索尼中国二十周年
心动时刻,索尼与你!

SONY

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Ce ZHENG
ce.zheng@sony.com
Research Scientist & 3GPP SA2 Delegate,
Wireless Network Research Department,
SONY R&D Center (China)

Tianming ZANG, **Ce ZHENG***, Shiyao MA, Chen SUN, Wei CHEN, "*A General Solution for Straggler Effect and Unreliable Communication in Federated Learning*".

CONFIDENTIAL

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

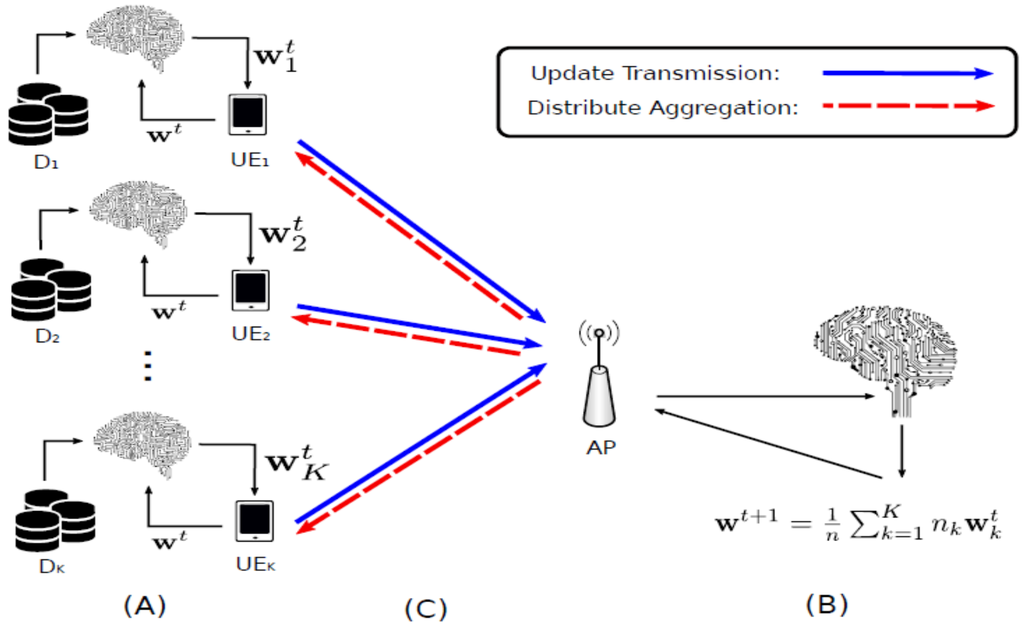


Fig.1

FedAvg algorithm:

1) UE selection and broadcasting (Server):

   initialize model w^0, candidate set S_r

2) Local model updating and training (Each Device#k):

$$w_k^{t+1} \leftarrow w_k^t - \eta \nabla F_k(w_k^t)$$

3) Aggregation:

$$w^{t+1} = \frac{1}{\sum_{k\in S_r} n_k} \sum_{k\in S_r} w_k^t$$

Repeat until convergence…

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

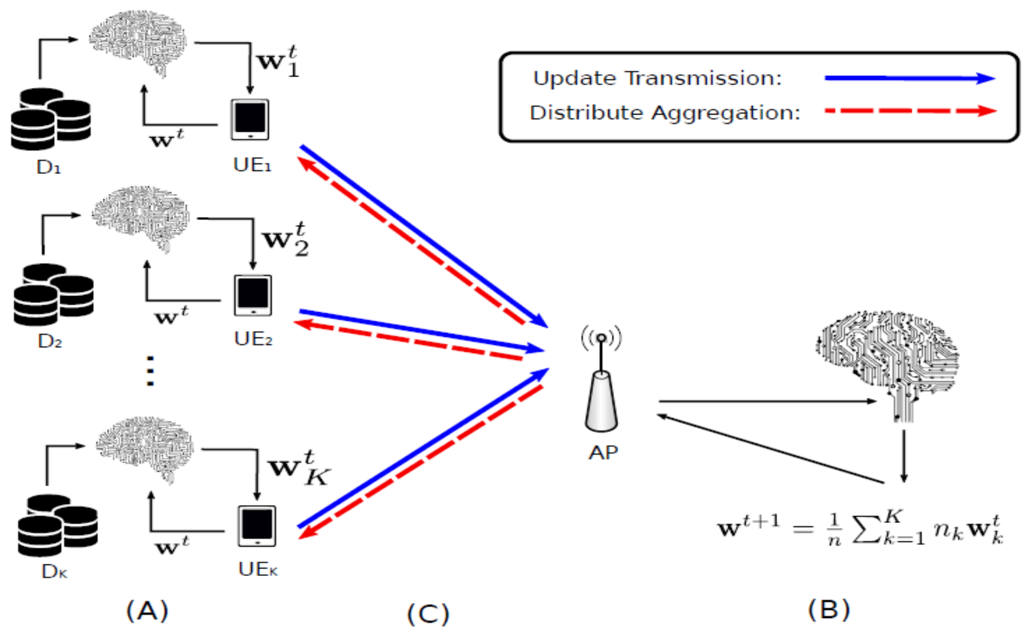


Fig.1

Scenario

Two problems

1. **The straggler effect** --- the performance of FL is decided on the "slowest" UE.
   →
   Prolonged training time

2. **Unreliable communication** --- models from UEs with transmission failure or errors will be abandoned.
   →
   aggregated model biased towards UEs with good channel quality.

$$w^{t+1} = \frac{1}{\sum_{k\in S_r} n_k} \sum_{k\in S_r} w_k^t$$

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

☐ Problem#1: Straggler Effect

The training time of FL is limited by the slowest UE.

Solution: Clustering Based on Training Time

--- put UEs of "*same training time*" into the same round, "straggler" has no UEs to straggle

Fig.2-a

Fig.2-b

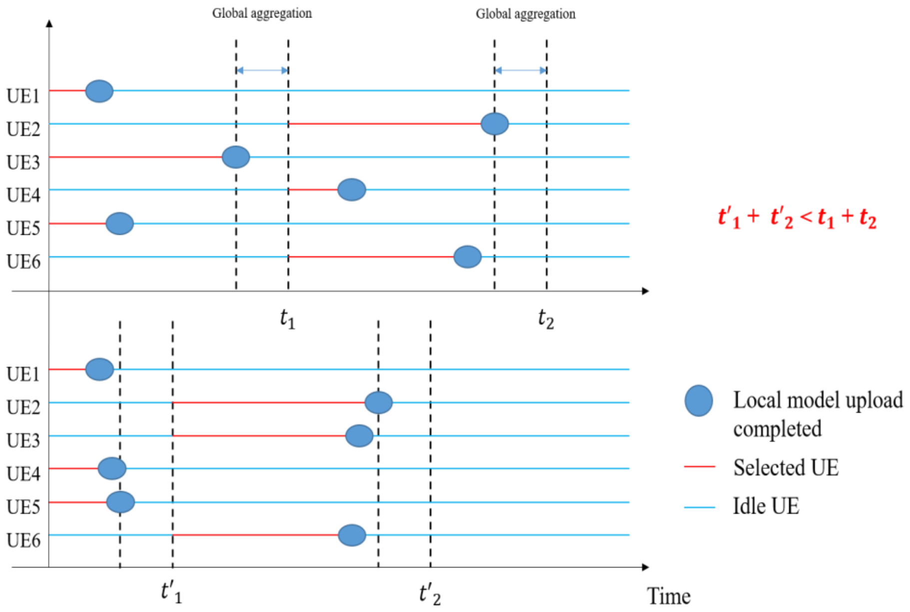


Fig.2

$t'_1 + t'_2 < t_1 + t_2$

In Fig.2-a:

First round: (UE#1, UE#3, UE#5) UE#3 is slower

Second round: (UE#2, UE#4, UE#6) UE#4 is faster.

In Fig.2-b: "straggler" is overcome

First round: faster UEs (UE#1, UE#4, UE#5)

Second round: second UEs (UE#2, UE#3, UE#6)

Why the training process becomes faster?

The divergence of training time is reduced!

However, most of the time:

Training time is unknown before UE selection

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Solution: Clustering Based on Communication Time

We can make FL more efficient in time simply by reducing **the divergence of upload time**

$$\boldsymbol{T^{uplod} = T^{comp} + T^{comm}}$$

T^{upload} -- upload time (training time)

T^{comp} -- computation time

T^{comm} -- communication time

T^{comp} and T^{comm} are independent.

The divergence of T^{up} is the sum of divergence of T^{comp} and divergence of T^{comm}

However, most of the time:

1. Computation time is unknown
2. Intuitively, clustering based on **communication time** can still mitigate the straggler effect as the divergence of upload time is still reduced by minimizing the divergence of communication time.

Divergence of T^{comm} ↘ $\longrightarrow$ Divergence of T^{up} ↘

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Problem#2: Unreliable Communication

Models from UEs with failed transmission will be abandoned

The aggregated model will be **biased** towards the model of UEs with good channel quality.

Example:

Round t: UE#1, UE#2, UE#3 (UE#6 abandoned)

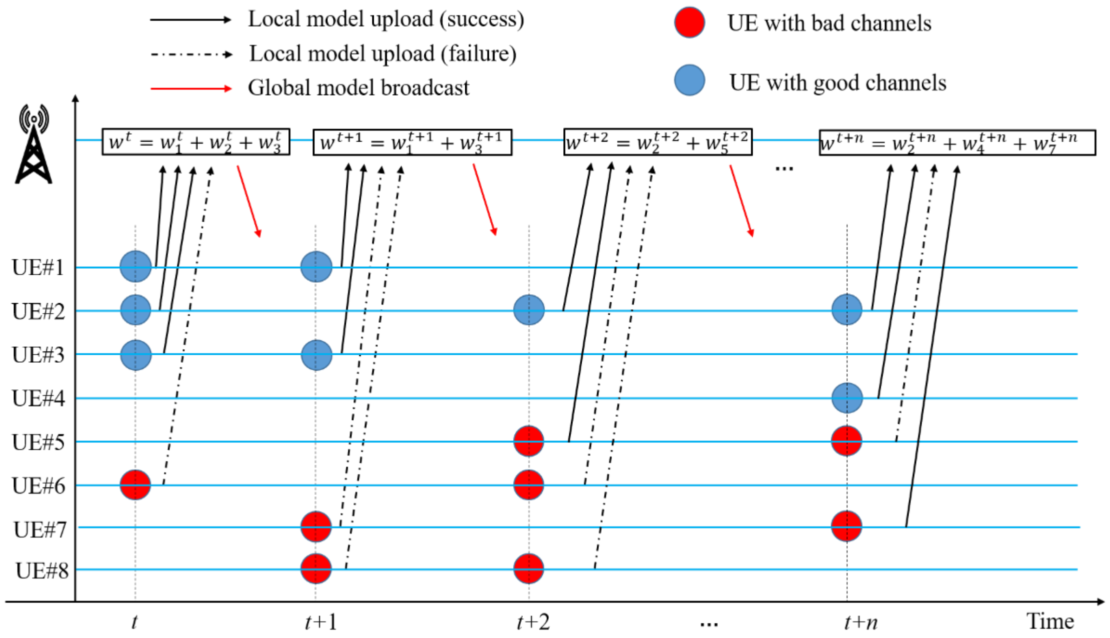
Round t+1: UE#1, UE#3 (UE#7, UE#8 abandoned)

Round t+2: UE#2, UE#5 (UE#6, UE#8 abandoned)

…

Round t+n: UE#2, UE#4, UE#7 (UE#5 abandoned)

Models from Group#1 are more likely to participate into the aggregation



Group#1: UE#1, UE#2, UE#3, UE#4 (Good channels)

Group#2: UE#5, UE#6, UE#7, UE#8 (Bad channels)

Fig.3

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Problem#2: Unreliable Communication

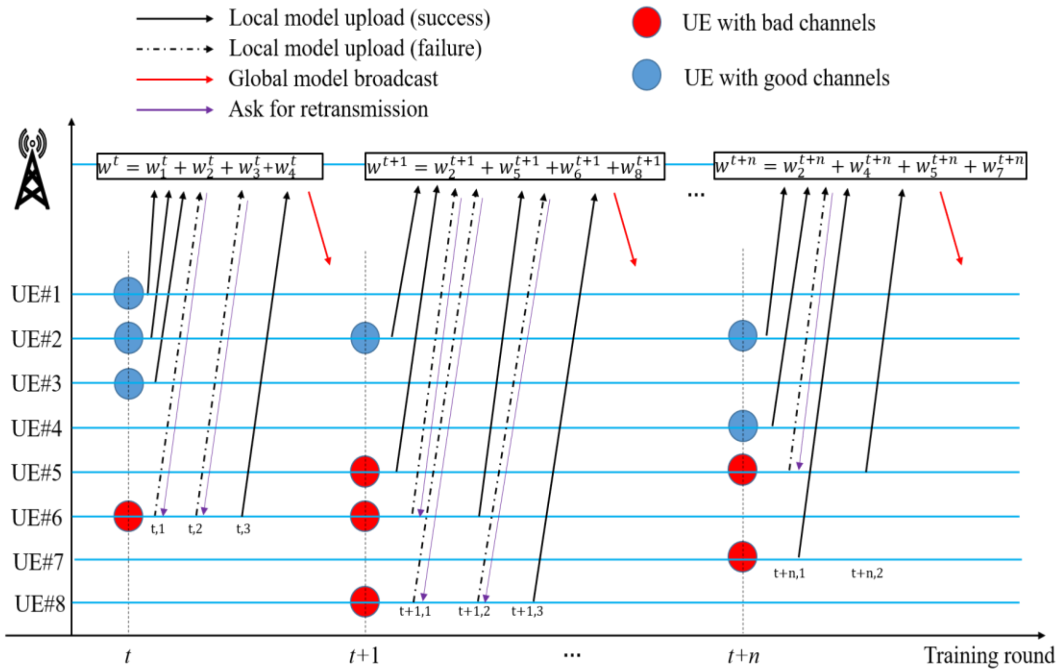
Bad-channel UEs have to retransmit till model successfully received → "straggler"

In round *t*, UE#1, UE#2, UE#3 have to wait. UE#6 transmits 3 times.

The training time is increased due to the retransmission of UE#6.

In this scenario, **the bad-channel UEs transmit more times and become straggler!**

Therefore, we can solve the problem of retransmission **via clustering by treating it as straggler problem**.



Group#1: UE#1, UE#2, UE#3, UE#4 (Good channels)

Group#2: UE#5, UE#6, UE#7, UE#8 (Bad channels)

Fig.4

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Solution: Clustering Based on NR Measurements

Divergence of channel qualities $\searrow$ → Divergence of retransmissions $\searrow$ → Divergence of training time $\searrow$

The NR measurements (e.g., SINR, RSRP, RSRQ, or RSSI, etc) represent channel quality

For UE#k, model is successfully received at *l*-th transmission:

$$P(L_k = l) = p_k(1 - p_k)^{l-1}\ ; \qquad \bar{L}_k = 1/p_k; \qquad \bar{\boldsymbol{T}}_{\boldsymbol{i}} = t/p_k$$

where

p_k ---- the probability that UE#k's model is successfully received with one transmission, i.e., $P(C_i \geq R_i)$;
L_k ---- number of transmissions of UE#k;
t ---- time taken for one transmission
$\bar{L}_i$ ---- average number of transmissions required for UE#i's model is successfully received;
$\bar{T}_i$ ---- average time required for UE#i's model is successfully received

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Solution: Clustering Based on NR Measurements

Divergence of channel qualities ↘ → Divergence of retransmissions ↘ → Divergence of training time ↘

The NR measurements (e.g., SINR, RSRP, RSRQ, or RSSI, etc) represent channel quality

$\bar{T}_k \leftrightarrow \bar{L}_k \leftrightarrow p_k \leftrightarrow$NR Measurement

$$P(L_k = l) = p_k(1 - p_k)^{l-1}\,; \qquad \bar{L}_k = 1/p_k; \qquad \bar{\boldsymbol{T}}_{\boldsymbol{i}} = t/p_k$$

where

p_k ---- the probability that UE#k's model is successfully received with one transmission, i.e., $p_k = P(C_i \geq R_i)$;

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Setup:

Assume we have 100 UEs uniformly distributed over the cell with radius 600 m. In each round, we choose 10 UEs. Consider a Noise-limited scenarios:

Bandwidth: $B = 20\text{MHz}$;
For each UE: $b = \frac{B}{N} = 2\text{MHz}$;
Path-loss exponent: α=3.76;
Noise power: $N_0 = -114$ dBm;
Rayleigh Fading: $|h_k|^2 \sim \exp(1)$;
Transmit power: $P = 10$ dBm

$$R_k = b\log_2\left(1 + \frac{P|h_k|^2 d_k^{-\alpha}}{bN_0}\right)$$

Data partitioning[1].

- **IID:** The MNIST 60,000 training images are partitioned into 100 datasets with each of 600 samples. Then each UE is allocated with one dataset
- **Non-IID:** The 60,000 training images are sorted by digit label and divided into 200 shards of size 300. Then we assign each of 100 clients 2 shards

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "*Communication-efficient learning of deep networks from decentralized data*," in Artificial intelligence and statistics. PMLR, 2017, pp. 1273– 1282.

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

For straggler effect:

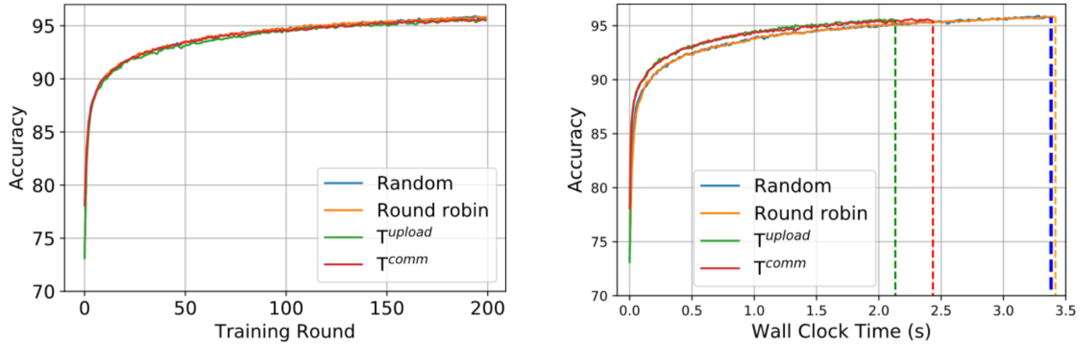
Consider 4 UE selection methods:

Random: 10 UEs are randomly selected
Round robin: 100 UEs are divided into 10 groups, each with 10 UEs. Each group joins FL consecutively
T^{upload}: Clustering based on the upload time
T^{comm}: Clustering based on communication time

Observation:

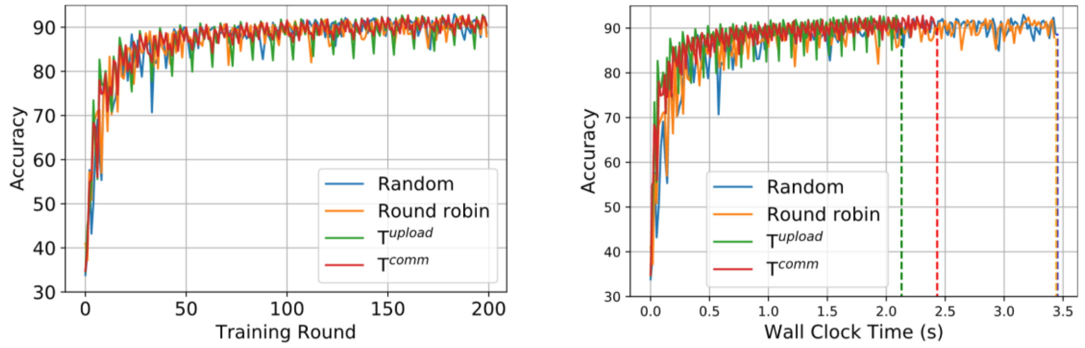
1. Little difference in terms of training round
2. For wall-clock time:

T^{upload} -- 2.13s
T^{comm} -- 2.46s
Others --- 3.40s
Our methods take less time



(a) Training round (b) Wall-clolck time

Fig.5 Straggler effect: comparison of four UE selection methods on **IID dataset** — random selection, round robin, clustering based on upload time and clustering based on communication time.



(a) Training round (b) Wall-clolck time

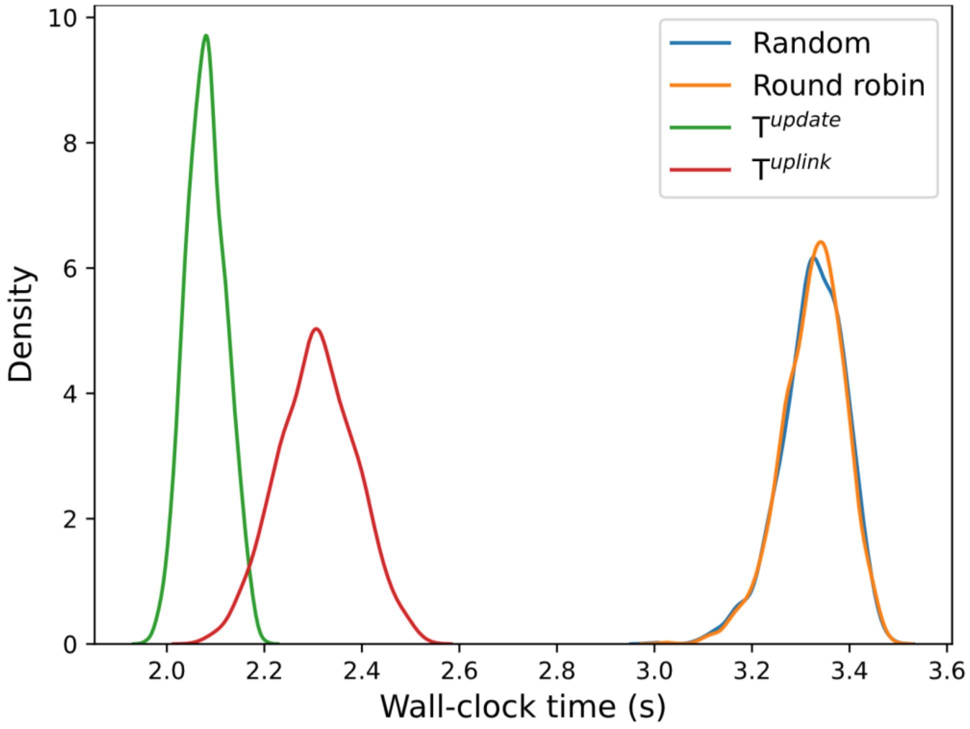
Fig.6 Straggler effect: comparison of four UE selection methods on **non-IID dataset** — random selection, round robin, clustering based on upload time and clustering based on communication time.

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

- For straggler effect:

Consider 4 UE selection methods:

Random: 10 UEs are randomly selected
Round robin: 100 UEs are divided into 10 groups, each with 10 UEs. Each group joins FL consecutively
T^{upload}: Clustering based on the upload time
T^{comm}: Clustering based on communication time



Observation:

Clustering based on upload time is better than that clustering based on communication time because the computation time is also accounted for reducing the time divergence

Fig.7 Straggler effect: PDF of wall-clock time for performing 200 training rounds under four UE selection methods: random selection, round robin, clustering based on upload time and clustering based on communication time.

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

- For unreliable communication:

$R = 15$ MB/s:

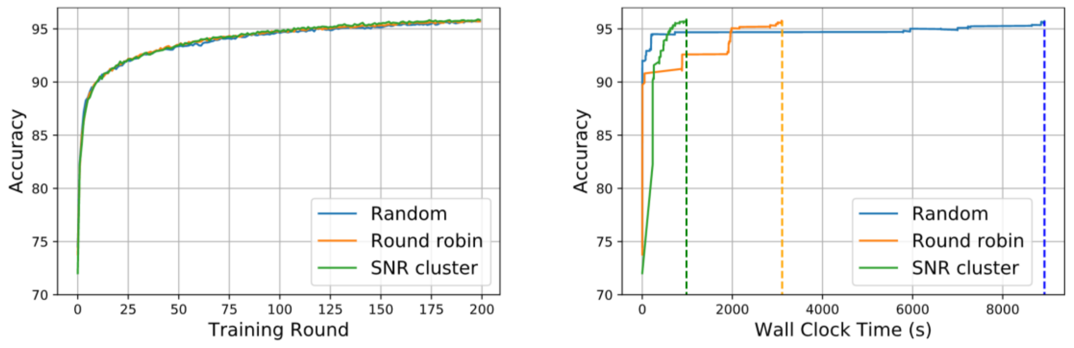
Uniformly choose a value from [7, 10, 13, 16, 19] dBm for each UE as its transmit power.

Consider 3 UE selection methods:

Random: 10 UEs are randomly selected
Round robin: 100 UEs are divided into 10 groups, each with 10 UEs. Each group joins FL consecutively
SNR: Clustering based on SNR

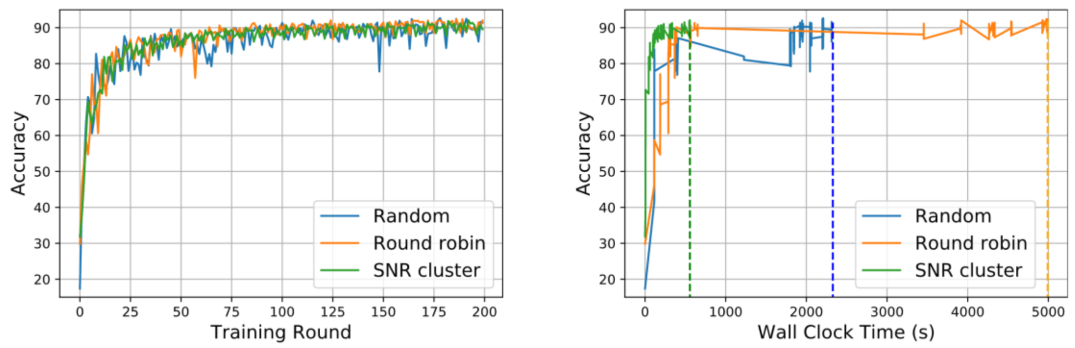
Observation:

Our methods take less time



(a) Training round (b) Wall-clock time

Fig.8 Unreliable communication: comparison of four UE selection methods on **IID dataset** — random selection, round robin, clustering based on upload time and clustering based on communication time.



(a) Training round (b) Wall-clock time

Fig.9 Unreliable communication: comparison of four UE selection methods on **non-IID dataset** — random selection, round robin, clustering based on upload time and clustering based on communication time.

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

- For unreliable communication:

$R = 15$ MB/s:

Uniformly choose a value from [7, 10, 13, 16, 19] dBm for each UE as its transmit power.

Consider 3 UE selection methods:

Random: 10 UEs are randomly selected
Round robin: 100 UEs are divided into 10 groups, each with 10 UEs. Each group joins FL consecutively
SNR: Clustering based on SNR

Observation:

Our methods take less time

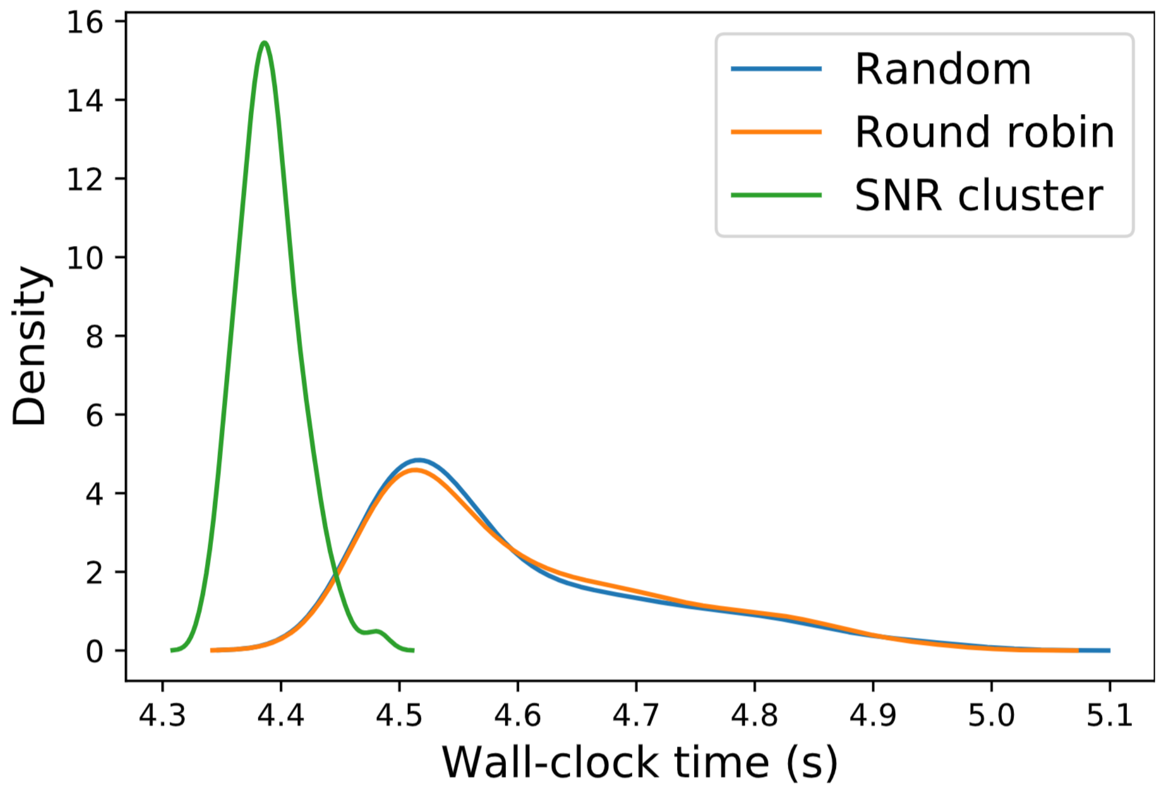


Fig.10 Unreliable Communication: PDF of wall-clock time for performing 200 training rounds under four UE selection methods: random selection, round robin, clustering based on upload time and clustering based on communication time.

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

- General Solution:

General Solution(Clustering based on performance metrics or physical parameters):

1). We cluster the UEs into ***K*** groups based on the ***Q*** where UEs with same or similar ***Q*** are put into the same group;

2). In each training round, only UEs from the same group are selected for FL operation.

Q could be performance metrics, i.e., computation time, communication time, transmission rate, and NR measurement.

It could also be extended to include other physical parameters that impact the time divergence directly or indirectly, e.g., distanc to BS

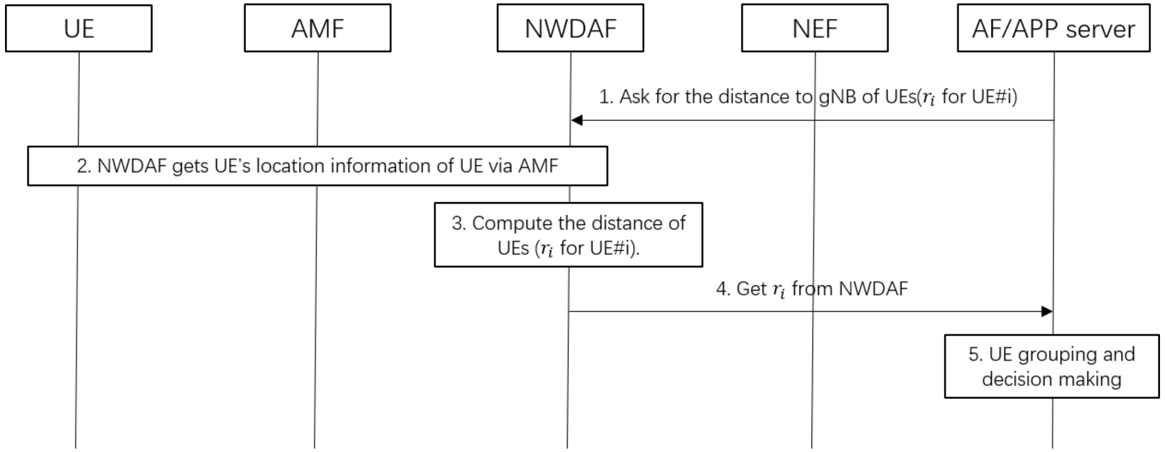
A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

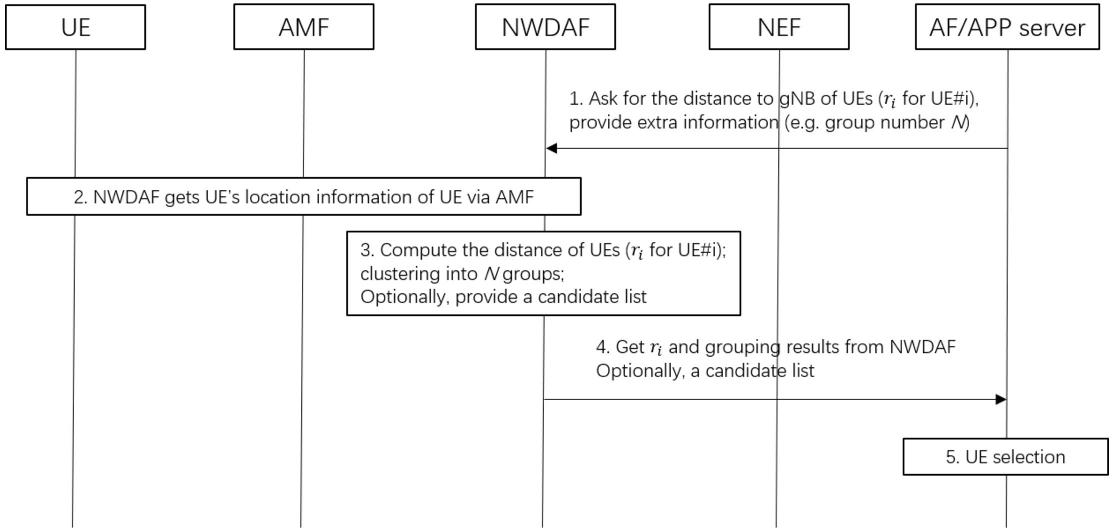
- General Solution (3GPP impact):

Benefits: get r_i directly from NWDAF based on the location of UE#i

1. NWDAF provides r_i and AF does the clustering and makes UE selections

2. NWDAF provides r_i and does clustering (optionally a candidate list)





A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

1. Straggler Problem

- Problem
- Solution

2. Unreliable Communication

- Problem
- Solution

3. Simulation Results

4. Standardization Impact

5. Conclusion

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

- Conclusion (Takeaway):

  Time divergence is the sin for straggler effect and prolonged retransmissions

  Mitigate the straggler or unreliable communication by reducing time divergence

  Combined with or “orthogonal” to other solutions

A General Solution for Straggler Effect and Unreliable Communication in Federated Learning

Thank you & Questions?