Adaptive Gradient Clipping for Robust Federated Learning

EPFL

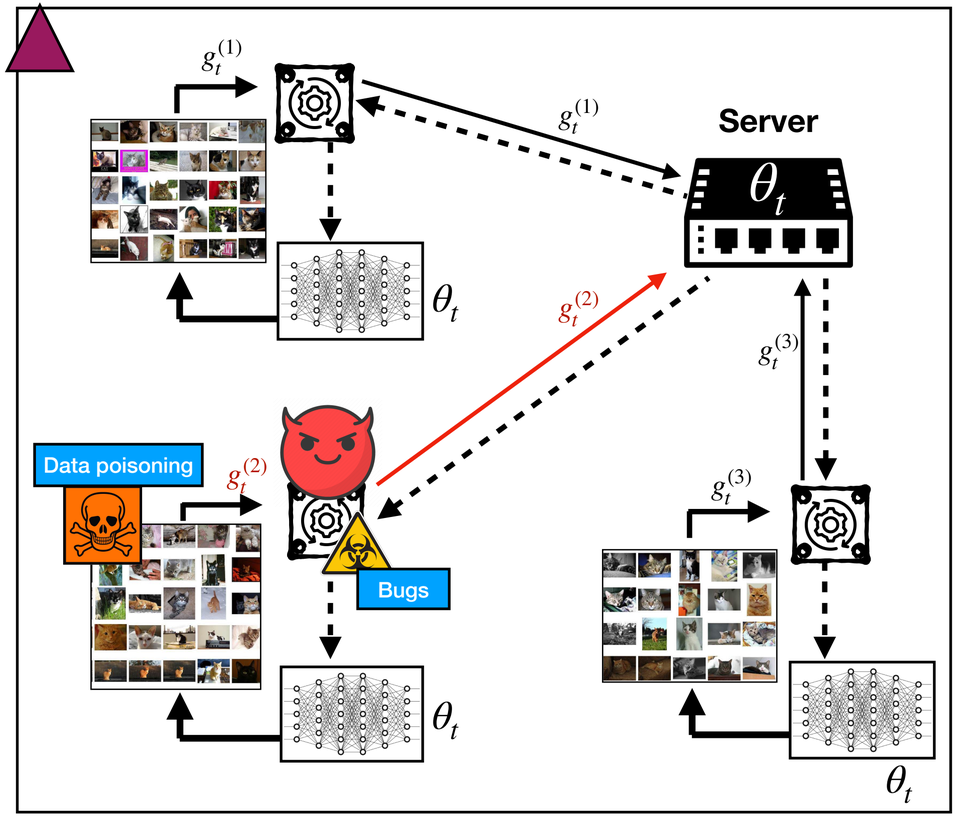
Youssef Allouah, Rachid Guerraoui, Nirupam Gupta, Ahmed Jellouli, Geovani Rizk & John Stephan*

nigu@di.ku.dk

UNIVERSITY OF COPENHAGEN

Federated learning is vulnerable to untrusted clients

f out of n clients are adversarial (Byzantine)



Robust aggregation (RobAgg)

Aggregation $F : (v_1, \dots, v_n) \mapsto \hat{v}$ for all $G \subseteq [n]$ of size $n - f$

$$\|\hat{v} - \bar{v}_G\|^2 \leq \kappa \, \frac{1}{n-f} \sum_{i \in G} \|v_i - \bar{v}_G\|^2$$

$\bar{v}_G = \frac{1}{n-f} \sum_{i \in G} v_i$

Coordinate-wise Trimmed Mean **(CWTM)**

$$\mathrm{CWTM}\left(\begin{matrix} 2 & 3 & 2 & \cancel{100} & \cancel{100} \\ 5 & \cancel{6} & 4 & 5 & \cancel{6} \end{matrix}\right) = \begin{pmatrix} 2.3 \\ 4.6 \end{pmatrix}$$

$\kappa \in \Theta\left(\frac{f}{n}\right)$

Lower bound

$\mathscr{L}_H \triangleq \frac{1}{|H|} \sum_{i \in H} \mathscr{L}_i$

$(G, B)-$gradient dissimilarity:

$$\frac{1}{|H|} \sum_{i \in H} \|\nabla \mathscr{L}_i(\theta) - \nabla \mathscr{L}_H(\theta)\|^2 \leq G^2 + B^2 \|\mathscr{L}_H(\theta)\|^2$$

A federated learning algorithm $\mathscr{A}$ with output $\widehat{\theta}$ can only tolerate $\frac{f}{n} < \frac{1}{2+B^2} \triangleq \mathrm{BP}$ and

$$\|\mathscr{L}_H(\widehat{\theta})\|^2 \geq \frac{1}{4}\left(\frac{f}{n-(2+B^2)f}\right) G^2 \triangleq \varepsilon_o$$

Proposed scheme: **Adaptive Robust Clipping (ARC)**

Given input vectors $v_1, \dots, v_n$, determine a permutation $\pi : [n] \to [n]$ such that

$$\|v_{\pi(1)}\| \leq \dots \leq \|v_{\pi(n)}\|$$

Set $k = 2\frac{f}{n}(n-f)$ and $C = \|v_{\pi(k+1)}\|$. Then,

$\mathrm{Clip}_C(v) = \min\left\{1, \frac{C}{\|v\|}\right\} v$

$$\mathrm{ARC}(v_1, \dots, v_n) = \left(\mathrm{Clip}_C(v_1), \dots, \mathrm{Clip}_C(v_n)\right)$$

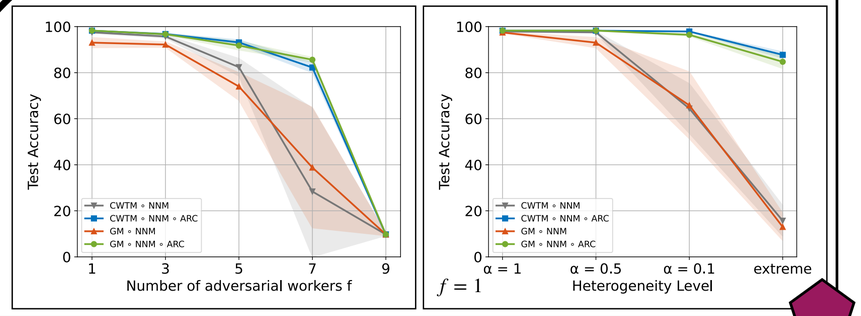
$\theta_{t+1} = \theta_t - \gamma_t \, \mathrm{RobAgg}\left(\mathrm{ARC}\left(v_1, \dots, v_n\right)\right)$. For $\upsilon \in (0,1)$ and $\frac{f}{n} = \left(1 - \frac{\upsilon}{\Psi(\theta_1)}\right) \mathrm{BP}$,

$$\frac{1}{T} \sum_{t=1}^{T} \|\mathscr{L}_H(\theta_t)\|^2 \leq \upsilon \, \varepsilon_o \,.$$

$\Psi(\theta_1)$: monotonically increasing function of $\|\nabla \mathscr{L}_H(\theta_1)\|$

Circumventing L.B. under **bounded initialization error**

Worst-case accuracies over MNIST, distributed over 10 honest clients



$g_t^{(i)} := \nabla \mathscr{L}_i(\theta_t)$, i **is honest**

$g_t^{(j)} =$ **arbitrary**, j **is Byzantine**

$$\theta_{t+1} = \theta_t - \gamma_t \, \cancel{\mathrm{Avg}\left(g_t^{(1)}, \dots, g_t^{(n)}\right)} \quad \longrightarrow \quad \theta_{t+1} = \theta_t - \gamma_t \, \mathrm{RobAgg}\left(g_t^{(1)}, \dots, g_t^{(n)}\right)$$

How to read this poster: ▲ → ● → ■ → ◆ → ★ → ⬟

Paper link:

* corresponding author (john.stephan@epfl.ch)

ICLR 2025 [Spotlight]