THE WEB CONFERENCE

Characterizing Impacts of Heterogeneity in Federated Learning upon Large-Scale Smartphone Data

Chengxu Yang, Qipeng Wang, Mengwei Xu, Zhenpeng Chen,
Kaigui Bian, Yunxin Liu and Xuanzhe Liu

北京大学
PEKING UNIVERSITY

Microsoft

Increasing Concerns on User Privacy

Cybersecurity Law of the People's Republic of China

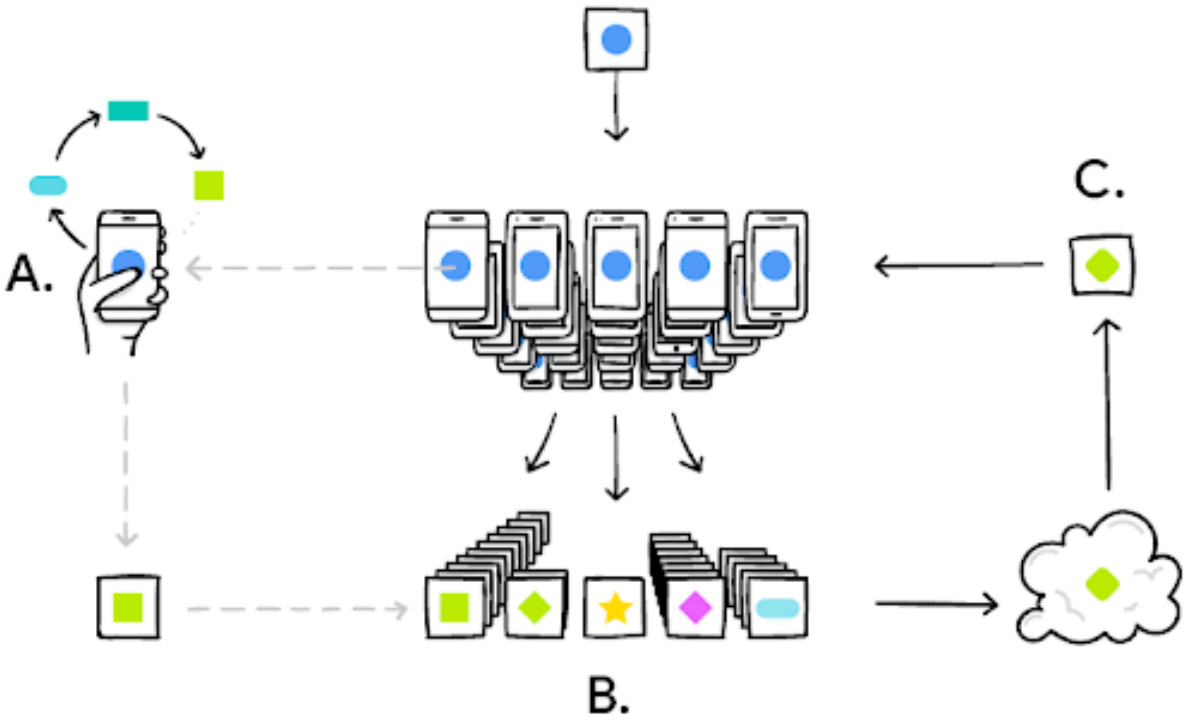
General Data Protection Regulation (GDPR)

California Consumer Privacy Act (CCPA)

Data

ML Model

Emerging Federated Learning (FL)



A. **Personalize the local model**

B. **Upload updates to the cloud**

C. **Aggregate and form a global model**

Heterogeneity – One of the Core Challenges

- Hardware heterogeneity

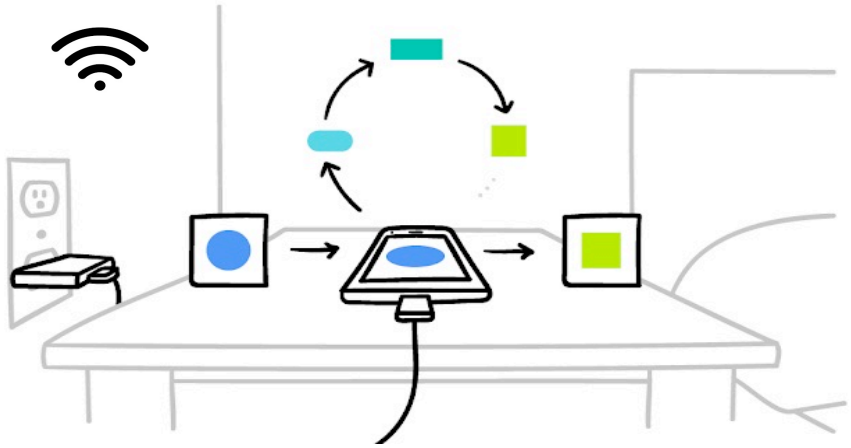
	CPU	RAM	Battery	…
Device A	Kirin 990	12GB	4000mAh	…
Device B	Snapdragon 630	4GB	3000mAh	…
…	…	…	…	…

Heterogeneity – One of the Core Challenges

- Hardware heterogeneity

	CPU	RAM	Battery	…
Device A	Kirin 990	12GB	4000mAh	…
Device B	Snapdragon 630	4GB	3000mAh	…
…	…	…	…	…

- State heterogeneity



A device participates only when it won't negatively impact user's experience

Required state criteria:

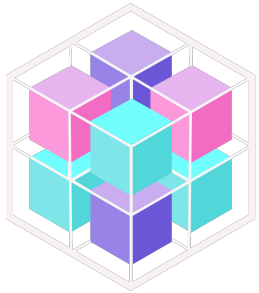
- CPU idle
- Charging
- Connected to WiFi
- …

Heterogeneity is Not Fully Considered

Leaf

PySyft

TensorFlow Federated



PFL
Paddle Federated Learning

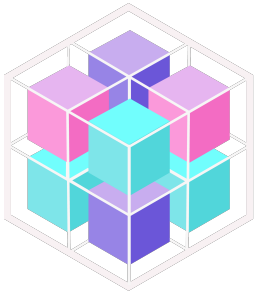
Paddle Federated Learning

Heterogeneity is Not Fully Considered

Leaf

PySyft

TensorFlow Federated



PFL
Paddle Federated Learning

Paddle Federated Learning

Homogeneous devices:

- Uniform hardware capacity
- Always available for training

Impacts of Heterogeneity?

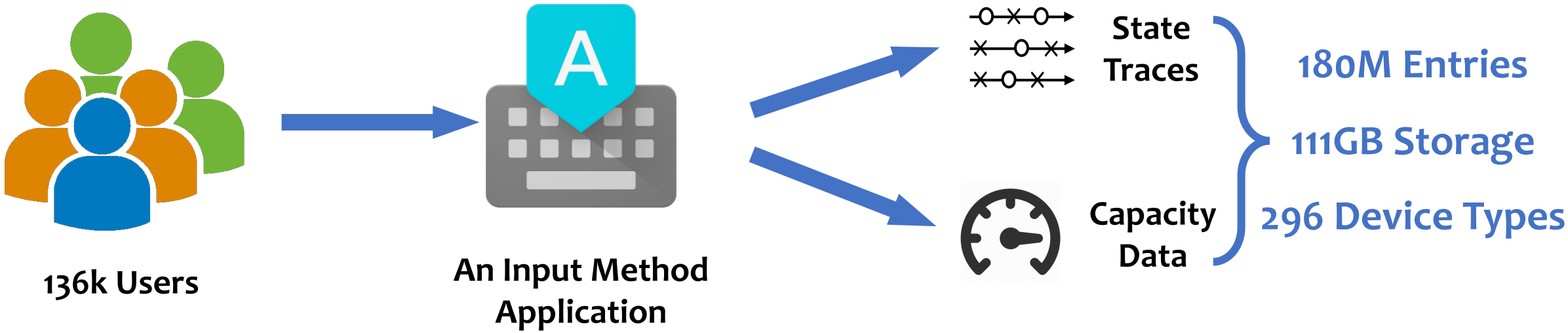
Incorporate Heterogeneity

We need

- Data that describe heterogeneity
- Heterogeneity-aware FL platform

Data Collection

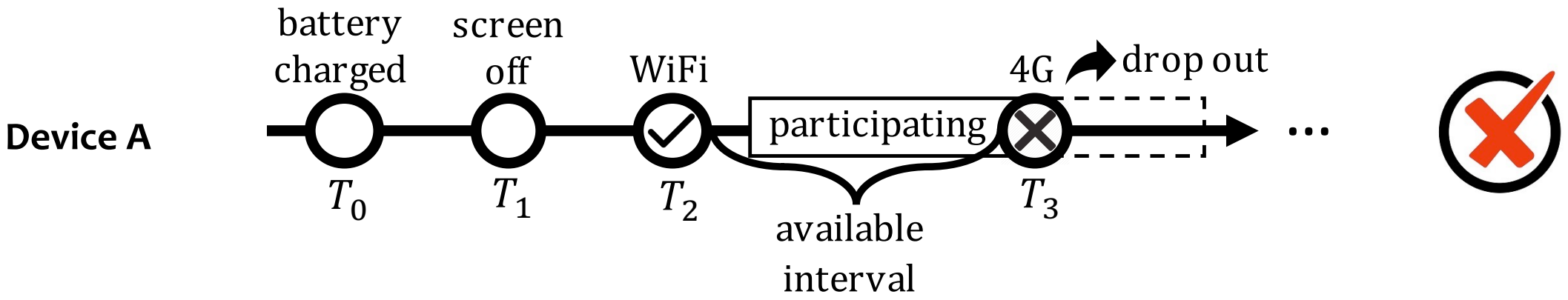
- Data that describe heterogeneity



We collect data from large scale real-world users through a commercial input method application

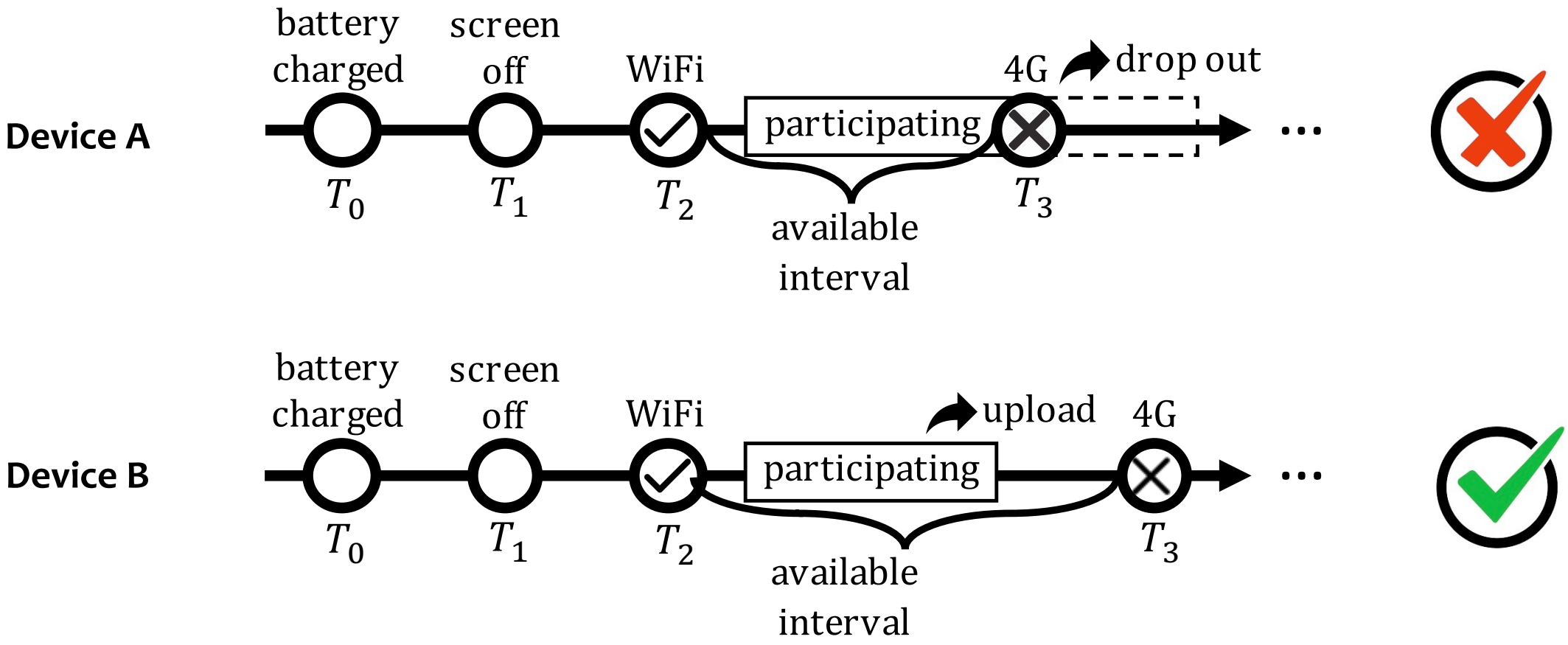
Describe State Heterogeneity

- **State traces** determine devices' checking in and dropping out



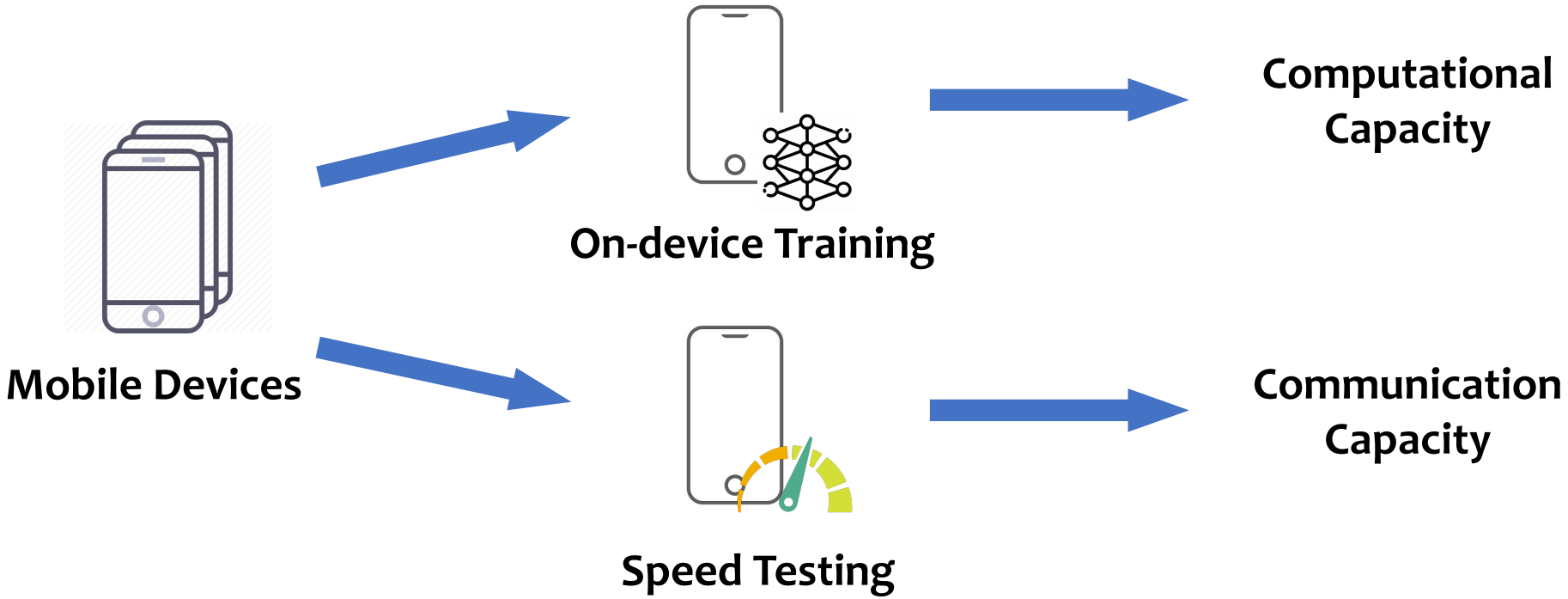
Describe State Heterogeneity

- **State traces** determine devices' checking in and dropping out



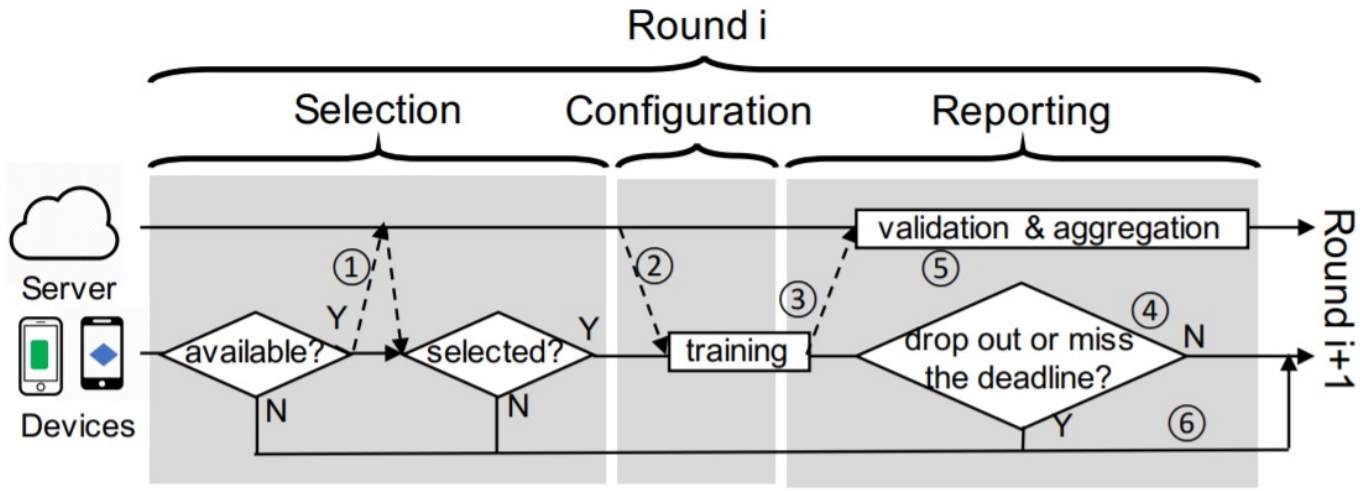
Describe Hardware Heterogeneity

- **Capacity data** determine devices' computational and communication capacity



Heterogeneity-aware FL Platform

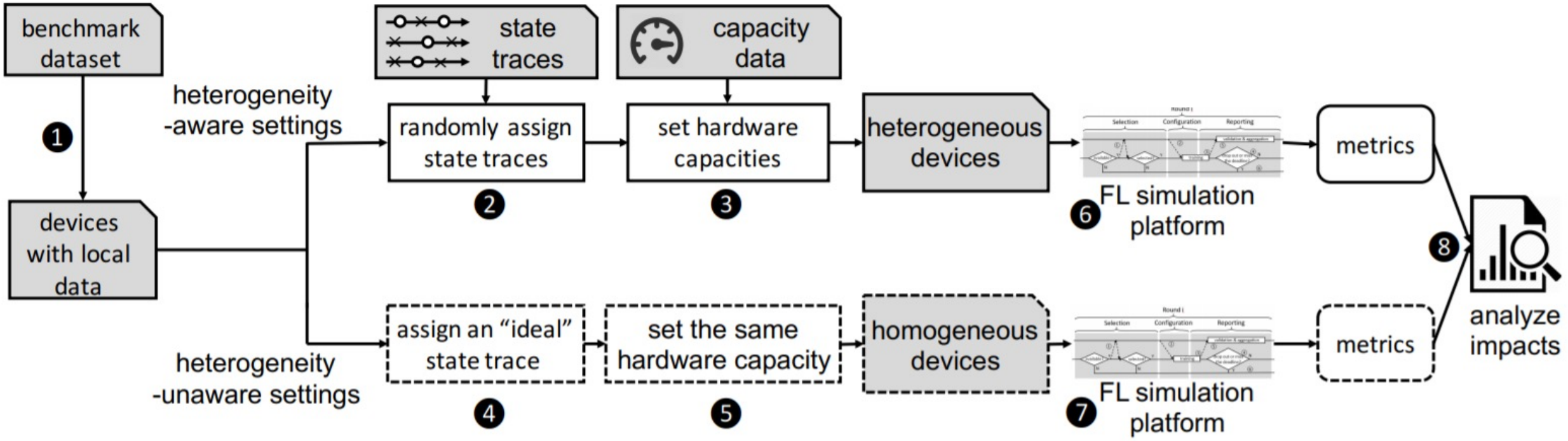
- *FLASH* -- designed according to industrial FL systems



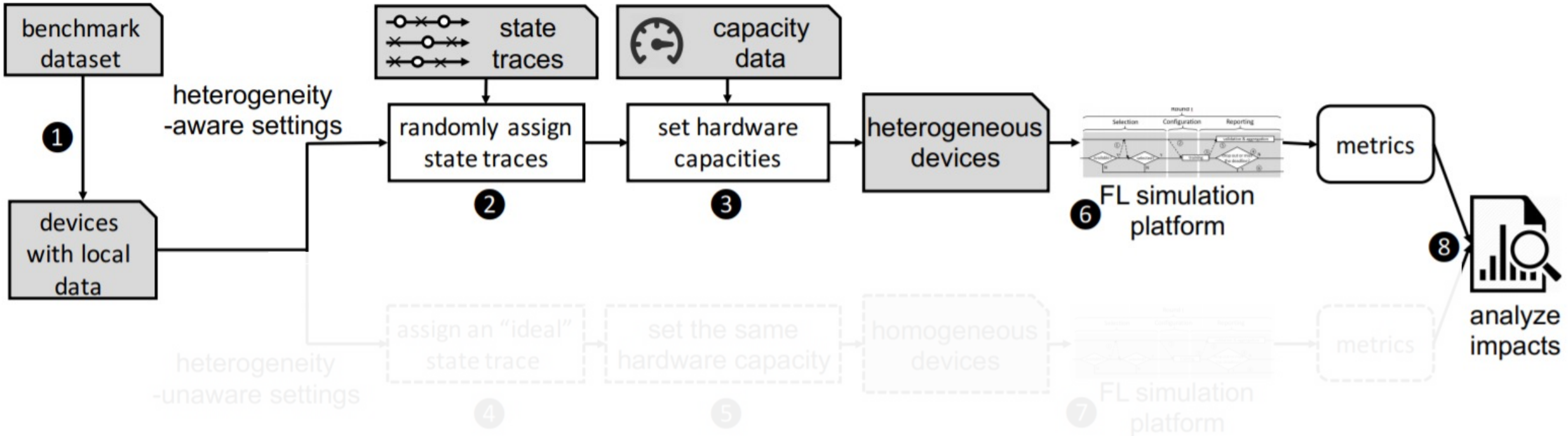
Differences from other platforms:

- Various training/communication time
- Check in and drop out

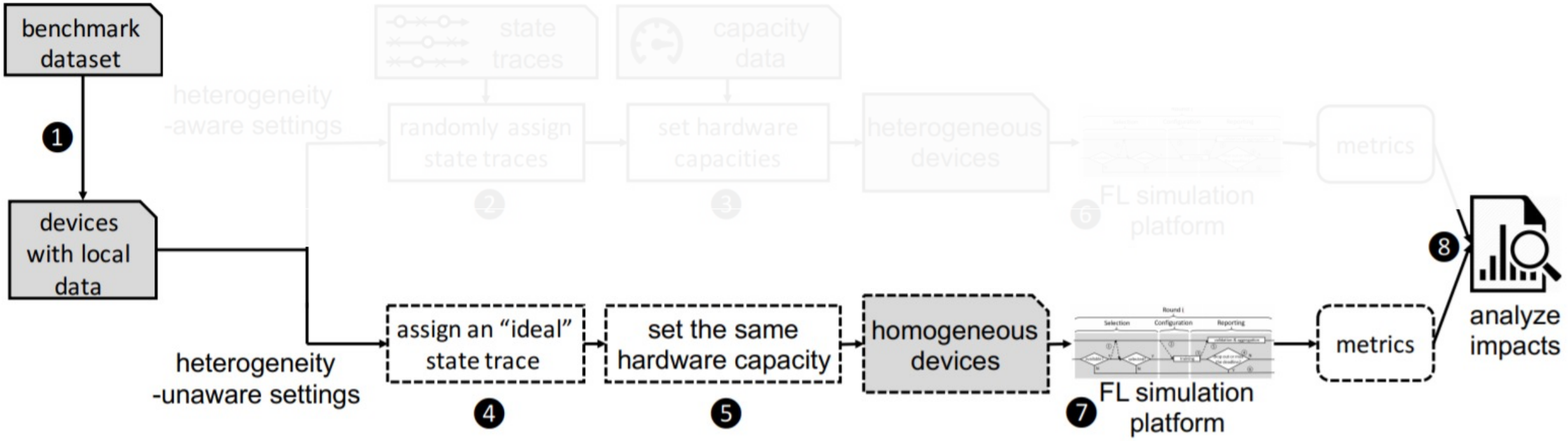
Workflow of the Measurement



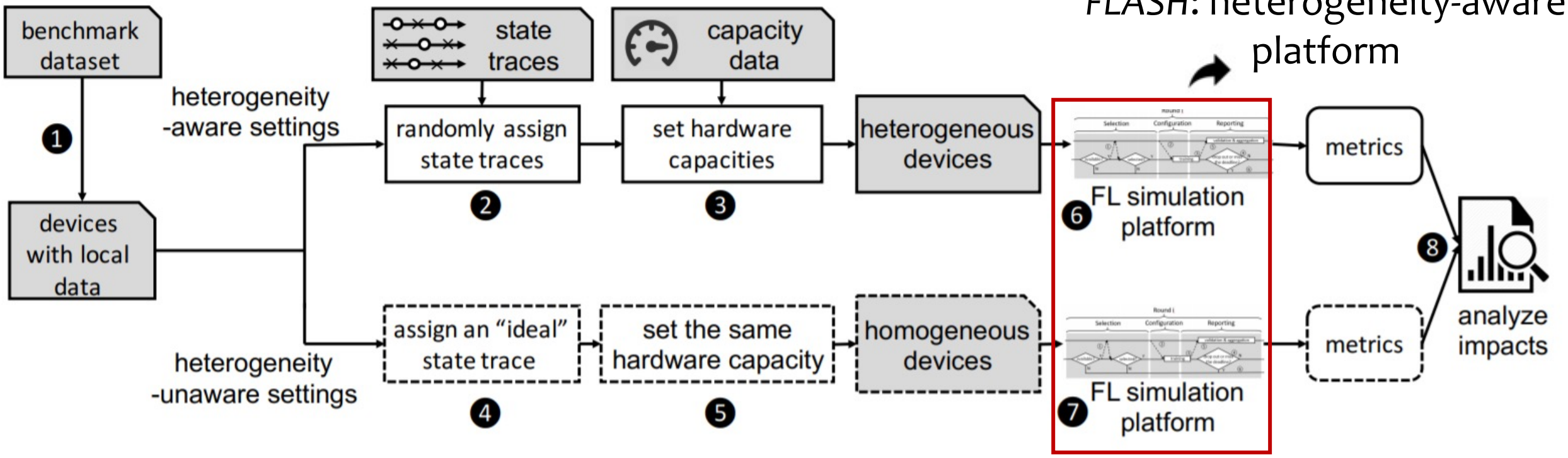
Workflow of the Measurement



Workflow of the Measurement



Workflow of the Measurement

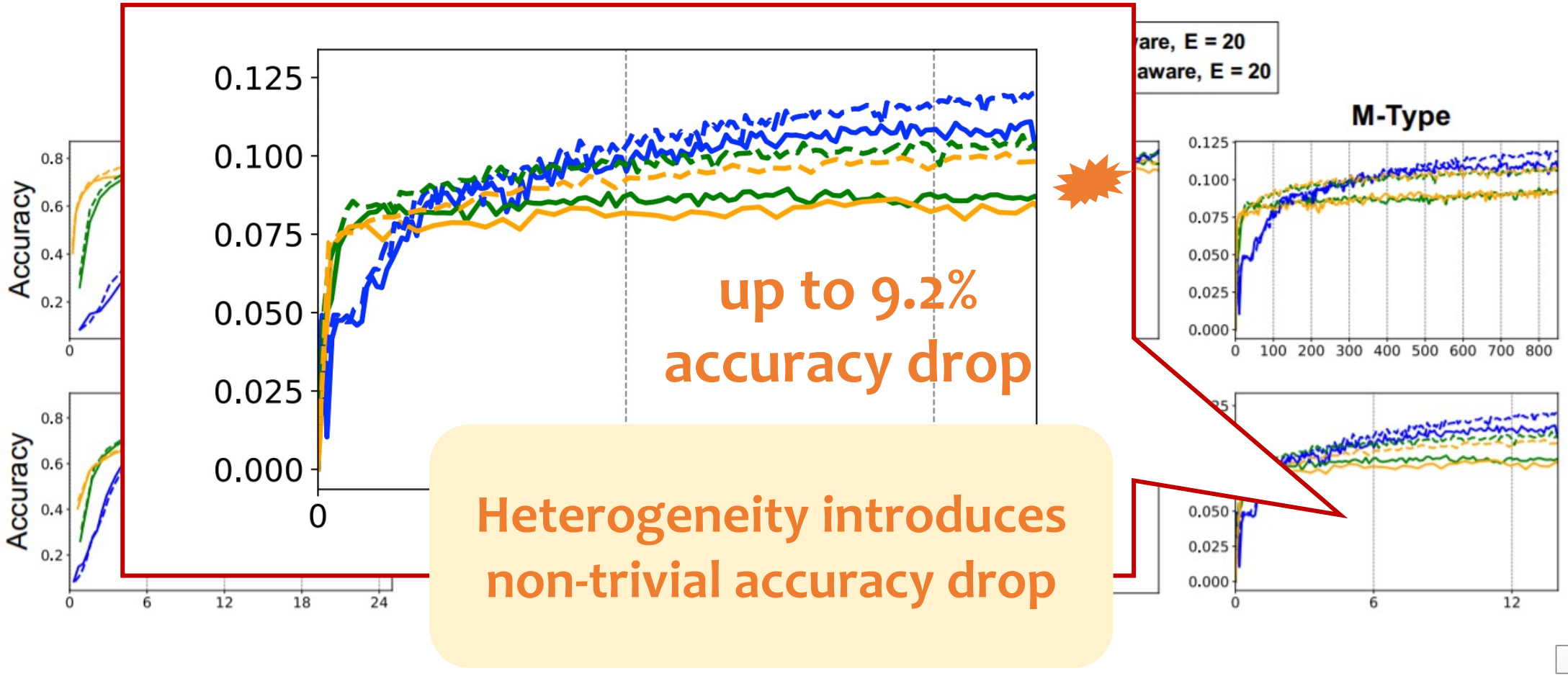


Algorithms and Metrics

Algorithms	Accuracy	Training Time/Round	Compression Ratio	Variance of Accuracy
FedAvg	✓	✓	-	-
Structured Updates	✓	✓	✓	-
GDrop	✓	✓	✓	-
SignSGD	✓	✓	✓	-
q-FedAvg	✓	✓	-	✓
FedProx	✓	✓	-	-

Results

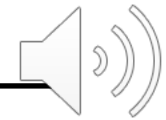
FedAvg's Accuracy on Four Datasets under Heter-aware/unaware Settings



Results

Accuracy for *q-FedAvg* and *FedAvg*. *q-FedAvg* is designed for addressing fairness issues.

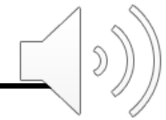
Dataset	Heterogeneity	Algorithm	Average	Worst 10%	Best 10%	Var. $\times 10^{-4}$
Femnist	Unaware	*FedAvg*	82.1%	61.1%	**97.2%**	213
		q-FedAvg	**82.7%**	**64.7%**	95.1%	**157 (26.3% ↓)**
	Aware	*FedAvg*	81.2%	61.1%	94.9%	203
		q-FedAvg	**81.2%**	**64.7%**	**95.1%**	**159 (21.7% ↓)**
M-Type	Unaware	*FedAvg*	**8.2%**	2.3%	**13.5%**	19.2
		q-FedAvg	7.8%	**2.3%**	13.0%	**17.2 (10.5% ↓)**
	Aware	*FedAvg*	7.5%	2.3%	12.3%	16.2
		q-FedAvg	**7.5%**	**2.3%**	**12.4%**	**15.6 (3.7% ↓)**



Results

Accuracy for *q-FedAvg* and *FedAvg*. *q-FedAvg* is designed for addressing fairness issues.

Dataset	Heterogeneity	Algorithm	Average	Worst 10%	Best 10%	Var. $\times 10^{-4}$
Femnist	Unaware	*FedAvg*	82.1%	61.1%	**97.2%**	213
		q-FedAvg	**82.7%**	**64.7%**	95.1%	**157 (26.3% ↓)**
	Aware	*FedAvg*	81.2%	61.1%	94.9%	203
		q-FedAvg	**81.2%**	**64.7%**	**95.1%**	**159 (21.7% ↓)**
M-Type	Unaware	*FedAvg*	**8.2%**	2.3%	**13.5%**	19.2
		q-FedAvg	7.8%	**2.3%**	13.0%	**17.2 (10.5% ↓)**
	Aware	*FedAvg*	7.5%	2.3%	12.3%	16.2
		q-FedAvg	**7.5%**	**2.3%**	**12.4%**	**15.6 (3.7% ↓)**



Results

Accuracy for *q-FedAvg* and *FedAvg*. *q-FedAvg* is designed for addressing fairness issues.

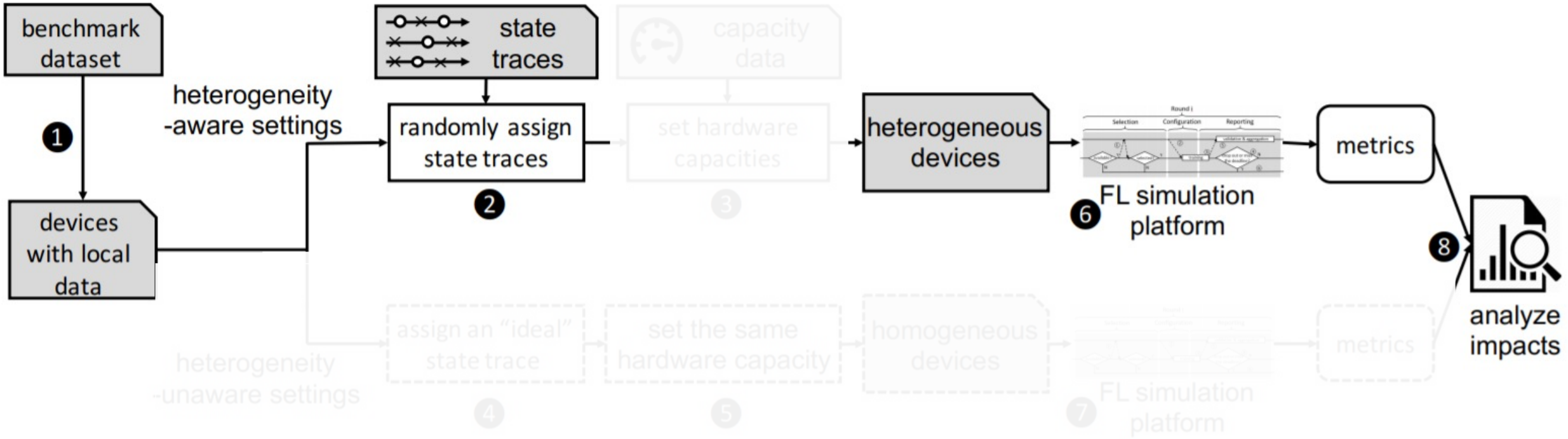
Dataset	Heterogeneity	Algorithm	Average	Worst 10%	Best 10%	Var. $\times 10^{-4}$
Femnist	Unaware	*FedAvg*	82.1%	61.1%	**97.2%**	213
		q-FedAvg	**82.7%**	**64.7%**	95.1%	**157 (26.3% ↓)**
	Aware	*FedAvg*			94.9%	203
		q-FedAvg			**95.1%**	**159 (21.7% ↓)**
M-Type	Unaware	*FedAvg*			**13.5%**	19.2
		q-FedAvg	7.8%	**2.3%**	13.0%	**17.2 (10.5% ↓)**
	Aware	*FedAvg*	7.5%	2.3%	12.3%	16.2
		q-FedAvg	**7.5%**	**2.3%**	**12.4%**	**15.6 (3.7% ↓)**

Heterogeneity hinders q-FedAvg from addressing fairness issues

Reasons for Negative Impacts?

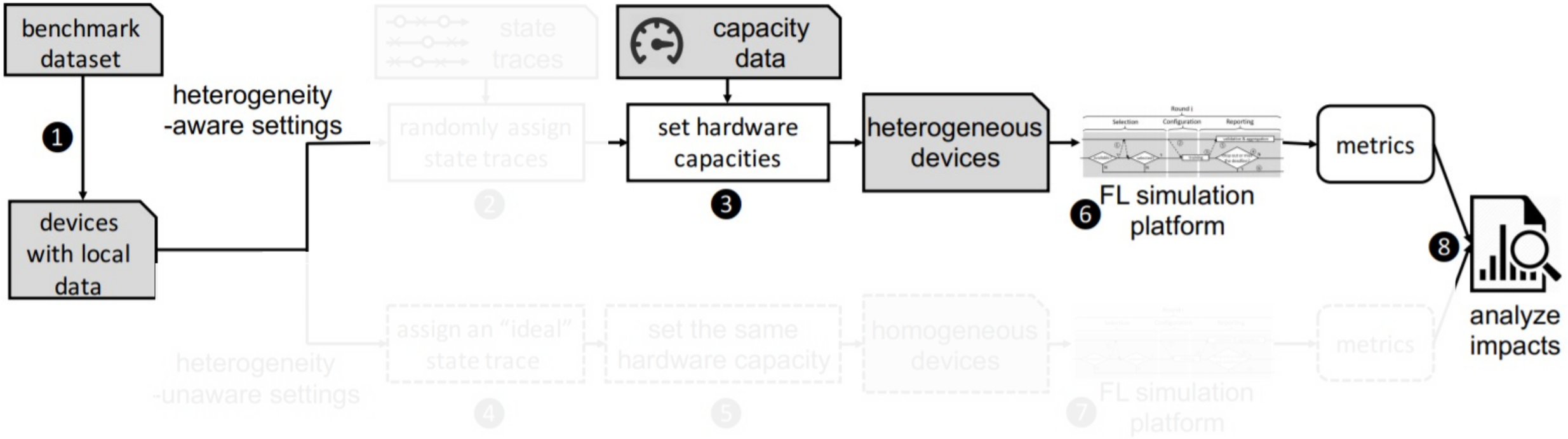
Which Type is More Influential?

Disable hardware heterogeneity



Which Type is More Influential?

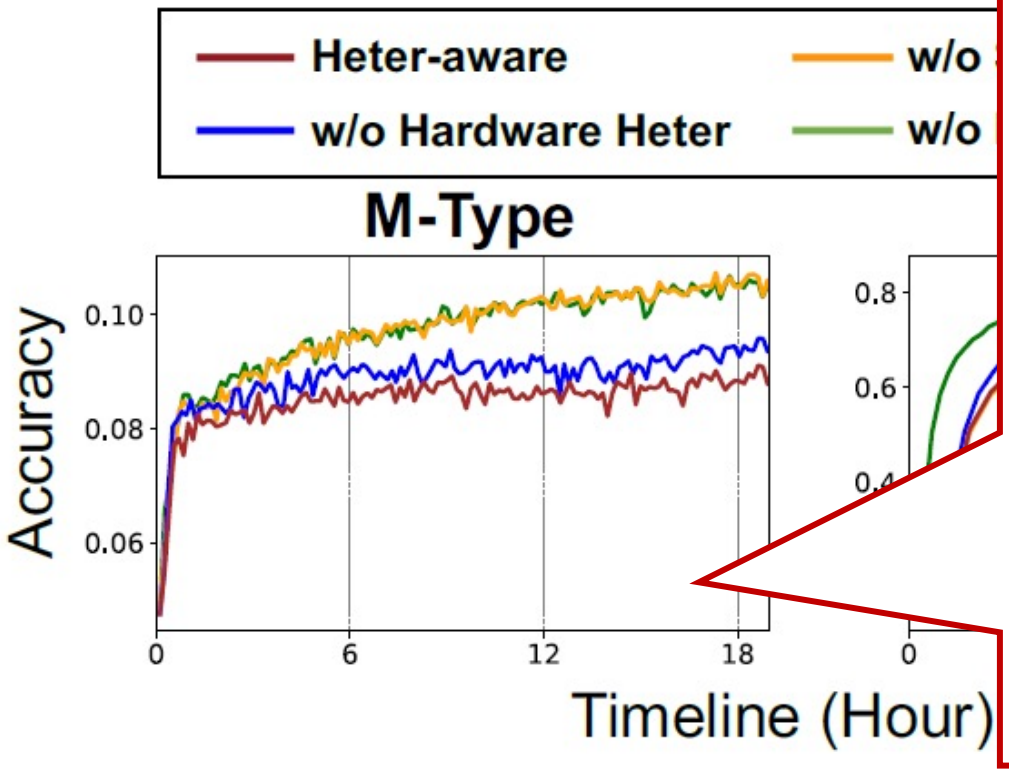
Disable state heterogeneity

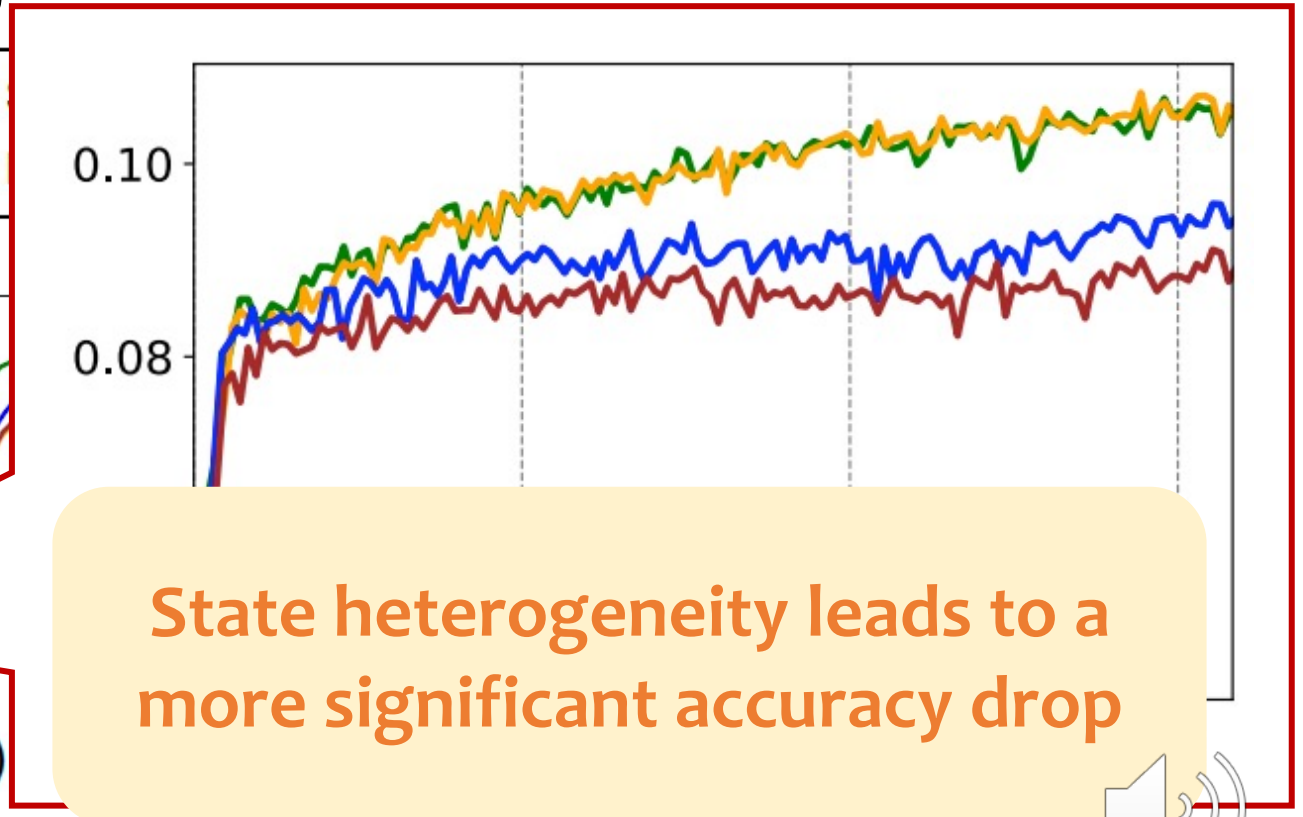


Which Type is More Influential?

State heterogeneity

A breakdown of the impacts of different types of heterogeneity.
(Algorithm: *FedAvg*)



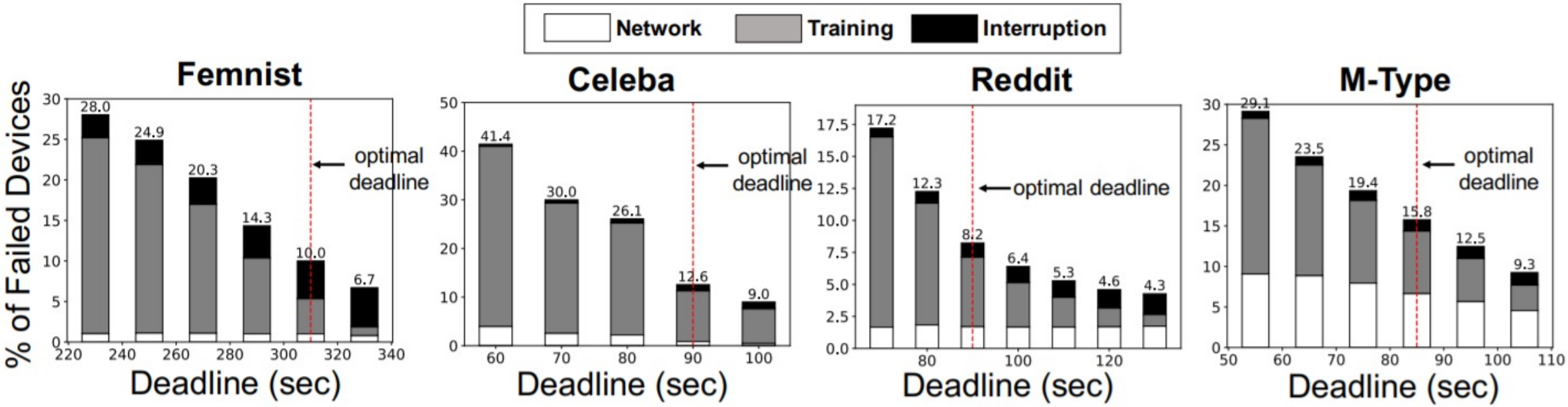


State heterogeneity leads to a more significant accuracy drop

What Causes the Performance Drop?

Device failure

The prevalence of different failure reasons.



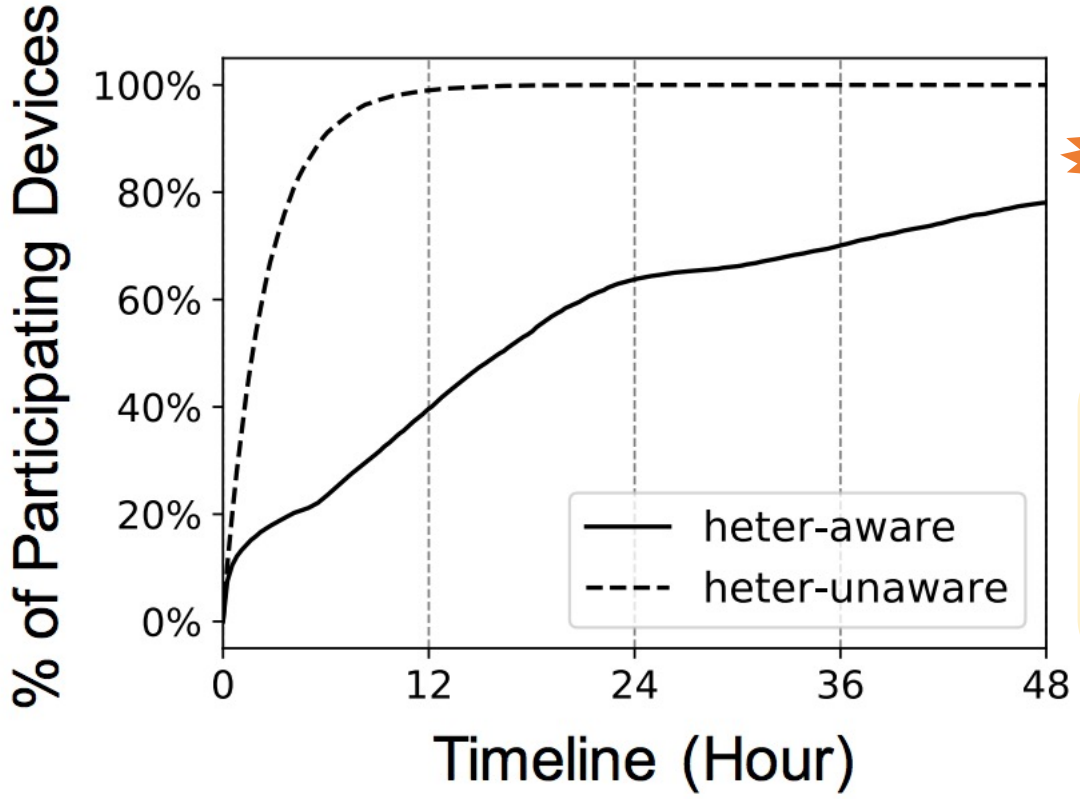
The overall proportion of the failed devices reaches 11.6% on average

What Causes the Performance Drop?

Device failure

Participation bias

Percentage of participating devices over time.



Up to 30% devices have not participated in FL process

Heterogeneity makes devices attend FL in a biased manner

Implications

Consider heterogeneity

Implications

- **Consider heterogeneity**

- **Reduce device failures through a “proactive alerting” technique**

Implications

- **Consider heterogeneity**
- **Reduce device failures through a “proactive alerting” technique**
- **Apply guided participant selection**

Implications

- Consider heterogeneity
- Reduce device failures through a “proactive alerting” technique
- Apply guided participant selection
- Optimize on-device training

Take Away

- A large-scale real-world dataset that describes heterogeneity in FL
- The first heterogeneity-aware FL platform
- Significant impacts of heterogeneity on FL

Thanks!

Data and platform:
https://github.com/PKU-Chengxu/FLASH

yangchengxu@pku.edu.cn

北京大学
PEKING UNIVERSITY

Microsoft