## FEDERATED MULTI-TASK LEARNING WITH DECENTRALIZED PERIODIC AVERAGING SGD

### Introduction



Federated Learning: training models collaboratively over a large number of distributed edge devices without centralizing their local data.

Statistical Challenges: 1) Non-IID. 2) Unbalanced Local Data

System Challenges: 1) Larger Worker Number; 2) Heterogeneous Networks; 3) Heterogeneous Computation.



(1)

Chaoyang He, Tian Xie, Zhengyu Yang, Zijian Hu, Shuai Xia <sup>†</sup>Department of Computer Science, University of Southern California

![](_page_0_Figure_8.jpeg)

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}\nabla F(\mathbf{u}_{t})^{2}\right] \leq \frac{2[F(\mathbf{x}_{1}) - F_{inf}]}{\eta T} + \frac{2\sigma^{2}}{\sqrt{KT}}$$
$$= \mathcal{O}(\frac{1}{\sqrt{KT}})$$

How to choose  $\tau$ ,  $\zeta$ : for a small number of well-connected workers, larger  $\tau$  is more preferable; for a large number of workers, using a sparse mixing matrix and small  $\tau$  gives better convergence

Effect of Extreme Large K: the iteration number T will be extreme large. To guarantee the convergence, try to reduce the worker number through system-wised optimization: 1) Streaming training; 2) upload on-device data to the edge data center.

(2)

![](_page_0_Picture_13.jpeg)

![](_page_0_Figure_14.jpeg)

Federated Learning Server/Client System Architecture

System Design

### Experiment

![](_page_0_Figure_17.jpeg)

![](_page_0_Figure_18.jpeg)

#### **Future Works**

![](_page_0_Figure_20.jpeg)

![](_page_0_Figure_21.jpeg)

# Worker2 onlysynchronizes w neighbors, without waiting for all other worker