CaPC Learning: Confidential & Private Collaborative Learning

Adam Dziedzic
Postdoctoral Fellow
ady@vectorinstitute.ai

cleverhans

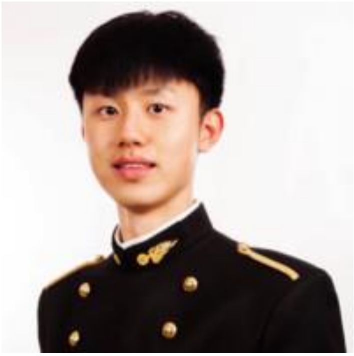
VECTOR INSTITUTE

UNIVERSITY OF TORONTO

Collaborators

Christopher A. Choquette-Choo

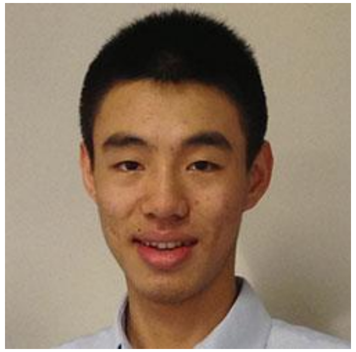
Natalie Dullerud



Yunxiang Zhang

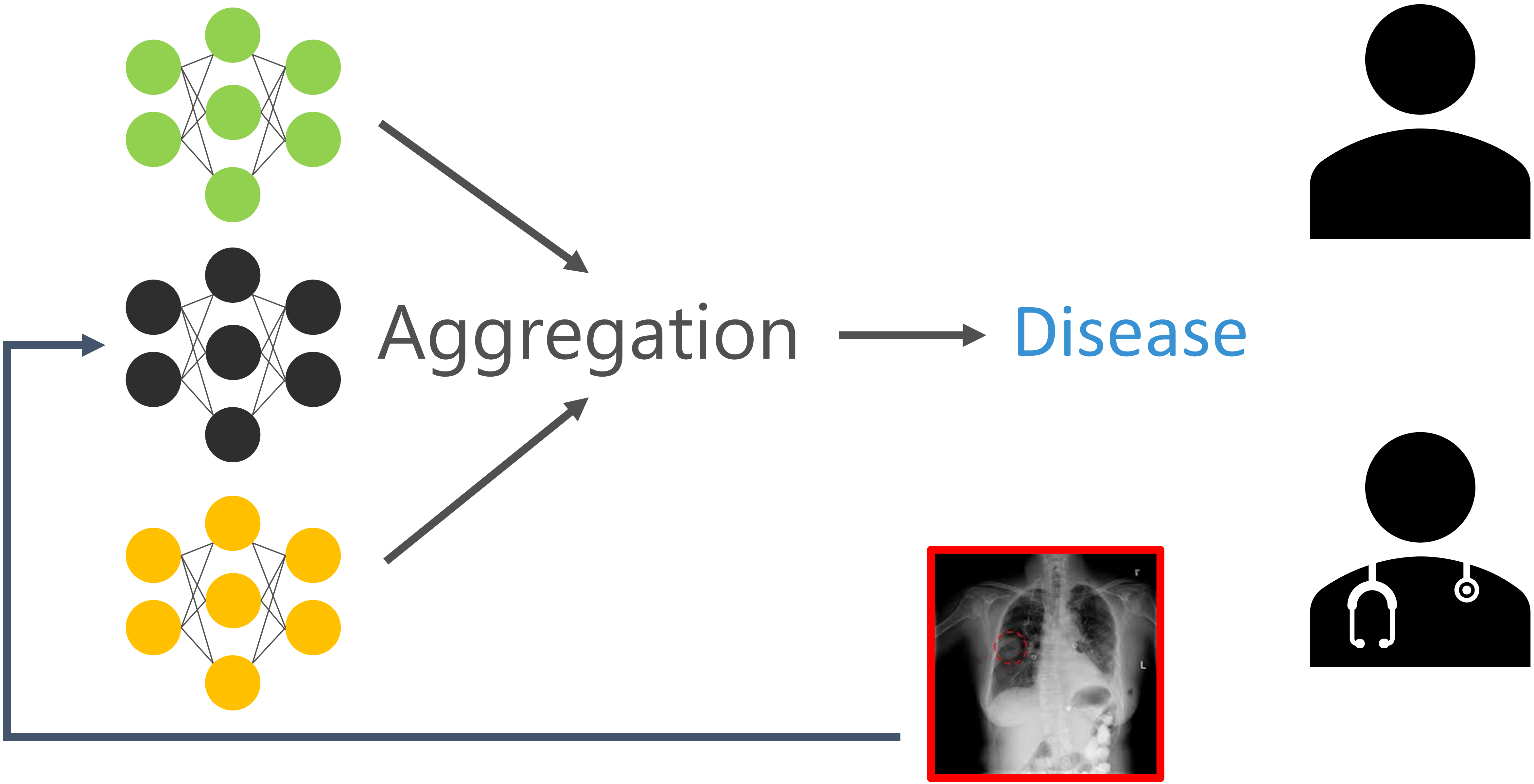
Somesh Jha

Nicolas Papernot

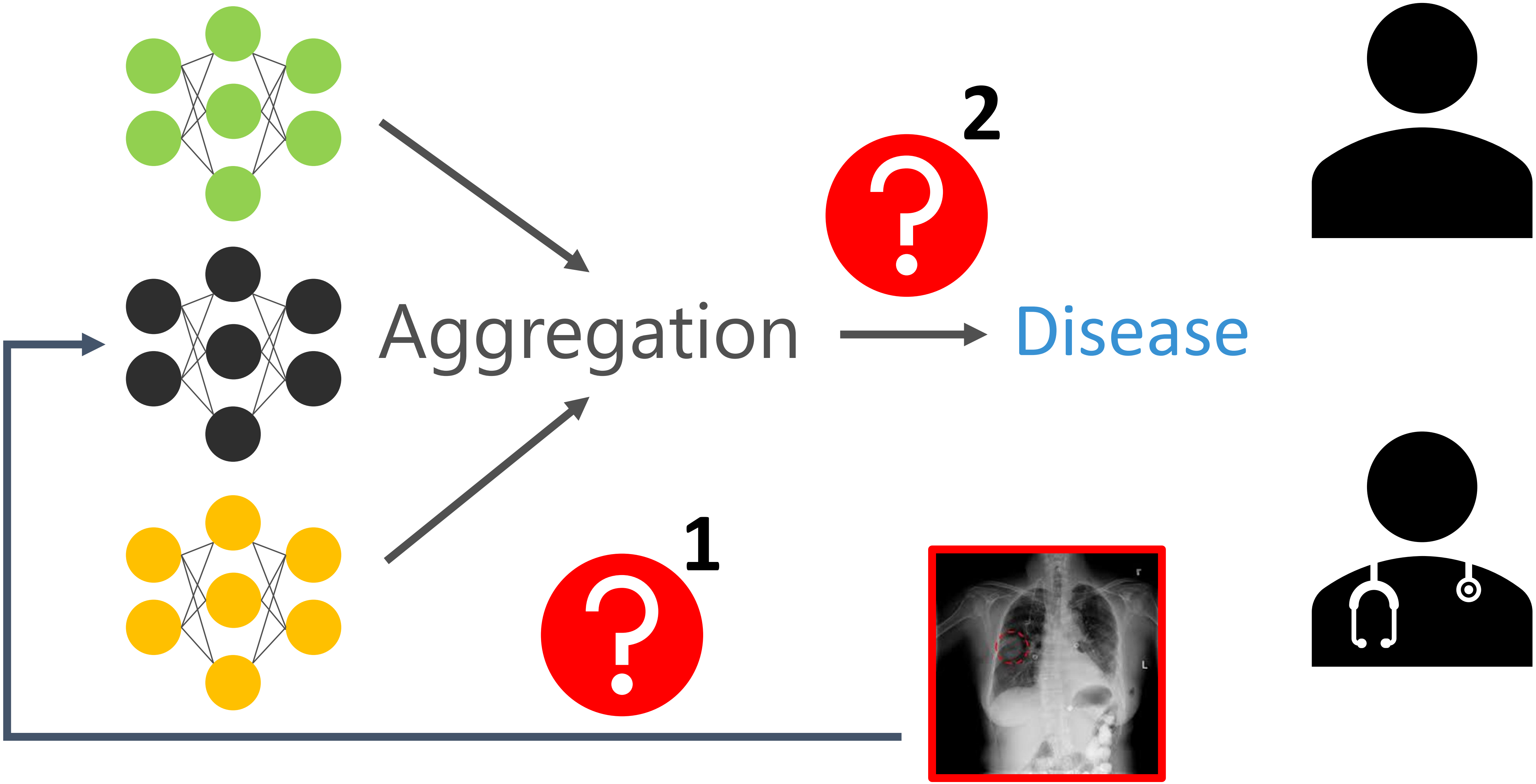


Xiao Wang

Private Consultation



How to Protect Confidentiality and Privacy?



1. Requirements & Overview

2. CaPC Protocol

3. Empirical Evaluation

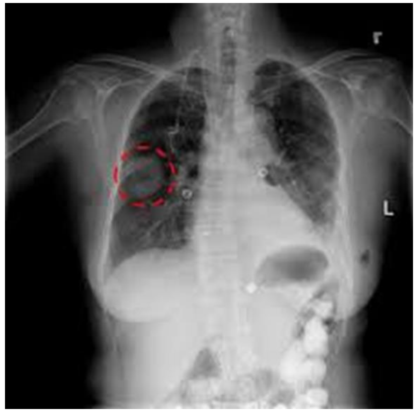
Requirements for CaPC

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.

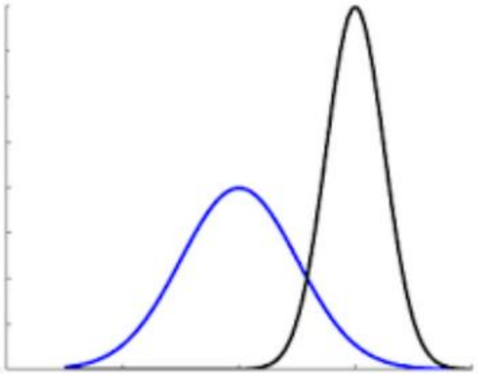
Use CaPC in Healthcare

- 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

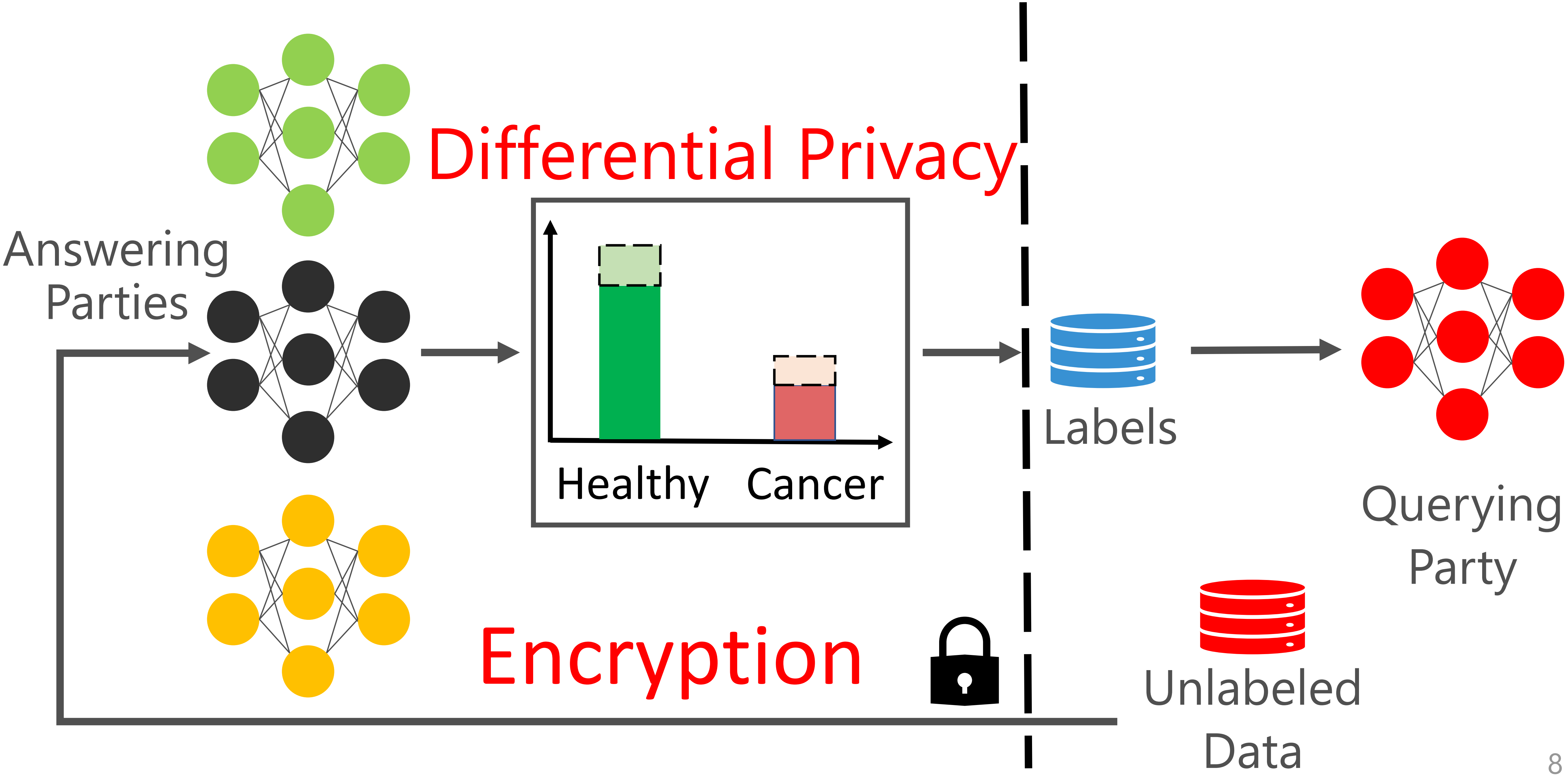


Private Consultation



Robustness to Distribution Shift

Privacy of Train & Confidentiality of Test Data



CaPC Workflow

1a Private Inference

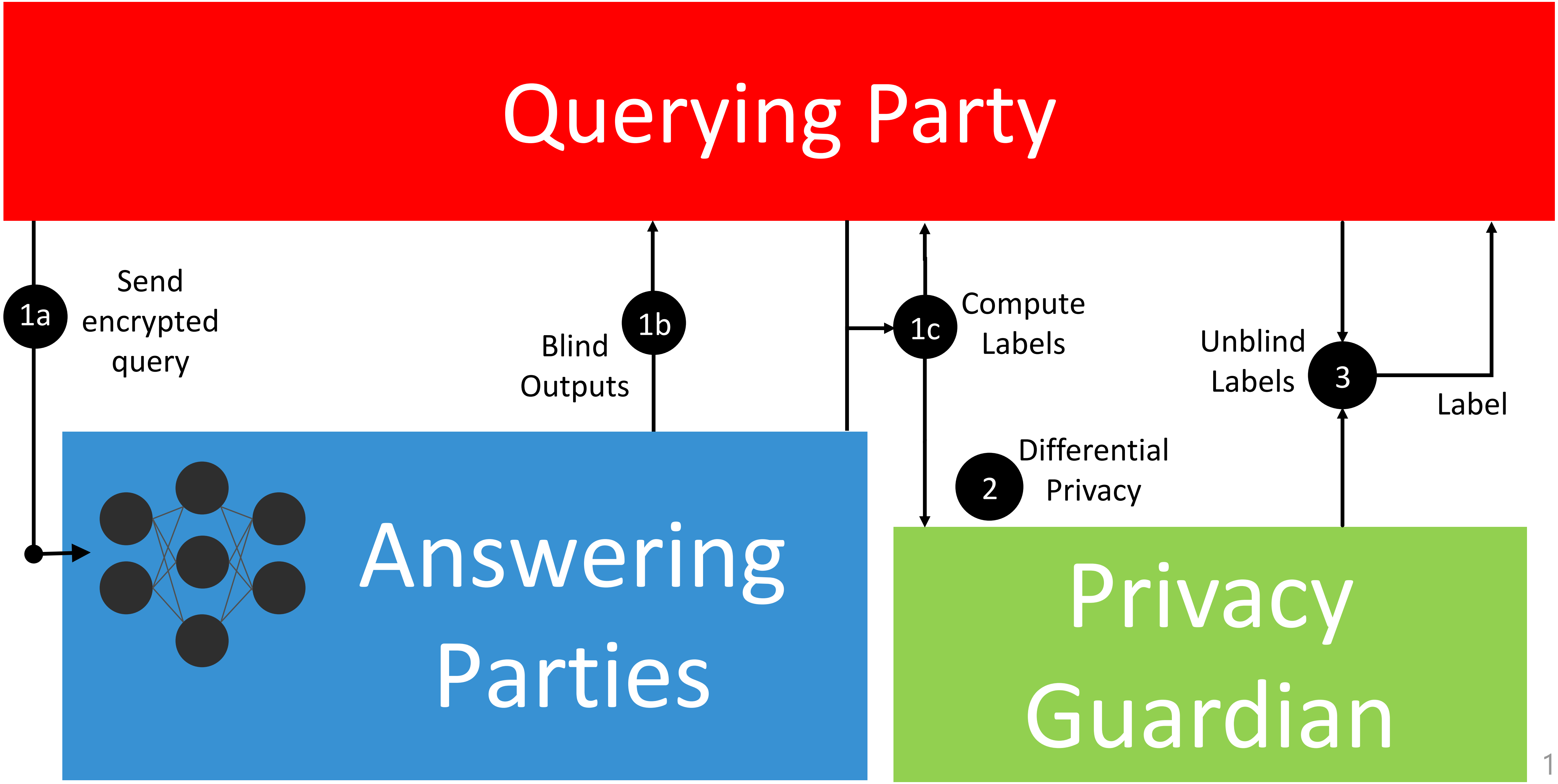
1b Blind Outputs

1c Compute Labels

2 Add DP Noise + Aggregate Labels

3 Unblind Final Label

Actors in CaPC

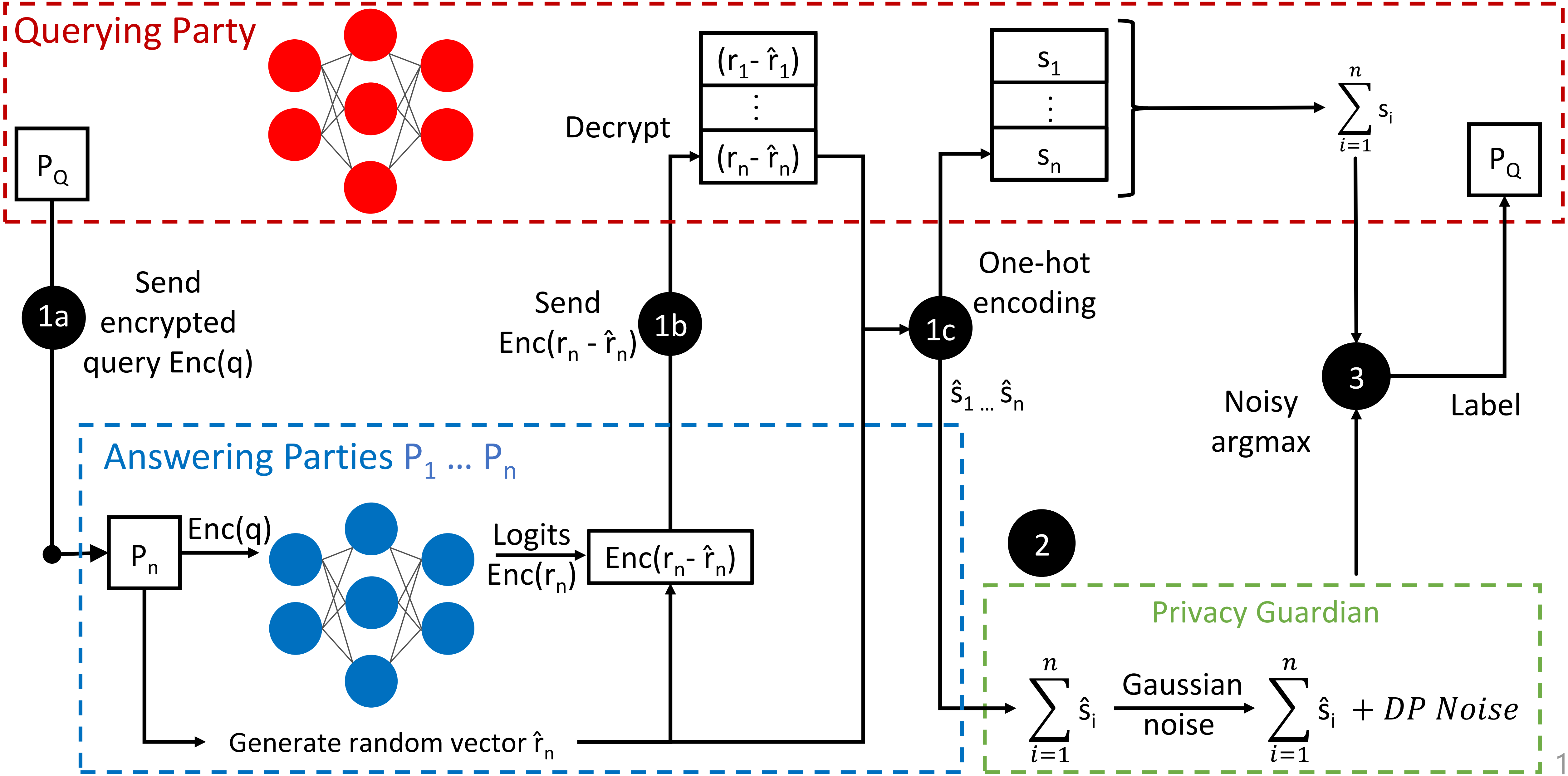


1. Requirements & Overview

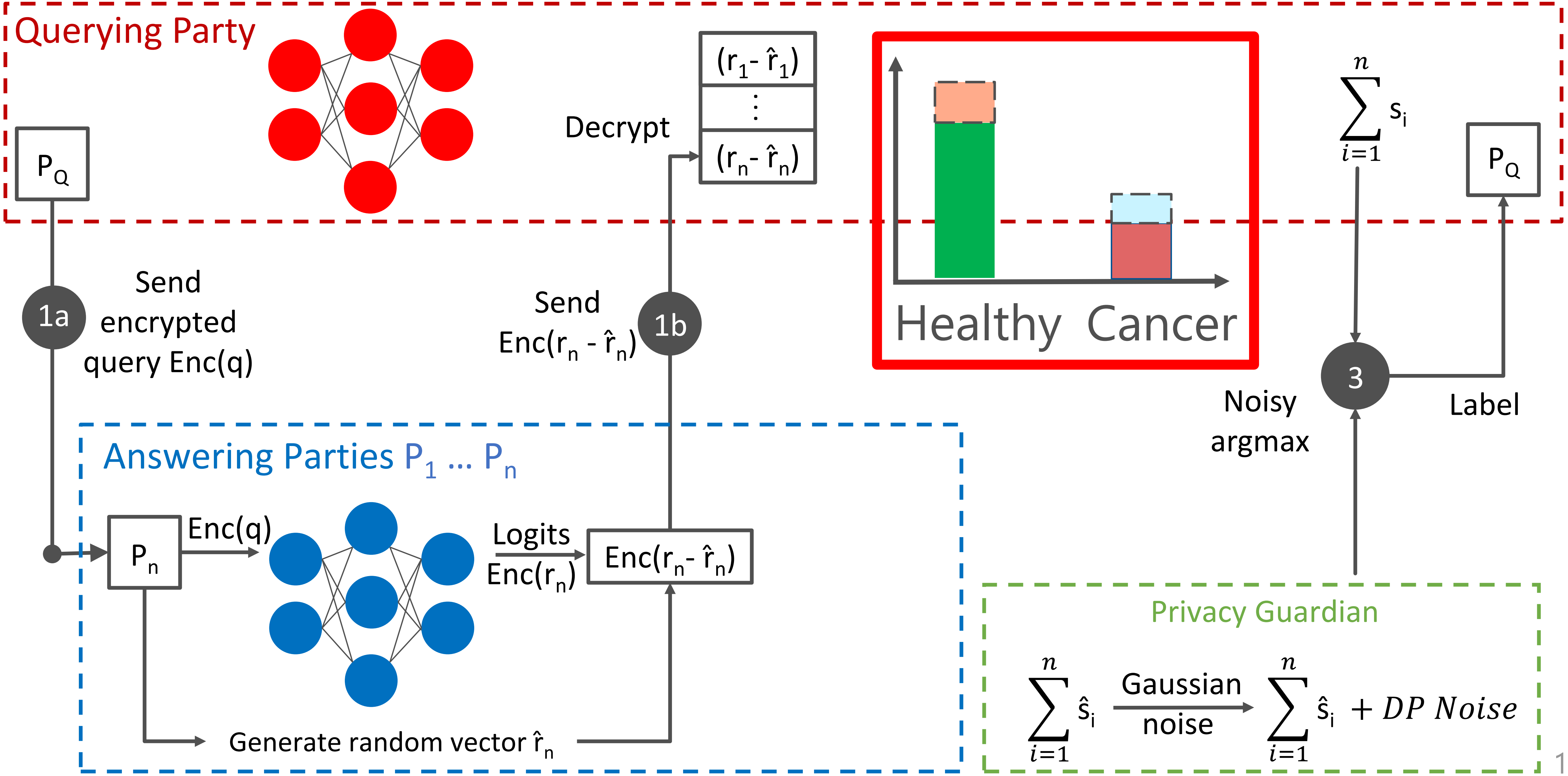
2. CaPC Protocol

3. Empirical Evaluation

CaPC Protocol



Noisy Argmax



Confidentiality & Privacy Guarantees

- **Honest-but-curious** setting – adversary follows the protocol but tries to infer information from the protocol transcript.
- **Semi-trusted third party** - PG (Privacy Guardian) does not collude with any other party.
  - If PG colludes with a querying party (and no noise added) – there is no privacy protection.
- **Perfect confidentiality** – assumed above, protocol reveals nothing except the final noised result to the querying party.
- **Strong privacy** – when at most 1 corrupted answering party.
  - Privacy degrades only and proportionally to the number of corrupted answering parties.
  - Privacy leakage only to the querying party when more than a single answering party is corrupted.

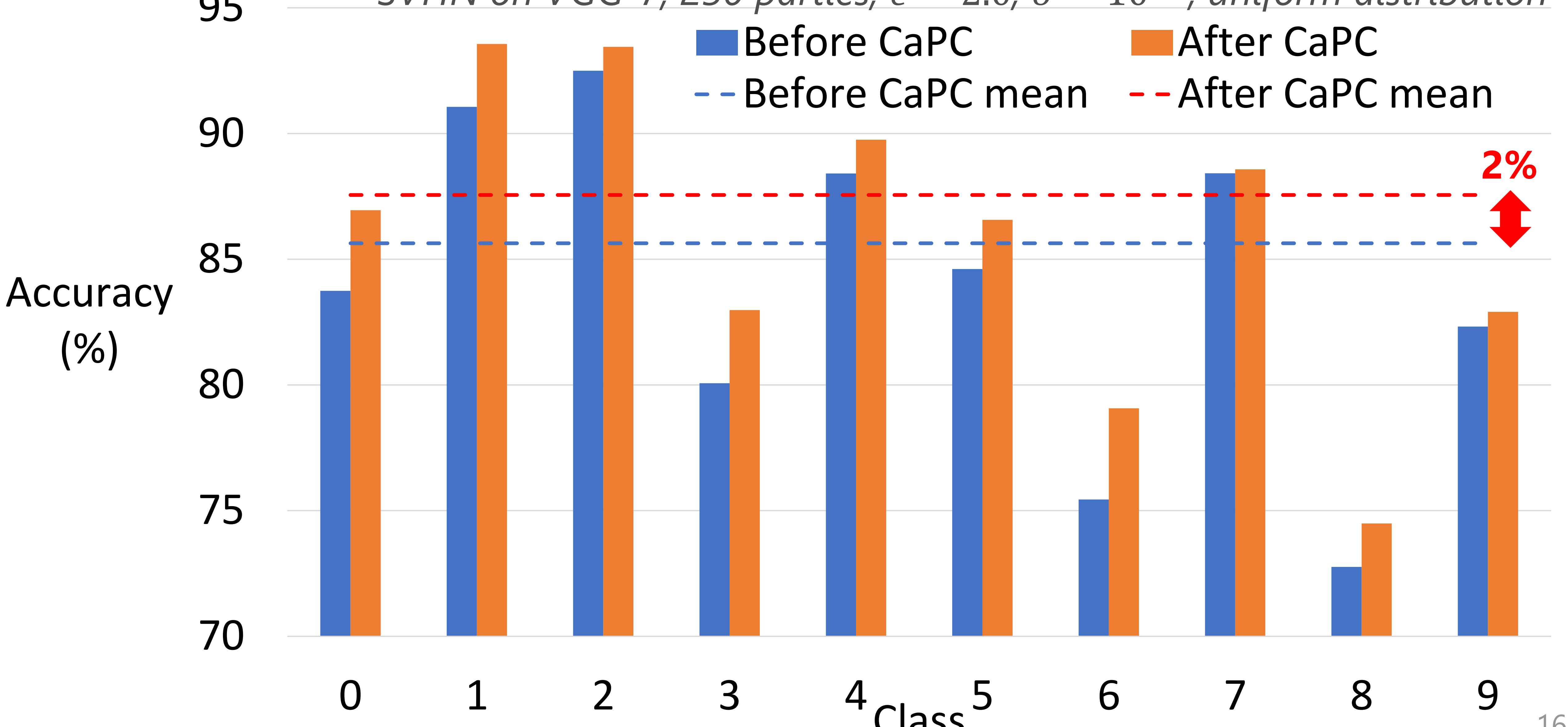
1. Requirements & Overview

2. CaPC Protocol

3. Empirical Evaluation

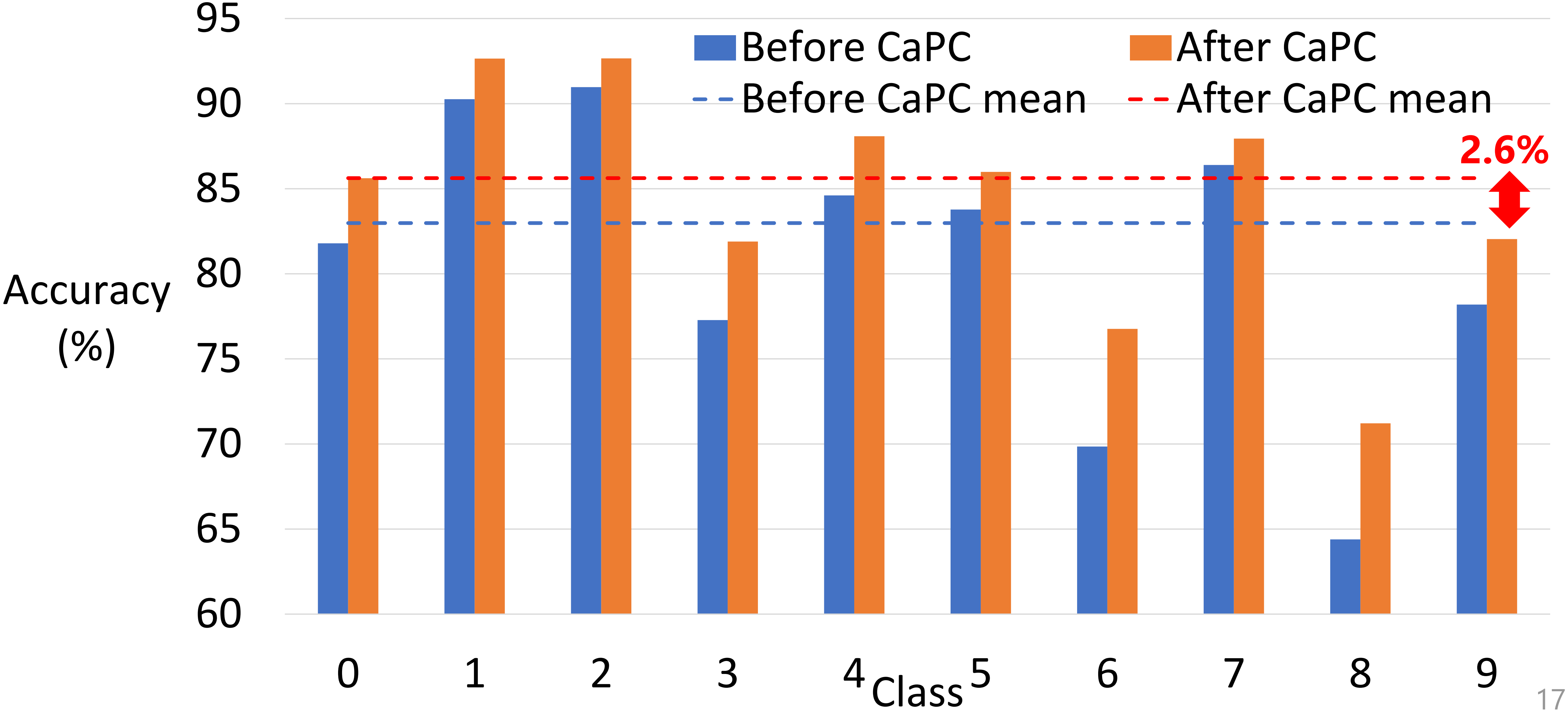
Homogenous Architecture

SVHN on VGG-7, 250 parties, $\epsilon = 2.0$, $\delta = 10^{-6}$, uniform distribution



Heterogenous Architecture

SVHN on **VGG-7, ResNet-8, ResNet-10**, *250 parties*, $\epsilon = 2.0$, $\delta = 10^{-6}$, *uniform distribution*



Active Learning

Active Learning for Query Selection

Given: an unlabeled dataset $\mathbf{d}$ and a classification model with conditional label distribution $\mathbf{P}_{\boldsymbol{\theta}}(\mathbf{y}|\mathbf{x})$, where $x \in d$.

Margin Sampling uses the gap between the most probable class and runner-up:

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbf{d}}{\mathbf{argmin}}\, \mathbf{P}_{\boldsymbol{\theta}}(\hat{\mathbf{y}}_1|\mathbf{x}) - \mathbf{P}_{\boldsymbol{\theta}}(\hat{\mathbf{y}}_2|\mathbf{x})$$

where $\hat{\mathbf{y}}_1$ and $\hat{\mathbf{y}}_2$ the most and second most probable labels for $\boldsymbol{x}$, according to the model.

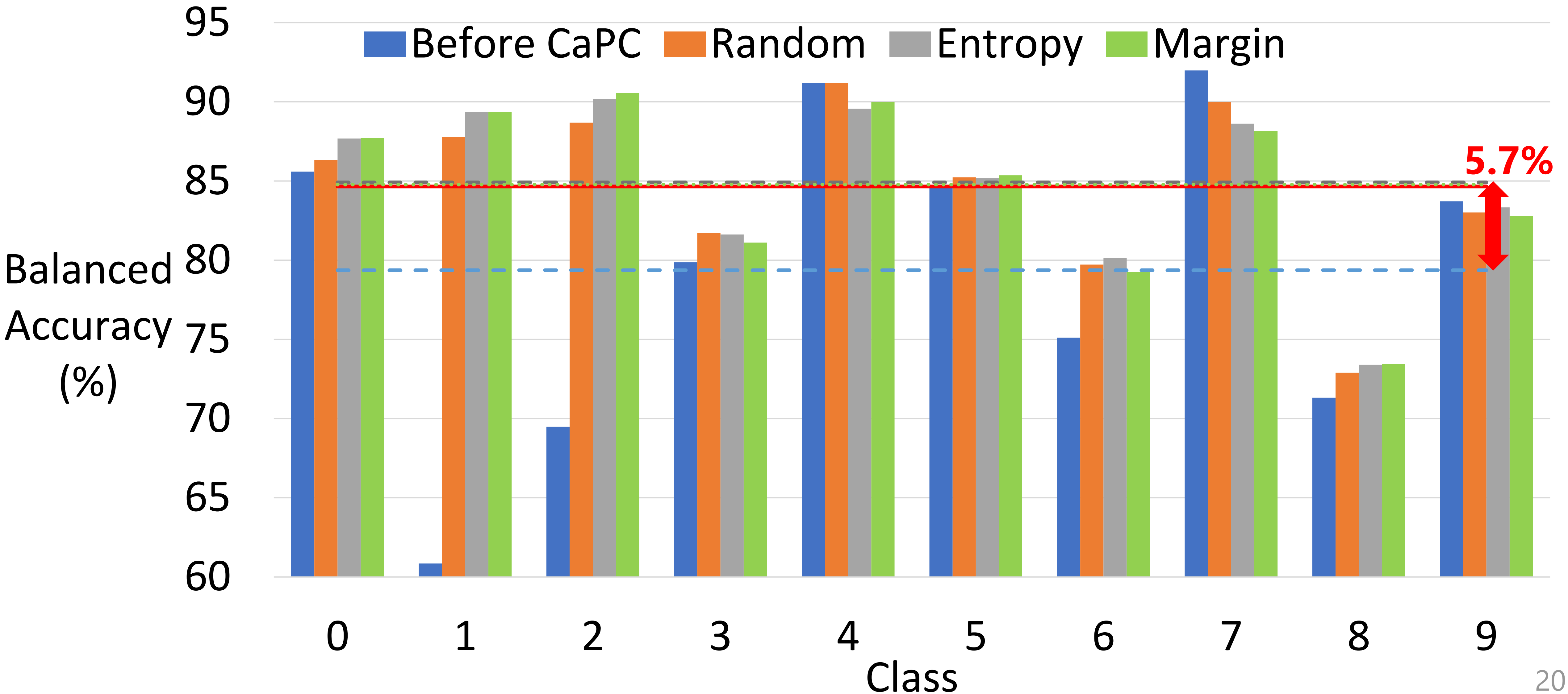
Entropy Sampling *uses entropy as an uncertainty measure:*

$$\boldsymbol{x}^* = \underset{\boldsymbol{x} \in \boldsymbol{d}}{\mathbf{argm}\boldsymbol{ax}} - \sum_{\boldsymbol{i}} \boldsymbol{P}_{\boldsymbol{\theta}}(\boldsymbol{y}_{\boldsymbol{i}}|\boldsymbol{x}) \,\mathbf{log}\, \boldsymbol{P}_{\boldsymbol{\theta}}(\boldsymbol{y}_{\boldsymbol{i}}|\boldsymbol{x})$$

where y_i ranges over all possible labels.

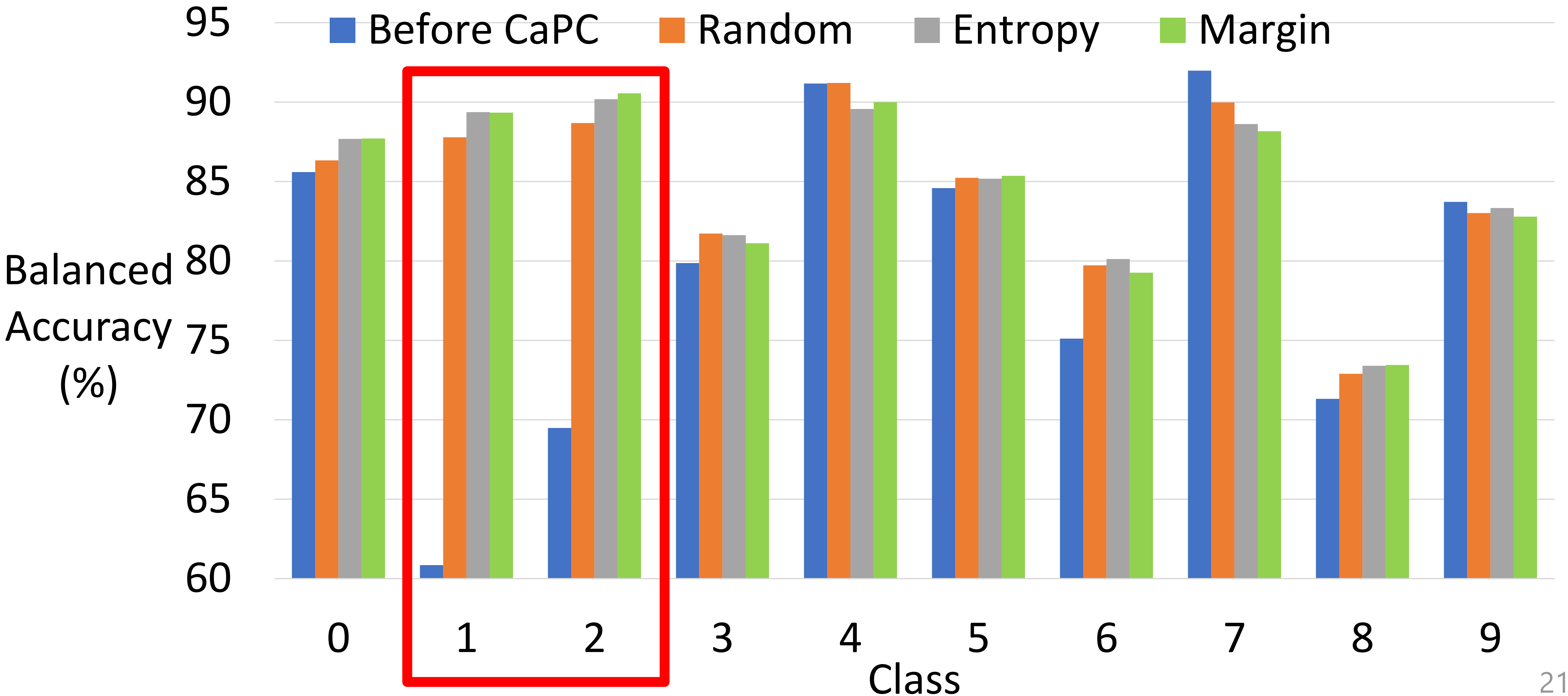
Non-uniform Data Distribution

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



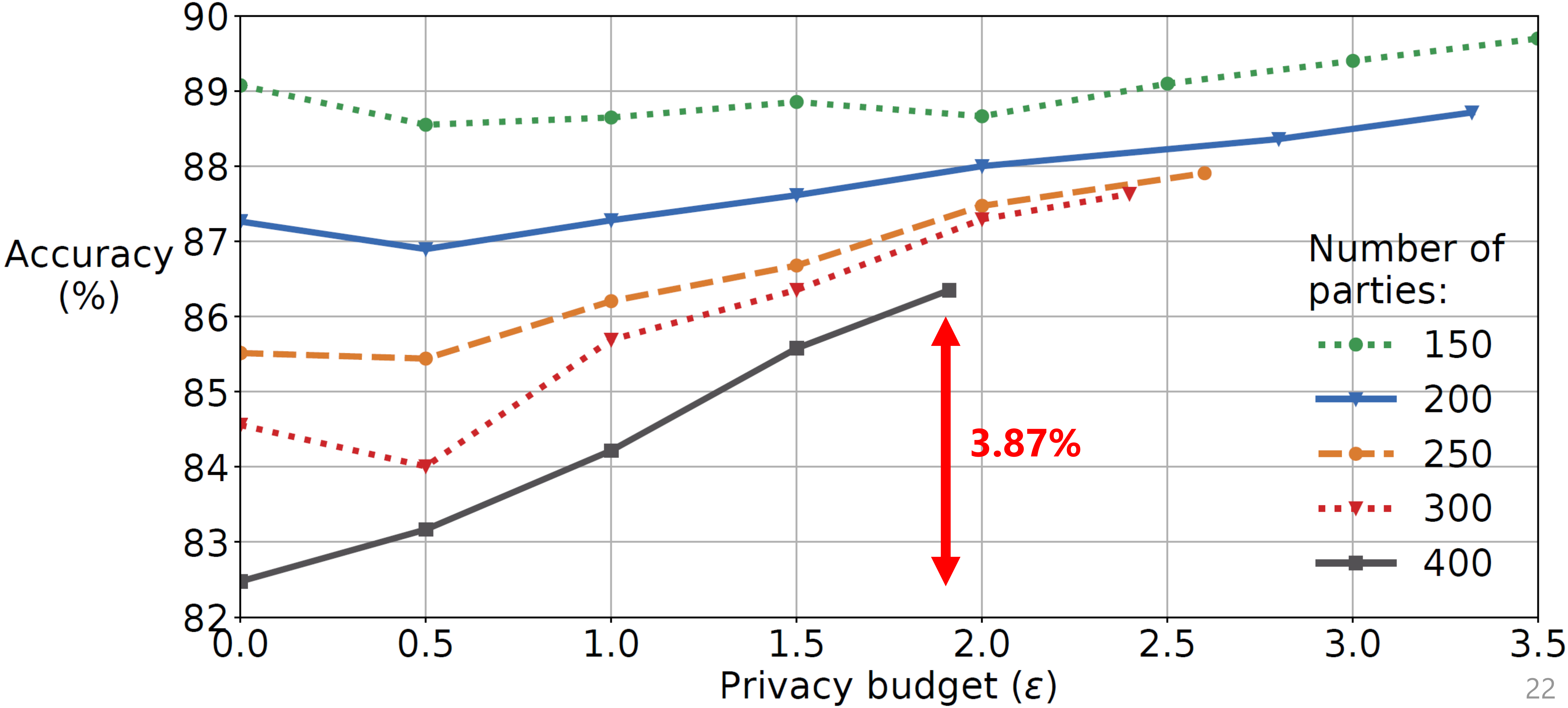
Non-uniform Data Distribution

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

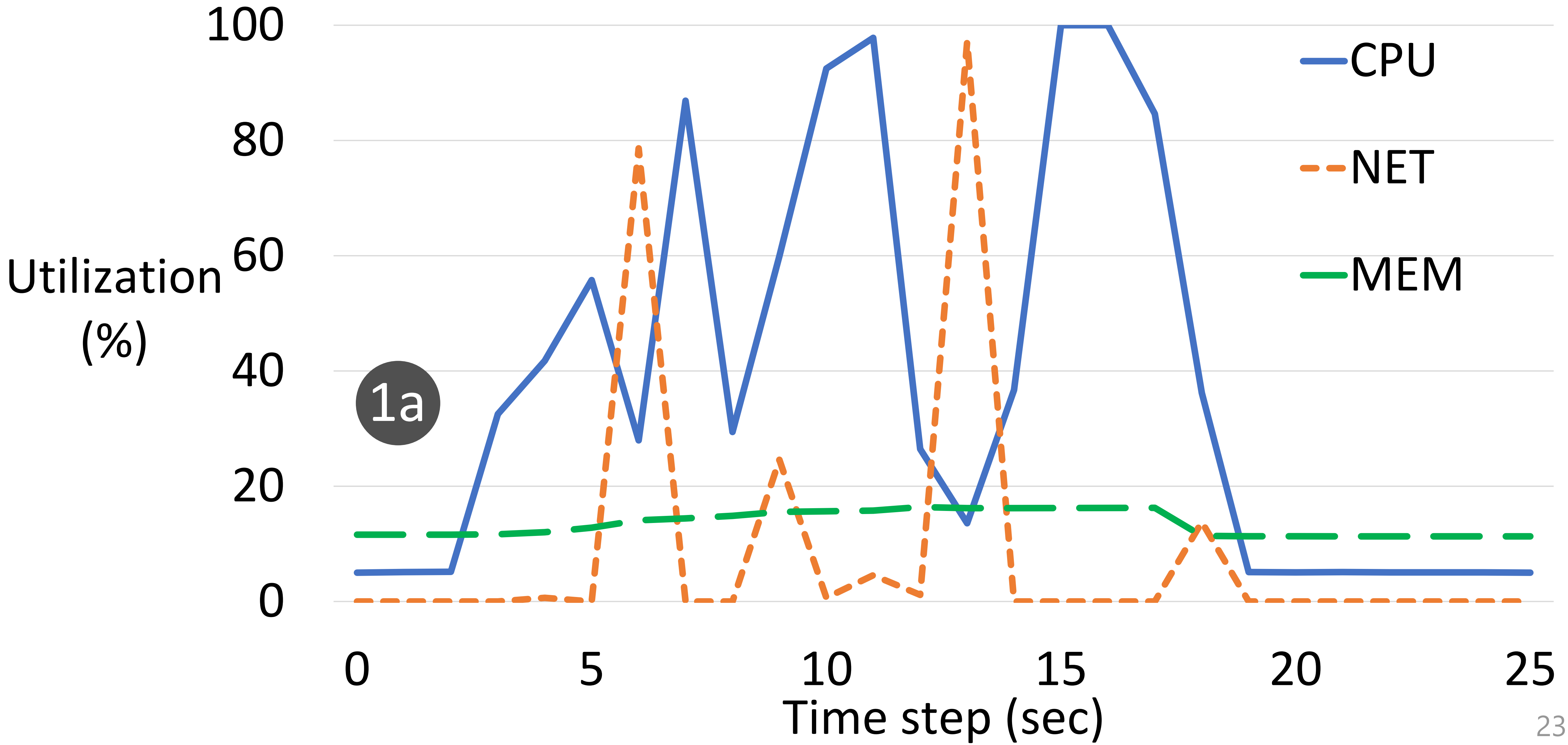


Accuracy Gain vs Number of Parties

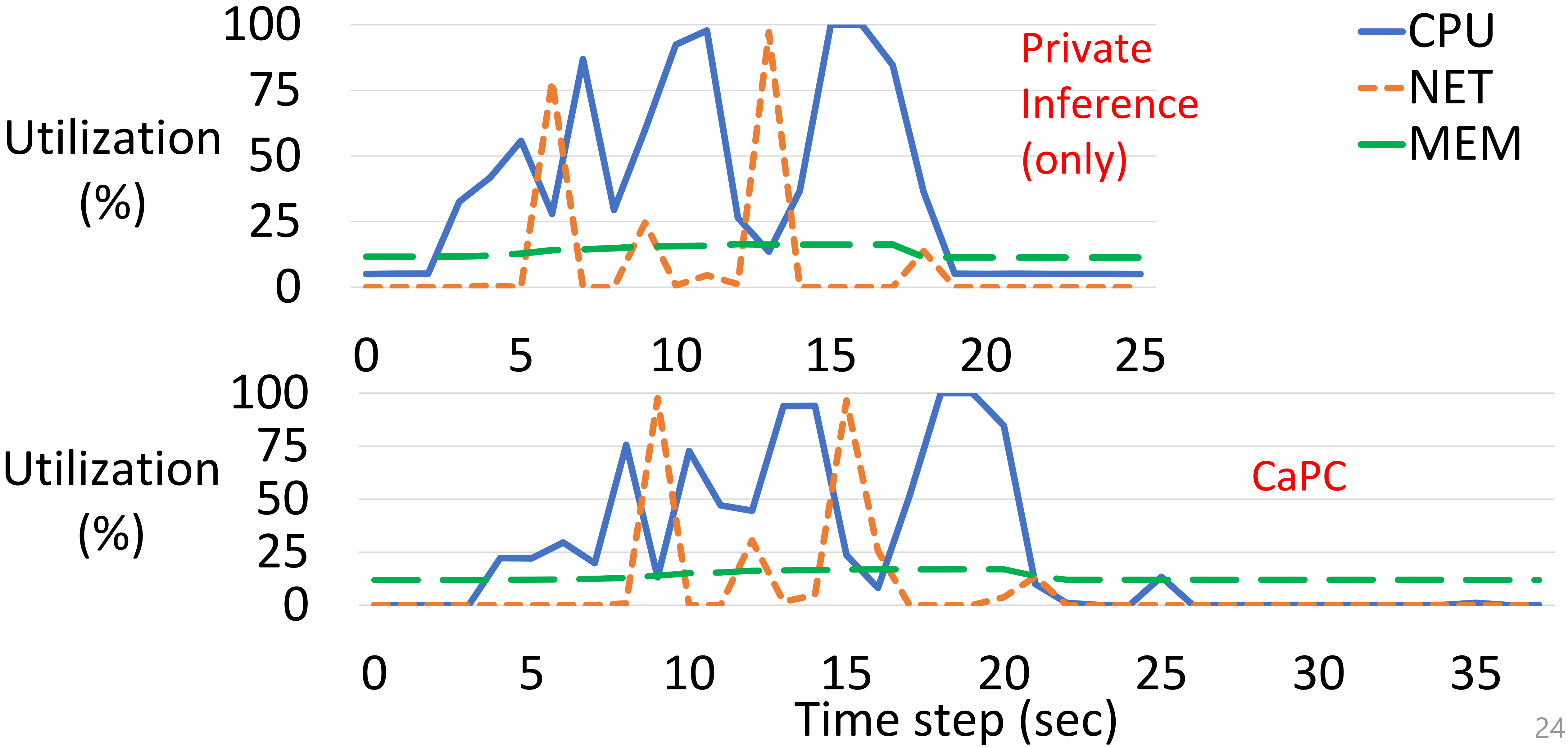
SVHN on VGG-7, 250 parties, $\epsilon = 2.0$, $\delta = 10^{-6}$, uniform distribution



Computational Cost of Private Inference



Small Computational Overhead of CaPC



CaPC	Federated Learning
Cross-silo setting, e.g., organizations	Cross-device setting, e.g., phones
Improve local models in each party by labeling new data points	Train central model without explicitly combining the parties' datasets
For heterogenous models and also non-differentiable models (trees)	Only for homogenous and differentiable models
Returns only predicted labels	Transfers gradients or parameters (large data transfer required)
Fewer parties required for privacy	Many more parties required
Provides confidentiality of data to be labeled & privacy (no gradients revealed) via Pâté.	Provides condifentiality but much higher cost of privacy (gradients shared allow us to infer private data)

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

Backup

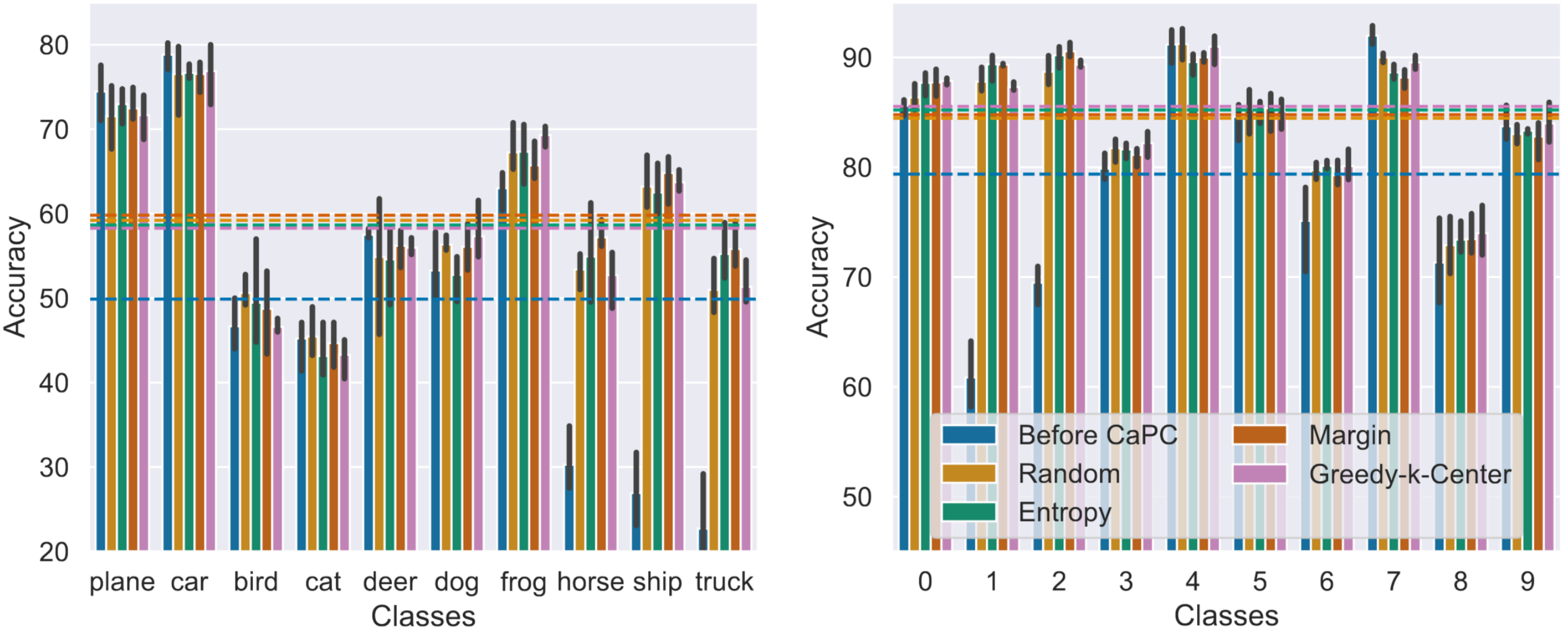
Definitions

- **Privacy** - aims to guarantee proper protection of personally identifiable information, against inference attacks.
- **Confidentiality** - aims to guarantee non-disclosure of sensitive information to unauthorized entities.
- **Integrity** - aims to prevent unauthorized modification of data and models.
- **Secure Multi-Party Computation** – a way for parties to compute a function jointly while keeping their inputs secret. In ML, this function can be a model's loss function during training, or the model itself in inference.

Definitions

- **Secret Sharing -** splitting the data into shares is the encryption, adding the shares back together is the decryption.

Balanced accuracy under non-uniform data distribution via Active Learning



Break-down of the execution time

Method	**Forward Pass (Step 1a)**
$CPU, P = 8192$	14.22 ± 0.11
$CPU, P = 16384$	29.46 ± 2.34
$CPU, P = 32768$	57.26 ± 0.39
GPU, no encryption	3.15 ± 0.22
CPU, no encryption	0.152 ± 0.0082

QP-AP (Steps 1b and 1c)	**QP-PG (Steps 2 and 3)**
0.12 ± 0.0058	0.030 ± 0.0045