COMET

Competence Centers for
Excellent Technologies

www.ffg.at/comet

SBA Research

Federated Learning and its' security aspects

SBA Research

Overview

Federated Learning promises advances over centralized learning:

- Data remains at user
  - Alleviating many risks and obstacles related to data privacy
- Computing resources at the data holders can be utilized, thus distributing the computation

Challenges

The three main challenges that Federated Learning currently faces are:

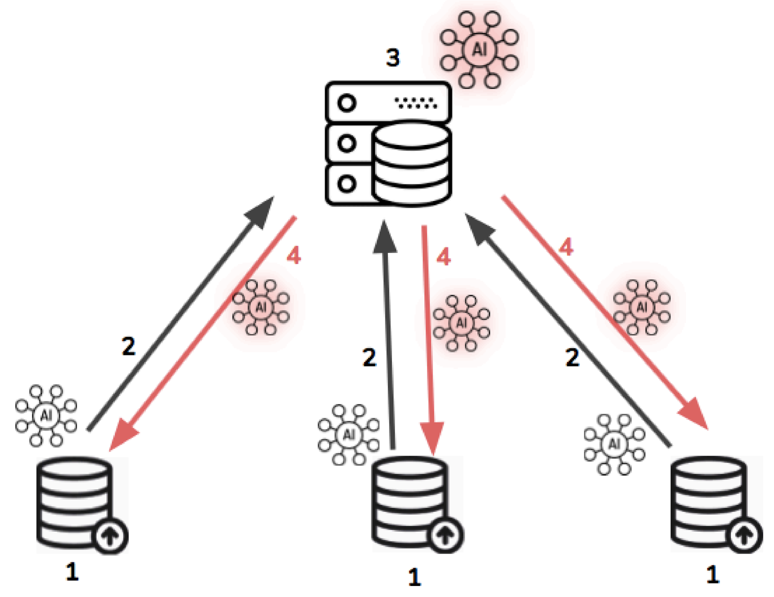
- Privacy risks: Distributed system ⟶ more and new attack surfaces and vectors like poisoning attacks
- Privacy risks: privacy attacks like membership inference attack are weaker but not completely eliminated

Federated Learning

In contrast to centralized machine learning where data is sent to an central server, no data is shared in the Federated Learning paradigm. In Federated Learning local models are trained with the data of the participants and those create the new model.

Parallel Federated Learning:

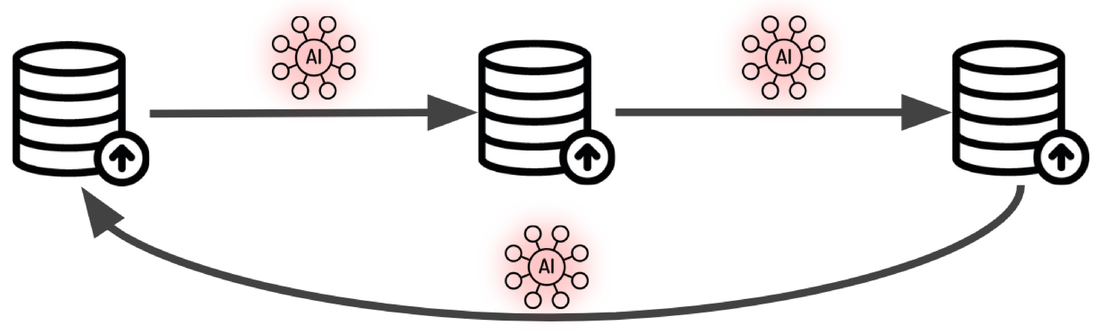
1. The aggregation server shared the global model with the clients
2. The clients train models based on their local data (1)
3. The local models are sent to the aggregation server (2) and aggregated into a new global model (e.g. by averaging the models' parameters) (3)
4. The new global model is shared with the clients (4)



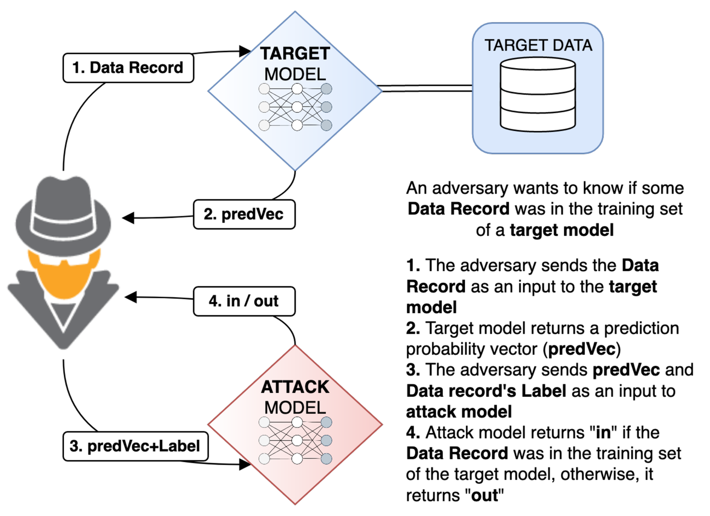
Sequential Federated Learning :

1. A client trains its model locally
2. Client sends model to the next client for further training
3. Step 1 and 2 are repeated until the last client is reached

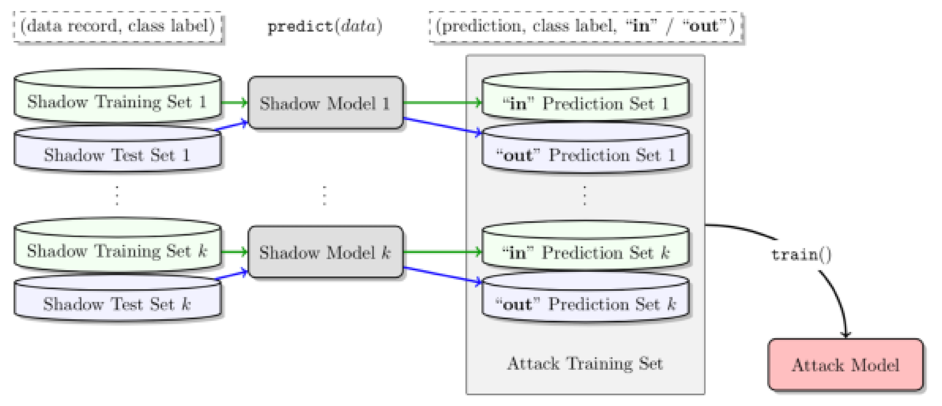
This does not require a central aggregation process.



Membership inference attack Attacks



If an adversary wants to determine if a data record was part of the training set of the target model, the attacker can use the membership inference attack. In order to be able to infer that knowledge the attacker needs to build an attack model which can be created from shadow models. These shadow models have the same structure as the target model, however, the adversary has to build his own shadow training and test data as he does not possess knowledge about the actual data.



Poisoning Attacks

Backdoor attacks are an attack targeting the model's integrity during the training phase. According to this strategy, an adversary poisons the training data by adding samples containing a certain pattern (the so-called "backdoor"). The goal is to trigger malicious behavior on data containing this pattern during the deployment phase.

(a) Original image (Traffic Sign dataset)   *(b)* Backdoor pattern „black square"

Benefits

Benefits of federated learning include:

- Data security: Keeps ownership of data ⟶ increased incentive to participate in collaborative learning ⟶More data and more data diversity
- Hardware efficiency: Uses computational power at edge

Conclusions

- Federated Learning improves privacy and security
- There are still challenges like poisoning and membership inference attacks
- Defenses against these attacks for centralized learning settings have to be modified for FL

E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov. How To Backdoor Federated Learning. In *23rd International Conference on Artificial Intelligence and Statistics (AISTATS)*, Palermo, Italy, June 2020.
F. Nuding and R. Mayer. Poisoning Attacks in Federated Learning – an Evaluation on Traffic Sign Classification. In *10th ACM Conference on Data and Application Security and Privacy (CODASPY)*, New Orleans, U.S., March 2020.

Bundesministerium Klimaschutz, Umwelt, Energie, Mobilität, Innovation und Technologie · Bundesministerium Digitalisierung und Wirtschaftsstandort · FFG Forschung wirkt. · wirtschafts agentur wien Ein Fonds der Stadt Wien · Europäische Kommission · FWF Der Wissenschaftsfonds. · netidee OPEN INNOVATIONS

SBA Research (SBA-K1) is a COMET Centre within the framework of COMET – Competence Centers for Excellent Technologies Programme and funded by BMK, BMDW, and the federal state of Vienna. The COMET Programme is managed by FFG.