Security Challenges for Federated Learning Systems: a Complex Attack Surface

Vuillod Bastien*[1], Pierre-Alain Moëllic*†[1], and Jean-Max Dutertre*‡[2]

[1]Commissariat à l'énergie atomique et aux énergies alternatives - Laboratoire d'Electronique et de Technologie de l'Information – Direction de Recherche Technologique (CEA) – France
[2]École des Mines de Saint-Étienne – Institut Mines-Télécom [Paris] – France

Résumé

Federated Learning (FL) is a decentralized learning paradigm consisting in training local models, usually on edge devices, that are sent to a central server to create a global model. The main advantages of FL are to not transmit potentially private local training data to the central server as well as gather different actors in a common learning goal. However, FL also comes with serious security issues regarding both the confidentiality of the data and the integrity of the system, typically with poisoning based threats. In this paper, we present the wide and complex attack surface of FL with a large panel of possibilities an adversary can exploit to break the confidentiality or integrity of these distributed systems. Since FL systems usually rely on edge devices that can be targeted by physical attacks, we also discuss advanced implementation-based threats.

*Intervenant
†Auteur correspondant: pierre-alain.moellic@cea.fr
‡Auteur correspondant: dutertre@emse.fr