HARMONY: Heterogeneity-Aware Hierarchical Management for Federated Learning System

Chunlin Tian[1] | Li Li[1] | Zhan Shi[2] | Jun Wang[3] | ChengZhong Xu[1]
[1]University of Macau
[2]The University of Texas at Austin
[3]Futurewei Technology

CrossFL 2022 Septembe 1st, 2022

INTRODUCTION

Harmony[1] a heterogeneity-aware hierarchical coordination framework for high-performance Federated Learning. It effectively directs the training process to make it proceed in harmony through intelligently mediating the conflict caused by the heterogeneity in the following four folds:

- The **static system heterogeneity** caused by different hardware configurations;
- The **dynamic system heterogeneity** caused by resource contention at runtime;
- The **data heterogeneity** in each **local** device ;
- The **data heterogeneity** in each **global** training round.

[1]The name of the system, Harmony, has two-fold implications: ① our FL system aims to manage all heterogeneous devices to work in harmony. ② our FL system aims to have the background training task execute harmoniously with the foreground applications of the devices

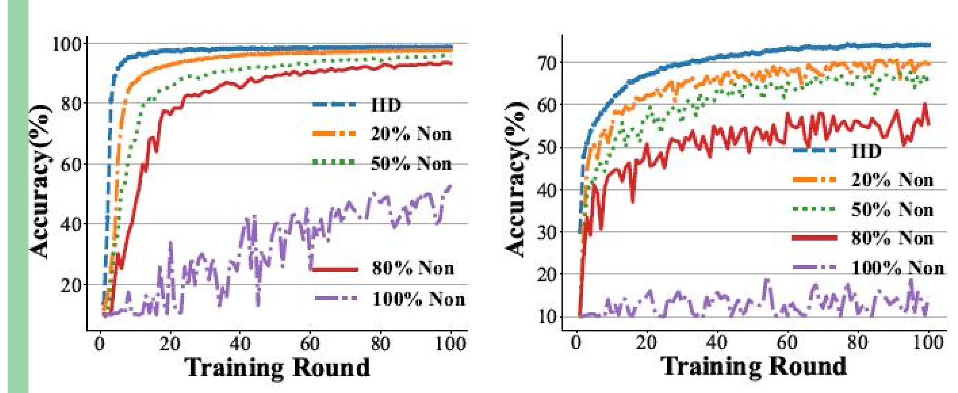
MOTIVATION

Q1: How is the system efficiency affected by the device type (**static system heterogeneity**) and the concurrently running foreground apps (**dynamic system heterogeneity**)?

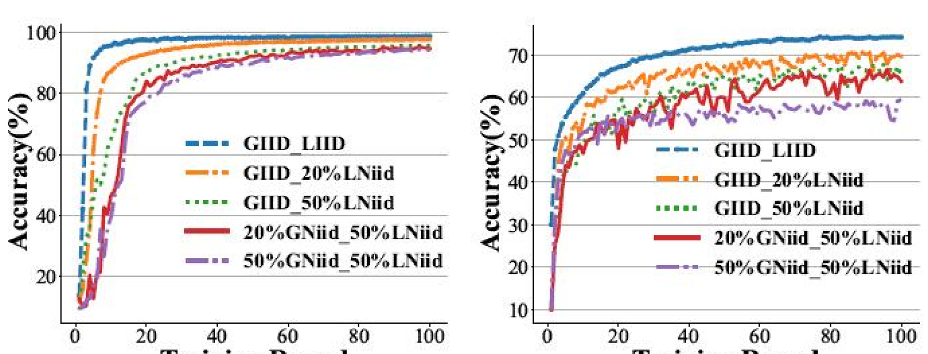
Foreground Task	Comp Time (s)	IPS (G)	CPU Load (%)	Perf Degradation (%)
Reading	2.3	5.0	0	0
Typing	2.9	3.8	4	0
2D-AngryBirds	4.2	2.9	23	1.3
3D-Basketball	5.0	1.8	30	3.8
Video Playing	4.7	3.2	21	0

Q2: How is the statistical efficiency affected by the local data distribution (**local data heterogeneity**) of each participating device?

Q3: How is the statistical efficiency affected by the global data distribution (**global data heterogeneity**) of the overall training data?



(a) MNIST-LeNet5 (b) CIFAR10-VGG11



(a) MNIST-LeNet5 (b) CIFAR10-VGG11

HIGH-LEVEL IDEAS

Harmony System Model $T(N) = \max\left(\frac{h_i^* m_i}{f_i^\delta * r_i}\right)$

Harmony Data Model $D_{KL} = \sum_{\forall \alpha_i = 1} [D_{KL_{local}}(P_i \| P_{exp}) * D_{KL_{global}}^{\chi/\mathbb{K}}(P_{glob} \| P_{exp})]$

Harmony's Utility Function

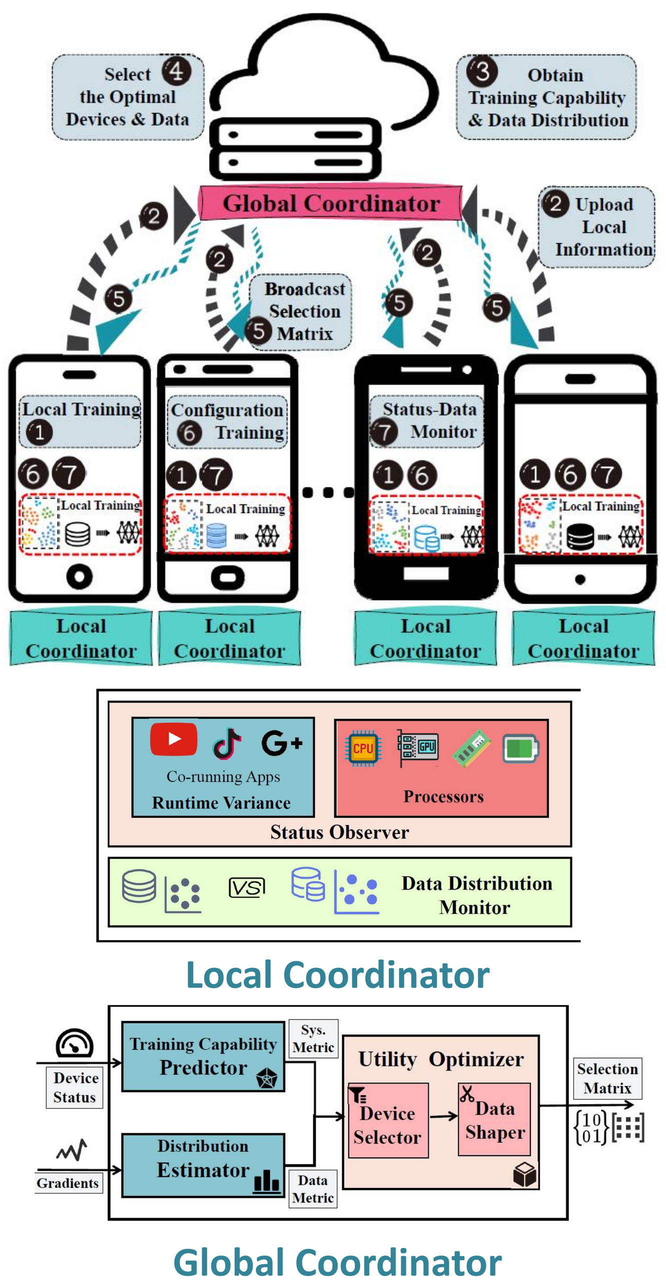
To unify both the system model and the data model for device and data selection, we propose a **heterogeneity-aware** utility function as follows:

$$Util(S) = \underset{(i \in \mathbb{K}, j \in \mathbb{C})}{\text{argmin}} \sum_{\forall \alpha_i = 1} \left[D_{KL_{local}} * D_{KL_{global}}^{\chi/\mathbb{K}} + \omega * T \right]$$

Subject to

$$\sum_{i=1}^{\mathbb{K}} \alpha_i * m_i > \mathbb{D}$$
$$\alpha_i \in \{0, 1\}$$
$$0 \le \beta_{i,j} \le 1, \quad \forall i \in \mathbb{K}, \forall j \in \mathbb{C}$$

SYSTEM DESIGN



System Overview

① All the mobile devices participate in the first training round and complete local training.
② Local coordinator sends the following information to the central server.
③ Global coordinator well estimates the data distribution and predicts the runtime training capability.
④ Global coordinator intelligently selects the participating devices by jointly considering the homogeneity of the local training data and runtime training capability. Moreover, global coordinator fine-tunes the distribution of the overall training data.
⑤ Global coordinator broadcasts the coordination result to the corresponding selected devices.
⑥ Local coordinator then conducts the local training process based on the coordination result.
⑦ Local coordinator monitors real-time status and data.

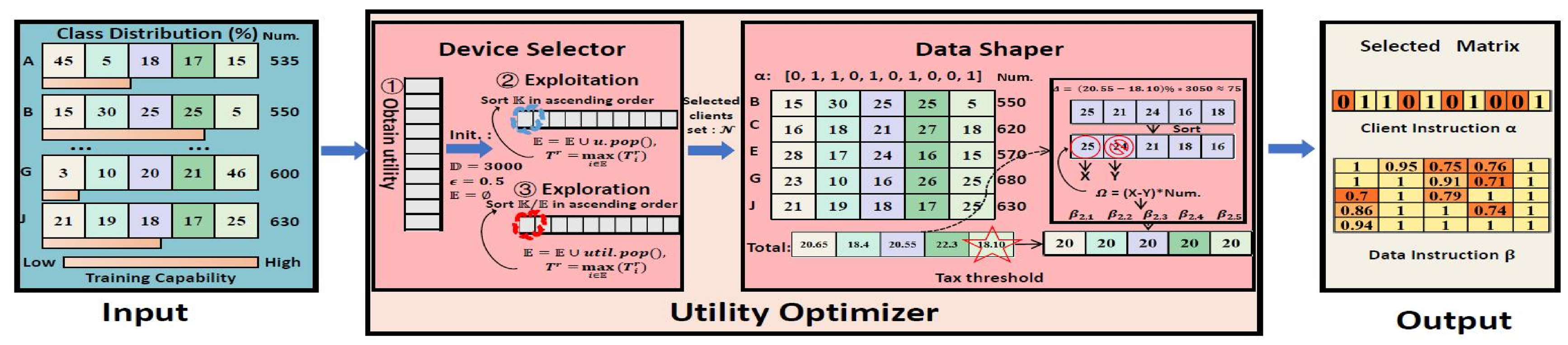
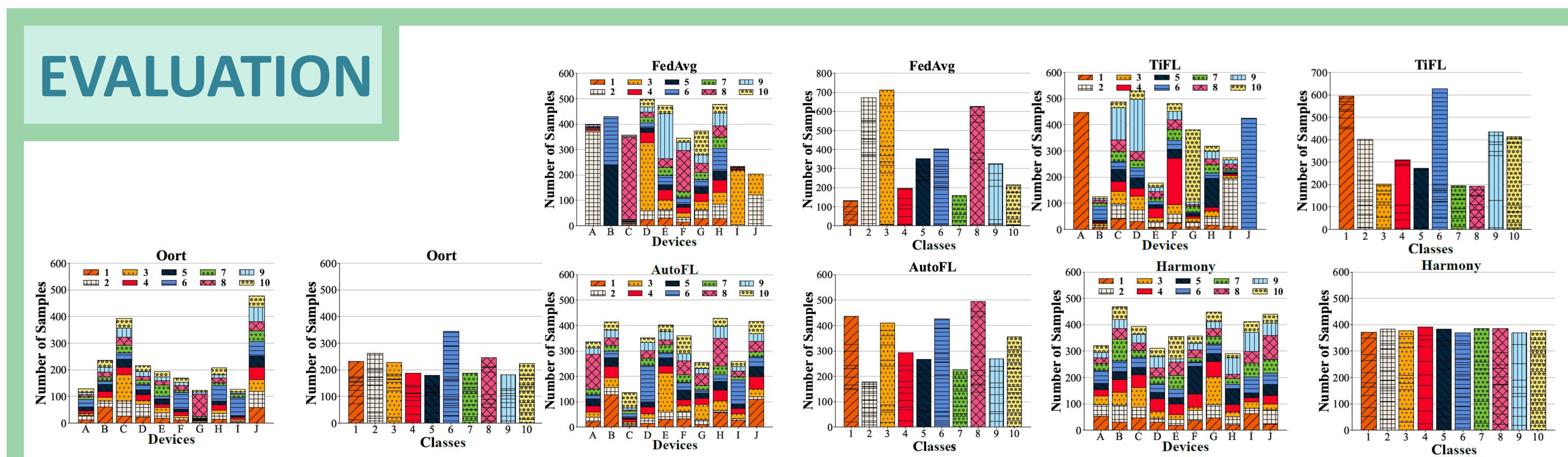
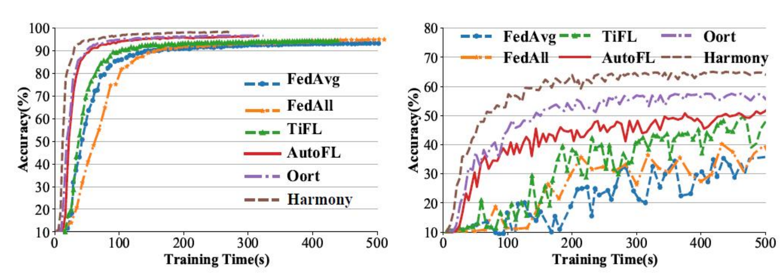


Illustration of the procedures of Utility Optimizer. Based on the estimated data distribution and the predicted device training capability, the Utility Optimizer first selects the participating device subset employing a Device Selector, then utilizes a Data Shaper to reshape the selected global data, and finally outputs the device-data selection matrix.

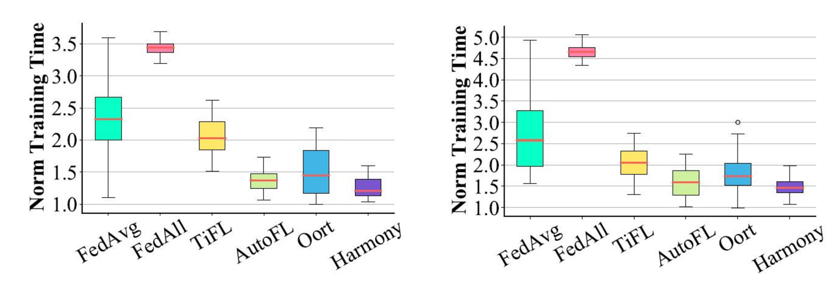
EVALUATION



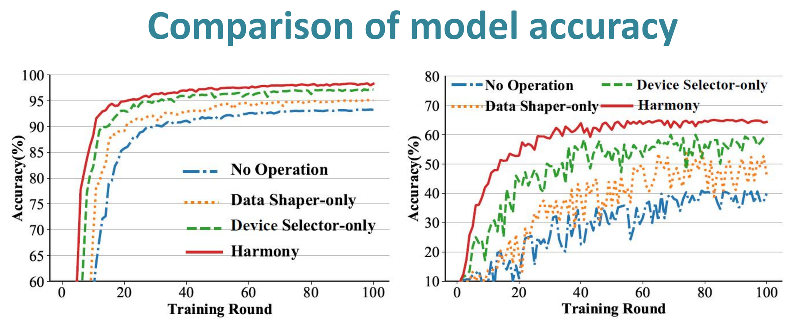
Data distribution within the selected devices of different schemes (left column), and the overall data distribution (right column)



Comparison of model accuracy

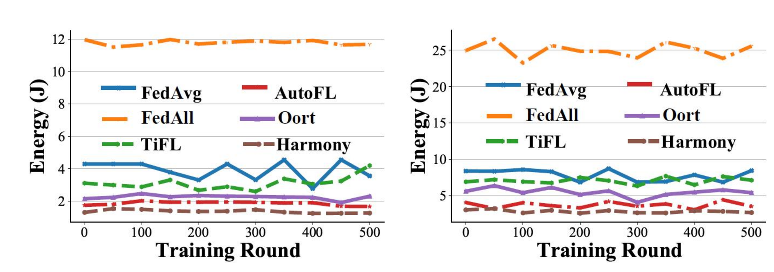


Comparison of per round training time



(a) MNIST-LeNet5 (b) CIFAR10-VGG11

Model accuracy of ablation study



(a) MNIST-LeNet5 (b) CIFAR10-VGG11

Comparison of energy consumption

CONCLUSION

Harmony intelligently balances the model performance and training progress in a highly dynamic and heterogeneous training environment from two perspectives.
The experiment results show that Harmony improves the model performance up to **27.62%**, effectively accelerates the training speed by up to **3.29×**, and achieves energy-saving up to **88.41%**.