Federated Learning under Intermittent and Correlated Client Availability

Angelo Rodio [1,2,3] Othmane Marfoq [1,2,4] Francescomaria Faticanti [1,2] Giovanni Neglia [1,2,3] Emilio Leonardi [5]

[1]Inria [2]Université Côte d'Azur [3]3IA Côte d'Azur [4]Accenture Labs [5]Politecnico di Torino

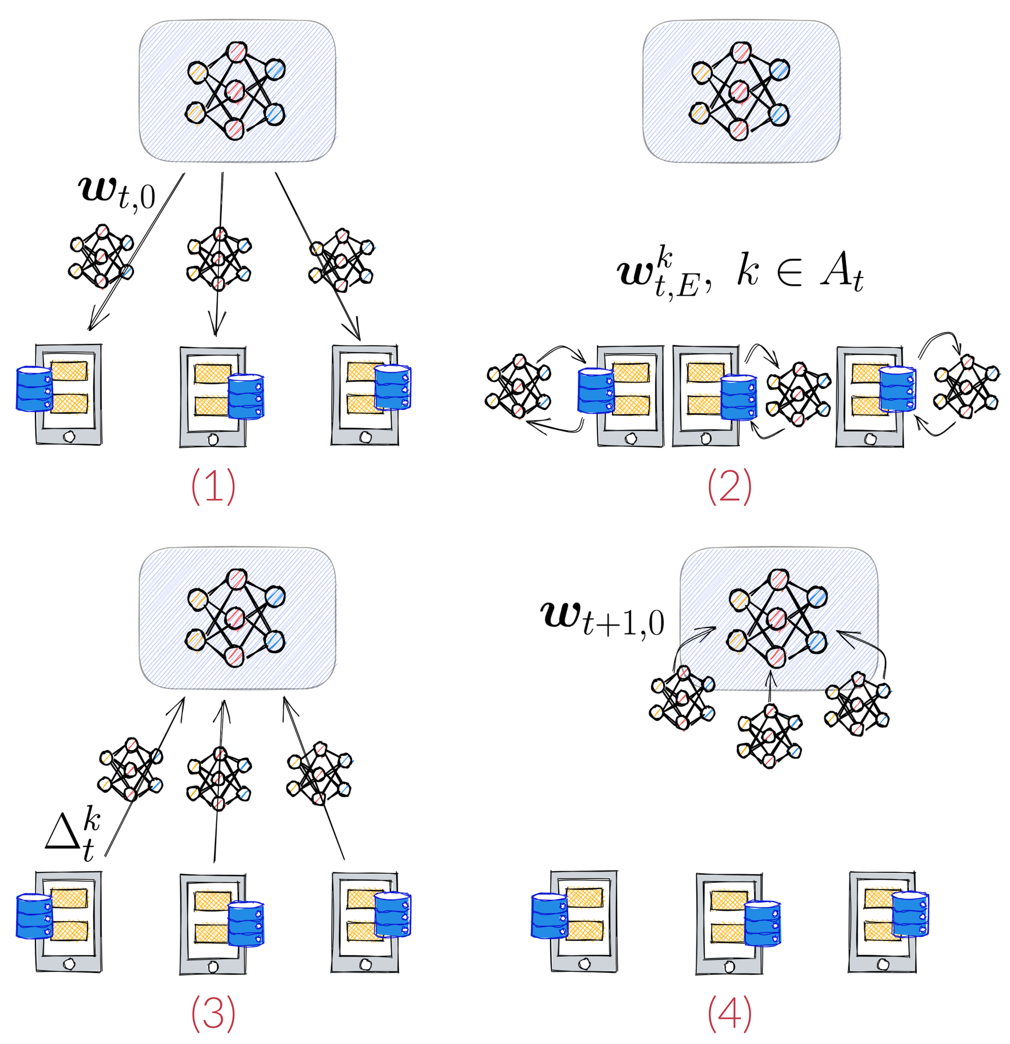
Problem description

- A population of clients $\mathcal{K} = \{1, \ldots, K\}$
- Each client $k \in \mathcal{K}$ holds a local dataset $D_k = \{\xi_{kl}\}_{l=1}^{n_k}$ of size n_k
- Clients learn the parameters $\boldsymbol{w}$ of a global ML model with loss function $f(\boldsymbol{w}; \xi)$
- Client $k \in \mathcal{K}$ has a local objective: $F_k(\boldsymbol{w}) \coloneqq \frac{1}{n_k} \sum_{l=1}^{n_k} f(\boldsymbol{w}; \xi_{kl})$
- In ***Federated Learning***, clients solve, under the orchestration of a central server:

$$\underset{\boldsymbol{w} \in W}{\text{minimize}} \;\; F(\boldsymbol{w}) \coloneqq \sum_{k=1}^{K} \alpha_k F_k(\boldsymbol{w}), \quad \|\boldsymbol{\alpha}\|_1 = 1 \tag{1}$$

$\boldsymbol{\alpha}$: importance weights

A common algorithm to solve (1) is `FedAvg`. For each training round $t > 0$:



$$\Delta^k_t = \boldsymbol{w}^k_{t,E} - \boldsymbol{w}_{t,0}$$

$$\Delta_t = \sum_{k \in A_t} q_k \Delta^k_t$$

$\boldsymbol{q}$: aggregation weights

$$\boldsymbol{w}_{t+1,0} = \boldsymbol{w}_{t,0} + \Delta_t$$

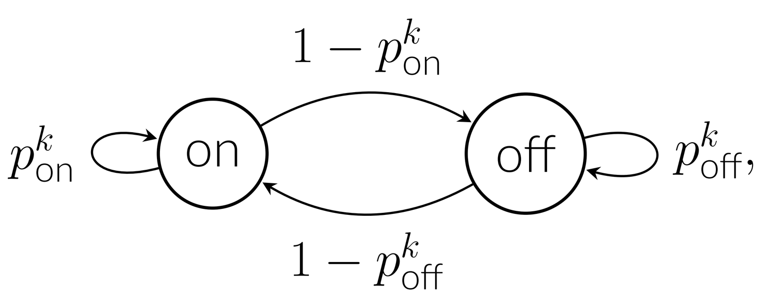
In real-world scenarios, the activity of clients $(A_t)_{t \geq 0}$ is dictated by exogenous factors beyond the control of the orchestrating server and hard to predict

- **Temporal correlation:** the activity of a client is correlated over time
- **Spatial correlation:** the activity is correlated across clients

Intermittent and Correlated Client Availability

Main assumption

Clients' activities follow a DTMC $(A_t)_{t \geq 0}$ with transition matrix $\boldsymbol{P}$ and stationary distribution $\boldsymbol{\pi}$. E.g., each client $k \in \mathcal{K}$ evolves independently according to $(A^k_t)_{t \geq 0}$



$$\boldsymbol{P} = \bigotimes_{k=1}^{K} \boldsymbol{P_k}, \;\; \boldsymbol{\pi} = \bigotimes_{k=1}^{K} \boldsymbol{\pi_k}, \;\; \lambda(\boldsymbol{P}) = \max_{k \in [K]} \lambda(\boldsymbol{P_k})$$

The intermittent availability introduces a model bias

Under intermittent availability $\boldsymbol{\pi}$, `FedAvg` converges to a biased objective $F_B(\boldsymbol{w})$:

$$F_B(\boldsymbol{w}) \coloneqq \sum_{k=1}^{K} p_k F_k(\boldsymbol{w}), \;\; p_k = \frac{\pi_k q_k}{\langle \boldsymbol{\pi}, \boldsymbol{q} \rangle} \quad \neq \quad F(\boldsymbol{w}) \coloneqq \sum_{k=1}^{K} \alpha_k F_k(\boldsymbol{w}) \tag{2}$$

$\boldsymbol{p}$: biased importance $\boldsymbol{\alpha}$: true importance

The correlated availability slows down convergence

$$\mathbb{E}[F_B(\bar{\boldsymbol{w}}_{T,0}) - F^*_B] \leq \mathcal{O}\left(\frac{1}{\sqrt{T}} \cdot \frac{1}{\ln(1/\lambda(\boldsymbol{P}))}\right) \tag{3}$$

where T is the total communication rounds and $\lambda(\boldsymbol{P})$ quantifies correlation

Convergence in terms of the true objective

$$\epsilon(\boldsymbol{q}) \coloneqq F(\boldsymbol{w}_{T,0}) - F^* \;\leq\; \underbrace{\mathcal{O}\left(F_B(\boldsymbol{w}_{T,0}) - F^*_B\right)}_{\coloneqq\, \epsilon_{\text{opt}}(\boldsymbol{q})} + \underbrace{\mathcal{O}\left(d^2_{TV}(\boldsymbol{\alpha}, \boldsymbol{p})\Gamma\right)}_{\coloneqq\, \epsilon_{\text{bias}}(\boldsymbol{q})} \tag{4}$$

where $d_{TV}(\boldsymbol{\alpha}, \boldsymbol{p}) = \frac{1}{2}\sum_{k=1}^{K} |\alpha_k - p_k|$, and $\Gamma = \max_{k \in [K]}\{F_k(\boldsymbol{w}^*_B) - F^*_k\}$

Objective: find the optimal aggregation weights $\boldsymbol{q}^*$ that minimize $\epsilon(\boldsymbol{q})$

Our algorithm: `CA-Fed`

From the optimization problem, we derive the following guidelines:

A) Some clients can be excluded from training, i.e., receive $q^*_k = 0$
B) Exclude clients with low availability π_k and high correlation $\lambda(\boldsymbol{P_k})$
C) Assign allocation $q_k = \alpha_k / \pi_k$ to the included clients

Combining these guidelines, we propose a **client aggregation strategy** (`CA-Fed`) that dynamically excludes clients from training and improves convergence rate

Experiments

Population with $K = 24$ clients, divided in:

- "More available" clients with large π_k
- "Less available, weakly correlated" clients with low π_k, low $\lambda(\boldsymbol{P_k})$
- "Less available, correlated" clients with low π_k, large $\lambda(\boldsymbol{P_k})$

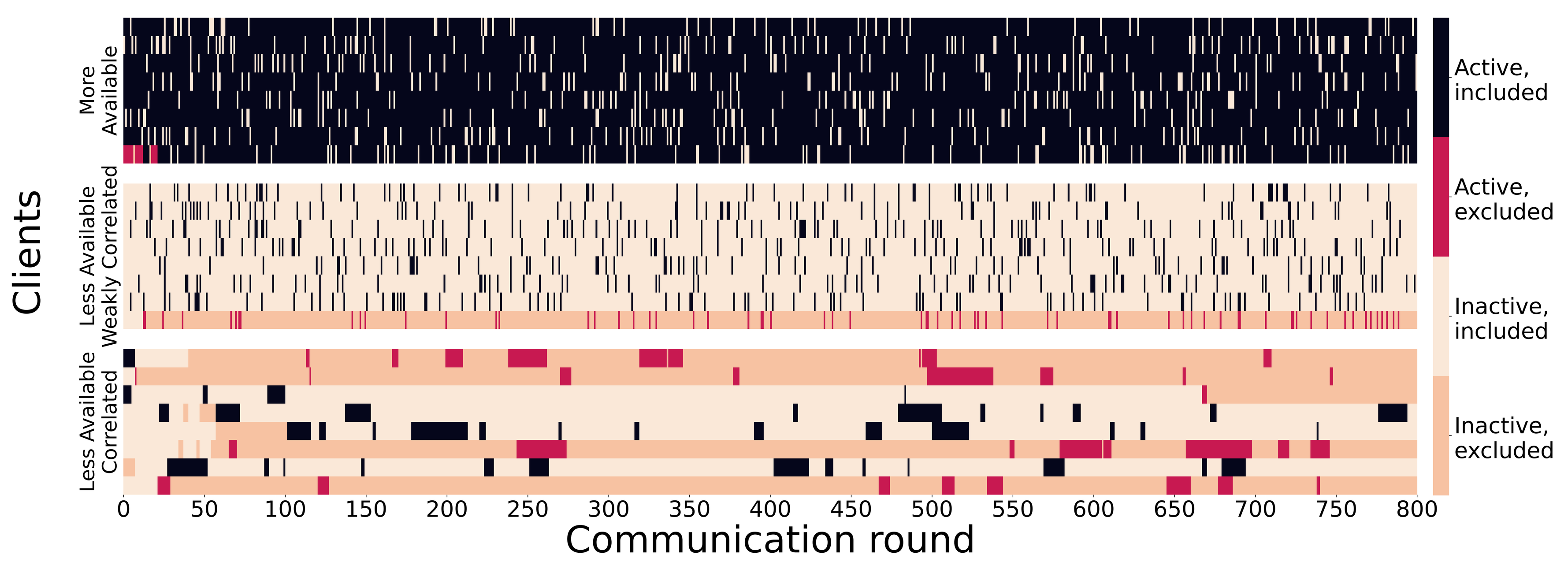


Figure 1. Clients' activities (active/inactive) and `CA-Fed`'s decisions (included/excluded)

We compare `CA-Fed` with the `Unbiased` baseline that assigns $q_k = \alpha_k / \pi_k \; \forall k \in \mathcal{K}$:

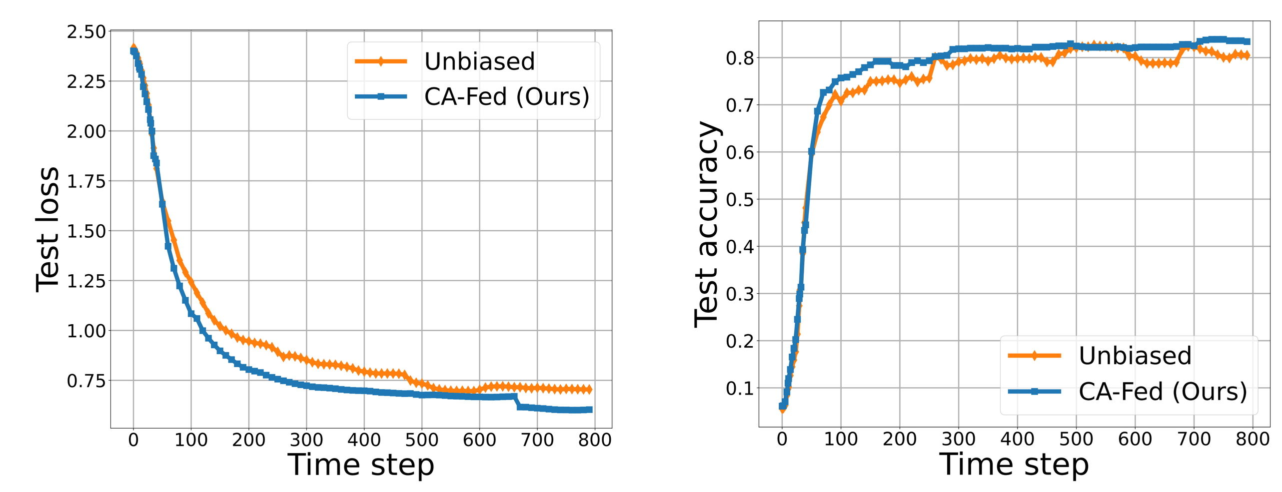
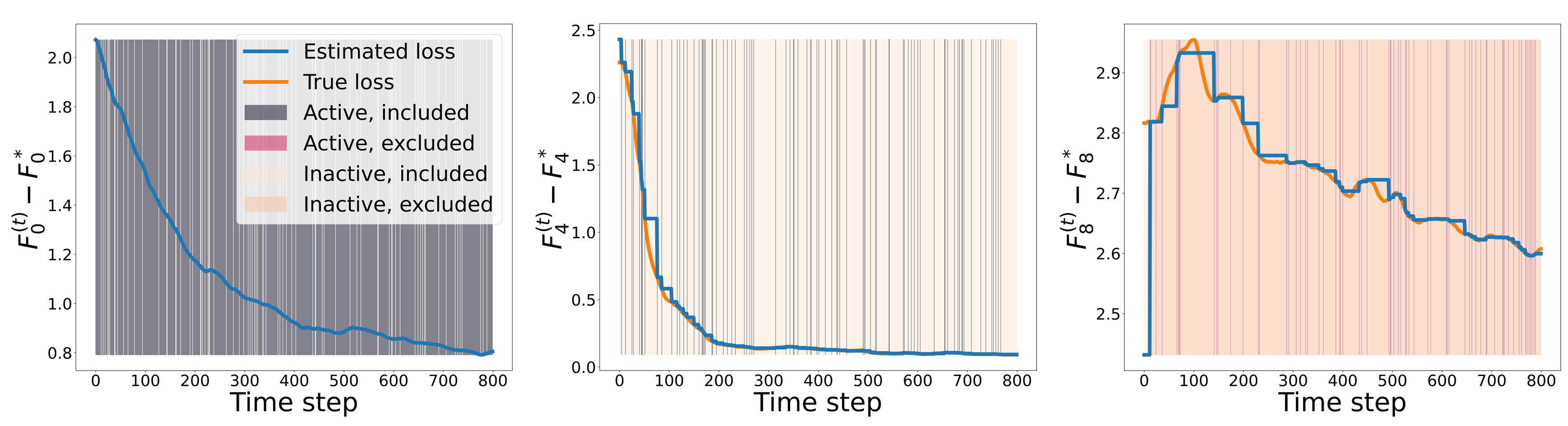
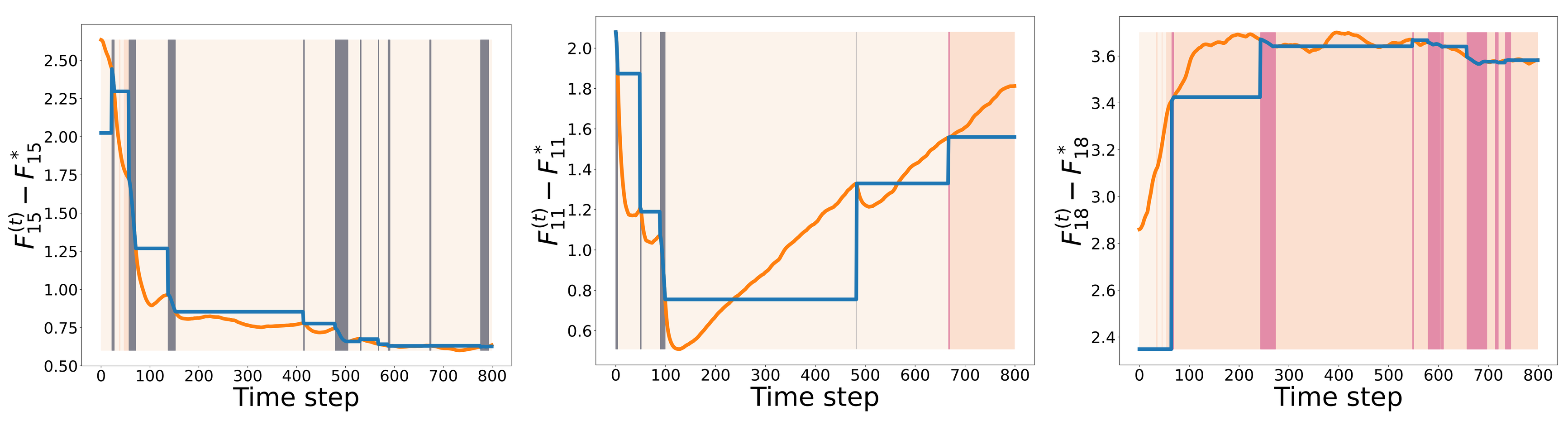


Figure 2. Test loss/accuracy vs communication round for `Unbiased` and `CA-Fed`



(1) More available (included) (2) Less available / weakly correlated (included) (3) Less available / weakly correlated (excluded)



(4) Less available / correlated (included) (5) Less available / correlated (borderline) (6) Less available / correlated (excluded)

Figure 3. Details on per-client losses vs communication round

`CA-Fed` excludes clients from training without performance drop

Conclusions

- Introducing a correlation process in the modeling of FL population
- First convergence analysis under intermittent and correlated client availability
- Adaptively excluding less available and correlated clients can be effective
- Excluding clients also reduces the overall training cost