Clustered Federated Learning via Gradient-based Partitioning

Heasung Kim, Hyeji Kim, and Gustavo de Veciana

6G@UT · Wireless Networking and Communications Group · University of Texas at Austin

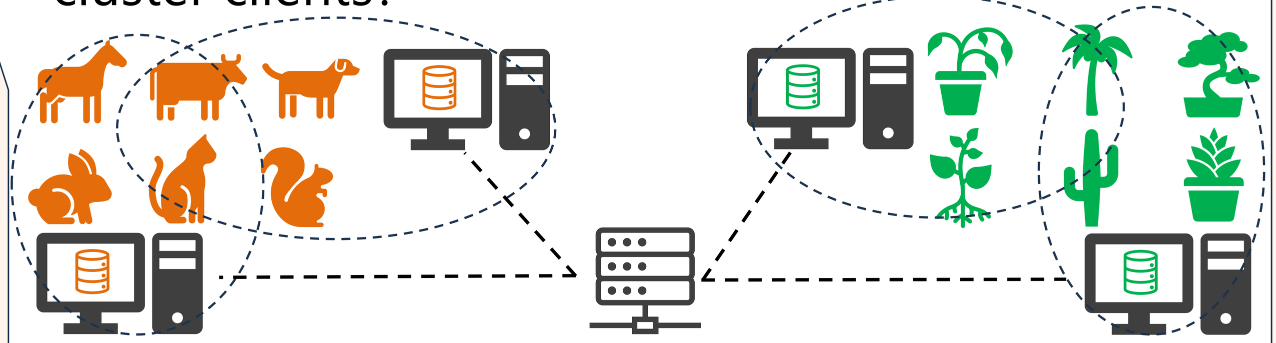
TEXAS The University of Texas at Austin

6G@UT WNCG interdigital

TL;DR

Q

CFL Problems: When training multiple models for diverse data distributions (clients), how do we cluster clients?



The Clients are connected to a central unit but no local dataset is shared.

CFL-GP (Ours)

① **Accumulate client gradients** over multiple learning iterations **for a set of models.**

② **Apply spectral clustering** to the accumulated gradient information

Q

What **advantages** does your approach offer over existing clustered federated learning algorithms?

CFL-GP (Ours)

① Rapidly identifies cluster identities, ***often within just a few rounds***.

② Highly robust (***both clustering/model convergence***) without dataset sharing

Theoretical results

Under assumptions (Gaussian SGN, convexity...),

① **Cluster convergence:** As $t \to \infty$, **all clients are correctly clustered** with high probability.

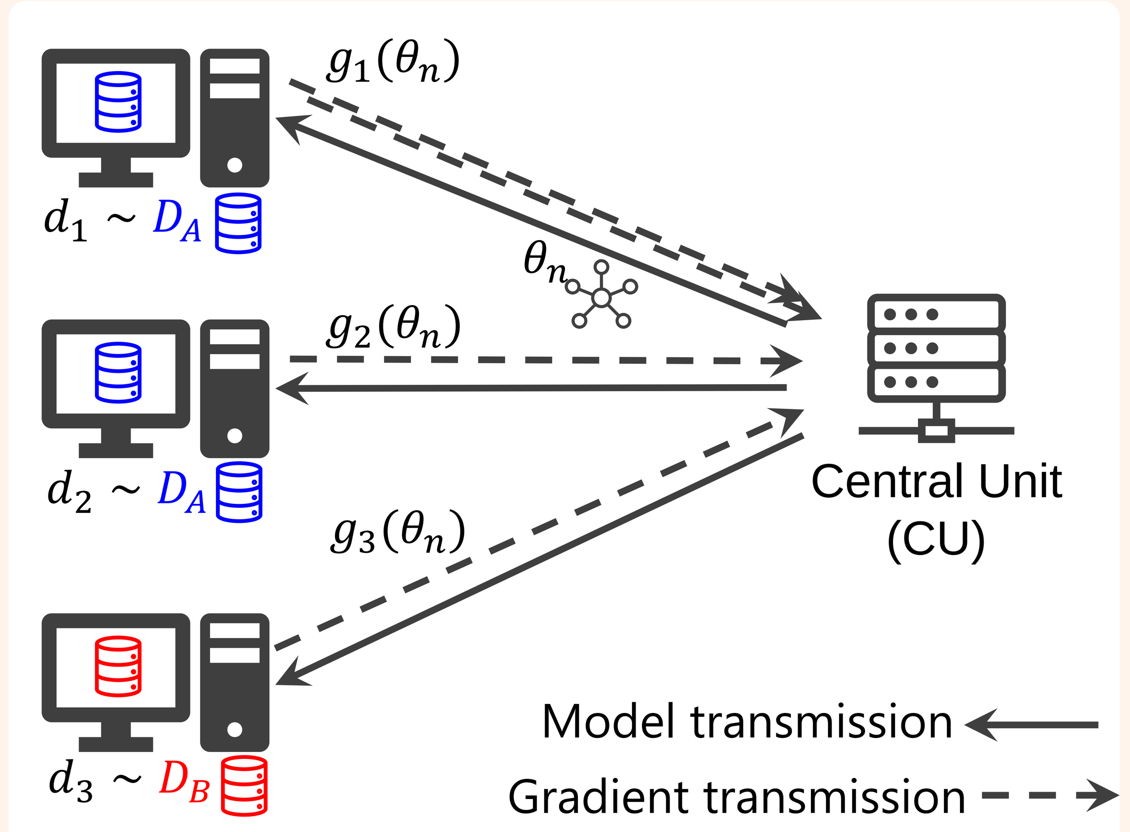
② **Contractive property:** As $t \to \infty$, the error rate of CFL-GP's gradient update converges to **the optimal error rate** with high probability.

③ **Model convergence**: As $t \to \infty$, all of models **converge to the optimal model** with high probability.

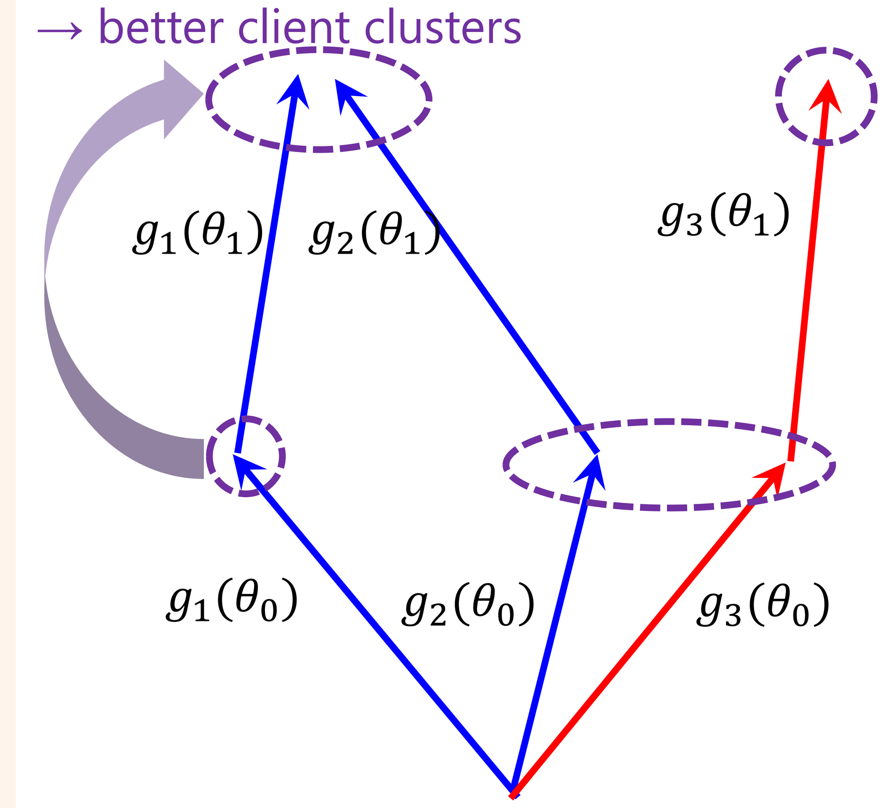
Key Findings / Algorithm (CFL-GP)

① **Accumulating clients' gradient information** for a set of models reveals their cluster identities by effectively **denoising gradient noise.**

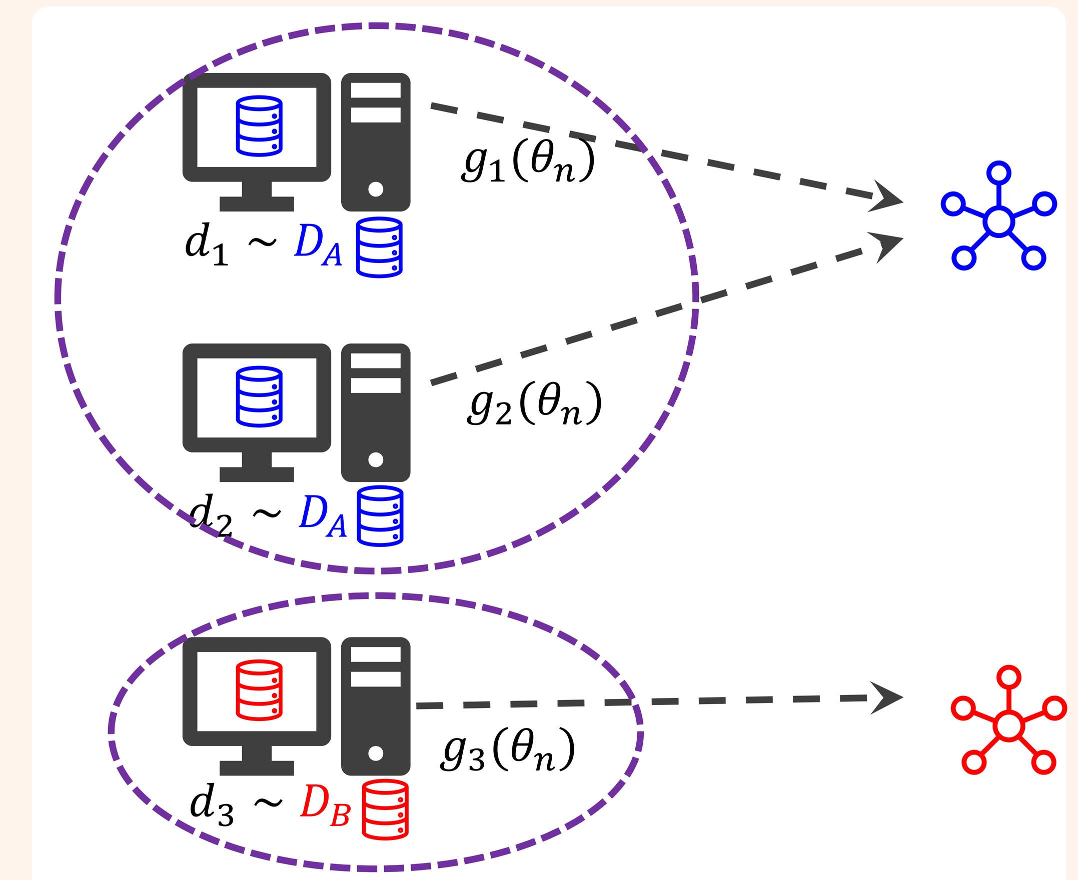
② **More accurate client clustering yields better task performance.**



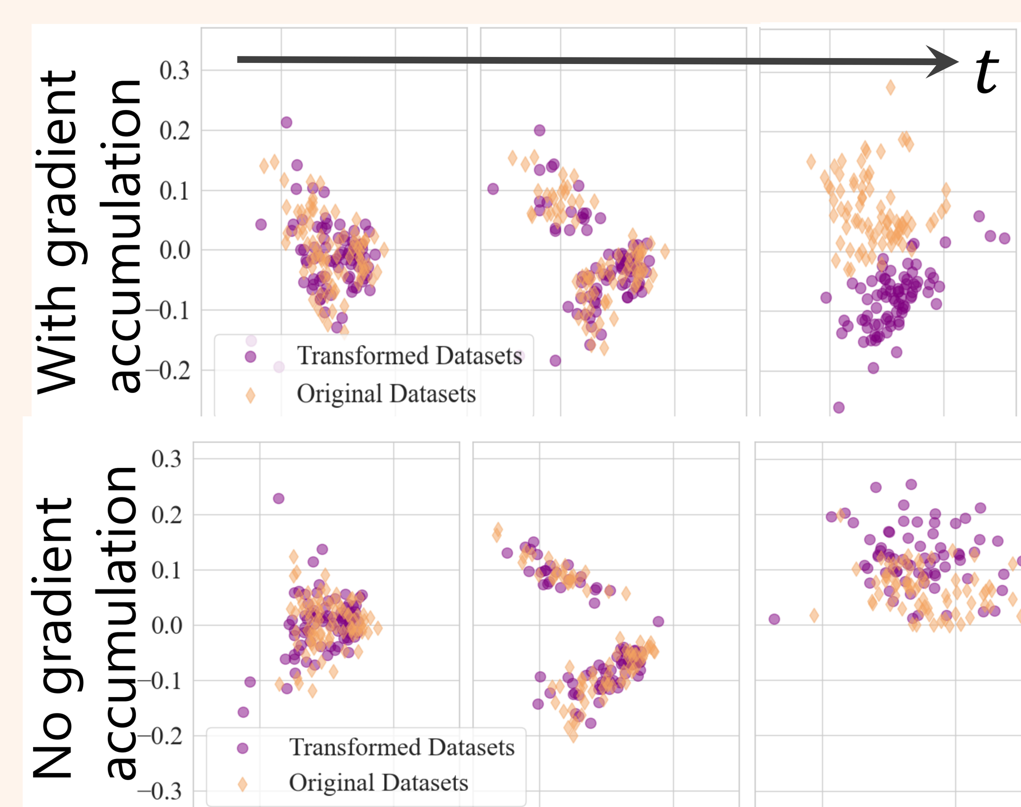
The CU broadcasts/transmits models and gather gradients.



Spectral clustering based on **aggregated gradients.**

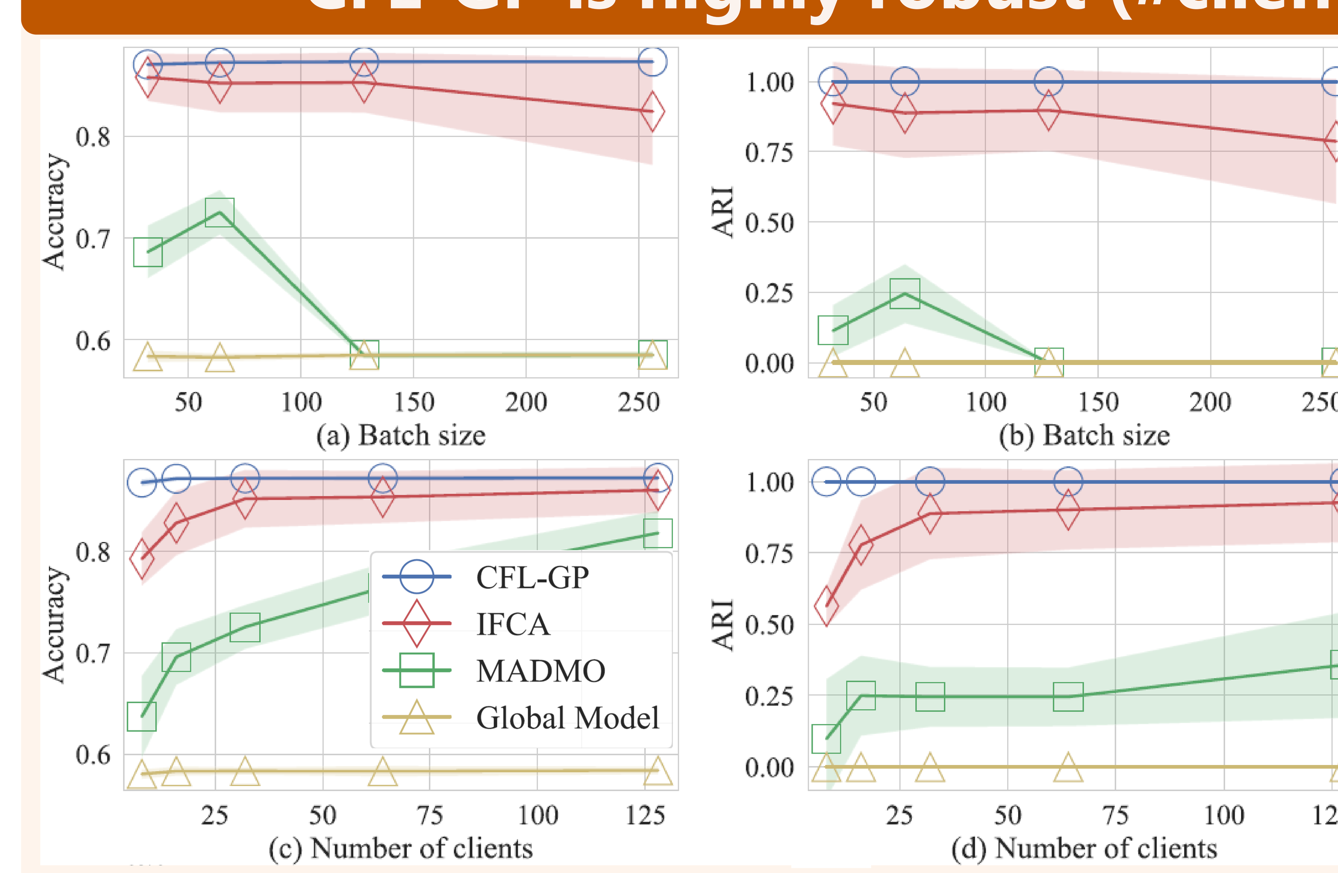


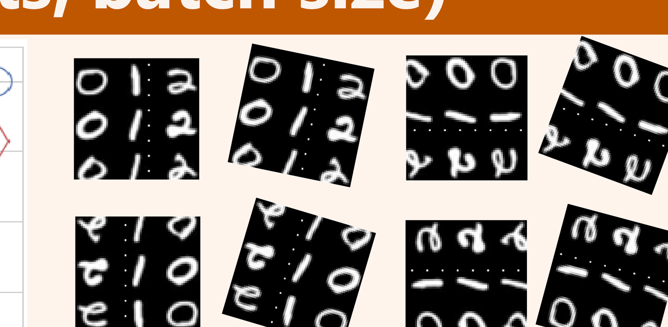
Update models based on updated grouping.



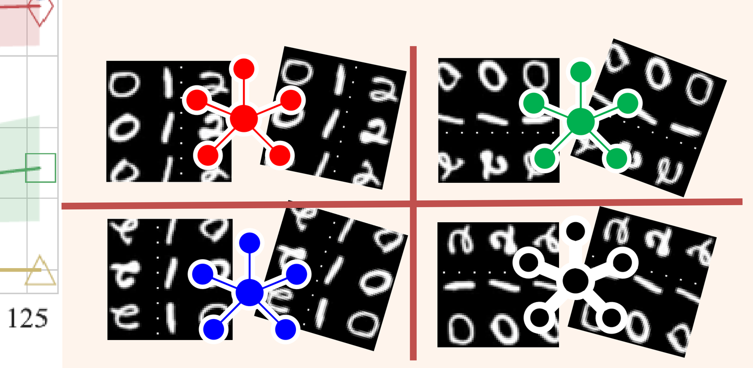
Client feature vs communication round t

CFL-GP is highly robust (#clients, batch size)

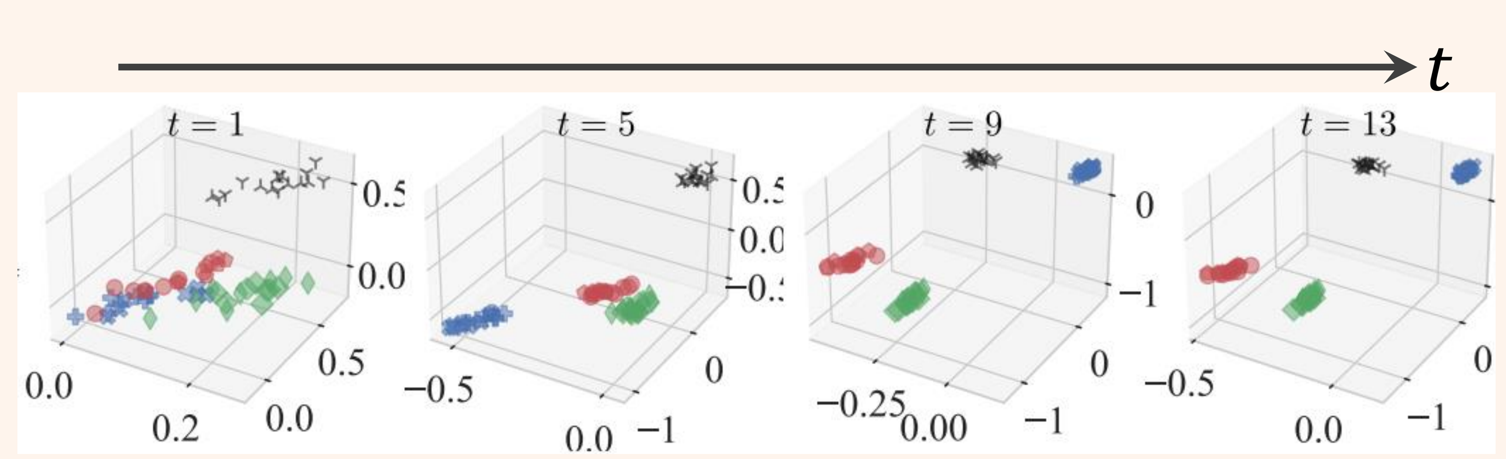




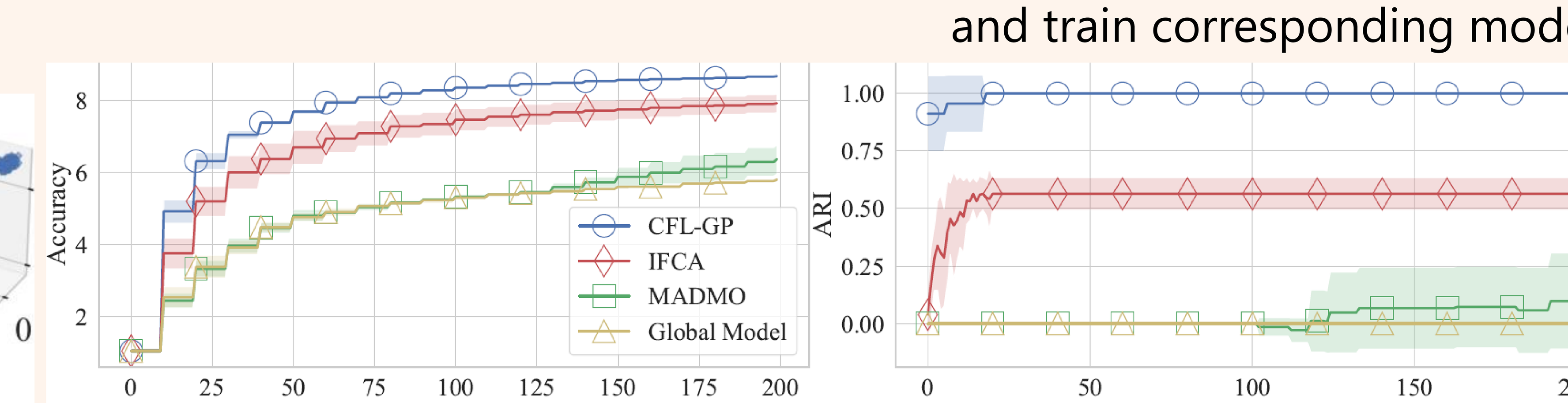
8~128 clients from 8 different data distributions



Goal: cluster them into 4 groups and train corresponding models



Client feature vs communication round t



Task Performance / Clustering Accuracy vs Communication round t