Towards Scalable Private Collaborative Learning

PIs: Nikos Triandopoulos, Stevens Institute of Technology, Hoboken, NJ

Co-PI: Alina Oprea, Northeastern University, Boston, MA

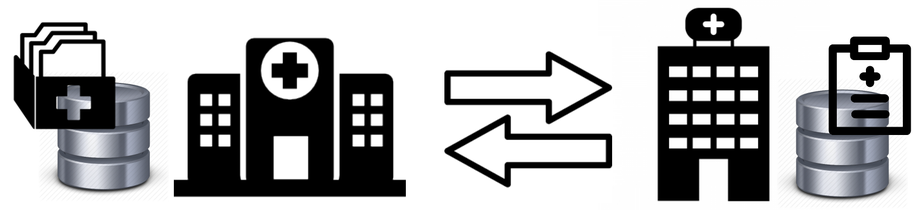
Collaborators: Samaneh Berenjian (Stevens), Xianrui Meng (Amazon), Dimitrios Papadopoulos (HKUST)

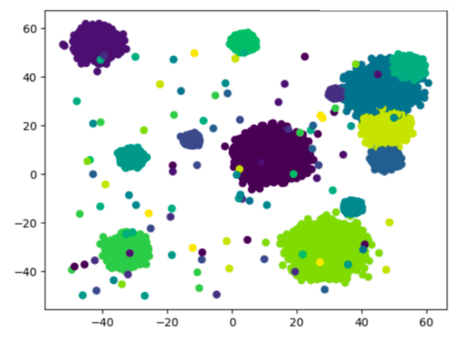
Private 2-Party Cluster Analysis: Formal Specification & Scalable Implementation

Collaborative learning allows entities to jointly deduce global ML models over their union dataset, but privacy risks often limit entities to individually learn local models using solely their own sensitive datasets.

Classification in heath-care and security analytics suffers from this natural Accuracy Vs. Privacy dichotomy.

Cluster analysis on medical data allows discovery of correlations that can improve health practices



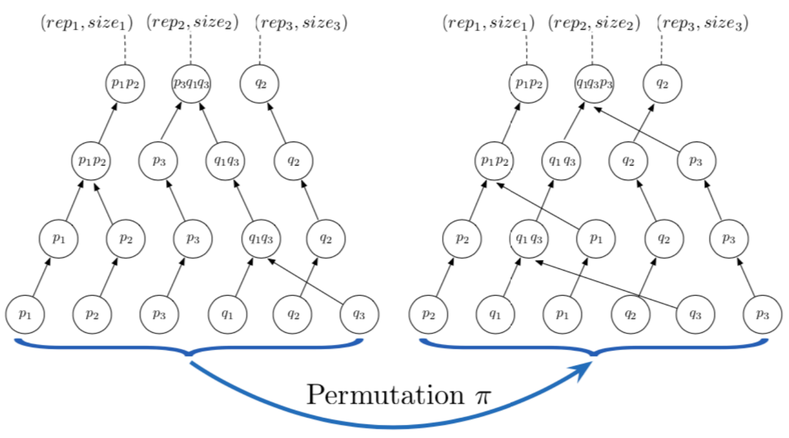


Main Goals/Challenges

1. Cast hierarchical clustering into a secure MPC instance
2. Design 2-party private protocol for joint cluster analysis
3. Devise secure implementations with practical efficiency

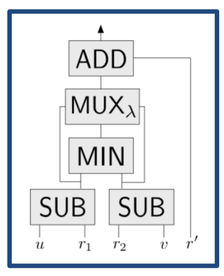
Key Ideas

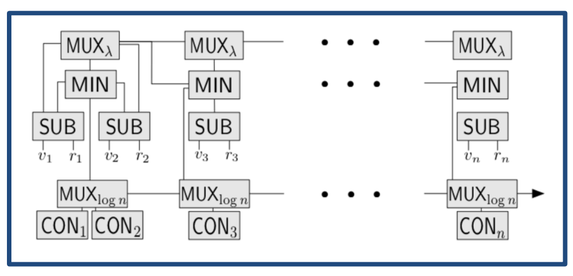
1. Use point-agnostic dendrograms



2. Employ "mixed" crypto protocols

PHE over Additive Secret Sharing





3. Integrate approximate clustering methods

CURE **approximate clustering**

Input: $\mathcal{D}$, parameters **Output:** Clusters $\mathcal{C}$ over $\mathcal{D}$

[**Sampling**] Randomly sample $\mathcal{D}$ into $\mathcal{S}$

[**Clustering A**] Cluster $\mathcal{S}$ & eliminate outlier clusters

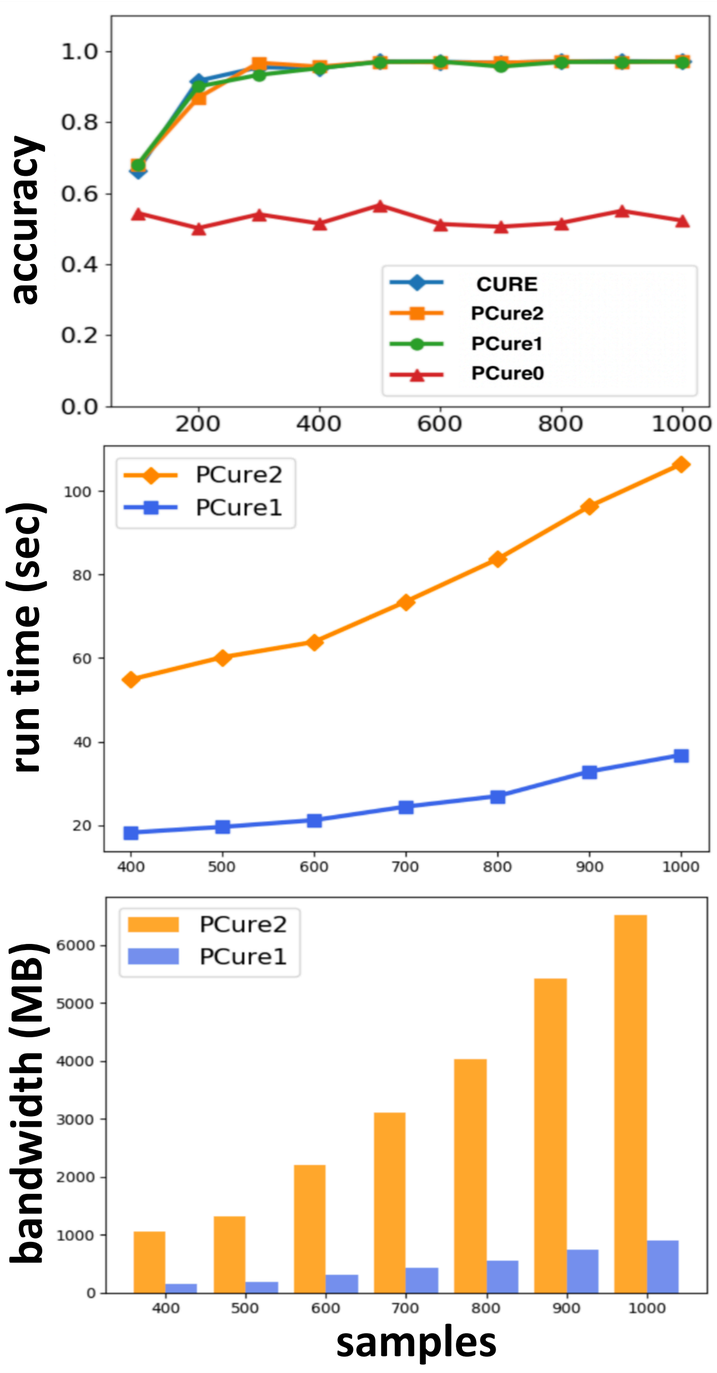
[**Clustering B**] Cluster A-clusters & eliminate outlier clusters

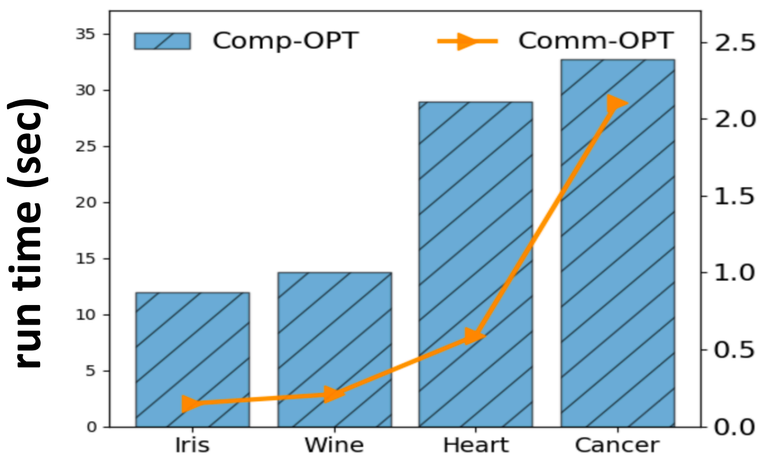
[**Classification**] Cluster singletons in $\mathcal{D}$ into closest B-clusters

Results

New design framework for private approximate clustering protocols

- provable security: strong privacy guarantees on parties' input
- flexibility: variety of schemes that balance accuracy with efficiency
- scalability: cluster analysis of 1M 10-dim records runs in 35sec, transferring only 896KB and achieving 97.09% accuracy





Impacts

Scientific: First complete study of secure cluster analysis, featuring formal specification and a variety of scalable implementations

Societal: Rendering benefits of data science available to anyone via safe joint analysis of high-volume and richly variate data, towards the vision of AI-assisted collaborations (e.g., community- and intelligence-based clinical trials and cyberattack defenses).

Future research: Explore further the merits of approximation in private collaborative learning

Outreach: Work currently under submission, available at https://arxiv.org/abs/1904.04475

Award ID#: 1718782