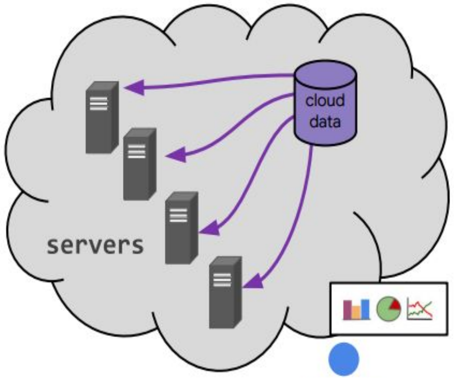
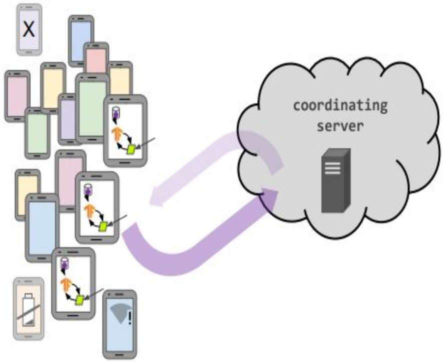
AI/ML for Distributed Sensing and Learning in a networked environment

Introduction

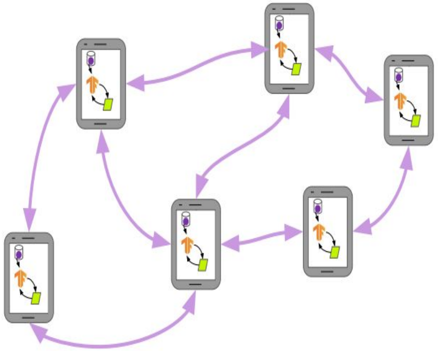
Learning Paradigms



Distributed datacenter machine learning
Parameter Server



Cross-device FL



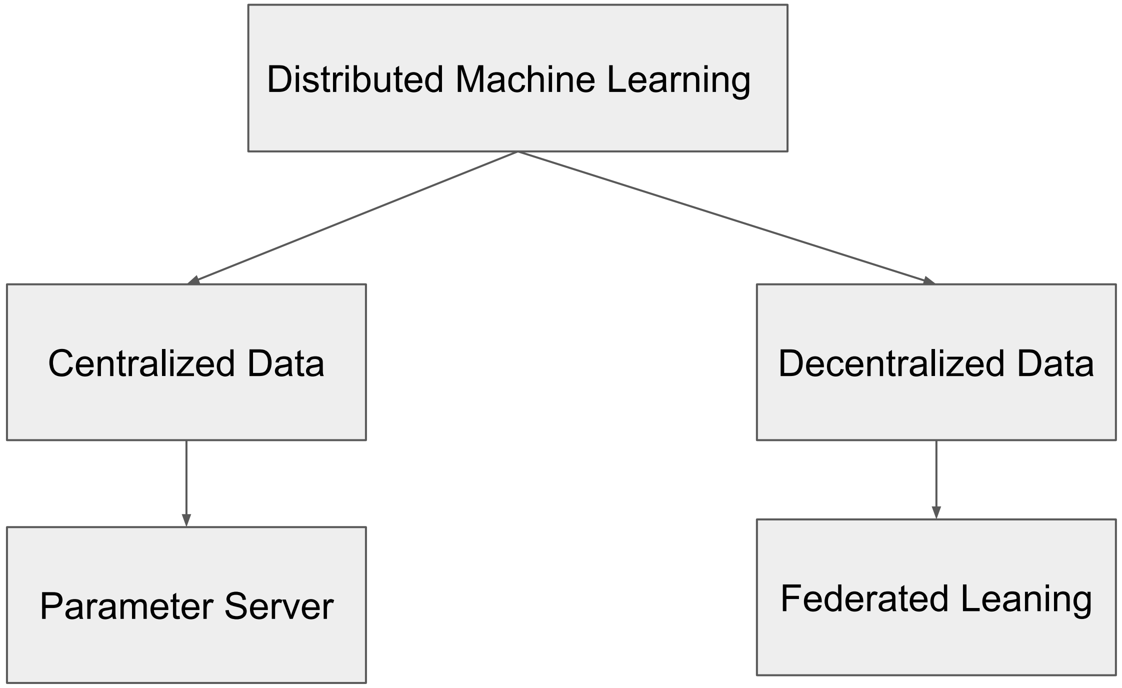
Fully decentralized/peer-to-peer learning

Images taken from NeurIPS 2020 FL Tutorial (https://sites.google.com/view/fl-tutorial)

Learning Paradigms- key distinctions

	Parameter Server	Cross-device FL	Decentralized Learning
Orchestration	Central sever organizes the training and the data distribution	Central server organizes the training but not the data	no central server is required
Data distribution	Centralized	Decentralized	Decentralized
Communication	None	Star topology	Peer-to-peer topology

Distributed Machine Learning



Distributed Machine Learning

Centralized Data

Decentralized Data

Parameter Server

Federated Leaning

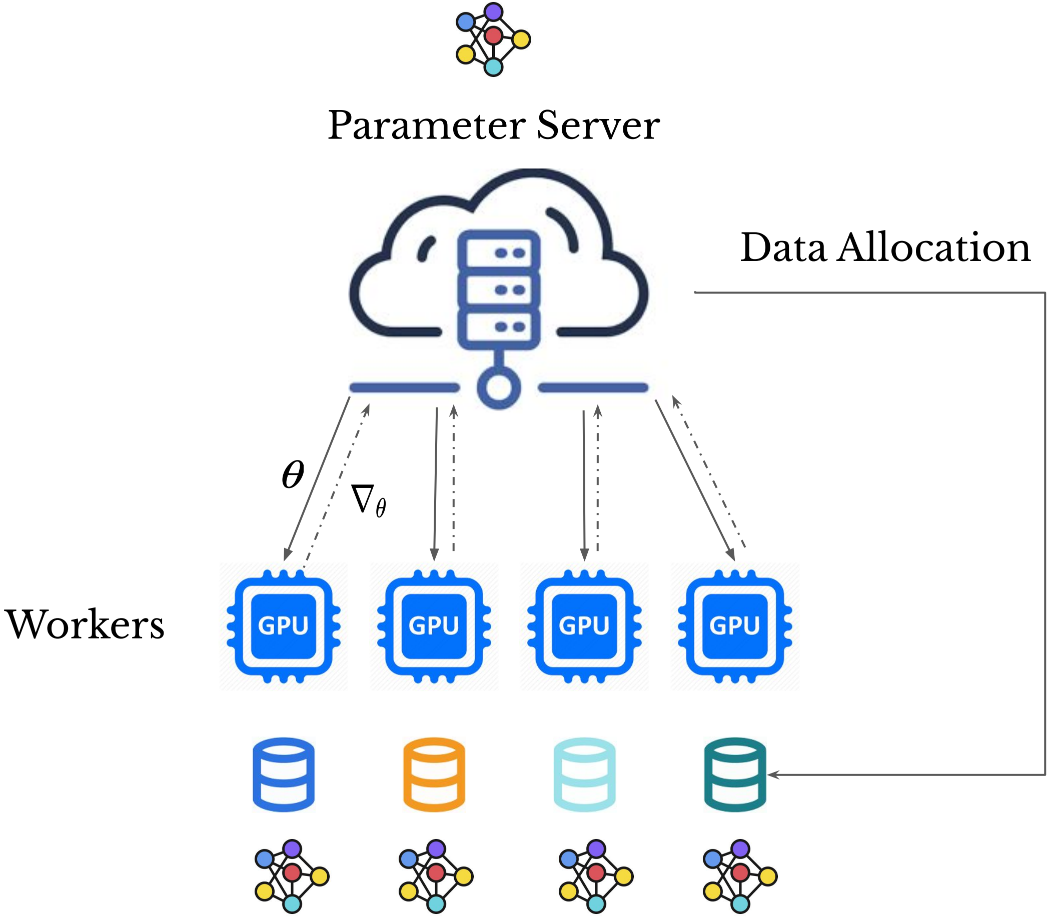
Distributed datacenter machine learning

Distributed Training

1- PS allocates different datasets to different workers and the PS model is copied to the different nodes

2- Workers compute the gradients w.r.t to the data they have and send it back to the PS

3- The PS aggregates the gradients and updates the model

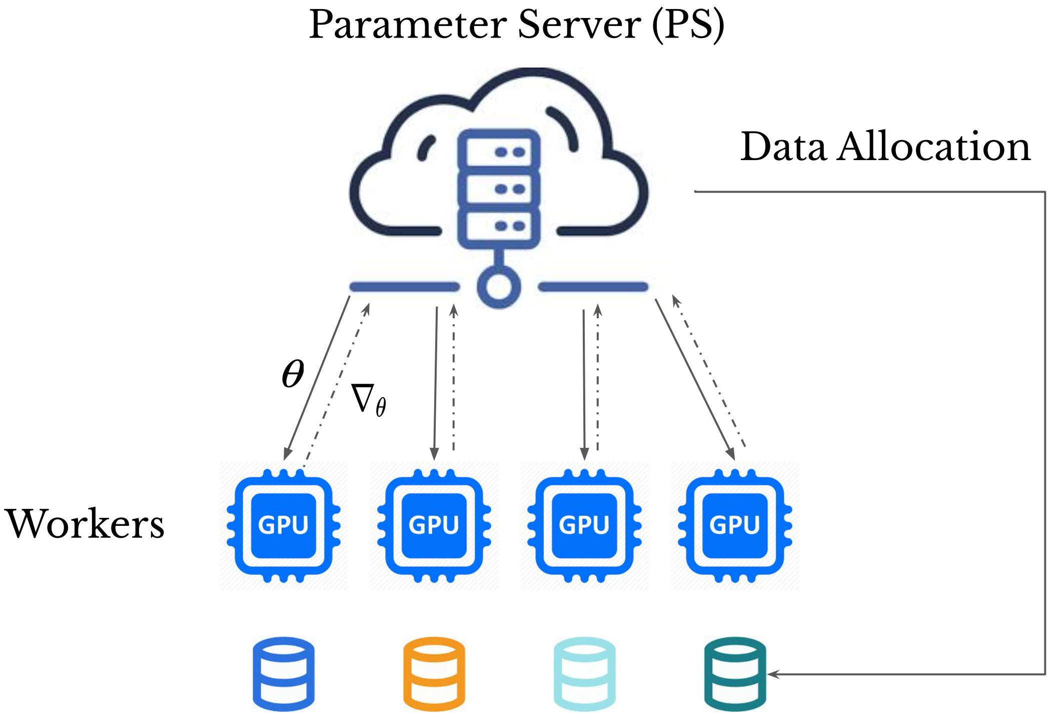


Distributed Training

- Scaling-up ML training [i.e 1]
- Centralized Data : PS owns the data and share it across workers

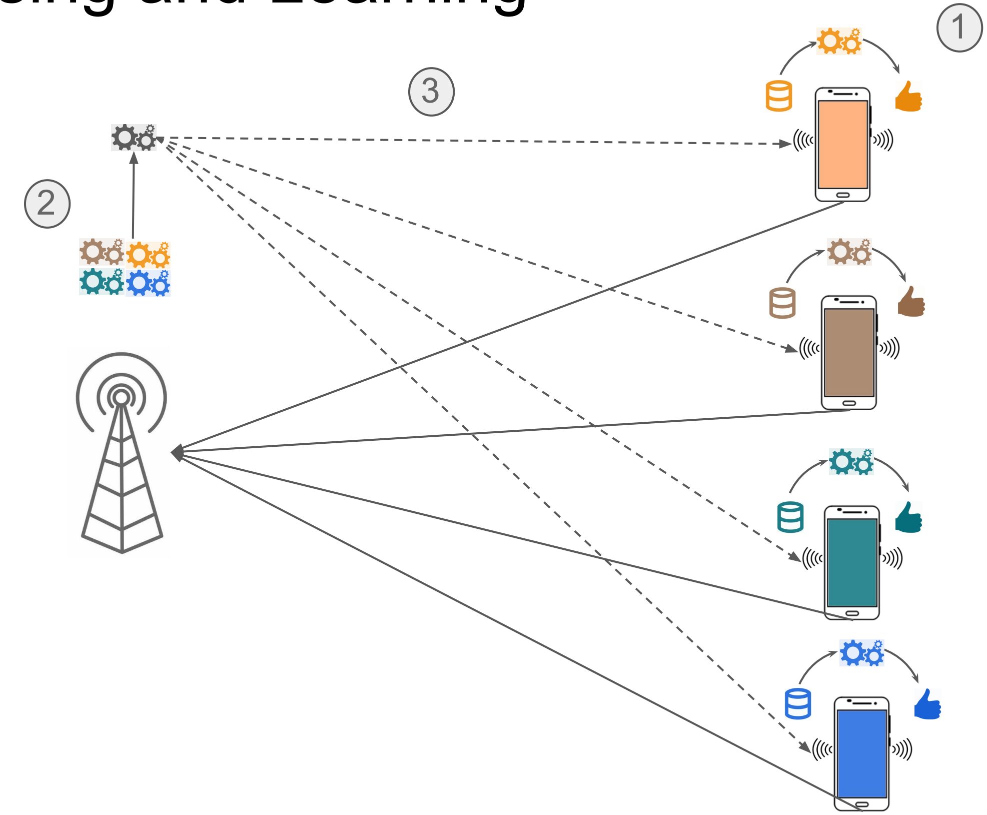
Challenges:

- Hardware/Software failures
- Communication failures
- Robustness to adversarial attacks [2,3]



1.Dean et al. Large Scale Distributed Deep Networks. NeurIPS (2012)
2.Blanchard et al. Machine learning with adversaries: Byzantine tolerant gradient descent. NeurIPS (2017).
3. Yin et al. Byzantine-Robust Distributed Learning: Towards Optimal Statistical Rates. ICML (2018).

Federated Sensing and Learning

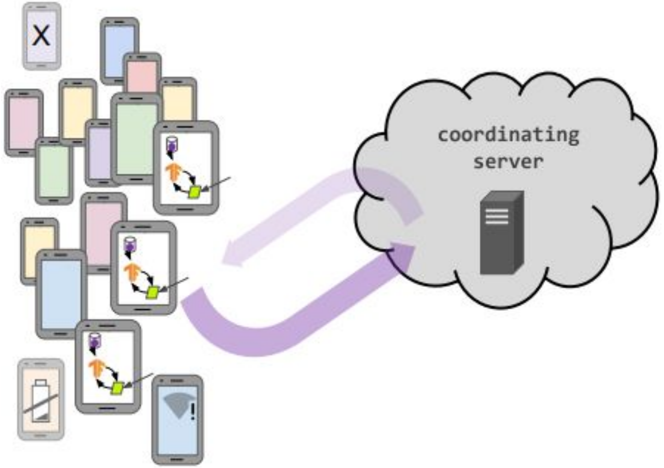


(1)Each edge device collects data through its **sensors** and train a **personalized** model **locally**. (2) Base station **aggregates** a **selection** of the users' models to construct a **global** one. (3) The updated global model is **broadcasted** to the users

Figure inspired from [1]

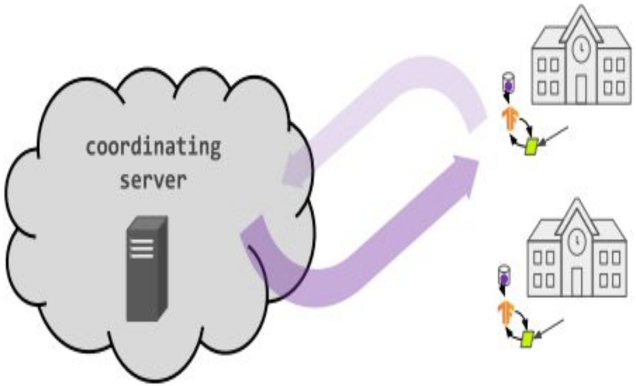
Federated Learning settings

Cross-device FL



- High number of available devices
- Only a random sample is available in each round
- horizontally partitioned data

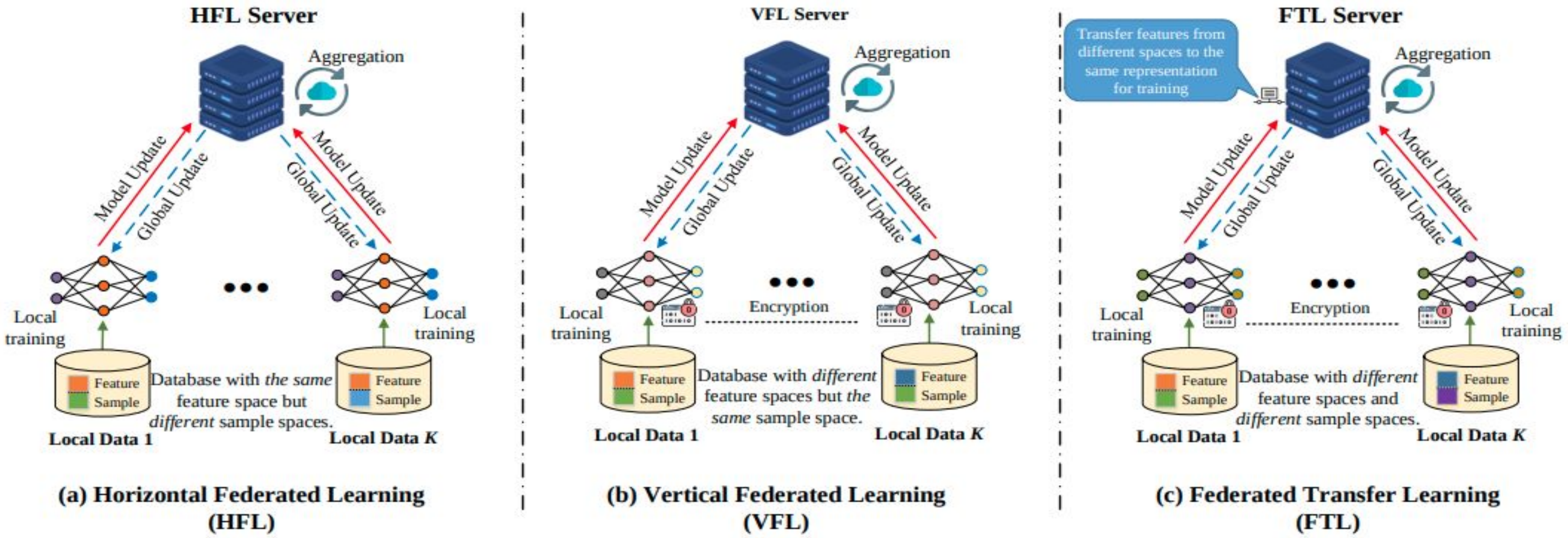
Cross-silo FL



- Small number of available clients (i.e institutions/ hospitals)
- Most clients participate in each round
- horizontally or vertically partitioned data

*slide inspired from Advances and Open Problems in Federated Learning (arxiv/1912.04977)

Data partitioning schemes in FL



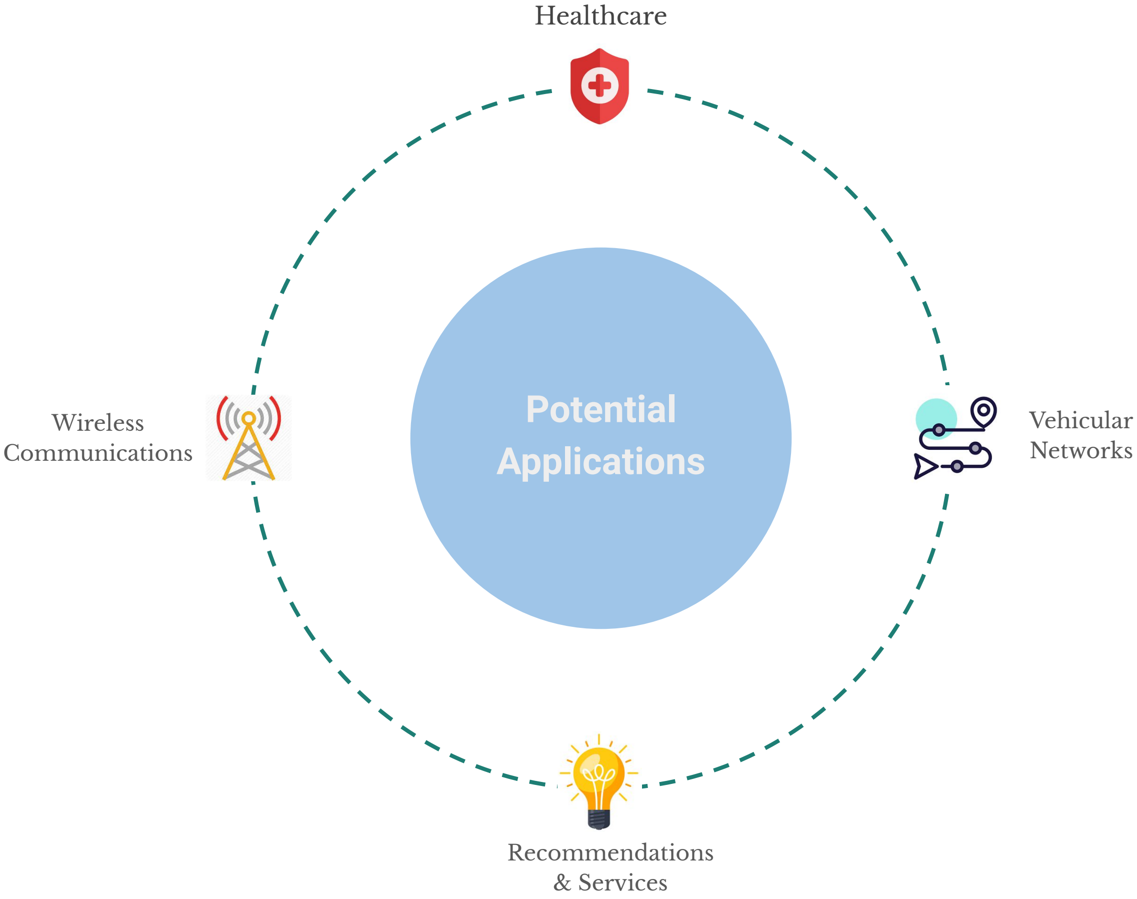
(a) Horizontal Federated Learning (HFL)

(b) Vertical Federated Learning (VFL)

(c) Federated Transfer Learning (FTL)

*Image taken from https://arxiv.org/pdf/2104.07914.pdf

Federated Sensing and Learning



Applications for wireless communication

Edge Computing and Caching

- Content popularity identification models based on the user-content interaction

- User preference prediction models

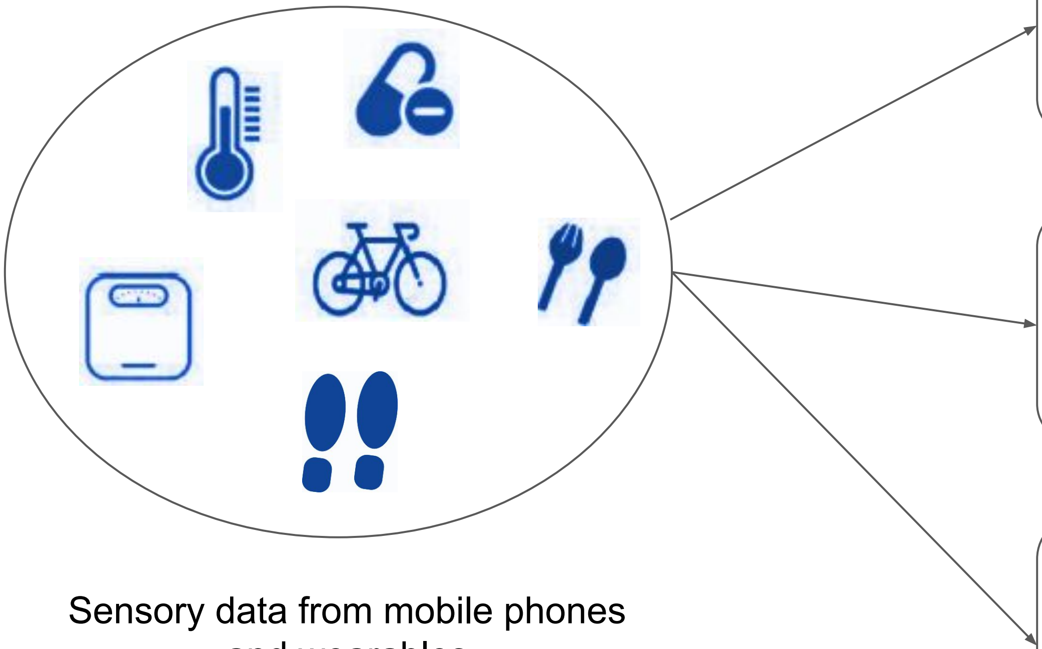
Spectrum Management

- Spectrum utilization predictive models
- Spectrum sharing models

Wireless resource allocation

- Power control

Applications in Medical IoT



Sensory data from mobile phones and wearables

Intelligent health monitoring and analytics: Given sensory information, smart and real-time diagnosis can be achieved. More accurate predictive models can be obtained using FL

Smart fitness programs: Based on activities of a group of persons in a geographic area, FL model can be used to suggest and recommend new fitness and workout programs

Disease prevention/management: Sensory data can be used to build insights about the health or condition of a person, or a group of individuals. Predictive models are learned to assist the users and inform the, about possible precautionary measures or prevention strategies.

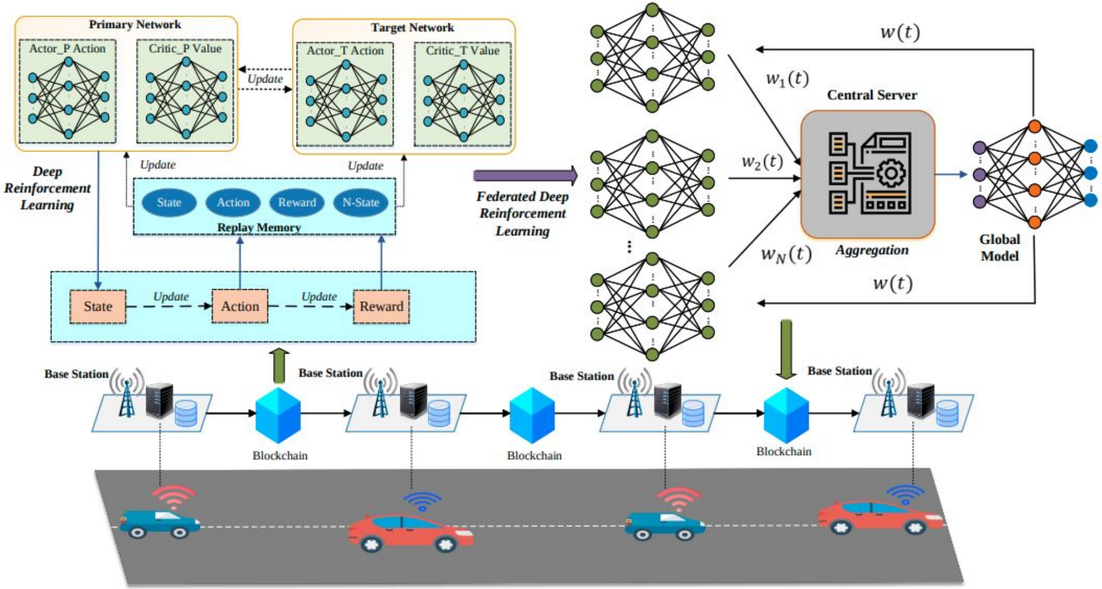
Applications in Vehicular Networks

1- Node selection using DRL

2- The participating vehicles send their local model updates to the nearby RSU and uploads it to the blockchain for further verification and aggregation.

3- The aggregator retrieves the updated local parameters from the permissioned blockchain and executes global aggregation by aggregating the local models

⇒ Traffic prediction
⇒ Smart navigation



The architecture of federated learning-based data sharing for IOV with blockchain*

* https://ieeexplore.ieee.org/document/8998397

Applications for mobile devices

- **Prediction on keyboard** : Enhance the suggestion quality and next-word prediction
- **Mobility prediction:** Use mobility motion sensor data to perform privacy aware mobility prediction [1] and human activity recognition [2]
- **Detection of hazards** in a smart home environment [3]
- **Active user authentication** using sensory data (i.e [4])

[1] Feng, J., Rong, C., Sun, F., Guo, D., & Li, Y. (2020). PMF: A privacy-preserving human mobility prediction framework via federated learning. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(1), 10:1–10:21. htt ps://doi.org/10.1145/3381006

[2] Sozinov, K., Vlassov, V., & Girdzijauskas, S. (2018). Human activity recognition using federated learning. In 2018 IEEE Intl Conf on Parallel Distributed Processing with Applications, Ubiquitous Computing Communications, Big Data Cloud Computing, Social Computing Networking, Sustainable Computing Communications (ISPA/ IUCC/BDCloud/SocialCom/SustainCom) (pp. 1103–1111). https://doi.org/10.1109 /BDCloud.2018.00164.

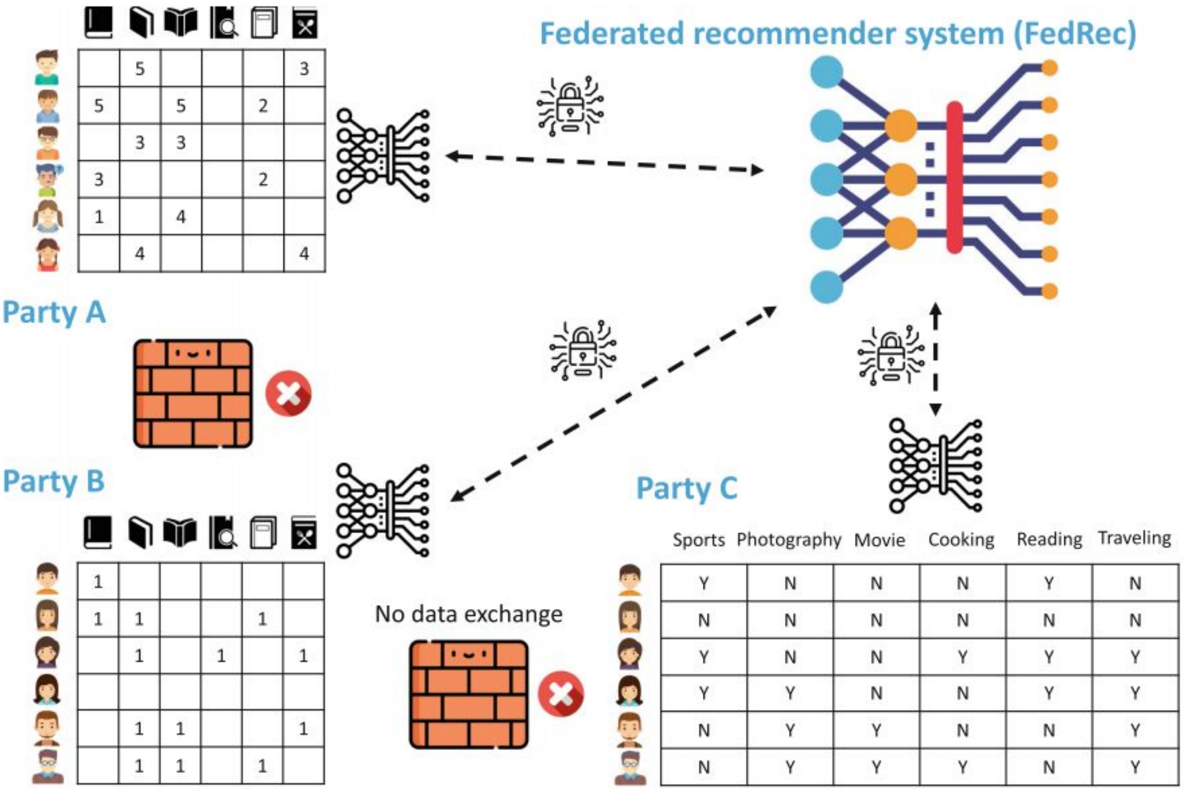
[3] Yu, T., Li, T., Sun, Y., Nanda, S., Smith, V., Sekar, V., et al. (2020). Learning contextaware policies from multiple smart homes via federated multi-task learning. IEEE/ ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), 2020, 104–115. https://doi.org/10.1109/IoTDI49375.2020.00017.

[4] P. Oza and V. M. Patel, “Federated learning-based active authentication on mobile devices,” CoRR, vol. abs/2104.07158, 2021.

Federated recommendation systems*

FedRec aims to collaboratively train recommendation model(s) among multiple parties without direct access to the private data of each other*

⇒Collaborative movie/apps recommendations



Party A

	5				3
5		5		2	
	3	3			
3				2	
1		4			
	4				4

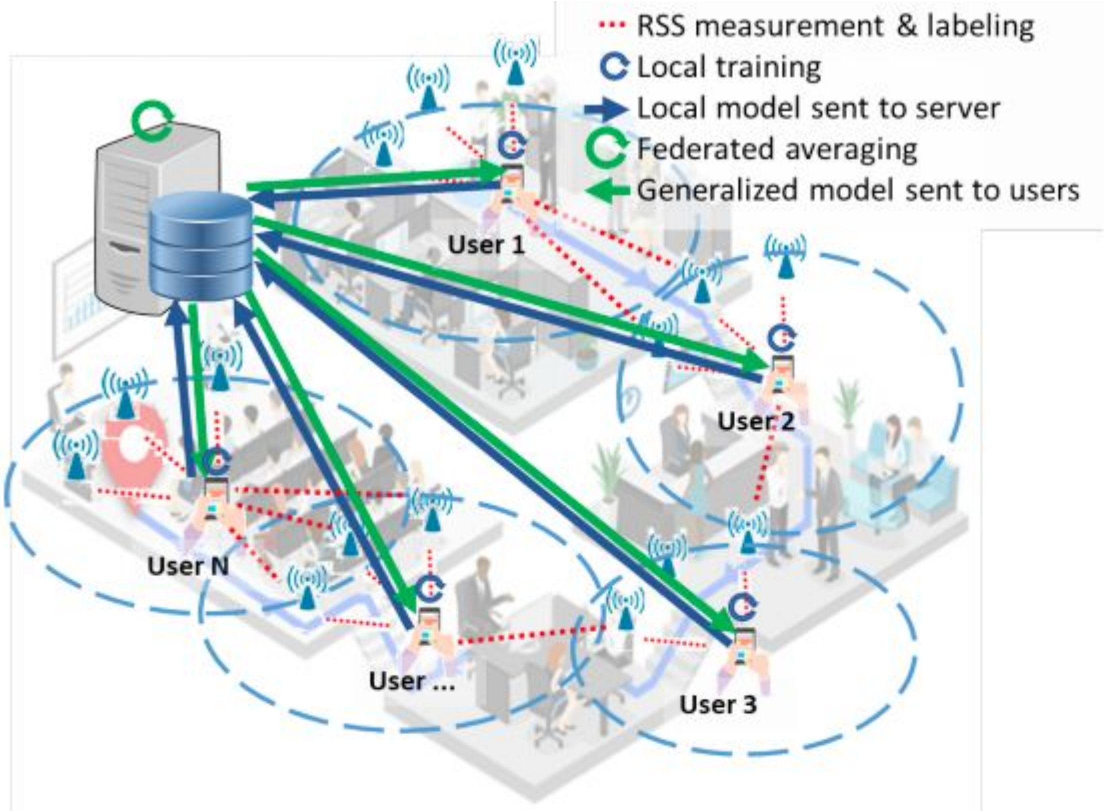
Party B

1					
1	1			1	
	1		1		1
	1	1			1
	1	1		1	

Party C

Sports	Photography	Movie	Cooking	Reading	Traveling
Y	N	N	N	Y	N
N	N	N	N	N	N
Y	N	N	Y	Y	Y
Y	Y	N	N	Y	Y
N	Y	Y	N	N	Y
N	Y	Y	Y	N	Y

* L. Yang, B. Tan, V. W. Zheng, K. Chen, and Q. Yang, “Federated recommendation systems,” in Federated Learning,pp. 225–239, Springer, 2020.

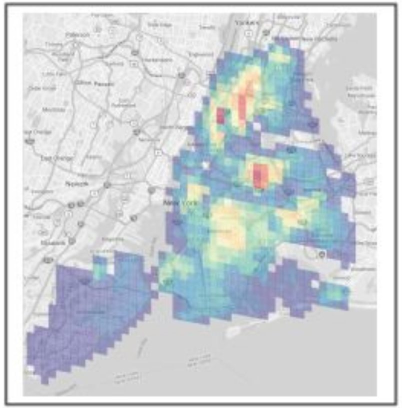

RSS measurement & labeling
Local training
Local model sent to server
Federated averaging
Generalized model sent to users
User 1
User 2
User 3
User ...
User N

Federated learning for Received Signal Strength (RSS) fingerprint-based localization

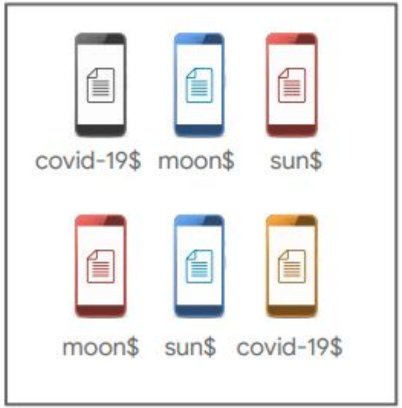
https://arxiv.org/pdf/2001.01911.pdf

Federated Analytics

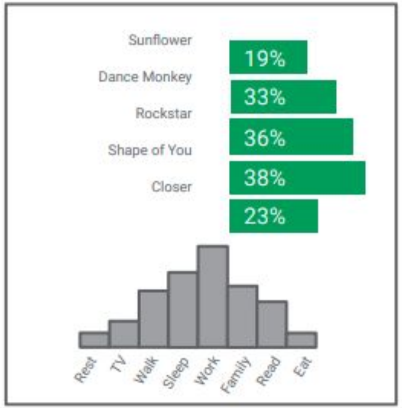
Federated analytics is the practice of applying data science methods to the analysis of raw data that is stored locally on users’ devices. Like federated learning, it works by running local computations over each device’s data, and only making the aggregated results — and never any data from a particular device



Geo-location heatmaps



Frequently typed out-of-dictionary words



Popular songs, trends, and activities

Definition proposed in https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.htm

Figure from (https://sites.google.com/view/fl-tutorial)

Federated Learning challenges

Communication

- Can be the primary bottleneck
  - Communication using wireless links is slower than datacenter links
  - Local updates are faster than communication
  - Massive number of connected devices
  - Limited bandwidth
- Can be reduced:
  - Limiting the number of devices in each communication round
  - Reducing the number of communication rounds (more local updating)
  - Reducing the size of messages (compression techniques)

Federated Learning challenges

Heterogeneity

- Non-IID across devices
- Lack of convergence guarantees for non-IID case (slower convergence, less stability, divergence)
- Heterogeneous devices (i.e hardware, network connectivity, and battery power) ⇒ stragglers
  - Asynchronous schemes help mitigating the stragglers
  - Active device sampling (i.e sampling based on the system resources)

Federated Learning challenges

Privacy : Model updates can reveal sensitive information

Robustness:

- **Data poisoning:** Adversary corrupt on-device data during local updates
- **Model poisoning:** Adversary corrupt the model updates

Fairness across devices:

- Reduce the model bias because of stragglers and device failures

Federated Learning Convergence

- N : Total number of users
- M : users per round
- T : total communication rounds
- K : local steps per round

Assumptions:

- $f(.\,, s_i)$ is H-smooth (i.e differentiable and has H-Lipschitz gradients)
- The stochastic gradient satisfies

$$\mathbb{E}_s ||\nabla_x f(x.\, s) - \nabla F(x)|| \leq \sigma^2$$

Federated Learning optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^d} \left[F(\mathbf{x}) := \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}; s_i) \right]$$

User i data

Federated Learning Convergence for IID case

convergence as function of number of iterations

Method	Convergence	Comments
mini-batch SGD	$\mathcal{O}\left(\frac{H}{T} + \frac{\sigma}{\sqrt{TKM}}\right)$	batch size K*M
SGD	$\mathcal{O}\left(\frac{H}{TK} + \frac{\sigma}{\sqrt{TK}}\right)$	1 worker
Fed-Averaging/Local SGD - [1] - [2]	$\mathcal{O}\left(\frac{HKM}{T}\frac{G^2}{\sigma^2} + \frac{\sigma}{\sqrt{TKM}}\right)$ $\mathcal{O}\left(\frac{HM}{T} + \frac{\sigma}{\sqrt{TKM}}\right)$	The bounded gradient norm assumption

[1] Hao Yu, Sen Yang, and Shenghuo Zhu. Parallel restarted SGD for non-convex optimization with faster convergence and less communication. arXiv preprint arXiv:1807.06629, 2018.

[2] Jianyu Wang and Gauri Joshi. Cooperative SGD: A unified framework for the design and analysis of communication-efficient SGD algorithms. preprint, August 2018. URL https://arxiv.org/abs/ 1808.07576.

Table inspired from P. Kairouz et al. , “Advances and open problems in federated learning,” arXiv preprint arXiv:1912.04977 , 2019.

Federated Learning Convergence for non-IID case

Method	Assumptions	Variant	Rate
Lian et al. [1]	BCGV/BLGV	Dec; AC; 1step	$O(1/T) + O(1/\sqrt{NT})$
PD-SGD [2]	BCGV/BLGV	Dec; AC	$O(N/T) + O(1/\sqrt{NT})$
MATCHA [4]	BCGV/BLGV	Dec	$O(1/\sqrt{TKM}) + O(M/KT)$
Khaled et al. [3]	BOGV/CVX	AC; LBG	$O(N/T) + O(1/\sqrt{NT})$
Li et al. [5]	BOBD/SCVX; BLGV; BLGN		$O(K/T)$
FedProx [6]	BGV/BNCVX	Prox	$O(1/\sqrt{T})$

[1] Xiangru Lian, Ce Zhang, Huan Zhang, Cho-Jui Hsieh, Wei Zhang, and Ji Liu. Can Decentralized Algorithms Outperform Centralized Algorithms? A Case Study for Decentralized Parallel Stochastic Gradient Descent. In NIPS, 2017

[2] Xiang Li, Wenhao Yang, Shusen Wang, and Zhihua Zhang. Communication efficient decentralized training with multiple local updates. arXiv preprint arXiv:1910.09126, 2019.

[3] Ahmed Khaled, Konstantin Mishchenko, and Peter Richtarik. First analysis of local GD on heterogeneous data, ´ 2019. URL https://arxiv.org/abs/1909.04715

[4] Jianyu Wang, Anit Sahu, Gauri Joshi, and Soummya Kar. MATCHA: Speeding Up Decentralized SGD via Matching Decomposition Sampling. preprint, May 2019. URL https://arxiv.org/abs/1905. 09435.

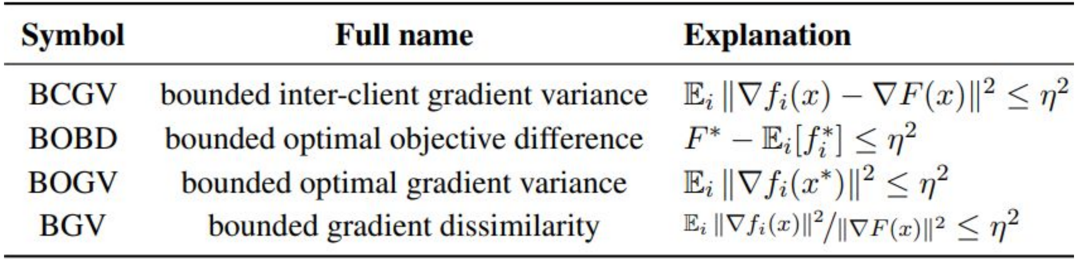
[5] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of FedAvg on non-IID data. arXiv preprint arXiv:1907.02189, 2019.

[6] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks, 2018. URL https://arxiv.org/abs/1812.06127.

Table inspired from P. Kairouz et al. , “Advances and open problems in federated learning,” arXiv preprint arXiv:1912.04977 , 2019.

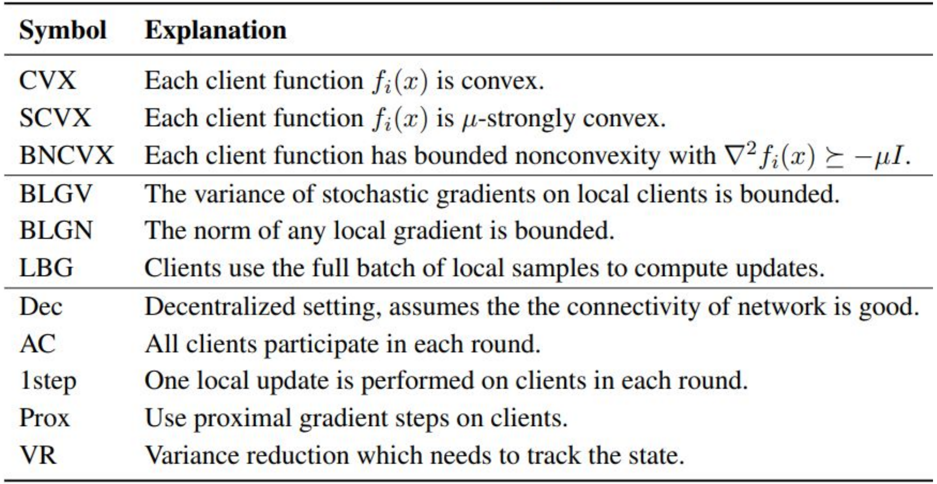
Federated Learning Convergence for non-IID case

Non-IID assumptions



Symbol	Full name	Explanation
BCGV	bounded inter-client gradient variance	$\mathbb{E}_i \|\nabla f_i(x) - \nabla F(x)\|^2 \le \eta^2$
BOBD	bounded optimal objective difference	$F^* - \mathbb{E}_i[f_i^*] \le \eta^2$
BOGV	bounded optimal gradient variance	$\mathbb{E}_i \|\nabla f_i(x^*)\|^2 \le \eta^2$
BGV	bounded gradient dissimilarity	$\mathbb{E}_i \|\nabla f_i(x)\|^2 / \|\nabla F(x)\|^2 \le \eta^2$

Other Assumptions and Federated Averaging variants

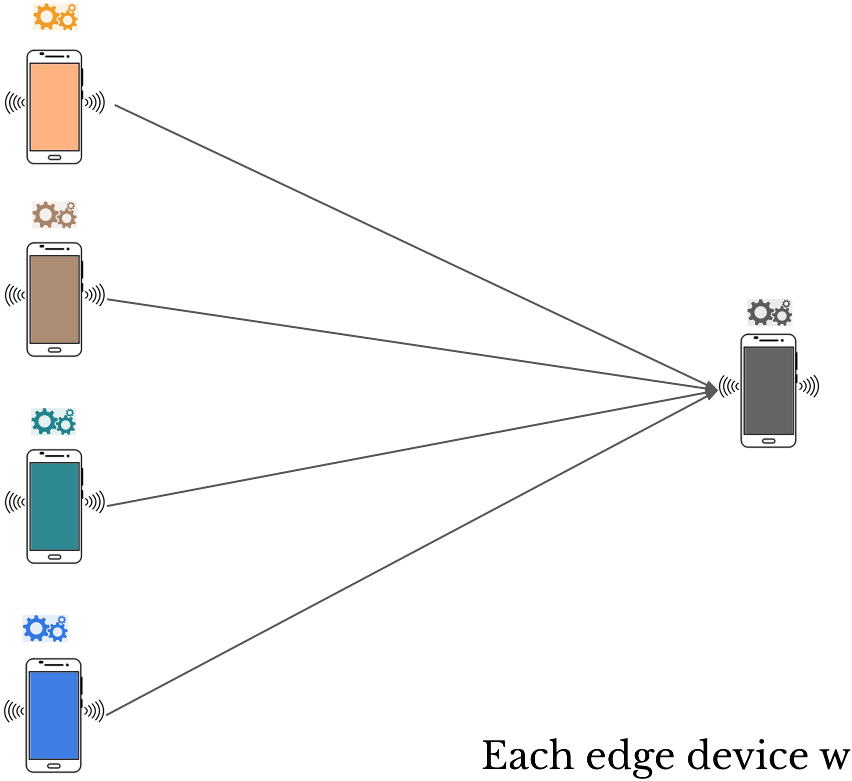


Symbol	Explanation
CVX	Each client function $f_i(x)$ is convex.
SCVX	Each client function $f_i(x)$ is μ-strongly convex.
BNCVX	Each client function has bounded nonconvexity with $\nabla^2 f_i(x) \succeq -\mu I$.
BLGV	The variance of stochastic gradients on local clients is bounded.
BLGN	The norm of any local gradient is bounded.
LBG	Clients use the full batch of local samples to compute updates.
Dec	Decentralized setting, assumes the the connectivity of network is good.
AC	All clients participate in each round.
1step	One local update is performed on clients in each round.
Prox	Use proximal gradient steps on clients.
VR	Variance reduction which needs to track the state.

More details in P. Kairouz et al. , "Advances and open problems in federated learning," arXiv preprint arXiv:1912.04977 , 2019.

Related Learning Approaches

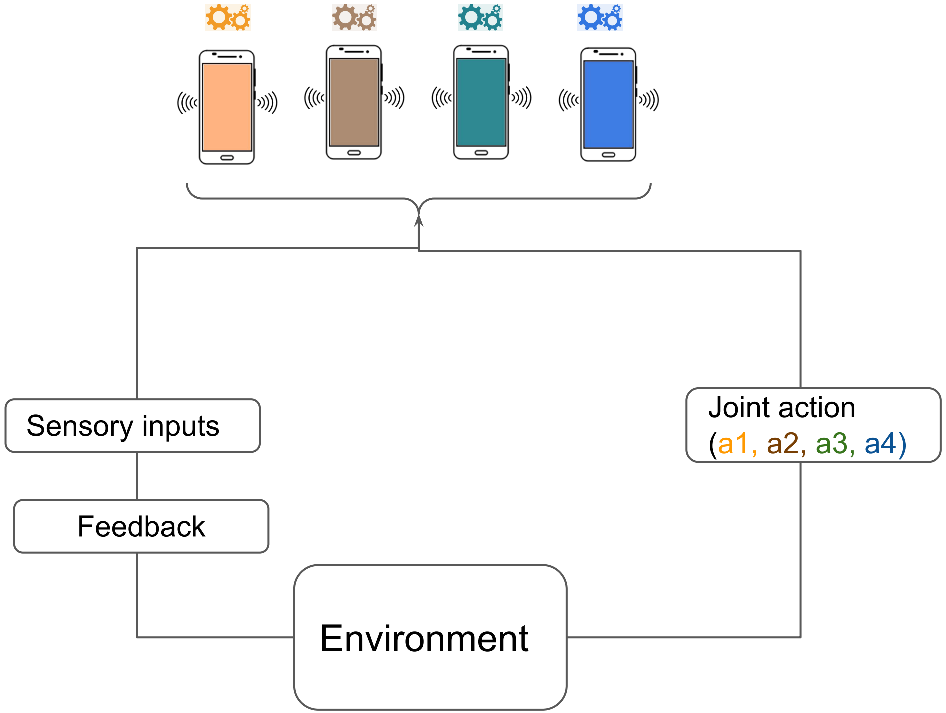
Transfer Learning



- Knowledge Distillation (supervised)
- Domain/Knowledge Adaptation (unsupervised)

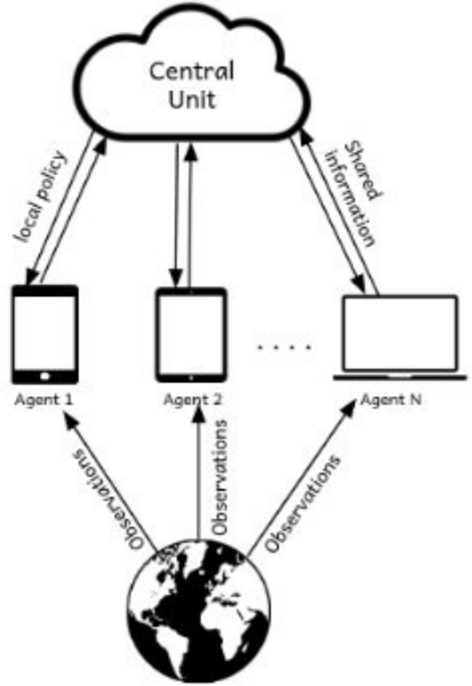
Each edge device will have its own personalized model, no global model. Useful if we want the users' models to be more focused on the users data

Multi-agent Reinforcement Learning (MARL)

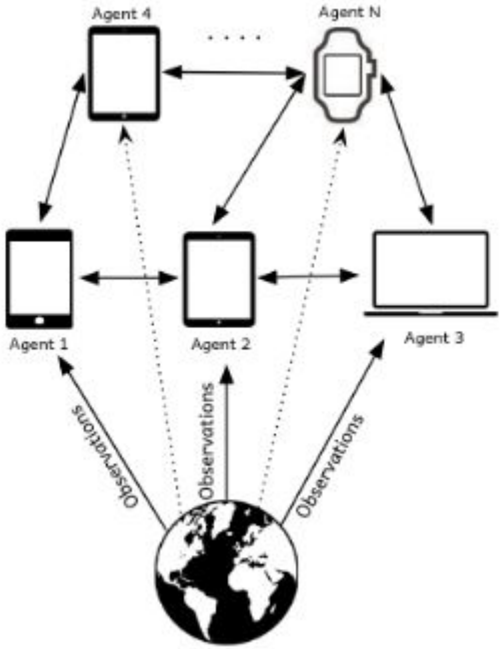


- each mobile device has its own policy/model (No global model)
- Possibility to cooperate between devices
- Each mobile device (i.e) agent receives sensory data from its environment and learns a model to control or act in the environment

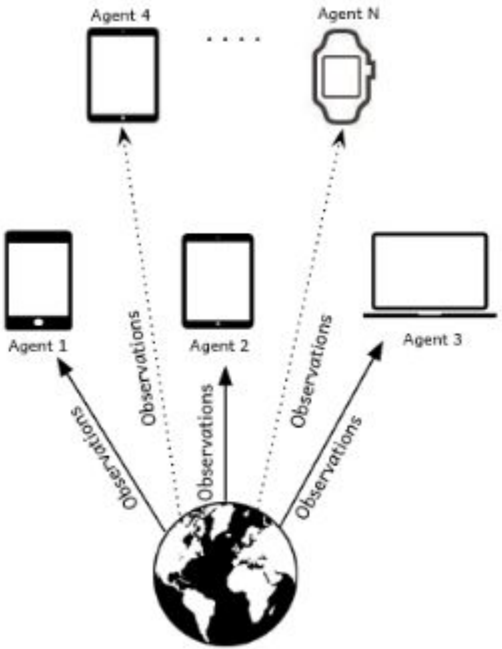
MARL learning frameworks



Centralized Training
Decentralized Execution



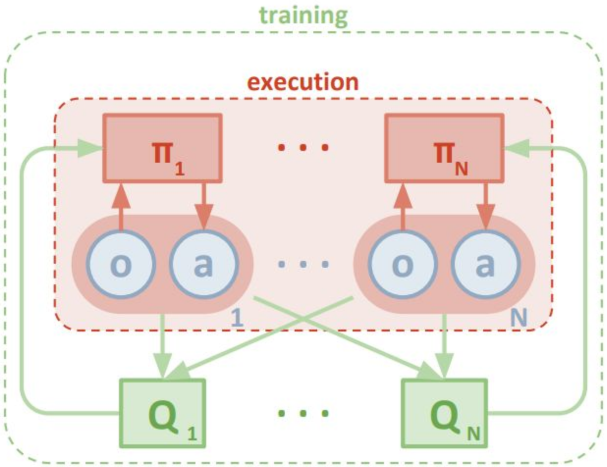
Networked
Agents/Peer-to-Peer
learning



Fully decentralized

Multi-Agent DDPG (MADDPG)

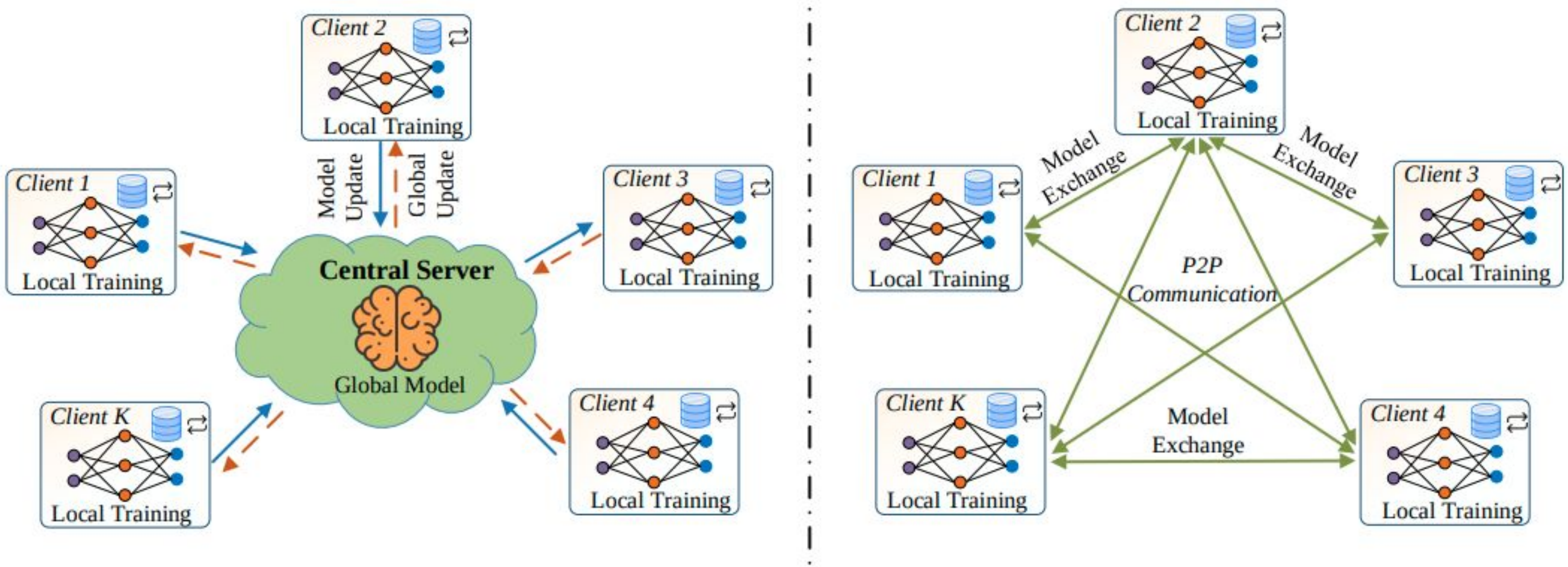
- Extension of the DDPG algorithm
- Training phase
  - **Centralized critic** based on the observations and actions of **all agents** to overcome the non-stationarity problem
- Execution phase
  - Learned actors use **local** information only to pick actions
  - No need for the central critic after training is finished
- Applicable for collaborative and competitive tasks
- Possibility to include communication between agents and/or opponent modeling



Reference

```
@article{xia2021survey,
  title={A Survey of Federated Learning for Edge Computing: Research Problems and Solutions},
  author={Xia, Qi and Ye, Winson and Tao, Zeyi and Wu, Jindi and Li, Qun},
  journal={High-Confidence Computing},
  pages={100008},
  year={2021},
  publisher={Elsevier}
}
```

Federated Sensing and Learning



(a) Centralized Federated Learning (CFL)

(b) Decentralized Federated Learning (DFL)

Types of FL models with networking structure*

*Image taken from https://arxiv.org/pdf/2104.07914.pdf