Scalable, Heterogeneity-Aware and Privacy-Enhancing Federated Learning

Yiran Chen

Department of Electrical and Computer Engineering, Duke University

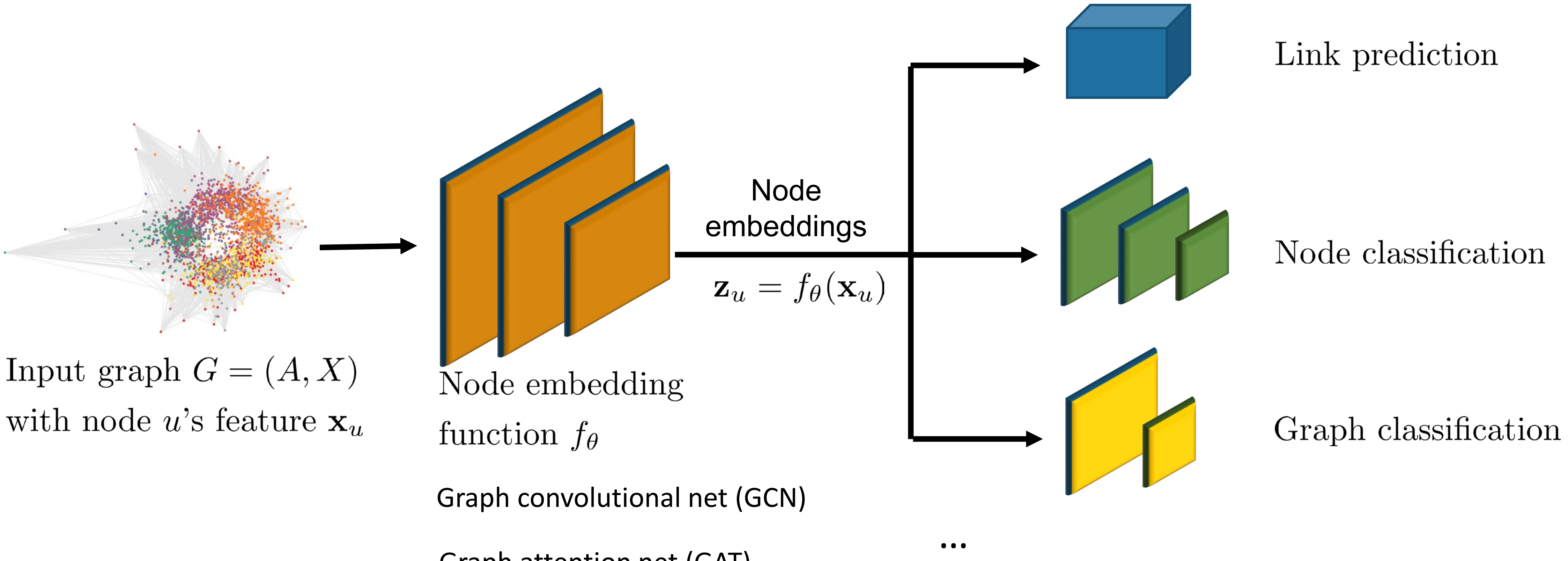
Duke

Outline

- Privacy-Preserving Representation Learning on Graphs: A Mutual Information Perspective (**KDD 2021**)
- Efficient and Heterogeneity-Aware Federated Learning
  - LotteryFL: Personalized and Communication-Efficient Federated Learning with Lottery Ticket Hypothesis on Non-IID Datasets (SEC’21)
  - FedMask: Joint Computation and Communication-Efficient Personalized Federated Learning via Heterogeneous Masking (SenSys’21)
- Privacy-Enhancing and Robust Federated Learning
  - Provable Defense against Privacy Leakage in Federated Learning from Representation Perspective (CVPR’21)
  - Enhancing Robustness against Model Poisoning Attacks in Federated Learning from a Client Perspective (NeurIPS’21)

Privacy-Preserving Representation Learning on Graphs: A Mutual Information Perspective (KDD’21)

Representation Learning on Graphs



Graph convolutional net (GCN)

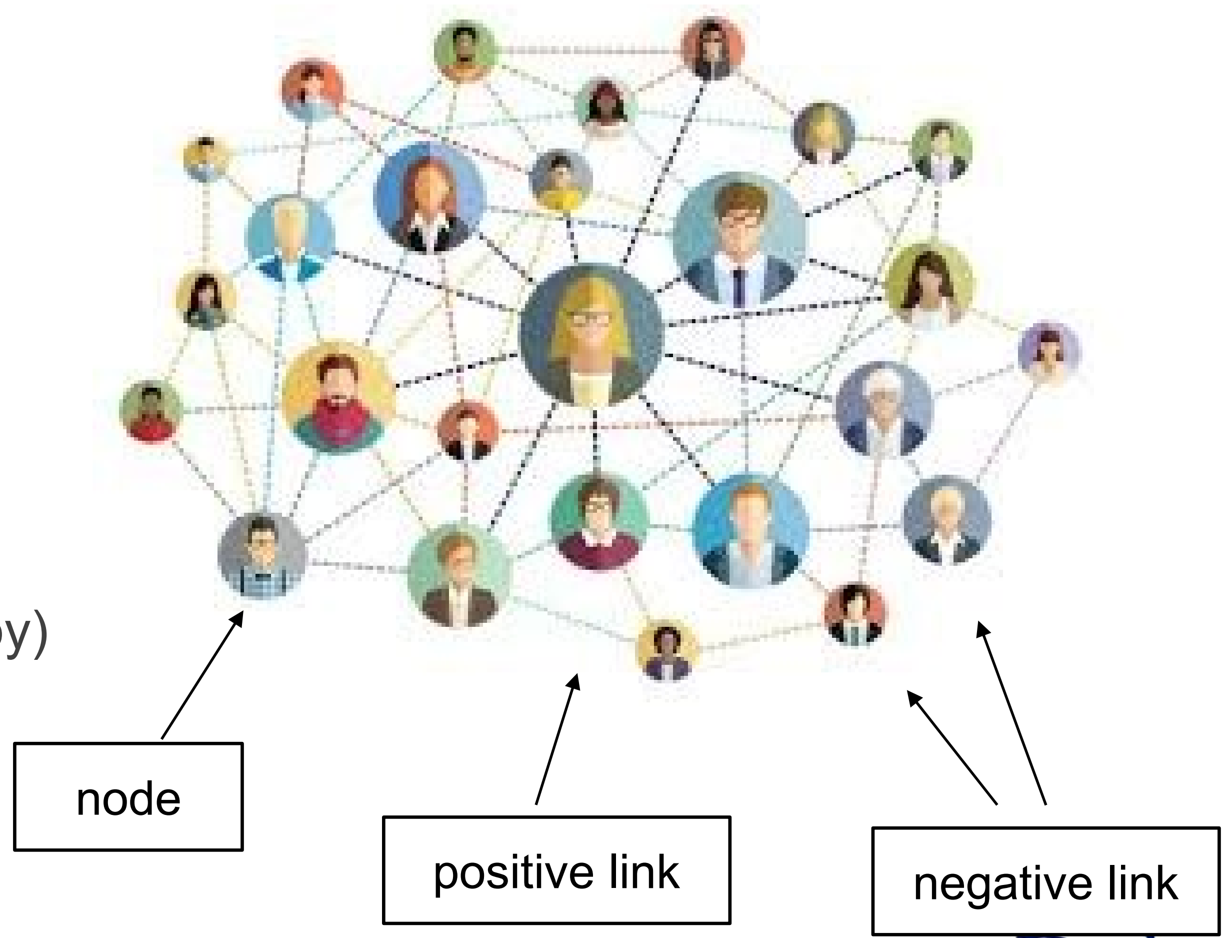
Graph attention net (GAT)

...

Duke

Example: Two Tasks in Social Networks

- Node classification
  - Infer user's private attributes
    (e.g., age, gender, sexual orientation, etc.)
- Link prediction
  - Predict relationship between users
    (e.g., whether two users have the same hobby)



Privacy Issues

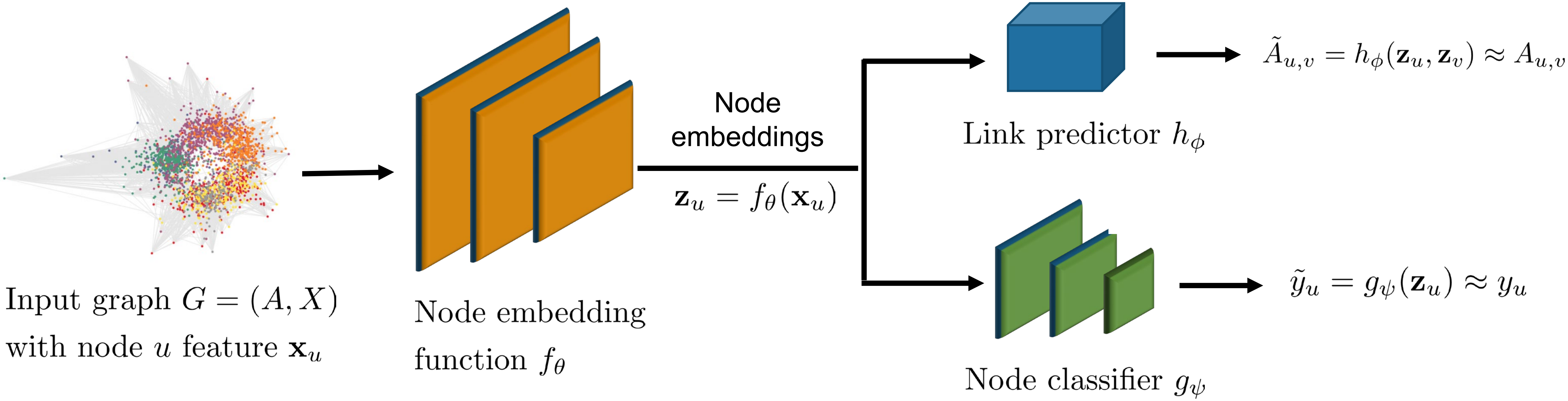
- One can accurately infer the links (node identity) from a node classifier (link predictor) trained on the learnt node embeddings
- Raise serious privacy issues (e.g., social network)
  - Celebrities just want to make their identities known to the public, but ***do not*** want to expose their private social (e.g., family) relationships
  - Malicious users do want to expose their social relationship with normal users to make themselves also look normal, but ***do not*** want to reveal their identities
  - Adversary can infer celebrities' private social relationship (malicious users' identities) based on user identity classification (social relationship prediction) system

Duke

Motivation

Primary learning task	**+**	**Privacy protection task**
Link prediction		Protect node privacy
Node classification		Protect link privacy

Problem Definition

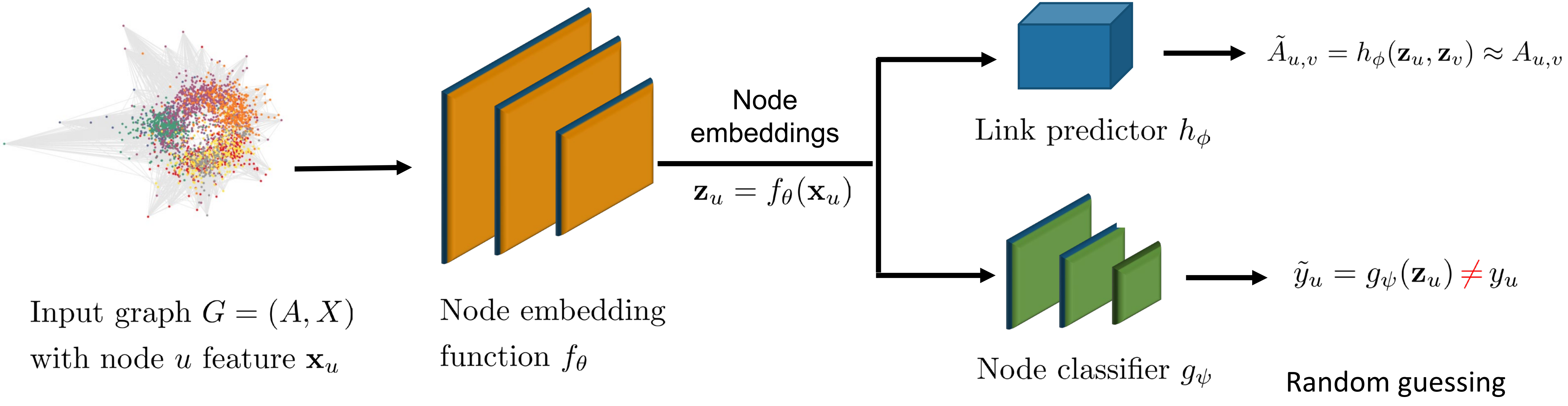


Problem 1: Link prediction with node privacy protection

Problem 2: Node classification with link privacy protection

Duke

Link Prediction with Node Privacy Protection



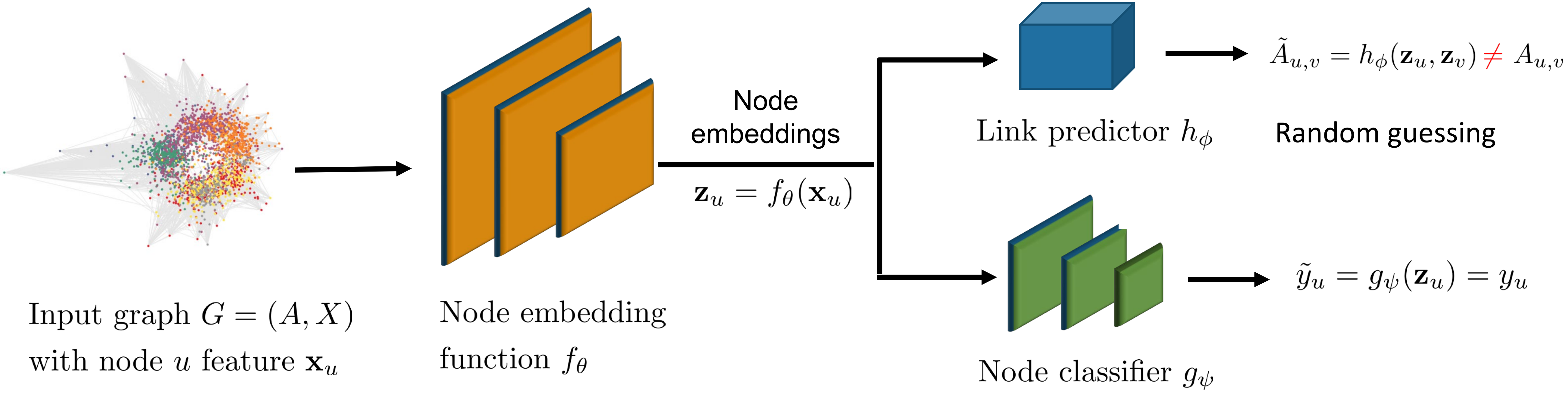
Mutual Information Objectives

Link prediction: $\max_\theta I(A_{uv}; \mathbf{z}_u, \mathbf{z}_v)$

Node privacy protection: $\min_\theta I(\mathbf{z}_u; y_u) = 0$

Duke

Node Classification with Link Privacy Protection



Mutual Information Objectives

Node classification: $\max_\theta I(\mathbf{z}_u; y_u)$

Link privacy protection: $\min_\theta I(A_{uv}; \mathbf{z}_u, \mathbf{z}_v) = 0$

Duke

Experimental Setup: Datasets + Metric

Datasets	#Nodes	#Edges	#Features	#Node Classes	#Link Classes
Cora	2,708	5,429	1,433	7	2
Citeseer	3,327	4,732	3,793	6	2
Pubmed	19,717	44,328	500	3	2

	Node classification	Link prediction
Training	20 per class	85% pos + 50% neg
Validation	500	5% pos + equal neg
Testing	1,000	10% pos + equal neg

Evaluation metric

Node classification: Accuracy

Link prediction: Area under curve (AUC)

Duke

Primary Learning + Privacy Protection Results

Primary task: link prediction	Link Prediction AUC			Node Accuracy		
	Cora	Citeseer	Pubmed	Cora	Citeseer	Pubmed
Without node privacy protection	89.33%	91.52%	91.43%	72.00%	67.40%	72.70%
With node privacy protection	84.12%	85.55%	84.24%	21.40%	17.40%	42.50%
Random guessing				14.29%	16.67%	33.33%

Primary task: node classification	Node Accuracy			Link Prediction AUC		
	Cora	Citeseer	Pubmed	Cora	Citeseer	Pubmed
Without link privacy protection	81.60%	67.50%	78.90%	82.73%	83.30%	78.80%
With link privacy protection	79.70%	65.80%	78.60%	50.50%	53.29%	49.57%
Random guessing				50.00%	50.00%	50.00%

Duke

Summary

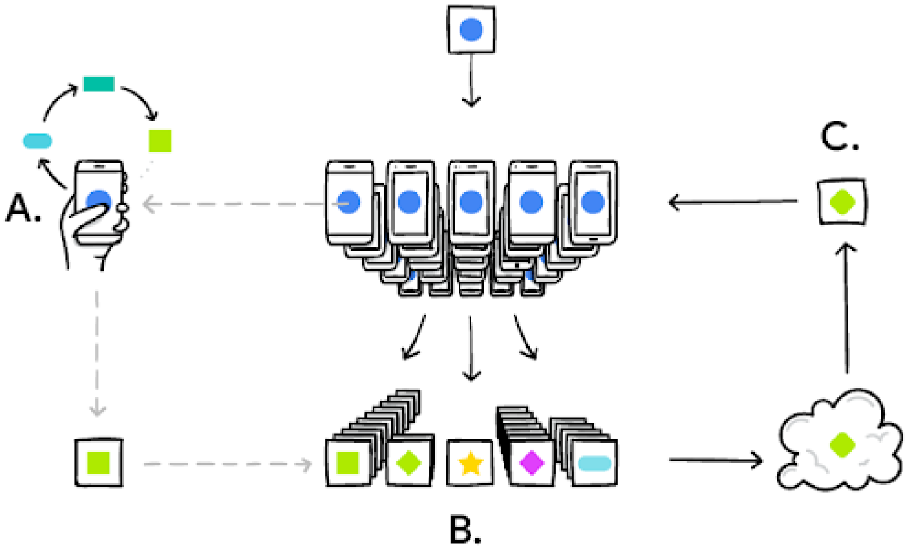
- We propose the first privacy-preserving representation learning framework on graphs
- Our framework is from the mutual information perspective and involves both a primary task and a privacy task
- We derive tractable mutual information bounds and train parameterized neural networks to estimate these bounds
- Our framework is effective to learn privacy-preserving node embeddings

Duke

LotteryFL: Empower Edge Intelligence with Personalized and Communication-Efficient Federated Learning (SEC’21)

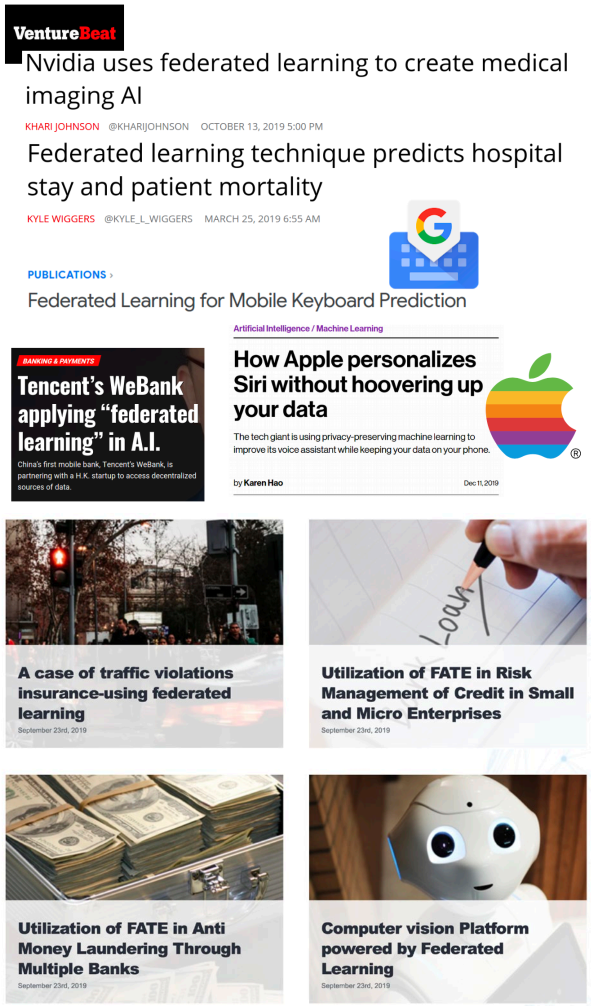
Background

- Federated learning (FL)



Your phone personalizes the model locally, based on your usage (A). Many users' updates are aggregated (B) to form a consensus change (C) to the shared model, after which the procedure is repeated.

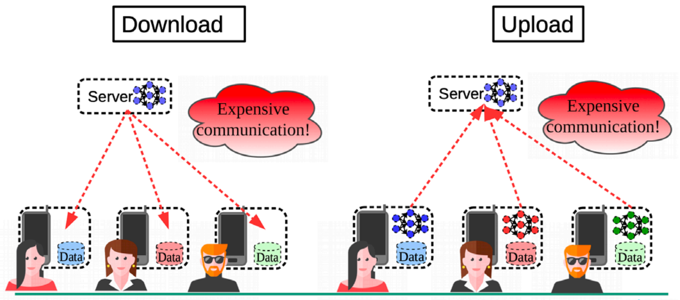
https://ai.googleblog.com/2017/04/federated-learning-collaborative.html



Duke

Challenges

- Communication efficiency
  - Total Communication = [#Communication Rounds] x [#Parameters] x [Avg. Codeword length]
- Case Study: VGG16 on ImageNet
  - Number of rounds until Convergence: 9,000
  - Number of Parameters: 138, 000, 000
  - Bits per Parameter: 32
  - Total Communication = **496.8 Terabyte** (round trip)



- Statistical heterogeneity
  - Devices frequently generate and collect data in a *non-identically distributed (non-IID)* manner across the network
  - The global model learned using FedAvg does not perform well when the data on different devices is heterogeneous

CIFAR-10 Settings	IID	Non-IID
Accuracy of FedAvg	89.21%	47.67%

Prior Arts

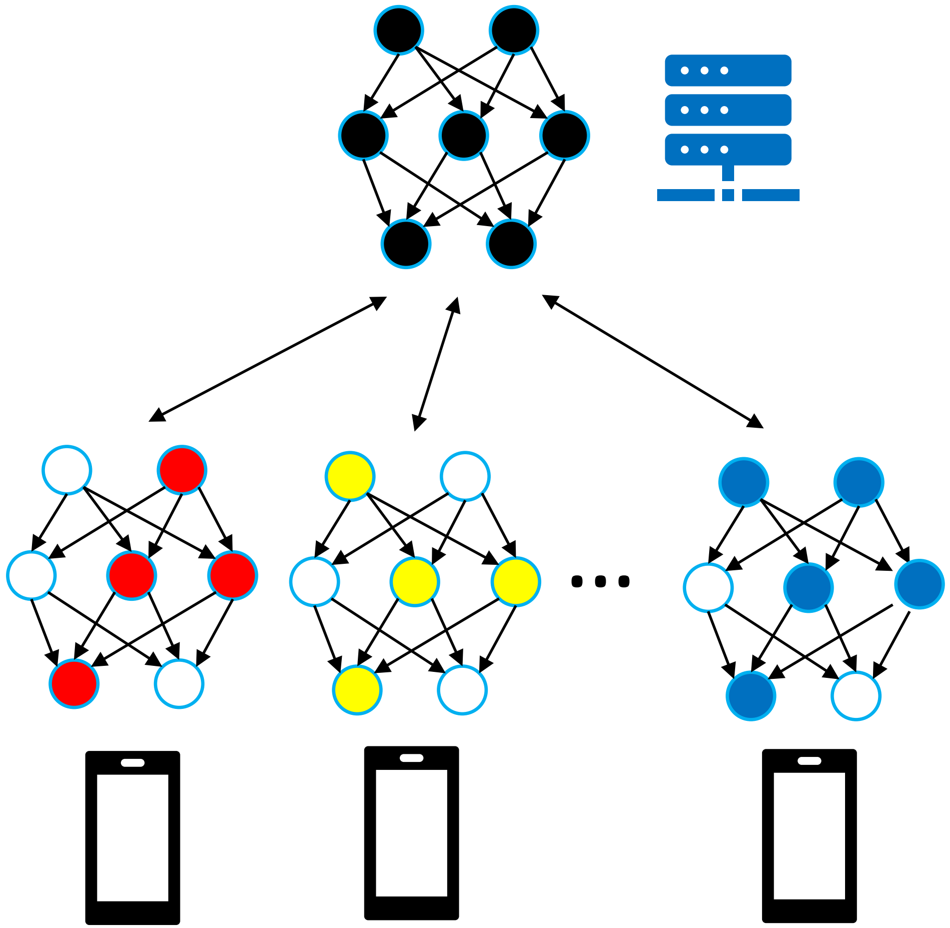
- Communication cost: compressing communicated data
  - Reduce communication frequency
  - Compress local updates, e.g., sparsity
  - Efficient encoding, e.g., quantization
- Statistical Heterogeneity
  - Mitigate the divergence between local models and the global model (*FedProx*) or make make activation vectors across multiple devices more similar (*FedMax*)
  - Personalization: meta learning, multi-task learning, transfer learning, etc.
- Limitations
  - Cannot address the two challenges simultaneously
  - Target unrealistic federate learning settings

Li, Tian, et al. "Federated optimization in heterogeneous networks." *MLSys.* 2020.

Chen, Wei, et al.. "FedMAX: Mitigating Activation Divergence for Accurate and Communication-Efficient Federated Learning." arXiv preprint arXiv:2004.03657 (2020).

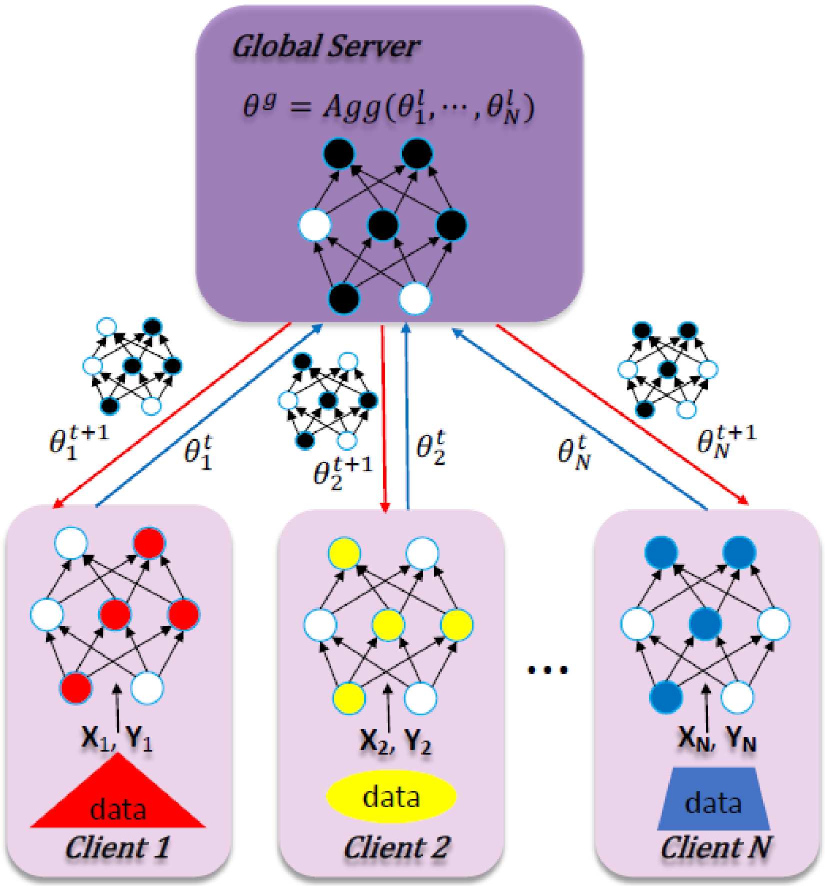
Motivation

- LotteryFL
  - Goal: improve *communication efficiency* and achieve *personalization* under *non-IID* settings
  - Non-IID+Personalization: seek device-specific "Lottery Ticket" subnets (LTN) for each device
  - Communication-efficient: only communicate the parameters of the subnets between devices and the central server



Design of LotteryFL

- Local Lottery Ticket Network Learning
  - Download subnet θ_k^t from the server
  - Prune and reset subnet θ_k^t if $acc > acc_{threshold}$ and $r_k^t < r_{target}$
  - Perform training using local data D_k and then update θ_k^{t+1}
- Personalization-Preserving Aggregation
  - Intuition: considering the non-IID data distribution across clients, the LTN of each client should not be significantly overlapped each other
  - Aggregation strategy: perform aggregation on the only *overlapped elements* among each LTN, while keeping the rest non-overlapped elements unchanged



Evaluations

- Realistic Non-IID settings
  - **Limited training data :** only 10-40 samples on each device
  - **Statistical heterogeneity:** only 2 classes of examples on each device
  - **Data unbalance:** data volumes are different across classes on each device
- Baselines
  - Standalone: local training only
  - FedAvg
  - LG-FedAvg: global model + local fine-tuning
- Evaluation metrics
  - *Inference accuracy*: we adopt the inference accuracy of each device's local test data to evaluate the performance of personalization, and report averaged accuracy over all devices
  - *Communication cost*: we use the data volume communicated between the clients and the server to measure communication costs

Duke

Extremely Limited Data Volumes

- Training on CIFAR-10 for 2000 communication rounds
  - Accuracy: increase by 13.48%-15.28% compared to LG-FedAvg
  - Communication cost: reduce 34%-53% compared to LG-FedAvg

Methods	**5 examples/class**		**10 examples/class**		**20 examples/class**	
	Acc (%)	Communication cost (MB)	Acc (%)	Communication cost (MB)	Acc (%)	Communication cost (MB)
Standalone	59.55	0	64.06	0	65.44	0
FedAvg	37.62	9425.35	43.20	9425.35	47.67	9425.35
LG-FedAvg	70.69	7174.58	72.09	7174.58	76.77	7174.58
LotteryFL	**85.97**	**3832.02**	**87.31**	**3069.95**	**90.61**	**2439.56**

Duke

Unbalanced Data

- Training on CIFAR-10 for 2000 communication rounds
  - Accuracy: increase by 13.84%-15.72% compared to LG-FedAvg
  - Communication cost: reduce by 59%-66% compared to LG-FedAvg

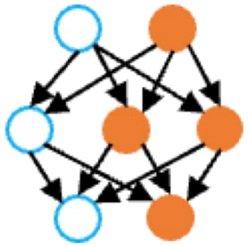
Methods	Balanced		Unbalanced (0.5)		Unbalanced (0.25)	
	Acc (%)	Communication cost (MB)	Acc (%)	Communication cost (MB)	Acc (%)	Communication cost (MB)
Standalone	65.44	0	55.60	0	50.33	0
FedAvg	47.67	9425.35	43.04	9425.35	40.19	9425.35
LG-FedAvg	76.77	7174.58	72.81	7174.58	69.03	7174.58
LotteryFL	**90.61**	**2439.56**	**88.53**	**2612.29**	**84.49**	**2973.22**

Duke

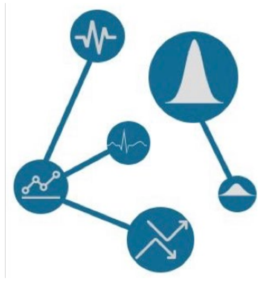
FedMask: Joint Computation and Communication-Efficient Personalized Federated Learning via Heterogeneous Masking (SenSys’21)

Overview of FedMask

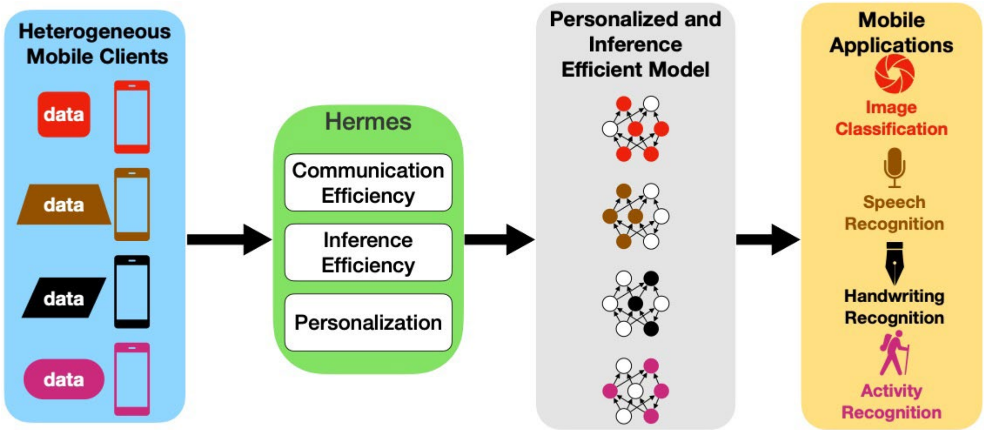
Minimize communication efficiency

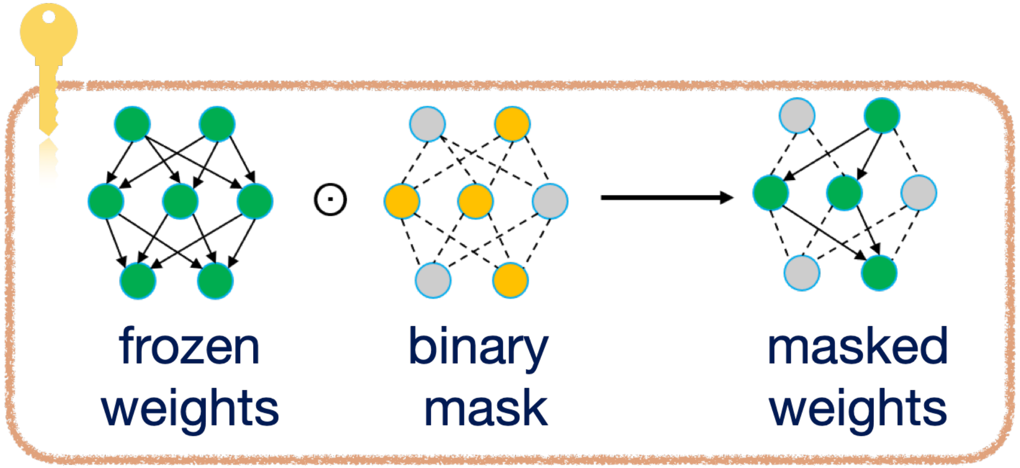


Reduce computation cost for both training and inference



Reduce computation cost for on-device inference

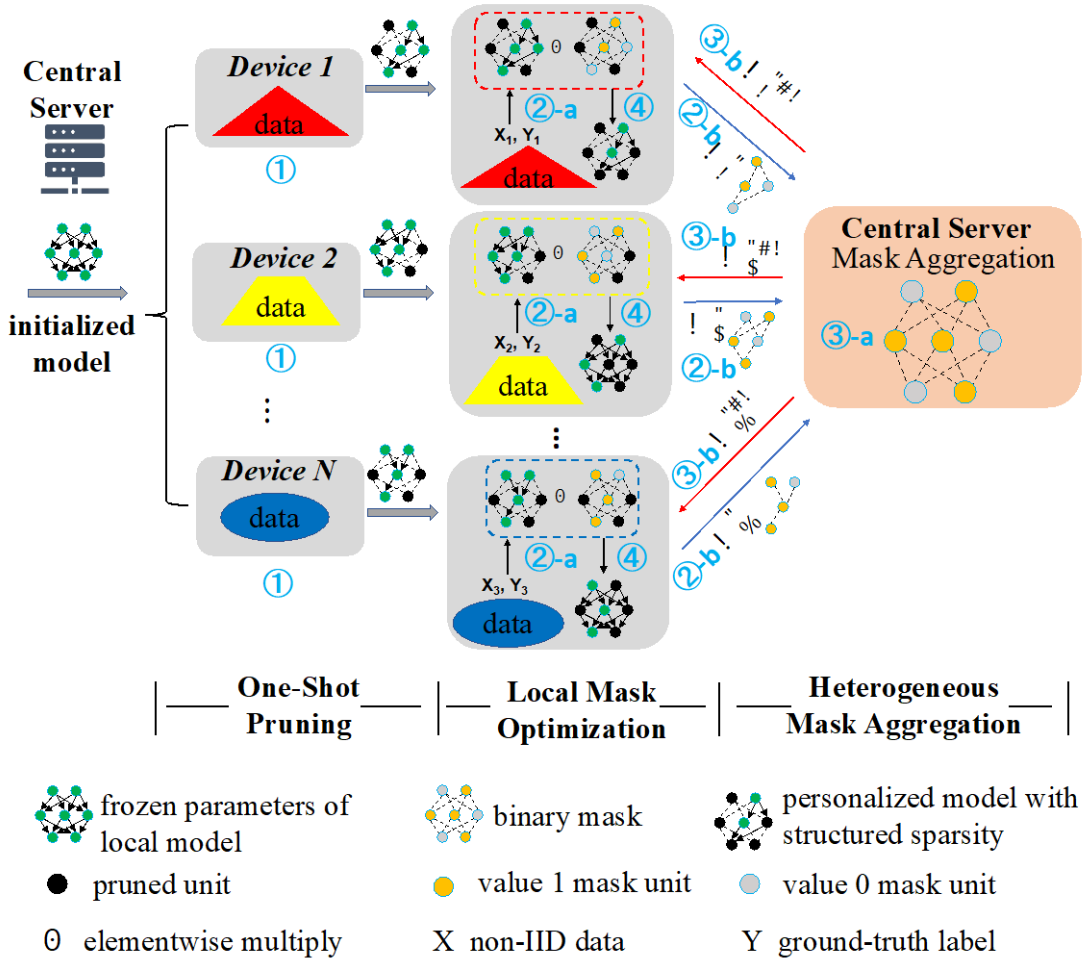




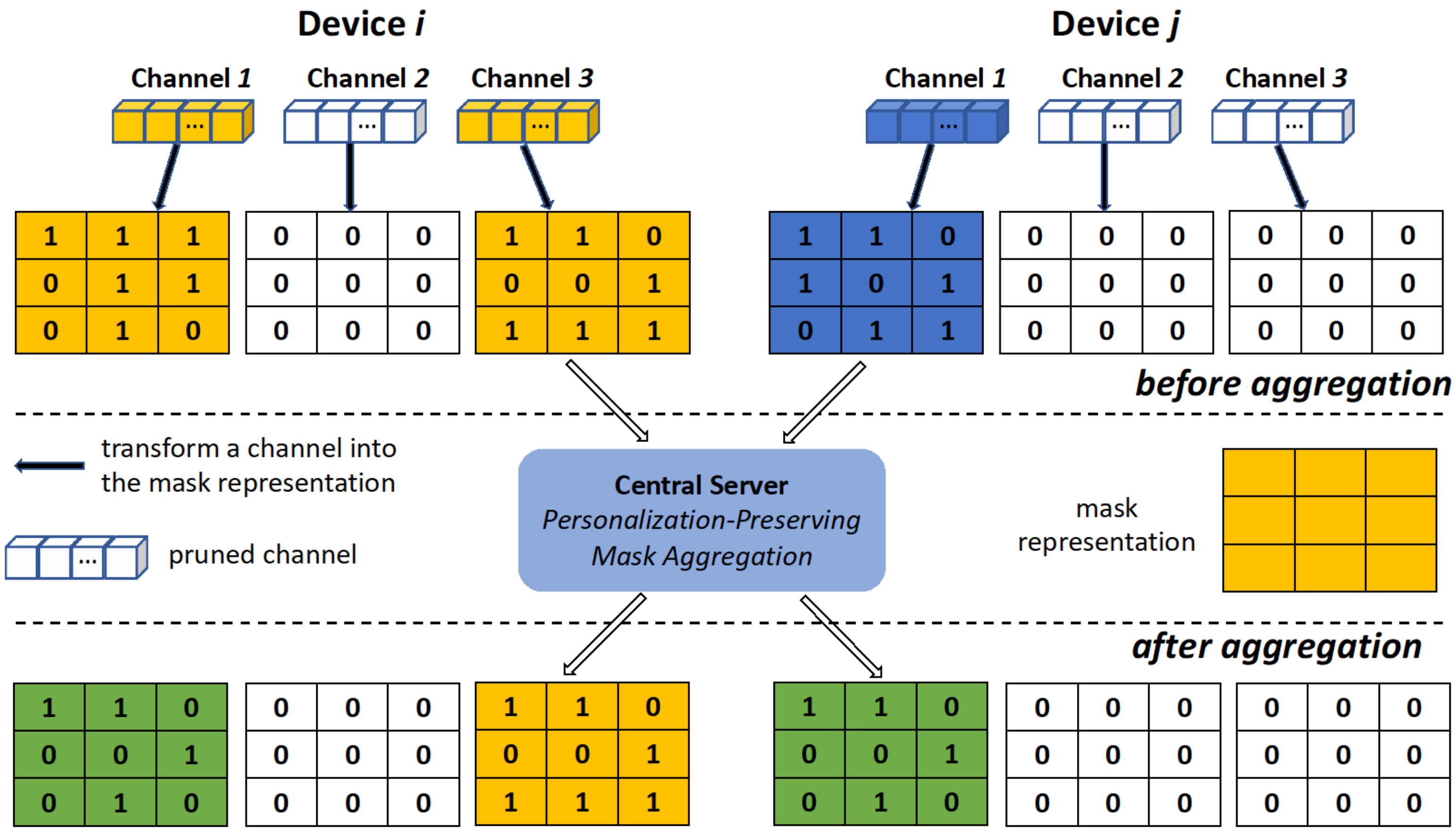
Duke

Design of FedMask

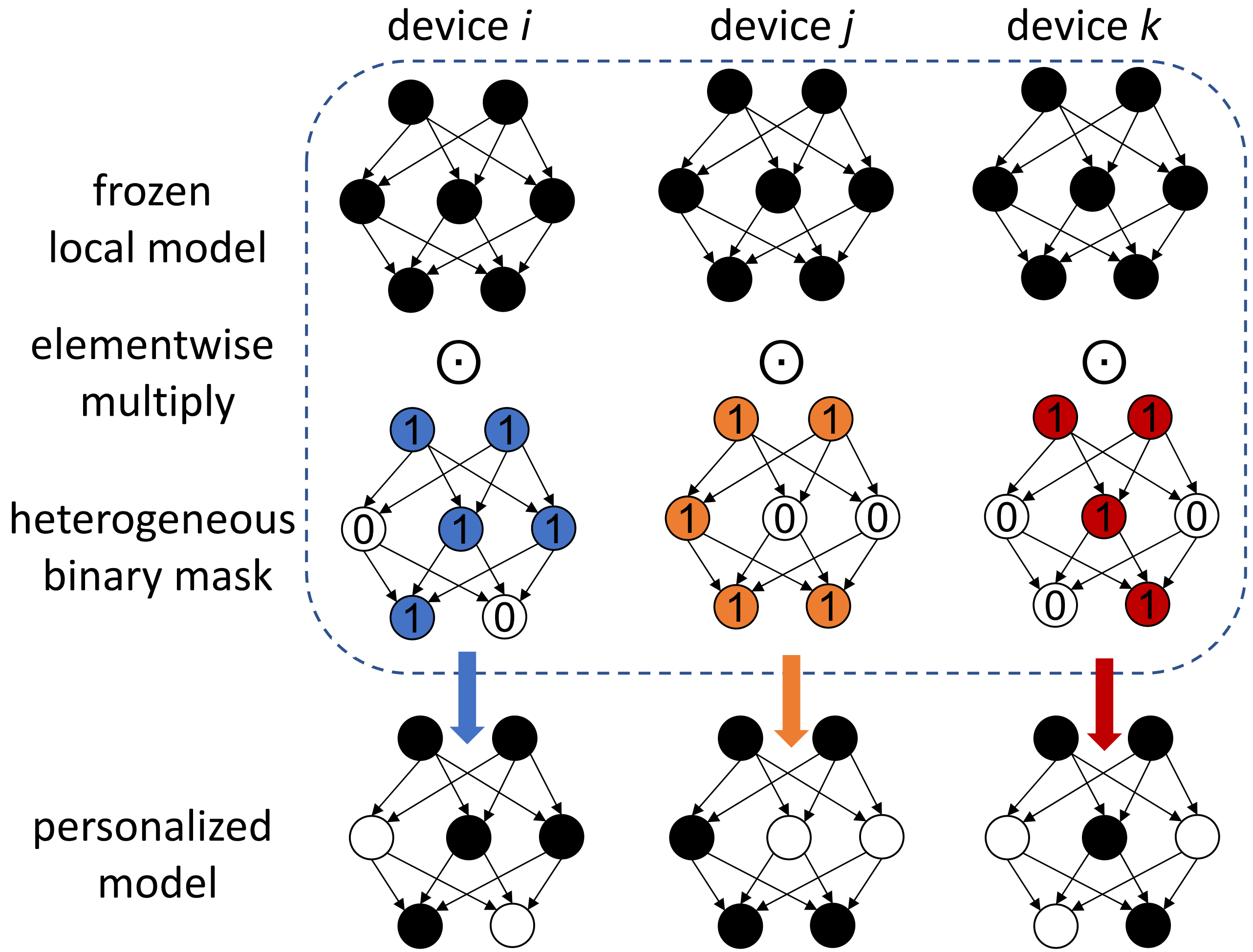
- Learns a heterogeneous and structured sparse binary mask
- Only communicate the binary mask
- The binary mask will be element-wise applied to the frozen parameters to generate a personalized and structured sparse model



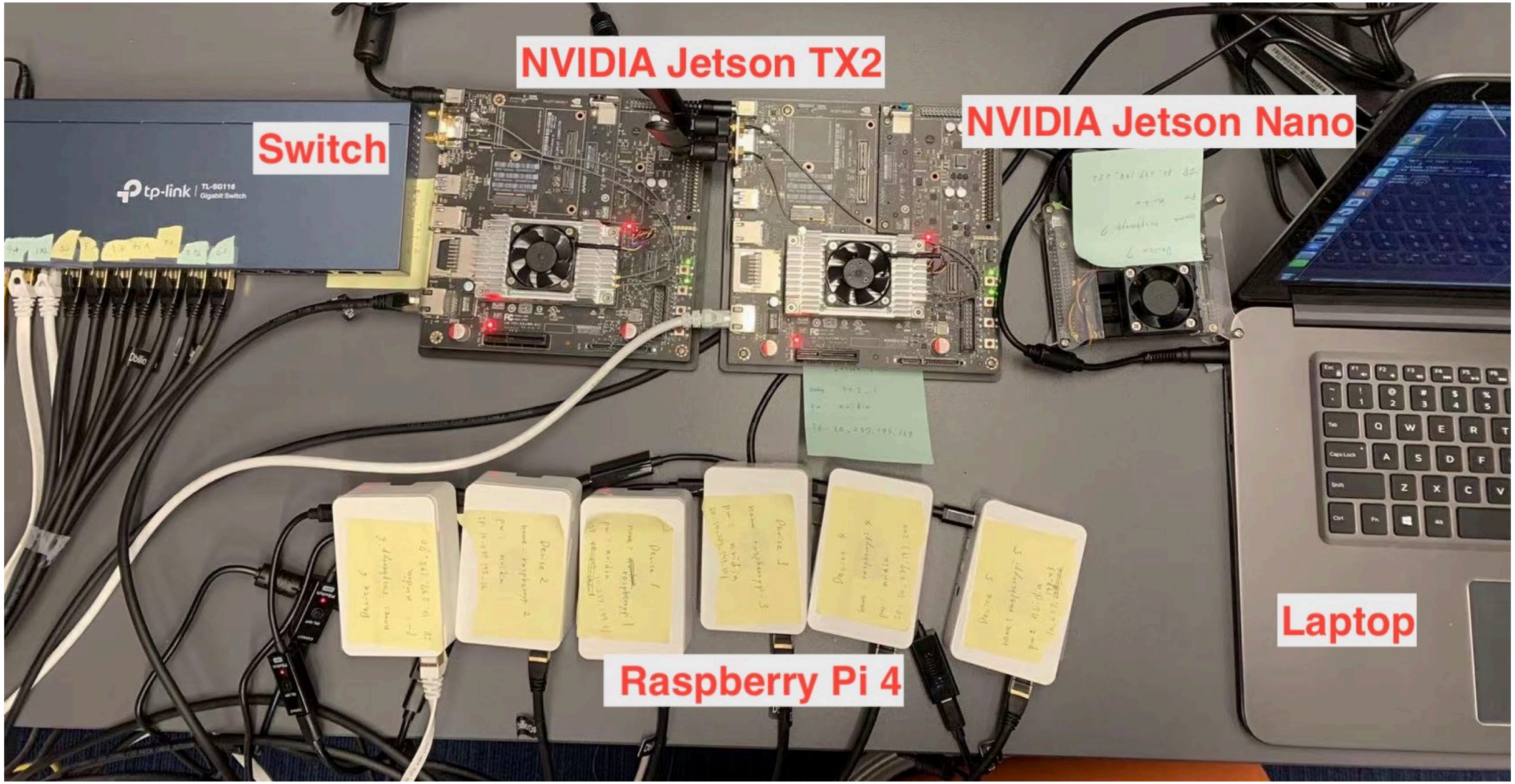
Personalization-Preserving Mask Aggregation



Achieving Personalization via Heterogeneous Masks

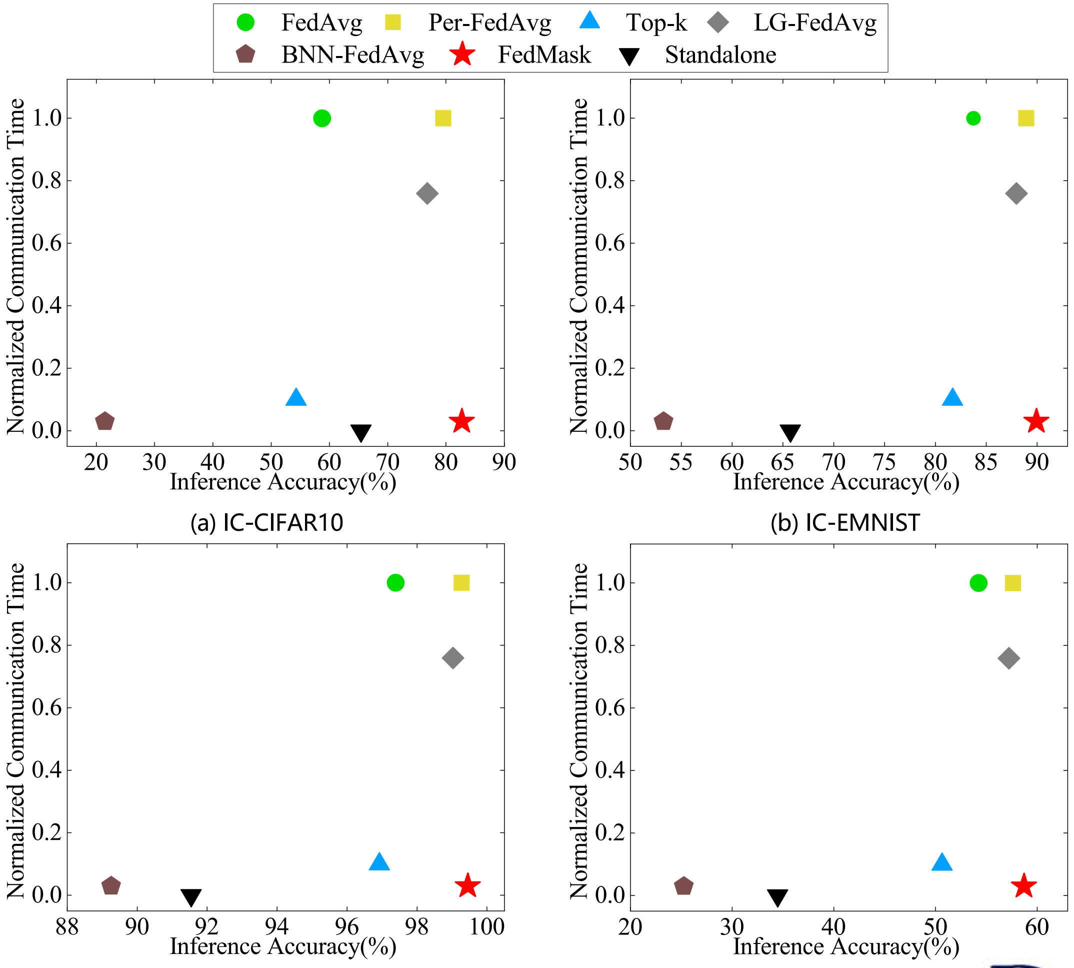


Experiment Setup



Evaluations

- Dataset
  - EMNIST, CIFAR10, HAR, Shakespeare
- Baselines
  - Standalone
  - FedAvg
  - Top-k (communication efficient)
  - BNN-FedAvg (binary neural network+FedAvg)
  - Per-FedAvg (FedAvg+MAML)
  - LG-FedAvg (personalization+communciation)

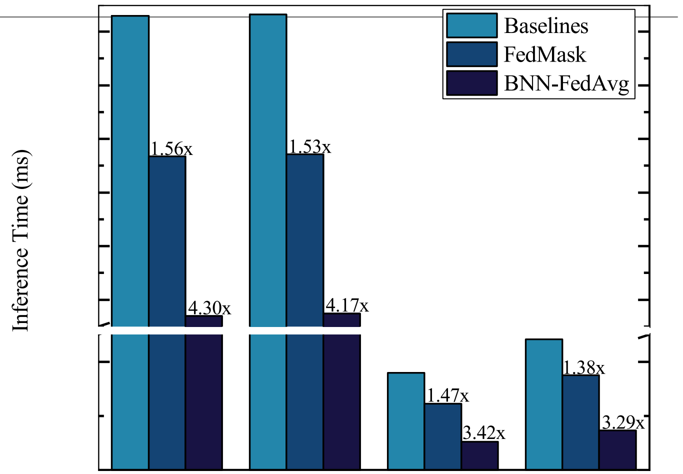


(a) IC-CIFAR10 (b) IC-EMNIST (c) HAR (d) NCP

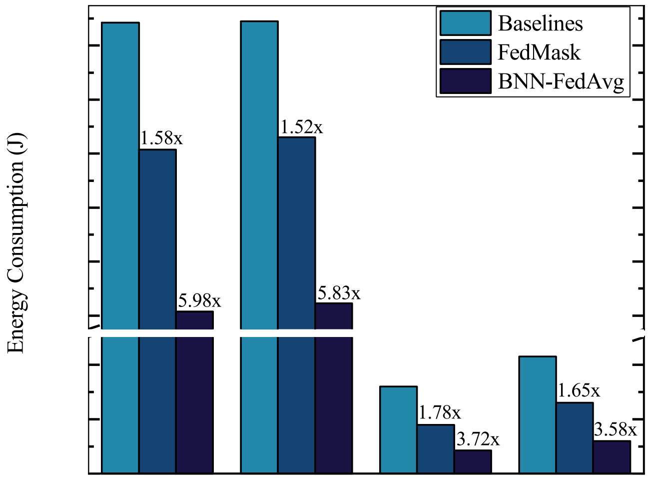
Runtime Performance

Memory Footprint

Application	FedMask Model Size (MB)	Baseline Model Size (MB)	BNN-FedAvg Model Size (MB)
IC-CIFAR10	365.30	537.21	16.78
IC-EMNIST	364.72	538.09	16.82
HAR	2.69	4.41	0.14
NCP	0.92	1.53	0.05
ALL Included	733.63	1081.24	33.79



inference speedup



energy savings

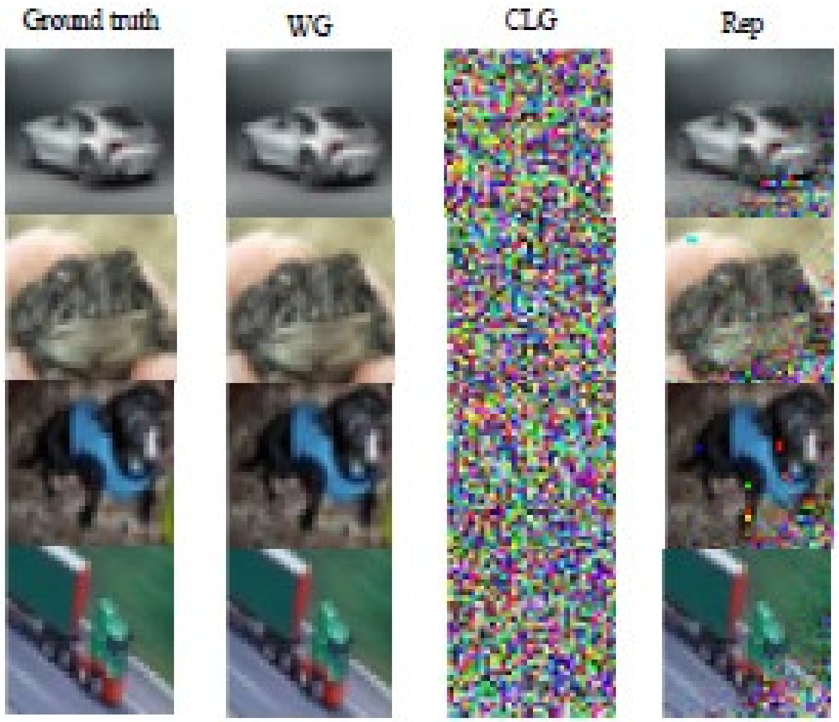
Soteria: Provable Defense against Privacy Leakage in Federated Learning from Representation Perspective (CVPR’21)

Introduction

- Motivations
  - Privacy preserving is the major motivation for proposing federated learning (FL)
  - Recent works demonstrated that sharing model updates or gradients also makes FL vulnerable to inference attack
  - Existing defensive approaches incur either significant computational overheads or unignorable accuracy loss
- Our work
  - Propose a defense approach against model inversion attack in FL based on the observation that the data representation leakage from gradients is the essential cause of privacy leakage in FL
- Key contributions
  - Explicitly reveal the essential cause of leaking private information from the communicated local updates in FL from the perspective of data representations
  - Develop an effective defense against model inversion attack by perturbing data representations

Method

- Data representation leakage in FL
  - Data representations are less entangled in FL
  - Allow us to explicitly reconstruct the input data utilizing the representation of each class on each device from the gradients
  - In practical FL applications, the numbers of batches and local training epochs of each device are both small
  - Reduce the data representation entanglement further



DLG attack results utilizing different parts of gradients.

Method

- Representation perturbation defense
  - Goal 1: To reduce the privacy information leakage, the reconstructed input X' through the perturbed data representations and the raw input X should be dissimilar
  - Goal 2: To maintain the FL performance, the perturbed data representation r' and the true data representations r without perturbation should be similar

Achieving Goal 1: $\max_{r'} ||X - X'||_p,$

Achieving Goal 2: s.t., $||r - r'||_q \leq \epsilon,$

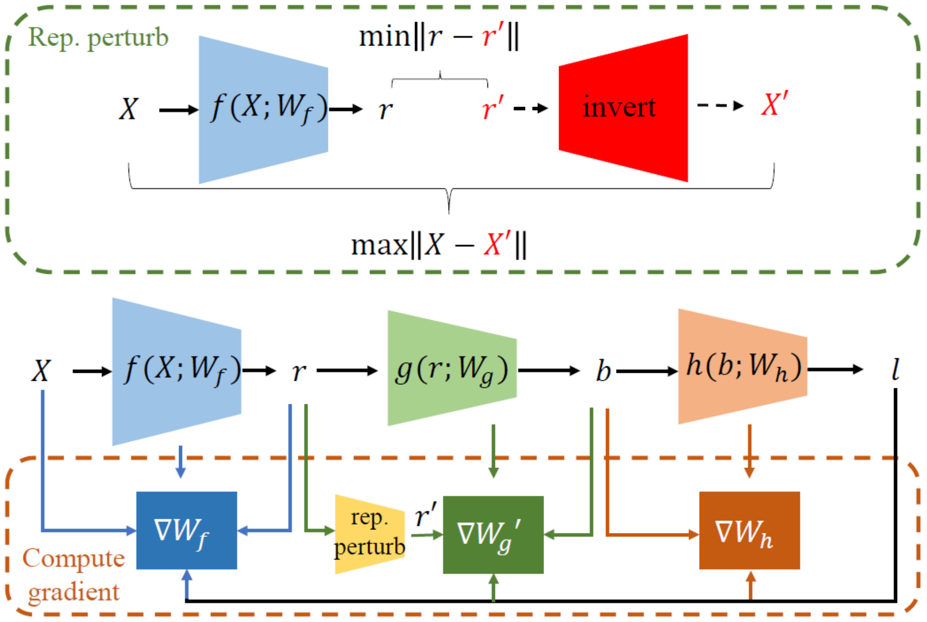


Illustration of our representation perturbation defense.

Method

- Defense Formulation

$$r' = \arg\max_{r'} ||(\nabla_X f)^{-1} \cdot (r - r')||_p, \ s.t. ||r - r'||_q \leq \epsilon$$

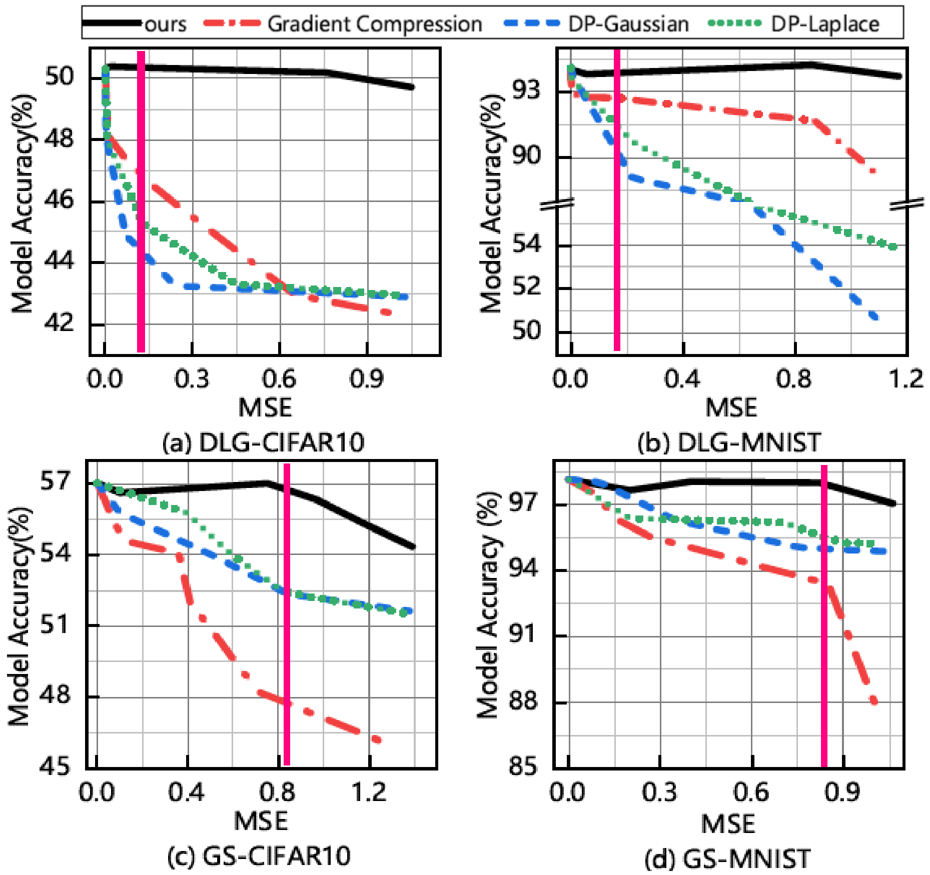
  - Different choices of $||.||_p$ and $||.||_q$ have different defense solutions and thus have different defense effects
  - We set p = 2 to maximize the MSE between the reconstructed input and the raw input. Meanwhile, we set q = 0 due to two reasons: our defense has an analytical solution and is communication efficient

- Certified Robustness Guarantee

$$||X - X'||_p \geq \frac{||r - r'||_p}{||\nabla_X f||_p}.$$

Evaluation

- Dataset:
  - Non-IID CIFAR10
  - Non-IID MNIST
- Attack methods:
  - Deep leakage from gradients (DLG) attack
  - Gradient Similarity (GS) attack
- Defense baselines:
  - Gradient compression (GC)
  - Differential privacy (DP)



Compared defenses on model accuracy and MSE between reconstructed image and raw image.

Duke

FL-WBC: Enhancing Robustness against Model Poisoning Attacks in Federated Learning from a Client Perspective (NeurIPS'21)

Introduction

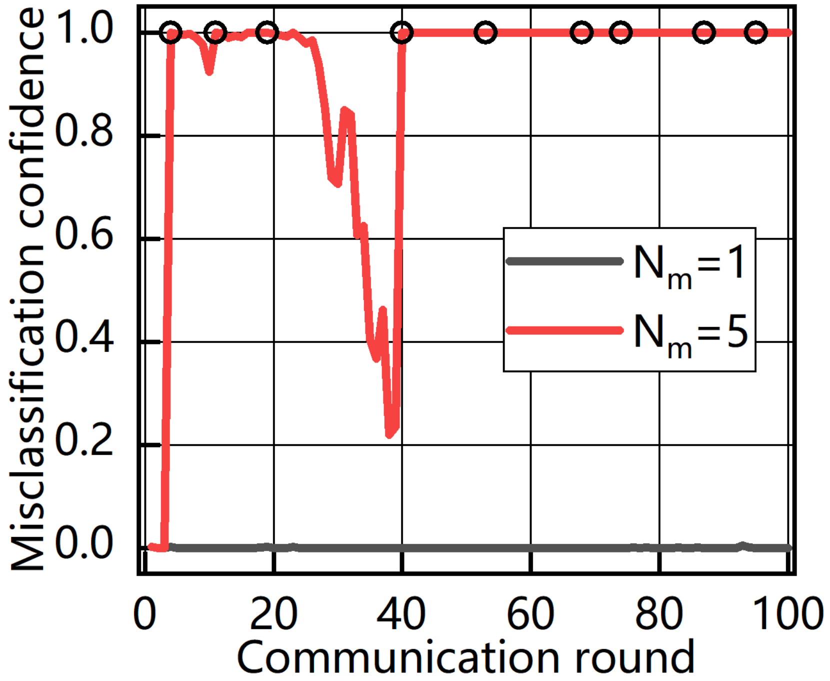
- Motivations
  - Model poisoning attacks fool the global model to produce adversarial misclassification on specific malicious dataset with high confidence
  - Current server-based defenses can not guarantee robustness when the attack is extremely strong
  - When the server-based defenses fail to defend the poisoning attacks, the attack effect will remain in the global model for subsequent rounds even without more attacks occurring.
- Our work
  - Reveal why model poisoning attack effect can persist in the global model for the subsequent rounds, and propose a defense to mitigate the long-lasting model poisoning attacks from a client perspective.
- Key contributions
  - We reveal the reason for the long-lasting effect of a model poisoning attack on the global model
  - Develop an effective defense against model poisoning attack from a client perspective by perturbing the part of the local training gradients where the attack effect resides in

Method

- Long-lasting model poisoning attacks in FL
  - Server-based defenses fail to defend the attacks
  - The attack effect remains in the global model even if no attacks occur in the subsequent rounds
- Attack effect on parameters (AEP)

$$\hat{\delta}_t = \frac{N}{K}[\sum_{k \in \mathbb{S}_t} p^k \prod_{i=0}^{I-1} (\boldsymbol{I} - \eta_{t,i} \boldsymbol{H}_{t,i}^k)]\hat{\delta}_{t-1}$$

  - The long-lasting attack effect resides in the kernel of hessian matrix during local training.



Misclassification confidence of the global model on the malicious data point applying Coordinate Median Aggregation.

Duke

Method

- FL-WBC: a client-based defense
  - Each client acts like a white blood cell in the FL system, i.e., mitigates the poisoning attack effect that is not defended by the server during aggregation.
  - Goal 1: To maintain the benign task's performance, loss of local benign task should be minimized.
  - Goal 2: To prevent AEP from being hidden in the kernel of Hessian matrices on benign devices, the rank of Hessian matrices should be maximized.

Achieving Goal 1: $\min_{W} F^k(W)$,

Achieving Goal 2: $\max_{W} \left\| ReLU\left(\left|\left(W - W^k_{t,i}\right) - \Delta W^k_{t,i}\right| / \eta_{t,i} - |\Upsilon|\right) \right\|_0$

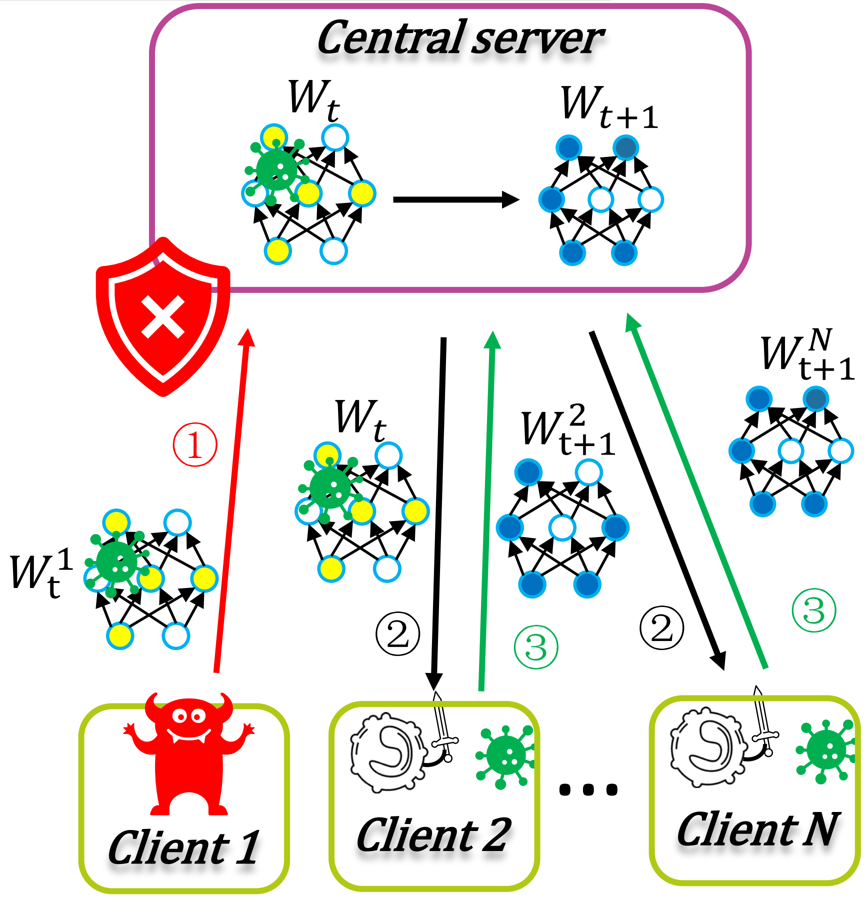
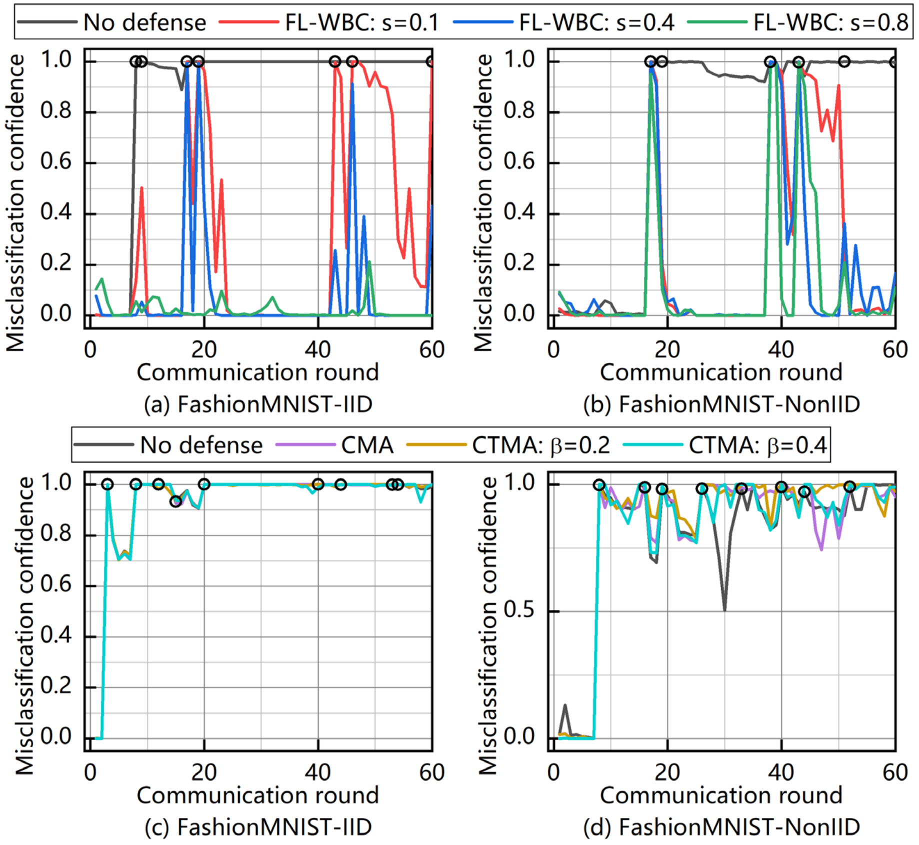


Illustration of FL-WBC.

Evaluation

- Dataset:
  - FasionMNIST
  - CIFAR10 (results not shown here)
- Important Hyperparameters:
  - 10 clients participate in training for each round
  - 5 malicious attackers in adversarial rounds
- Defense baselines:
  - Coordinate Median Aggregation (CMA)
  - Coordinate Trimmed-Mean Aggregation (CTMA)
  - Local Differential Privacy (LDP)
  - Central Differential Privacy (CDP)



Misclassification confidence of the global model on the malicious data point.

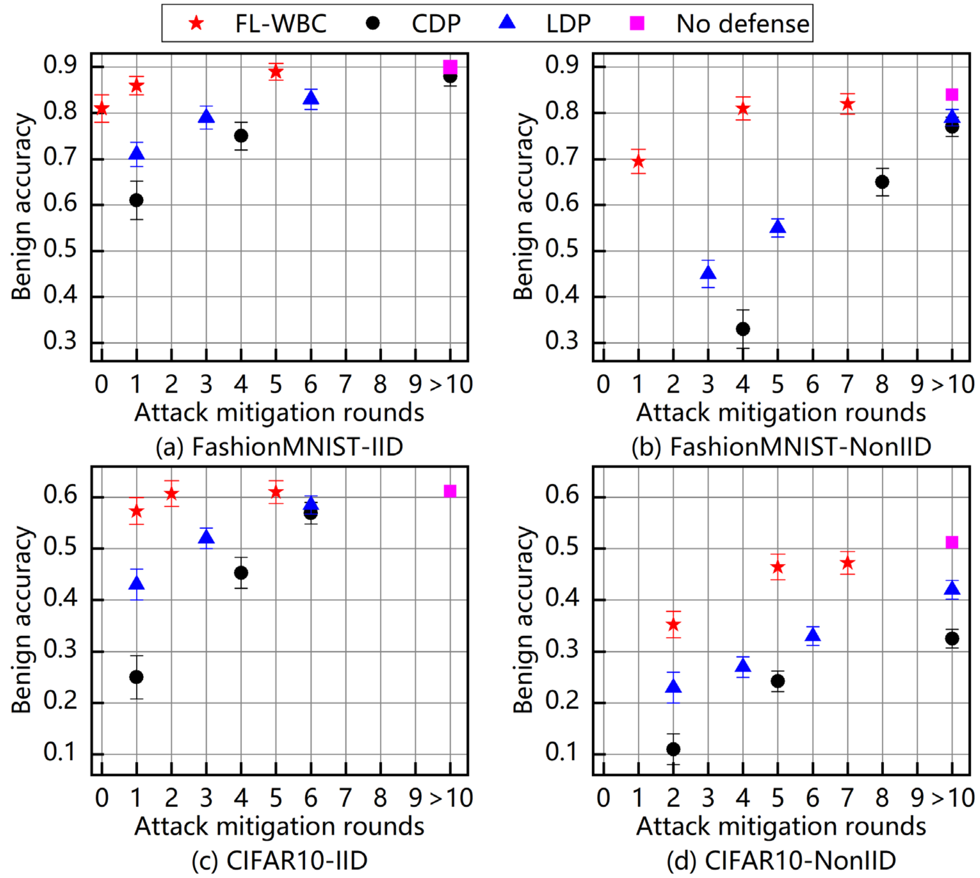
Duke

Evaluation

- Standard deviation of noise

  $s \in [0.1,1]$
  $\sigma_{LDP} \in [0.1, 1]$
  $\sigma_{CDP} \in [0.1,10]$

- FL-WBC only inject perturbations to the parameter space where the long-lasting AEP resides in instead of perturbing all the parameters like DP methods.



Benign accuracy vs. Attack mitigation rounds.

Duke

Recap

- Privacy-Preserving Graph learning
  - Privacy-Preserving Representation Learning on Graphs: Preserve node/link privacy by minimizing the information of node/link variables kept in embeddings.
- Efficient and Heterogeneity-Aware Federated Learning
  - LotteryFL: Realize personalization and communication efficiency by seeking and optimizing Lottery Ticket Networks (LTNs) of each device.
  - FedMask: Improve communication efficiency tremendously by optimizing and transmitting binary masks.
- Privacy-Enhancing and Robust Federated Learning
  - Soteria: Reveal how privacy is leaked through the representations embedded in the gradients and propose a defense against the privacy leakage by perturbing representations.
  - FL-WBC: Reveal why model poisoning attack effect can be long-lasting in the global model and design a client-based defense to mitigate such long-lasting attack effect.

Duke

Athena: AI Institute for Edge Computing Leveraging Next-generation Networks

- **Athena Institute capitalizes and responds to these challenges by advancing Artificial Intelligence (AI) technologies to transform the design, operation, and service of future mobile networks.**
- Athena is a multi-university and trans-disciplinary AI center including **seven academic institutions** (Duke, Yale, Wisconsin, Michigan, Princeton, MIT, and N.C. A&T State University); and **five industry collaborators** (AT&T, Microsoft, Motorola Solutions, EdgeMicro and 5NINES).
- The research activities of Athena are organized under four synergistic thrusts: **Networking**, **Computer Systems**, **AI**, and **Services**.
- More info: https://athena.duke.edu

ATHENA

Duke