

Yaqi Wang
Hulu, Beijing

M. Nikkhah
6D.ai

X. Zhu, W. Tan, R. Liston
Innovations Lab, Cisco Systems Inc.

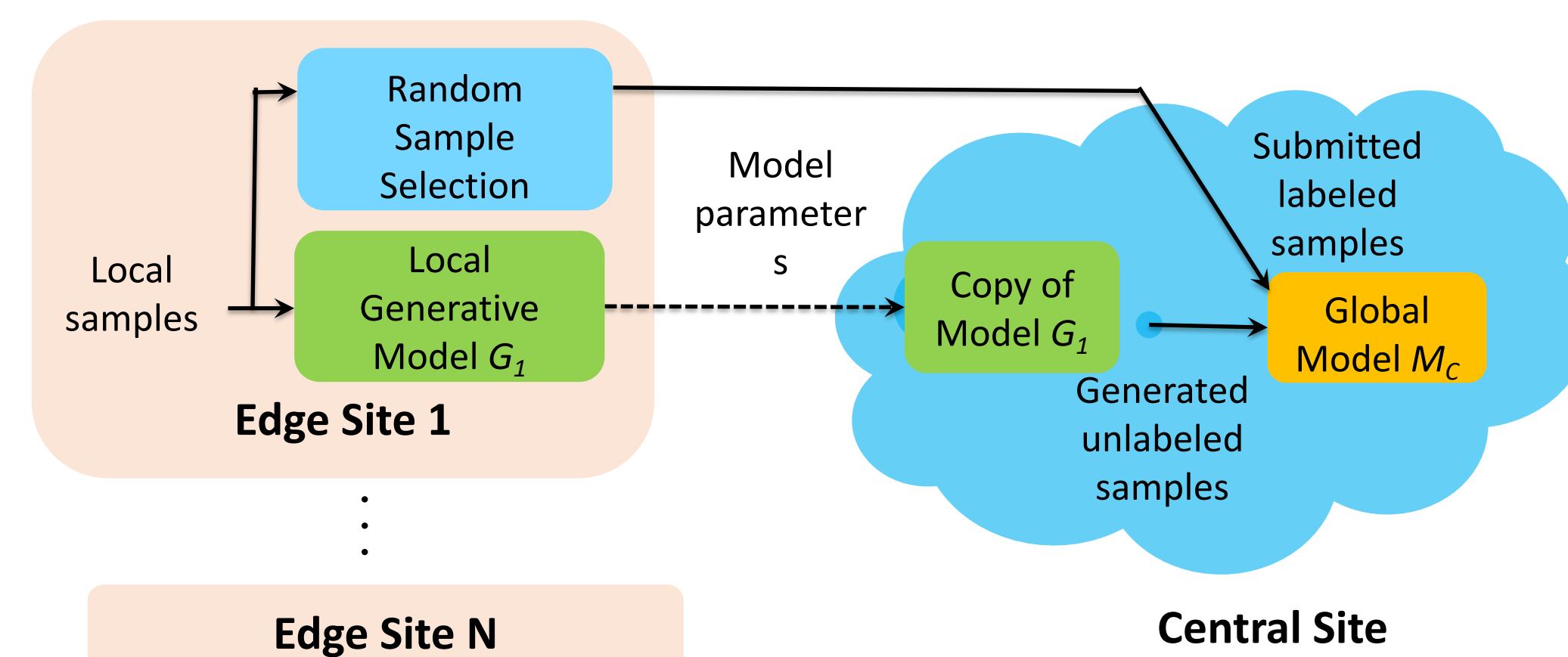
Learning from Geographically Distributed Data



Goals:

- Bandwidth efficiency
- Site scalability
- Supports multiple learning tasks

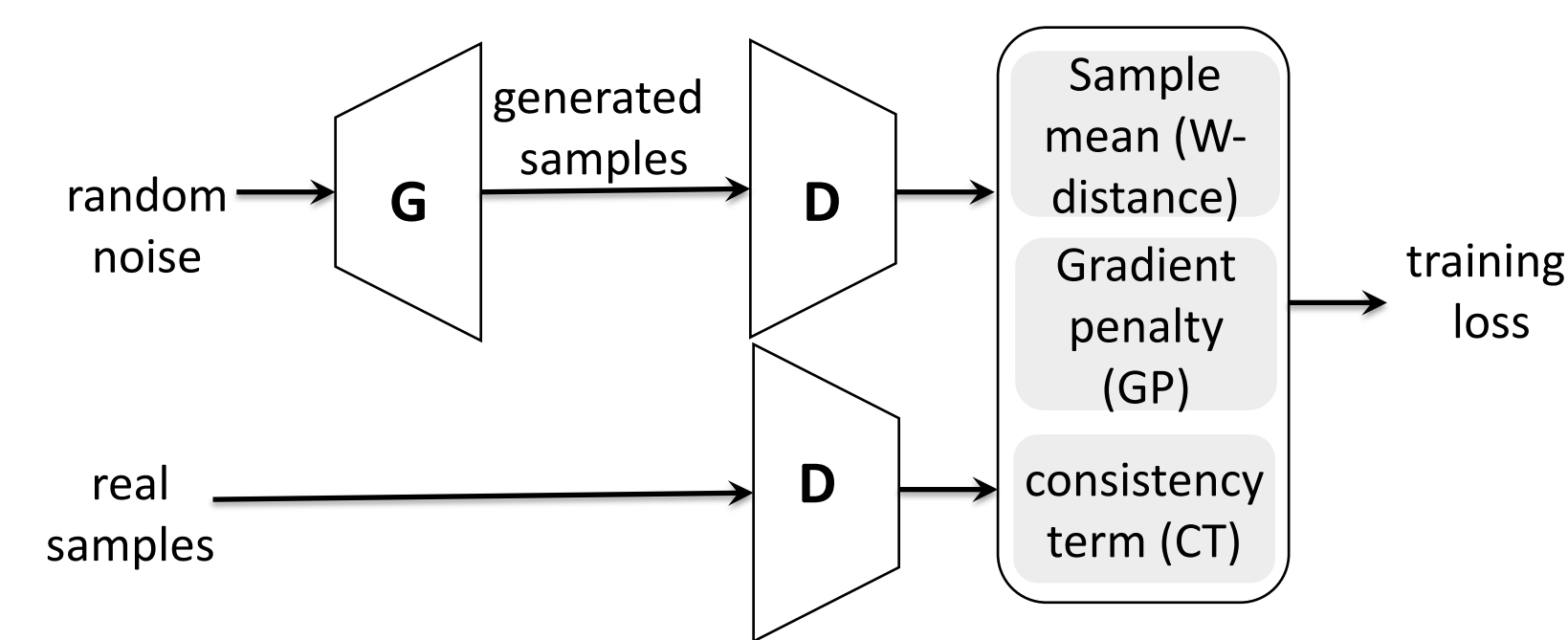
A Hybrid GAN-based Approach



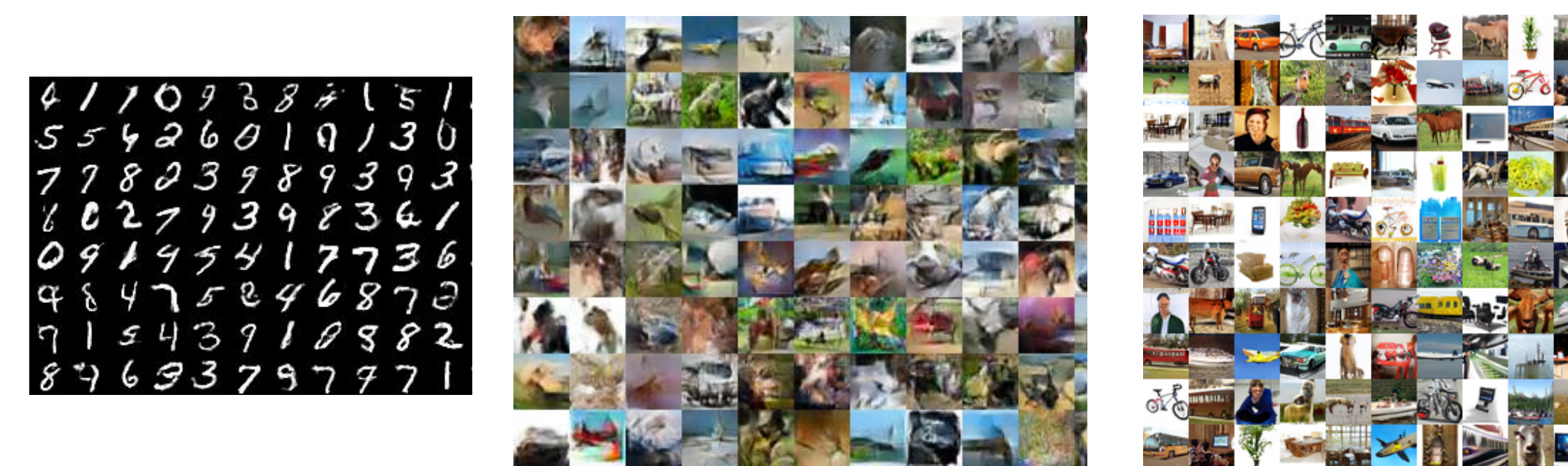
- Local training of generative models
- Centralized training of global discriminative model using *generated* and real samples
- Requires no feedback to local sites

Local Training of GAN Models

- Training of CT-GAN [Wei et al., ICML'18]



- Generated image samples



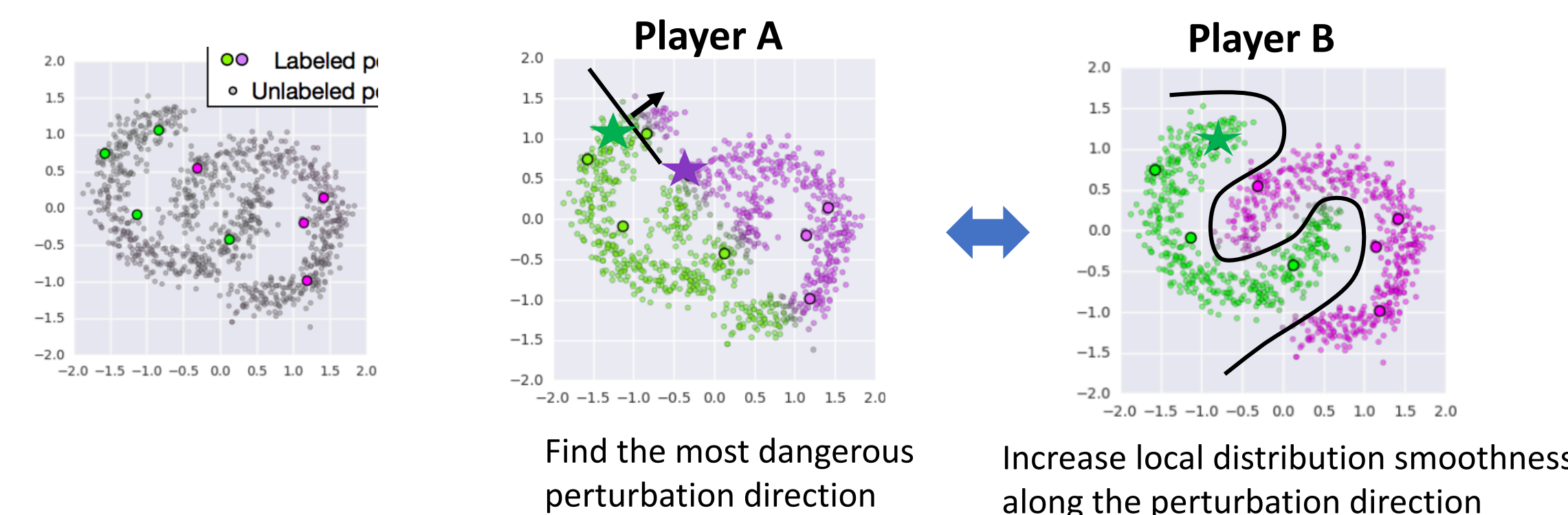
MNIST

CIFAR-10

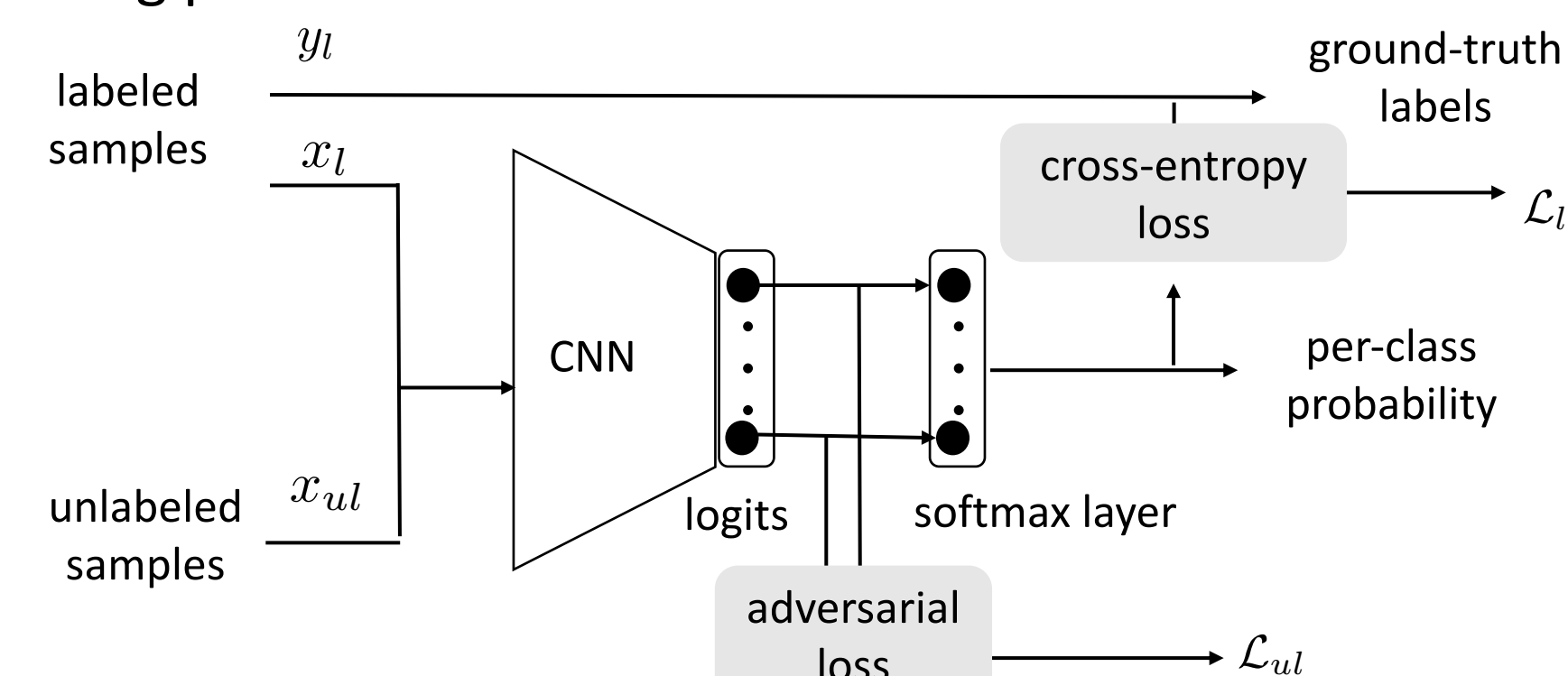
LSUN-20

Centralized Semi-supervised Learning

- Virtual Adversarial Training [Miyato et al., 2017]

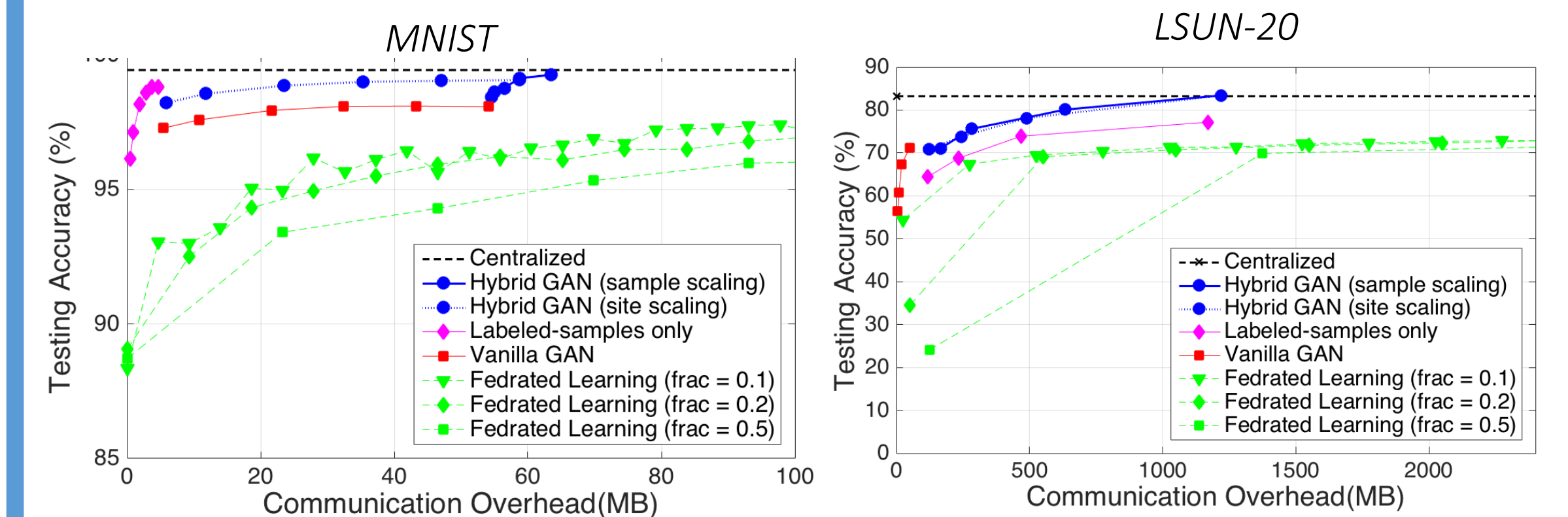


- Training procedure

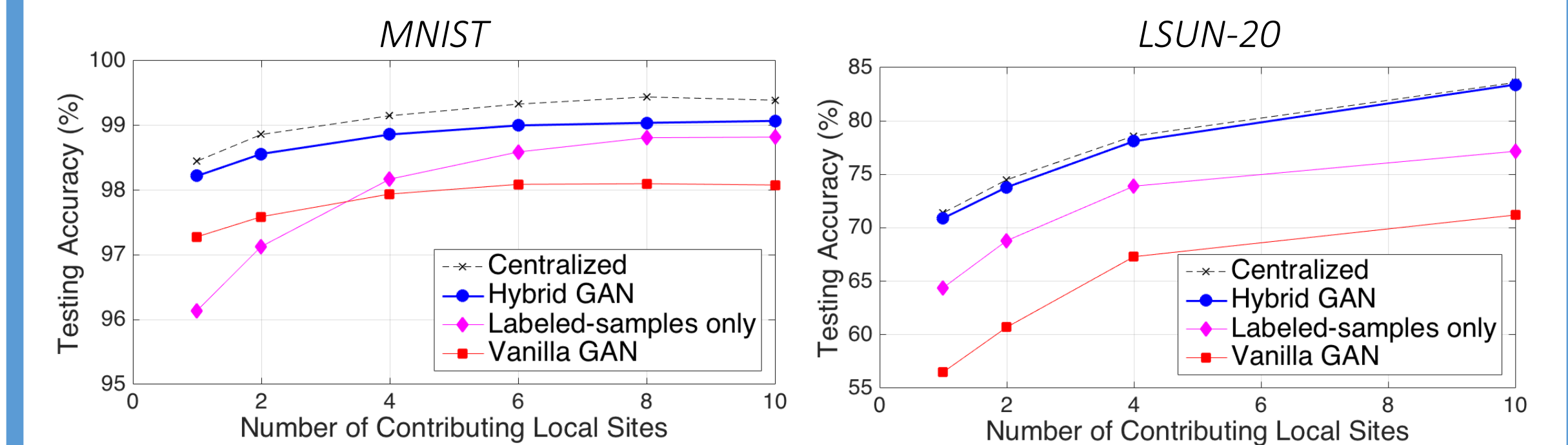


Results

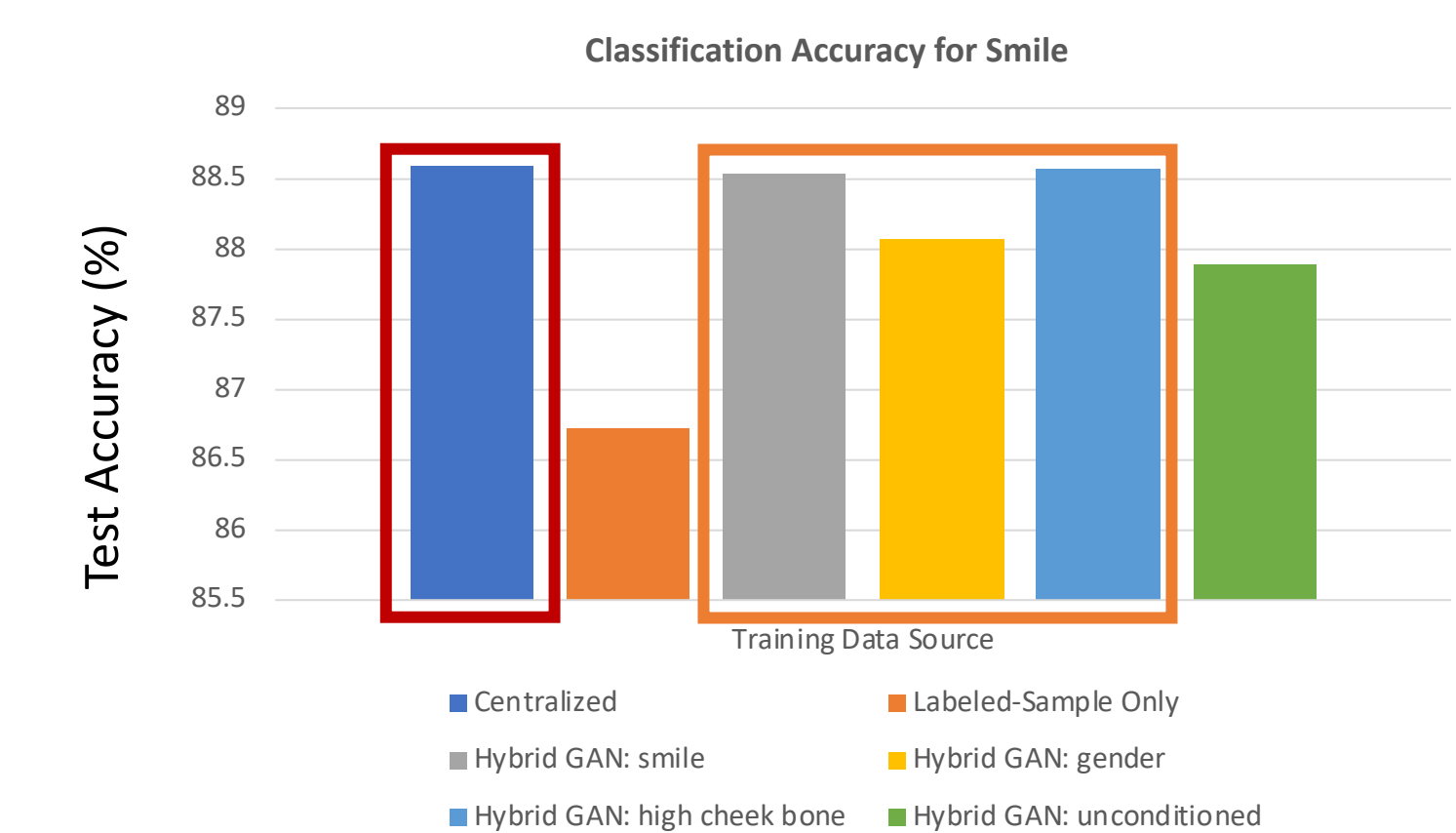
- Communication efficiency



- Scaling with multiple sites



- Learning multiple tasks: *Celeb-A*



Key Takeaways

- GAN models as parsimonious representation of local data
- Architecturally supports multi-task learning – eliminates need for multiple rounds of data collection
- Suitable for scenarios where each site has substantial amount of data (complementary to Federated Learning)