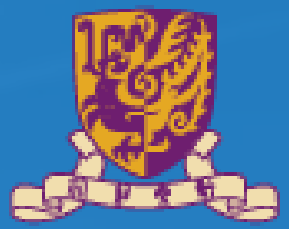
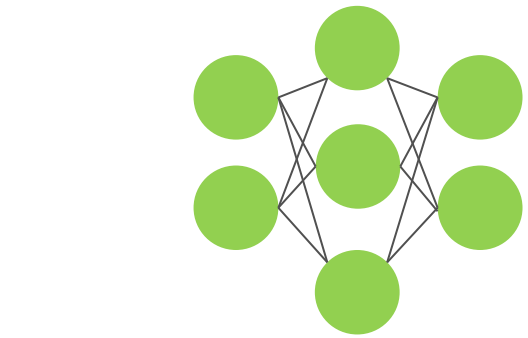
CaPC Learning: Confidential & Private Collaborative Learning

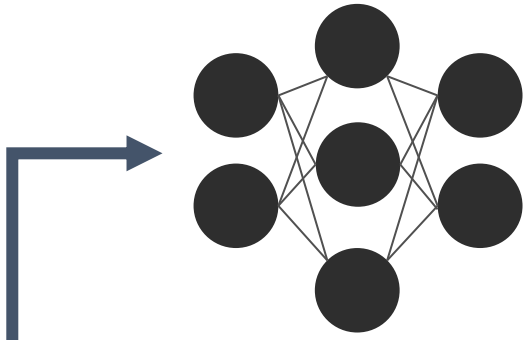
Christopher A. Choquette-Choo, Natalie Dullerud, Adam Dziedzic, Yunxiang Zhang, Somesh Jha, Nicolas Papernot, Xiao Wang

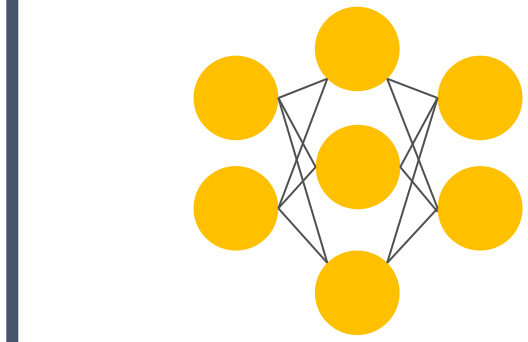

UNIVERSITY OF TORONTO
VECTOR INSTITUTE


WISCONSIN
UNIVERSITY OF WISCONSIN-MADISON
Northwestern University

Private Consultation

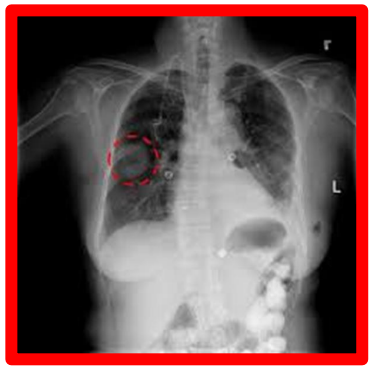


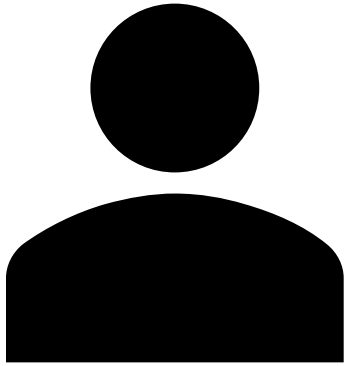


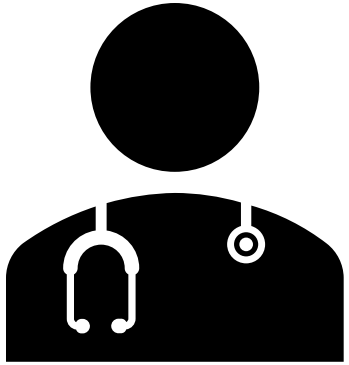


Aggregation

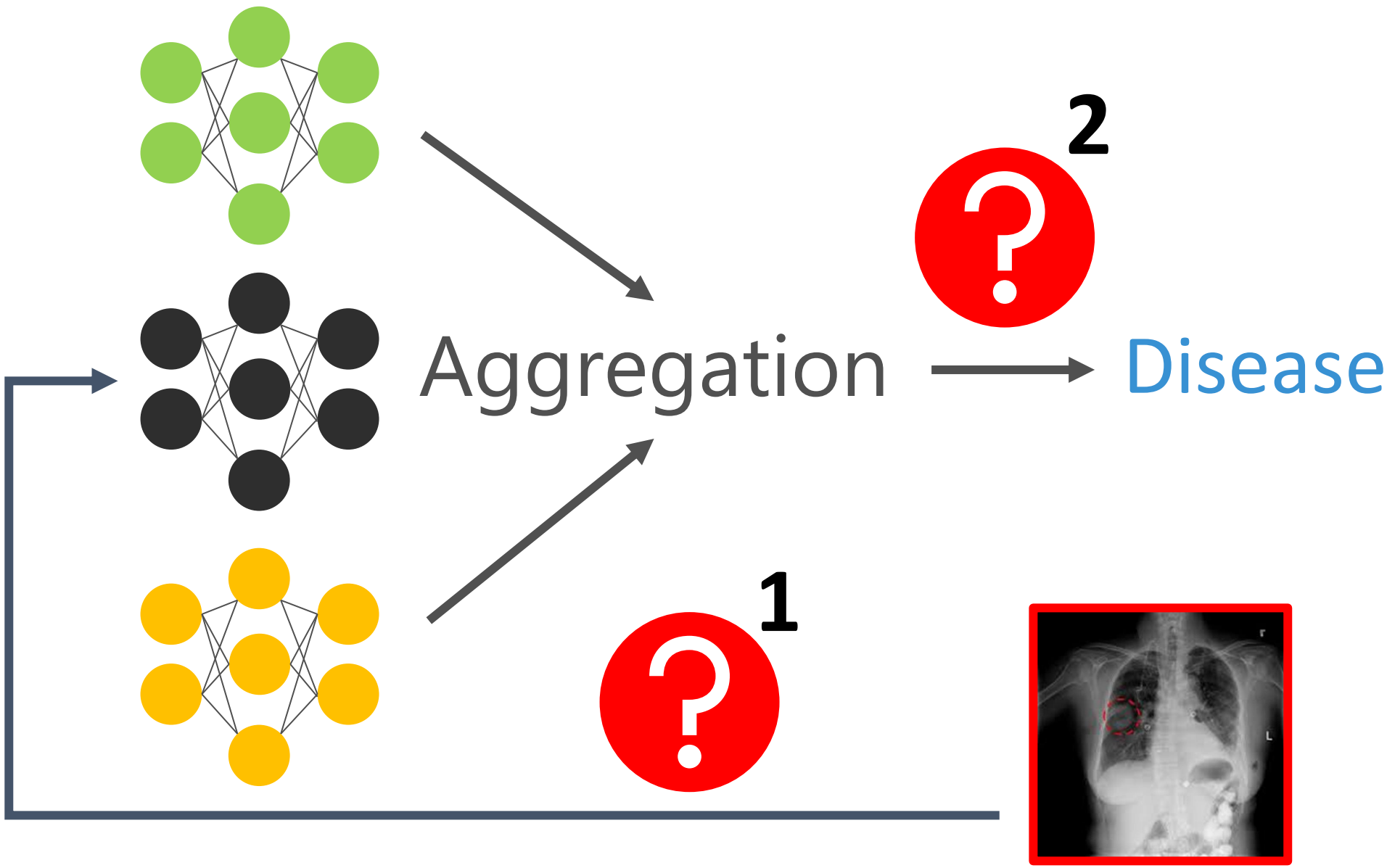
Disease







How to Protect Confidentiality and Privacy?



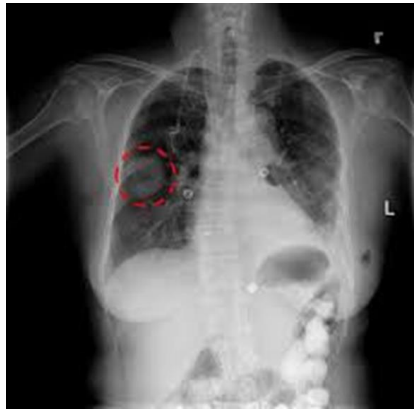
Requirements for CaPC

Requirement	What do we do?
Privacy of training data	Guarantee protection of personally identifiable information contained in training data via Differential Privacy.
Query confidentiality	Encrypt input data and do inference on encrypted data using Homomorphic Encryption and Secure Multi-Party Computation.
Model confidentiality	Prevent leakage of the answering parties' models to the querying party.

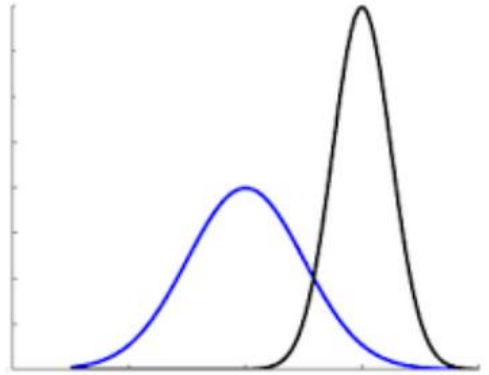
Use CaPC in Healthcare

- Hospitals act as collaborating parties.
- Protect privacy & confidentiality of patients' data.
- Using collaborative learning setup to investigate and possibly address some of the issues in healthcare.

Strong Privacy Guarantees

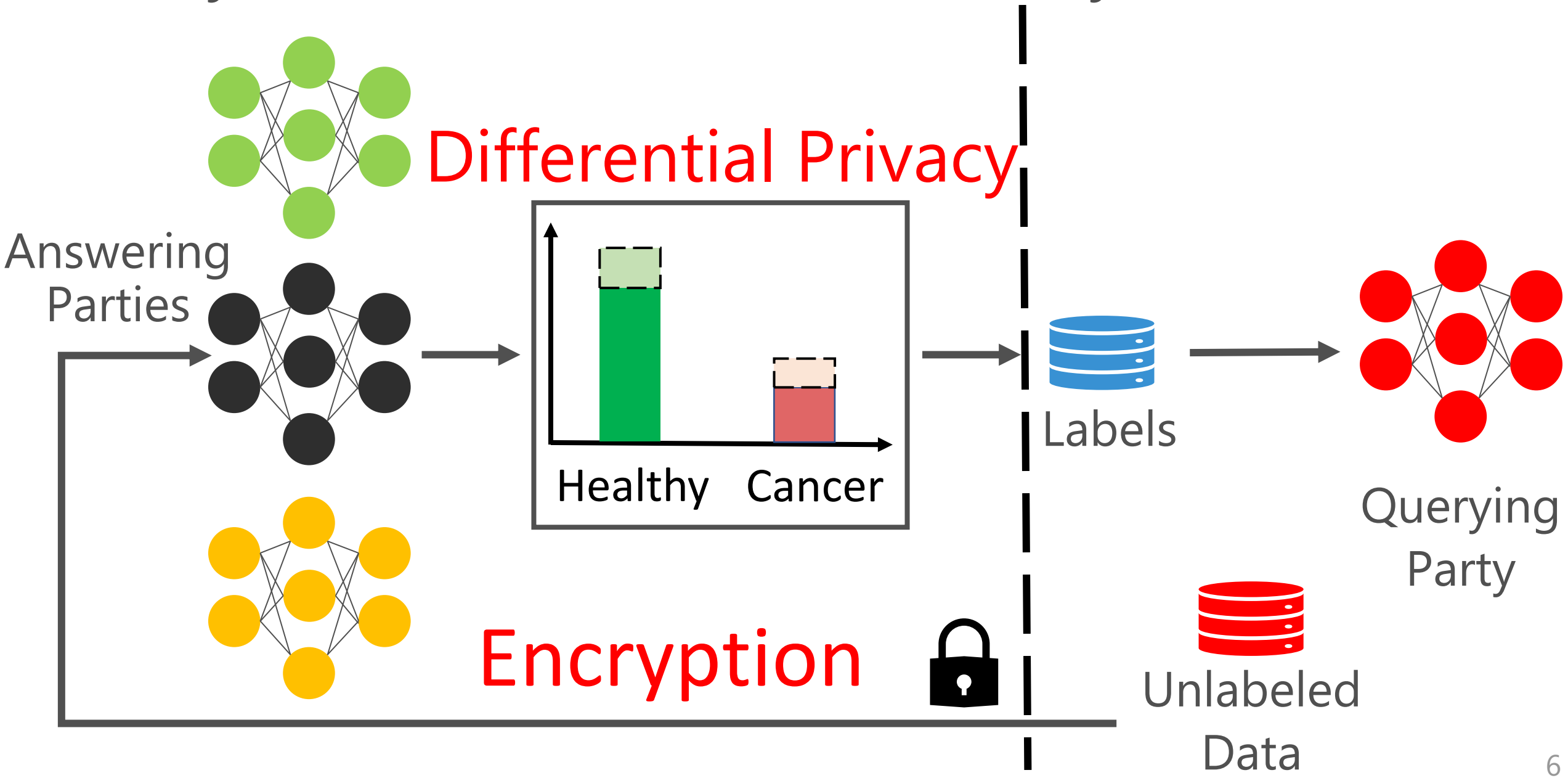


Private Consultation

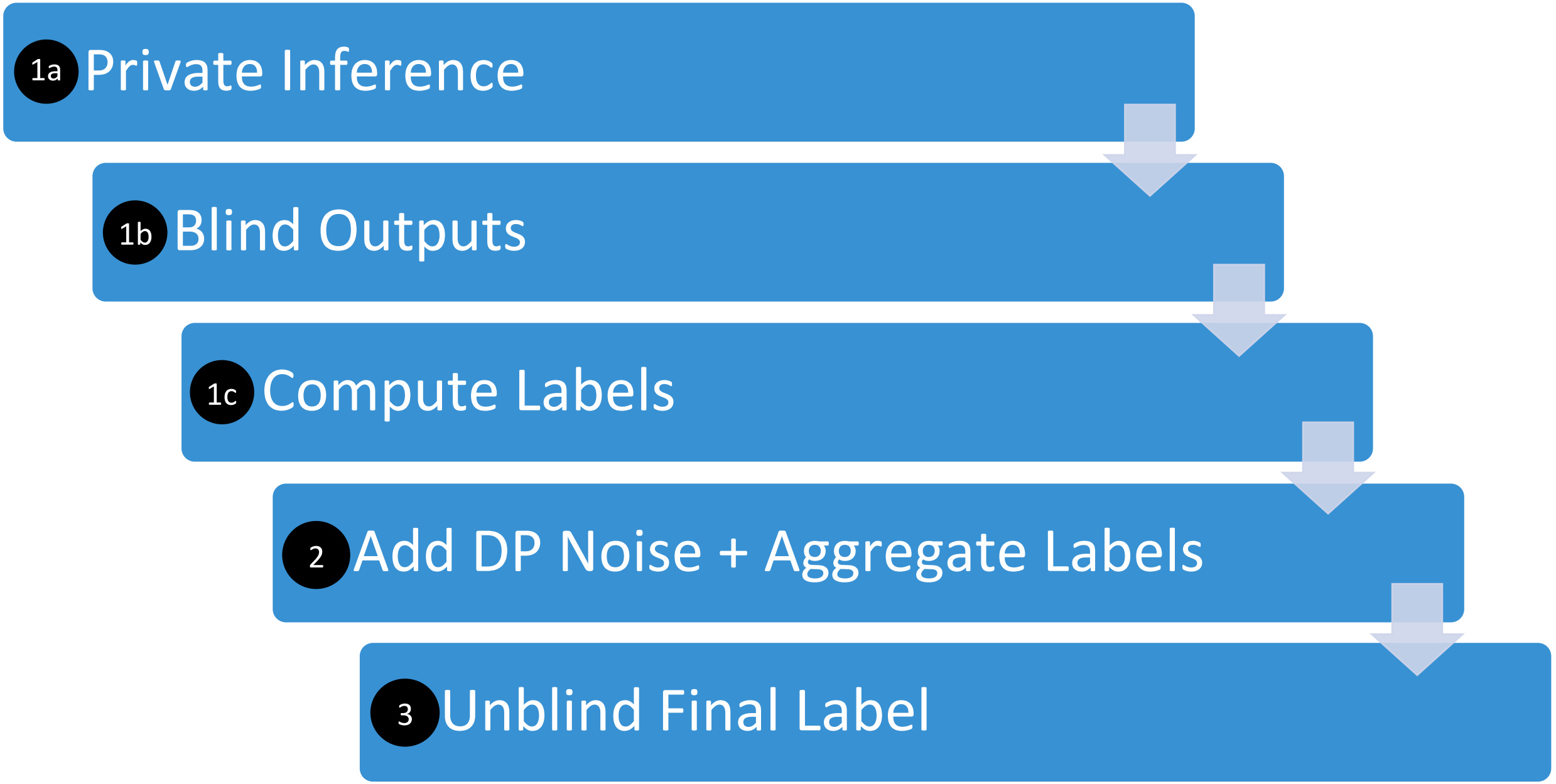


Robustness to Distribution Shift

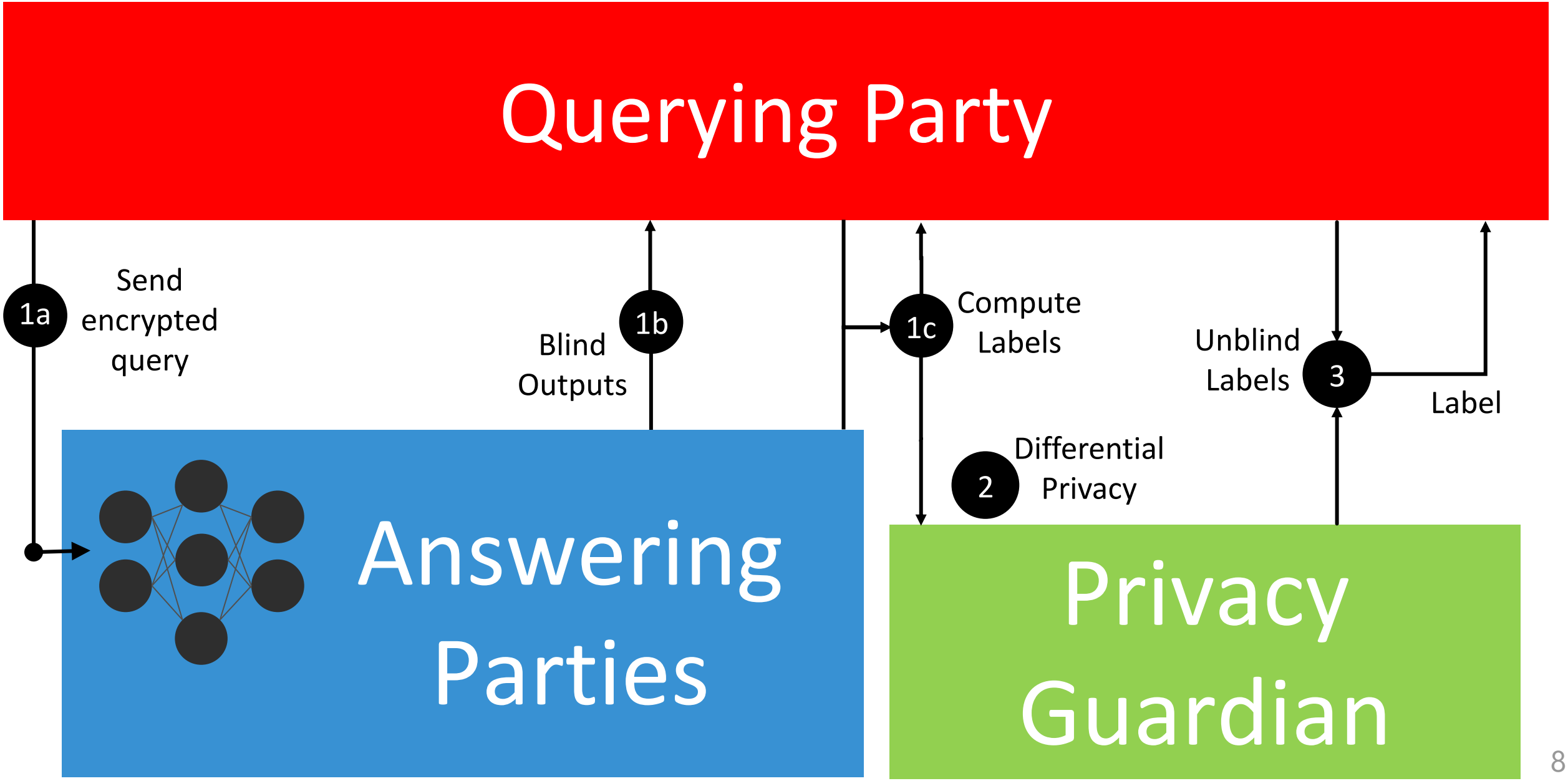
Privacy of Train & Confidentiality of Test Data



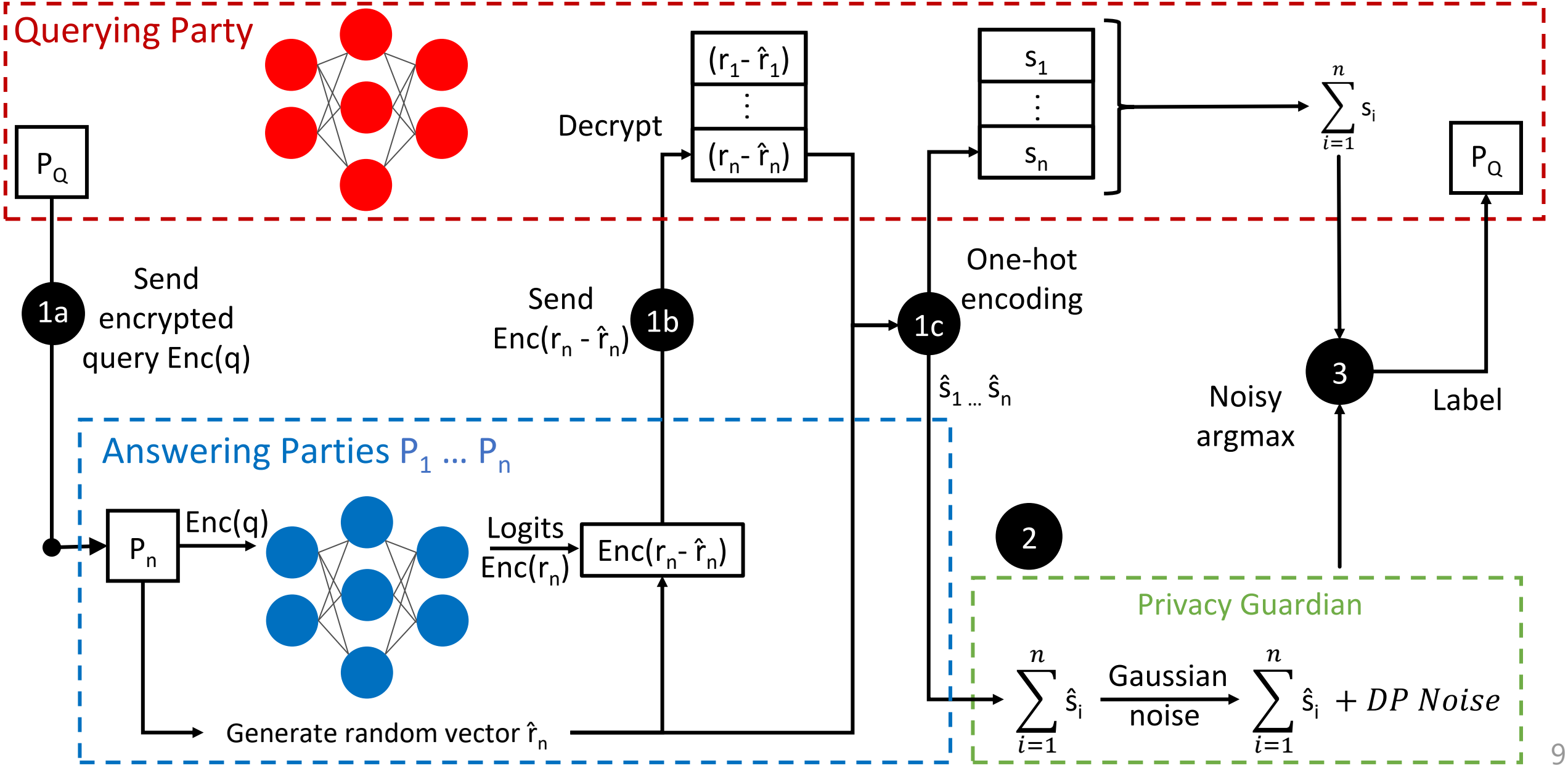
CaPC Workflow



Actors in CaPC

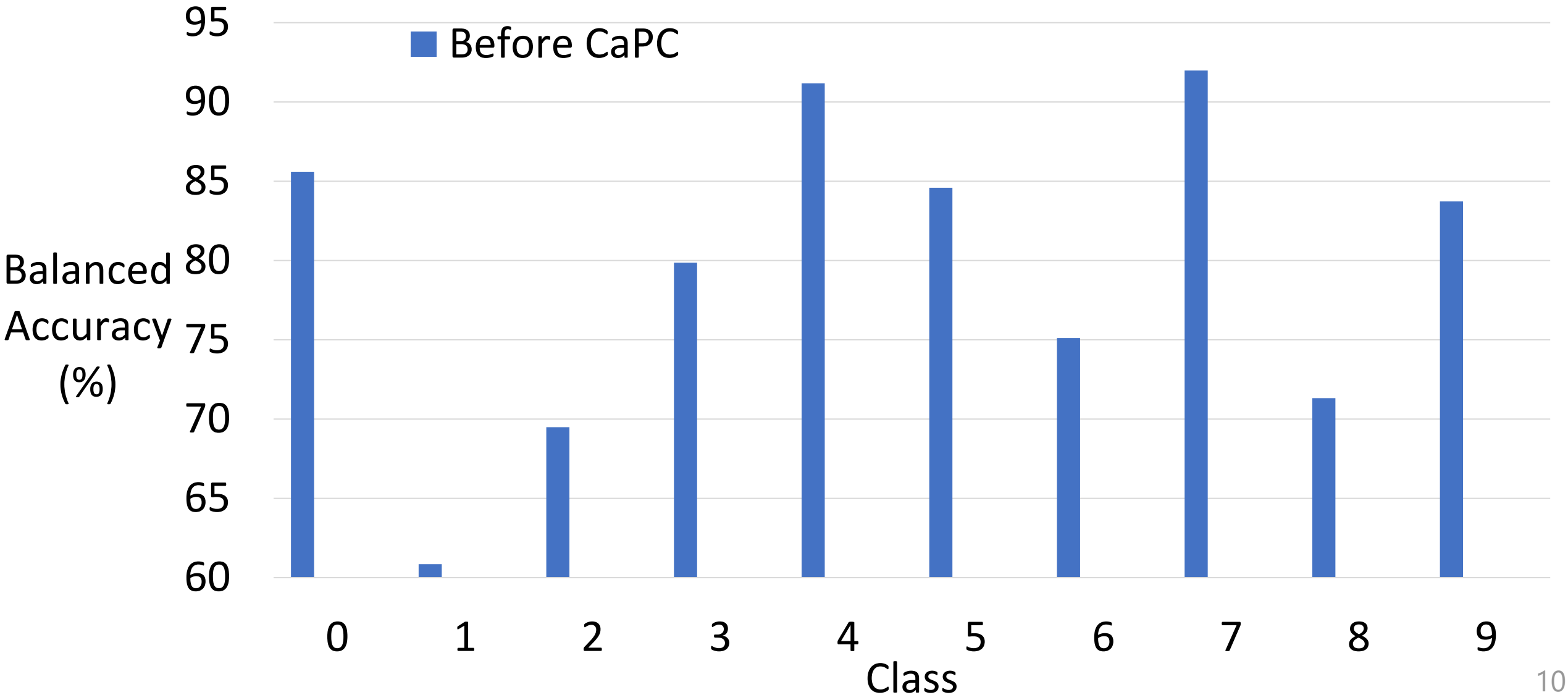


CaPC Protocol



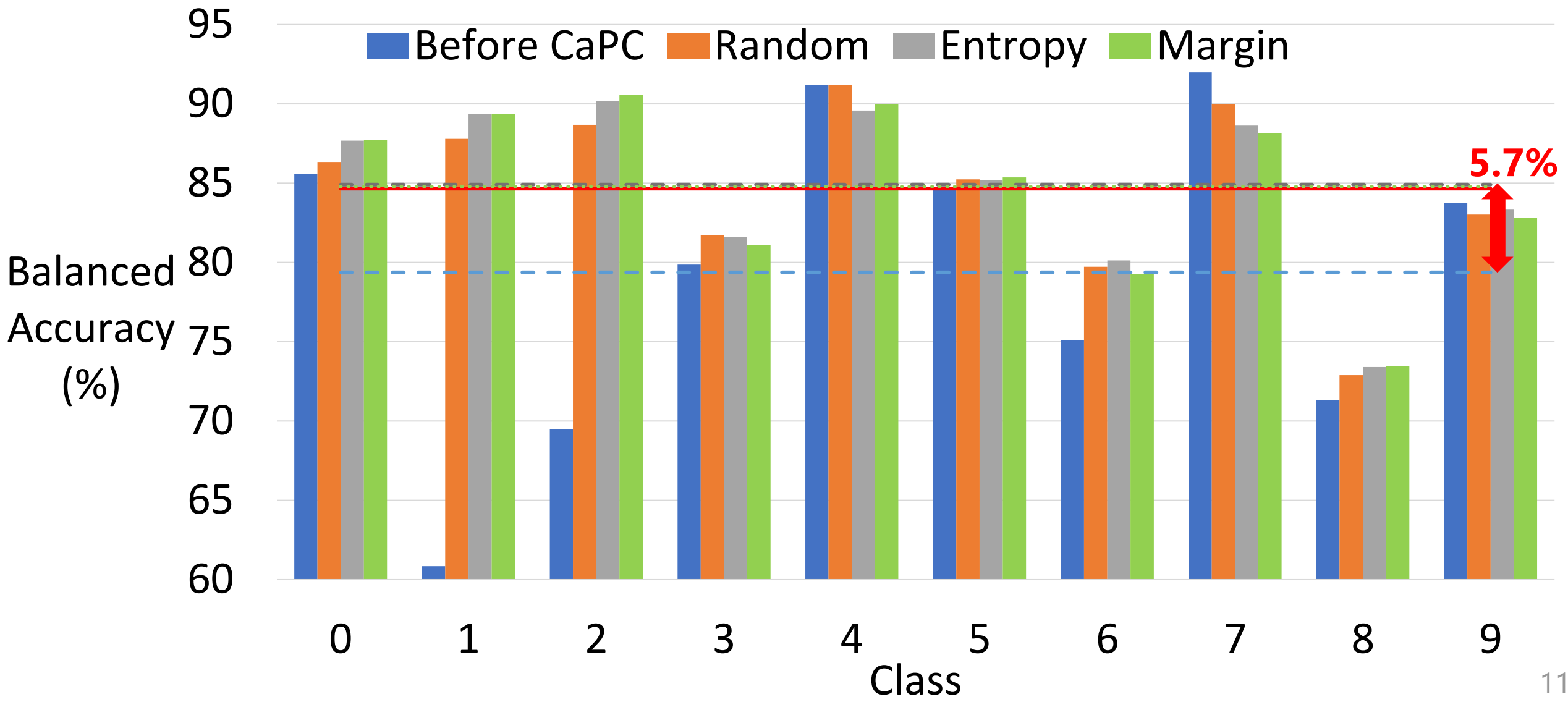
Non-uniform Data Distribution

SVHN on VGG-7, 250 parties, $\epsilon = 2.0$, $\delta = 10^{-6}$, ***classes 1 & 2 are under-represented***



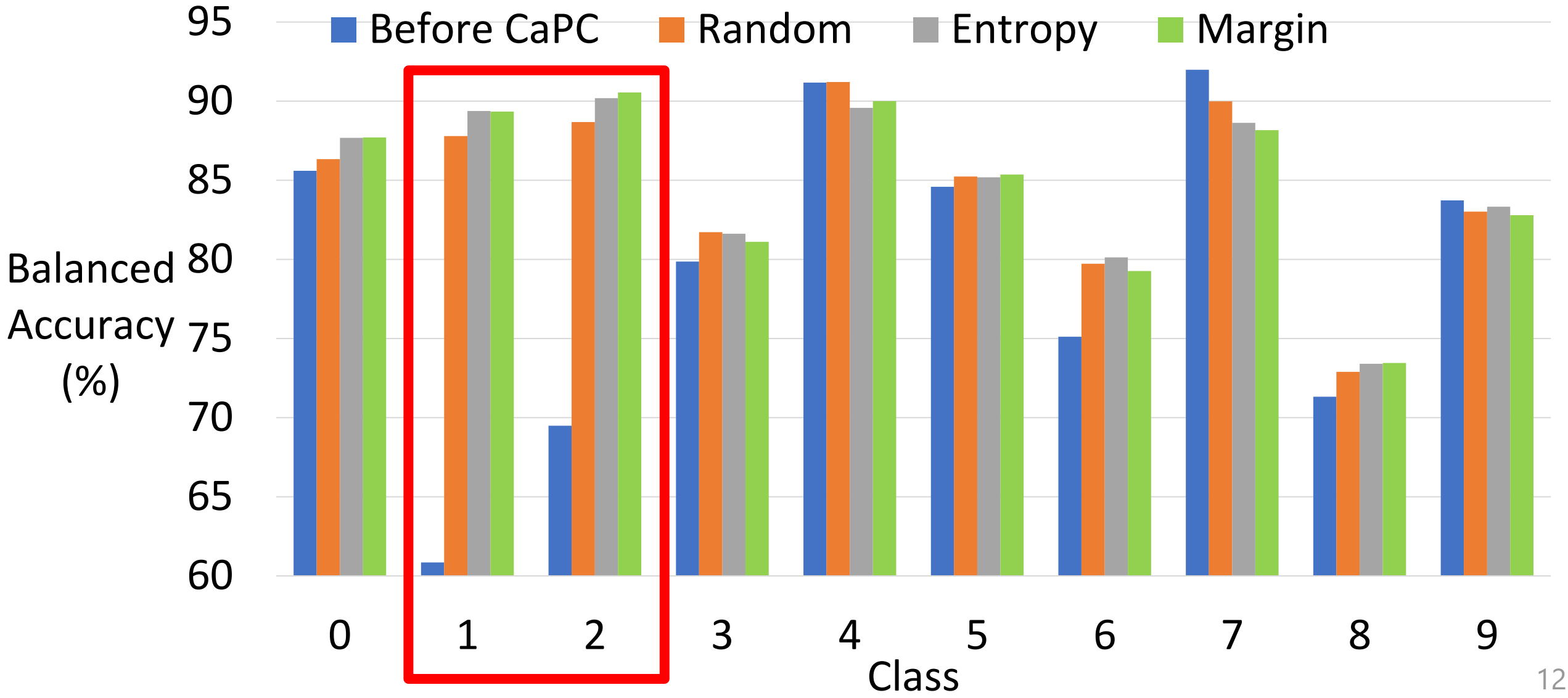
Non-uniform Data Distribution

SVHN on VGG-7, 250 parties, $\epsilon = 2.0$, $\delta = 10^{-6}$, ***classes 1 & 2 are under-represented***



Non-uniform Data Distribution

SVHN on VGG-7, 250 parties, $\epsilon = 2.0$, $\delta = 10^{-6}$, ***classes 1 & 2 are under-represented***



Conclusions

- CaPC protocol for privacy preserving collaboration and learning.
- Privacy of train data with differential privacy & Pâté.
- Confidentiality of test data via secure multi-party computation and homomorphic encryption.
- Participants label their new data items and use them to improve their own ML models.
- CaPC improves performance of models with heterogenous architectures and when there is skew in data.

Thank you

https://arxiv.org/abs/2102.05188