Towards Energy Efficient Federated Meta-learning in Edge Network

Xubo Li∗, Yuanjie Jia†, Yingyu Li‡, Yong Xiao ∗ § ¶

∗School of Elect. Inform. & Commun., Huazhong Univ. of Science & Technology, Wuhan, China

†Wuhan Research Institute of Posts and Telecommunications, Wuhan, China

‡School of Mech. Eng. and Elect. Inform., China University of Geosciences, Wuhan, China

§Pengcheng National Laboratory, Shenzhen, China

¶Pazhou Lab, Guangzhou, China

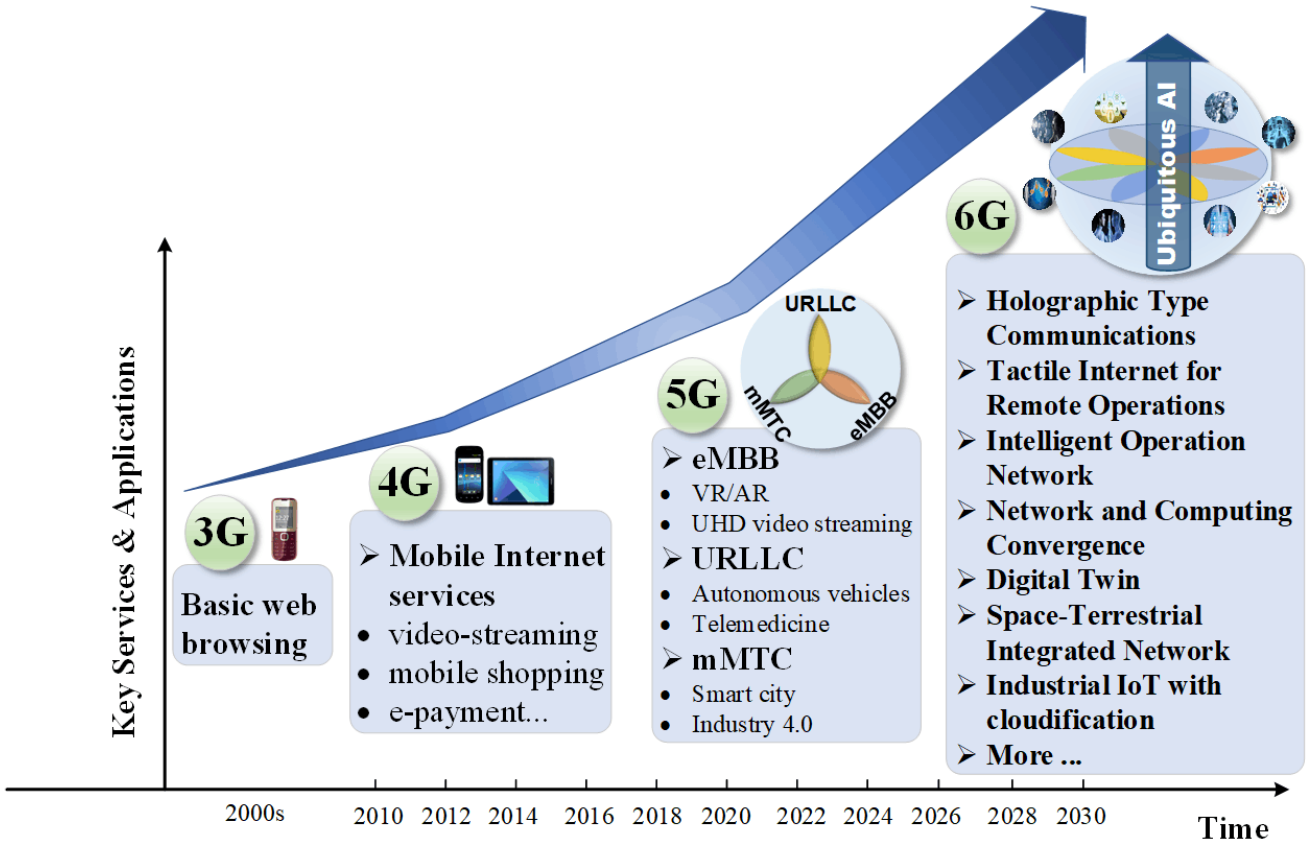
IEEE VTC'24Spring

November 2, 2025

Outline

- ☐ Background and Motivation
- ☐ System Model and Problem Formulation
- ☐ Analytical Solutions
- ☐ Experiments and Simulation Results
- ☐ Conclusions

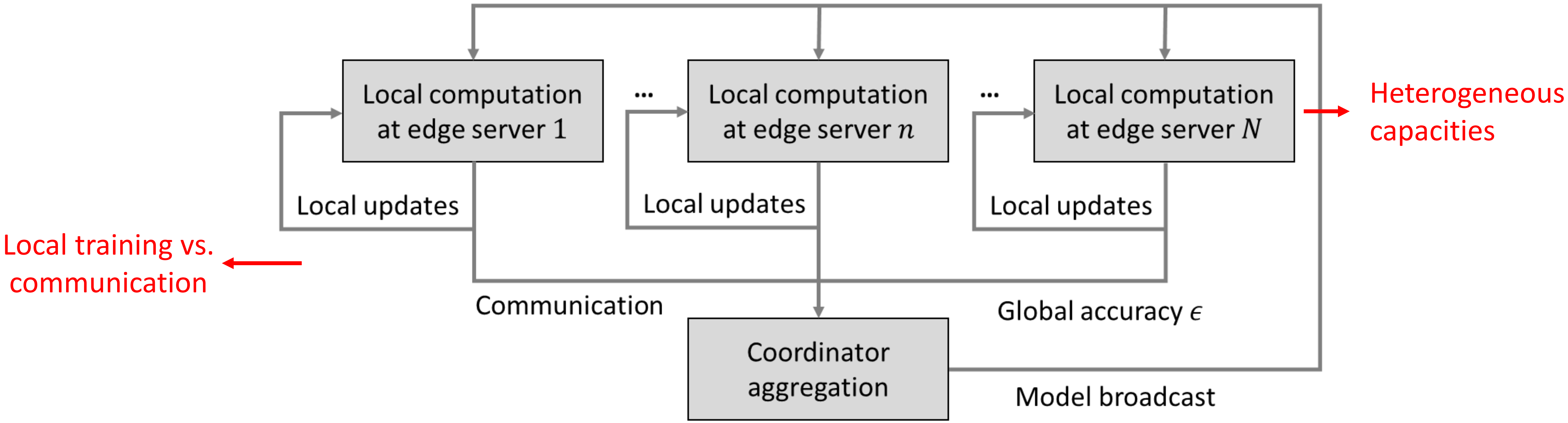
Background



- ➢ More and more network access
- ➢ Progressively increasing computing power at the edge of the network
- ➢ Growing need for privacy

- ☐ 6G takes AI as the core driving force
- ☐ Federated Edge Intelligence(FEI) is a candidate technology for network AI in 6G

Motivation



Challenges for FEI System

- Heterogeneous storage, communication, and computation capacities
- Multiple rounds of model training and model coordination
- Lack of a simple and comprehensive framework

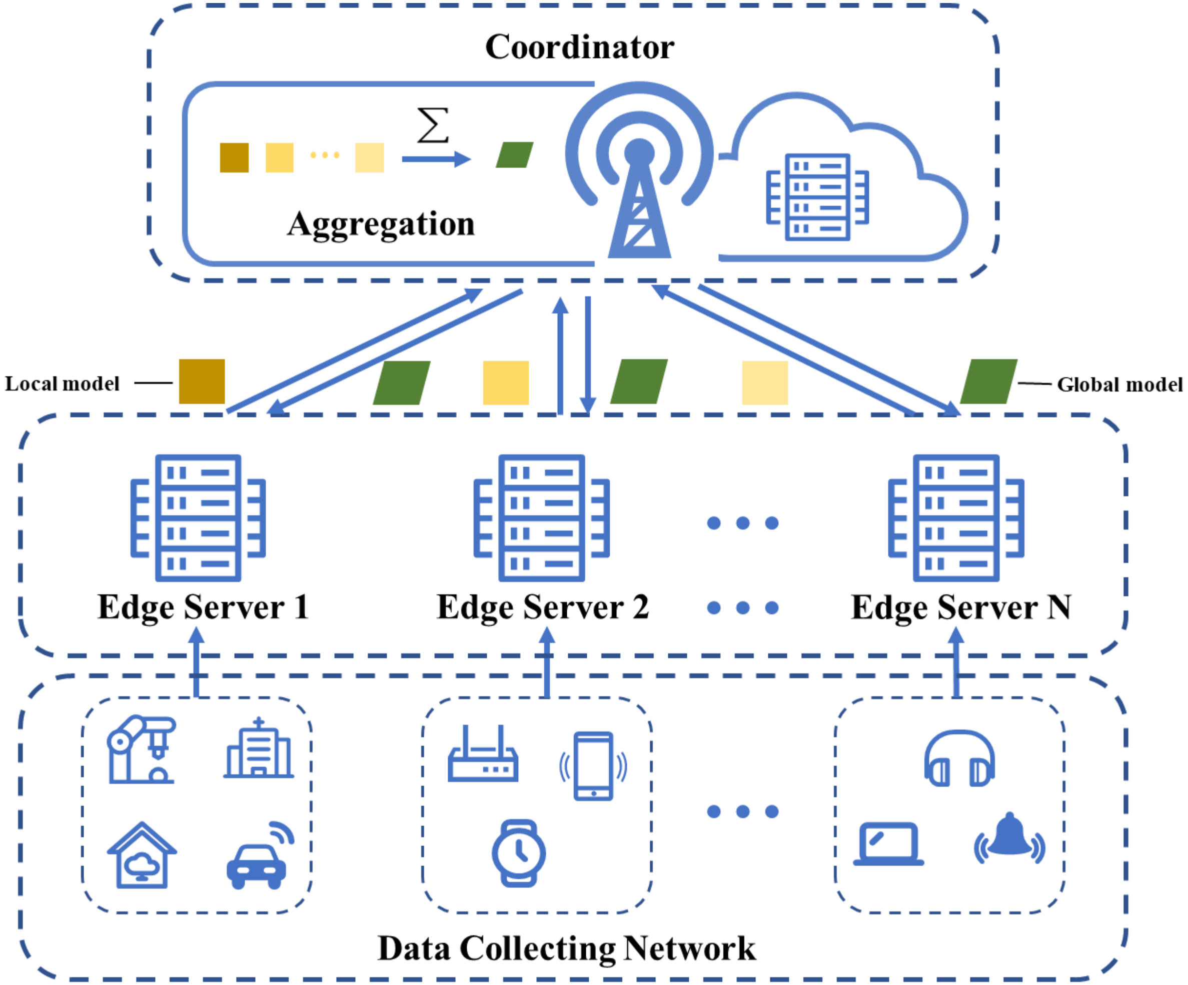
Key Contributions

- ✓ Propose an energy consumption model
- ✓ Derive the analytical solutions for the energy consumption of FEI networks with two popular distributed algorithms, FedAvg and FedMeta
- ✓ Develop a hardware prototype and conduct extensive measurements on real-world dataset

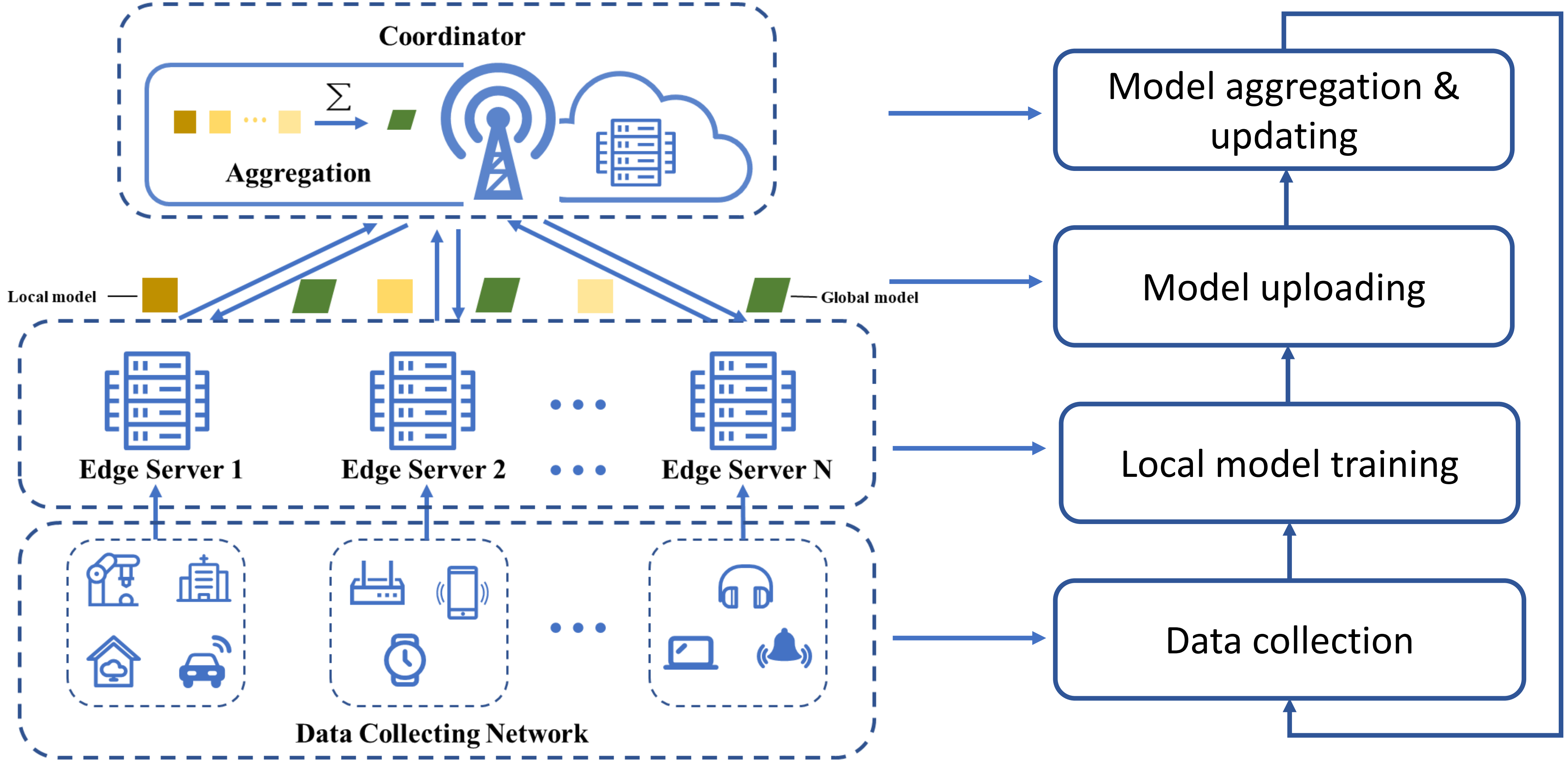
Outline

- ☐ Background and Motivation
- ☐ System Model and Problem Formulation
- ☐ Analytical Solutions
- ☐ Experiments and Simulation Results
- ☐ Conclusions

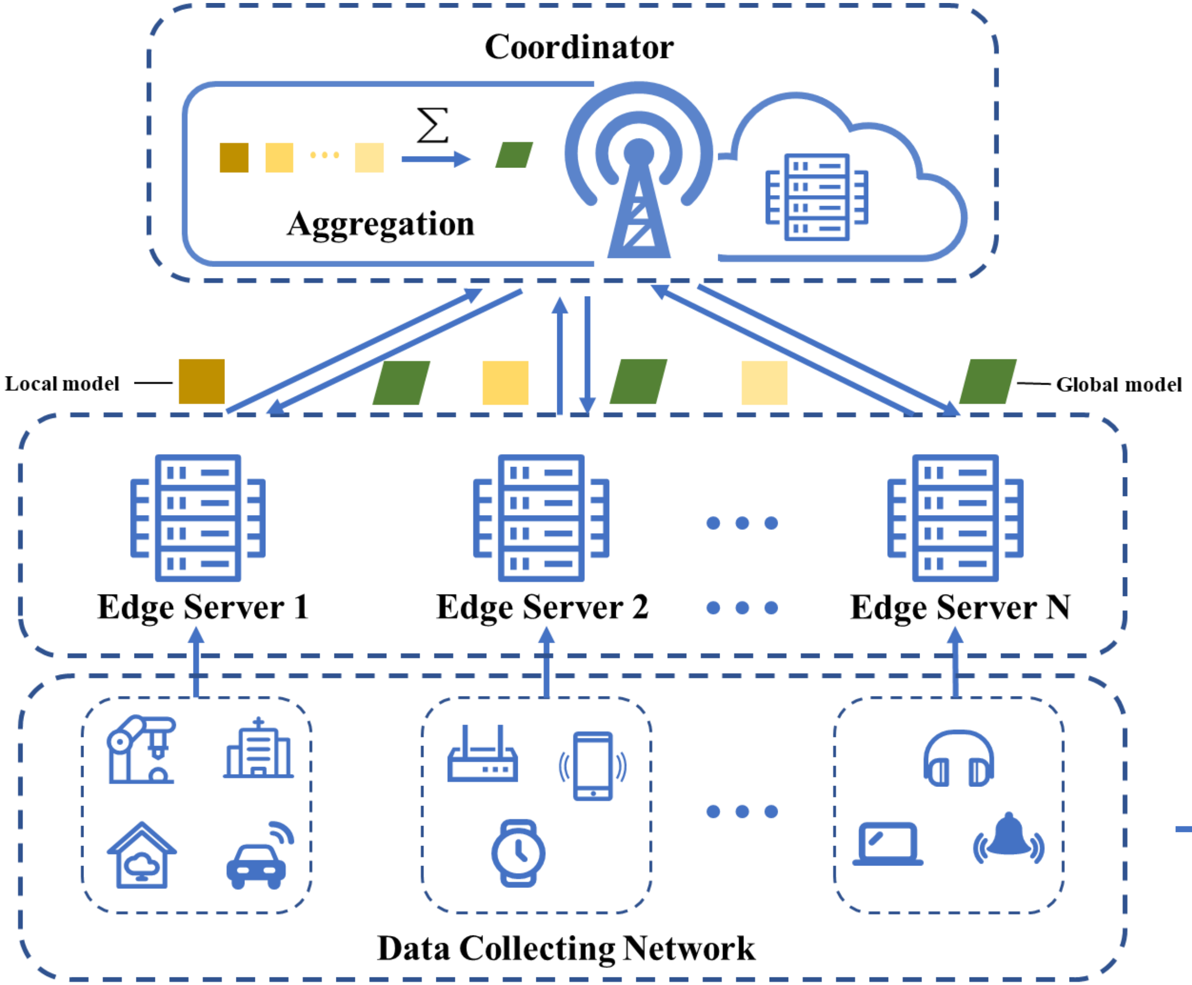
System Model



System Model



Problem Formulation

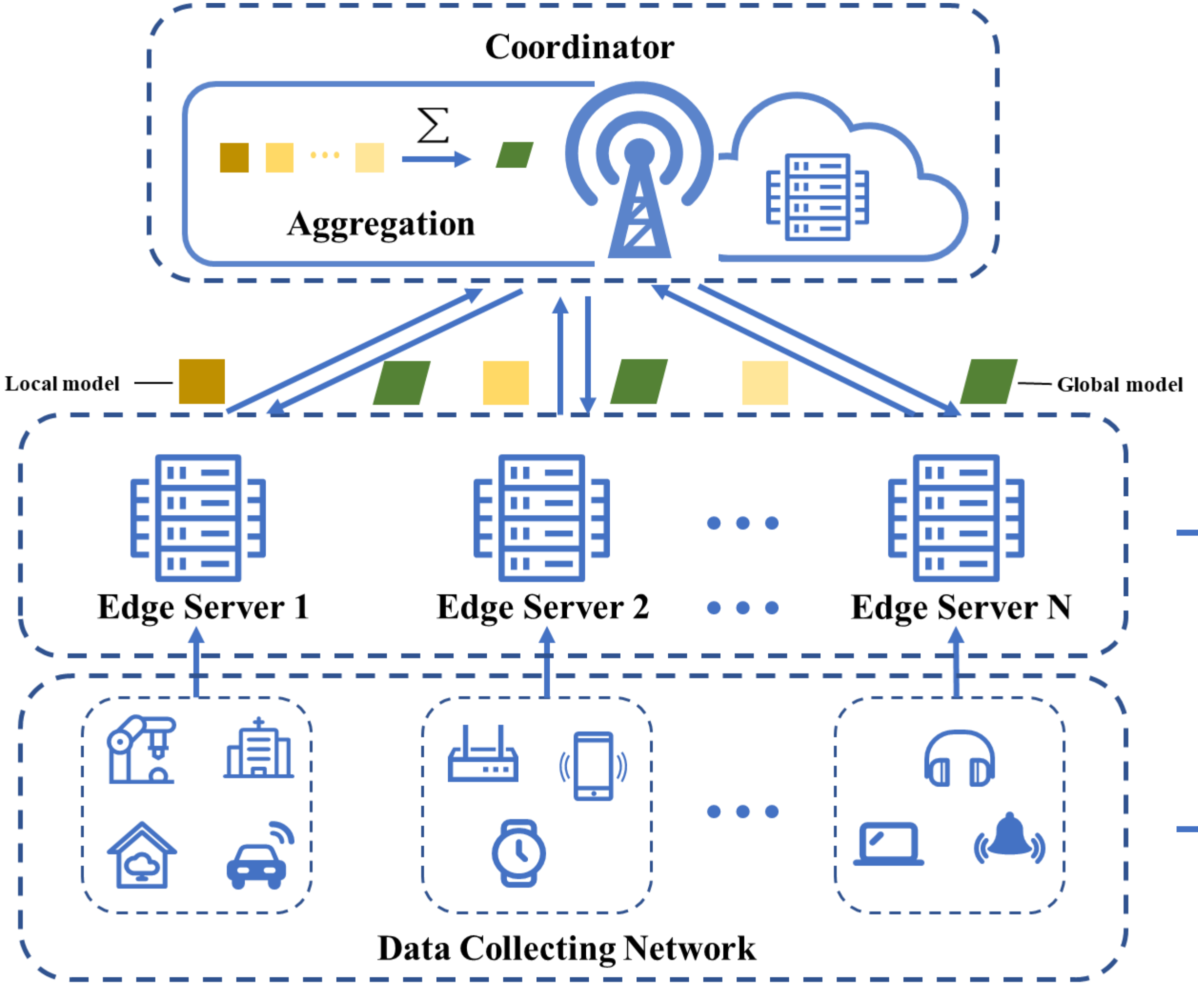


Energy consumption of data collection:

$$E_n^{dc}(K) = \eta K$$

K: the required number of data samples

Problem Formulation



Energy consumption of local model training:

$$E_n^{mt}(e, K) = c_0 eK + c_1 K$$

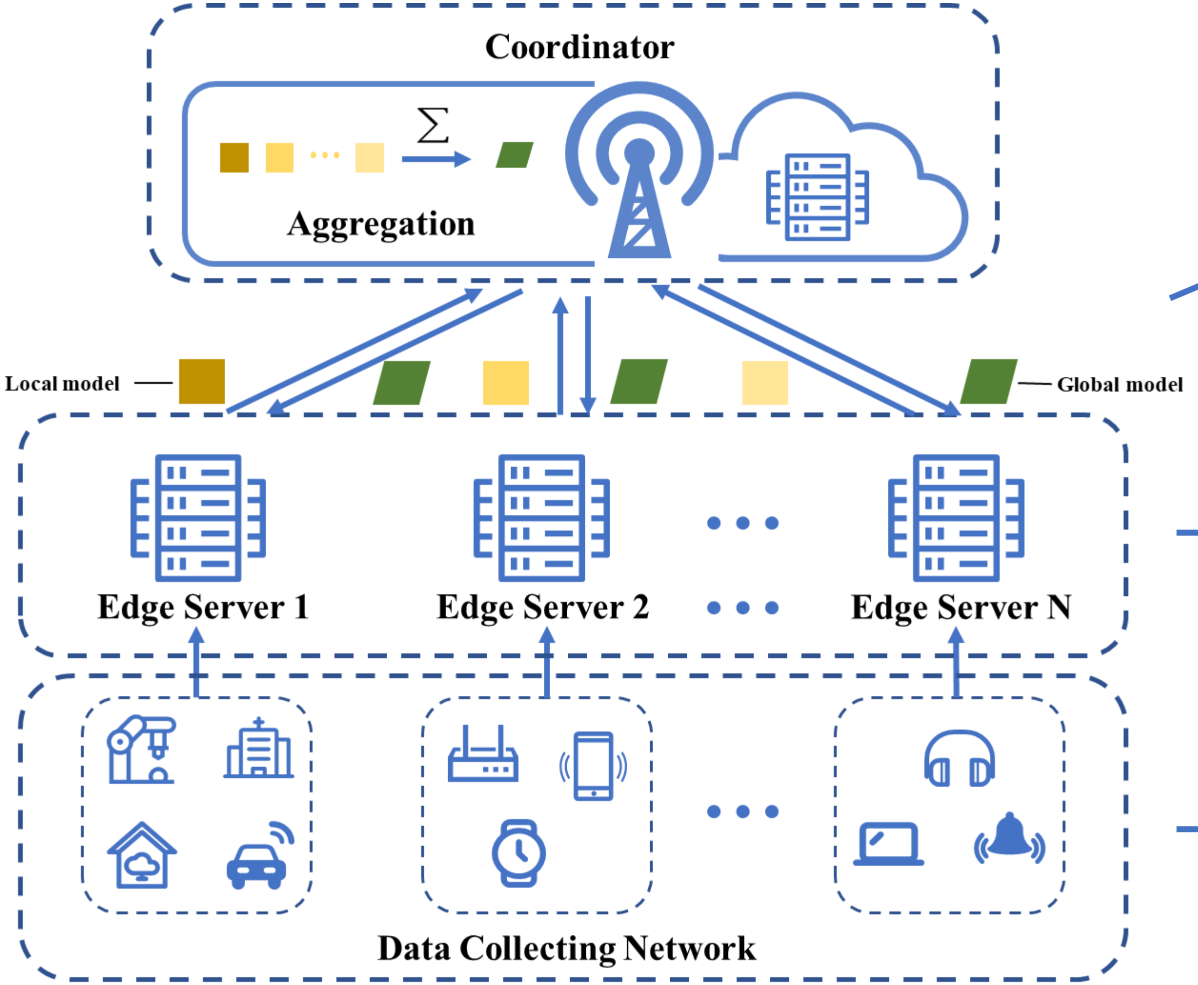
Energy consumption of data collection:

$$E_n^{dc}(K) = \eta K$$

K: the required number of data samples

e: the steps of local training

Problem Formulation



Energy consumption of model uploading:

$$E_n^{mu}$$

Energy consumption of local model training:

$$E_n^{mt}(e, K) = c_0 eK + c_1 K$$

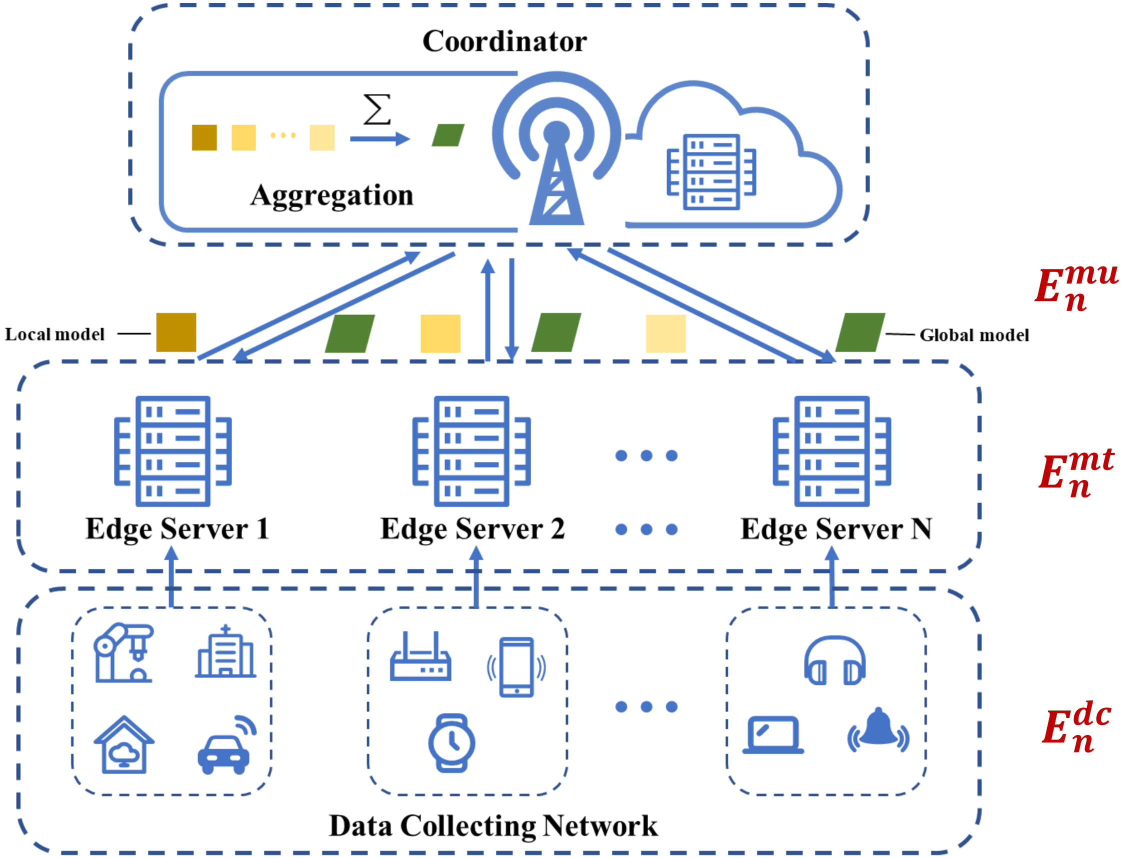
Energy consumption of data collection:

$$E_n^{dc}(K) = \eta K$$

K: the required number of data samples

e: the steps of local training

Problem Formulation



Overall energy consumption of FEI system

$$E(K,T) = \sum_{t=1}^{T} \sum_{n \in \mathcal{N}} \left(E_n^{dc} + E_n^{mt} + E_n^{mu}\right)$$

Energy minimization problem of FEI system

$$\min_{K,T} [E(K,T)]$$

$$s.t.\ \frac{1}{eT} \sum_{t=0}^{T-1} \sum_{\tau=0}^{e-1} \mathbb{E}[\|\nabla \mathcal{R}(\bar{w}_{t+1}^{\tau})\|^2] \le \epsilon$$

where $\bar{w}_{t+1}^{\tau} = \frac{1}{N} \sum_{n \in \mathcal{N}} w_{n,t+1}^{\tau}$.

Outline

- ☐ Background and Motivation
- ☐ System Model and Problem Formulation
- ☐ Analytical Solutions
- ☐ Experiments and Simulation Results
- ☐ Conclusions

Training Procedure

- Objective function of FedAvg

$$\mathcal{R}_n(w_t) = f_n(w_t, \mathcal{S}_n)$$

- Objective function of FedMeta

$$\mathcal{R}_n(w_t) = f_n\left(w_t - \alpha \nabla f_n\left(w_t, \mathcal{S}_n^{in}\right), \mathcal{S}_n^{out}\right)$$

Step size

$$f_n(w, \mathcal{S}_n) = \frac{1}{|\mathcal{S}_n|} \sum_{x \in \mathcal{S}_n} \ell_n(w; x)$$

[1] M. Brendan, M. Edier, H. Seth, et al, "Communication-efficient learning of deep networks from decentralized data," in AISTATS, Florida, Apr., 2017.

[2] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in NeurIPS, Virtual, Dec. 2020.

Training Procedure

- Objective function of FedAvg

$$\mathcal{R}_n(w_t) = f_n(w_t, \mathcal{S}_n)$$

- Objective function of FedMeta

$$\mathcal{R}_n(w_t) = f_n\left(w_t - \alpha \nabla f_n(w_t, \mathcal{S}_n^{in}), \mathcal{S}_n^{out}\right)$$

Step size

$$f_n(w, \mathcal{S}_n) = \frac{1}{|\mathcal{S}_n|} \sum_{x \in \mathcal{S}_n} \ell_n(w; x)$$

Local update

$$w_{n,t+1}^{\tau} = w_{n,t+1}^{\tau-1} - \beta \nabla \mathcal{R}_n\left(w_{n,t+1}^{\tau-1}\right)$$

Local learning rate

[1] M. Brendan, M. Edier, H. Seth, et al, "Communication-efficient learning of deep networks from decentralized data," in AISTATS, Florida, Apr., 2017.
[2] A. Fallah, A. Mokhtari, and A. Ozdaglar, "Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach," in NeurIPS, Virtual, Dec. 2020.

Energy Consumption

FedAvg

Convergence constraint

$$\frac{1}{eT}\sum_{t=0}^{T-1}\sum_{\tau=0}^{e-1}\mathbb{E}[\|\nabla\mathcal{R}(\overline{w}_{t+1}^{\tau})\|^2] \leq \frac{A_1}{T}+\frac{A_2}{K}+A_3$$

Under a given accuracy

$$\frac{A_1}{T}+\frac{A_2}{K}+A_3 \leq \epsilon$$

The optimal T^*

$$T_{Avg}^{*}=\frac{KA_1}{K\epsilon-A_2-KA_3}$$

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)]=\frac{KA_1}{K\epsilon-A_2-KA_3}N(B_0K+B_1)$$

FedMeta

Convergence constraint

$$\frac{1}{eT}\sum_{t=0}^{T-1}\sum_{\tau=0}^{e-1}\mathbb{E}[\|\nabla\mathcal{R}(\overline{w}_{t+1}^{\tau})\|^2] \leq \frac{C_1}{T}+\frac{C_2}{K}+C_3$$

Under a given accuracy

$$\frac{C_1}{T}+\frac{C_2}{K}+C_3 \leq \epsilon$$

The optimal T^*

$$T_{Meta}^{*}=\frac{KC_1}{K\epsilon-C_2-KC_3}$$

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)]=\frac{KC_1}{K\epsilon-C_2-KC_3}N(2B_0K+B_1)$$

[3] H. Yu, S. Yang, and S. Zhu, "Parallel restarted SGD with faster convergence and less communication: Demystifying why model averaging works for deep learning," in AAAI, Hawaii, USA, Jan. 2019.

Energy Consumption

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \frac{KA_1}{K\epsilon - A_2 - KA_3} N\boxed{(B_0K + B_1)}$$

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \frac{KC_1}{K\epsilon - C_2 - KC_3} N\boxed{(2B_0K + B_1)}$$

$$B_0 = \mathbb{E}[c_0]e$$

$$B_1 = \mathbb{E}[\eta]K + \mathbb{E}[c_1]e + \mathbb{E}[E_n^{mu}]$$

- ***Note1:***

  Objective function of FedAvg:

  $$\mathcal{R}_n(w_t) = f_n(w_t, \mathcal{S}_n)$$

  Objective function of FedMeta:

  $$\mathcal{R}_n(w_t) = f_n\big(w_t - \alpha\nabla f_n(w_t, \mathcal{S}_n^{in}), \mathcal{S}_n^{out}\big)$$

To perform the same number of local updates

The number of data samples:

FedMeta > FedAvg

More local computation

- ***Remark1:*** Within a single communication round,

  Energy consumption: **FedMeta > FedAvg**

Energy Consumption

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \boxed{\frac{KA_1}{K\epsilon - A_2 - KA_3}} N \boxed{(B_0K + B_1)}$$

T^*_{Avg}

$B_0 = \mathbb{E}[c_0]e$

$B_1 = \mathbb{E}[\eta]K + \mathbb{E}[c_1]e + \mathbb{E}[E_n^{mu}]$

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \boxed{\frac{KC_1}{K\epsilon - C_2 - KC_3}} N \boxed{(2B_0K + B_1)}$$

T^*_{Meta}

- ***Note2:***

By setting the number of local updates as $e = \mathcal{O}(\epsilon^{-1/2})$, the number of samples as $K = \mathcal{O}(\epsilon^{-1})$, the stepsize as $\beta = \mathcal{O}(\epsilon)$.

To achieve an $\mathcal{O}(\epsilon)$-first-order stationary point of $\mathcal{R}$

FedAvg: $T = \mathcal{O}(\epsilon^{-4})$ rounds

FedMeta: $T = \mathcal{O}(\epsilon^{-2})$ rounds

- ***Remark1:*** Within a single communication round,

  Energy consumption: **FedMeta > FedAvg**

- ***Remark2:*** To achieve a satisfactory model,

  Communication rounds: **FedMeta < FedAvg**

Energy Consumption

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \underbrace{\boxed{\frac{KA_1}{K\epsilon - A_2 - KA_3}}}_{T^*_{Avg}} N\boxed{(B_0K + B_1)}$$

$$B_0 = \mathbb{E}[c_0]e$$

$$B_1 = \mathbb{E}[\eta]K + \mathbb{E}[c_1]e + \mathbb{E}[E_n^{mu}]$$

The optimal overall energy consumption

$$\mathbb{E}[\hat{E}(K)] = \underbrace{\boxed{\frac{KC_1}{K\epsilon - C_2 - KC_3}}}_{T^*_{Meta}} N\boxed{(2B_0K + B_1)}$$

- ***Note2:***

  By setting the number of local updates as $e = \mathcal{O}(\epsilon^{-1/2})$,

  the number of samples as $K = \mathcal{O}(\epsilon^{-1})$,

  the stepsize as $\beta = \mathcal{O}(\epsilon)$.

  To achieve an $\mathcal{O}(\epsilon)$-first-order stationary point of $\mathcal{R}$ $\longrightarrow$

  FedAvg: $T = \mathcal{O}(\epsilon^{-4})$ rounds

  FedMeta: $T = \mathcal{O}(\epsilon^{-2})$ rounds

- ***Remark1:*** Within a single communication round,

  Energy consumption: **FedMeta > FedAvg**

- ***Remark2:*** To achieve a satisfactory model,

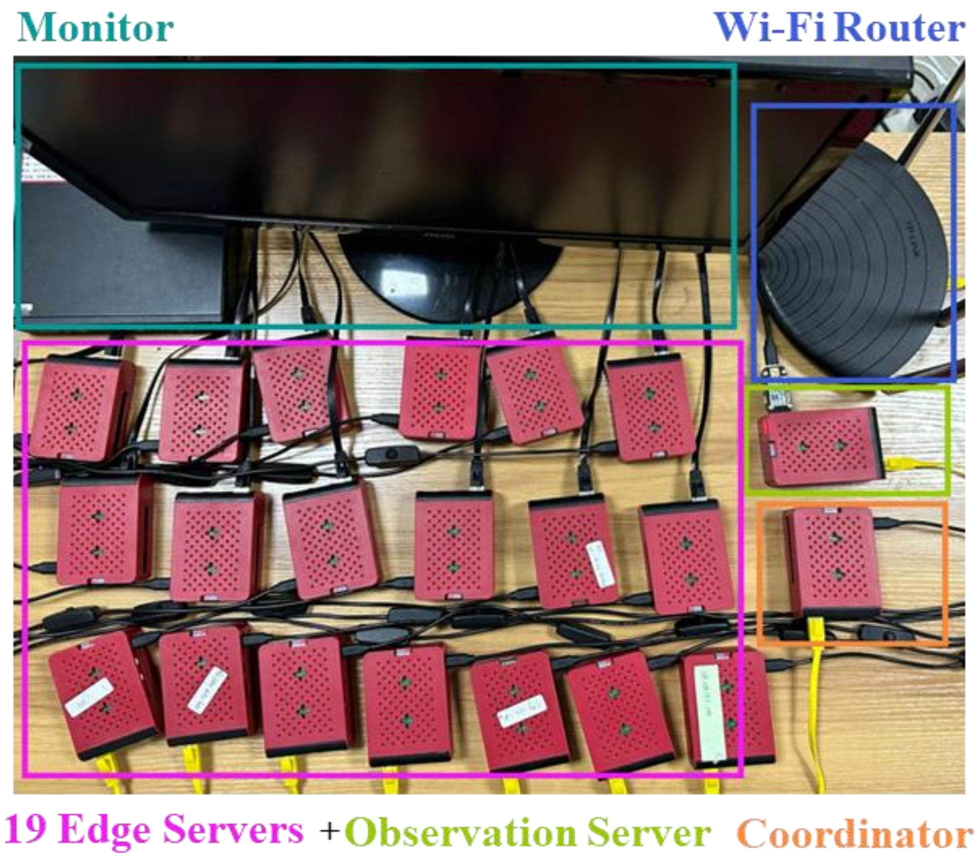
  Communication rounds: **FedMeta < FedAvg**

$\xrightarrow{?}$ **the overall energy consumption**

Outline

- ☐ Background and Motivation
- ☐ System Model and Problem Formulation
- ☐ Analytical Solutions
- ☐ Experiments and Simulation Results
- ☐ Conclusions

Experiment Setup



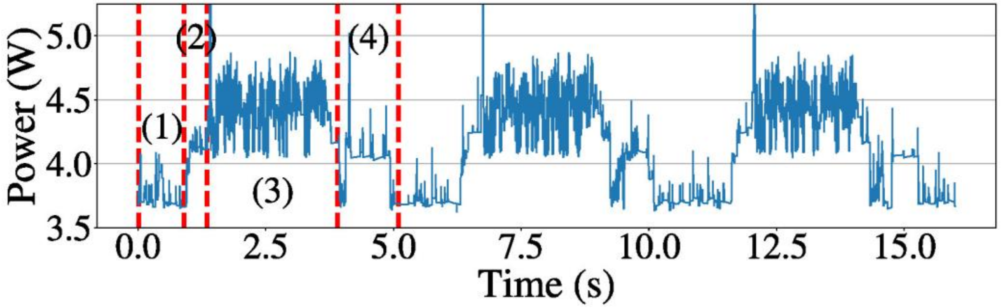
Hardware prototype

Algorithm	FedAvg	FedMeta
Dataset	MNIST	
Number of devices	20	
Number of labels per device	2	
Number of samples per device	3000	
Learning rate	0.005	Inner layer: 0.004 Outer layer: 0.005
Model architecture	2-layer CNN with ReLu	
Communication rounds	100	

- Fix a test performance,
- Compare the energy consumption to achieve it.

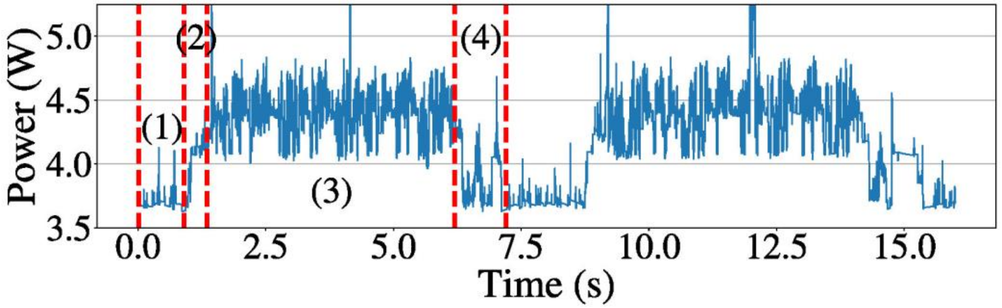
Numerical Results

- **FedAvg**



Stage	Average power	Energy consumption
Waiting	3.7W	3.33J
Model downloading	4.1W	1.845J
Local training	4.4W	11.22J
Model uploading	3.9W	3.9J

- **FedMeta**

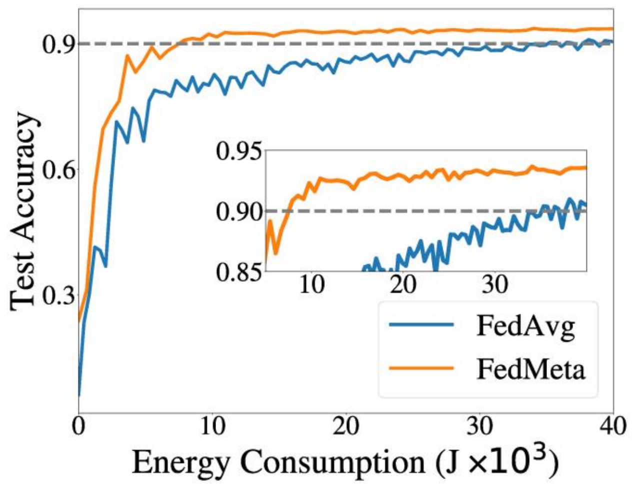


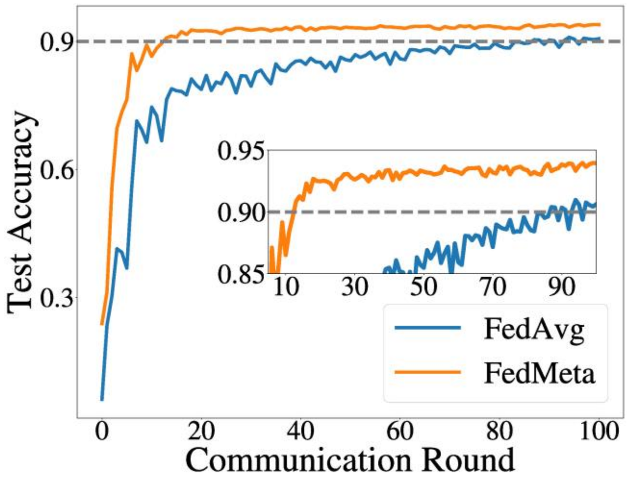
Stage	Average power	Energy consumption
Waiting	3.7W	3.33J
Model downloading	4.1W	1.845J
Local training	4.4W	21.34J
Model uploading	3.9W	3.9J

✓ Within a single communication round, FedMeta requires more energy consumption

Numerical Results

Algorithm	FedAvg	FedMeta
Energy consumption	36,125J (4.5x)	7,970J
Communication Rounds	89 (6.8x)	13





- ✓ Although within a single round, FedMeta requires more energy consumption
- ✓ For a fixed test accuracy, **FedMeta** requires **fewer communication rounds** and achieves **less energy consumption** than FedAvg.

Outline

- ☐Background and Motivation
- ☐System Model and Problem Formulation
- ☐Analytical Solutions
- ☐Experiments and Simulation Results
- ☐Conclusions

Conclusions

- ✓ Proposed an energy consumption model
- ✓ Derived the analytical solutions for the energy consumption of FEI networks with two popular distributed algorithms, FedAvg and FedMeta
- ✓ Proved that energy consumption is related to many factors, and large energy consumption within a single communication round does not mean large overall energy consumption

Thank You!

For more information, please contact:

Xubo Li
(xuboli@hust.edu.cn)
Yong Xiao
(yongxiao@hust.edu.cn)