MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

CaPriDe Learning: Confidential and Private Decentralized Learning Based on Encryption-friendly Distillation Loss

Nurbek Tastan, Karthik Nandakumar
MBZUAI, Abu-Dhabi, UAE

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

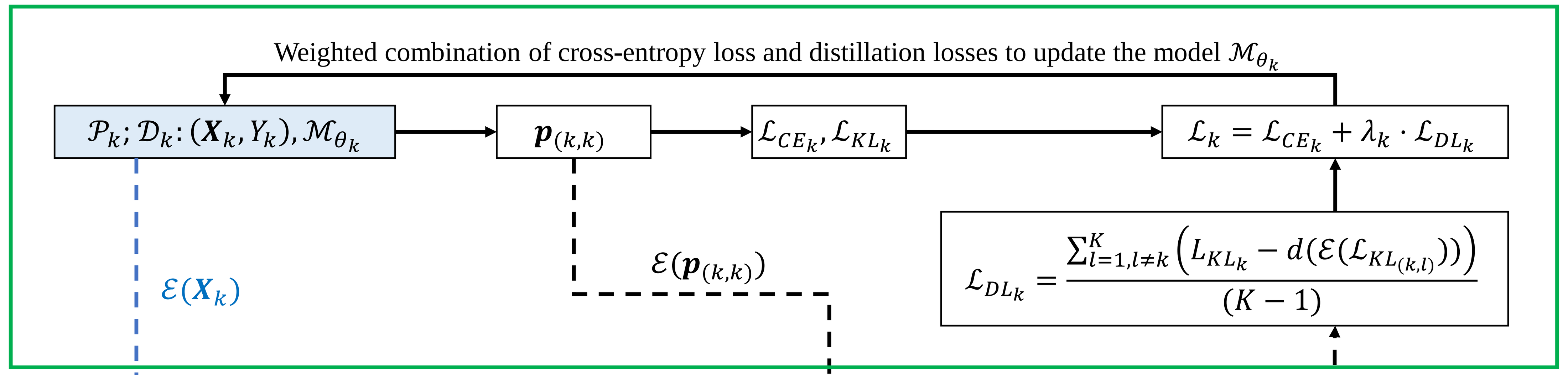
Motivation:

- Large volume of data required to train deep neural networks (DNNs) is seldom available to one single entity.
- Data sharing between entities or with third-party service providers is constrained by privacy concerns and regulations.
- Federated learning allows collaborative training of DNNs without data sharing, but has unacceptable utility-privacy trade-off.

Contributions:

- A collaborative learning algorithm based on encrypted inference and knowledge distillation to achieve confidentiality and privacy without any central orchestration and non-private shared data.
- An encryption-friendly distillation loss that estimates the approximate KL divergence between model predictions and a protocol to securely compute the loss in the encrypted domain.

Method



$\boldsymbol{p}_{(k,k)}$ - probability distribution computed using $\mathcal{M}_{\theta_k}$ on $\boldsymbol{X}_k$
$\mathcal{E}(\boldsymbol{z}_{(k,l)})$ - logits vector computed using $\mathcal{M}_{\theta_l}$ on $\mathcal{E}(\boldsymbol{X}_k)$

Pairwise distillation loss:

$$\mathcal{L}_{DL_{(k,l)}} = \sum_{j=1}^{N_k}(\boldsymbol{p}_{j,(k,k)} \cdot \log \boldsymbol{p}_{j,(k,k)}) - \sum_{j=1}^{N_k}\boldsymbol{p}_{j,(k,k)} \cdot \left(\frac{\boldsymbol{z}_{j,(k,l)}}{T} - \log\left(\sum_{j'=1}^{N_k}\exp\left(\frac{\boldsymbol{z}_{j',(k,l)}}{T}\right)\right)\right)$$

The total distillation loss for $\mathcal{P}_k$:

$$\mathcal{L}_{DL_k} = \frac{1}{K-1}\sum_{l=1, l\neq k}^{K}\mathcal{L}_{DL_{(k,l)}}$$

Participant $\mathcal{P}_k$:

$$\mathcal{L}_k = \mathcal{L}_{CE_k} + \lambda_k \mathcal{L}_{DL_k}$$

Experimental Setup

Datasets & Architectures:

Architecture	Dataset	Batch size	λ_k	Description
ResNet-18	CIFAR-10	128	50	60000 (10 classes), 32x32 images
	CIFAR-100	128	50	60000 (100 classes), 32x32 images
	HAM10000	32	20	10015 (7 classes), 224x224 images

Fully Homomorphic Encryption (FHE) Algorithm:

Tile Tensor Framework that relies on the CKKS scheme (Aharoni et al., Complex Encoded Tile Tensors: Accelerating Encrypted Analytics, IEEE S&P, 2022)

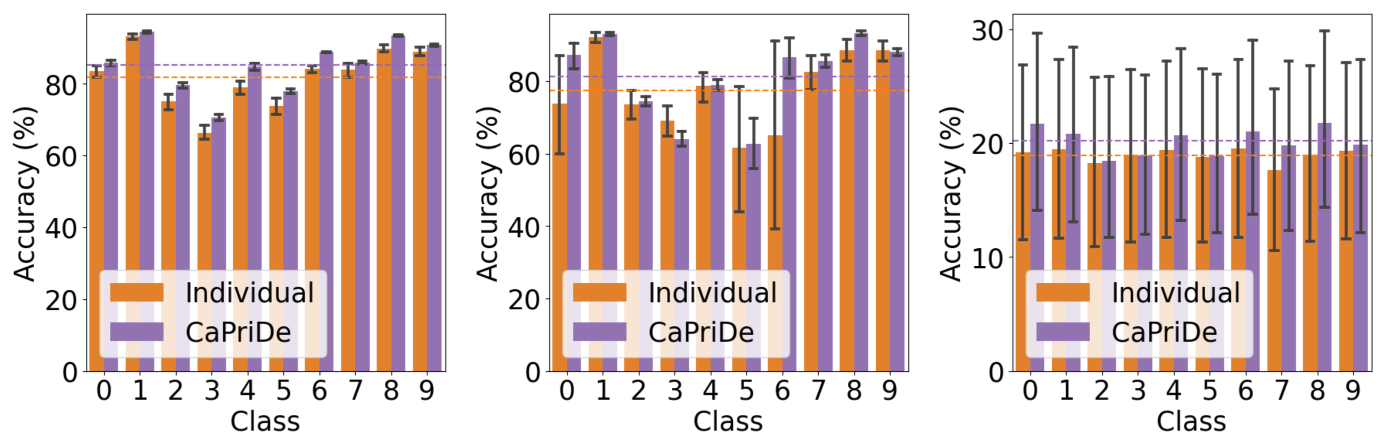
Data Partition:

- **Homogeneous:** each participant has an equal number of samples per class;
- **Heterogeneous:** each participant has an unequal number of samples determined randomly (both total number and number of samples per class);
- **Non-overlapping class distribution:** each participant has samples from a non-overlapping subset of classes.

Experiments & Results

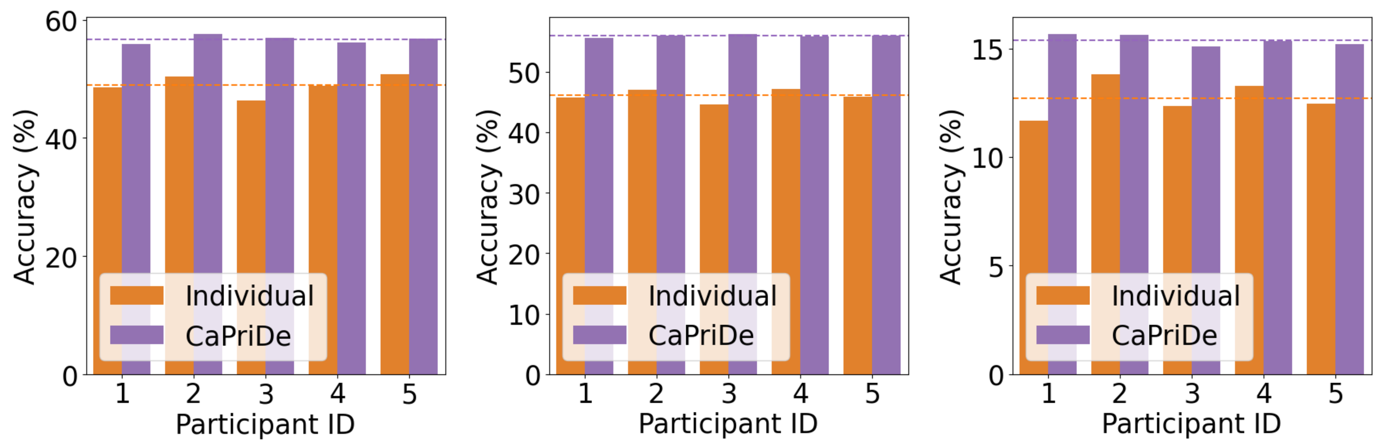
Collaborative Learning Results:

Dataset	Setting	K	FedAvg	FedAvg+DP	CaPriDe
CIFAR-10	Homogeneous	2	3.46	-20.54	1.58
		5	9.22	-11.8	3.55
		10	15.44	2.58	5.52
	Heterogeneous	2	6.12	-17.04	0.93
		5	13.78	-7.84	5.06
		10	21.81	5.69	6.20
	No class overlap	2	29.85	11.08	8.16
CIFAR-100	Homogeneous	2	7.08	-30.97	4.19
		5	22.68	-13.81	9.10
		10	26.10	-0.11	11.69
	Heterogeneous	2	9.59	-28.89	6.28
		5	22.50	-13.26	9.23
		10	34.38	3.86	10.68
	No class overlap	2	19.40	5.14	7.75
HAM10000	Homogeneous	2	0.97	0.48	1.81
	Heterogeneous	2	1.93	-1.21	1.67



(a) Homogeneous (b) Heterogeneous (c) No overlap

Per-class accuracy on CIFAR-10



(a) Homogeneous (b) Heterogeneous (c) No overlap

Per-participant accuracy on CIFAR-100

FHE Results:

	CIFAR-10	CIFAR-100	HAM10000
Security Level	128	128	128
Number of slots	16384	16384	16384
Time taken to encrypt one sample	90 ms	103 ms	619 ms
Ciphertext size of one sample	29.101 KB	29.152 KB	1.359 MB
Time taken to encrypt a batch of 32 samples	1.31 s	1.26 s	15.57 s
Encrypted inference of a batch of 32 samples	110.21 s	112.09 s	896.12 s

Setting	K	Individual	CaPriDe (KL)	CaPriDe (L_2)
Homogeneous	2	91.050	**92.155**	91.025
Homogeneous	5	81.770	**85.194**	82.710
Homogeneous	10	68.065	**72.580**	68.272
Heterogeneous	2	87.485	**88.424**	87.320
Heterogeneous	5	77.336	**81.324**	76.070
Heterogeneous	10	64.320	**70.520**	65.010

Proposed approx. KL loss in comparison with L_2 loss