DP-TwoLevel: Two-Stage Gradient Subspace Learning for Differentially Private Federated Learning

Scientific Achievement

- Differential privacy (DP) in federated learning (FL) often reduces accuracy, especially for high-dimensional models.
- A paper featuring ORNL researchers introduces DP-TwoLevel, which learns a two-stage principal component analysis subspace of gradients to represent updates in fewer dimensions.
- DP-TwoLevel adds privacy noise in the low-dimensional representation and achieves higher accuracy than DP-FedAvg.

Significance and Impact

- The method improves the privacy–utility tradeoff most in strong-privacy settings where standard DP-FL degrades the most.
- It helps enable collaborative learning on sensitive scientific and institutional datasets that cannot be centrally shared, while also reducing communication burden.

Technical Approach

- The approach builds a shared gradient basis during a short warmup phase and splits it into coarse and fine components to capture dominant and residual structure.
- Each client projects gradients, clips the coefficients, and adds Gaussian noise while the server reconstructs and aggregates the updates without changing the DP guarantee.
- The study identifies a variance-retention threshold that predicts success and motivates per-layer projections to scale beyond the limits of a single global basis.

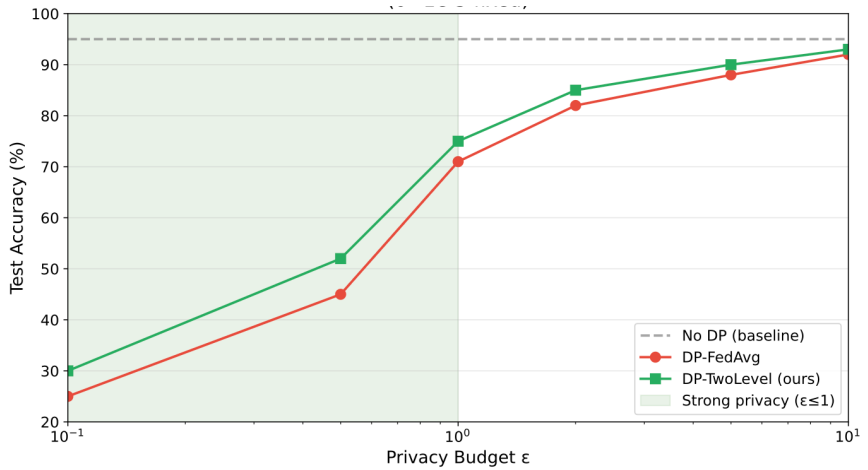

Figure 1. Privacy-utility relationship: Improvement magnitude over privacy budget. Lower ε (stronger privacy) yields larger gains due to noise reduction benefit.

PI(s)/Facility Lead(s): K. Kim (PI), O. Kotevska (ORNL's PI)
ASCR Program: AI for Science
ASCR PM: Xujing Davis
Publication(s) for this work: Kotevska, O., Patton, P., Jha, S., Balaprakash, P. (2026). DP-TwoLevel: Two-Stage Gradient Subspace Learning for Differentially Private Federated Learning. In *Assurance and Security for AI-enabled Systems. SPIE.*