Geometric Median Based Fed REINFORCE

Ankit Pratap Singh [1] Ronak Tali [1]

[1]Iowa State University

IOWA STATE UNIVERSITY
Electrical and Computer Engineering

Background

Federated Reinforcement Learning [FRL] overcomes some of the problems associated with traditional Reinforcement learning, but most importantly it addresses the issue of Sample Efficiency, especially, in the context of Real-Time Systems. However, for FRL to be truly successfully deployed, the following two criterias need to be met:

- Efficient sampling of trajectories at each client node.
- Account for client node failures or byzantine attacks; sometimes up to 50% of the originally available nodes.

Problem Statement

- REINFORCE is a policy gradient method that learns a parameterized policy that can select actions without consulting a value function.
- For an MDP $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \rho\}$ and the actual policy π_θ, the trajectory is said drawn from the density distribution $p(\tau|\pi_\theta, M)$.
- The target is to learn θ in a multi-agent setting where each agent $\ell \in [L]$ samples trajectories $\{\tau_i\}_{i=1}^N$ using policy π_θ and calculates the gradient $\widehat{\nabla}_N J(\theta_t)^\ell$. The server receives the gradients, performs a policy update step, and then sends back the new policy parameter θ.
- One of the challenges in this setup is the presence of adversarial attacks on the agents. Byzantine attack is one such form of attack. In such attacks, the attacking agent has complete knowledge of both the incoming data and the value of the gradients of all other agents, while these agents are communicating the same to the central server.
- Such wide scale knowledge allows the byzantine agent to craftily manipulate the value of the true gradient before sending that over to the central server.
- The objective is to robustly aggregate the gradients, received from all the linked agents, at the central server and thereby learn, with a high probability, a new policy parameter θ that ultimately (over several iterations of REINFORCE algorithm) allows convergence to the correct policy parameter θ^*.

GM-FedREINFORCE

Algorithm 1 GM-FedREINFORCE

1: $\tau = \{\langle s_t, a_t \rangle\}_{t=0}^H = \{z_t\}_{t=0}^H$ is a $(H+1)-$steps trajectory which depends on the MDP $M = \{\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \rho\}$ and the actual policy π_θ, the trajectory is said drawn from the density distribution $p(\tau|\pi_\theta, M)$
2: **Initialize:** θ_0
3: **Agents** $\ell = 1, ..., L$
4: **pull** θ_k from server
5: Collect $\mathcal{D}_N = \{\tau_i\}_{i=1}^N$ using policy π_θ
6: Compute $\widehat{\nabla}_N J(\theta_k)^\ell \leftarrow \frac{1}{N}\sum_{n=1}^N \left(\sum_{h=0}^H \nabla \log \pi_{\theta_k}(z_h^n)\right)\left(\sum_{h=0}^H \gamma^h r_h^n - b(z_h^n)\right)$
7: **push** $\widehat{\nabla}_N J(\theta_k)^\ell$ to server
8: **Server**
9: **pull** $\widehat{\nabla}_N J(\theta_k)^\ell$'s from $\ell = 1, ..., L$
10: compute $\theta_{k+1} \leftarrow \theta_k + \eta$**Aggregate**$\{\widehat{\nabla}_N J(\theta_k)^\ell\}_{\ell=1}^L$
11: **push** θ_{k+1} to nodes

Main Theorem

Consider Algorithm 1 with aggregate step as geometric median and with the following step size:

$$\eta = \frac{1}{\tilde{L}}\left(1 - \frac{\epsilon(\delta_k)}{\|GM\{\widehat{\nabla}_N J(\boldsymbol{\theta}_k)^\ell\}\|}\right), \tag{1}$$

where $\epsilon(\delta_k) = C_\alpha 2WR_T\sqrt{2d\log(6/\delta_k)}/\sqrt{N}$. If $L_{byz} < L$ agents are byzantine such that $\frac{L_{byz}}{L} = \tau \in (0, \frac{1}{2})$. And each agent ℓ samples N trajectories such that:

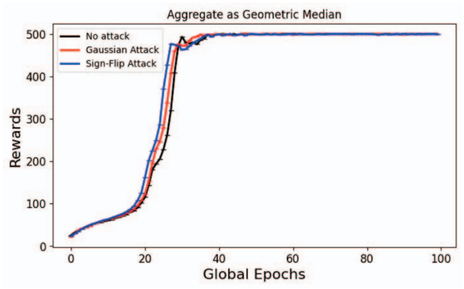
$$N \geq \frac{\left(C_\alpha 2WR_T\sqrt{2d\log(6/\delta_k)}\right)^2}{\|GM\{\widehat{\nabla}_N J(\boldsymbol{\theta}_k)^\ell\}\|^2}, \tag{2}$$

Then the performance improvement of $\boldsymbol{\theta}_{k+1}$ wrt $\boldsymbol{\theta}_k$ can be lower bounded, with probability at least $1 - \tilde{\delta}_k$, as follows

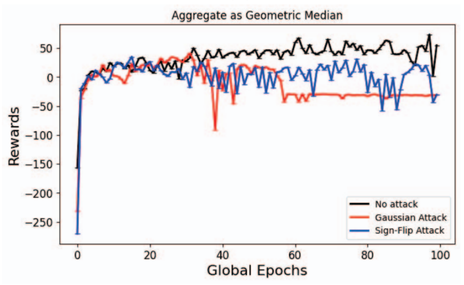
$$J(\boldsymbol{\theta}_{k+1}) - J(\boldsymbol{\theta}_k) \geq \frac{\left(\|GM\{\widehat{\nabla}_N J(\boldsymbol{\theta}_k)^\ell\}\| - \epsilon(\delta_k)\right)^2}{2\tilde{L}} \tag{3}$$

where $\tilde{L} = \frac{R}{(1-\gamma)^2}\left(\frac{2\gamma\xi_1^2}{1-\gamma} + \xi_2 + \xi_3\right)$, $\alpha \in (\tau, \frac{1}{2})$

Key Results



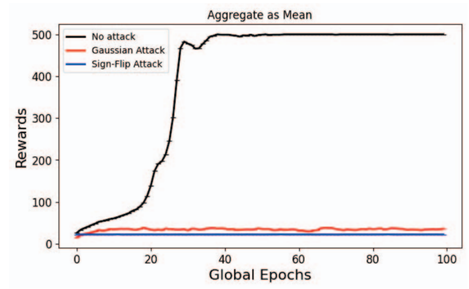
(a) CartPole



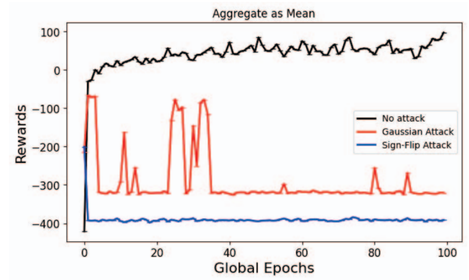
(b) CartPole Swing-up

Figure 1. Geometric Median as aggregate step with $L = 10$ and $L_{byz} = 3$

Key Results



(a) CartPole



(b) CartPole Swing-up

Figure 2. Mean as aggregate step with $L = 10$ and $L_{byz} = 3$

Our Novelty

- Our proposed approach (GM - FedREINFORCE) is the first proposed algorithm that leverages geometric median technique that provides theoretical guarantees of a robust defense against byzantine attack in federated policy gradient settings.
- We do not make any strong assumptions about the nature of byzantine attack.
- We maintain the accuracy and consistency of the learned model even in the face of significant (large proportion up to 50% of malicious agents) attack.

References

[Chen et al.(2017)Chen, Su, and Xu] Yudong Chen, Lili Su, and Jiaming Xu. Distributed statistical machine learning in adversarial settings: Byzantine gradient descent. *Proceedings of the ACM on Measurement and Analysis of Computing Systems*, 1(2):1–25, 2017.

[Fan et al.(2021)Fan, Ma, Dai, Jing, Tan, and Low] Xiaofeng Fan, Yining Ma, Zhongxiang Dai, Wei Jing, Cheston Tan, and Bryan Kian Hsiang Low. Fault-tolerant federated reinforcement learning with theoretical guarantee. *Advances in Neural Information Processing Systems*, 34: 1007–1021, 2021.