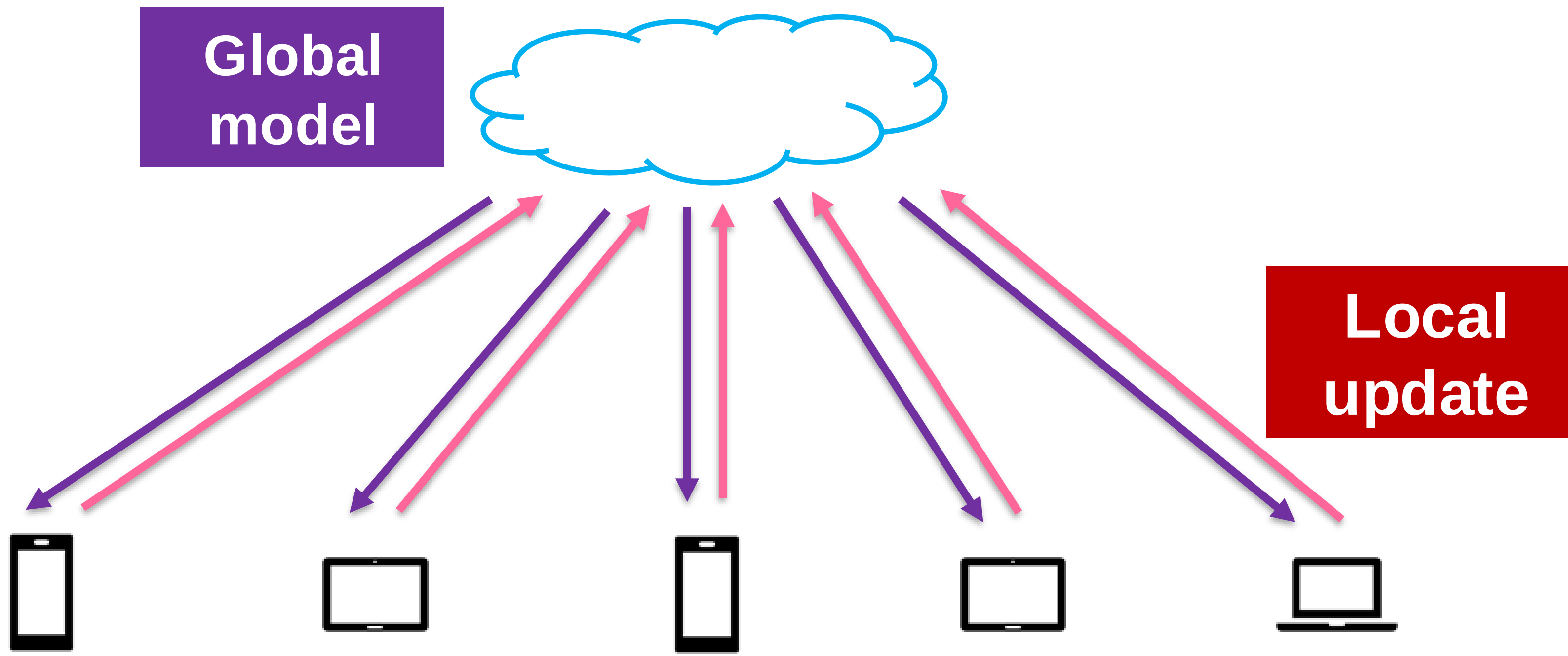
Accelerated Training via Device Similarity in Federated Learning

Yuanli Wang, Joel Wolfrath, Nikhil Sreekumar, Dhruv Kumar, Abhishek Chandra

Distributed Computing Systems Group

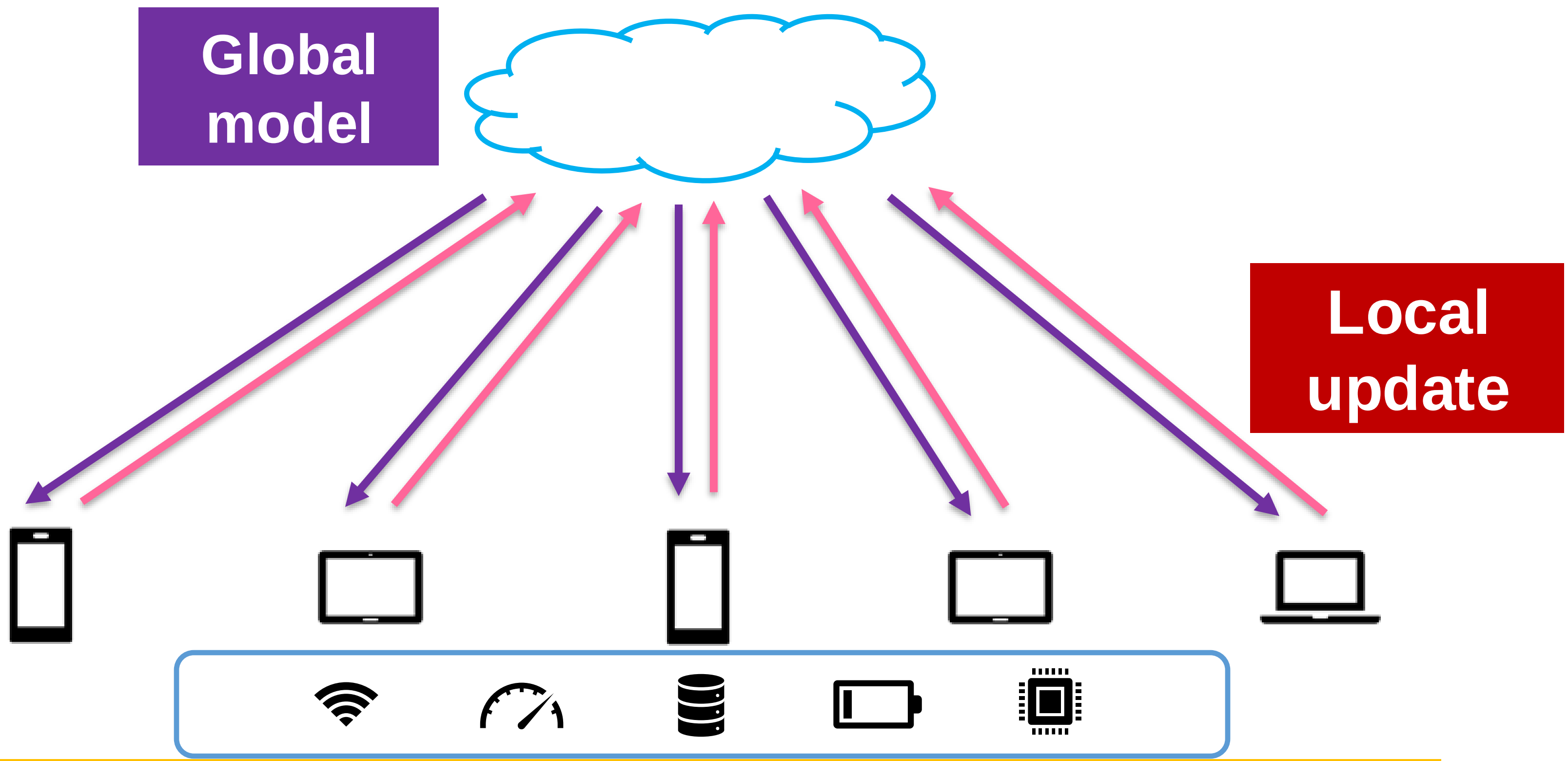
University of Minnesota
Driven to Discover

Federated Learning



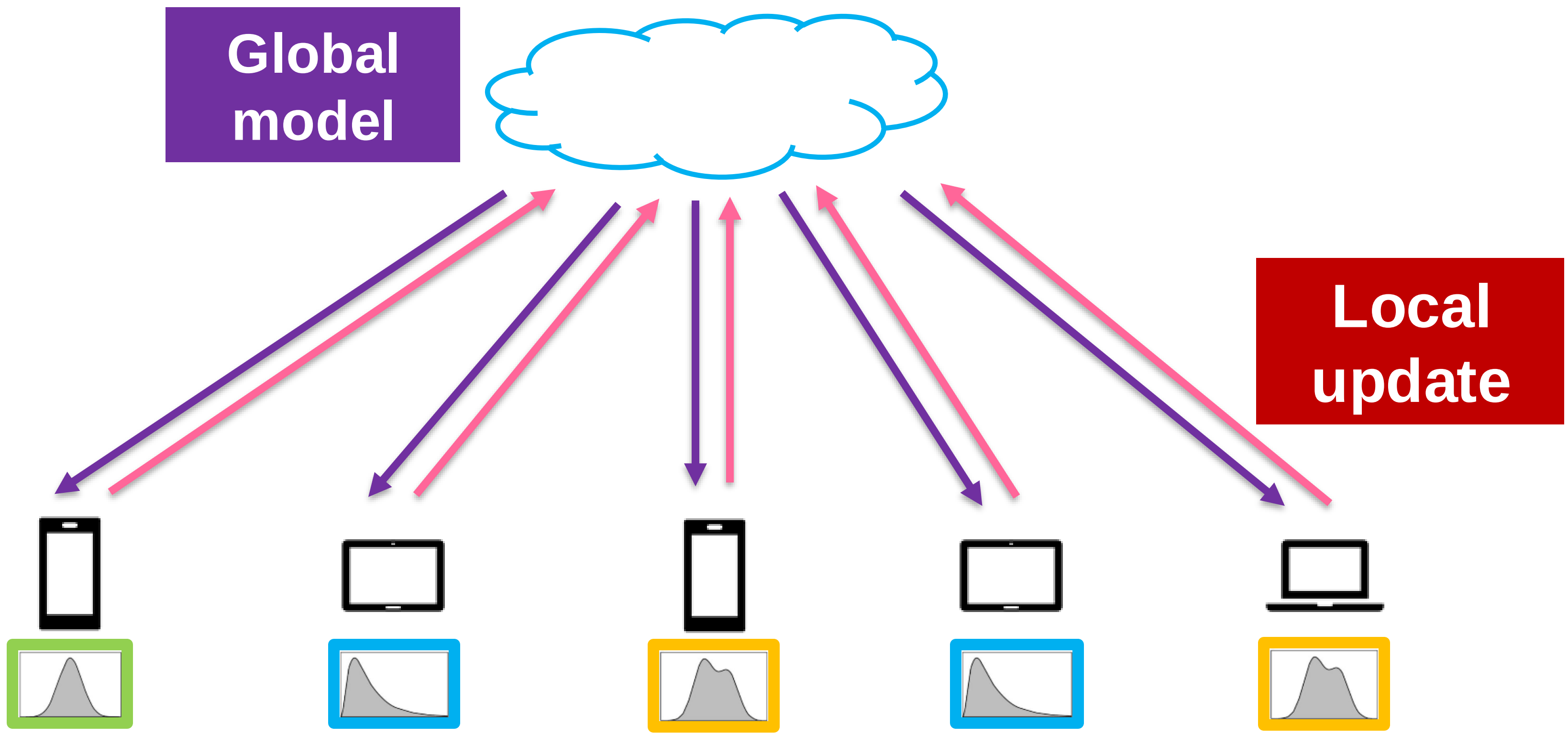
Learn a shared ML model together without uploading private training data

Federated Learning



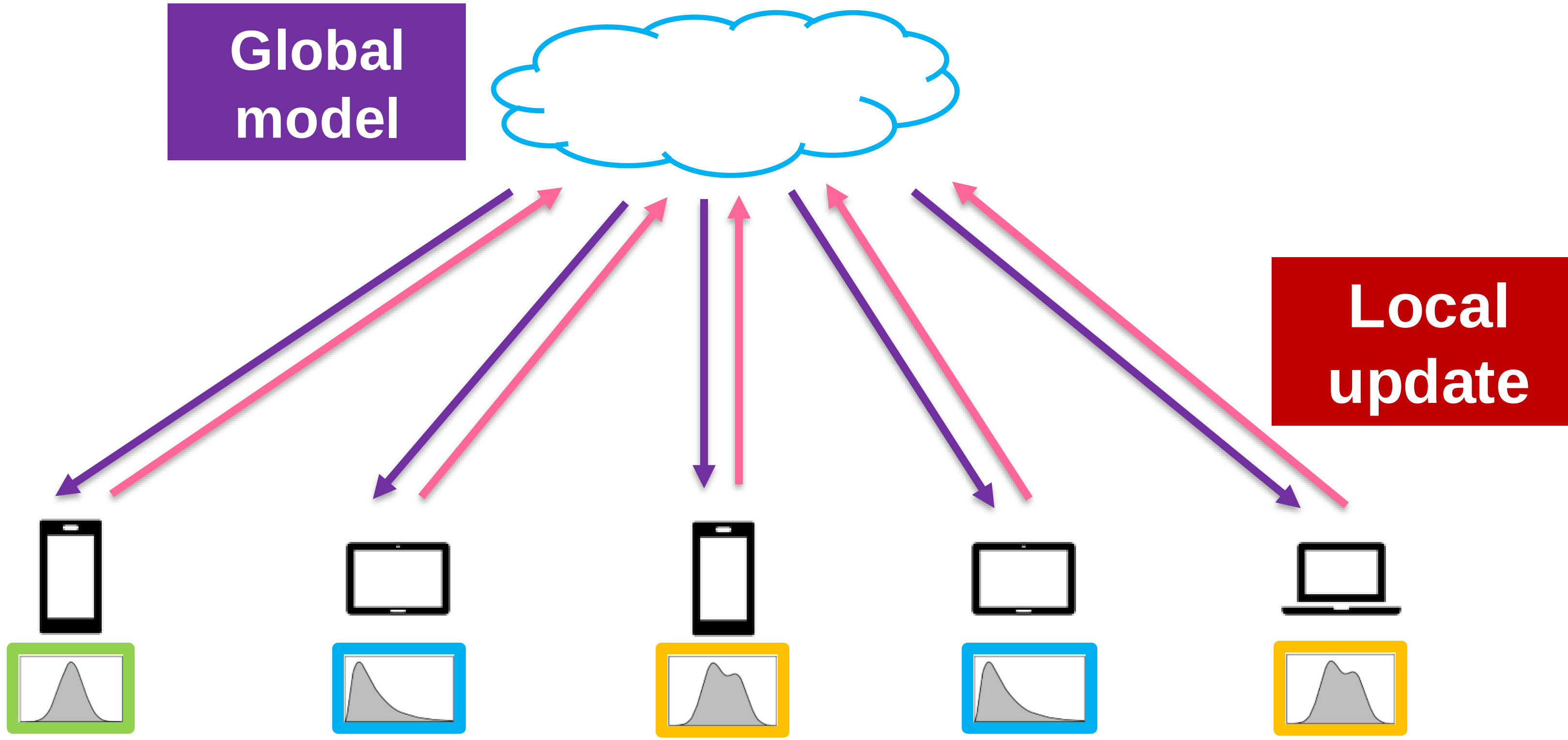
System Heterogeneity: Different devices have different computation resources

Federated Learning



Data Heterogeneity: The dataset of different devices have different statistical distribution(non-IID)

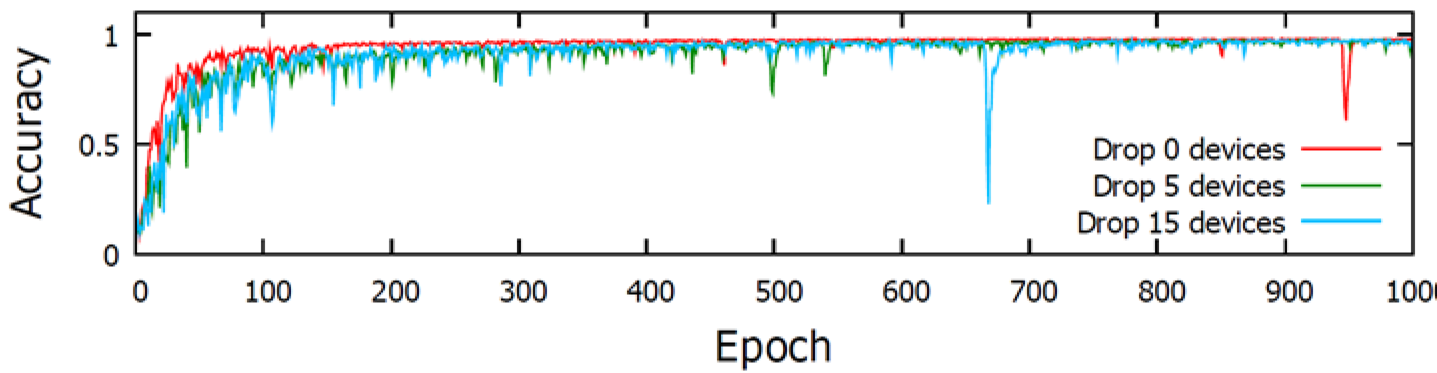
Impact of data heterogeneity



Question: What is the impact of data heterogeneity on Federated Learning?

Intermittent failure of devices

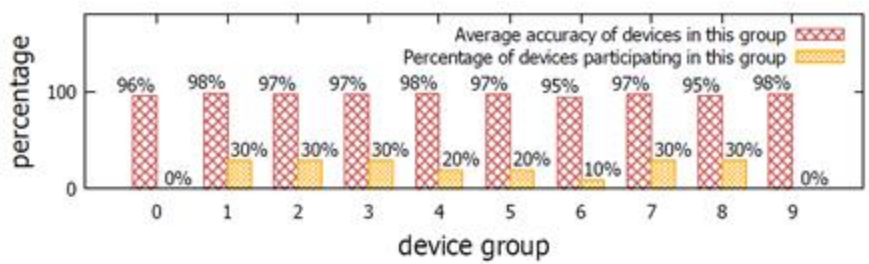
- Divide MNIST dataset into 100 partitions. Each partition is assigned to one client.
- Randomly select 20 clients from these 100 clients for each training epoch.
- Drop 0, 5, or 15 clients out of 20 clients in each epoch



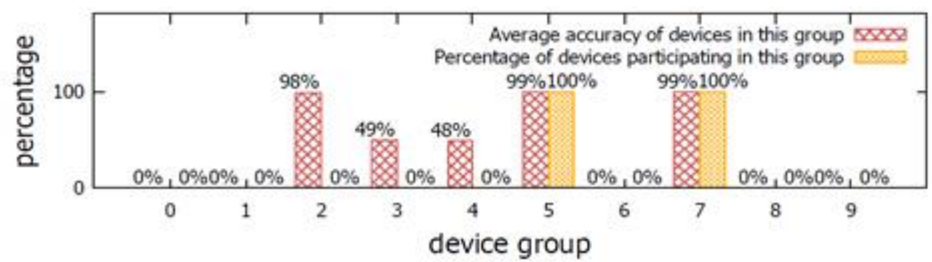
Federated Learning is robust to intermittent failure of devices

Permanent failure of devices

- Partition 100 clients into 10 groups. Each group contains ten clients and will be assigned only two classes from MNIST dataset.
- Drop 80 out of 100 devices under 2 different patterns.
- Measure the trained global model’s accuracy on the local test dataset of each device.



randomly pre-select some clients to drop

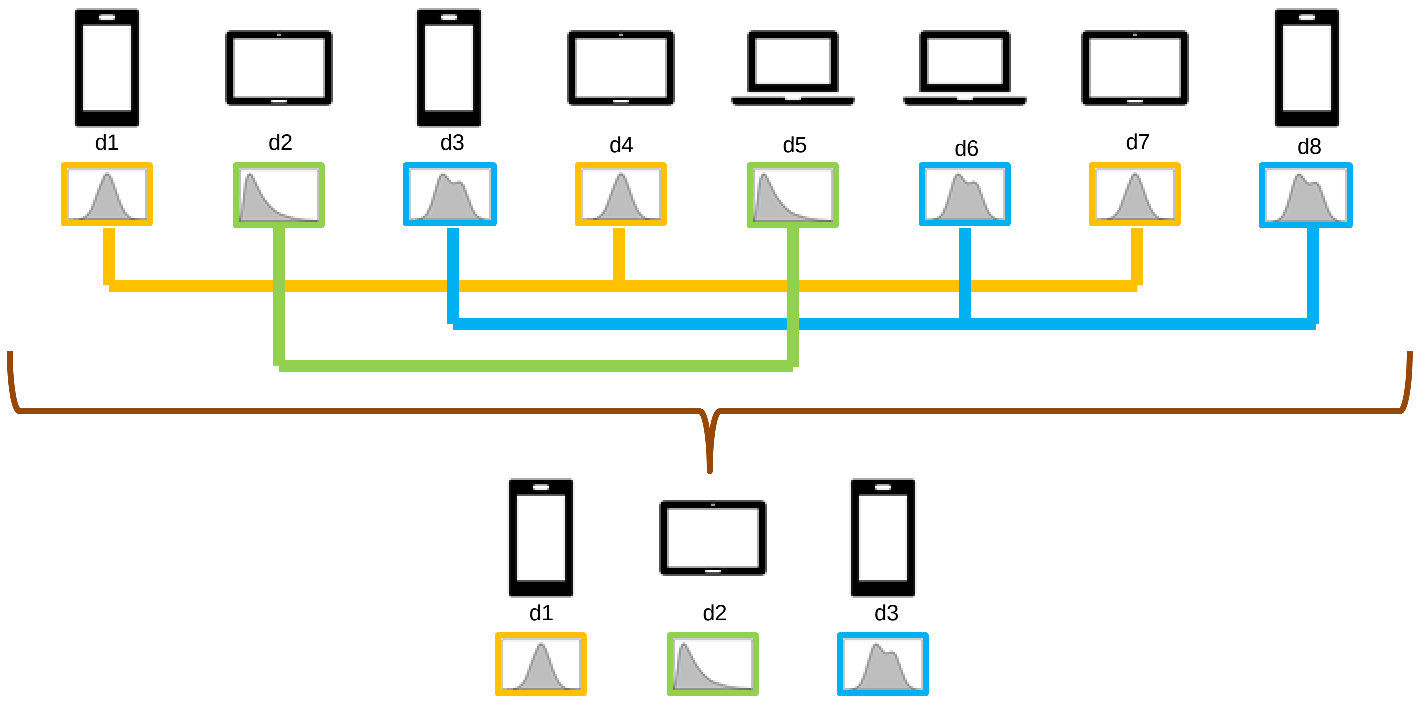


pre-select an entire group of devices to drop

FL is robust to permanent failures, as long as the data heterogeneity is well represented

Exploiting data heterogeneity

Idea: Accelerate training by identifying subsets of devices with "sufficiently similar" data distributions



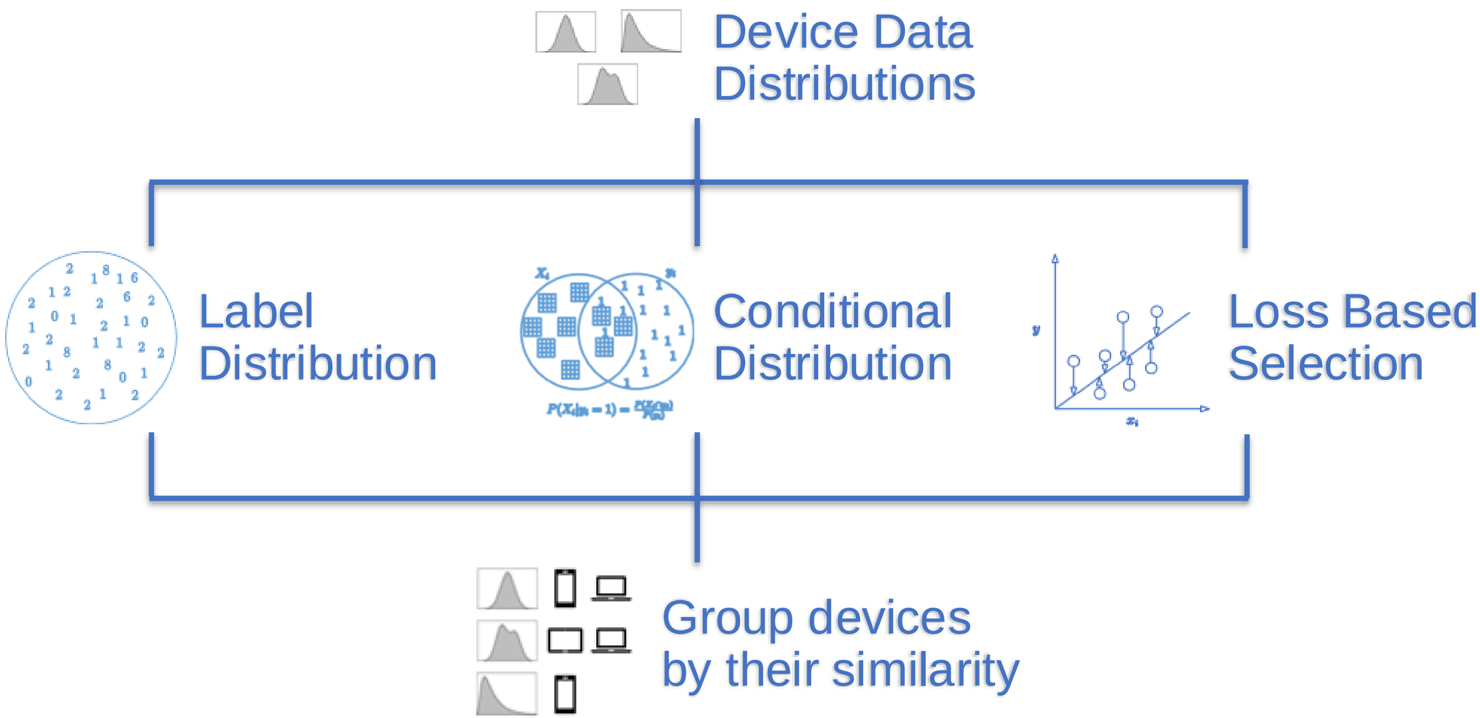
Types of data heterogeneity

Training data at each device drawn from a joint distribution $p(x, y)$

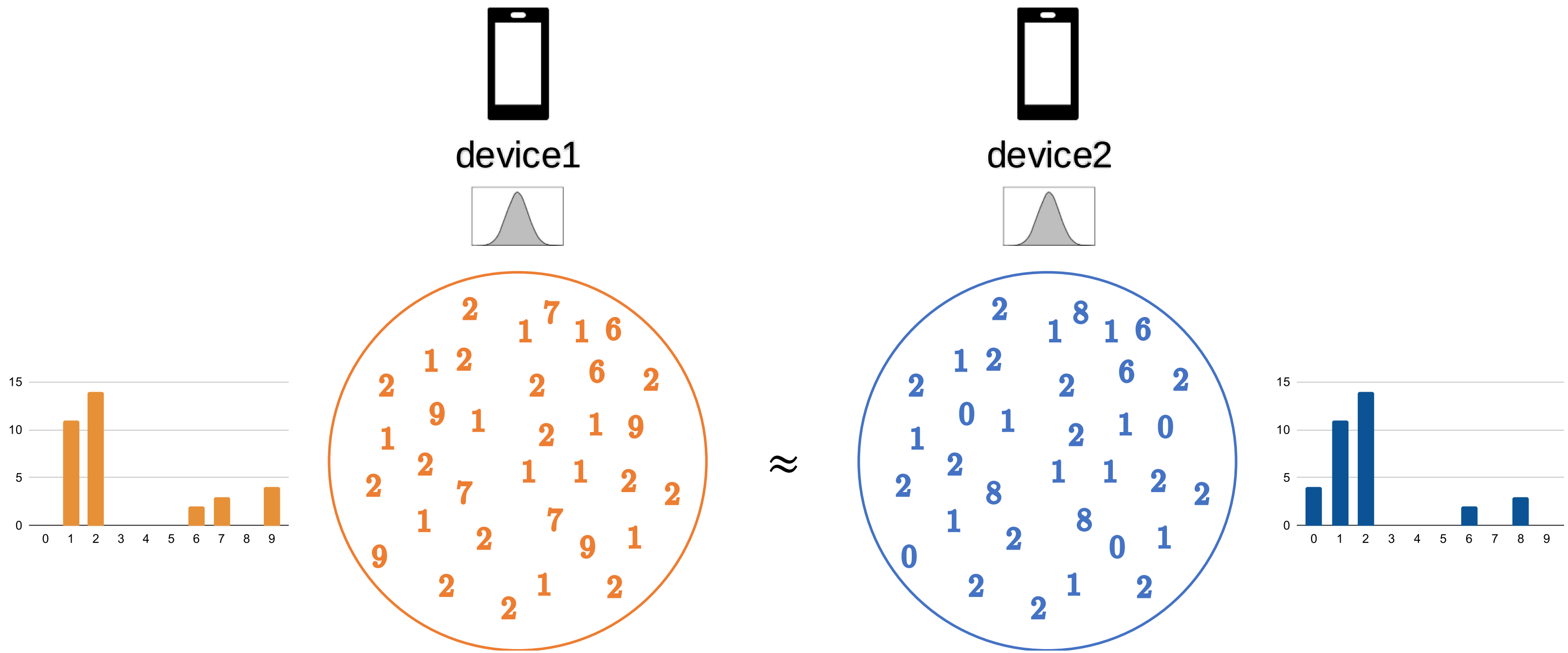
$$p(x, y) = p(y \mid x)\, p(x)$$
$$p(x, y) = p(x \mid y)\, p(y)$$

$p(x)$	Features have different distributions
$p(y)$	Labels have different distributions
$p(x \mid y)$	Different data generates the same labels
$p(y \mid x)$	Given the same data, labels might differ

Our Solution: Identify Data Similarity

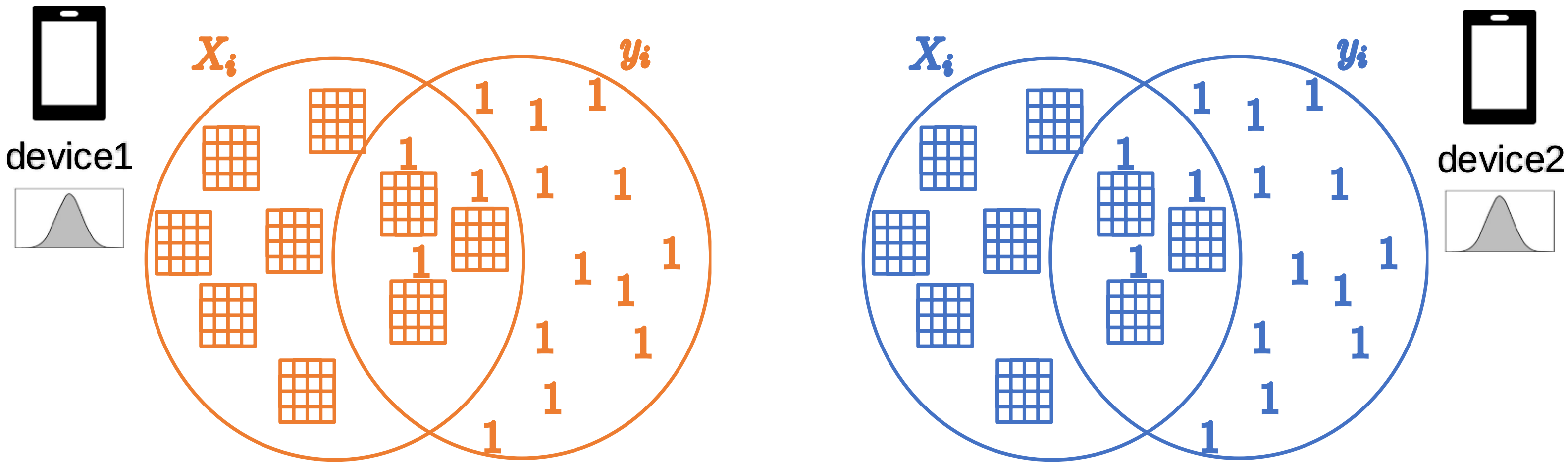


Method #1: Class Label Distribution



device1 and device2 similar

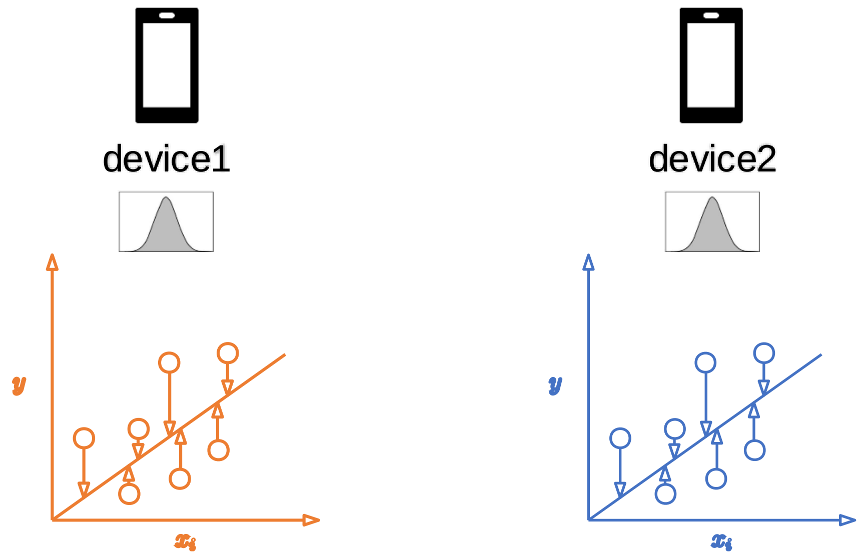
Method #2: Conditional Data Distribution



$$P(X_i|y_i = 1) = \frac{P(X_i \cap y_i)}{P(y_i)} \approx P(X_i|y_i = 1) = \frac{P(X_i \cap y_i)}{P(y_i)}$$

device1 and device2 similar

Method #3: Loss based Selection



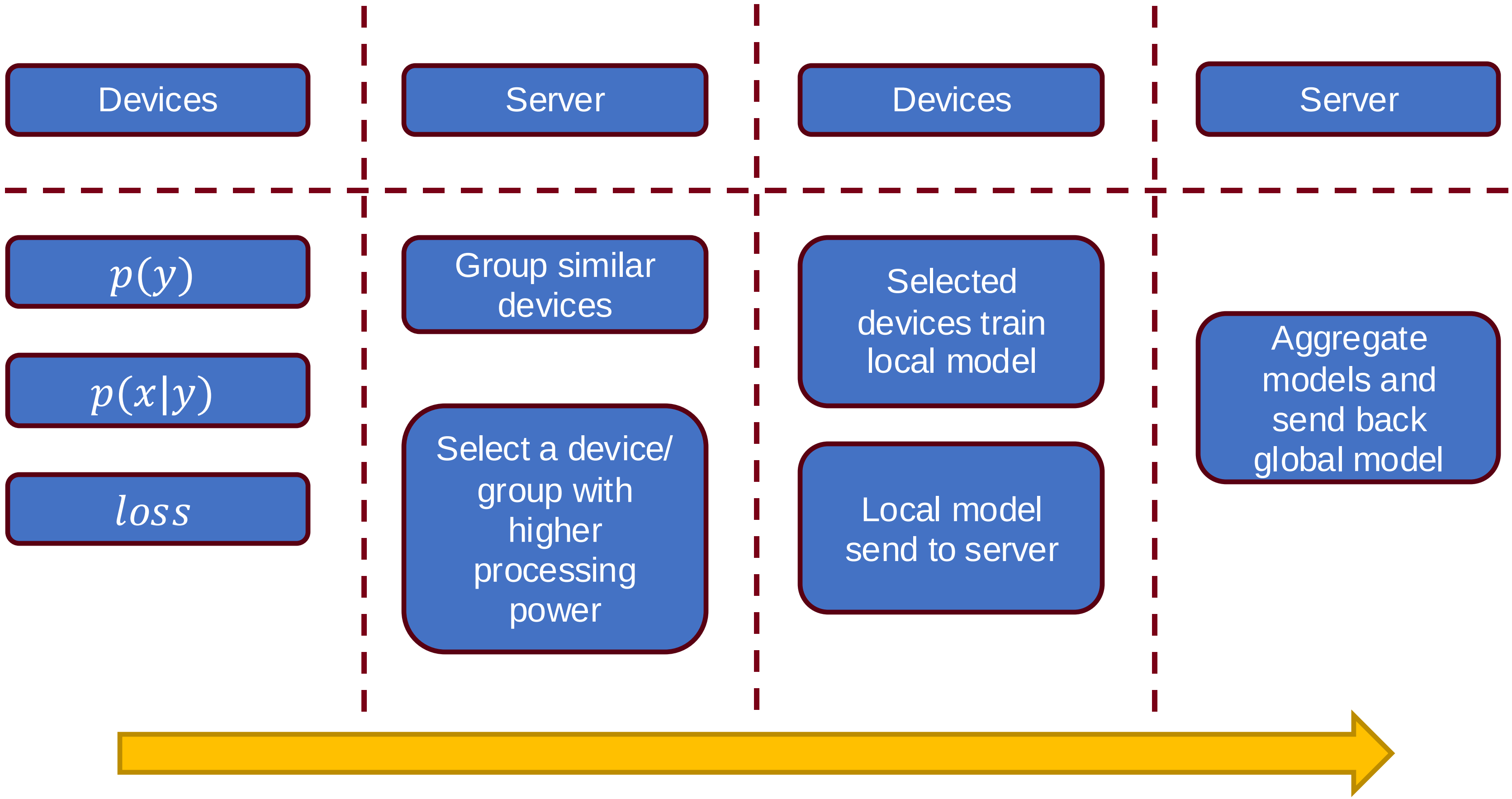
$$Loss_{device1} \approx Loss_{device2}$$

Hypothesis: device1 and device2 similar

Trade-off between data size and privacy

Data summary method	Data Size sent from each device	Privacy
Random	None	Complete Privacy
Distribution of Class Labels	$\Theta(m_i)$	Partial
Conditional Data Distribution	$\Theta(s_i m_i)$	Partial
Empirical Loss on Global Model	$\Theta(1)$	Stronger Privacy

Training Pipeline: One Epoch

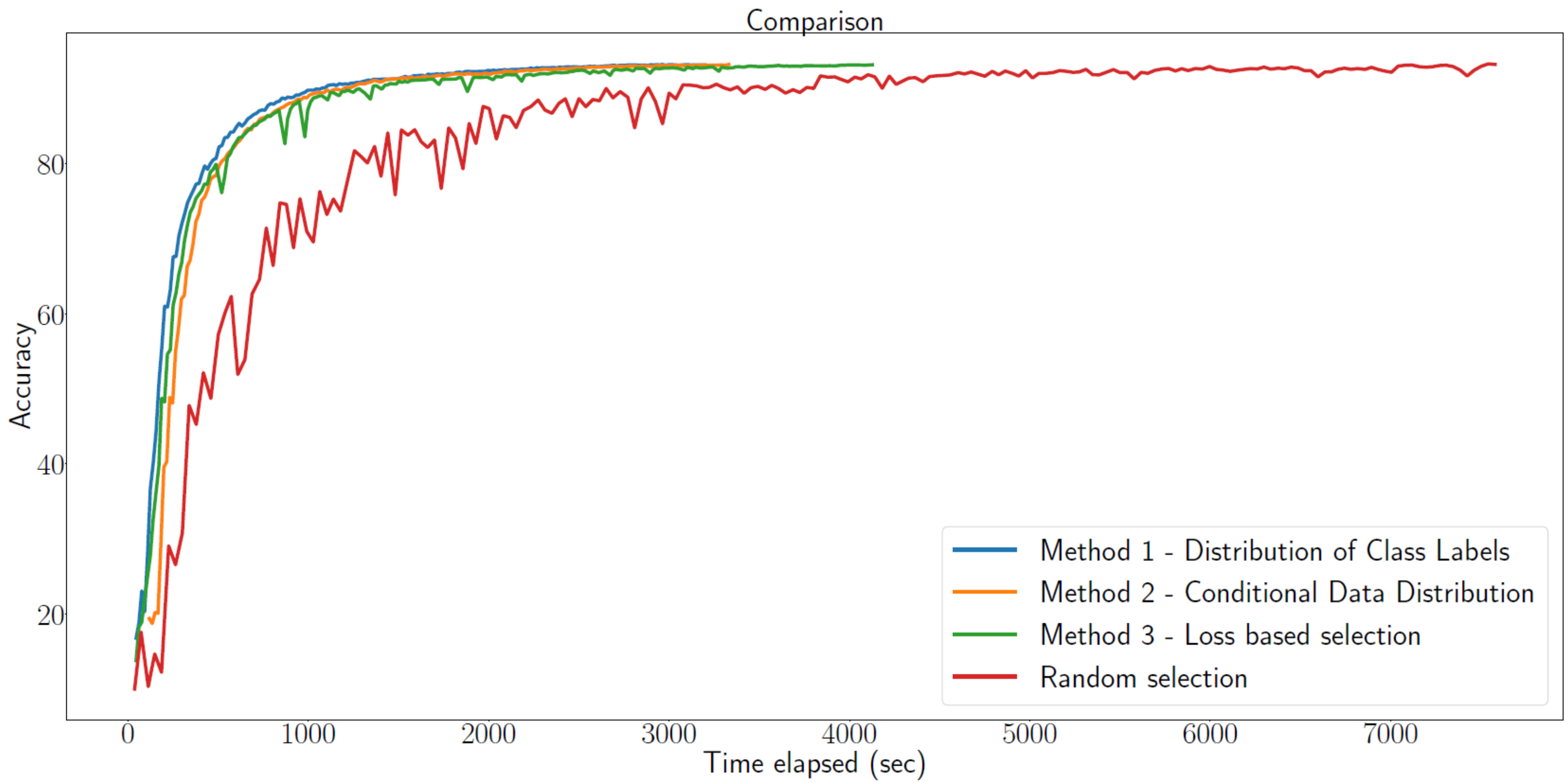


Experimental Setup

- 20 devices (10 slow and 10 fast)
  - 2 for each class label. 1 slow and 1 fast
- Dataset: MNIST - Handwriting
- Distribution
  - Each device data has one prominent digit and noise from 3 other digits
  - Same distribution ratio maintained across training and testing

Prominent digit	0	1	2	3	4	5	6	7	8	9
device speed	Slow	Slow	Slow	Slow	Slow	Slow	Slow	Slow	Slow	Slow
device speed	Fast	Fast	Fast	Fast	Fast	Fast	Fast	Fast	Fast	Fast

Evaluation



46% - 58% reduction in training time to reach the same level of accuracy

Conclusion

- Explore the impact of data heterogeneity in federated learning
- Propose 3 methods for evaluating the similarity of device data
- 46% to 58% reduction of training speed via device similarity

Future work

- Fault tolerance
- Profile more system heterogeneity factors
- Explore other data summaries