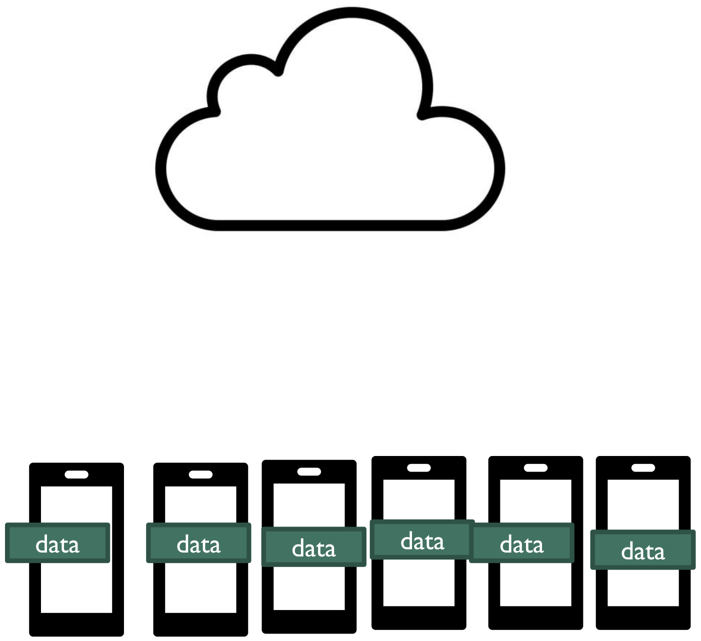
INTRODUCTION: FEDERATED LEARNING

Motivating Examples



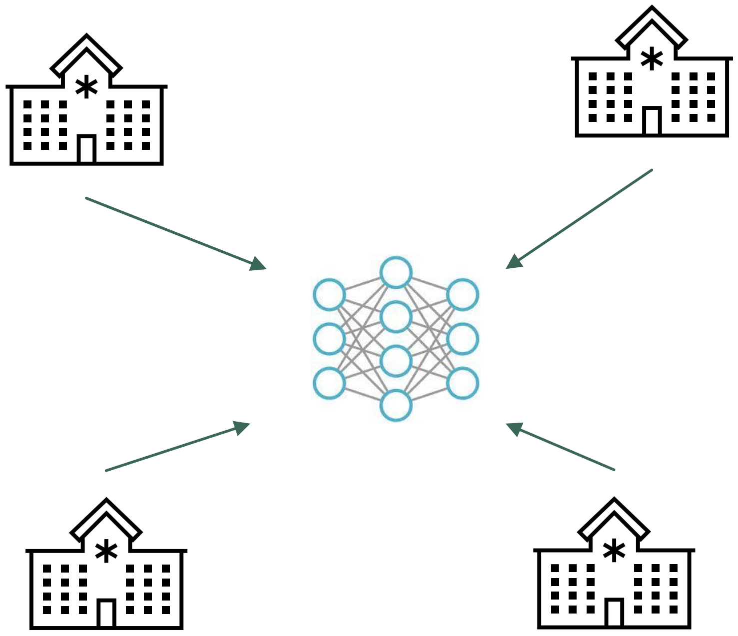
Problem: Google wants to train a model using users' mobile data.

Possible solution: Centralized learning
- Collect users' data.
- Train a model on the cluster.

Challenge: Users may refuse to upload their data, especially sensitive data, to Google's server.

INTRODUCTION

Motivating Examples



Problem: Hospitals want to jointly train a model using medical data.

Possible solution: Centralized learning

- Aggregate the data.
- Train a model on the server.

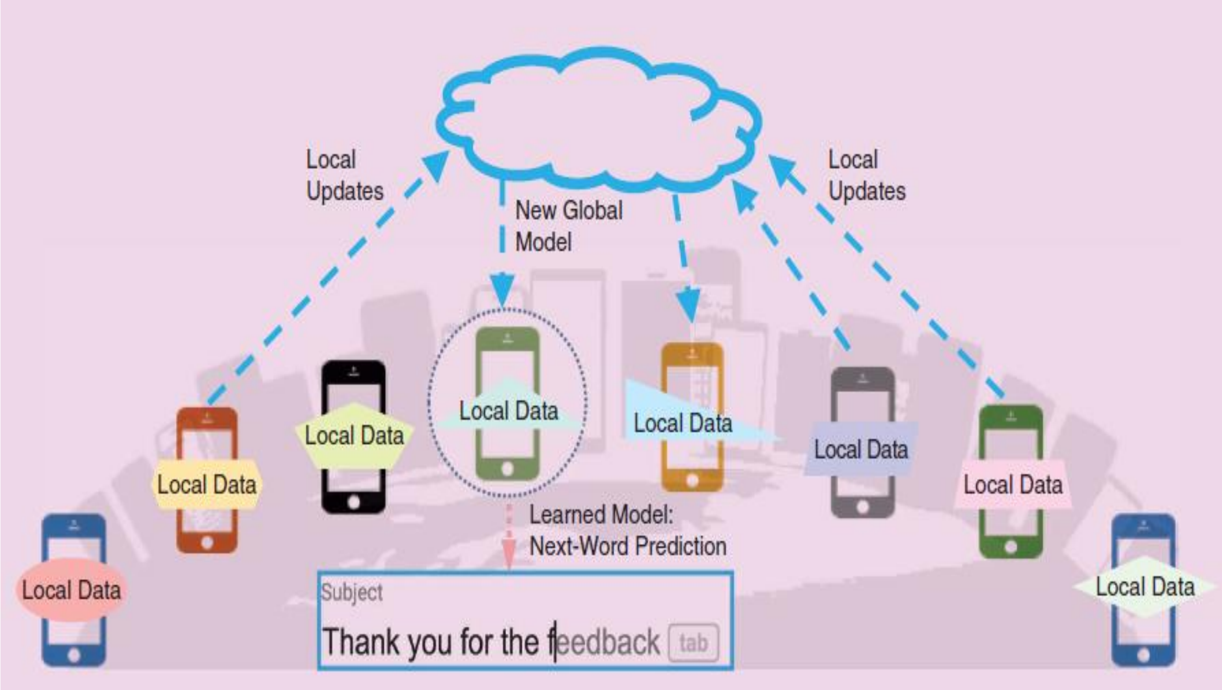
Challenge: Laws or policies may forbid giving patients' data to others.

INTRODUCTION

- The standard federated learning problem involves learning a single global statistical model to potentially millions of remote devices.

- Challenges:
  - *Expensive communication*
  - *Privacy concerns*
  - *Systems heterogeneity*
  - *Statistical heterogeneity*

    (Non-IID, Independent and identically distributed)



WHAT IS FEDERATED LEARNING?

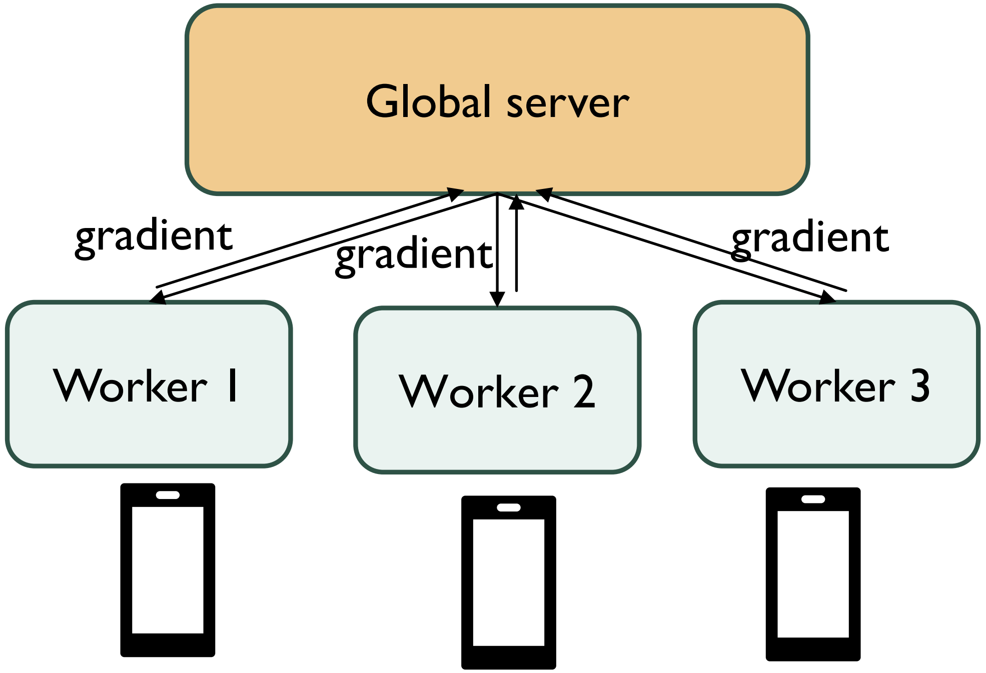
- User have control over their device and data
- Worker nodes are unstable
- Communication cost is higher than computation cost
- Data stored on worker nodes are not IID
- The amount of data is severely imbalanced

Reference :

H. Brendan McMahan, Eider Moore and others : **Communication-Efficient Learning of Deep Networks from Decentralized Data ,** *17 Feb 2016* , in Artificial Intelligence and Statistics (AISTATS) 2017

COST FUNCTION IN FEDERATED SYSTEM

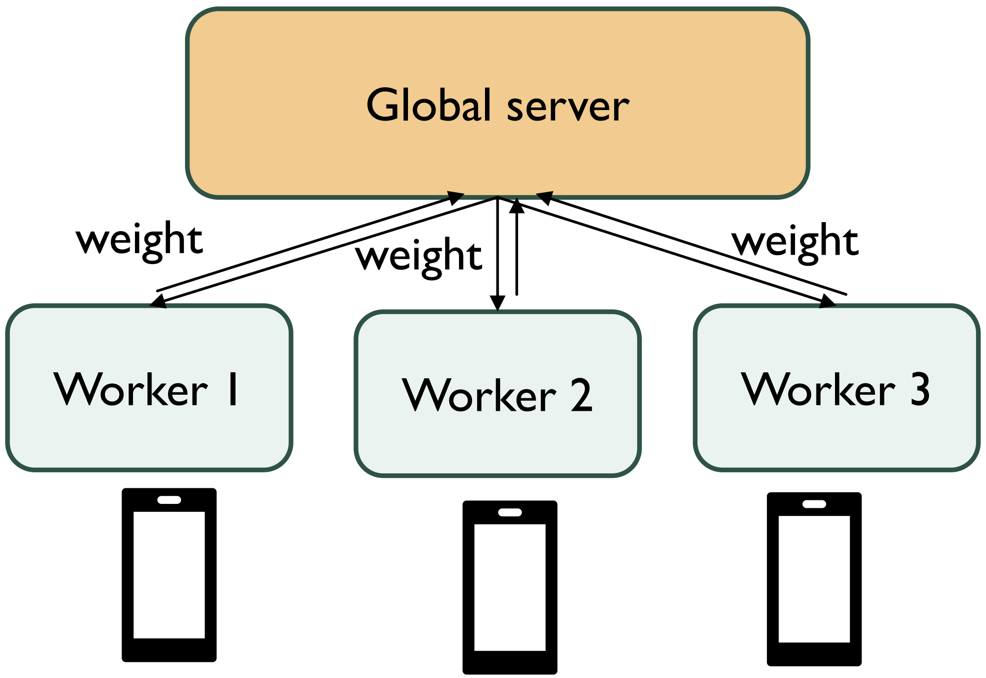
- Gradient Descent



1. Server 下放weight給client去學習
2. Client 上傳 gradient g1, g2,…..
3. Server計算 $G=g1+g2+\ldots\ldots$
4. $W = W - \varepsilon * G$

COST FUNCTION IN FEDERATED SYSTEM

- FedAVG

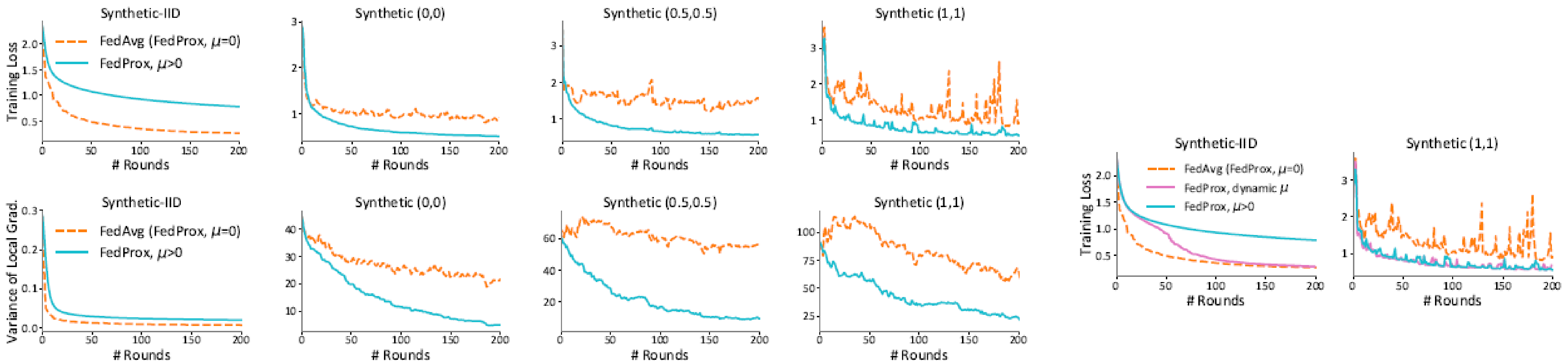


1. Server 下放weight給client去學習
2. Client 學習N個 epochs
3. Client計算 $W = W - \varepsilon * G$後上傳 weight w1,w2….
4. Server計算 $W=(w1+w2+\ldots\ldots.)/m$

優點: 減少communicate的次數
缺點: epoch過多會使global weights偏向Local

COST FUNCTION IN FEDERATED SYSTEM

- FedProx
  - 在cost 計算中加入proximal term $$\min_{w} h_k(w;\ w^t) = F_k(w) + \boxed{\frac{\mu}{2}\|w - w^t\|^2}\,.$$
  - The proximal term addresses the issue of statistical heterogeneity by restricting the local updates to be closer to global model



TOPICS IN STATISTICAL HETEROGENEITY

- Federated Learning with Matched Averaging [9](ICLR 2020, IBM)
  - FedMA constructs the shared global model in a layer-wise manner by matching and averaging hidden elements (i.e. channels for convolution layers; hidden states for LSTM; neurons for fully connected layers) with similar feature extraction signatures
  - 基於排列不變性(Permutation invariance)，單純weight相加取平均可能會出錯，因此本方法會以層數為單位，針對不同client相似的filter進行置換後平均再上傳server，直到所有層數執行結束。

TOPICS IN STATISTICAL HETEROGENEITY

- Federated Learning With Matched Averaging

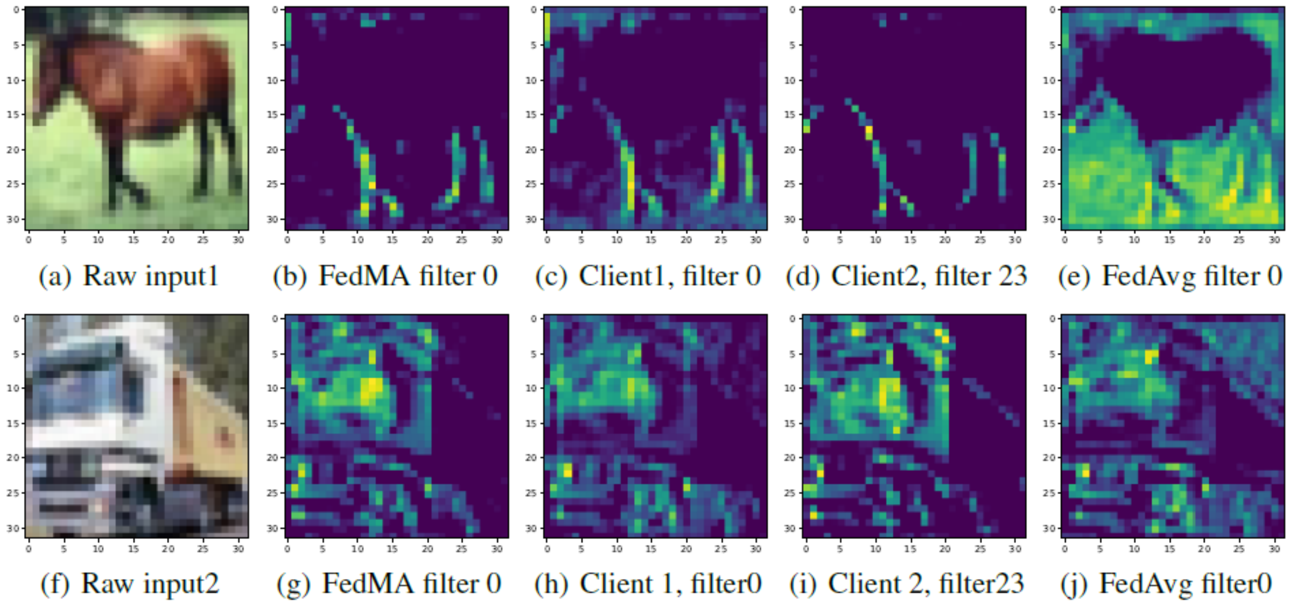
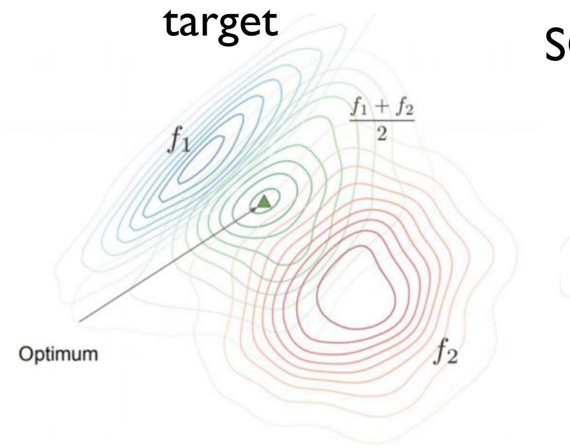
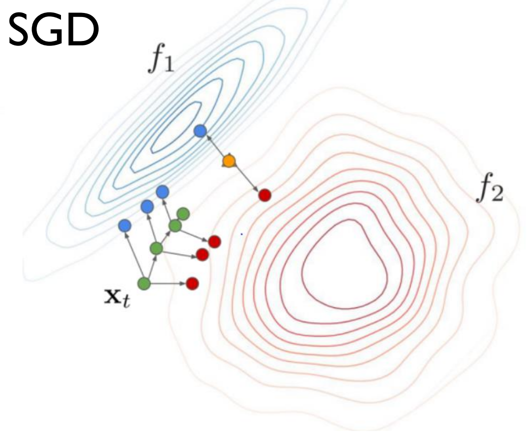


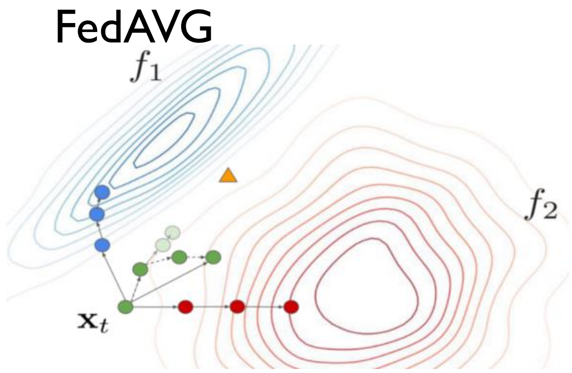
Figure 5: Representations generated by the first convolution layers of locally trained models, FedMA global model and the FedAvg global model.

TOPICS IN STATISTICAL HETEROGENEITY

- SCAFFOLD: Stochastic Controlled Averaging for Federated Learning [10] (PMLR 2020)
  - SCAFFOLD uses control variates (variance reduction) to correct for the 'client-drift' in its local updates. We prove that SCAFFOLD requires significantly fewer communication rounds and is not affected by data heterogeneity
  - 同樣是改進FedAVG，在算client的cost時增加一個控制項 $y_i \leftarrow y_i - \eta_l(g_i(y_i) + c - c_i)\,.$







$$c_i^+ \leftarrow \begin{cases} \text{Option I.} & g_i(x)\,,\text{ or} \\ \text{Option II.} & c_i - c + \frac{1}{K\eta_l}(x - y_i)\,. \end{cases}$$

TOPICS IN STATISTICAL HETEROGENEITY

- SCAFFOLD:

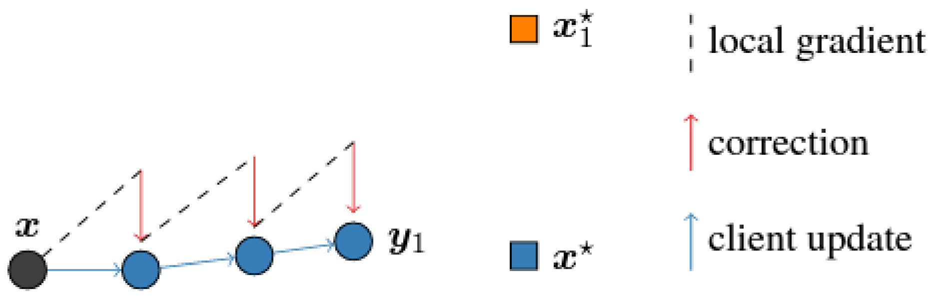
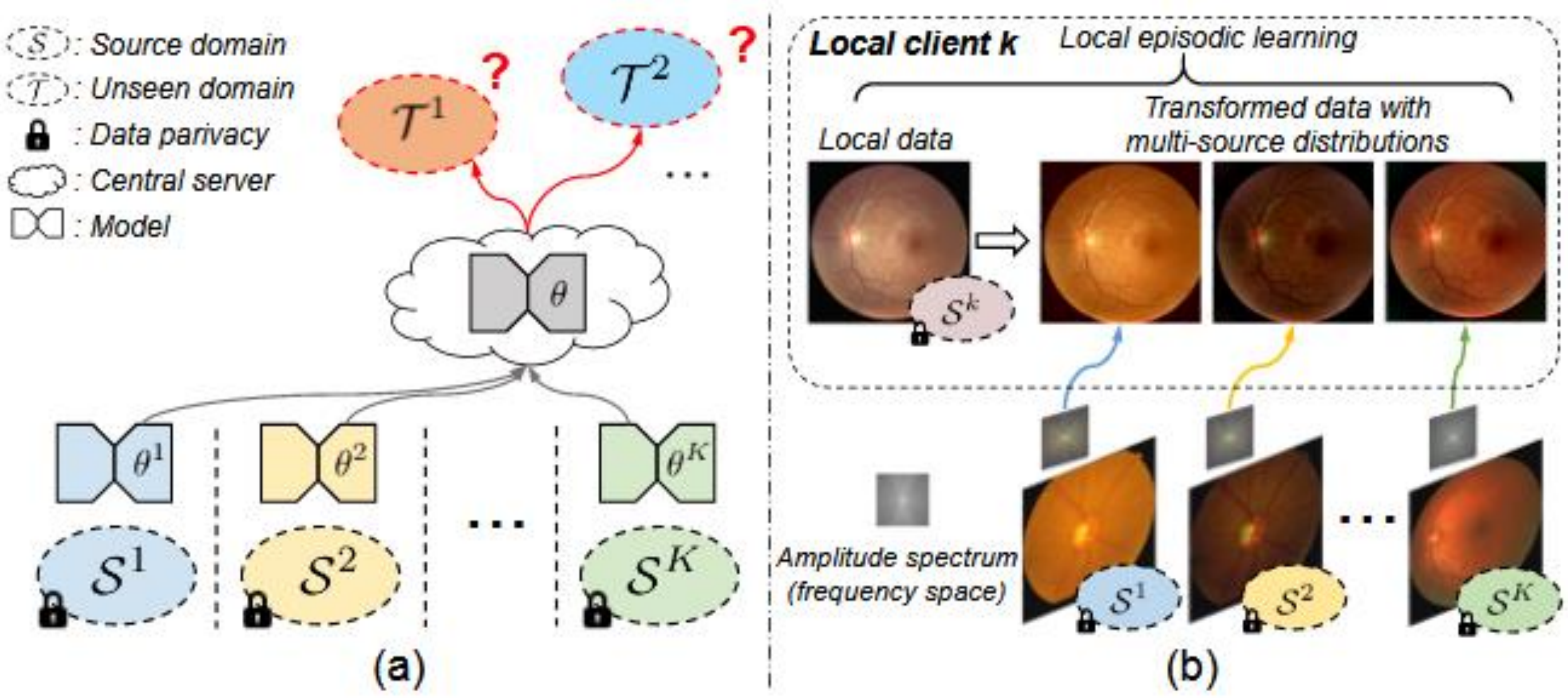


Figure 2. Update steps of SCAFFOLD on a single client. The local gradient (dashed black) points to $x_1^\star$ (orange square), but the correction term $(c - c_i)$ (in red) ensures the update moves towards the true optimum $x^\star$ (black square).

Algorithm 1 SCAFFOLD: Stochastic Controlled Averaging for federated learning

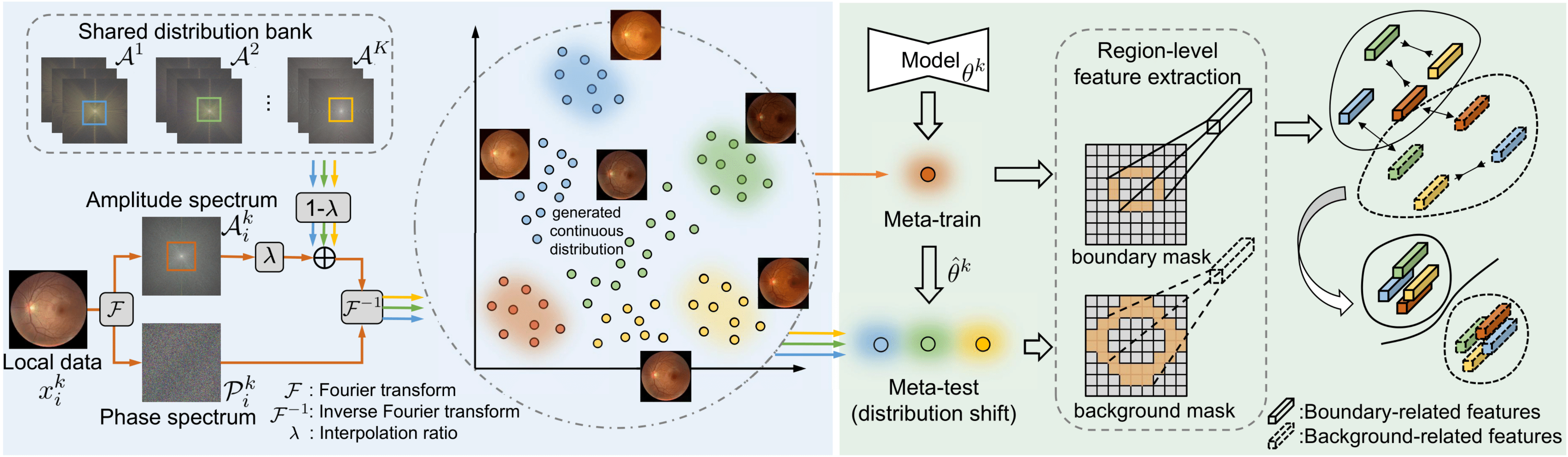
```
server input: initial x and c, and global step-size η_g
client i's input: c_i, and local step-size η_l
for each round r = 1, ..., R do
    sample clients S ⊆ {1, ..., N}
    communicate (x, c) to all clients i ∈ S
    on client i ∈ S in parallel do
        initialize local model y_i ← x
        for k = 1, ..., K do
            compute mini-batch gradient g_i(y_i)
            y_i ← y_i − η_l (g_i(y_i) − c_i + c)
        end for
        c_i^+ ← (i) g_i(x), or (ii) c_i − c + 1/(K η_l) (x − y_i)
        communicate (Δy_i, Δc_i) ← (y_i − x, c_i^+ − c_i)
        c_i ← c_i^+
    end on client
    (Δx, Δc) ← 1/|S| Σ_{i∈S} (Δy_i, Δc_i)
    x ← x + η_g Δx and c ← c + |S|/N Δc
end for
```

FEDDG [8]

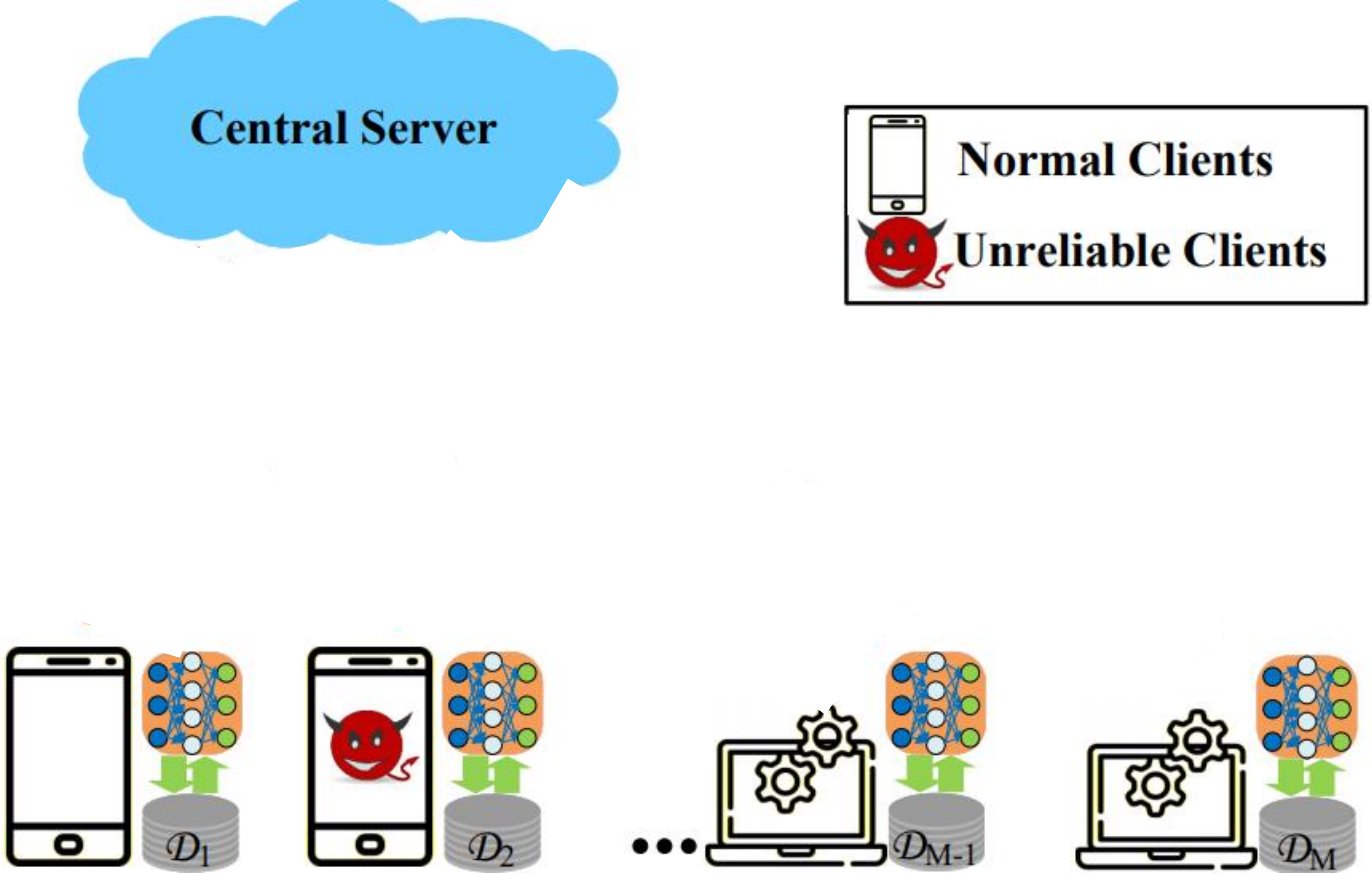


CVPR 21'

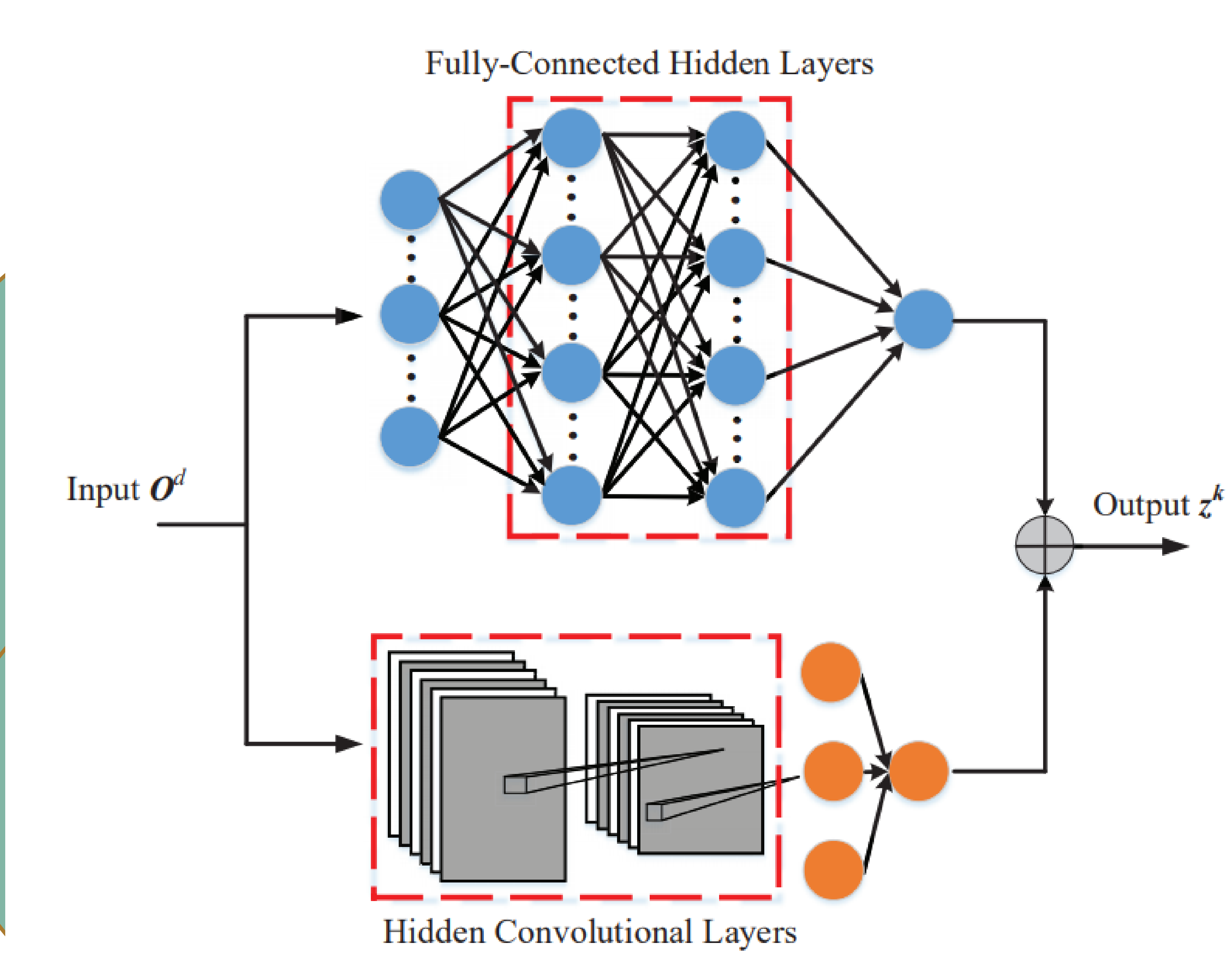
FEDDG [8]



CVPR 21'

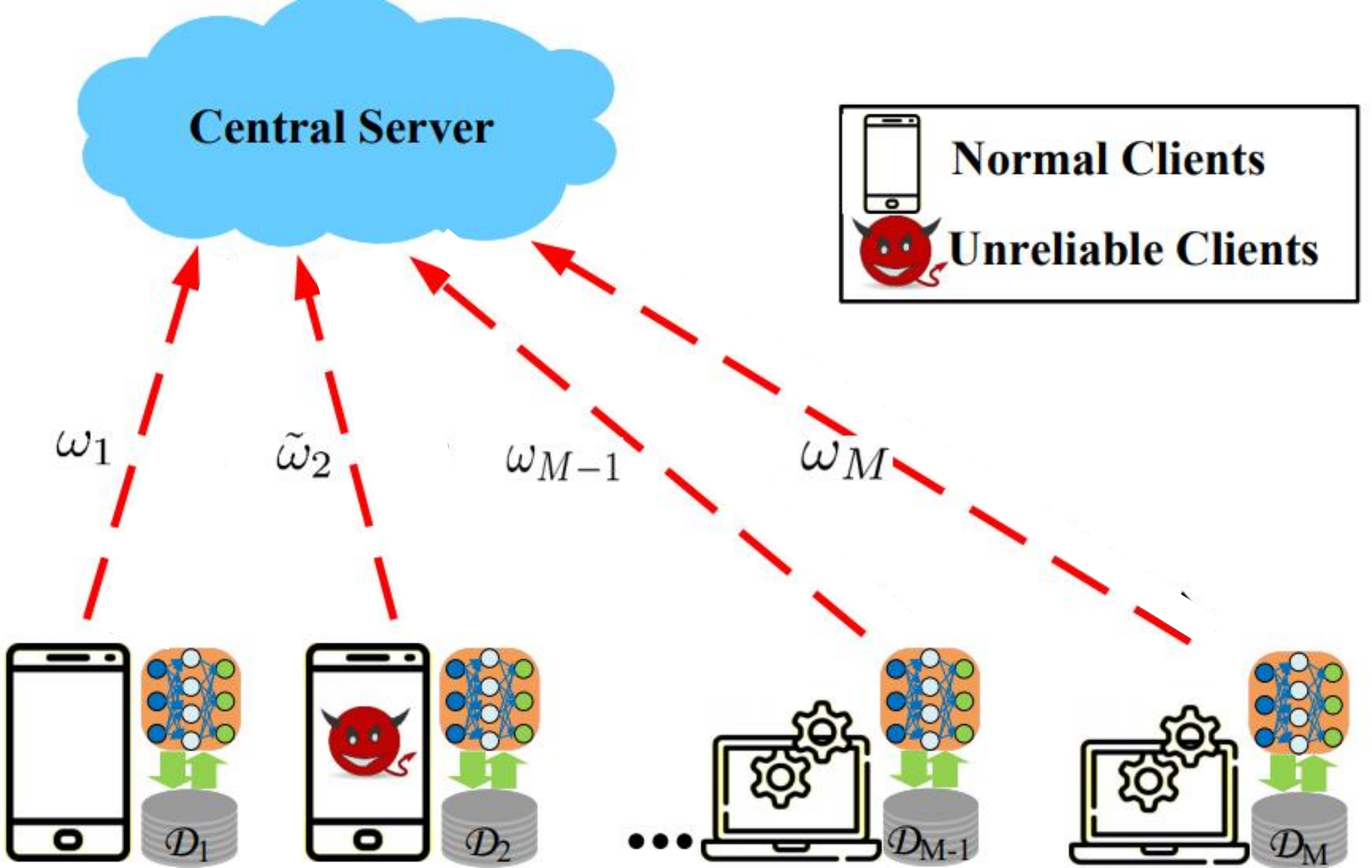


Central Server
Normal Clients
Unreliable Clients
$\mathcal{D}_1$
$\mathcal{D}_2$
$\mathcal{D}_{M-1}$
$\mathcal{D}_M$



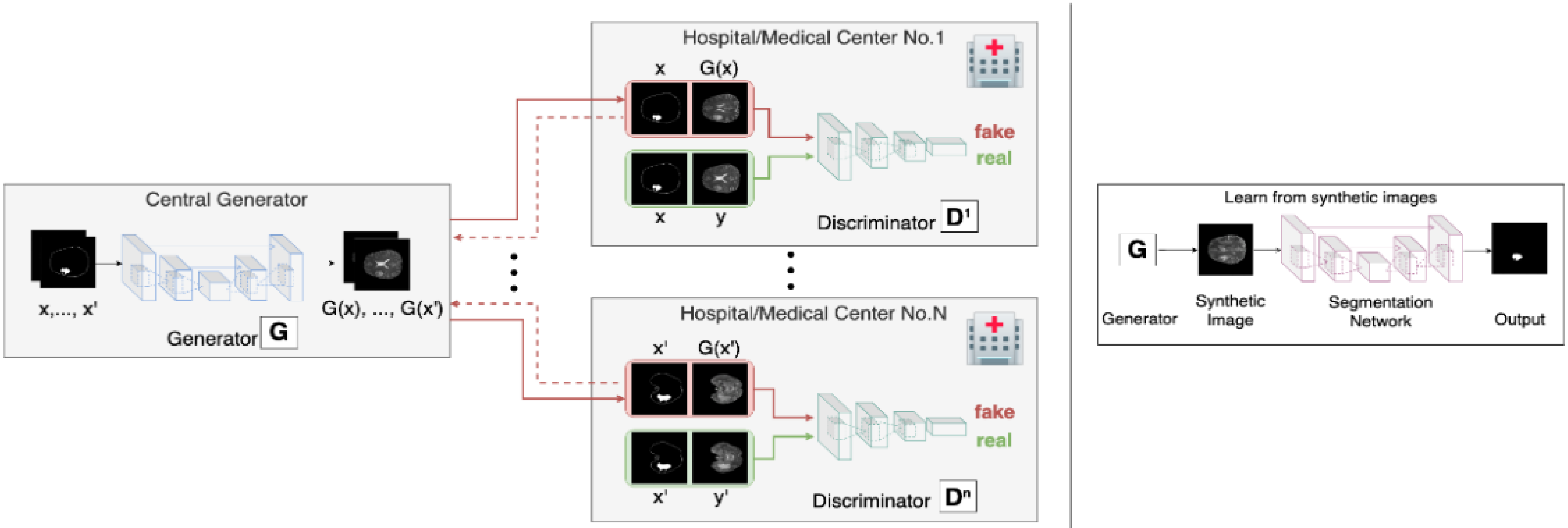
Average of Benign Models

$$\boldsymbol{w}^{k\tau} = \sum_{i=1}^{M} z_i p_i \boldsymbol{w}_i^{k\tau}$$



Central Server
Normal Clients
Unreliable Clients
ω_1
$\tilde{\omega}_2$
ω_{M-1}
ω_M
$\mathcal{D}_1$
$\mathcal{D}_2$
$\mathcal{D}_{M-1}$
$\mathcal{D}_M$

ASYNDGAN [6]



CVPR 20'

Learn without sharing medical image | Segmentation from synthetic image

TOPICS IN STATISTICAL HETEROGENEITY

- Robust Aggregation for Federated Learning (University of Washington) [11]
  - The proposed approach relies on a robust secure aggregation oracle based on the geometric median

- Fast-convergent federated learning with class-weighted aggregation(Journal of Systems Architecture) [12]
  - Since existing scheme may degrade the representative of local models after aggregation, this paper proposed a reallocate weights aggregator based on contributions to each class

- Differentially Private Learning with Adaptive Clipping (google) [13]

UNRELIABLE CLIENTS

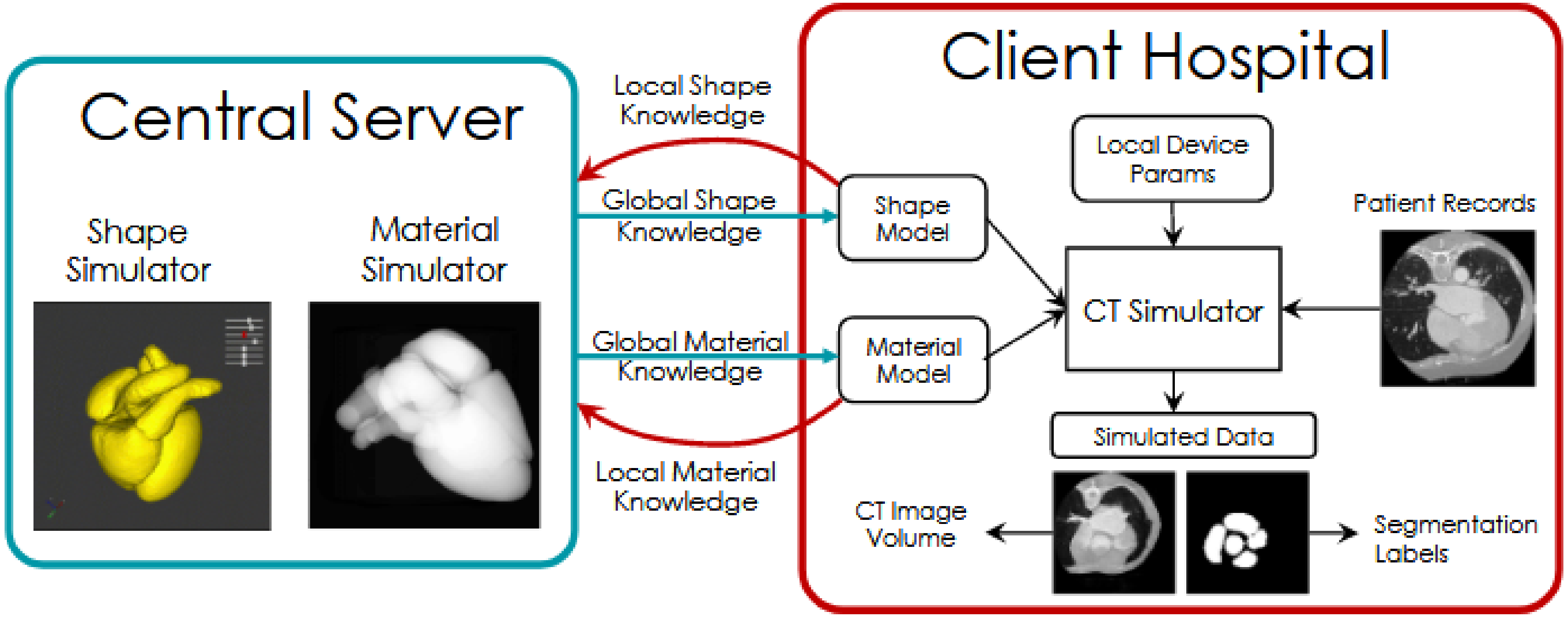
- Server does not have full control of clients' behaviors
- Client may deviate from normal behavior
- Unreliable Clients may:
  - manipulate outputs sent to server to dominate the training process
  - -> Make global model deviate from optimal solution

- DeepSA (Deep Neural Network Based Secure Aggregation)

FEDERATED LEARNING IN MEDICAL IMAGING

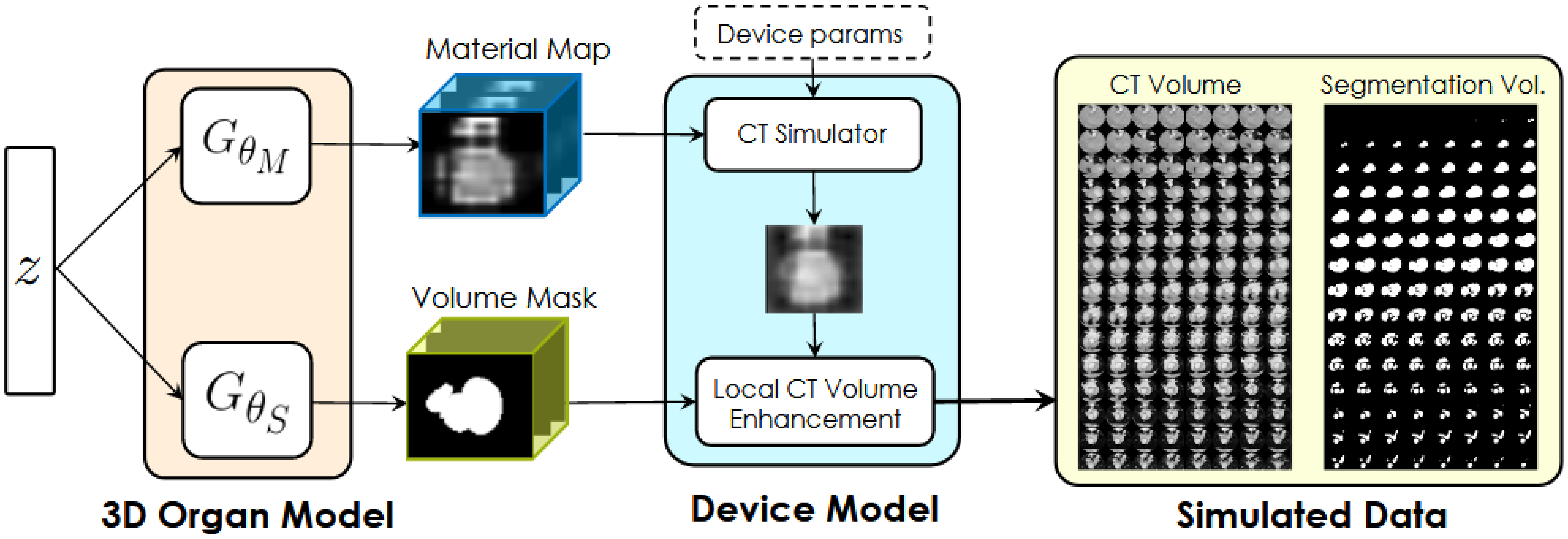
- MICCAI 18' [1]
  - 1st federated learning paper on medical images
- MICCAI 19' [2]
  - Enhance privacy of federated learning
- Nature MI 20' [3]
  - Enhance security and privacy of federated learning
- MICCAI 20' [4]
  - Real-world implementation

FED-SIM [5]



Central Server
Shape Simulator
Material Simulator
Local Shape Knowledge
Global Shape Knowledge
Global Material Knowledge
Local Material Knowledge
Client Hospital
Local Device Params
Shape Model
Material Model
CT Simulator
Patient Records
Simulated Data
CT Image Volume
Segmentation Labels

FED-SIM [5]



Material Map
Device params
z
G_{θ_M}
G_{θ_S}
CT Simulator
CT Volume
Segmentation Vol.
Volume Mask
Local CT Volume Enhancement
3D Organ Model
Device Model
Simulated Data

FEDERATED META-ANALYSIS [7]

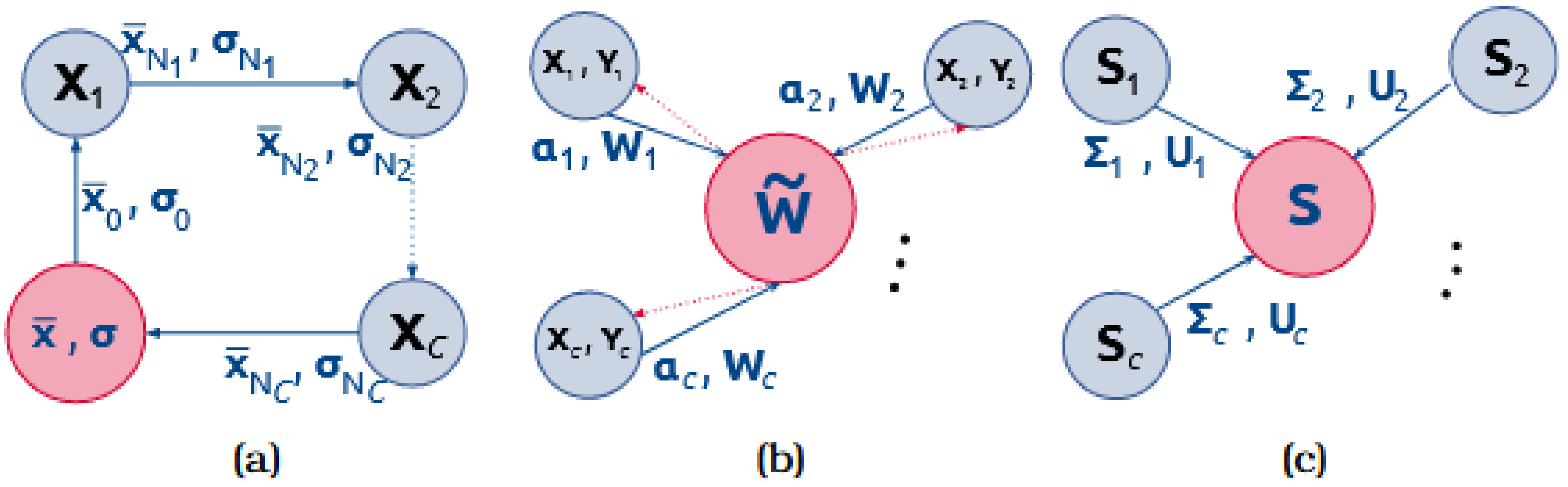


Figure 1: Data flow to obtain: (a) the global statistics $\bar{x}$ and σ, (b) the shared parameter matrix $\widehat{W}$ to correct from covariates and (c) the approximated global covariance matrix S. Red node: master; blue nodes: local centers. Arrows denote the data flows from centers (blue) and from the master (red).

REFERENCE

1. M. J Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Brain Lesion Workshop, MICCAI*, pp. 92–104, 2018.
2. W. Li *et al*., "Privacy-preserving federated brain tumour segmentation," in *Intl. Workshop on Machine Learning in Medical Imaging*, pp. 133–141, 2019.
3. G. A. Kaissis, M. R. Makowski, D. R ̈uckert, and R. F. Braren, "Secure, privacy-preserving and federated machine learning in medical imaging," Nature Machine Intelligence, pp. 1–7, 2020.
4. H. R. Roth et al., "Federated learning for breast density classification: A real-world implementation," in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, pp. 181–191, 2020.
5. D. Li, A. Kar, N. Ravikumar, A. F. Frangi, and S. Fidler, "Federated simulation for medical imaging," in *Intl. Conf Medical Image Computing and Computer-Assisted Intervention*, pp. 159–168, 2020.
6. Q. Chang *et al*., "Synthetic learning: Learn from distributed asynchronized discriminator GAN without sharing medical image data," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition*, pp. 13856–13866, 2020.
7. S. Silva, B. A. Gutman, E. Romero, P. M. Thompson, A. Altmann, and M. Lorenzi, "Federated learning in distributed medical databases: Meta-analysis of large-scale subcortical brain data," in *2019 IEEE 16th Intl. Symposium on Biomedical Imaging*, pp. 270–274, 2019.
8. Q. Liu, C. Chen, J. Qin, Q. Dou, and P.-A. Heng, "FedDG: federated domain generalization on medical image segmentation via episodic learning in continuous frequency space," in *IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2021.
9. Wang, Hongyi, et al. "Federated learning with matched averaging." *arXiv preprint arXiv:2002.06440* (2020).
10. Karimireddy, Sai Praneeth, et al. "SCAFFOLD: Stochastic controlled averaging for federated learning." *International Conference on Machine Learning*. PMLR, 2020.
11. Pillutla, Krishna, Sham M. Kakade, and Zaid Harchaoui. "Robust aggregation for federated learning." *arXiv preprint arXiv:1912.13445* (2019).
12. Ma, Zezhong, et al. "Fast-convergent federated learning with class-weighted aggregation." *Journal of Systems Architecture* 117 (2021): 102125.
13. Andrew, Galen, et al. "Differentially private learning with adaptive clipping." *arXiv preprint arXiv:1905.03871* (2019).