Quality Inference in Federated Learning with Secure Aggregation

Balázs Pejó & Gergely Biczók - CrySyS Lab

Villamosmérnöki és Informatikai Kar

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.

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.

MEMBERSHIP INFERENCE
From the model updates it is possible to determine whether a particular data sample was user for training or not.

MEMBERSHIP INFERENCE FOR FL WITH SA COULD LEAD TO ATTRIBUTION
Training an NLP Model
- Mail Address ⇒ D
- Location ⇒ B or F

HEADLINE
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.