UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Federated Learning and Analysis In Multi-Access Edge Computing

Zhu Han,
John and Rebecca Moores Professor, IEEE Fellow, AAAS Fellow
Department of Electrical and Computer Engineering
University of Houston, TX, USA

Thanks to Dawei Chen, Latif U. Khan, Choong Seon Hong, Nguyen Tran, Lixin Li, Minh Nguyen, Tra Huong Thi Le, Chunxiao Jiang, Yue Yu, Zibo Wang, Wenbo Wang, Yifei Zhu, Siping Shi, and Dan Wang, supported by National Science Foundation, Toyota, and Amazon

Outline

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

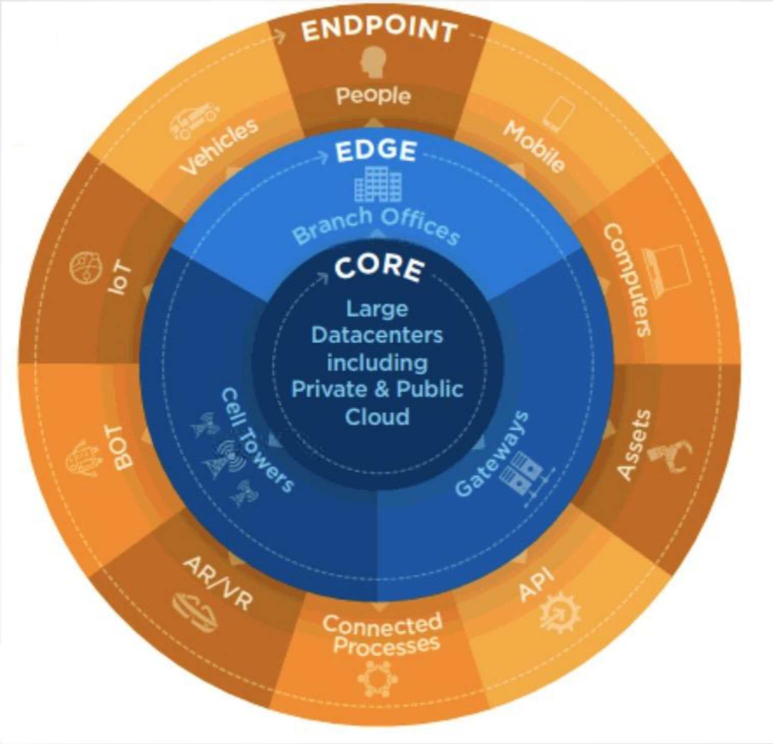
- Background and Fundamentals of Federated Paradigm
  - Background
  - Machine Learning (ML) Point of View
  - Optimization Point of View
- Federated Learning for Wireless Networks
  - Unsupervised Federated Learning for Unbalanced Data
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - Federated Skewness Analytics in Heterogeneous Decentralized Data Environments
  - Federated Anomaly Analytics for Local Model Poisoning Attack
- Open Problems and Conclusions

Background

- Can data live at the edge?
  – Billions of phones & IoT devices constantly generate data
  – Data processing is moving on device:
    - ➢ Improved latency
    - ➢ Works offline
    - ➢ Better battery life
    - ➢ Privacy advantages

What about analytics?

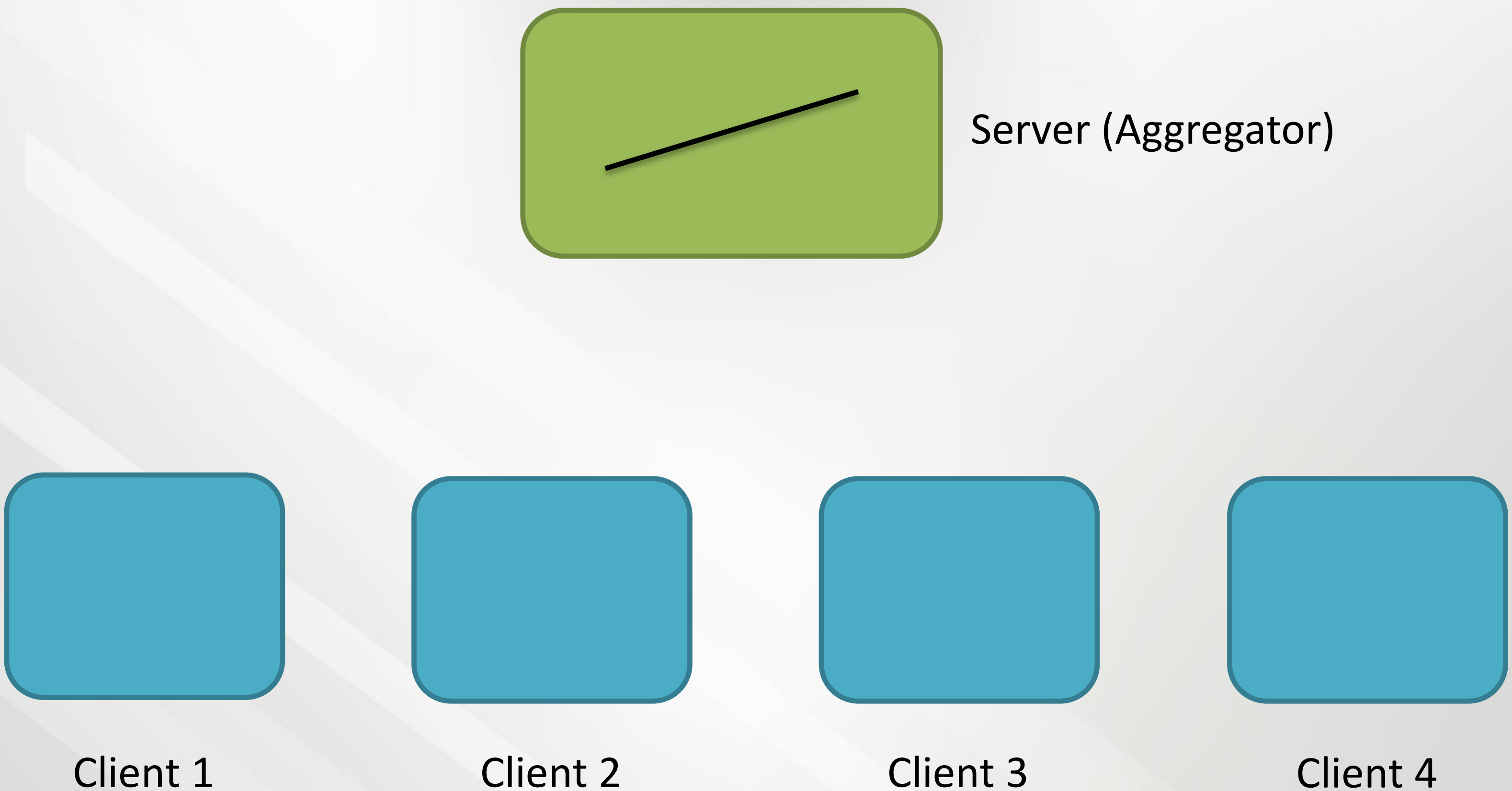
What about learning?



Sources: D. Reinsel, J. Gantz, and J. Rydning, "The digitization of the world from edge to core," IDC White Paper, 2018.

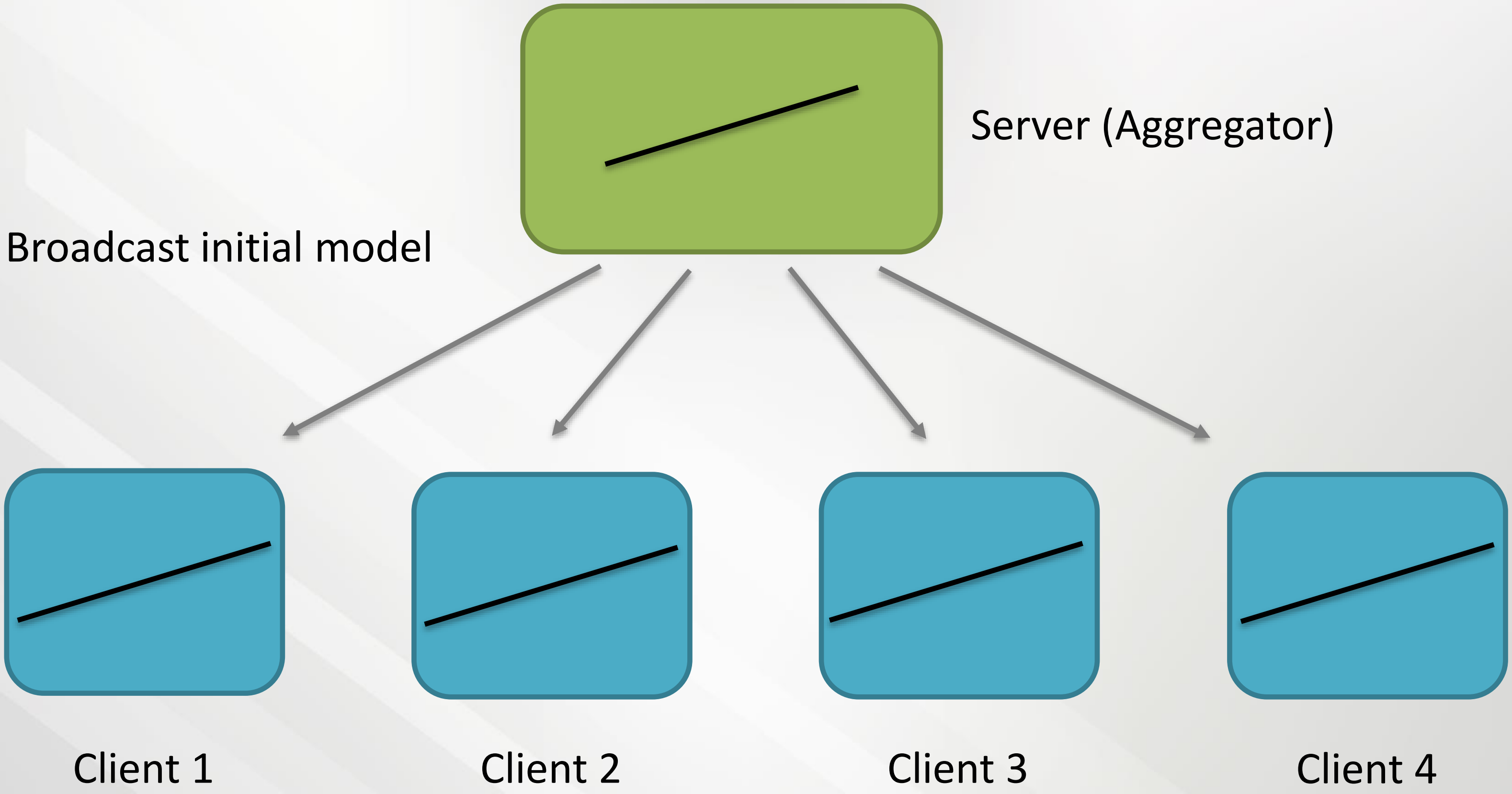
ML Point of View

- What is Federated Learning?
  - General workflow



ML Point of View

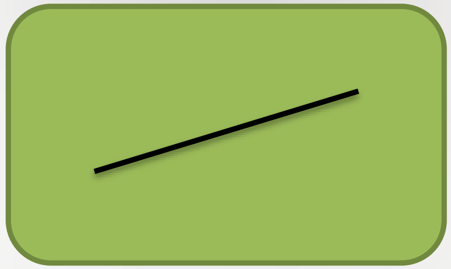
- What is Federated Learning?
  - General workflow



ML Point of View

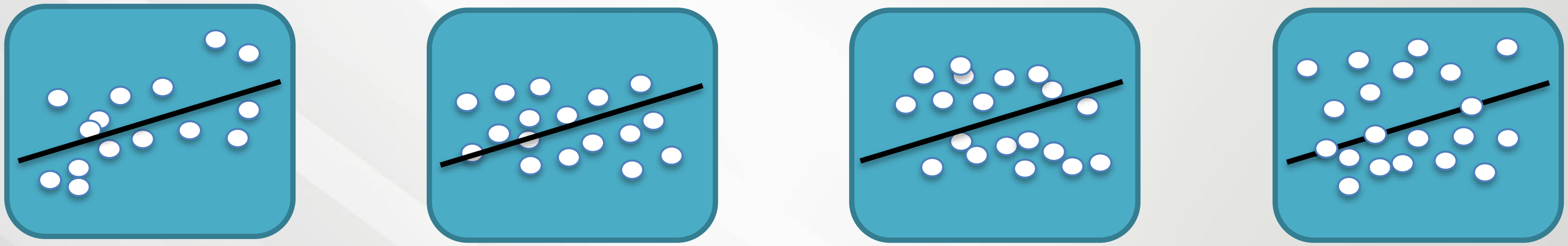
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- What is Federated Learning?
  - General workflow



Server (Aggregator)

Clients generate local data

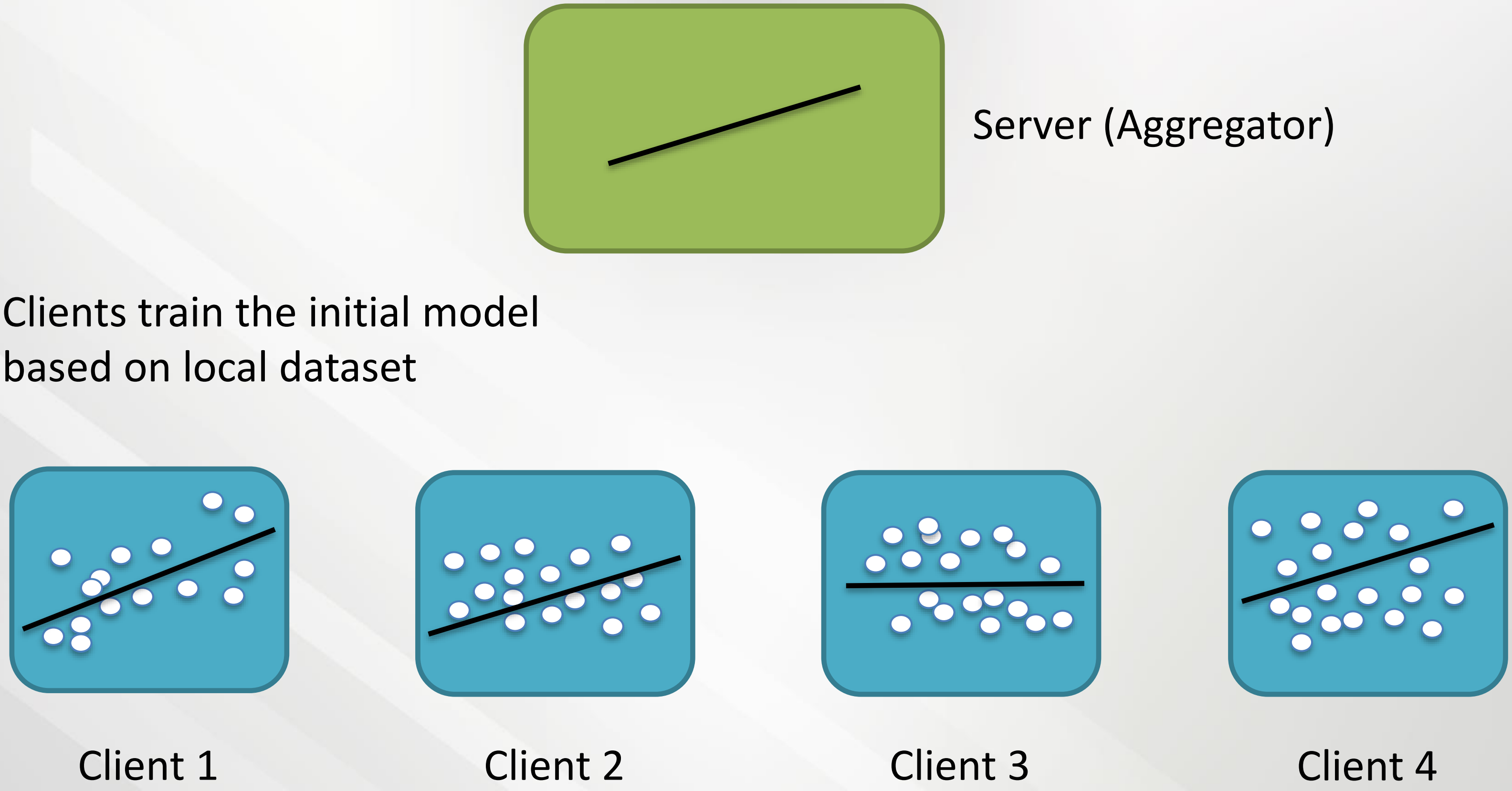


Client 1 Client 2 Client 3 Client 4

ML Point of View

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

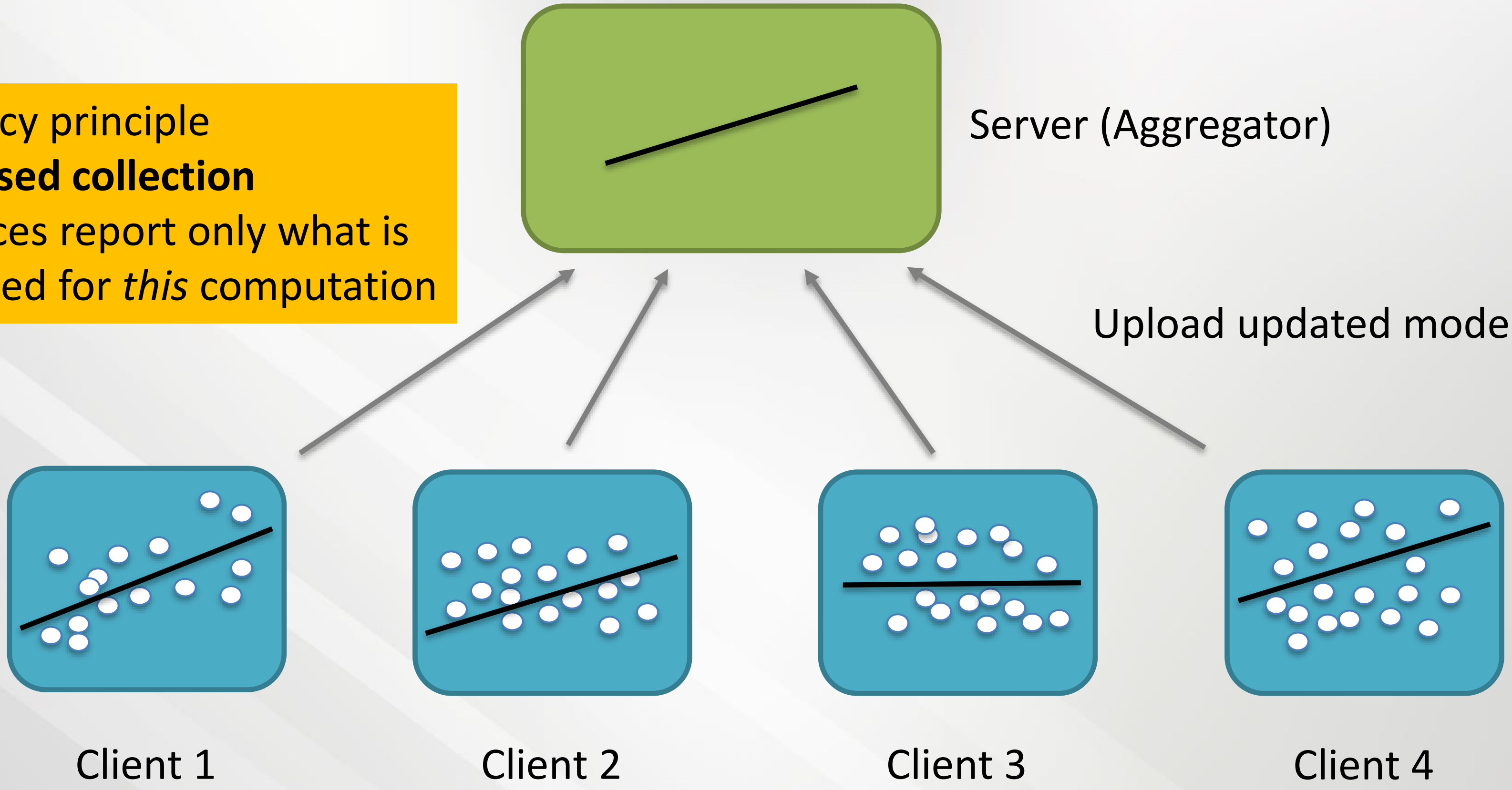
- What is Federated Learning?
  - General workflow



ML Point of View

- What is Federated Learning?
  - General workflow

Privacy principle
Focused collection
Devices report only what is needed for *this* computation



ML Point of View

- What is Federated Learning?
  - General workflow

Combine in

or)

Repeat these pr
convergence

Client 1 | Client 2 | Client 3 | Client 4

Advantages

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Advantages:

 1. Generally, the data generated by different users are non-i.i.d. data due to the various behavior characteristics. However, the task aims at obtaining a model that is suitable for each individual user. FL has been proved to be an effective way to tackle with non-i.i.d. data [1], which is perfectly suitable for multi-user scenario.

 2. Communication cost can be easily relieved by FL because what are transmitted between edge devices and datacenter are the machine learning model or the model parameters, whose data size is greatly smaller than the original dataset [2].

 3. In addition, because the original data will not be uploaded, FL is an effective way to reduce the probabilities of eavesdropping, which means the user's privacy can be ensured [3].

[1]. Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, “Federatedlearning with non-iid data,”arXiv preprint arXiv:1806.00582, 2018.
[2]. J. Koneˇcn`y, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, andD. Bacon, “Federated learning: Strategies for improving communicationefficiency,”arXiv preprint arXiv:1610.05492, 2016.
[3]. R. C. Geyer, T. Klein, and M. Nabi, “Differentially private federatedlearning: A client level perspective,” inthe 31st Conference on NeuralInformation Processing Systems, Long Beach, CA, December 2017.

Optimization POV

- Characteristics (Major challenges)

  - Non-IID
    - ✓ The data generated by each user are quite different
  - Unbalanced
    - ✓ Some users produce significantly more data than others
  - Limited communication
    - ✓ Unstable mobile network connections
  - Massively distributed
    - ✓ # mobile device owners >> avg # training samples on each device

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Essentially, FL aims at collaboratively obtain a global machine learning model for N users.

- Individually, each participant perform local training process to optimize its own model

$$\min_{\omega_i \in \mathbb{R}^m} f_i\left(x_i, \omega_i; y_i\right)$$

- Server aggregates these local models

Global weight

$$\omega^s = \frac{\sum_{i=1}^{N} D_i \omega_i^T}{D}$$

Local data size and weight

Total data size

- The federated objective function

$$\min_{\omega \in \mathbb{R}^m} J(\omega^s) = \frac{1}{N} \sum_{i=1}^{N} f_i(\omega_i^T)$$

Number of clients

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

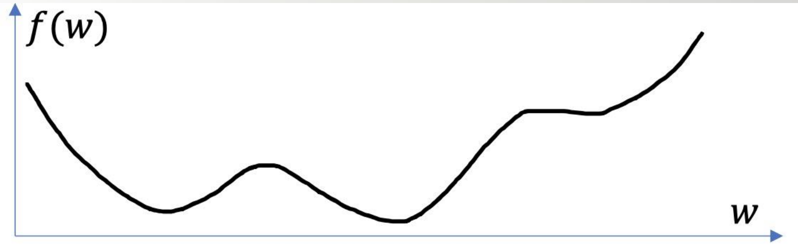
- Recall - deep learning training method

- For a training dataset containing n samples (x_i, y_i), $1 \leq i \leq n$, the training objective is:

$$\min_{w \in \mathbb{R}^d} f(w) \qquad \text{where } f(w) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} f_i(w)$$

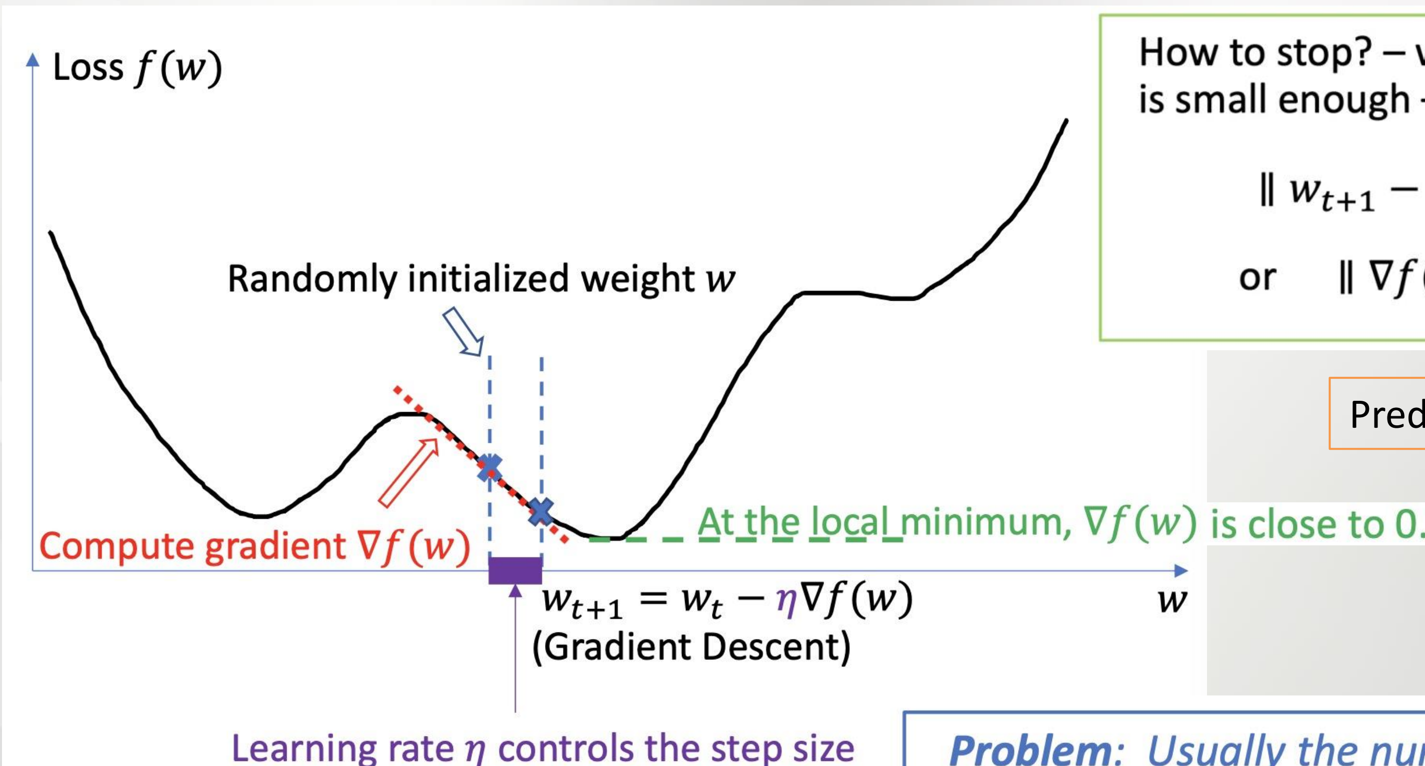
- $f_i(w) = l(x_i, y_i, w)$ is the loss of the prediction on example (x_i, y_i).

- **No closed-form solution**: in a typical deep learning model, w may contain millions of parameters.
- **Non-convex**: multiple local minima exist.



Optimization POV

- Recall – gradient descent



How to stop? – when the update is small enough – converge.

$$\| w_{t+1} - w_t \| \leq \epsilon$$

or $$\| \nabla f(w_t) \| \leq \epsilon$$

Predefined threshold

***Problem**: Usually the number of training samples n is large – slow convergence*

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Recall – stochastic gradient descent

- At each step of gradient descent, instead of compute for all training samples, randomly pick a small subset (mini-batch) of training samples (x_k, y_k)

$$w_{t+1} \leftarrow w_t - \underbrace{\eta}_{\text{Learning rate}} \nabla f(w_t; x_k, y_k)$$

- Compared to gradient descent, SGD takes more steps to converge, but each step is much faster.

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Baseline solution for FL – FedSGD
- In a round t:
  - The central server broadcasts current model w_t to each client; each client k computes gradient: $g_k = \nabla F_k(w_t)$, on its local data.
    - ✓ Approach 1: Each client k submits g_k; the central server aggregates the gradients to generate a new model:

$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w)$$

$$w_{t+1} \leftarrow w_t - \eta \nabla f(w_t) = w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k.$$

    - ✓ Approach 2: Each client k computes: $w_{t+1}^k \leftarrow w_t^k - \eta g_k$; the central server performs aggregation:

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

For multiple times $\Longrightarrow$ Federated Averaging (FedAvg)

Optimization POV

- Federated learning - deal with limited communication
  - Due to the enormous number of end devices and limited bandwidth, the communication cost dominates the federated learning process

- Increase computation
  - ✓ Select more clients for training between each communication round
  - ✓ Increase computation on each client

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Federated Averaging (FedAvg)
- In a round t
  - The central server broadcasts current model w_t to each client; each client k computes gradient: $g_k = \nabla F_k(w_t)$, on its local data.
    - ✓ Approach 2:
      - ◆Each client k computes for E epochs: $w_{t+1}^k \leftarrow w_t^k - \eta g_k$
      - ◆The central server performs aggregation: $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$
      - ◆Suppose B is the local mini-batch size, #updates on client k in each round: $u_k = E\frac{n_k}{B}$

Optimization POV

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Federated Averaging (FedAvg)

Algorithm 1 `FederatedAveraging`. The K clients are indexed by k; B is the local minibatch size, E is the number of local epochs, and η is the learning rate.

Server executes:
initialize w_0
for each round $t = 1, 2, \dots$ **do**
  $m \leftarrow \max(C \cdot K, 1)$
  $S_t \leftarrow$ (random set of m clients)
  for each client $k \in S_t$ **in parallel do**
    $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$
  $w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$

ClientUpdate(k, w)**:** *// Run on client k*
$\mathcal{B} \leftarrow$ (split $\mathcal{P}_k$ into batches of size B)
for each local epoch i from 1 to E **do**
  for batch $b \in \mathcal{B}$ **do**
    $w \leftarrow w - \eta \nabla \ell(w; b)$
return w to server

Overall procedures:

1. At first, a model is randomly initialized on the central server.
2. For each round *t:*
   i. *A random set of clients are chosen;*
   ii. *Each client performs local gradient descent steps;*
   iii. *The server aggregates model parameters submitted by the clients.*

Toyota/Amazon Project

Outline

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Background and Fundamentals of Federated Paradigm
  - Background
  - Machine Learning (ML) Point of View
  - Optimization Point of View
- Federated Learning for Wireless Networks
  - Unsupervised Federated Learning for Unbalanced Data
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - Federated Skewness Analytics in Heterogeneous Decentralized Data Environments
  - Federated Anomaly Analytics for Local Model Poisoning Attack
- Open Problems and Conclusions

Motivation

- Unsupervised federated learning framework

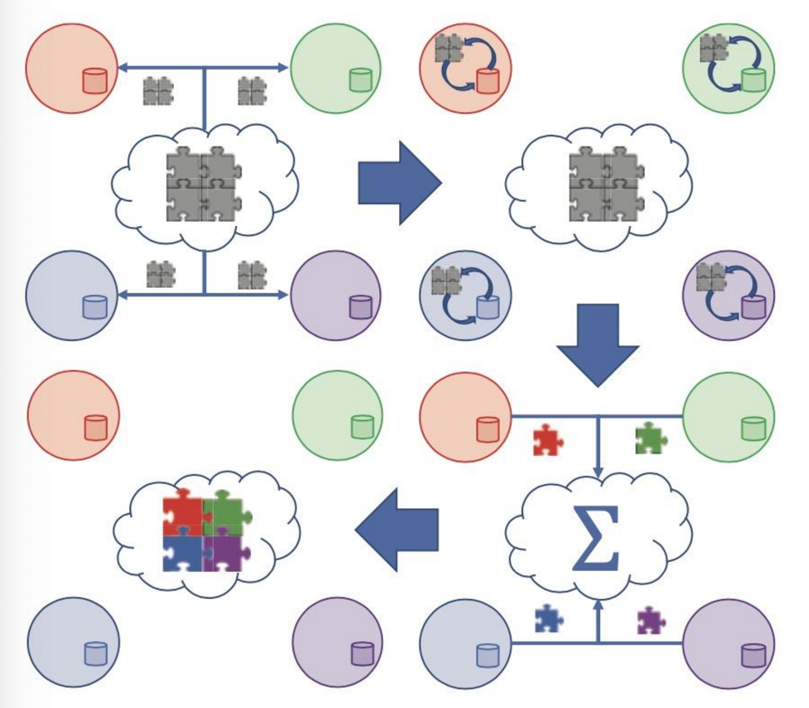


Fig. 1. Unsupervised federated learning procedure.

Challenges:
the accuracy of federated learning training reduces significantly when the data is nonuniformly distributed across devices.

Mykola Servetnyk, Carrson C. Fung, and Zhu Han, ``Unsupervised Federated Learning for Unbalanced Data," IEEE Globecom.

Formulation

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Goal: assign each observation point to a particular cluster and estimate the cluster centroid

$\mathcal{J}$: agents set
$\mathcal{K}$: class set
$\mathcal{N}$: data set

Centroid

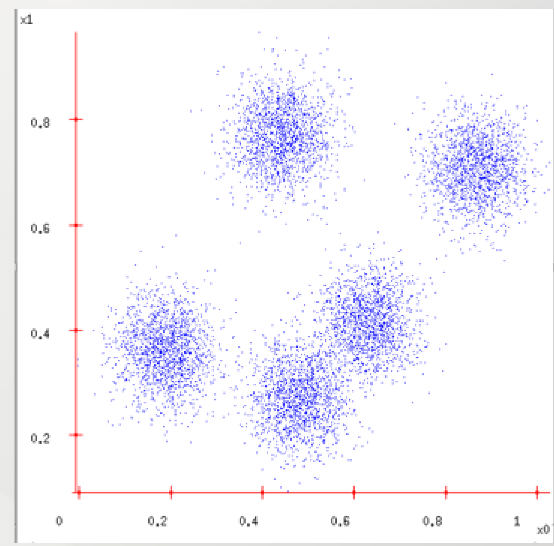
Data sample

$$\min_{\mu_{jnk},\mathbf{m}_k} \sum_{j\in\mathcal{J}} \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}_j} \frac{1}{2}\mu_{jnk}\|\underline{\mathbf{m}_k} - \underline{\mathbf{x}_{jn}}\|^2$$

$$\text{s.t.} \sum\nolimits_{j\in\mathcal{J}} \mu_{jnk} = 1,\ k\in\mathcal{K}, n\in\mathcal{N}_j,$$

$$\underline{\mu_{jnk}} \in \{0,1\}, j\in\mathcal{J}, k\in\mathcal{K},\ n\in\mathcal{N}_j$$

Class membership indicator



Methodology

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Dual Averaging
  - Step 1: Data labeling

$$\mu_{jnk} = \begin{cases} 1, \text{if } k = \arg\min\, (1 + \frac{\xi}{t_1}) \|\mathbf{x}_{jn} - \mathbf{m}_k^{(t_1)}\| \\ 0, \text{otherwise}, \end{cases}$$

    where ξ is random variable drawn from uniform distribution $\xi \sim \mathcal{U}(0; \xi_{\max})$ between 0 and $\xi_{\max}$. $\frac{\xi}{t_1} \|\mathbf{x}_{jn} - \mathbf{m}_{jk}^{(t_1)}\|$

  - Step 2: DA-based centroid computation
    - Each node calculates a gradient $\mathbf{g}_{jk}^{(t_2)} = \sum_{n \in \mathcal{N}_j} \mu_{jnk} (\mathbf{m}_k^{(t_2)} - \mathbf{x}_{jn})$
    - Centroid update

$$\mathbf{z}_k^{(t_2+1)} = \mathbf{z}_k^{(t_2)} + \sum_{j \in \mathcal{J}} [\mathbf{w}]_j \mathbf{g}_{jk}^{(t_2)},$$

accumulated gradient for cluster k at iteration t_2

$$\mathbf{m}_k^{(t_2+1)} = \underline{\arg\min\, \langle \mathbf{z}_k^{(t_2+1)}, \mathbf{m}_k \rangle} + \alpha^{(t_2)} \underline{\|\mathbf{m}_k\|^2}$$

first-order approximation of the objective

regularization term

Methodology

- Dual Averaging
  - Step 3(a): Weight computation via bin method
    - assign the weights by dividing the data space into a grid with uniform-sized bins and calculate the number of points (as weight) falling into a particular bin (a region of the grid) at each node.

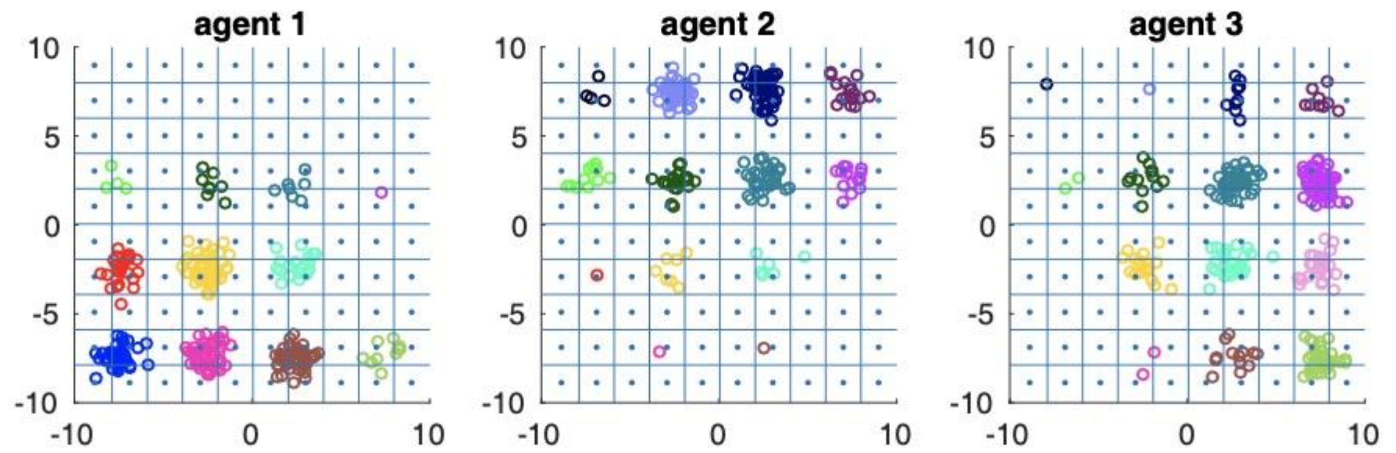


Fig. 2. Unbalanced data sets observed at different agents.

Methodology

- Dual Averaging
  - Step 3(b): Weight computation via self-organizing maps
    - All neurons are initialized with small values of their weights.

    1. For each data point, the neurons compute the distance to the data point and the closest neuron is declared as the winner.
    2. The winning neuron determines the neighborhood of excited neurons and these neurons adjust their individual weights towards the data point.
    3. Neurons decrease neighborhood radius and learning rate.

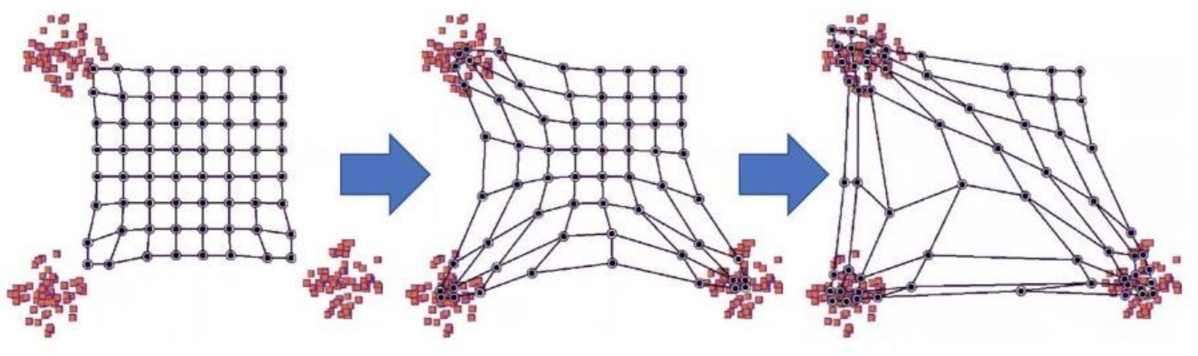


Fig. 3. Self-organizing maps training.

Repeat until convergence

Simulation Results

- Data are generated at random from K = 16 classes, with vectors from each class generated from a symmetric Gaussian distributions.

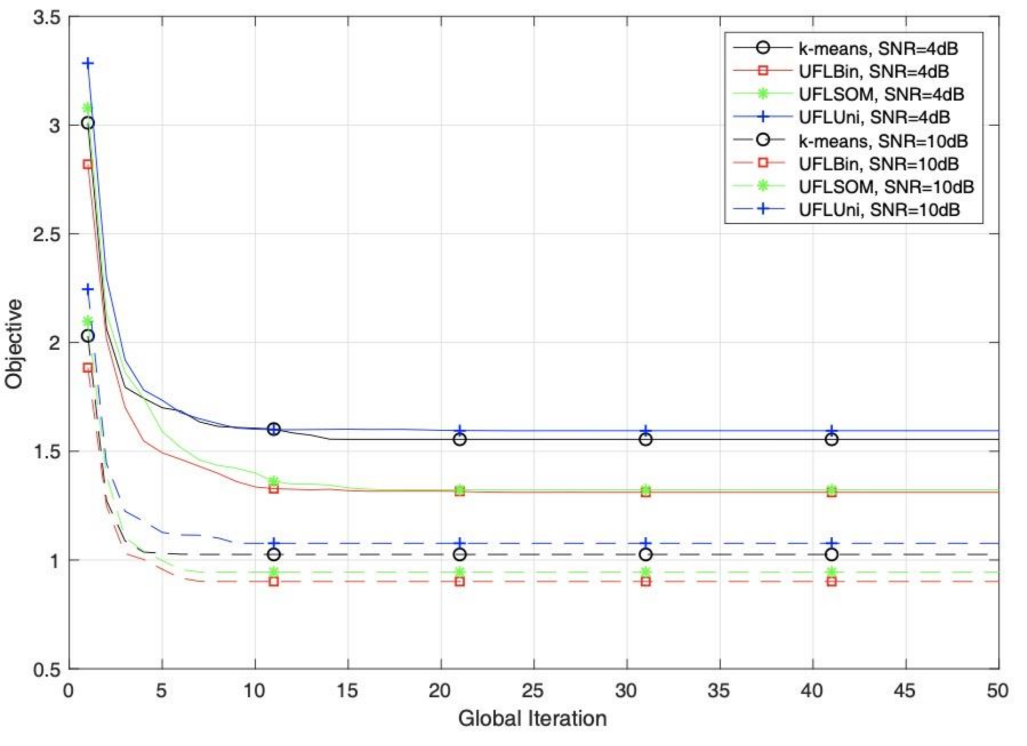


Fig. 4. Convergence of the proposed algorithms.

UFLBin: Bin based DA
UFLSOM: SOM based DA
UFLUNI: uniform gradient weighting
K-means as baseline

The proposed methods close optimal to centralized k-means.

Outline

- Background and Fundamentals of Federated Paradigm
  - Background
  - Machine Learning (ML) Point of View
  - Optimization Point of View
- Federated Learning for Wireless Networks
  - Unsupervised Federated Learning for Unbalanced Data
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - Federated Skewness Analytics in Heterogeneous Decentralized Data Environments
  - Federated Anomaly Analytics for Local Model Poisoning Attack
- Open Problems and Conclusions

Motivation

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Challenges:
  – Once the end devices are invited, they will unconditionally take part in the federated learning tasks which ignores their willingness.
    - Computation cost, remained energy…
  – There are many available edge nodes in a MEC network, how to parallelly perform multiple federated learning tasks needs to be considered.
  – Information exchanging cannot be done entirely in large scale IoTs scenarios.
  – Matching Game Framework with incomplete preference list

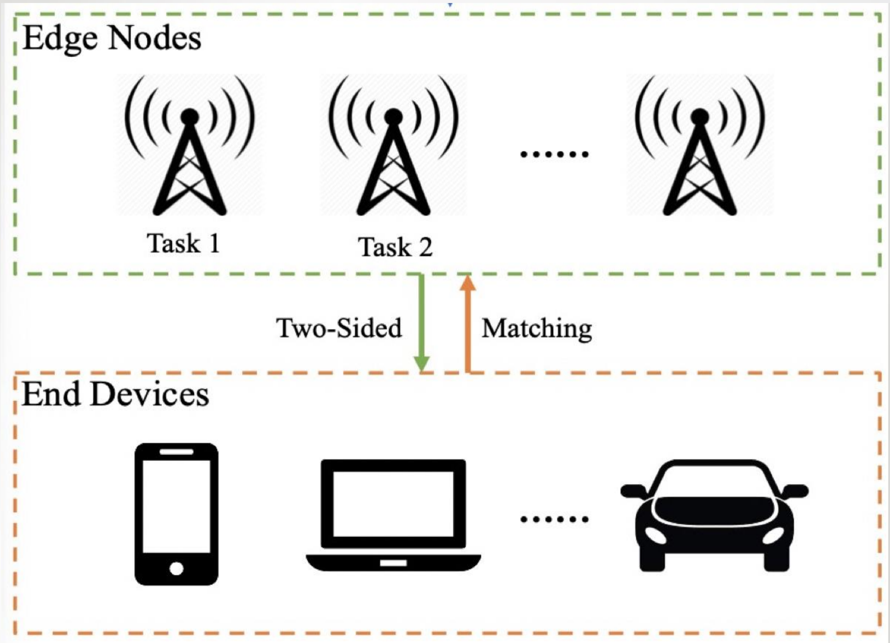


Fig. 2. The multi-task federated learning framework in MEC scenario.

Dawei Chen, Choong Seon Hong, Li Wang, Yiyong Zha, Yunfei Zhang, Xin Liu and Zhu Han, ``Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks," IEEE Transactions on Mobile Computing, 2021.

Stable Marriage Matching

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Basic elements (***Stable Marriage***):
  - ***Agents***: A set of men, and a set of women;
  - ***Preference list:*** A sorted list of men/women based on her/his preferences;
  - ***Blocking pair (BP)*** (m,w):
    - 1). m is unassigned or prefers w to his current partner;
    - 2). w is unassigned or prefers m to her current partner;
  - ***Stable matching***: A matching admit no BPs.
  - ***Gale-Shapley*** Algorithm: find a stable matching in SM.

GS algorithm

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Geeta, Heiki, Irina, Fran

Adam

Irina, Fran, Heiki, Geeta

We reach a stable marriage!

Bob

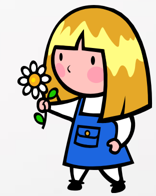
Geet i, Irin

Ca

Fran

David

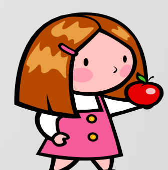
Challenge: What if the preference list is incomplete?



Fran

Geeta

Heiki



Irina

Simulation Results

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Impact of user numbers and edge node numbers

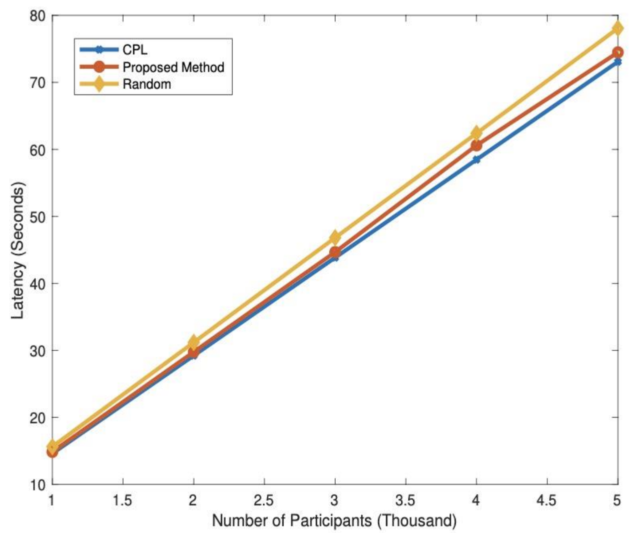


Fig. 4. Network latency with different number of users.

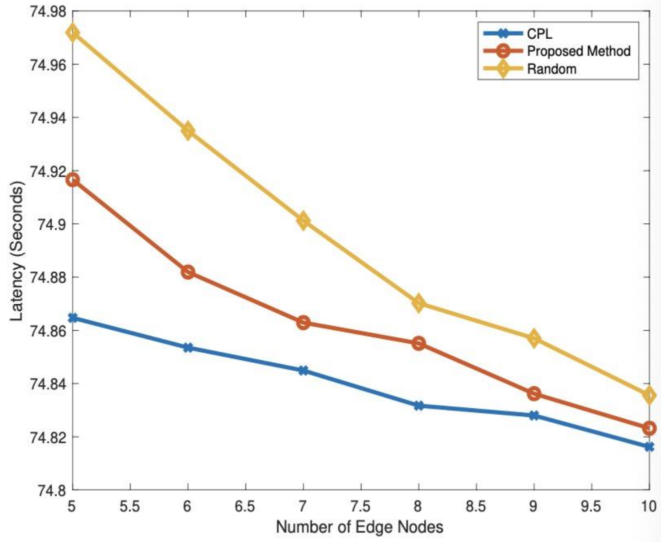


Fig. 5. Network latency with different number of edge nodes.

Evidently, the network latency is positively related to the number of participants while is negatively correlated with the number of edge nodes.

Our proposed method is close to the performance of complete preference list (CPL) case.

Outline

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

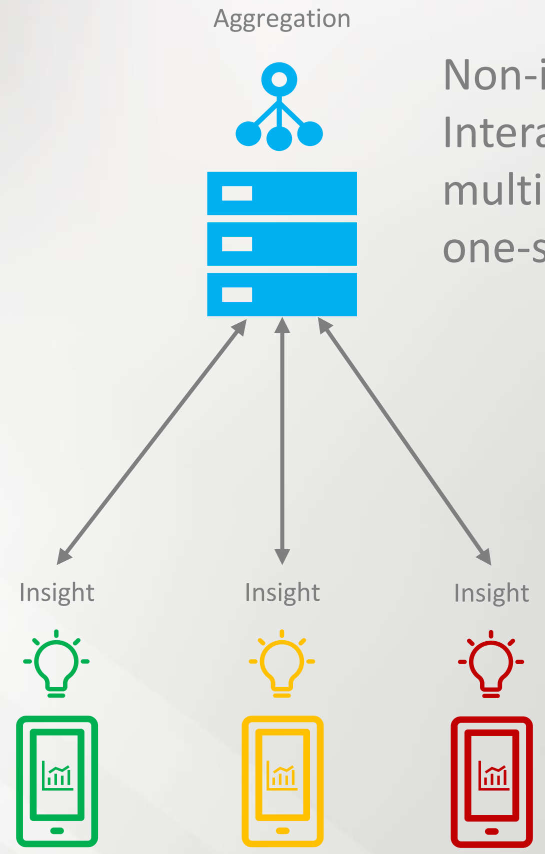
- Background and Fundamentals of Federated Paradigm
  - Background
  - Machine Learning (ML) Point of View
  - Optimization Point of View
- Federated Learning for Wireless Networks
  - Unsupervised Federated Learning for Unbalanced Data
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - Federated Skewness Analytics in Heterogeneous Decentralized Data Environments
  - Federated Anomaly Analytics for Local Model Poisoning Attack
- Open Problems and Conclusions

Beyond Federated Learning: Federated Analytics

__Federated Analytics__ is the practice of applying __data science__ methods to the analysis of raw data that is stored locally on users' devices.

Originally defined in https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html

- Terminology
  - *Insights* are derived by clients and sent to server
  - *Aggregation* are performed by server for global knowledge construction
- Characteristics
  - No raw data exchange
  - Focus on population-level knowledge
  - Interactive/non-interactive
  - Privacy guaranteed
- An Analogy Example between FL and FA ☺



Non-interactive:
Interactive:
multiple rounds
one-shot analytic

Federated learning Workflow

Federated Analytics vs. Others

- To Federated Learning

	Federated Learning	Federated Analytics
Goal	Training ML models	Non-training tasks (data science)
Aggregation approach	FedAvg	Task dependent
		Tree | Bayesian | MPC | etc.
Local insights	Model weights	Task dependent
		Partial info | Distilled info | etc.

- To Distributed Data Mining

	Distributed Data Mining	Federated Analytics
Raw data transmission	Redistribution assumed	Stay where it origins
Clients (nodes) and server	Trusted	Untrusted (privacy & Byzantine attack)
Data & device heterogeneity	Little concerned	Focused

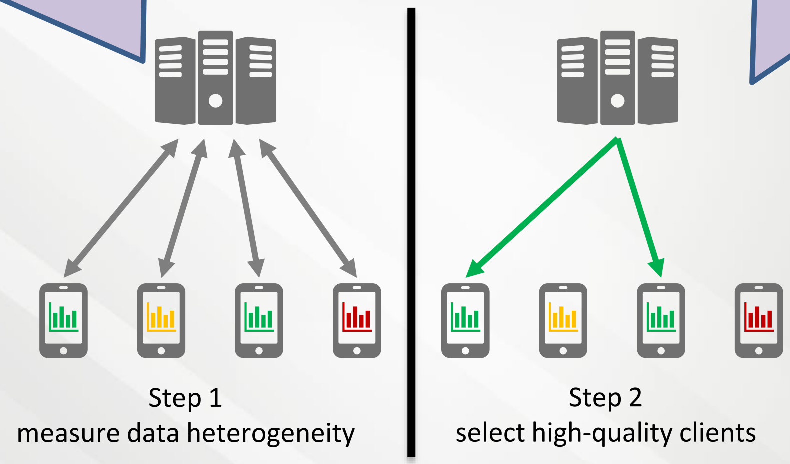
FedACS: an Example of Federate Analysis

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- FedACS: a stand-alone federated analysis instance assisting some other federated tasks
  - Goal: measuring data heterogeneity (skewness) and create a client-pool with low data skewness

Goal: data heterogeneity measurement
Insight: weight reuse
Aggregation: Hoeffding inequality based

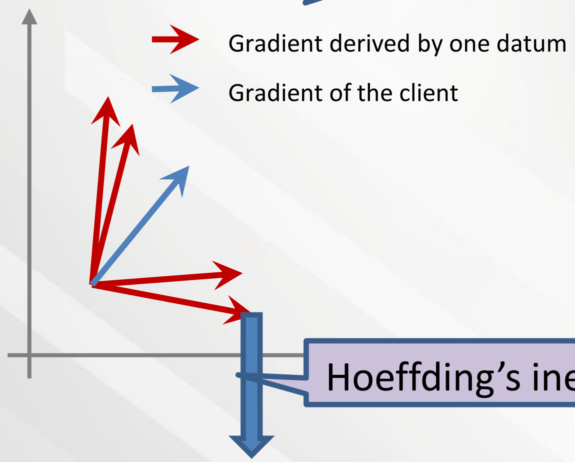
Goal: client selection
Challenge: non-stationary measurement
Solution: dueling bandit



"FedACS: Federated Skewness Analytics in Heterogeneous Decentralized Data Environments", Z. Wang, Y. Zhu, D. Wang, Z. Han, IWQoS 2021

FedACS: Design Overview

Client gradient is the average of datum gradients



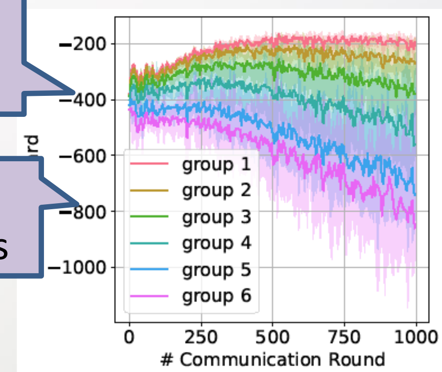
Hoeffding's inequality

$$Skewness_i = \|\Delta w_i - \overline{\Delta w}\|_2$$

Step 1
measure data heterogeneity

Skewness estimate is drifting during the training procedure

Relative preference holds between different client groups



Dueling bandit

$R_i = -2$	$R_j = -3$	$R_k = -10$
$win = 2, lose = 0$	$win = 1, lose = 1$	$win = 0, lose = 2$

Step 2
select high-quality clients

- When assisting FL, FedACS reduces 65.6% of accuracy loss and speeds up for 2.4x

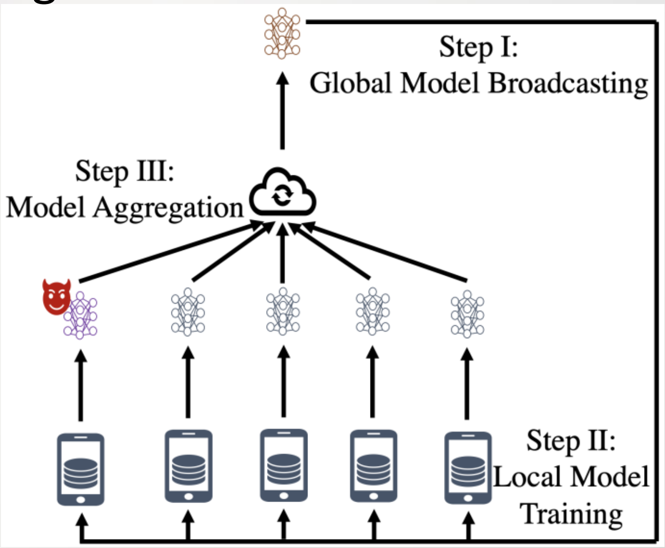
Outline

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Background and Fundamentals of Federated Paradigm
  - Background
  - Machine Learning (ML) Point of View
  - Optimization Point of View
- Federated Learning for Wireless Networks
  - Unsupervised Federated Learning for Unbalanced Data
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - Federated Skewness Analytics in Heterogeneous Decentralized Data Environments
  - Federated Anomaly Analytics for Local Model Poisoning Attack
- Open Problems and Conclusions

Federated Anomaly Analytics for Local Model Poisoning Attack

- Local model poisoning attack
  - ➢ Threat model
    - The attacker can manipulate the shared local models not the local data during the process of federated learning.
  - ➢ Impact
    - Slowing down the convergence rate of the learning process.
    - Degrading the prediction accuracy of the learned global mode.
  - ➢ Most defense methods are passive
    - Treat the normal local models and the poisoned local models indiscriminately, such as GeoMed and Trimmed Mean.



(a) Federated Learning under local model poisoning attack

    - It cannot eliminate all the poisoned local models, thus the training performance is affected to some degree, i.e., the accuracy of learned global model is reduced.

Siping Shi, Chuang Hu, Dan Wang, Yifei Zhu, and Zhu Han, ``Federated Anomaly Analytics for Local Model Poisoning Attack," to appear IEEE Journal on Selected Areas in Communication

Motivation, Challenges and Methodology

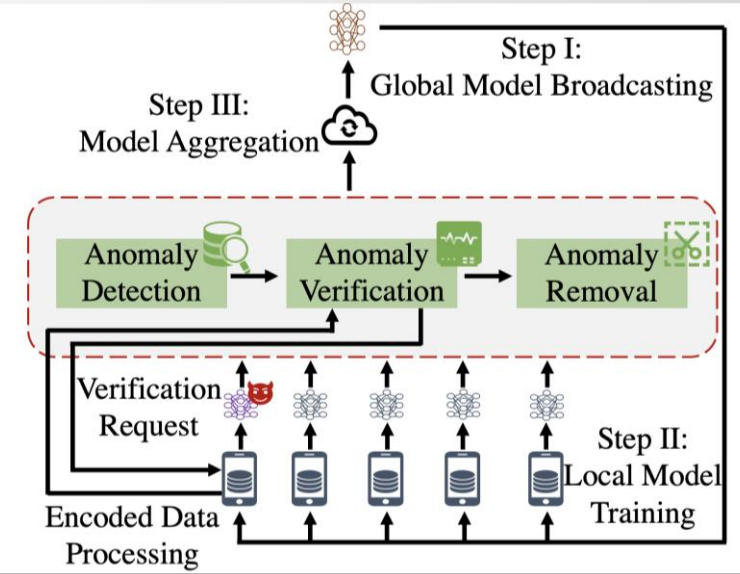
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Motivation
  - ➢ Leverage the new ***federated analytics*** paradigm to develop a ***proactive defense*** method with privacy and performance guarantee.
- Challenge
  - ➢ Data heterogeneity caused by federated scenarios increases the difficulty in anomaly analytics.
- Methodology of federated analysis framework with three modules

1) Anomaly detection module

✓ Identify the potential malicious local model updates with a light-weight anomaly detection algorithm

✓ FAA-DL allows greater compatibility with various anomaly detection algorithms. Support Vector Machine (SVM) is selected in our paper.



Methodology (cont.)

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

2) Anomaly verification module

- ✓ Request encrypted local data from potential malicious clients which identified in the anomaly detection module.
- ✓ The server computes the corresponding gradient with the receiving encrypted local data based on functional encryption method.
- ✓ Verify whether the potential malicious client is true malicious by comparing the gradients.

3) Anomaly removal module

- ✓ Remove the verified malicious local model updates from aggregation.

Algorithm 2: FAA-DL

Input: Number of participated clients: n;
Local training data of client i: D_i;
Number of global iterations: R
Number of selected clients: k; Set of local model update :
$\boldsymbol{G} = \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k\}$; Learning rate: α;
Proportion of malicious client: β.

Output: Global model: $\boldsymbol{w}$;

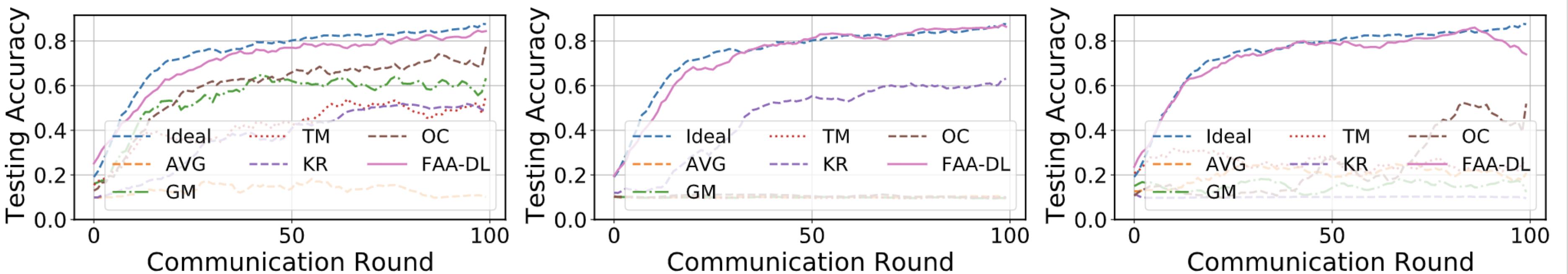
$\boldsymbol{w} \leftarrow$ random initialization.
for $r = 1, 2, \ldots, R$ **do**
  // Step I: Global model broadcasting
  The server randomly selects k clients from n clients and sends them $\boldsymbol{w}$.
  // Step II: Local model training
  Client side:
  for $i = 1, 2, \ldots, k$ **do**
    $\boldsymbol{g}_i = \texttt{ModelUpdate}(\boldsymbol{w}, D_i)$,
    Send $\boldsymbol{g}_i$ to server.
  // Step III: Global model aggregation
  Server side:
  $\boldsymbol{G}'_m \leftarrow \texttt{AnomalyDetection}(\boldsymbol{G}_m, \beta)$,
  for $\boldsymbol{g}_i \in \boldsymbol{G}'_m$ **do**
    $VR_i \leftarrow \texttt{AnomalyVerification}(\boldsymbol{g}_i, \boldsymbol{w}, \alpha, D_i)$
    if $VR_i == True$ **then**
      $\boldsymbol{G}_m.\texttt{add}(\boldsymbol{g}_i)$
  $\boldsymbol{G}_b \leftarrow \texttt{AnomalyRemoval}(\boldsymbol{G}_m, \boldsymbol{G})$,
  $\boldsymbol{g} \leftarrow \texttt{FedAvg}(\boldsymbol{G}_b)$,
  $\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha \cdot \boldsymbol{g}$

Experiments

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Accuracy:

- FAA-DL outperforms other defense methods on the accuracy of the learnt global model.
- The performance gap of FAA-DL is within 0.92% –2.48% of the ideal baseline across all tested attacks.



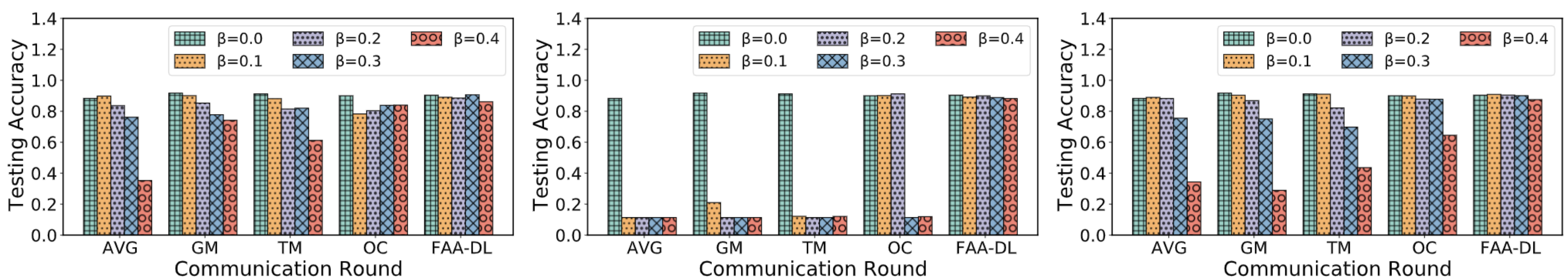
(a) Gaussian attack (b) Sign-flipping attack (c) ALIE attack

Fig. 4: The accuracy of defense to different attacks with different methods.

Experiments

Robustness :

- FAA-DL remains nearly the same accuracy as the ideal baseline when the proportion of attacked devices increased from 10% to 40%,
- while other methods decreased greatly epically in sign-flipping attack.



(a) Gaussian attack (b) Sign-flipping attack (c) ALIE attack

Fig. 7: Top-1 accuracy of different defense methods to different attacks with various fraction of malicious devices(from 0.1 to 0.4)

Open Problems

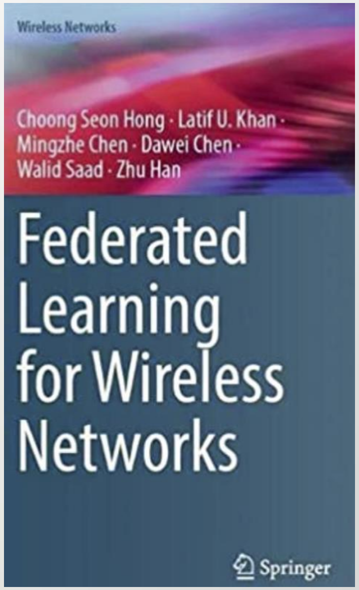
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Some open areas in Federated learning
  - ✓ Optimization algorithms for FL, particularly communication-efficient algorithms tolerant of non-IID data
  - ✓ Approaches that scale FL to larger models, including model and gradient compression techniques
  - ✓ Novel applications of FL, extension to new learning algorithms and model classes.
  - ✓ Not everyone has to have the same model (multi-task and pluralistic learning, personalization, domain adaptation)
  - ✓ Bias and fairness in the FL setting (new possibilities and new challenges)
  - ✓ Enhancing the security and privacy of FL, including cryptographic techniques and differential privacy

Conclusions

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Federated learning will be a major part of learning paradigm
  - Mobile massively decentralized, naturally arising (non-IID) partition
  - Availability of distributed clients; Address communication bottleneck
  - Privacy concern
- Explore different aspects and applications of federated learning and wireless networks
  - Formulations, Problem specific solution
  - Link machine learning, computation, communication, networking, and operational research together
  - From federated learning to federate analysis
- Some other federated works
  - Satellite Communications Based Federated Learning with Mean-field Game
  - Collaborative Frequent Pattern Mining
  - Protecting Inference Privacy



Amigo Lab

http://wireless.egr.uh.edu/

http://www2.egr.uh.edu/~zhan2


THANK YOU
UNIVERSITY of HOUSTON | ENGINEERING