Appendix of "FedPerturb: Covert Poisoning Attack on Federated Learning via Partial Perturbation"

Tongsai Jin, Zhihui Fu, Dan Meng, Jun Wang, Yue Qi, Guitao Cao

•Non-IID experiment

We adopted a Dirichlet distribution with a hyper-parameter α to generate non-IID data, lower values of α correspond to higher non-IIDness.

Table 1 shows the attack performance of FedPerturb on the CIFAR10 dataset against multiple defense schemes under different non-IID degrees, which uses ResNet18 as the test model. Except for the non-IID setting, the other experimental settings are consistent with those expressed in Section 4.3 of our paper.

Experimental results show that FedPerturb in non-iid settings is still able to circumvent multiple state of the art defense schemes and cause global model divergence. For Mean, Signguard, DnC, and CC, FedPerturb causes global model divergence with a 2% malicious client ratio for all α settings. For Flame, FedPerturb's attack effect increases as α decreases.

Table 1: Accuracy of the global model under our attack with different α and different malicious client percentages, where lower values of α correspond to higher non-IIDness, **D** indicates that the test loss of the global model is greater than 1000 or becomes NaN, making the global model diverge before the end of training.

α	Ratio	Mean	Median	Multikrum	Signguard	DnC	CC	FLAME	FLTrust
0.1	0	52.98	35.91	49.18	52.53	53.51	57.29	45.26	30.22
	2%	10.00(D)	31.94	48.14	16.42(D)	10.88(D)	10.06(D)	16.04(D)	33.86
	4%	10.00(D)	36.41	44.86	10.00(D)	14.59(D)	10.00(D)	10.53(D)	13.01(D)
	10%	10.00(D)	35.03	20.28(D)	10.02(D)	10.00(D)	10.00(D)	12.78(D)	10.00(D)
	20%	10.00(D)	34.02	16.98(D)	10.00(D)	10.00(D)	10.00(D)	14.03(D)	10.00(D)
0.3	0	76.52	73.04	76.07	76.52	76.61	78.67	69.91	64.51
	2%	10.00(D)	75.50	75.90	55.74(D)	50.96(D)	10.00(D)	24.92(D)	13.89(D)
	4%	10.00(D)	71.71	77.98	10.00(D)	20.12(D)	10.00(D)	14.61(D)	10.00(D)
	10%	10.00(D)	71.77	12.60(D)	10.00(D)	14.95(D)	10.00(D)	14.60(D)	10.00(D)
	20%	10.00(D)	72.46	11.69(D)	10.00(D)	11.28(D)	10.00(D)	11.28(D)	10.00(D)
0.5	0	82.21	80.19	79.72	80.68	81.89	81.02	76.52	75.34
	2%	17.77(D)	80.27	78.51	33.60(D)	42.51(D)	14.67(D)	78.60	10.00(D)
	4%	10.00(D)	81.53	70.53(D)	10.42(D)	10.00(D)	10.00(D)	22.61(D)	10.00(D)
	10%	10.00(D)	79.58	32.97(D)	10.10(D)	10.00(D)	10.00(D)	23.97(D)	10.00(D)
	20%	10.50(D)	80.14	15.14(D)	10.00(D)	10.00(D)	10.00(D)	27.89(D)	10.00(D)
0.7	0	83.91	83.12	81.89	83.39	82.62	83.70	81.79	74.70
	2%	10.00(D)	82.58	82.00	10.00(D)	13.63(D)	10.00(D)	80.48	75.01
	4%	10.00(D)	82.22	82.34	10.00(D)	10.91(D)	10.00(D)	41.24(D)	10.06(D)
	10%	10.00(D)	83.50	34.25(D)	10.00(D)	14.90(D)	10.00(D)	36.59(D)	10.12(D)
	20%	12.42(D)	82.53	13.73(D)	10.00(D)	10.61(D)	10.10(D)	33.88(D)	10.07(D)
0.9	0	84.62	83.92	83.82	84.08	84.2	84.33	83.49	84.91
	2%	13.77(D)	84.21	83.23	18.58(D)	10.00(D)	14.63(D)	82.12	10.00(D)
	4%	10.00(D)	84.32	82.87	10.00(D)	11.52(D)	10.00(D)	83.38	10.00(D)
	10%	10.00(D)	84.29	49.75(D)	10.00(D)	17.32(D)	10.00(D)	33.18(D)	10.00(D)
	20%	10.00(D)	84.06	11.61(D)	10.00(D)	12.74(D)	12.89(D)	32.11(D)	10.00(D)