UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Federated Learning and Analysis In Mobile 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, Dan Wang

National Science Foundation, Toyota and Amazon

Outline

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

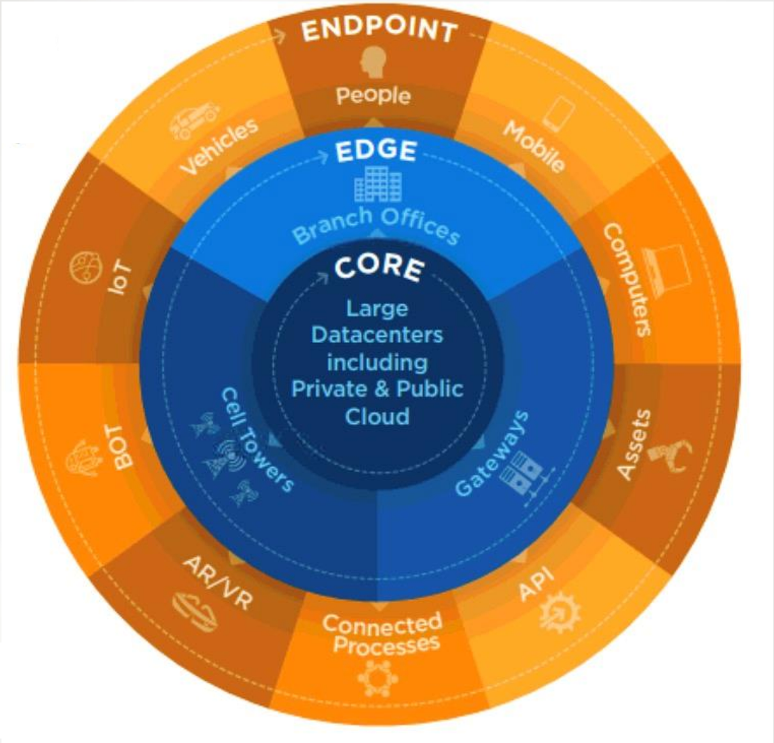
- Background and Fundamentals
  - Background
  - Machine Learning and Optimization Point of Views
- Federated Learning for Wireless Networks
  - Toyota Example
  - Matching Theory Based Low-Latency Scheme for Multi-Task Federated Learning in MEC Networks
- From Federated Learning to Federated Analysis
  - 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

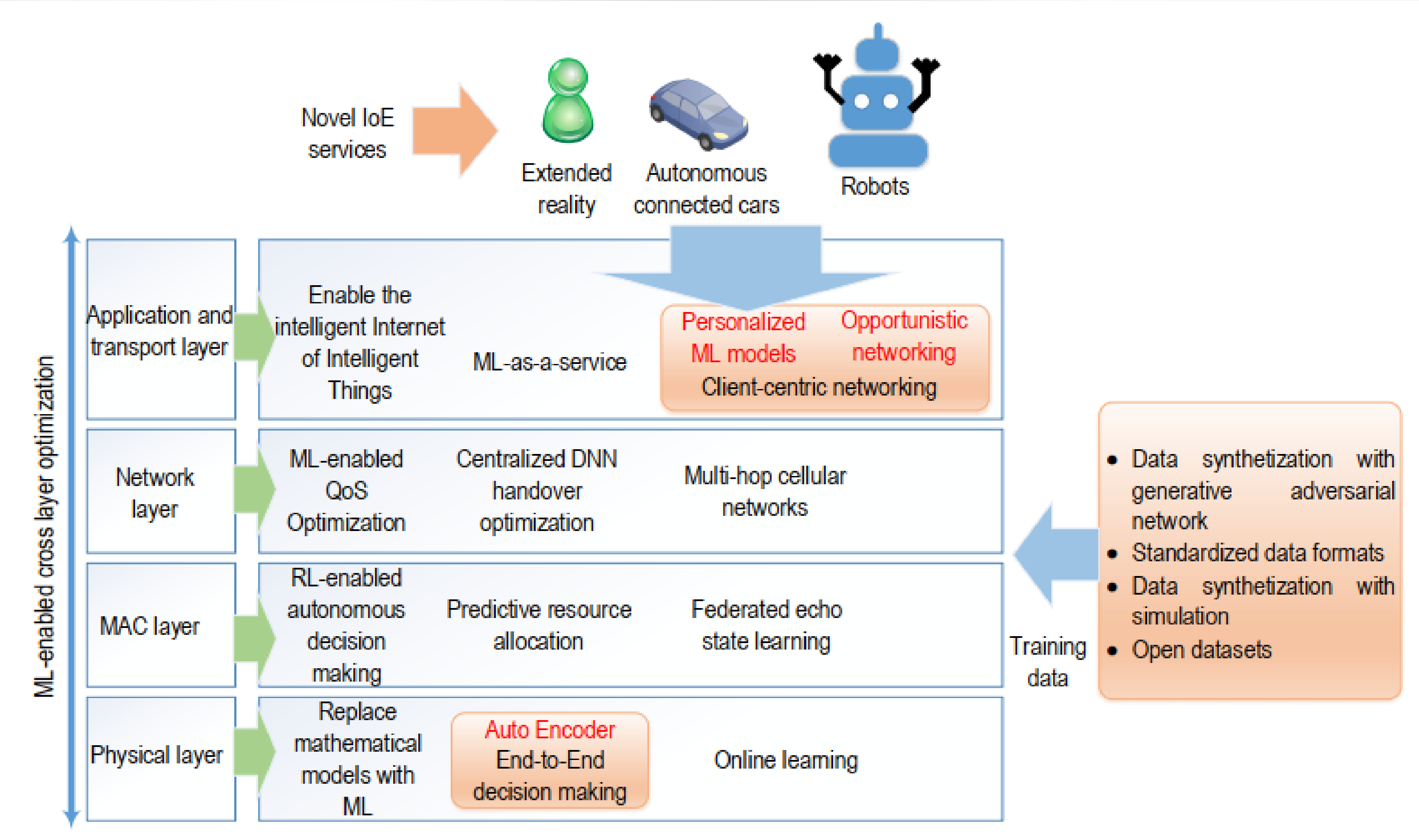
What about analytics?

What about learning?



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

Background

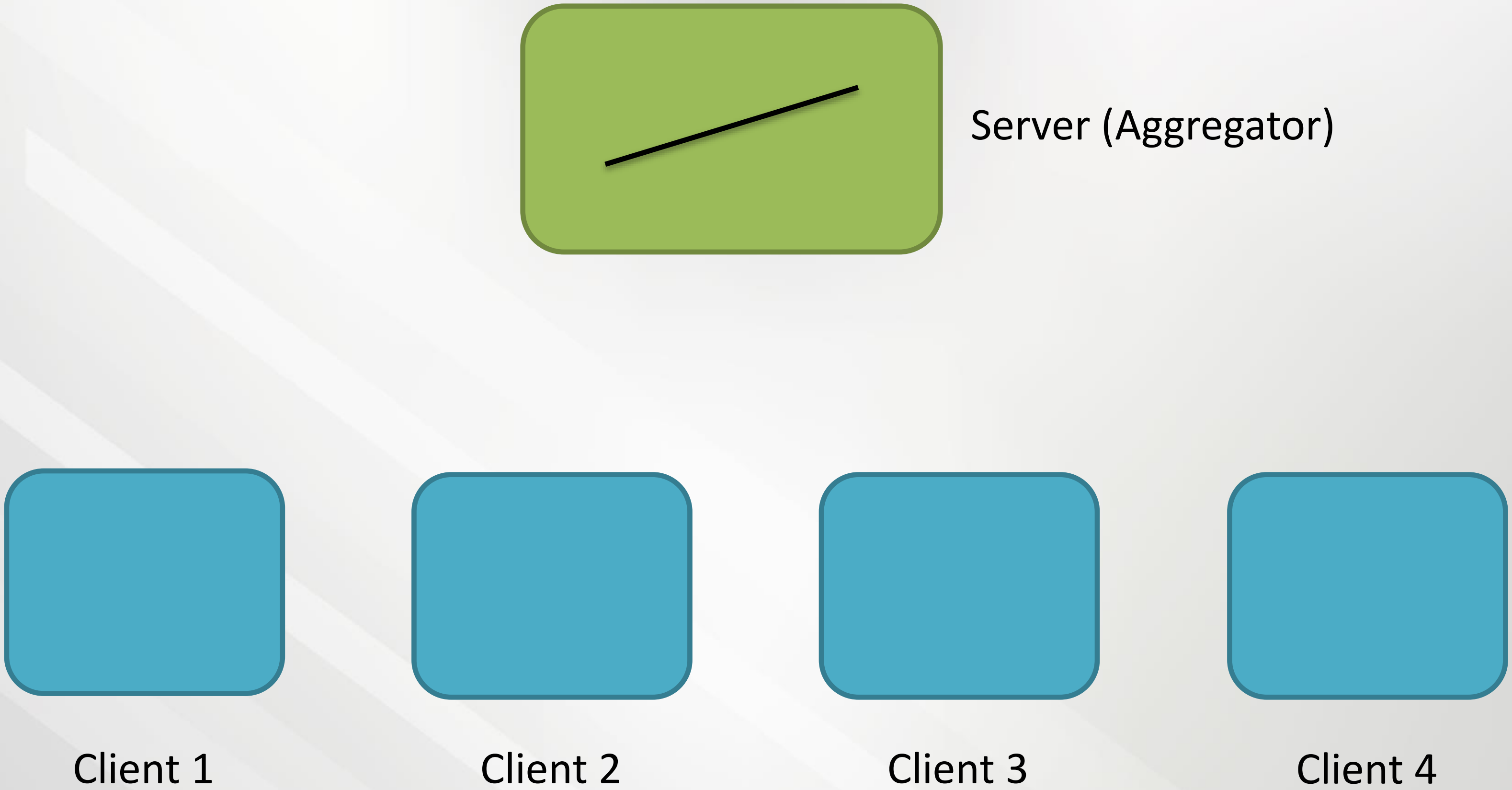


S. Ali, W. Saad, N. Rajatheva, K. Chang, D. Steinbach, B. Sliwa, C. Wietfeld, K. Mei, H. Shiri, H.-J. Zepernick *et al.*, "6g white paper on machine learning in wireless communication networks," *arXiv preprint arXiv:2004.13875*, 2020

ML Point of View

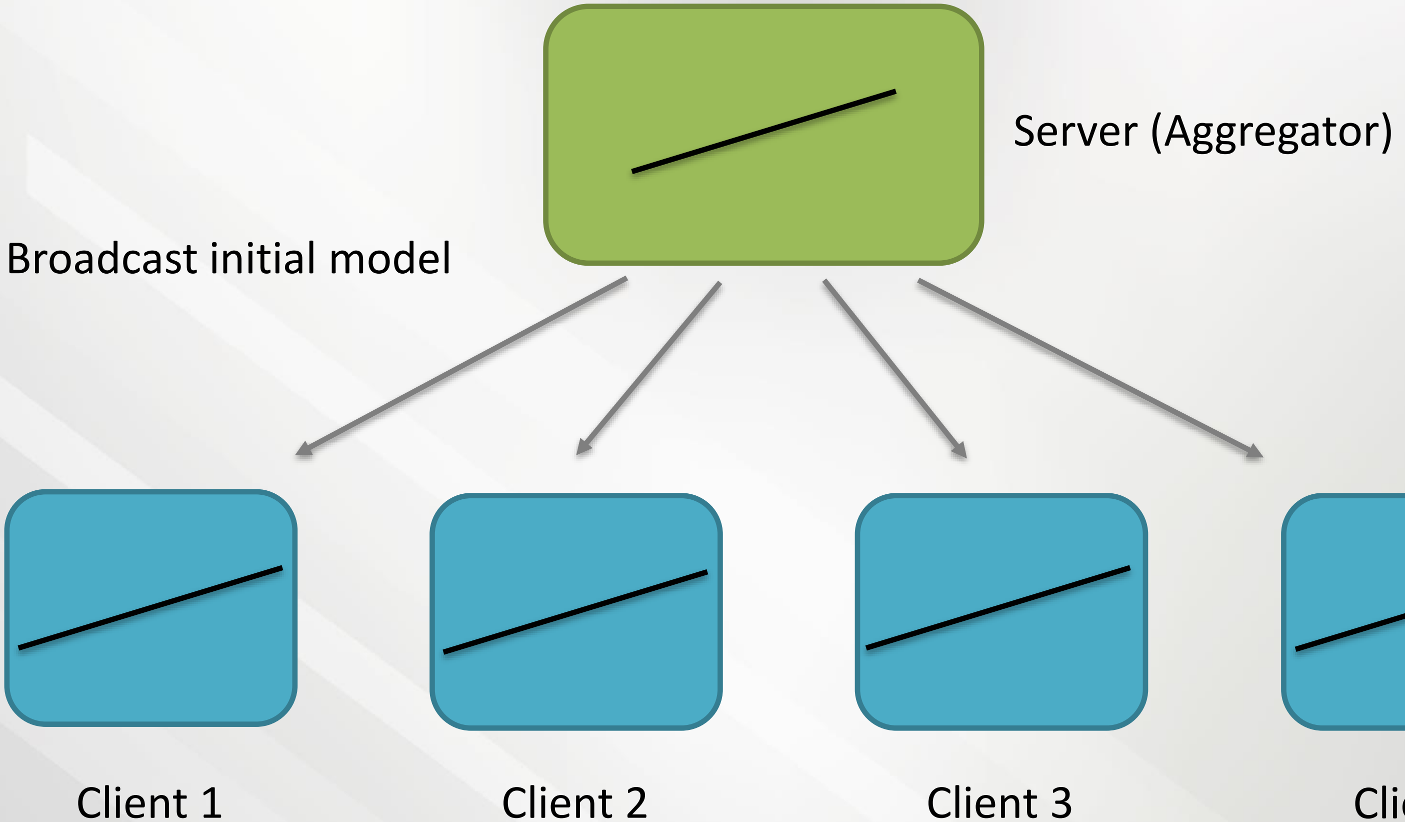
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- What is Federated Learning?
  - General workflow



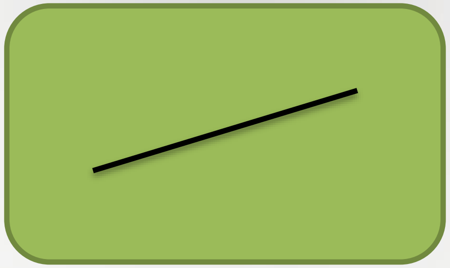
ML Point of View

- What is Federated Learning?
  - General workflow



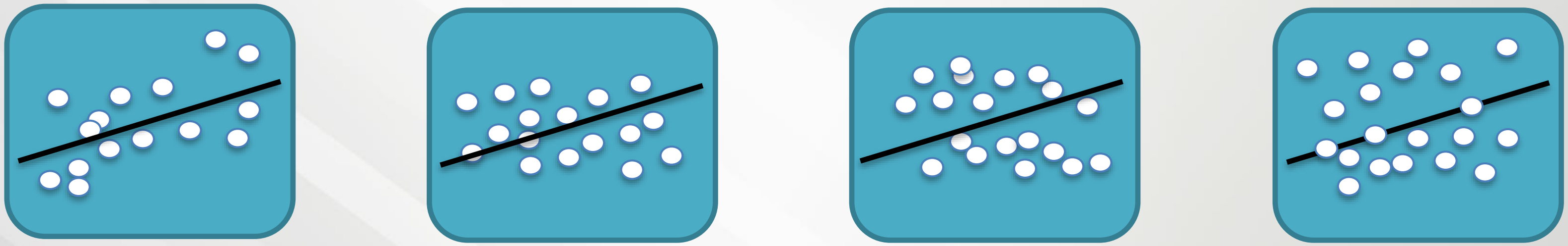
ML Point of View

- What is Federated Learning?
  - General workflow



Server (Aggregator)

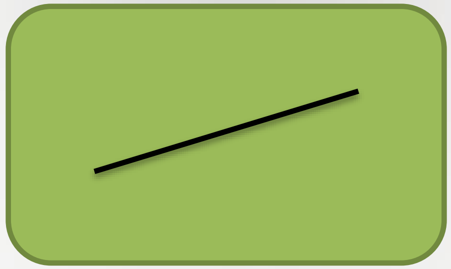
Clients generate local data



Client 1 Client 2 Client 3 Client 4

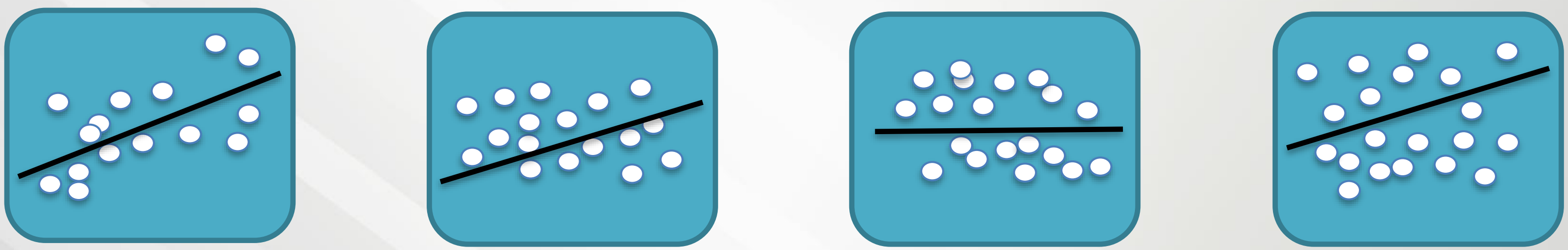
ML Point of View

- What is Federated Learning?
  - General workflow



Server (Aggregator)

Clients train the initial model
based on local dataset

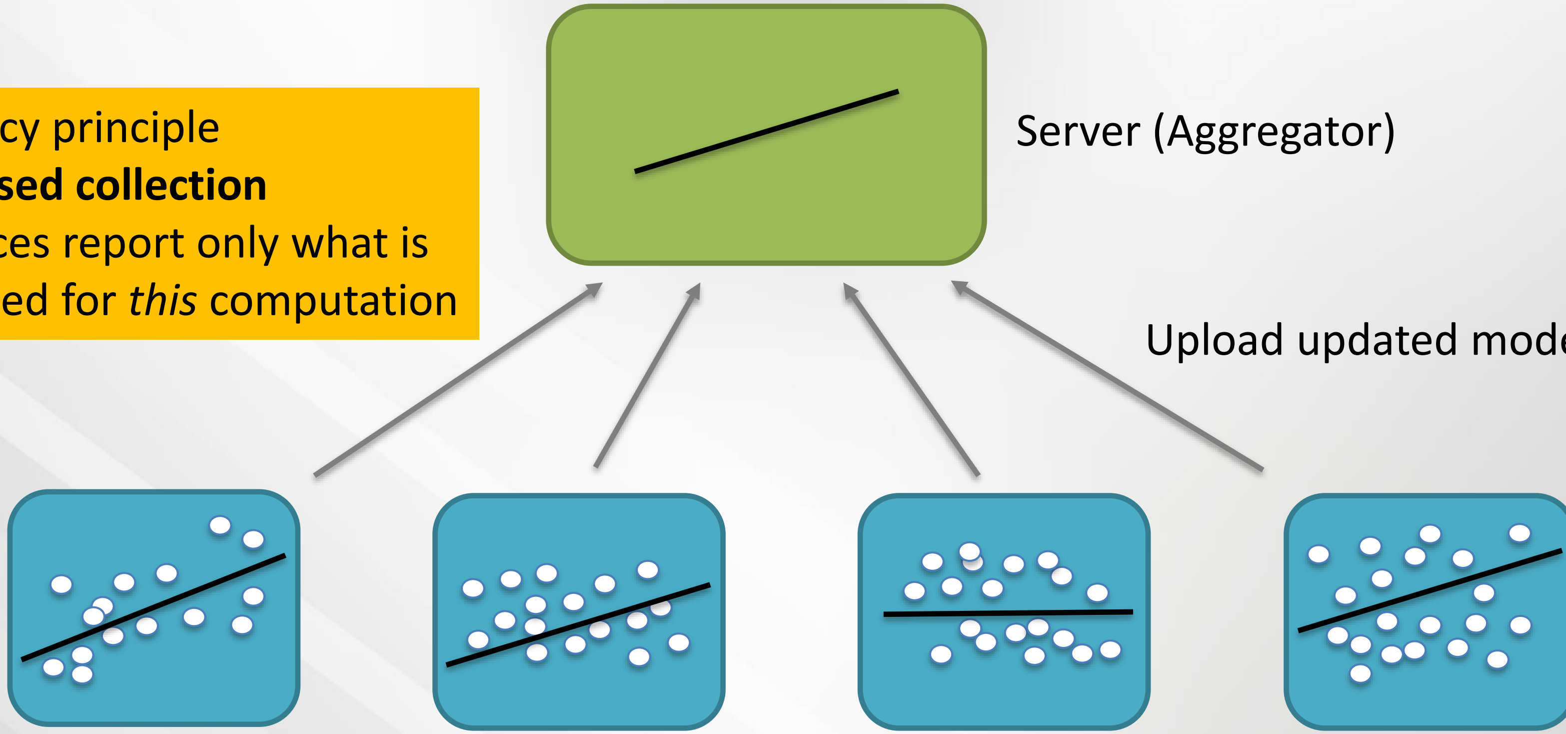


Client 1 Client 2 Client 3 Client 4

ML Point of View

- What is Federated Learning?
  - General workflow

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



Server (Aggregator)

Upload updated model

Client 1 | Client 2 | Client 3 | Client 4

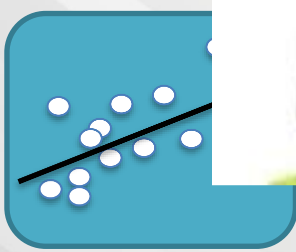
ML Point of View

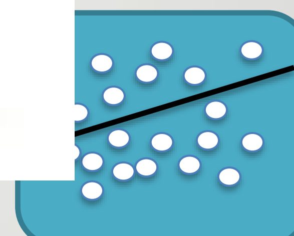
- What is Federated Learning?
  - General workflow

Combine in

or)

Repeat these pr
convergence





Client 1 Client 2 Client 3 Client 4

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.*

How to handle our research group

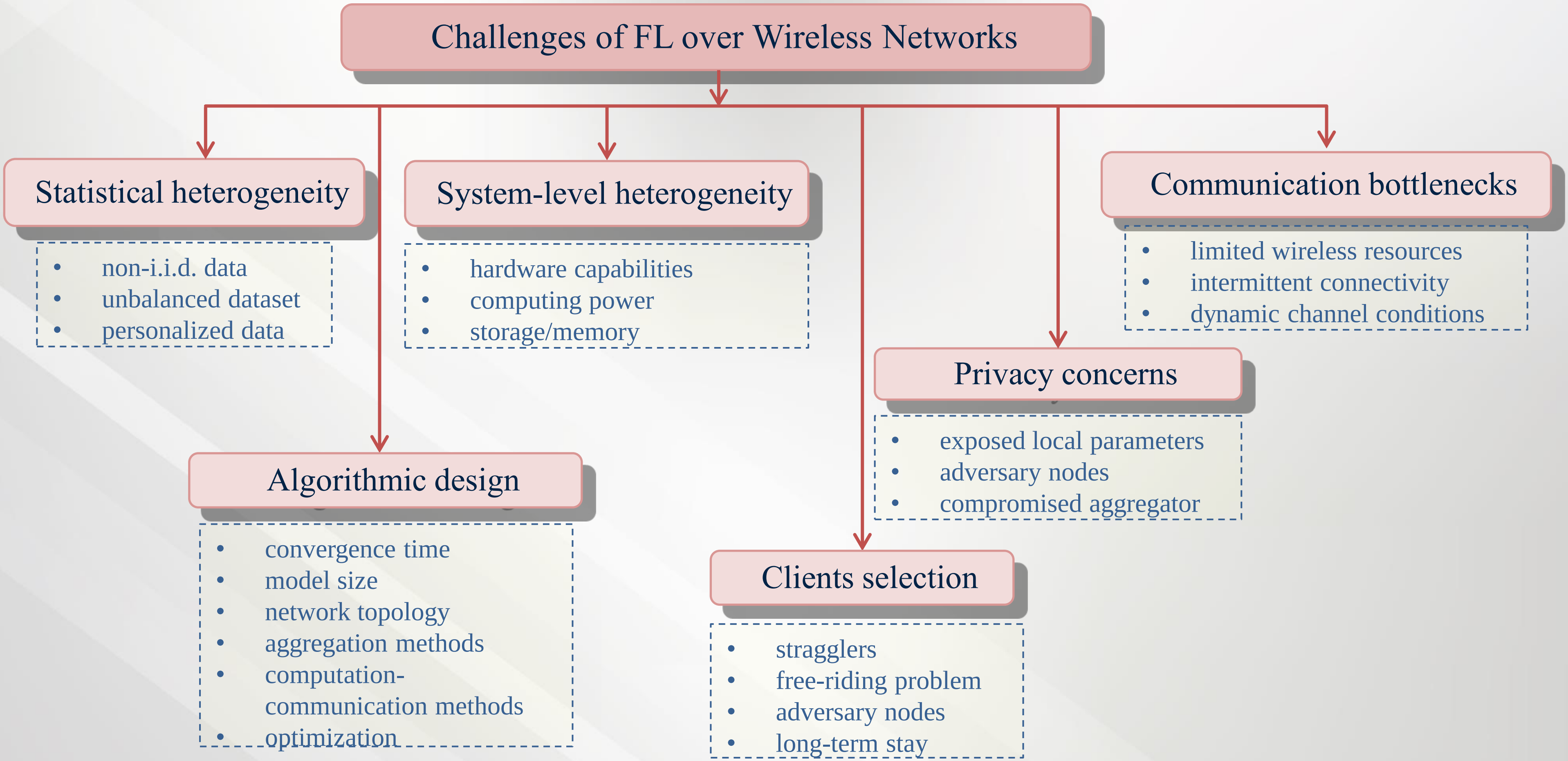
FL Advantages

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

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.

FL Challenges



Outline

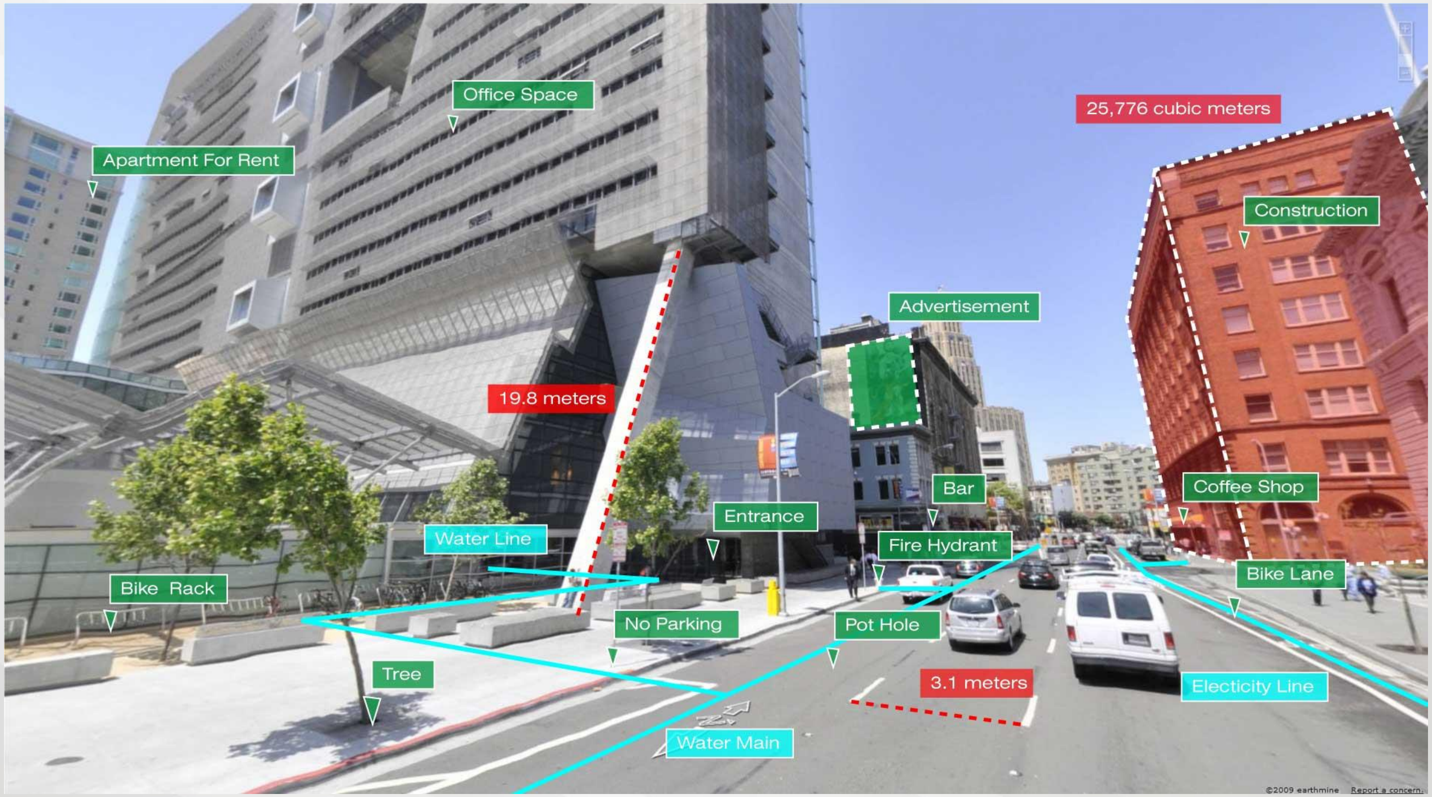
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Background and Fundamentals
  - Background
  - Machine Learning and Optimization Point of Views
- Federated Learning for Wireless Networks
  - Toyota Example
  - 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

Toyota Example

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- What AR does is to implant 3-D virtual objects in a real-world context.
- Challenges:
  - ✓ Latency: Real-time interaction; Dizziness
  - ✓ Accuracy: Object recognition and matching



Methodology

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Example 2: Matching Motivation

- 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

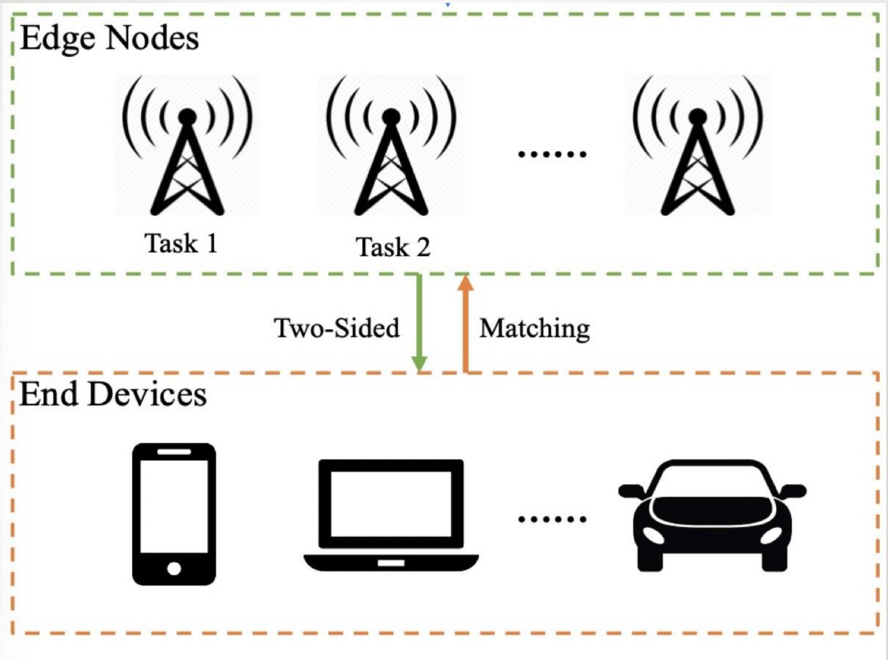


Fig. 2. The multi-task federated learning framework in MEC scenario.

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

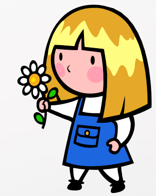
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- 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

Geeta, Heiki, Irina, Fran

Adam



Fran

Irina, Fran, Heiki, Geeta

We reach a stable marriage!

Bob

Geeta

Geet i, Iri

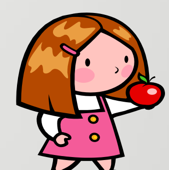
Ca

Heiki

Challenge: What if the preference list is incomplete?

Ge Fran

David



Irina

Simulation Results

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Impact of user numbers and edge node numbers

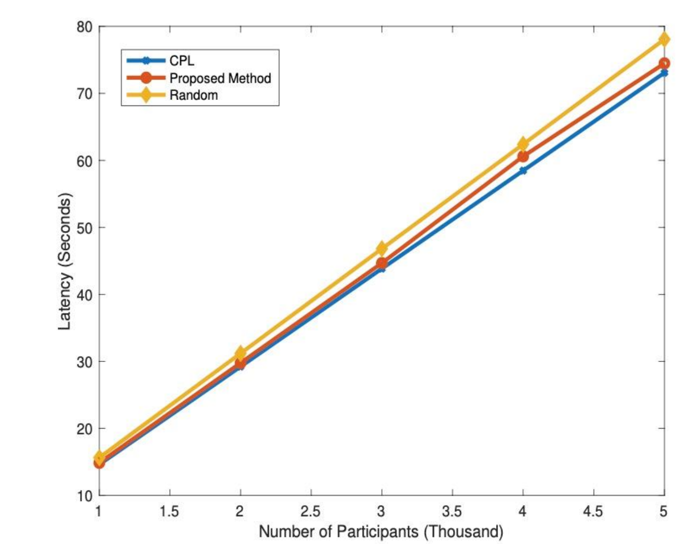


Fig. 4. Network latency with different number of users.

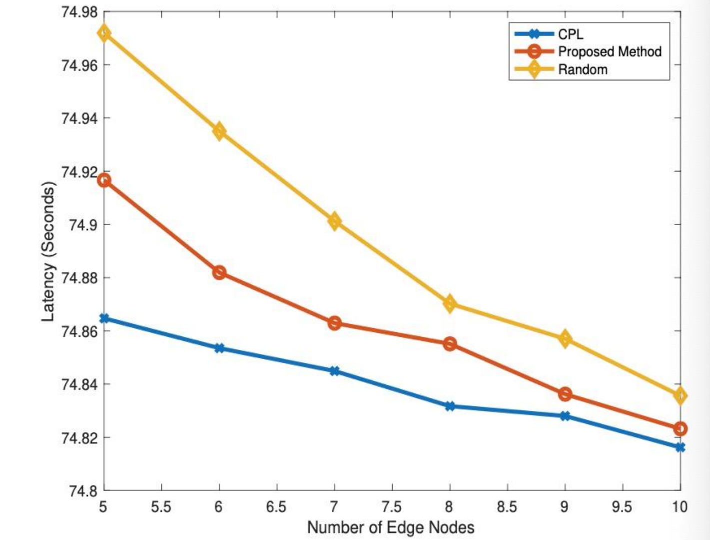


Fig. 5. Network latency with different number of edge nodes.

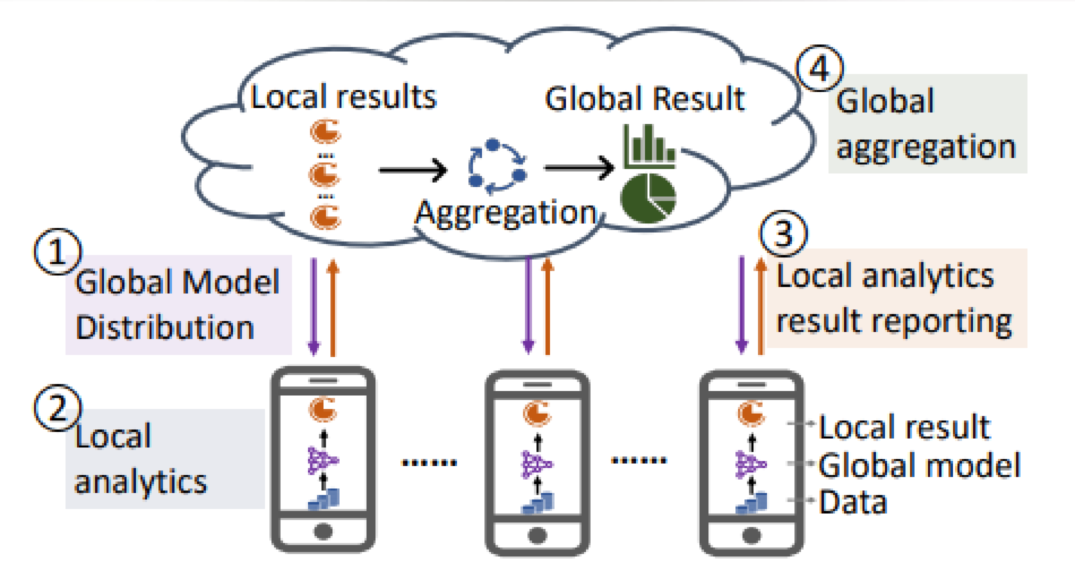
Evidently, the network latency is positively related to the number of participants while is negatively correlated with the number of edge nodes.

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

Outline

- Background and Fundamentals
  - Background
  - Machine Learning and Optimization Point of Views
- Federated Learning for Wireless Networks
  - Toyota Example
  - 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

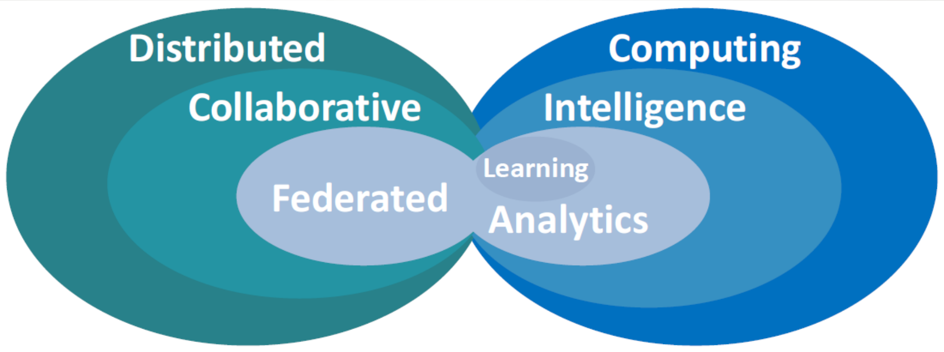
Beyond Federated Learning: Federated Analytics

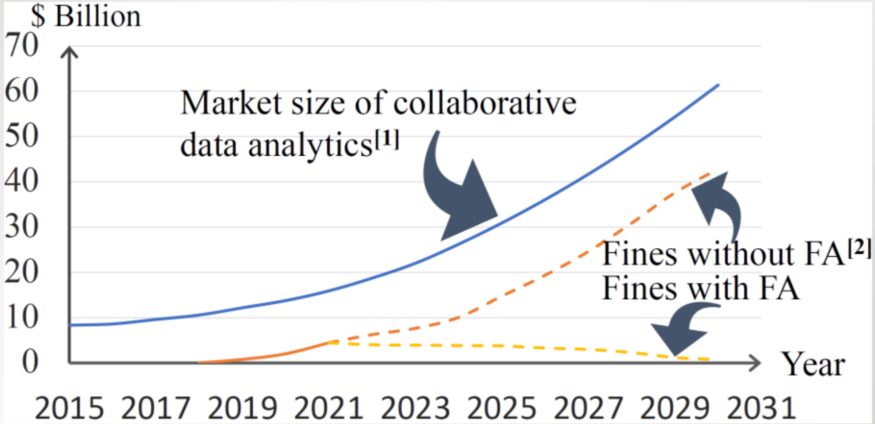


- Google proposed Federated Analytics in May 2020
  - Also for the Gboard application
  - Federated learning for model training
  - Federated analytics for model testing
- Google's definition on federated analytics:
  - Collaborative data science without data collection
  - https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html
- My two examples of federated analytics

What is Federated Analytics: Taxonomy

- Federated: how nodes collaborate
- Analytics: what the computing task is





Collaboration Model Computing Model

- Data analytics: to draw conclusions from data
- Federated analytics: A collaborative computing paradigm that performs data analytic computing tasks across multiple decentalized devices where the raw data should be kept local
- Market: Increasing demands on collaborative data analytics vs. Increasing concerns on privacy and confidentiality

Federated Analytics vs. Others

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- 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

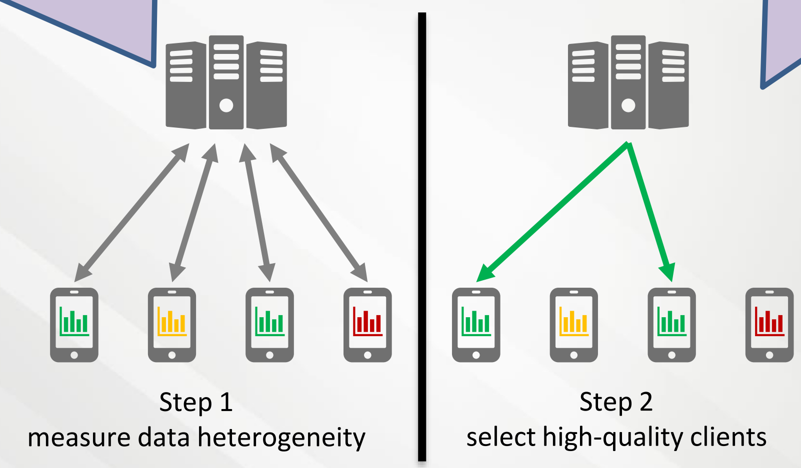
	Distributed Data Mining	Federated Analytics
Raw data transmission	Redistribution assumed	Stay where it origins
Clients (nodes) and server	Trusted	Untrusted (privacy & Byzantine attack)
Data & device heterogeneity	Little concerned	Focused

FA Example1: FedACS

- 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

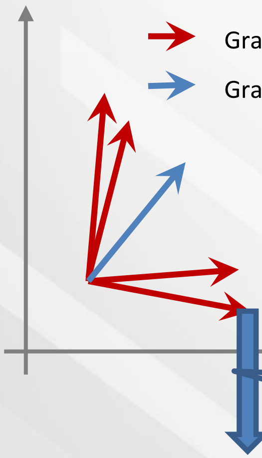


“FedACS: Federated Skewness Analytics in Heterogeneous Decentralized Data Environments”, Z. Wang, Y. Zhu, D. Wang, Z. Han, IWQoS 2021

FedACS: Design Overview

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Client gradient is the average of datum gradients



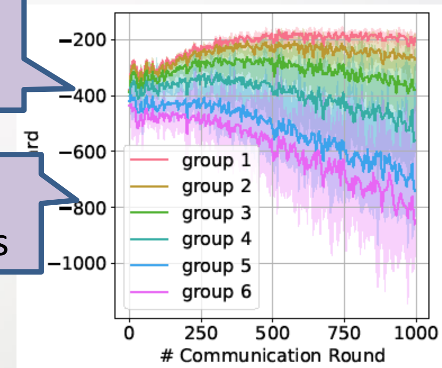
Hoeffding's inequality

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

Step 1
measure data heterogeneity

Skewness estimate is drifting during the training procedure

Relative preference holds between different client groups



Dueling bandit

$R_i = -2$ → $win = 2, lose = 0$

$R_j = -3$ → $win = 1, lose = 1$

$R_k = -10$ → $win = 0, lose = 2$

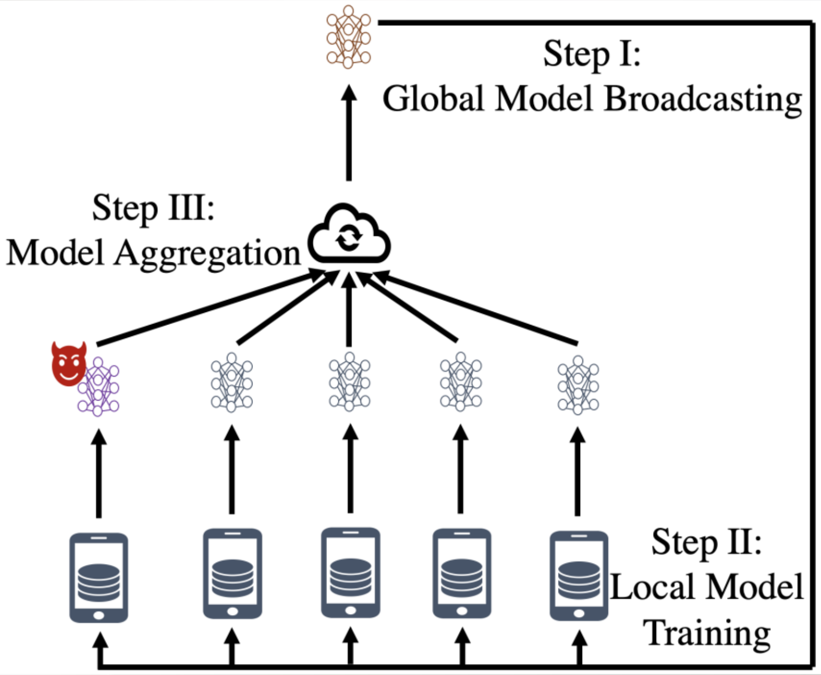
Step 2
select high-quality clients

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

Example 2: Local Model Poisoning Attack

FA is vulnerable to attacks

- Local model poisoning attack
  - A single malicious worker can arbitrarily manipulate the uploaded local models during the process of federated learning
- Harmful effect on the whole FL
  - Broadly slowing down the convergence rate[1]
  - Significantly degrading the prediction accuracy of the learned global model[2]



Shi, Siping, et al. "Federated anomaly analytics for local model poisoning attack." IEEE Journal on Selected Areas in Communications. 2021.

[1] . Blanchard, et al, “Machine learning with adversaries: Byzantine tolerant gradient descent,” NeurPIS 2017

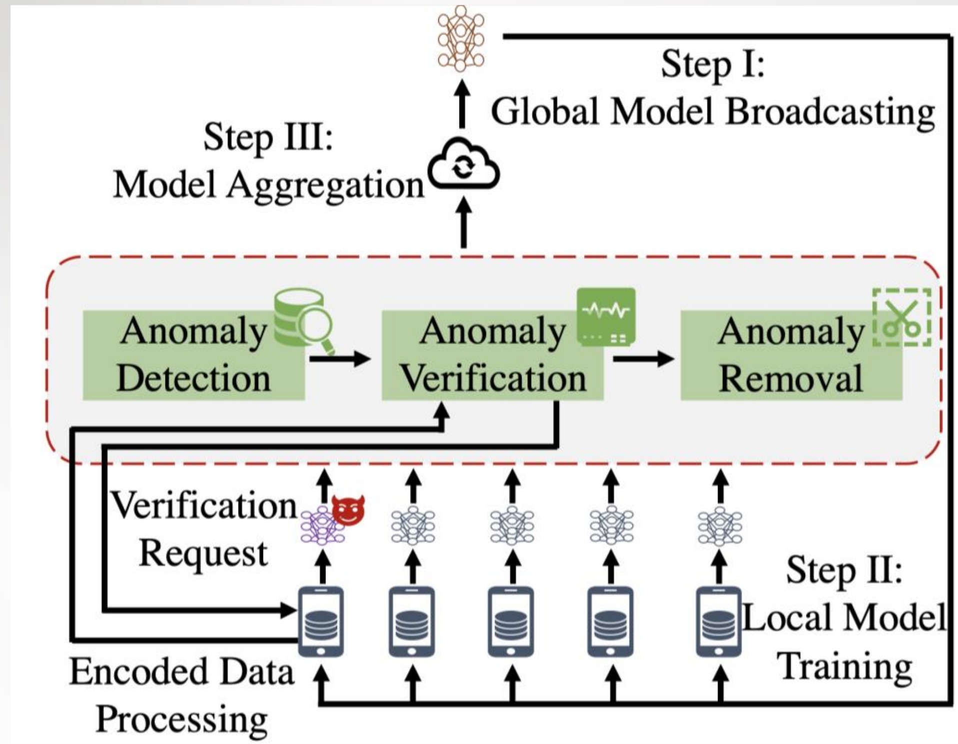
[2] . Bagdasaryan, et al, “Howto backdoor federated learning,” AISTATS 2020

Motivation, Challenges and Methodology

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Motivation:

Everything Everywhere All at Once, Oscar 2023



Modules:

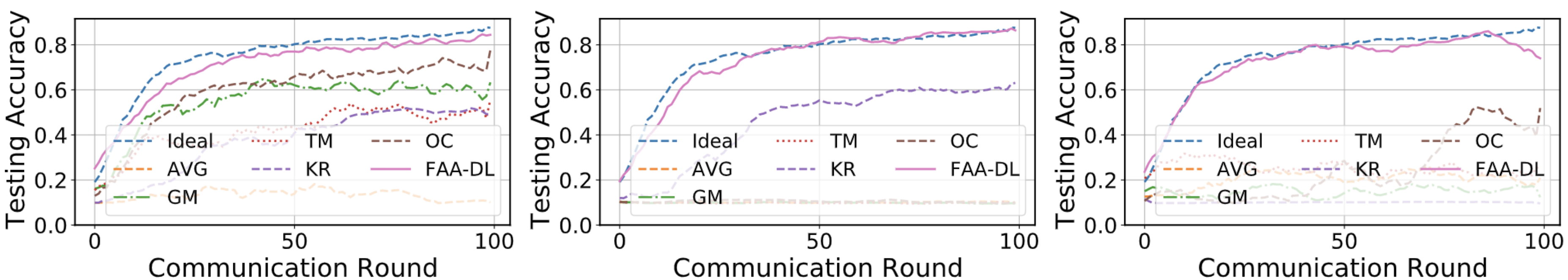
- Anomaly Detection Module
- Anomaly Verification Module
- Anomaly Removal Module

Experiments

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

Results:

- FAA-DL outperforms other defense methods on the accuracy of the learned global model, with an accuracy improvement up to 2.03X
- 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.

Open Problems

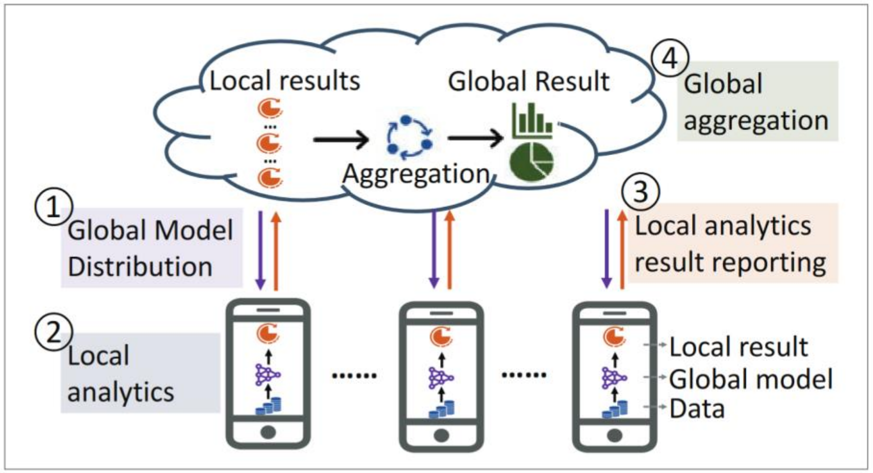
UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

- Resource optimization
  - Optimization algorithms for FL, particularly communication-efficient algorithms tolerant of non-IID data
- Scalability
  - Approaches that scale FL to larger models, including model and gradient compression techniques
- Convergence improvement
  - There is a need to devise approaches that converge fast.
- Fairness-enabled FL
  - Bias and fairness in the FL setting (new possibilities and new challenges)
- Secure FL
  - Enhancing the security and privacy of FL, including cryptographic techniques and differential privacy

Open Areas

UNIVERSITY of HOUSTON
CULLEN COLLEGE of ENGINEERING

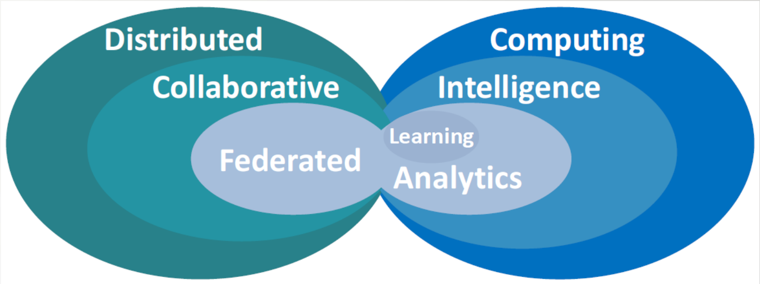
- ✓ Application/algorithm level: more applications call for redesign
  - ✓ Federated statistics
  - ✓ Federated visualization (e.g. histogram, heatmap)
  - ✓ Federated global/local model evaluation
  - ✓ Federated database query
  - ✓ Federated data sketching
  - ✓ Federated data publication
  - ✓ and more …
- ✓ System level
  - ✓ Communication efficiency
  - ✓ Device heterogeneity
  - ✓ and more …
- ✓ Privacy, incentive algorithm, and more…



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
- We explore different aspects and applications to integration of federated learning and wireless networks
  - Formulations
  - Problem specific solution
  - Link machine learning, computation, communication, networking, and OR
  - From federated learning to federate analysis



Join or Visit Our Lab

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

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


THANK YOU
UNIVERSITY of HOUSTON | ENGINEERING