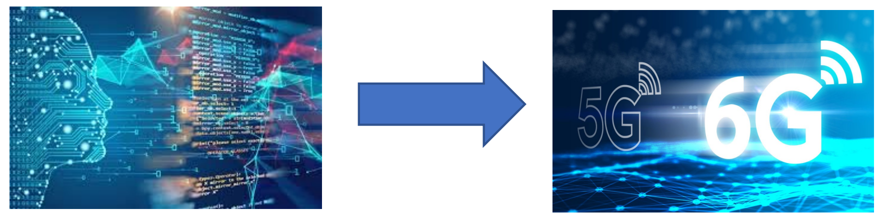
Machine Learning at the Wireless Edge

H. Vincent Poor

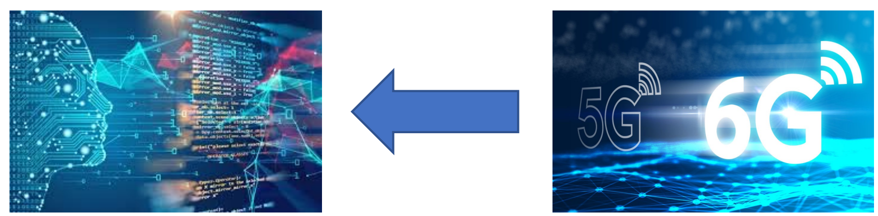
Princeton University

Machine Learning & Wireless Networks

- Two Aspects:



  - Using machine learning to optimize communication networks
  - Learning on mobile devices (the focus of today's talk)

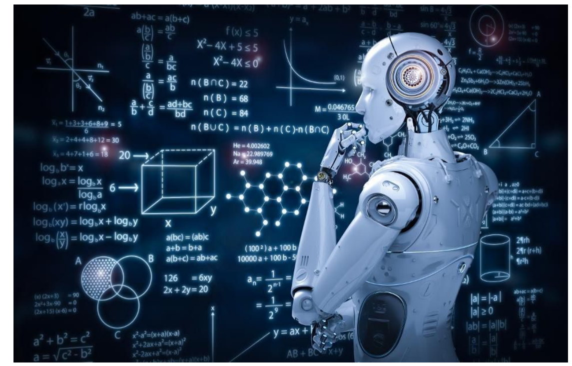


Today's Talk: Focus on Federated Learning

- Motivation
- Federated Learning over Wireless Channels (Scheduling)
- Privacy Protection in Federated Learning (Differential Privacy)
- Some Research Issues

Motivation

Machine Learning (ML): State-of-the-Art



- Tremendous progress in recent years
    - More and more data is available
    - Significant increase in computational power

- "Standard" ML
    - Implemented in a centralized manner (e.g., in a data center/cloud)
    - Full access to the data
- State-of-the art models (e.g., Deep Neural Networks) run in the cloud
    - Managed and operated by standard software tools (e.g., TensorFlow, etc.)
    - Accelerated by specialized hardware (e.g., Nvidia's GPUs, Google's TPUs)

Machine Learning at the Wireless Edge

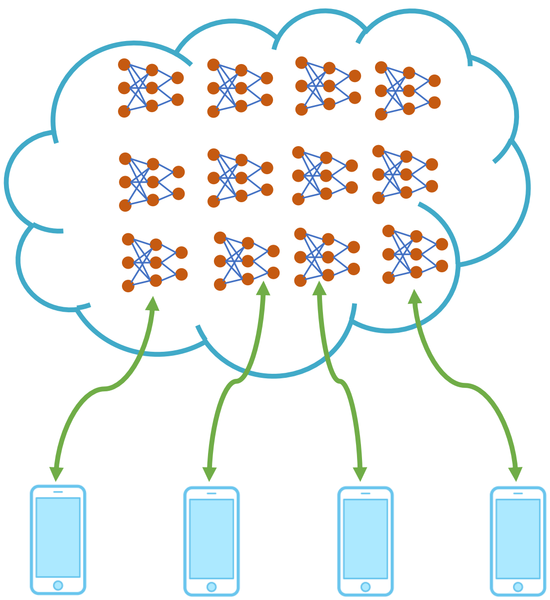
- Centralized ML may not be suitable for many emerging applications, e.g.,
  - Self-driving cars
  - First responder networks
  - Healthcare networks

- What makes these applications/situations different
  - Data is born at the edge (phones and IoT devices)
  - Limited capacity uplinks
  - Low latency & high reliability
  - Data privacy / security
  - Scalability & locality

- Motivates moving learning closer to the network edge

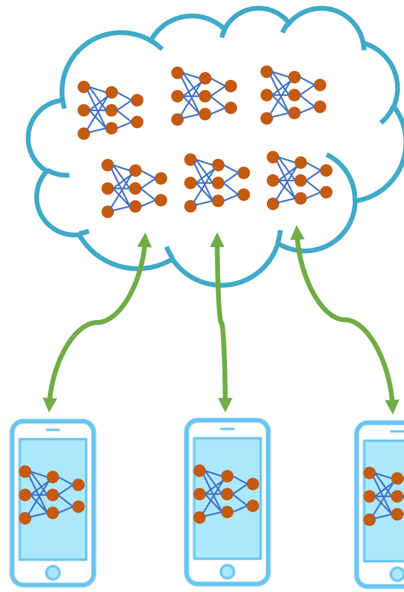
Networked ML Models

"Standard" ML



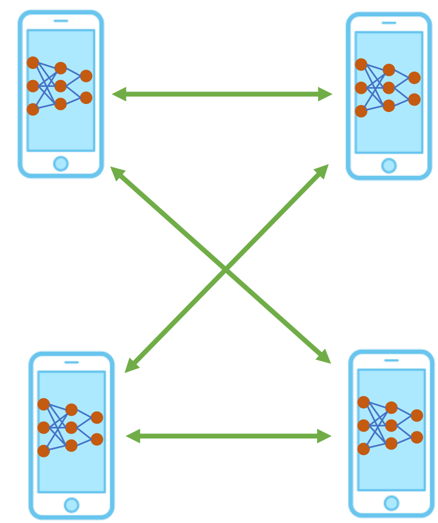
- ML in the cloud with dumb end-user devices
- All data is in the cloud
- Inference and decision making in the cloud
- No data privacy

Federated ML



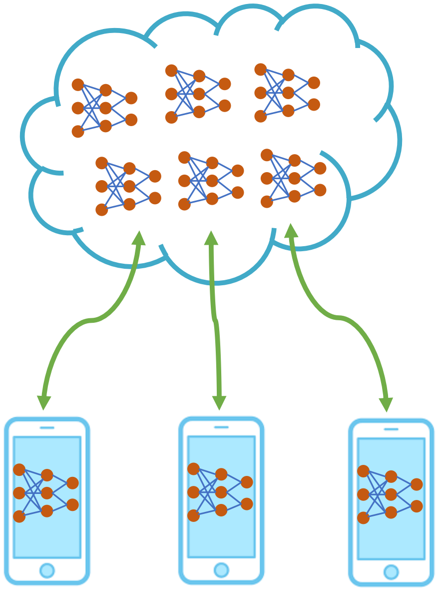
- ML in the cloud + on-user-device ML
- Only part of the data is in the cloud
- Use the cloud but smartly
- Privacy-promoting

Decentralized ML



- No infrastructure (e.g., cloud) needed
- Data is fully distributed
- Collaborative intelligence
- Privacy-promoting (sharing models instead of data)

Federated Learning over Wireless Channels (Scheduling)



Federated Learning: Basic Architecture

- Federated Learning
  - Enable end-user devices to do ML without centralizing data
  - Key features
    - On-device datasets: end users (UEs) keep raw data locally
    - On-device training: end-user devices perform training on a shared model
    - Federated computation: an edge node (AP) collects trained weights from end users and updates the shared model; then the process is iterated to convergence

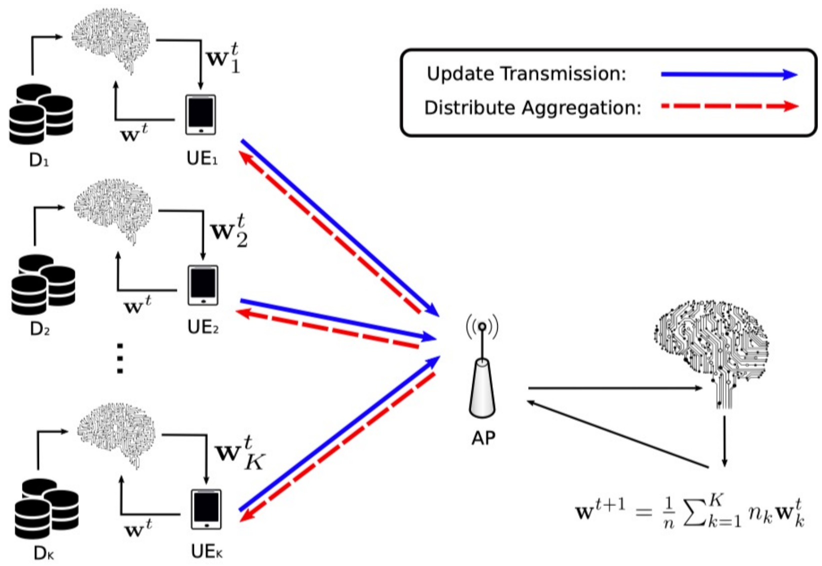
Federated Learning: Issues to Address

- Living on the edge
  - Communication to the AP needs to go through wireless channels
  - The wireless medium is shared and resource-constrained
    - Only a limited number of devices can be selected in each update round
    - Transmissions are not reliable due to interference

- Questions
  - How should we schedule devices to update trained weights?
  - How does the interference affect the training?

Scheduling Mechanisms

- Scheduling mechanisms
  - Random Scheduling: AP uniformly selects N out of K UEs at random
  - Round Robin: AP groups UEs into $G=K/N$ groups, sequentially selecting each group
  - Proportional Fair: AP selects N out of K UEs with the strongest SNRs:



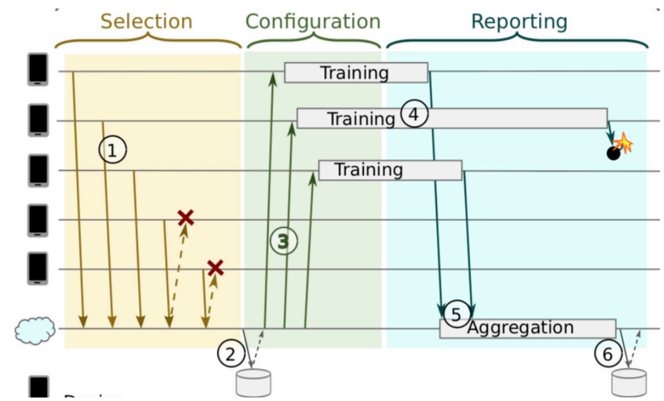
$$\mathbf{m}^* = \arg\max_{\mathbf{m}\subset\{1,2,\ldots,K\}} \left\{ \frac{\tilde{R}_{m_1}}{\bar{R}_{m_1}}, \ldots, \frac{\tilde{R}_{m_N}}{\bar{R}_{m_N}} \right\}$$

Yang, et al. (2020), “Scheduling Policies for Federated Learning in Wireless Networks”, *IEEE T-COM*

Performance Metric

- Federated Learning in a mobile edge network
  - The trained update can be successfully received by AP if and only if
    - The UE is selected by the AP, and
    - The received SINR exceeds a decoding threshold

$$\gamma_{k,t} = \frac{P_{\mathrm{ut}} h_k \|z_k\|^{-\alpha}}{\sum_{z \in \tilde{\Phi}_{\mathrm{u}}^k} P_{\mathrm{ut}} h_z \|z\|^{-\alpha} + \sigma^2} > \theta,$$



- Metric to quantify the effectiveness of training:
  - The number of communication rounds required to reach an ε-accurate solution

Convergence Rates of Federated Learning

Theorem 1: *Under RS policy, for any given convergence target ε, choosing the T_{RS} such that*

$$T_{\mathrm{RS}} \geq \frac{\log(\varepsilon/n)}{\log\left(1 - \frac{(1-\beta)/G}{1+\mathcal{V}(\theta,\alpha)}\right)}, \tag{28}$$

we have the expected duality gap satisfies $\mathbb{E}[P(\mathbf{w}(\mathbf{a}^{T_{\mathrm{RS}}})) - D(\mathbf{a}^{T_{\mathrm{RS}}})] < \varepsilon$.

Theorem 2: *Under RR policy, for any given convergence target ε, choosing the T_{RR} such that*

$$T_{\mathrm{RR}} \geq \frac{G\log(\varepsilon/n)}{\log\left(1 - \frac{1-\beta}{1+\mathcal{V}(\theta,\alpha)}\right)}, \tag{31}$$

we have the expected duality gap satisfies $\mathbb{E}[P(\mathbf{w}(\mathbf{a}^{T_{\mathrm{RR}}})) - D(\mathbf{a}^{T_{\mathrm{RR}}})] < \varepsilon$.

Theorem 3: *Under PF policy, for any given convergence target ε, choosing the T_{PF} such that*

$$T_{\mathrm{PF}} \geq \frac{\log(\varepsilon/n)}{\log\left(1-(1-\beta)\sum_{i=1}^{K-N+1}\binom{K-N+1}{i}\frac{(-1)^{i+1}/G}{1+\mathcal{V}(i\theta,\alpha)}\right)}, \tag{33}$$

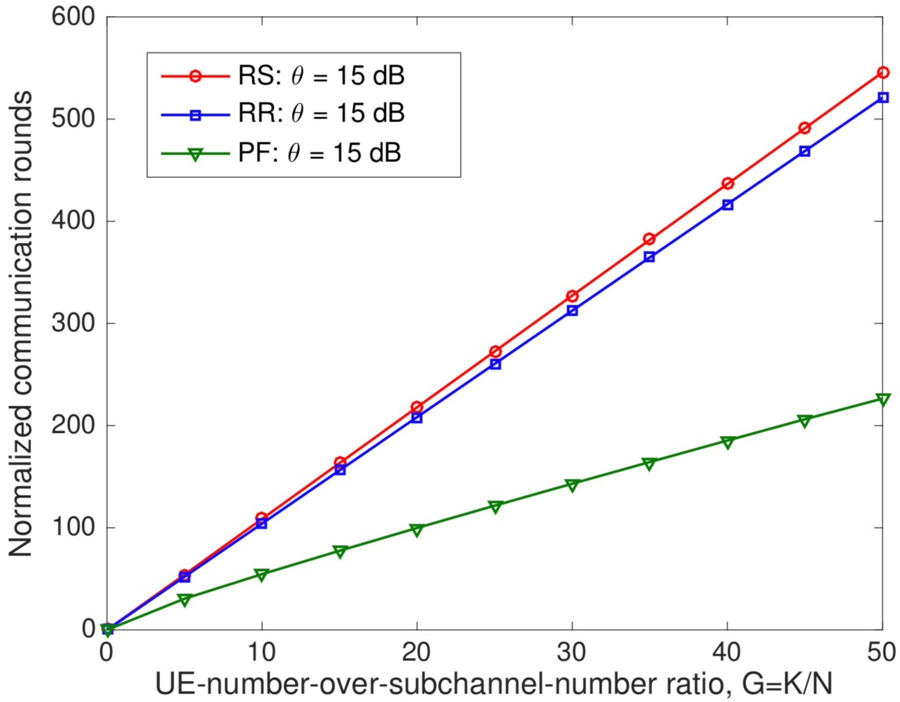
we have the expected duality gap satisfies $\mathbb{E}[P(\mathbf{w}(\mathbf{a}^{T_{\mathrm{PF}}})) - D(\mathbf{a}^{T_{\mathrm{PF}}})] < \varepsilon$.

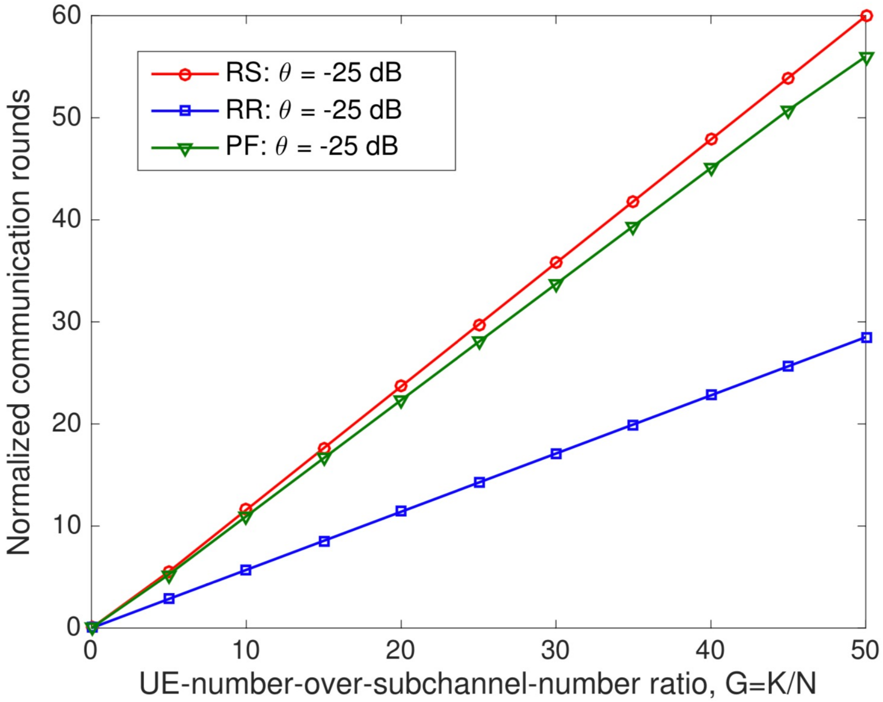
α = path loss exponent
β = precision level at UEs
n = total # exemplars

Numerical Example

- PF works the best in high SINR condition
- RR works the best in low SINR condition

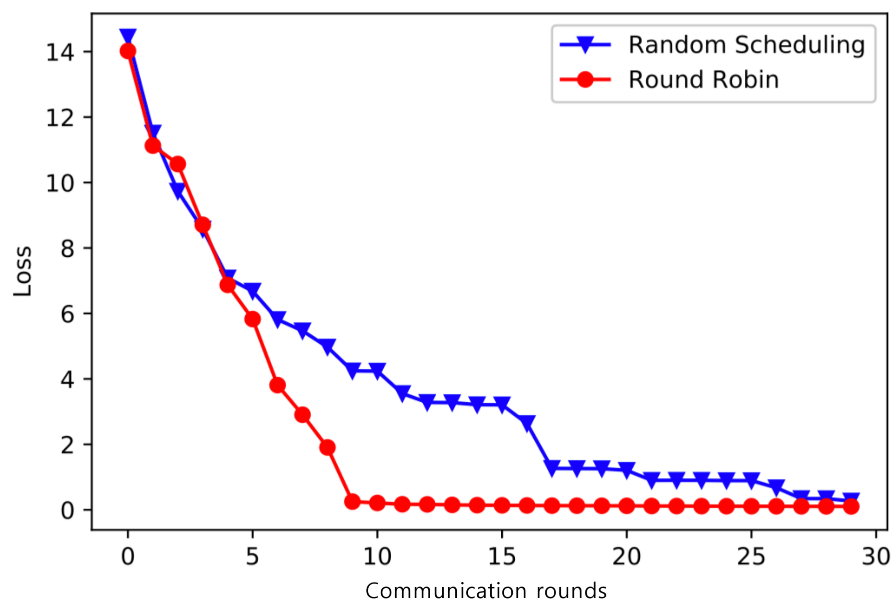
- High SINR vs low SINR threshold

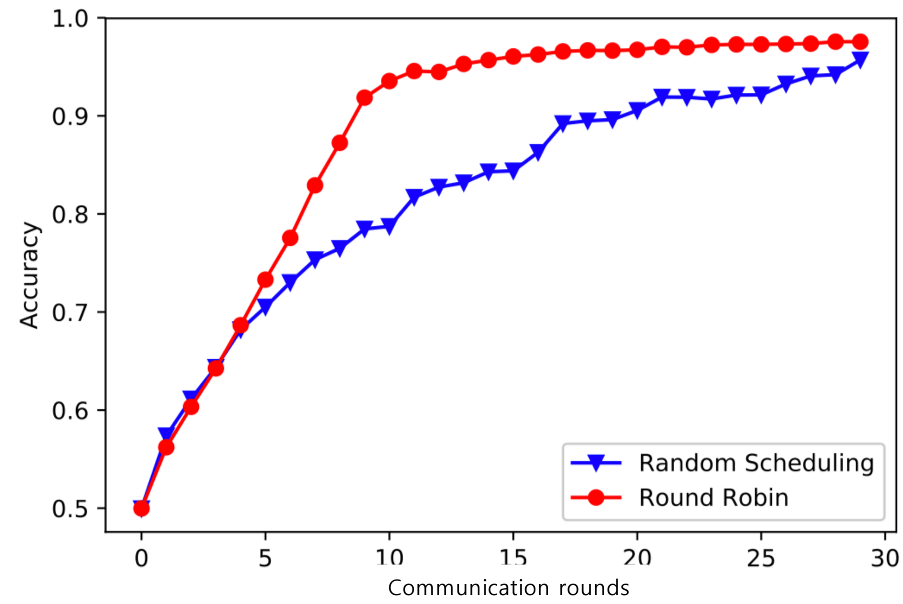




A Conclusion: Scheduling Protocol Matters

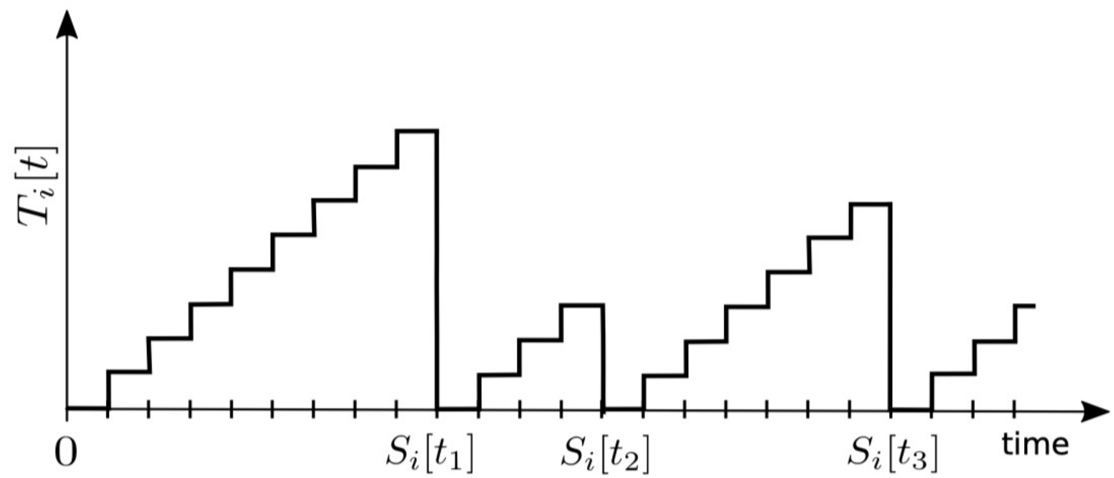
- SVM on MNIST data set
- 10,000 sample points distributed on 100 devices
- Select 20 out of 100 each global aggregation
- Low SINR threshold





Can we optimize scheduling?

Design Metric: Age of Information

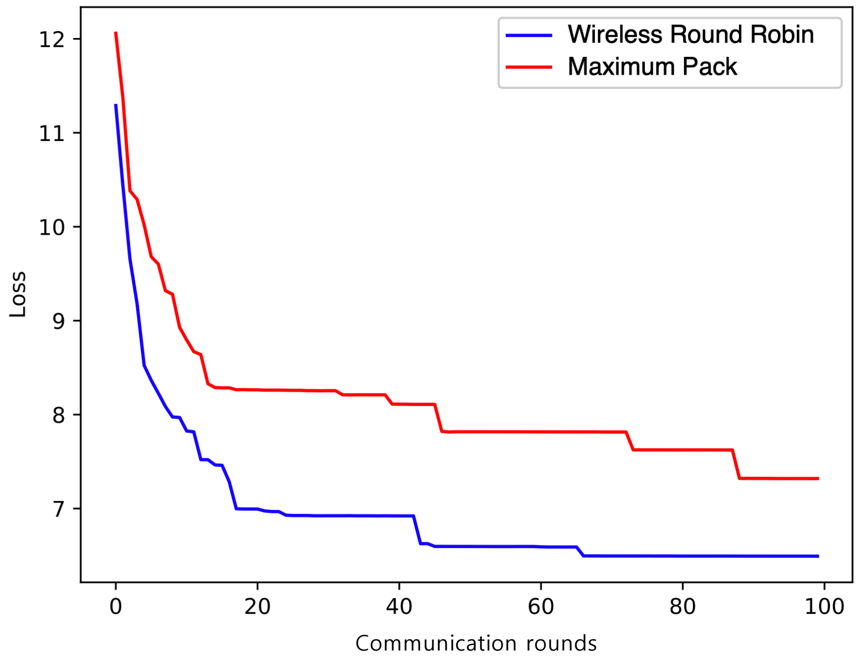


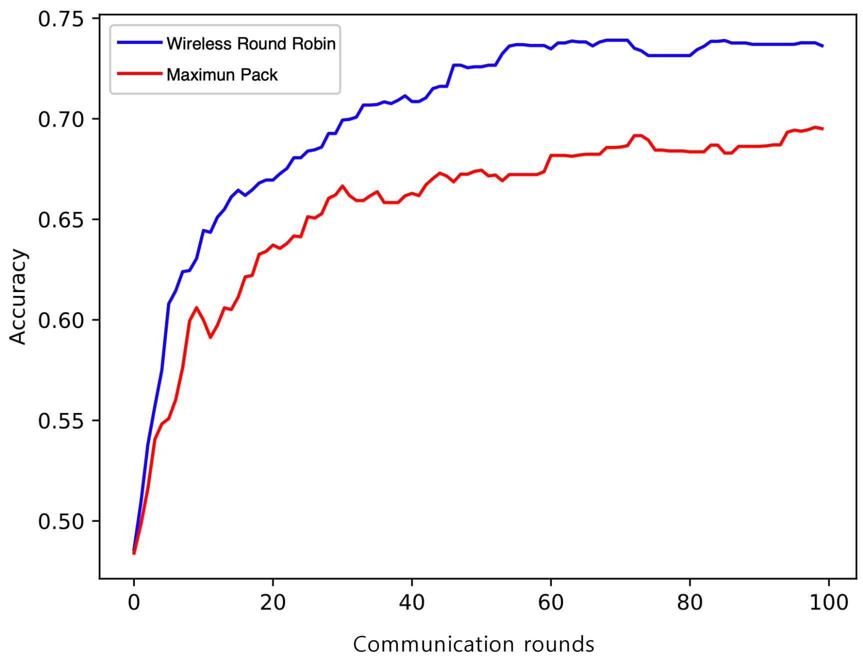
- Metric
  - Age-of-Information (AoI) at a UE i
    - During each communication round, if selected, the AoI drop to 0. Otherwise, the AoI increases by 1: $T_i[t+1] = (T_i[t]+1)(1-S_i[t]), \quad S_i[t] \in \{0,1\}$

Yang, et al. (2020), "Age-Based Scheduling Scheme for Federated Learning in Mobile Edge Networks," *ICASSP*

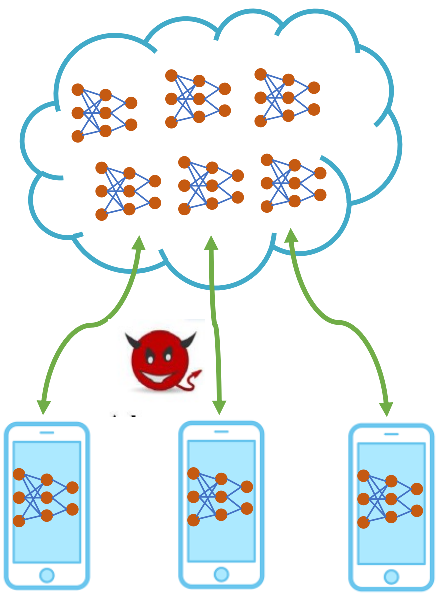
Numerical Results – Minimizing Average AoI

- SVM on MNIST data set
- 10,000 sample points distributed on 100 devices
- Available subchannels: 20





Privacy Protection in Federated Learning (Differential Privacy)



Privacy in Federated Learning

- An early claim for federated learning was that it was "privacy preserving" because the data remains on the end-user devices.

- Subsequent studies have shown that this is not the case, and that end-user data can be inferred from parameter (or gradient) updates.

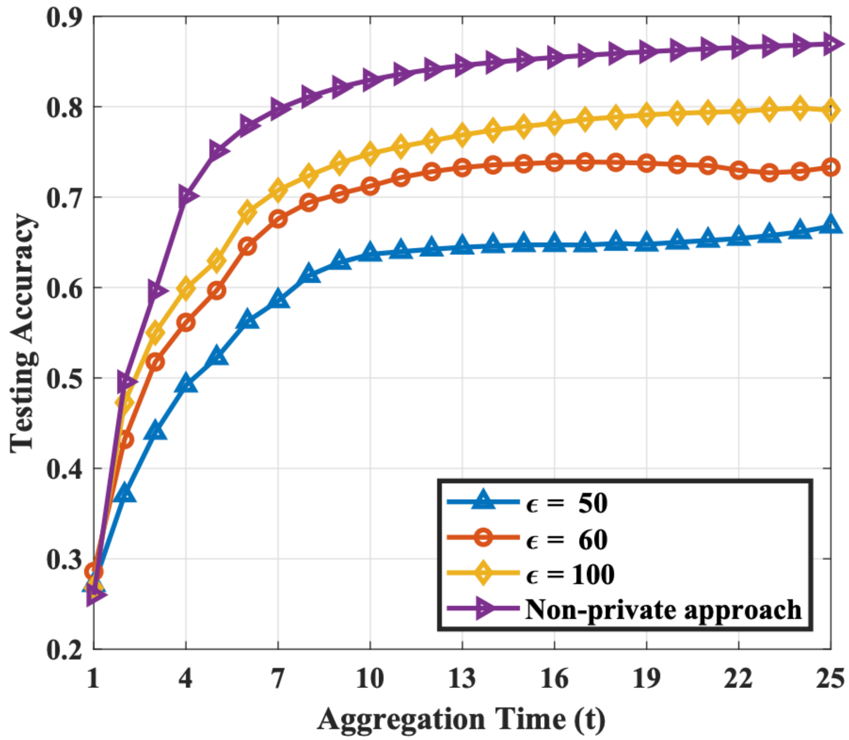
- So, privacy of end-user data is a concern with federated learning.

- One approach is to use differential privacy to protect end-user data.

Differential Privacy in Federated Learning: The Basic Idea

- Generally speaking, differential privacy refers to a type of privacy in which two datasets, one with private information and one without it, but otherwise identical, cannot be distinguished by a statistical query (with high probability).
- Differential privacy can be achieved in many cases by adding noise to data.
- This approach can be used in federated learning.
- This creates a tradeoff between privacy and performance.

Differential Privacy in Federated Learning: An Example



Parameter setting:

- CNN on MNIST data set
- 10,000 sample points distributed on 50 devices

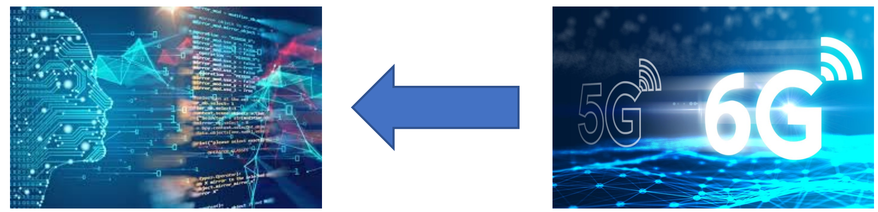
Observations:

- Convergence under differential privacy
- Tradeoff between privacy and accuracy

Wei, et al. (2020), "Federated Learning with Differential Privacy: Algorithm and Performance Analysis," *T-IFS*

Some Research Issues

- <u>Device limitation</u>
  - Resources on end-user devices are limited (e.g., energy, storage, computational power)
  - Fundamental trade-offs between, e.g., # of layers, # of neurons per layer, energy expenditure, accuracy, …
  - Heterogeneous datasets and device capabilities



- <u>Communication efficiency</u>
  - Coded distributed machine learning

- <u>Limited data at the edge</u>
  - Local data is sparse → training sets are usually small
  - Incorporating domain and physics knowledge

- <u>Security & Privacy</u>
  - Robustness to malicious end-user devices & adversarial training examples
  - Server-less implementations (e.g., with blockchain)

Some Recent Papers of Interest

Privacy and Security:

Nguyen, et al. (2021) "Federated Learning Meet Blockchain in Edge Computing," *IEEE IoTJ*

Wei, et al. (2020) "Federated Learning with Differential Privacy: Algorithms and Performance Analysis," *IEEE T-IFS*

Communications Efficiency:

Chen, et al. (2021), "Communication Efficient Federated Learning," *PNAS*

Shlezinger, et al. (2021), "UVeQFed: Universal Vector Quantization for Federated Learning," *IEEE T-SP*

Yang, et al. (2020), "Scheduling Policies for Federated Learning in Wireless Networks," *IEEE T-COM*

Thank You!