

QUALITY INFERENCE IN FEDERATED LEARNING WITH SECURE AGGREGATION

Balázs Pejó & Gergely Biczók – Laboratory of Cryptography and System Security, Department of Networked

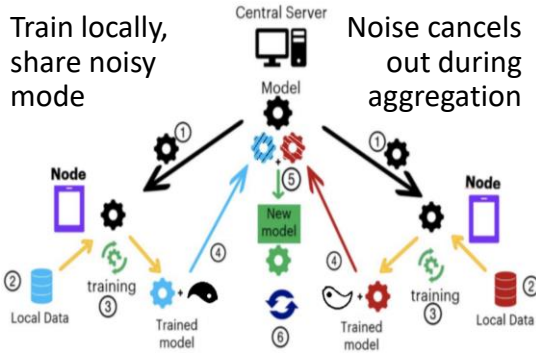
Systems and Services, Faculty of Electrical Engineering and Informatic, Budapest University of Technology and Economics

1 of March 2022, Budapest, Hungary



Federated Learning with Secure Aggregation

Train locally, share noisy mode



Membership Inference Attack + FL&SA = Attribution Attack

Training an NLP Model @



Mail Address ⇒ D & Location ⇒ B or F

Goal

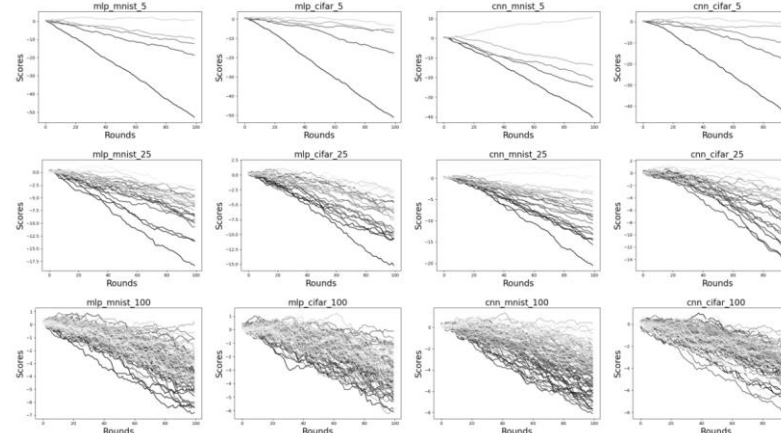
Study the possibility of inferring the quality of the individual datasets when Secure Aggregation is in place.

Quality inference is different from poisoning attack detection, as that merely interested in classifying participants as malicious or benign, while our goal is to enable the fine-grained differentiation of the honest participants with respect to input quality.

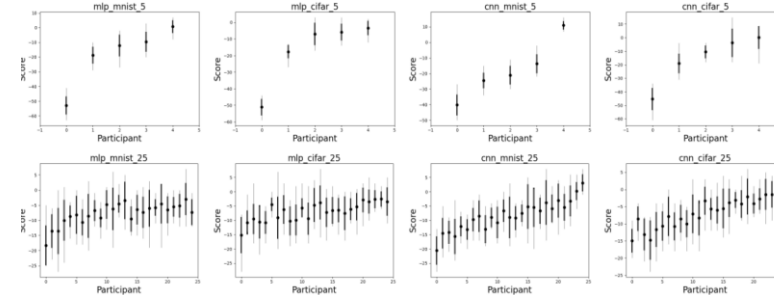
Scoring Rules

- **The Good:** each participant contributing in a round which improves the model more than the previous round receives +1.
- **The Bad:** each participant contributing in a round which improves the model less than the following round receives -1.
- **The Ugly:** each participant contributing in a round which does not improve the model receives -1.

Results (MNIST/CIFAR10-MLP/CNN)



The round-wise change of the participants' scores: the lighter the better (the darker the worse) corresponding dataset quality.



Quality scores of the participants after 50 rounds where the data quality grows with x axis.

Applications

On-the-fly performance boosting: carefully weighting the participants based on the inferred quality smooths the learning curve as well as improves the trained model's accuracy.

Misbehavior detection: the scores can be used to detect both malicious misbehavior and free-riding.

Shapley-Value Approximation: The scoring rules might be used for contribution score computation, which is currently not solved when Secure Aggregation is enabled.

Conclusion

Due to the design of federated learning, naïve secure aggregation is not safe: a few simple quality scoring rules were able to successfully recover the relative ordering of the participant's dataset qualities.



M Ű E G Y E T E M 1 7 8 2