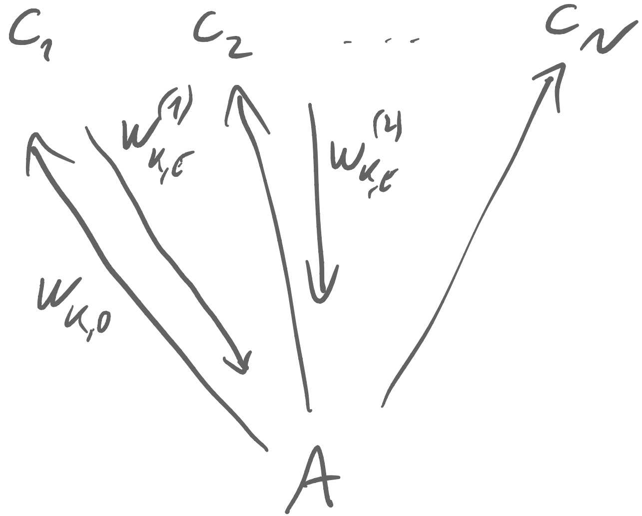
FL & DATA PRIVACY
7/2/23



AGGREGATOR

$$w_{k+1,0} = \sum_{i=1}^{N} \alpha_i \, w_{k,E}^{(i)}$$

$$\sum_i \alpha_i = 1$$

EACH CLIENT

$$\underbrace{w_{k,h+1}^{(i)} = w_{k,h}^{(i)} - \eta_h \nabla F(w_{k,h}^{(i)})}$$

$$\alpha_i = \frac{1}{N}$$ PER-CLIENT FAIRNESS

$$\alpha_i = \frac{|S_i|}{|\bigcup_j S_j|}$$ PER-SAMPLE FAIRNESS

CLIENTS ARE NOT ALWAYS AVAILABLE

A_k : SET OF CLIENTS AVAILABLE AT TIME K

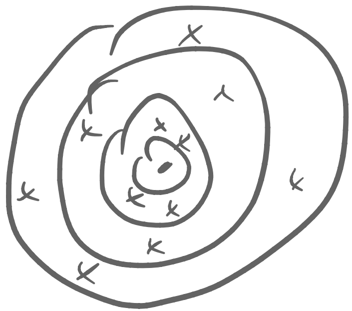
USE ALL CLIENTS IN A_k

$$w_{k+1,0} = \frac{\sum_{i \in A_k} q_i \, w_{k,E}^{(i)}}{\sum_{i \in A_k} q_i}$$

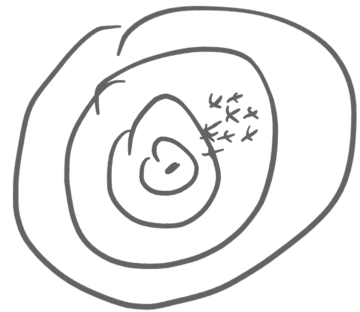
q_i : AGGREG. WEIGHTS

WE CONSIDER THIS AGGR. SCHEME UNTIL WE SAY OW.

$$w_{k+1,0} = w_{k,0} + \eta_s \sum_{i \in A_k} q_i \underbrace{\left(w_{k,E}^{(i)} - w_{k,0}^{(i)}\right)}_{\simeq -\nabla F^{(i)}}$$



UNBIASED



BIASED

BUT LESS VARIANCE

α_i : IMPORTANCE WE TARGET FOR CLIENT i

q_i : AGGREGATION WEIGHT OF CLIENT i

k	1	2	3	4	5	6
C_1	—	—		—		
C_2			—			
C_3	—	—	—	—	—	—

$q_1 = 2$ $\quad$ $\alpha_i = \frac{1}{3}$

$q_2 = 6$

$q_3 = 1$

$\pi_1 = \frac{1}{2}$

$\pi_2 = \frac{1}{6}$

$\pi_3 = 1$

π_i : PROB THAT CLIENT i IS AVAILABLE AT A GIVEN TIME

$$w_{k+1,0} = w_{k,0} - \eta_s \sum_{i \in A_k} q_i \left(w_{k,E}^{(i)} - w_{k,0}^{(i)} \right)$$

WE OPTIMIZE

$$F_B(w) = \sum_i p_i F^{(i)}(w) \neq \sum_i \alpha_i F^{(i)}(w)$$

BIASED FUNCTION

$$\boxed{p_i \propto \pi_i q_i}$$

$$p_i = \frac{\pi_i q_i}{\sum_j \pi_j q_j}$$

HOW OFT THE CLIENT PARTICIPATES

WHAT IMP YOU GIVE TO THE CLIENT AT THE AGGREGATION

WHEN ALL CLIENTS ARE AVAILABLE ALL THE TIME

$\pi_i = 1 \quad \forall i$

$p_i \propto \pi_i q_i = q_i$

IF WE WANT $p_i = \alpha_i$

$q_i = ?$ $\quad q_i = \alpha_i$

OR ALL CLIENT ARE EQUALLY AVAILABLE

$\pi_i = \pi \quad \forall i$

$$p_i = \frac{\pi q_i}{\sum_j \pi q_j}$$

IF WE TAKE

$q_i = \alpha_i$

$\Rightarrow p_i = \alpha_i$

$F_B = F_T$

π_i IN GENERAL DIFFERENT

BUT WE WANT $p_i = \alpha_i$

(WE WANT TO OPTIMIZE OUR TARGET OBJECTIVE)

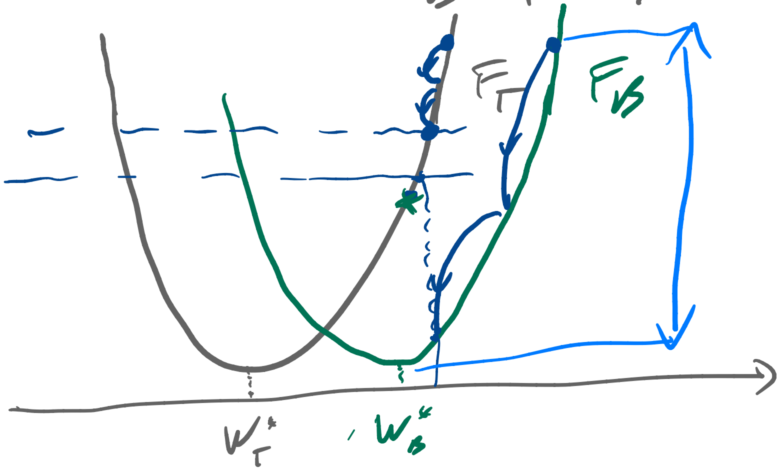
HOW TO PICK q_i???

$$p_i \propto \pi_i q_i$$

ANSWER

$$q_i = \frac{\alpha_i}{\pi_i} \quad \rightarrow \quad p_i \propto \cancel{\pi_i} \frac{\alpha_i}{\cancel{\pi_i}} = \alpha_i$$

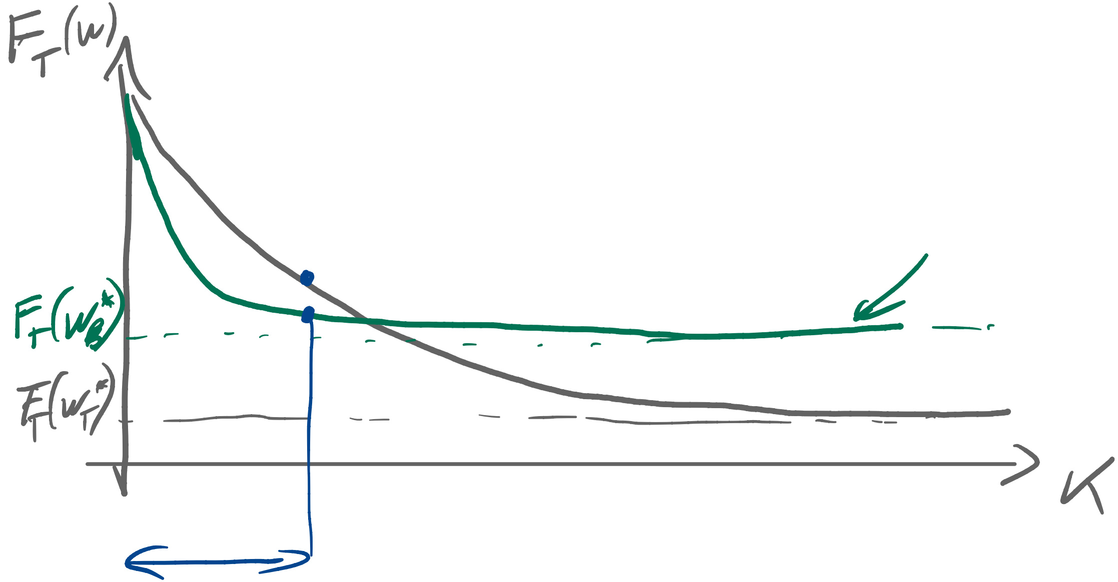
MAY WE EVER WANT TO OPTIMIZE A BIASED $F_B \neq F_T$



$$F_T = \sum_i \alpha_i \cdot F^{(i)}$$

K COMMUNICATION ROUNDS

MAY MAKE SENSE IF OPTIMIZING F_B IS SOMEWAY EASIER THAN OPTIMIZING F_T ~~AND~~ SO YOU MAY BE BETTER OFF TARGETING F_B



IF IT MAY BE EASIER TO MINIMIZE F_B

ANGELO RODIO: FL UNDER HETEROGENEOUS AND CORRELATED CLIENTS' AVAILABILITY

$$F_B(\boldsymbol{w}) := \sum_{k=1}^N p_k F_k(\boldsymbol{w}), \text{ with } p_k := \frac{\pi_k q_k}{\sum_{h=1}^N \pi_h q_h}, \tag{4}$$

$F \equiv F_T$ $\quad F_u \equiv F^{(u)}$

$w^* \in \arg\min F_T(w)$

$F_u^* = \min_w F_k(w)$

$$\Gamma := \max_{k\in\mathcal{K}}\{F_k(\boldsymbol{w}^*) - F_k^*\}.$$

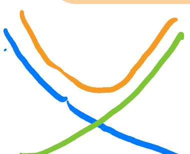
QUANTIFIES THE HETEROGENEITY

$F_{u'} = F_{u''}$ ALL EQUAL

$F_T = \sum \alpha_k F_k =$
$= F_k$

$\Rightarrow \Gamma = 0$

Theorem 1 (Decomposing the total error). *Under Assumptions 2–4, the optimization error of the target global objective* $\epsilon = F(\boldsymbol{w}) - F^*$ *can be bounded as follows:*



$$\epsilon \leq \underbrace{2\kappa^2(F_B(\boldsymbol{w}) - F_B^*)}_{:=\epsilon_{opt}} + \cancel{\underbrace{2\kappa^4\chi^2_{\boldsymbol{\alpha}\|\boldsymbol{p}}\Gamma}_{:=\epsilon_{bias}}}, \tag{10}$$

where $\kappa := L/\mu$, *and* $\chi^2_{\boldsymbol{\alpha}\|\boldsymbol{p}} := \sum_{k=1}^N (\alpha_k - p_k)^2/p_k$.

$q_k = \frac{\alpha_k}{\pi_k}$ $\quad p_k = \alpha_k, \chi^2_{\alpha\|p} = 0$

Theorem 2 (Convergence of the optimization error ϵ_{opt}). *Let Assumptions 1–5 hold and the constants $M, L, D, G, H, \Gamma, \sigma_k, C_P, T_P, \lambda(\boldsymbol{P})$ be defined as above. Let $Q = \sum_{k \in \mathcal{K}} q_k$. Let the stepsizes satisfy:*

$$\sum_t \eta_t = +\infty, \quad \sum_t \ln(t) \cdot \eta_t^2 < +\infty. \tag{11}$$

Let T denote the total communication rounds. For $T \geq T_P$, the expected optimization error can be bounded as follows:

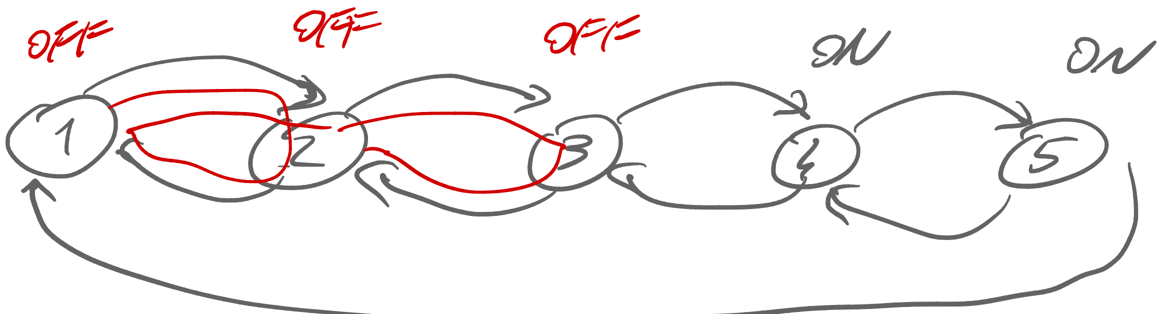
$$\mathbb{E}[F_B(\bar{\boldsymbol{w}}_{T,0}) - F_B^*] \leq \frac{\frac{\frac{1}{2}\boldsymbol{q}^\intercal \boldsymbol{\Sigma} \boldsymbol{q} + \upsilon}{\boldsymbol{\pi}^\intercal \boldsymbol{q}} + \psi + \frac{\phi}{\ln(1/\lambda(\boldsymbol{P}))}}{\left(\sum_{t=1}^{T} \eta_t\right)}, \tag{12}$$

$\lambda(P) = \max_k |\lambda_2(P_k)|$

$\frac{1}{\pi^\intercal q}$

minimize $\frac{1}{\pi^\intercal q}$

$\Rightarrow$ HIGHER q_k TO ON OFF MODE AVAILABLE CLIENTS



P TRANS. MATRIX

P IS TRANS. MATRIX

$$|\lambda_1(P)| \geq |\lambda_2(P)| \geq \dots |\lambda_n(P)|$$

$$\lambda_1(P) = 1$$

AFFECTS TIME CORRELATION

$|\lambda_2(P)| \simeq 1$ HIGH CORRELATED

$|\lambda_2(P)| \to 0$ NO CORRELATION

WHAT WE LEARN FROM THEOREM 2

IN ORDER TO SPEED UP THE TRAINING
WE MAY WANT

1) TO IGNORE CLIENTS WITH LOW AVAILABILITY

2) " " CLIENTS HIGHLY CORRELATED (OVER TIME)

$$w_{k+1,0} = w_{k,0} + \eta_s \underbrace{\sum_{i \in A_k} q_i \left(w_{k,E}^{(i)} - w_{k,0}^{(i)}\right)}$$

VARIABLE # OF CLIENTS

QUITE NOISY

$\Rightarrow$ η_s SMALL

1st PROBLEM

2nd " : YOU MAY NOT BE ABLE TO MANAGE ALL AVAILABLE CLIENTS

GOOGLE ALGO (FedAvg)

FOR $k = 1$ TO K

A_k : CLIENTS AVAILABLE

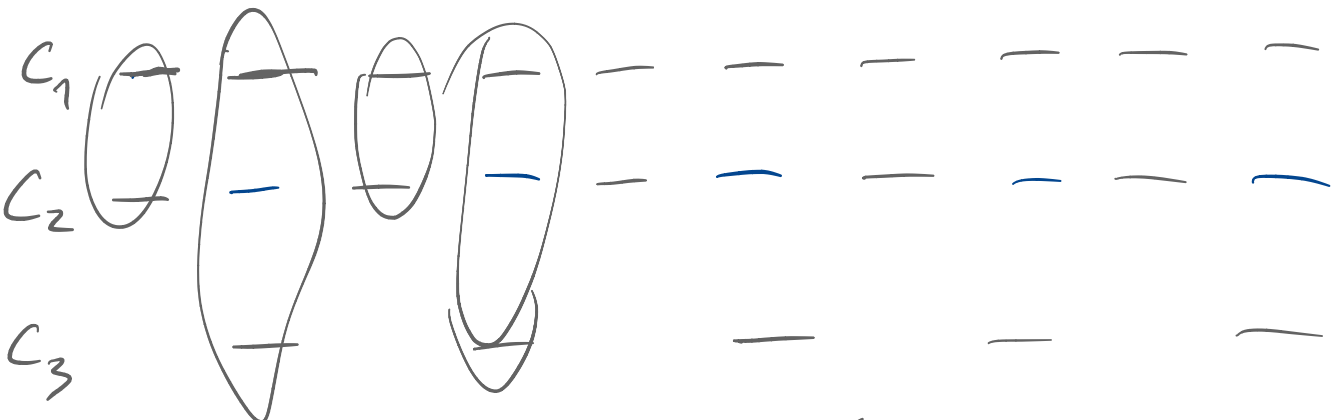
PICK N CLIENTS UNIFORMLY FROM A_k
(SAMPLING) $B_k \subset A_k$

SEND $w_{k,0}$ TO THE N CLIENTS

$$w_{k+1,0} = \frac{1}{N} \sum_{i \in B_k} q_i \, w_{k,E}^{(i)}$$

$$\left(= w_{k,0} + \eta_s \frac{1}{N} \sum_{i \in B_k} q_i \left(w_{k,E}^{(i)} - w_{k,0}^{(i)} \right) \right)$$

STILL PRONE TO THE BIAS DUE TO HETEROG. CLIENT AVAILAB.



$\pi_1 = 1$

$\pi_3 = \frac{1}{2}$

IF $N=2$ $\quad Prob(1 \in B_a) = \pi_1 = 1$

$$Prob(1 \in B_a) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \frac{2}{3} = \frac{5}{6} < 1$$

MUCH MORE DIFFICULT TO STUDY
HOW TO ACCOUNT THE EFFECT
OF HETEROG. AVAILAB. ?

Li CONVERG. OF FedAvg UNDER NON-IID DISTRIB.

M CLIENTS

1) , EVERY CLIENT HAS
PROB. $\frac{N}{M}$ TO BE SELECTED AT ANY SLOT
AND $q_k = \alpha_i$

2) EVERY CLIENT IS SELECTED W. P. α_i
AND THE $q_i = \frac{1}{N}$

RECENT PAPER FROM

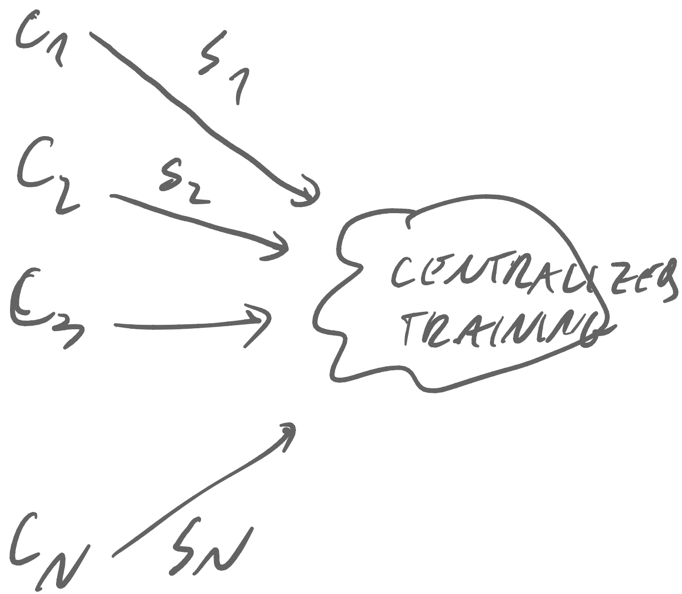
RIBERO, DE VECIANA, ?

PRACTICAL NUMBERS

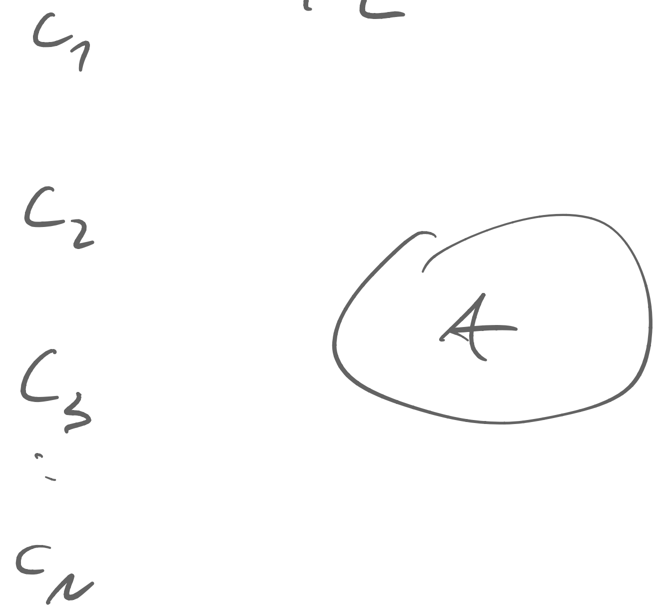
- $N = 100-500$
- TRAINING DURATION: WEEKS
- 1 COMM. ROUND : 1h

$20 \times 7 \times 2 \times 100 \simeq 30K$ CLIENTS INVOLVED

COMMUNICATION COST REDUCED?



FL



COMM. COST

IS $|S_1| + |S_2| + \dots |S_N|$

COMM. COST

$$\sum_{k=1}^{K} |B_k| \cdot \text{SIZE MODEL}$$

$$= K N \cdot \text{SIZE MODEL}$$

REDUCE COMMUNICATION COST

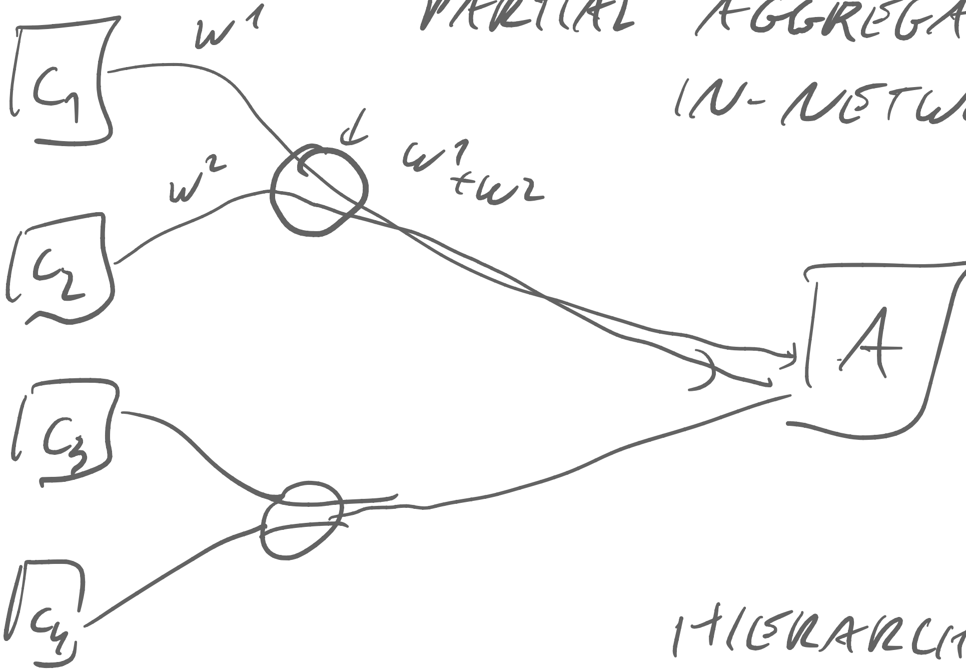
COMPRESS $\rightarrow$ ENCODE NUMBERS WITH LESS BYTES

$\rightarrow$ TOP-K

NOT NECESSARILY TO $w_{k,E}^{(i)}$

BUT $w_{k,E}^{(i)} - w_{k,0}$

PARTIAL AGGREGATION IN-NETWORK



HIERARCHICAL FL

LARGE SENSOR NETWORK