SSRDBA: A Secret Sharing-Inspired Robust Distributed Backdoor Attack to Federated Learning

Response to the reviewers' suggestion on reorganizing the results

We reorganize the figures in the submission and report the attack performance using the results in the final round. The default setting of the important hyperparamters are: global trigger = rectangle, local trigger = rectangle, poison ratio = 10%, rebel_id = 1, rebel's target labels =1. When studying the impact of a hyperparamter, we fix the other hyperparameters to be the default value.

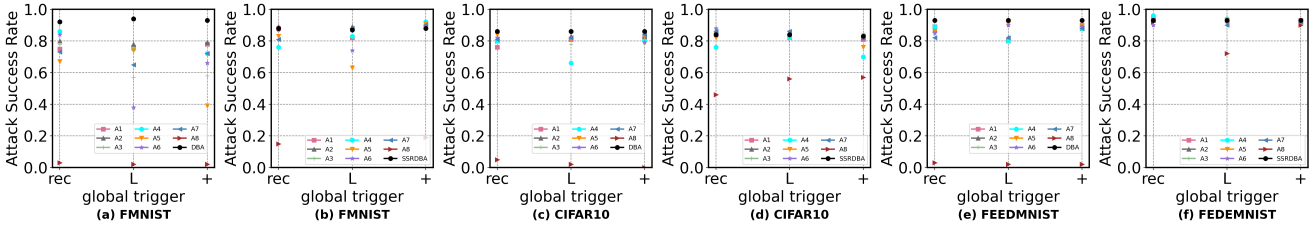

Figure 1: Impact of global trigger (rectangle, "L", and "+") on DBA vs. SSRDBA against the rebels. Figure (a)(c)(e) show the results of DBA against the rebel, while Figure(b)(d)(f) show the results of SSRDBA against the rebel on the three datasts, respectively. We can see: 1) the ASRs of r-DBA (i.e., DBA w.r.t. the rebel) are consistently and significantly lower than those of the vanilla DBA, indicating DBA is vulnerable to the rebels; 2) the ASR of SSRDBA is similar to that of DBA without rebel, meaning the backdoor performance is not affected based on our secret sharing; 3) the ASRs of r-SSRDBA are (much) higher than that of r-DBA, implying SSRDBA is a more robust attack than DBA against the rebels.

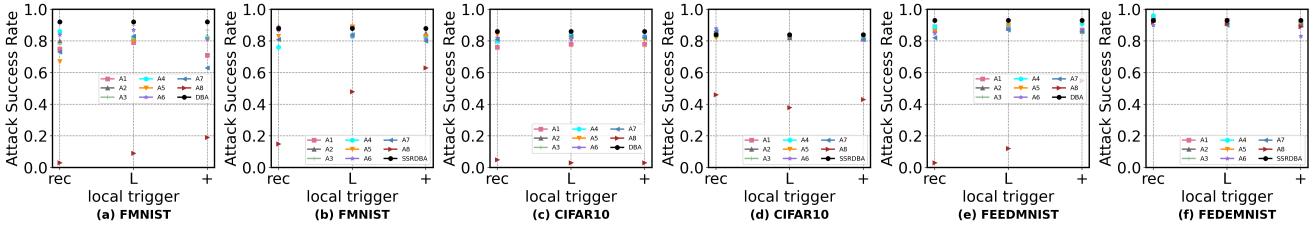

Figure 2: Impact of local trigger (rectangle, "L", and "+") on DBA vs. SSRDBA against the rebels. Figure (a)(c)(e) show the results of DBA against the rebel, while Figure(b)(d)(f) show the results of SSRDBA against the rebel on the three datasts, respectively. Similarly, DBA is vulnerable to the rebels with different local triggers, while SSRDBA is more robust against local triggers.

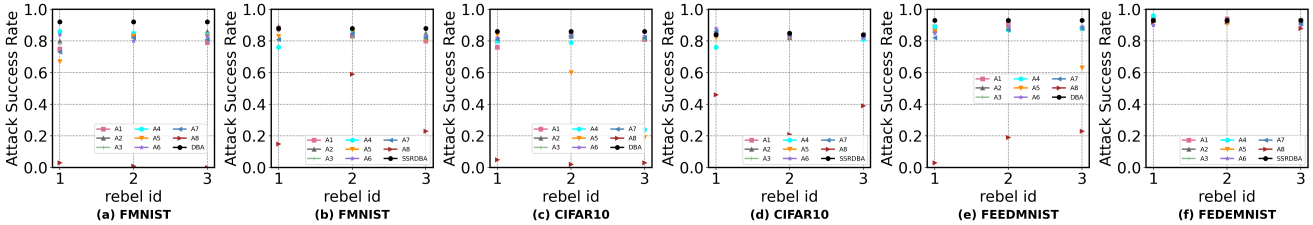

Figure 3: Impact of the rebel_id (1, 2, 3). on DBA vs. SSRDBA against the rebels. We observe that SSRDBA is insensitive to which malicious client becomes the rebel. Instead, DBA is more sensitive by observing that the ASRs are different w.r.t. a different rebel id.

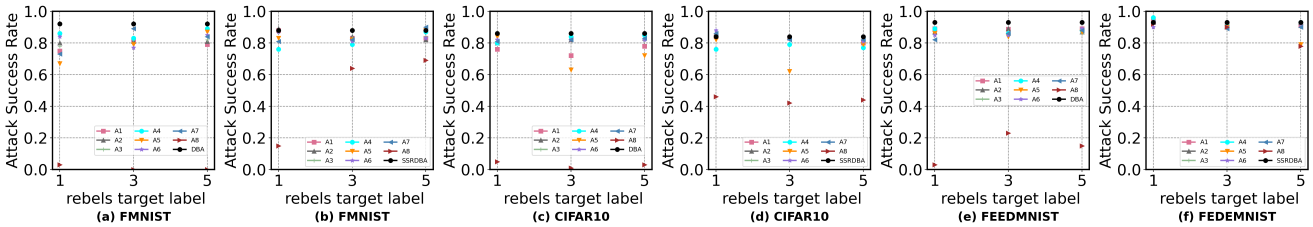

Figure 4: Impact of the rebel's target label (1,3,5) on DBA vs. SSRDBA against the rebels. We observe that SSRDBA is insensitive to which target label a rebel is chosen.

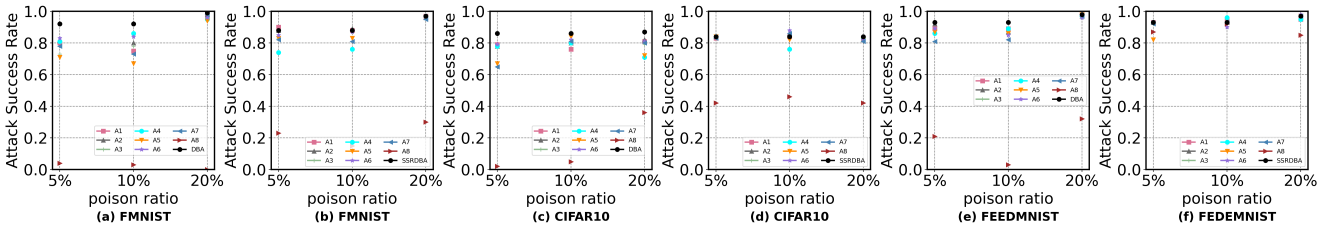

Figure 5: Impact of the poison ratio on DBA vs. SSRDBA against the rebels. We observe that a larger poison ration yields a larger ASR in general.