Attacks on Federated Learning

- Attack 1: Data poisoning attack [1].
- Attack 2: Model poisoning attack [2].

- Defense 1: Server check validation accuracy.
- Defense 2: Server check gradient statistics.

References

1. Shafah and others: Poison frogs! targeted clean-label poisoning attacks on neural networks. In *NIPS*, 2018.
2. Bhagoji and others: Analyzing federated learning through an adversarial lens. In *ICML*, 2019.