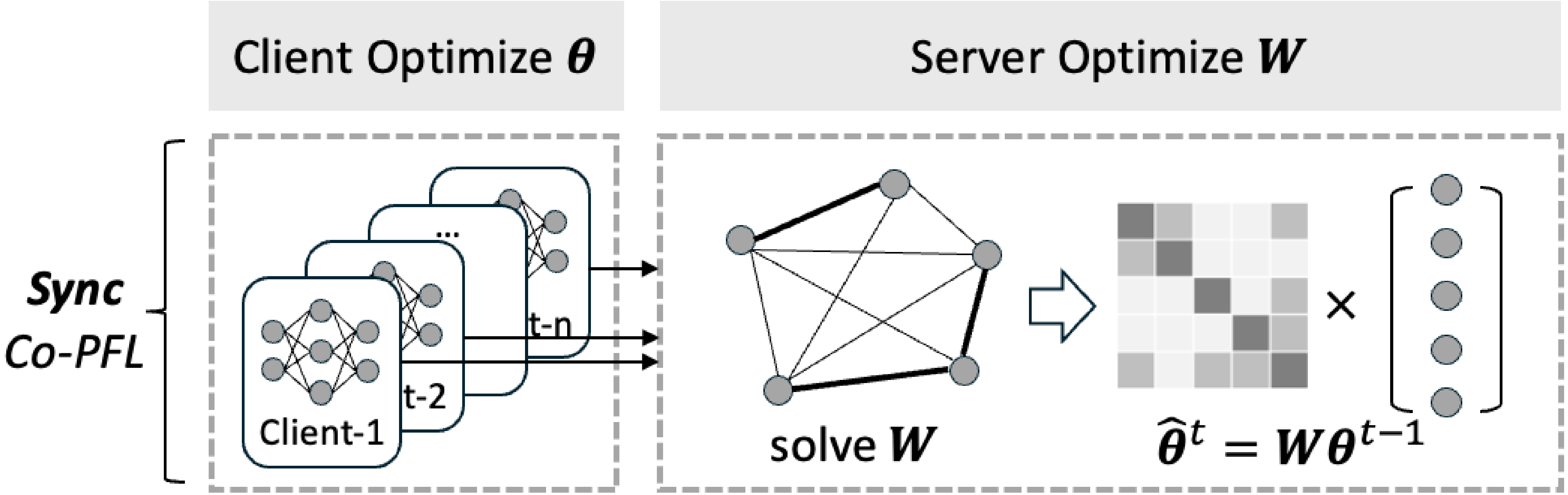
Background & Motivation

- **Collaboration-based Personalized Federated Learning (Co-PFL)**
  - Step 1: Collaboration Estimation
  - Step 2: Personalized Aggregation



Background & Motivation

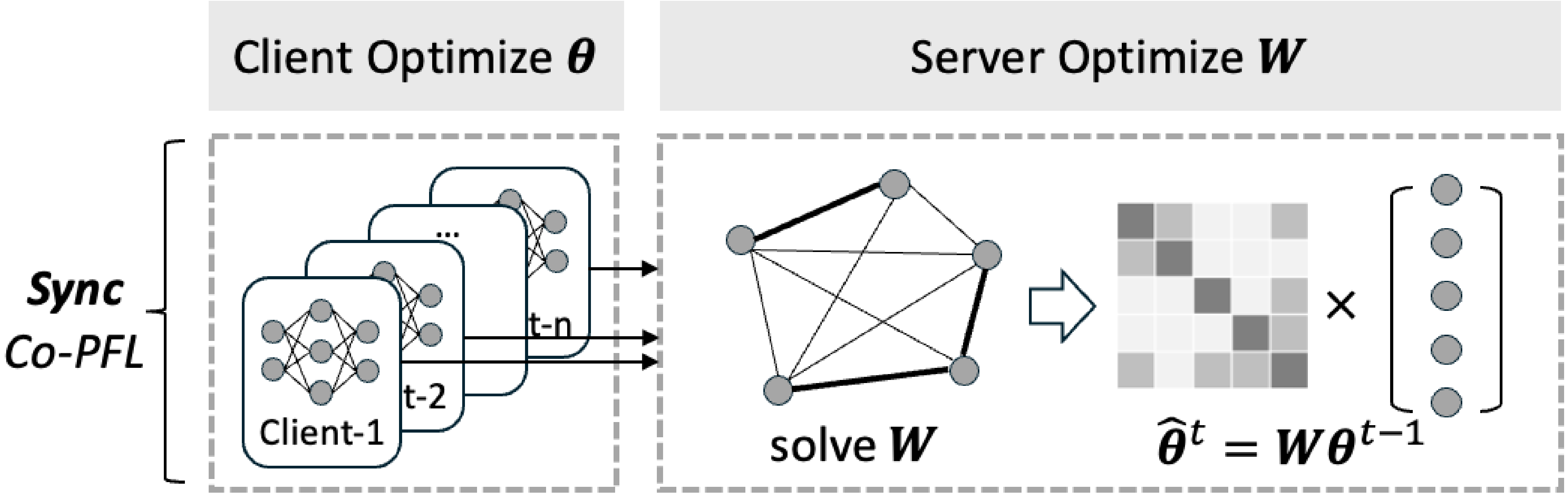
- **Collaboration-based Personalized Federated Learning (Co-PFL)**



Data heterogeneity, *however,* ***Device heterogeneity***

Background & Motivation

- **Collaboration-based Personalized Federated Learning (Co-PFL)**



✅ ***Data heterogeneity,*** *however,* ❌ ***Device heterogeneity***

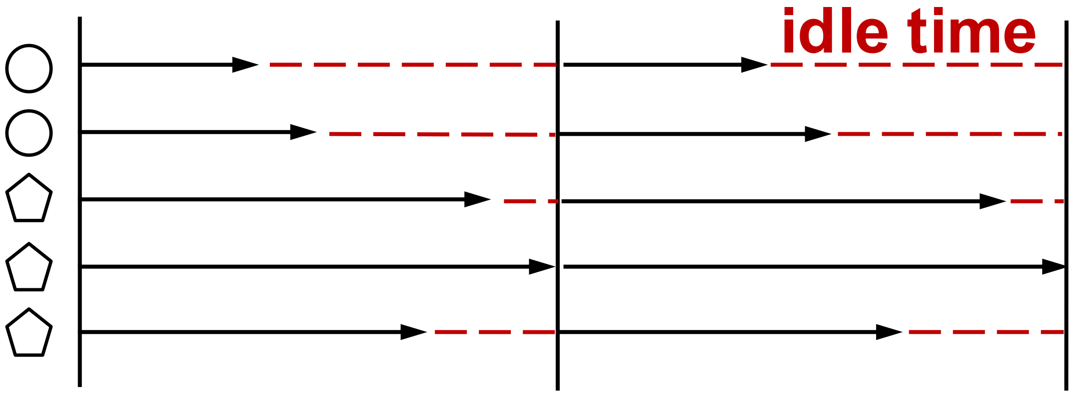
Can Co-PFL manage device heterogeneity?

Outline

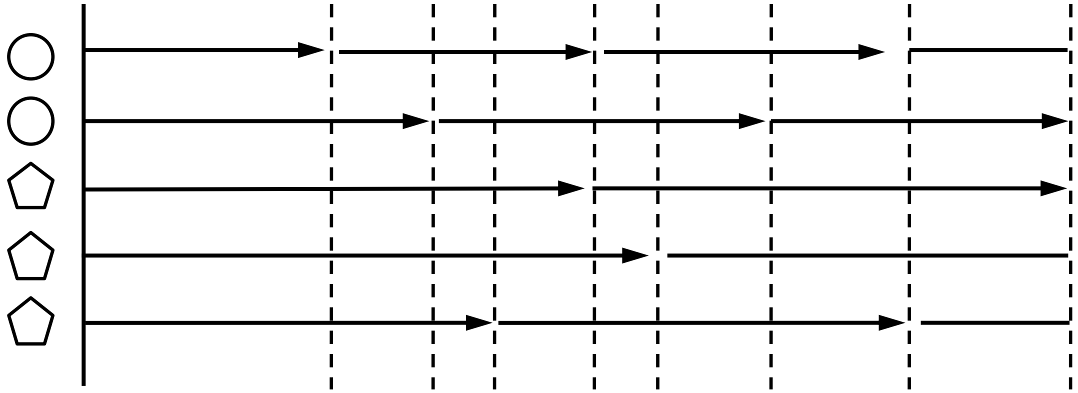
- Background & Motivation
- Problem Statement
- Our Solutions
- Experiments
- Conclusion

Problem Statement

- **Exploring Asynchronous Update Scheme**



synchronous update

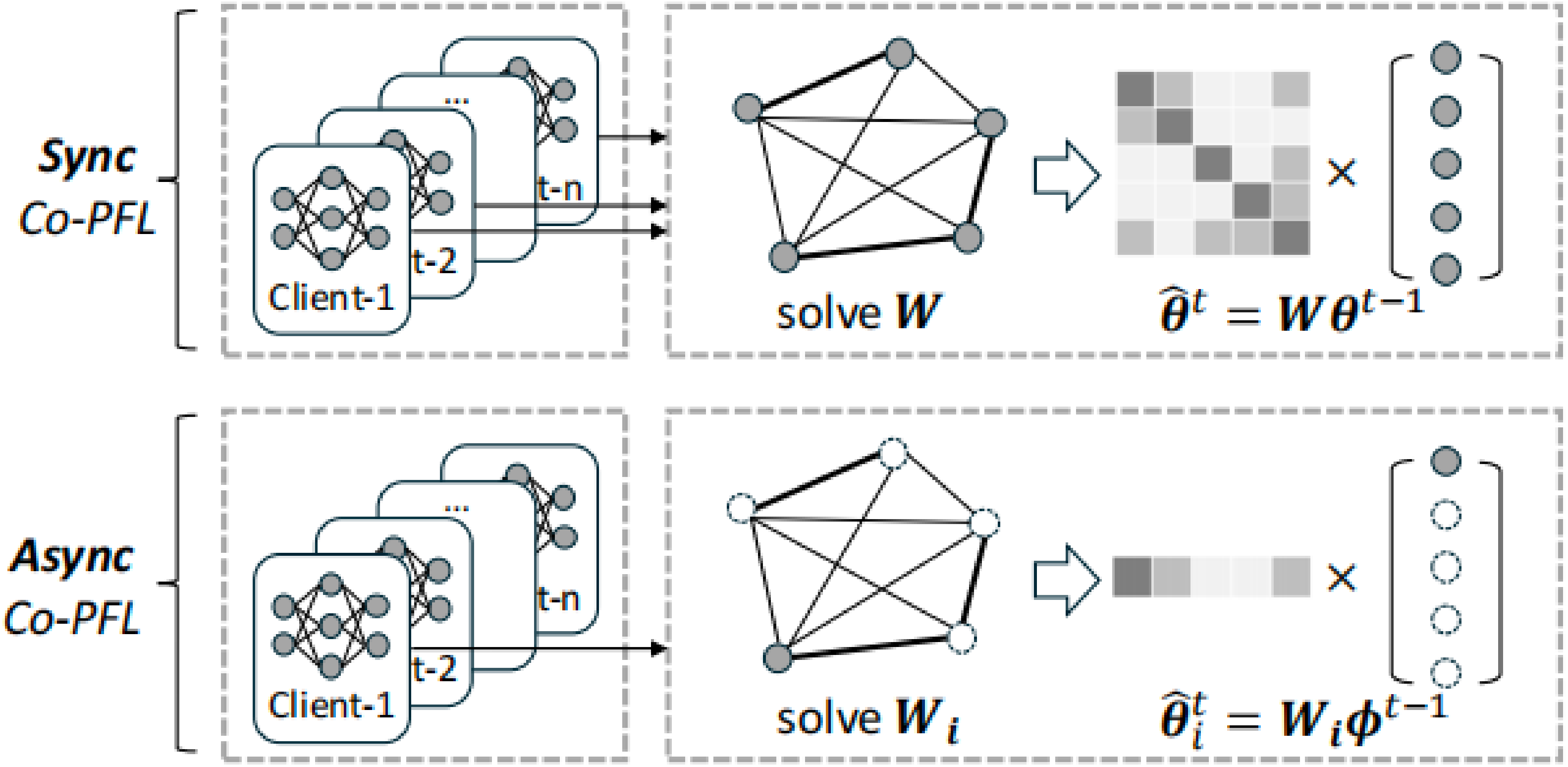


asynchronous update

With asynchronous update, we don't have to wait stragglers!

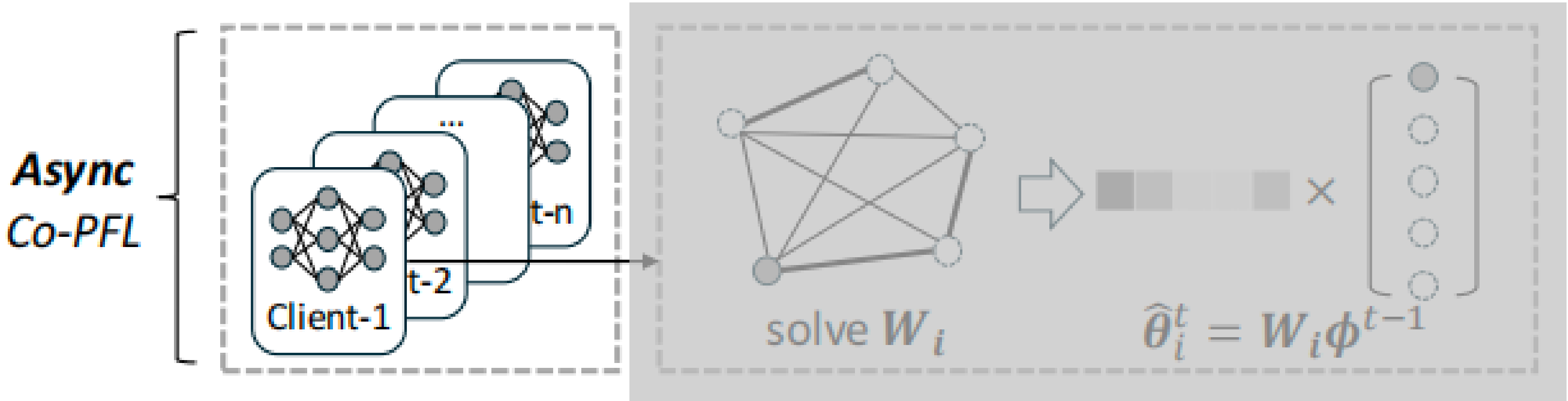
Problem Statement

- **Sync Co-PFL v.s. Async Co-PFL**



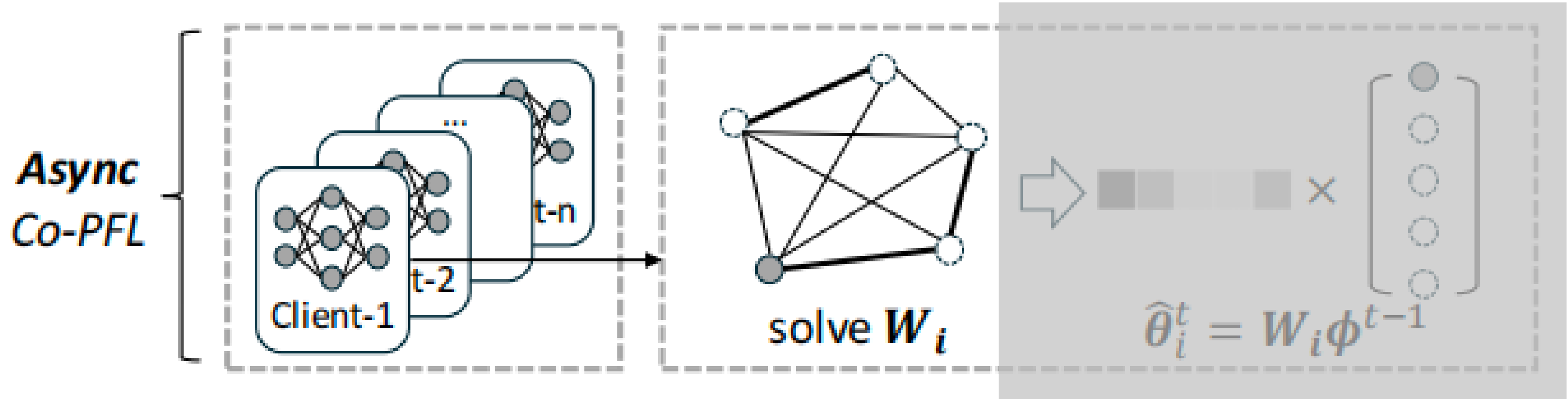
Problem Statement

- **Workflow of asynchronous Co-PFL naïve adaptation**
  - **Preliminary: the server maintains n buffered models $\{\phi_1, \dots, \phi_n\}$ for each client**
  - **Step 1: Client i finishes training, and uploads θ_i**



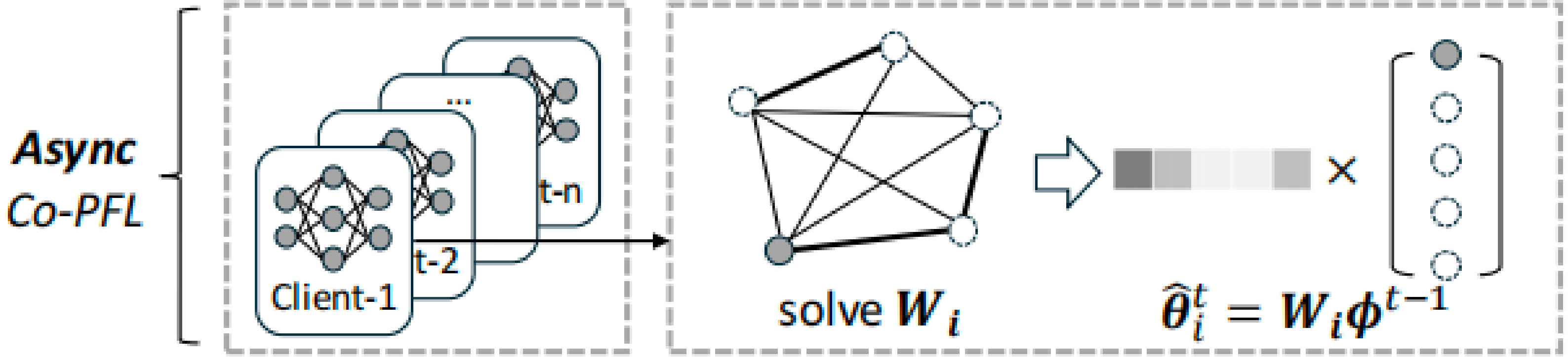
Problem Statement

- **Workflow of asynchronous Co-PFL naïve adaptation**
  - **Preliminary: the server maintains n buffered models $\{\phi_1, \dots, \phi_n\}$ for each client**
  - **Step 1: Client i finishes training, and uploads θ_i**
  - **Step 2: The server estimates collaboration relationship W_i with $\{\phi_1, \dots, \phi_n\}$ and θ_i**



Problem Statement

- **Workflow of asynchronous Co-PFL naïve adaptation**
  - **Preliminary: the server maintains n buffered models $\{\phi_1, \dots, \phi_n\}$ for each client**
  - **Step 1: Client i finishes training, and uploads θ_i**
  - **Step 2: The server estimates collaboration relationship W_i with $\{\phi_1, \dots, \phi_n\}$ and θ_i**
  - **Step 3: The server updates buffer ϕ_i with θ_i, and aggregates $\phi_i \leftarrow \sum_{j=1}^{n} W_{ij}\, \phi_j$**



Problem Statement

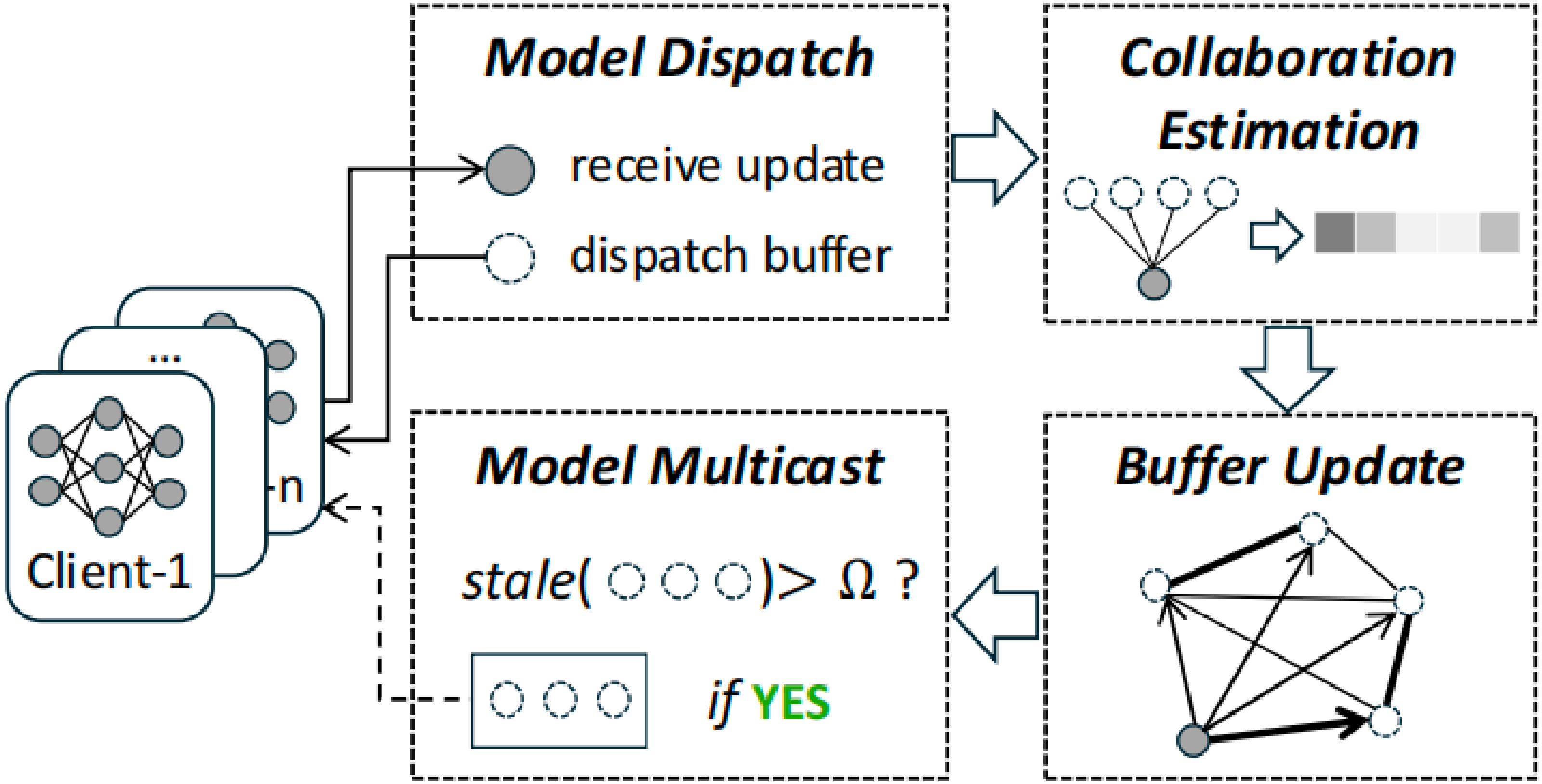
- **Challenges**
  - **Change 1: Inaccurate Collaboration Estimation**
    - **Similarity estimation $s(\theta_i, \phi_j)$ applies models from different start point**
  - **Change 2: Biased Model Aggregation**
    - **The buffers $\{\phi_1, \ldots, \phi_n\}$ for aggregation are stale**

Outline

- Background & Motivation
- Problem Statement
- Our Solutions
- Experiments
- Conclusion

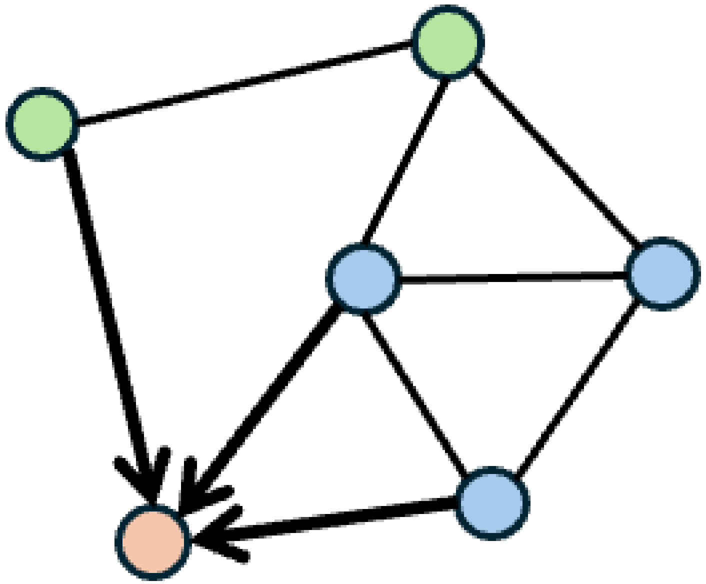
Our Solutions

- **PACE**
  - **Personalization under Asynchronous Collaboration Estimation**
  - **Overview**

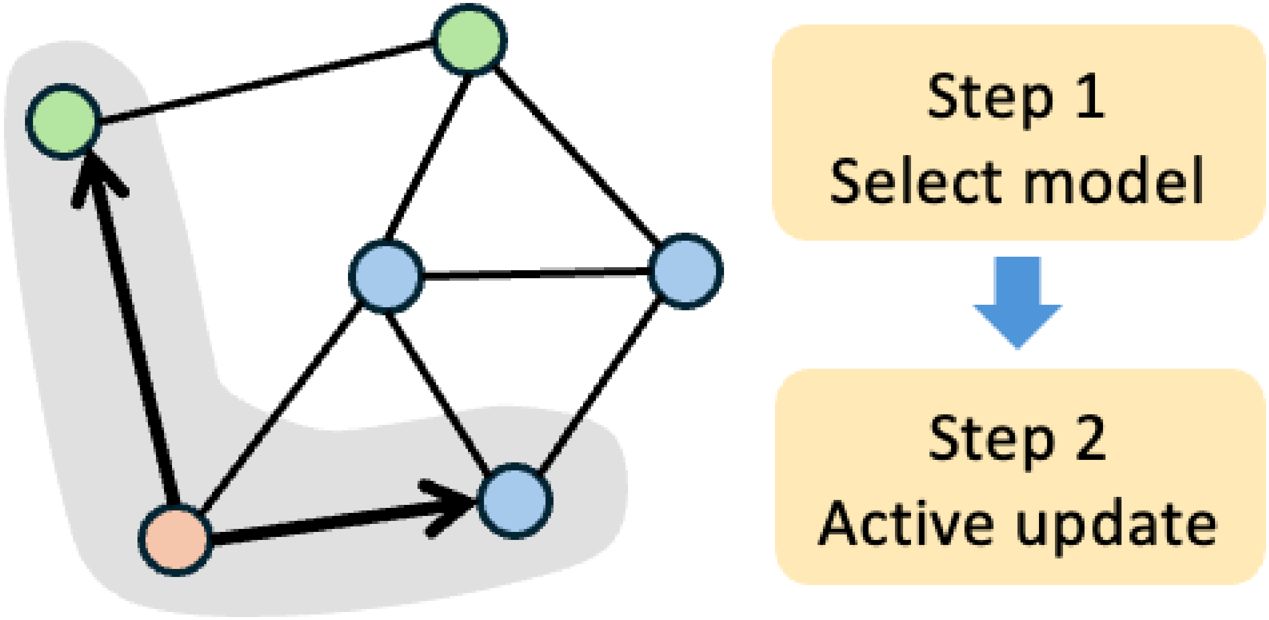


Our Solutions

- **Collaboration-Aware Buffer Update**
  - **Principle: Actively update the buffer models to keep freshness**



Naive collaboration



Our solution

Our Solutions

- **Collaboration-Aware Buffer Update**
  - **How to determine the weight of active update?**
  - **Collaboration-aware decay**

$$\alpha_{ij} = \boxed{\frac{W_{ji}}{\sum_{\pi \in \boldsymbol{\pi}_j \cup \{j\}} W_{j\pi}}} \cdot \boxed{(1+\tau_i)^{-a}.}$$

Collaboration *Staleness*

Our Solutions

- **Collaboration-Aware Buffer Update**
  - **Proof sketch**
    - Suppose buffer ϕ_j is updated by a sequence of clients $\pi_j = \{\pi_j^1, \dots, \pi_j^L\}$, and corresponding models aggregated are $\psi_j = \{\psi_j^1, \dots, \psi_j^L\}$

(1) Buffer ϕ_j is updated as:

$$\phi_j^L = \alpha_j^L \psi_j^L + (1-\alpha_j^L)\alpha_j^{L-1}\psi_j^{L-1} + \cdots + \left(\prod_{l=2}^{L}(1-\alpha_j^l)\alpha_j^1\psi_j^1\right) + \left(\prod_{l=1}^{L}(1-\alpha_j^l)\phi_j^0\right).$$

(2) For correct collaboration relationship:

$$\begin{cases} \frac{W_{jj}}{W_{jj}+\sum_{\pi\in\boldsymbol{\pi}_j} W_{j\pi}} = \prod_{l=1}^{L}(1-\alpha_j^l), \\ \frac{W_{j\pi_j^1}}{W_{jj}+\sum_{\pi\in\boldsymbol{\pi}_j} W_{j\pi}} = \prod_{l=2}^{L}(1-\alpha_j^l)\alpha_j^1, \\ \vdots \\ \frac{W_{j\pi_j^L}}{W_{jj}+\sum_{\pi\in\boldsymbol{\pi}_j} W_{j\pi}} = \alpha_j^L. \end{cases}$$

(3) Solve the equality:

$$\boxed{\alpha_j^l = \frac{W_{j\pi_j^l}}{W_{jj} + \sum_{q=1}^{l} W_{j\pi_j^q}}, \quad \forall\ 1 \le l \le L.}$$

IEEE INFOCOM
May 18 – 21, 2026
Tokyo, Japan

Our Solutions

- **Staleness-Triggered Model Multicast**
  - **Understanding the impact of asynchrony on collaboration**

Claim 1. *(Collaboration estimation error). The divergence between the estimated collaboration vector $\tilde{W}_i$ and the ground-truth W_i for client i is:*

staleness impact collaboration

$$\|W_i - \tilde{W}_i\| \leq \sum_{j=1}^{n} \frac{\boxed{\tau_i}\delta\eta E G}{2\sum_{\pi \in \boldsymbol{\pi}_j \cup \{j\}} W_{j\pi} G_j}, \tag{12}$$

where δ is defined in Definition 1 of Appendix A, η is the learning rate, E is the number of local epochs, G is the upper bound of gradient, and $G_j = \|\phi_j\|$.

Our Solutions

- **Staleness-Triggered Model Multicast**
  - **Our solution: Model Multicast to limit staleness**
    - **Select a client set $\mathcal{G}^t$ to actively multicast the buffered model**

$$\sum_{i \in \mathcal{G}^t} \tau_i^2 > \Omega, \quad |\mathcal{G}^t| \times \theta.size() \leq \mathcal{P}$$

Outline

- Background & Motivation
- Problem Statement
- Our Solutions
- Experiments
- Conclusion

Experiments

- **Setup**
  - **Task**
    - **Image classification: EMNIST/CIFAR-10/CIFAR-100**
    - **Human activity recognition: HARBox**
    - **Text classification: AGNews**
  - **Baseline**
    - **Sync FL: FedAvg/FedProx**
    - **Async FL: FedAsync/FedBuff/PORT/FedAC/ASAFL**
    - **Sync Co-PFL: FedAMP/pFedGraph/FedSaC**
    - **Async Co-PFL: Async version of Sync Co-PFL baselines**

Experiments

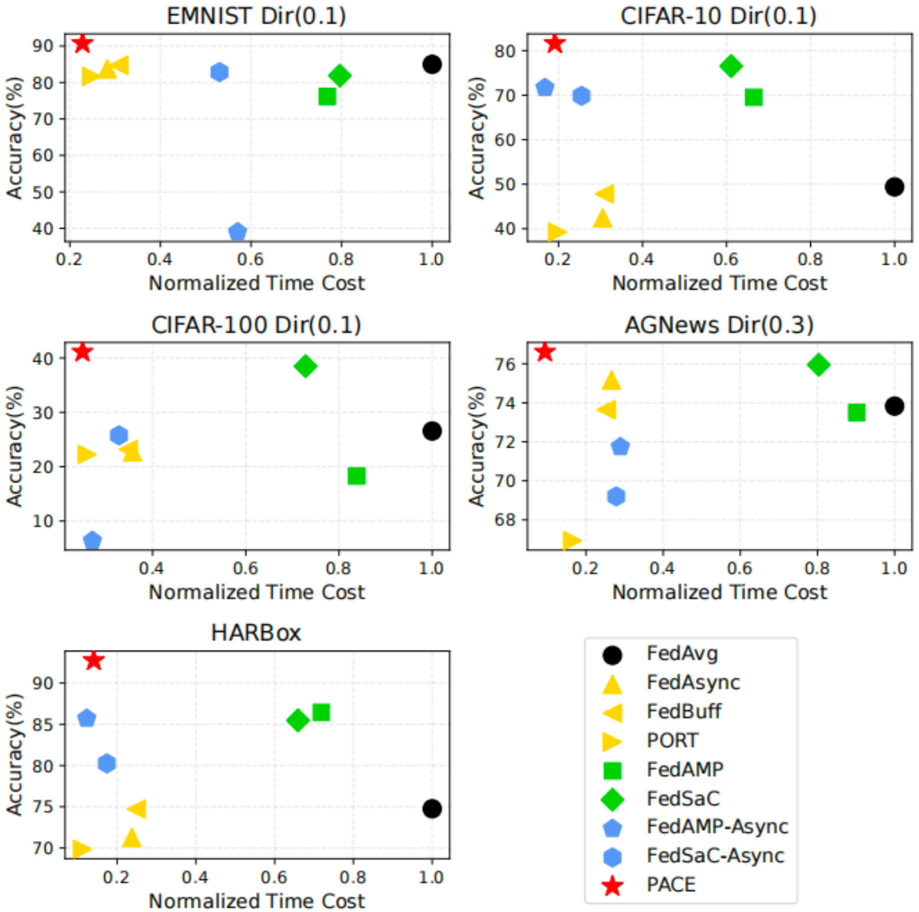
- **Accuracy**

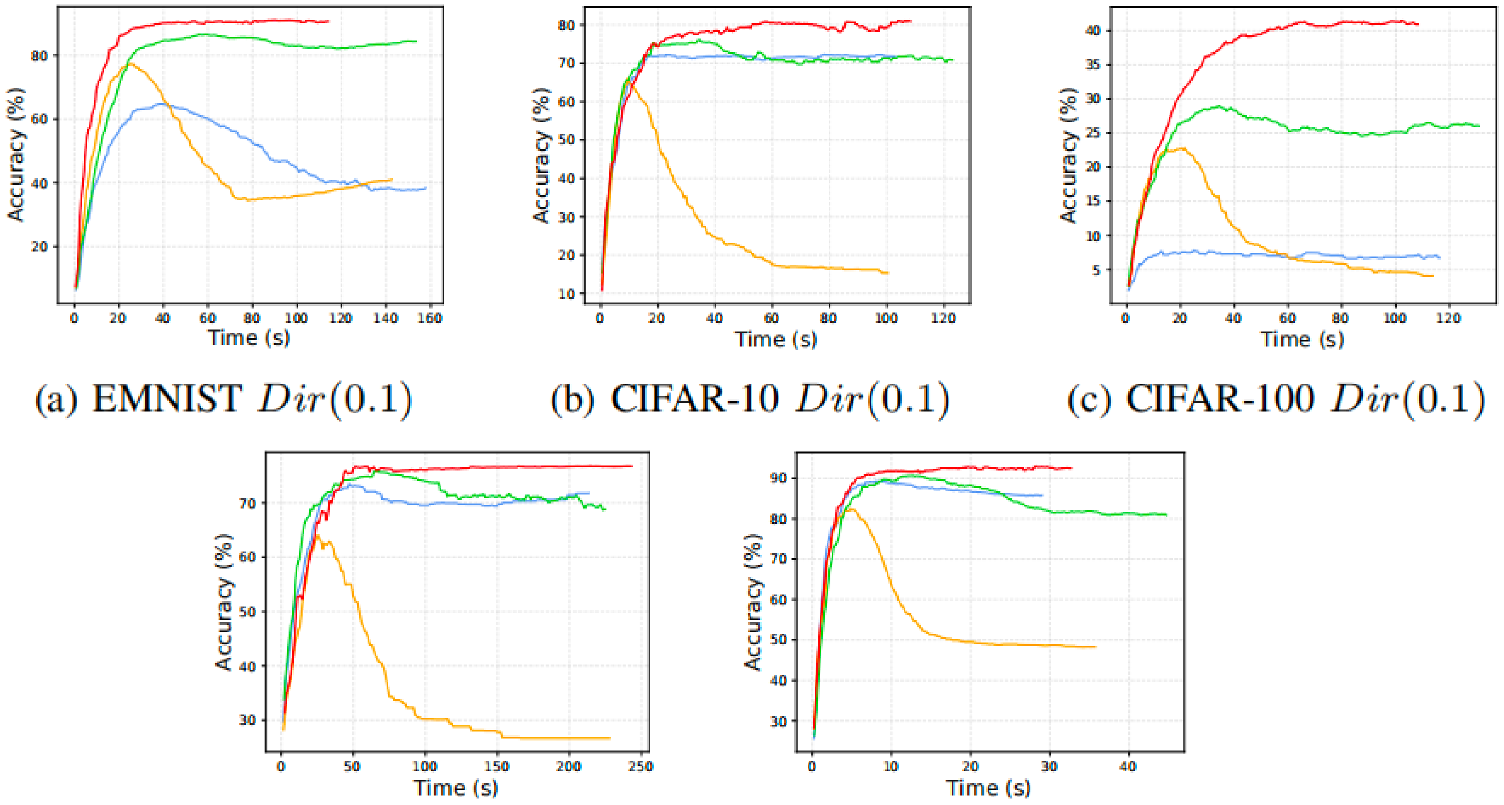
Method	EMNIST		CIFAR-10		CIFAR-100		AGNews	HARBox
	$\#C = 10$	$Dir(0.1)$	$\#C = 2$	$Dir(0.1)$	$\#C = 20$	$Dir(0.1)$	$Dir(0.3)$	*Real-world*
FedAvg [29]	$82.38_{\pm 0.35}$	$84.98_{\pm 0.23}$	$45.32_{\pm 0.07}$	$49.40_{\pm 0.07}$	$23.77_{\pm 0.14}$	$26.56_{\pm 0.50}$	$72.27_{\pm 1.56}$	$74.76_{\pm 0.14}$
FedProx [30]	$82.65_{\pm 1.17}$	$84.43_{\pm 0.16}$	$46.08_{\pm 0.45}$	$51.07_{\pm 1.56}$	$23.34_{\pm 0.63}$	$26.55_{\pm 0.27}$	$67.97_{\pm 0.55}$	$74.90_{\pm 0.09}$
FedAsync [17]	$80.62_{\pm 0.08}$	$83.74_{\pm 0.20}$	$40.76_{\pm 0.91}$	$42.45_{\pm 0.44}$	$20.12_{\pm 0.28}$	$22.76_{\pm 0.19}$	$74.20_{\pm 0.96}$	$71.28_{\pm 1.42}$
FedBuff [24]	$81.24_{\pm 1.19}$	$84.47_{\pm 0.17}$	$43.89_{\pm 0.17}$	$47.83_{\pm 1.92}$	$17.04_{\pm 0.34}$	$23.16_{\pm 0.45}$	$70.62_{\pm 3.03}$	$74.72_{\pm 0.70}$
PORT [18]	$68.43_{\pm 0.01}$	$81.67_{\pm 0.19}$	$32.48_{\pm 0.75}$	$39.26_{\pm 0.09}$	$17.43_{\pm 0.25}$	$22.26_{\pm 0.25}$	$64.90_{\pm 2.04}$	$69.85_{\pm 1.16}$
FedAC [19]	$83.36_{\pm 0.42}$	$84.65_{\pm 0.17}$	$45.48_{\pm 0.19}$	$52.81_{\pm 0.47}$	$20.61_{\pm 0.07}$	$24.09_{\pm 0.38}$	$75.77_{\pm 0.01}$	$76.14_{\pm 0.15}$
ASAFL [25]	$80.71_{\pm 0.28}$	$81.23_{\pm 0.03}$	$42.81_{\pm 0.70}$	$41.12_{\pm 1.54}$	$10.14_{\pm 0.13}$	$13.79_{\pm 0.30}$	$61.56_{\pm 0.82}$	$71.93_{\pm 0.85}$
FedAMP [1]	$80.70_{\pm 0.31}$	$76.17_{\pm 0.68}$	$56.71_{\pm 0.31}$	$69.53_{\pm 0.09}$	$23.58_{\pm 0.37}$	$18.29_{\pm 0.18}$	$73.50_{\pm 0.24}$	$86.44_{\pm 0.06}$
pFedGraph [2]	$85.32_{\pm 0.01}$	$81.71_{\pm 0.00}$	$77.43_{\pm 0.09}$	$76.80_{\pm 0.03}$	$22.82_{\pm 0.06}$	$38.05_{\pm 0.01}$	$75.85_{\pm 0.12}$	$85.52_{\pm 0.10}$
FedSaC [3]	$85.54_{\pm 0.31}$	$81.71_{\pm 0.09}$	$77.63_{\pm 0.20}$	$76.54_{\pm 0.17}$	$22.69_{\pm 0.04}$	$38.48_{\pm 0.09}$	$75.95_{\pm 0.07}$	$85.47_{\pm 0.19}$
FedAMP-Async	$16.77_{\pm 0.41}$	$38.97_{\pm 0.92}$	$50.33_{\pm 0.11}$	$71.67_{\pm 0.15}$	$1.07_{\pm 0.18}$	$6.39_{\pm 0.52}$	$71.60_{\pm 0.14}$	$85.70_{\pm 0.10}$
pFedGraph-Async	$72.64_{\pm 1.55}$	$42.98_{\pm 2.20}$	$59.90_{\pm 0.33}$	$16.70_{\pm 1.38}$	$4.80_{\pm 0.25}$	$3.91_{\pm 0.18}$	$26.83_{\pm 0.14}$	$45.84_{\pm 2.48}$
FedSaC-Async	$86.59_{\pm 0.26}$	$82.85_{\pm 1.53}$	$68.89_{\pm 0.27}$	$69.85_{\pm 0.93}$	$12.19_{\pm 0.38}$	$25.76_{\pm 0.33}$	$67.49_{\pm 1.72}$	$80.26_{\pm 0.82}$
PACE (Ours)	$\mathbf{91.97}_{\pm 0.13}$	$\mathbf{90.75}_{\pm 0.21}$	$\mathbf{79.98}_{\pm 0.05}$	$\mathbf{81.65}_{\pm 0.62}$	$\mathbf{25.03}_{\pm 0.22}$	$\mathbf{41.11}_{\pm 0.11}$	$\mathbf{76.69}_{\pm 0.04}$	$\mathbf{92.73}_{\pm 0.20}$

PACE outperform existing Sync & Async Co-PFL algorithms

Experiments

- **Convergence Time**





(a) EMNIST $Dir(0.1)$ (b) CIFAR-10 $Dir(0.1)$ (c) CIFAR-100 $Dir(0.1)$

(d) AGNews $Dir(0.3)$ (e) HARBox

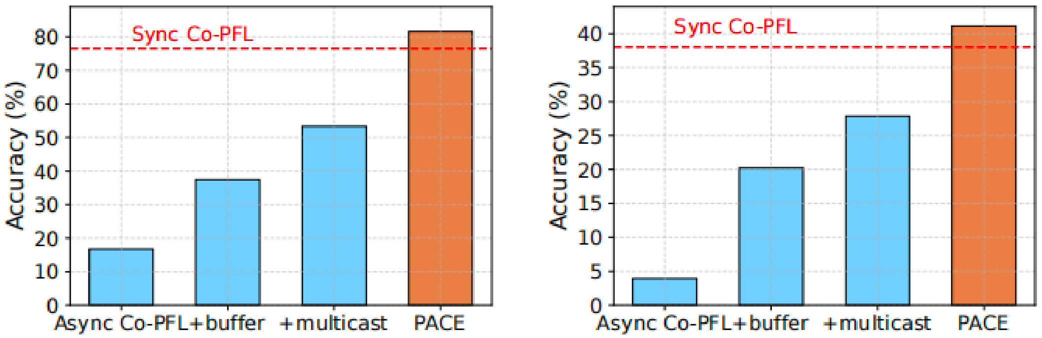
PACE converges faster than existing Sync & Async Co-PFL algorithms

Experiments

- **Effective of PACE's module**

Method	CIFAR-10 $Dir(0.1)$	CIFAR-100 $Dir(0.1)$	AGNews $Dir(0.3)$
FedAMP-Async	71.67	6.39	71.60
FedAMP-PACE	**79.87**(+8.20)	**37.21**(+30.82)	**77.39**(+5.79)
FedSaC-Async	69.85	25.76	67.49
FedSaC-PACE	**81.58**(+11.73)	**40.97**(+15.21)	**76.87**(+9.38)

Extending Co-PFL methods with PACE



(a) CIFAR-10 (b) CIFAR-100

Impact of each modules

PACE's each module works well

Outline

- **Background & Motivation**
- **Problem Statement**
- **Our Solutions**
- **Experiments**
- **Conclusion**

Conclusions

- We focus on ***Collaboration-based Personalized Federated Learning (Co-PFL)***, and explore its ***asynchronous*** variant

- We propose ***PACE***, a new framework that solves the impact with ***Collaboration-Aware Buffer Update*** and ***Staleness-Triggered Model Multicast***

- Extensive experiments on various datasets validate the performances on ***accuracy and efficiency***

THANK YOU