BYZANTINE ROBUSTNESS AND PARTIAL PARTICIPATION CAN BE ACHIEVED SIMULTANEOUSLY: JUST CLIP GRADIENT DIFFERENCES

Grigory Malinovsky
KAUST

Eduard Gorbunov
MBZUAI

Samuel Horváth
MBZUAI

Peter Richtárik
KAUST

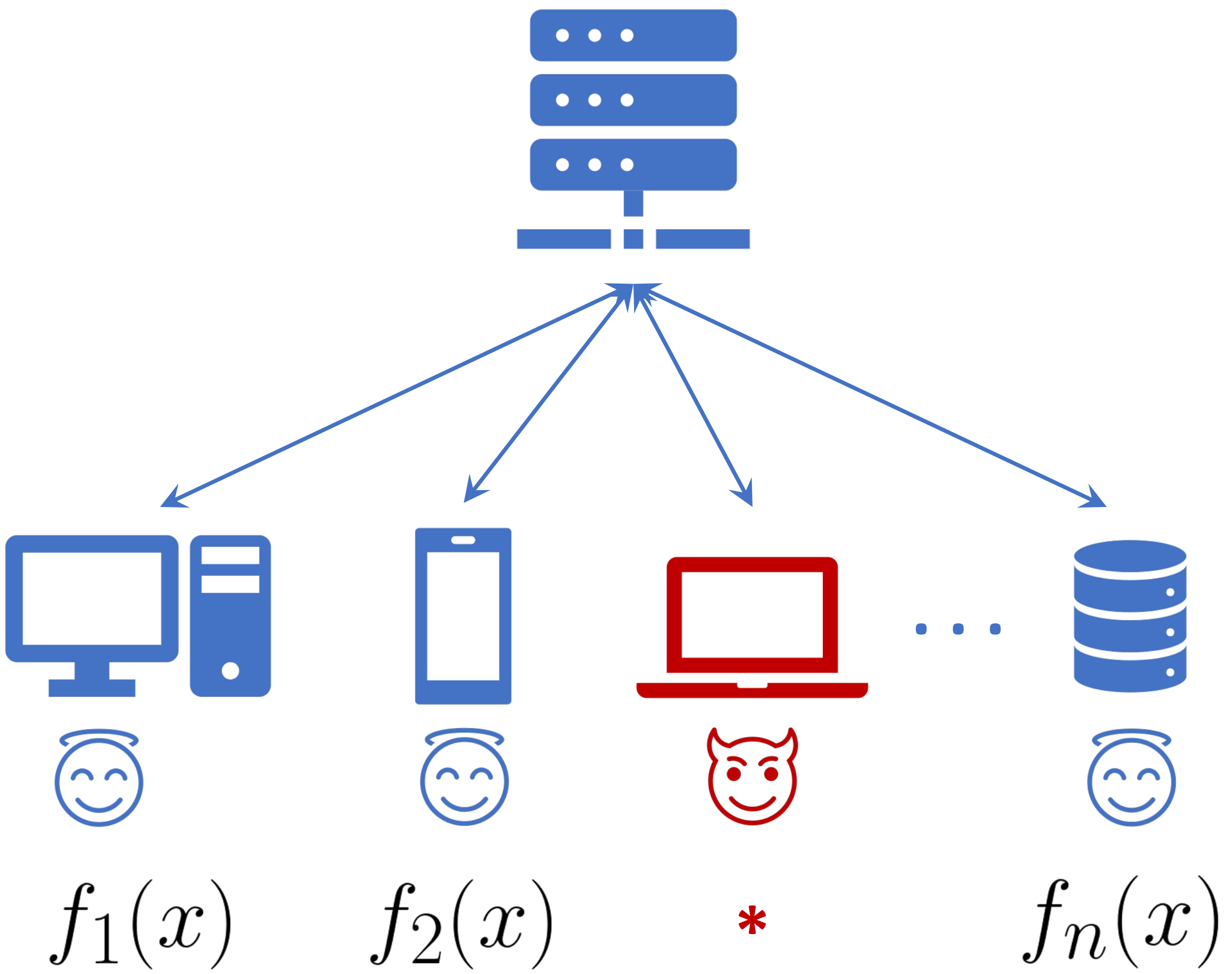
NeurIPS 2024

KAUST

MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

February 21, 2024

Refined Problem Formulation



$$\min_{x \in \mathbb{R}^d} \left\{ f(x) := \frac{1}{G} \sum_{i \in \mathcal{G}} f_i(x) \right\}$$

Good workers form the majority:

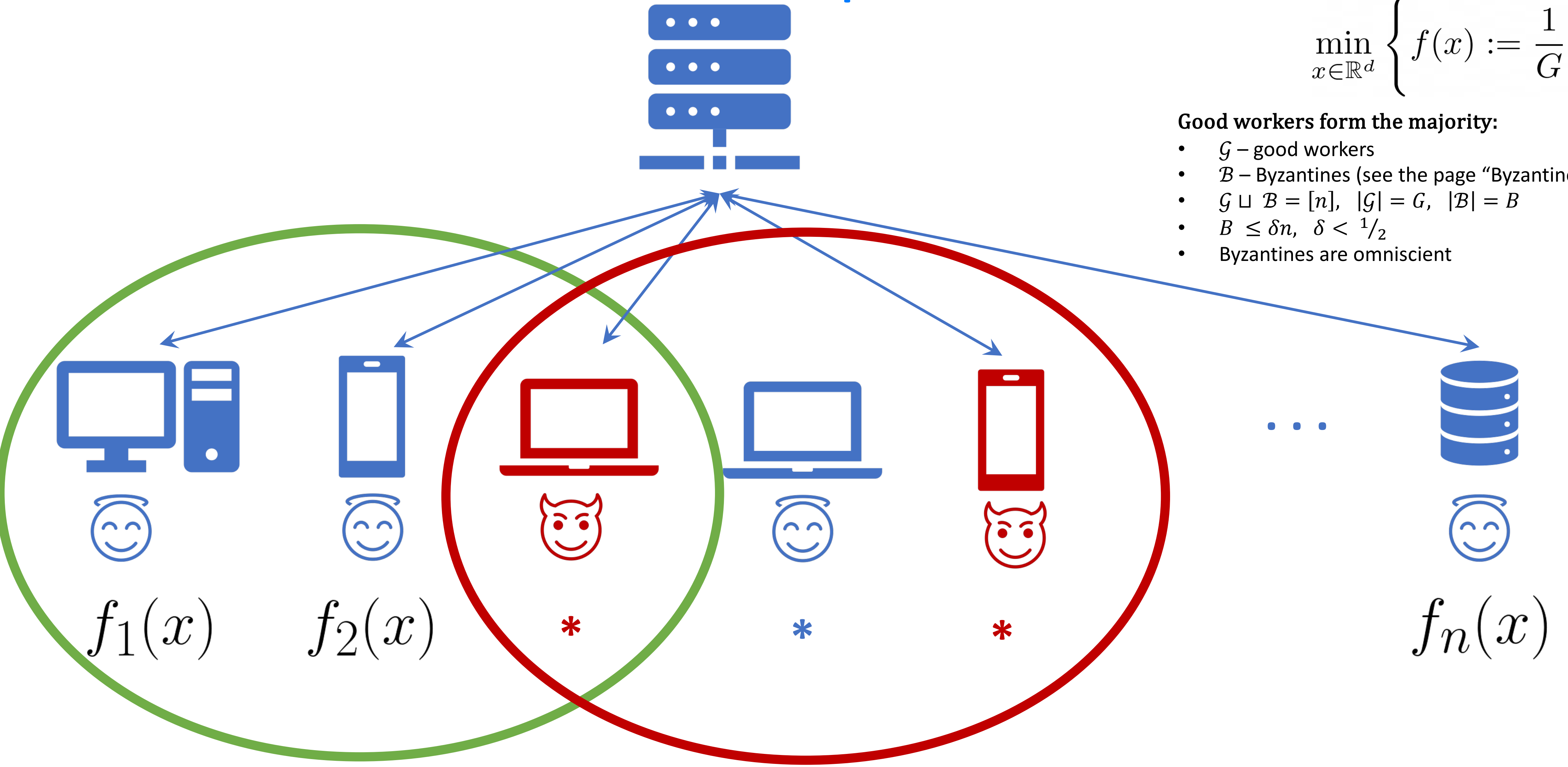
- $\mathcal{G}$ – good workers
- $\mathcal{B}$ – Byzantines (see the page "Byzantine fault" in Wikipedia)
- $\mathcal{G} \sqcup \mathcal{B} = [n], \ \ |\mathcal{G}| = G, \ \ |\mathcal{B}| = B$
- $B \leq \delta n, \ \ \delta < {}^{1}/_{2}$
- Byzantines are omniscient

Partial Participation of clients

$$\min_{x\in\mathbb{R}^d}\left\{f(x) := \frac{1}{G}\sum_{i\in\mathcal{G}} f_i(x)\right\}$$

Good workers form the majority:

- $\mathcal{G}$ – good workers
- $\mathcal{B}$ – Byzantines (see the page “Byzantine fault” in Wikipedia)
- $\mathcal{G} \sqcup \mathcal{B} = [n], \quad |\mathcal{G}| = G, \quad |\mathcal{B}| = B$
- $B \leq \delta n, \quad \delta < {}^1\!/_2$
- Byzantines are omniscient



Question: how to solve such problems?

Remedy to majority of Byzantines

$$\min_{x\in\mathbb{R}^d}\left\{f(x) := \frac{1}{G}\sum_{i\in\mathcal{G}} f_i(x)\right\}$$

Good workers form the majority:

- $\mathcal{G}$ – good workers
- $\mathcal{B}$ – Byzantines (see the page "Byzantine fault" in Wikipedia)
- $\mathcal{G} \sqcup \mathcal{B} = [n], \ |\mathcal{G}| = G, \ |\mathcal{B}| = B$
- $B \le \delta n, \ \delta < {}^1/_2$
- Byzantines are omniscient

Clipping of updates

$$\mathrm{clip}_{\lambda_{k+1}}\left(\nabla f_i(x^{k+1}) - \nabla f_i(x^k)\right)$$

Level of clipping

$$\lambda_{k+1} = \alpha_{k+1}\left\|x^{k+1} - x^k\right\|$$

Lipschitz smoothness

$$\|\nabla f_i(x) - \nabla f_i(y)\| \leqslant L_i\|x - y\|$$

Byz-VR-MARINA-PP

Contributions

1. Application of clipping to Byz-VR-MARINA

2. General non-convex and PL analysis

3. Additional variants of algorithm

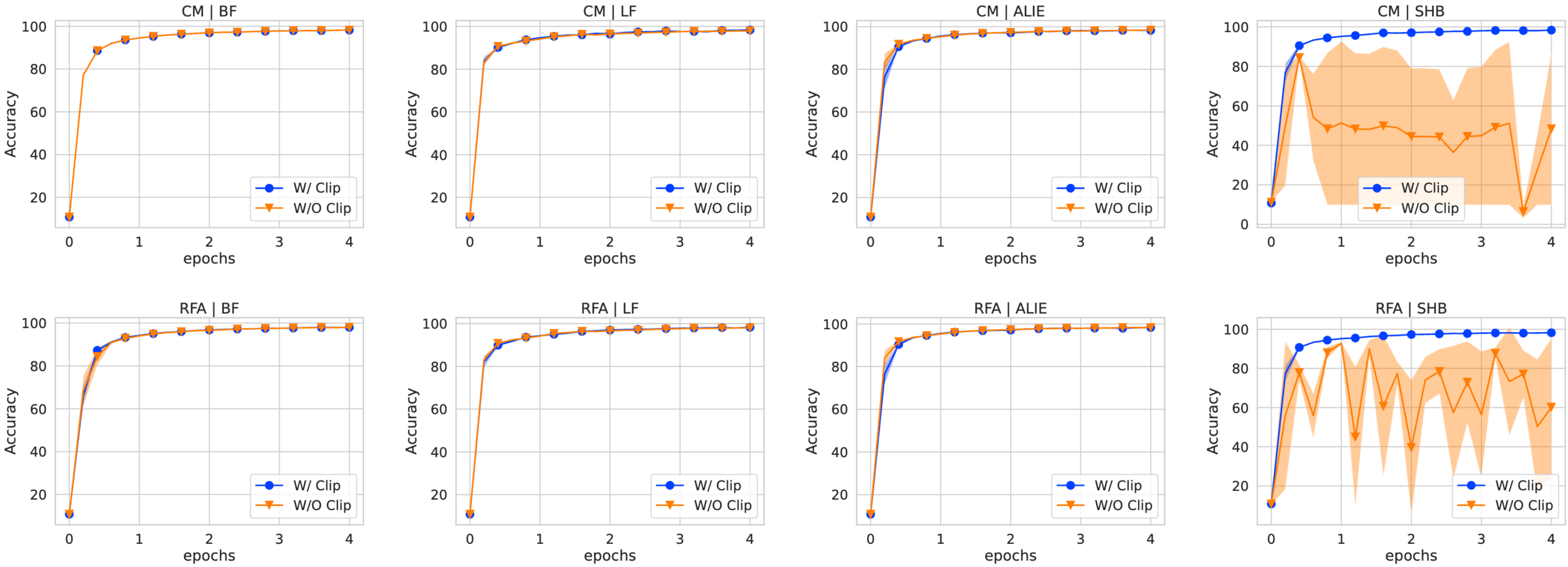
4. General framework

$$\min_{x \in \mathbb{R}^d} \left\{ f(x) := \frac{1}{G} \sum_{i \in \mathcal{G}} f_i(x) \right\}$$

Good workers form the majority:

- $\mathcal{G}$ – good workers
- $\mathcal{B}$ – Byzantines (see the page "Byzantine fault" in Wikipedia)
- $\mathcal{G} \sqcup \mathcal{B} = [n], \ |\mathcal{G}| = G, \ |\mathcal{B}| = B$
- $B \leq \delta n, \ \delta < 1/2$
- Byzantines are omniscient

Numerical Results: Neural Network Training



- Clipping does not spoil the convergence
- Clipping helps when Byzantine workers form majority (see SHB attack)