Titre : Decentralized Federated Learning for Edge Intelligence

Supervisors :

Van-Tam Nguyen (van-tam.nguyen@telecom-paris.fr)

Summary:

In this PhD Track project, we will investigate solutions to improve cybersecurity mechanisms based on participant and application scenarios in decentralized federated learning. We will explore architectures that treat participants differently based on their privacy restrictions, allowing for a more personalized and secure approach.

Description:

AI/ML benefits from the vast amount of data. However, data from a massive number of IoT devices are usually stored in a distributed manner. In some use cases, such as remote medical monitoring or connected vehicles, data collection in central entities, as is traditionally the case in MLs, is often unfeasible or impractical due to limited communication resources, data privacy concerns or national regulations. Federated Learning (FL) solves this problem by enabling data owners to collaborate on AI model training without exposing private data. FL deploys collaborative AI model training on end-user devices where the data is stored, removing the communication overhead and privacy issues associated with transmitting raw data. It therefore moves the learning capability from a central ML engine to the data sources, enabling a large number of end-user devices to participate in collaborative learning. However, FL requires that sufficient computing and communication capacity be deployed close to end-user devices.

Edge computing deploys cloud computing capabilities at the edge of the network. It has been widely integrated into 5G networks and is expected to continue to play a crucial role in the 6G network as a key component of various applications. It is an efficient approach to providing the computing resources required by the LF close to data sources, enabling the FL to take full advantage of distributed computing and networking capabilities between the various devices in the network.

Federated learning can be divided into two categories, depending on how federated models are created: centralized federated learning (CFL) and decentralized federated learning (DFL). CFL considers that a central orchestration server creates and distributes a global model to the rest of the participants. More precisely, participants train their models with local data, then send local model parameters to a central server, where a global model is created by aggregating and combining individual model parameters. DFL distributes the aggregation of model parameters between neighboring participants. DFL's operation focuses on transmitting fast updates calculated locally by each node (e.g. model parameters or gradients) and metadata (e.g. activation functions in neural networks) to the rest of the nodes in the federation. Compared with CFL, therefore, DFL improves on single-point-of-failure limitations, trust dependencies and bottlenecks at the server node. In addition, DFL enables more efficient use of resources by distributing the computing power required to aggregate model parameters among all participating nodes, rather than relying on a single parameter server under centralized control.

Despite the benefits offered by DFL, it also introduces new challenges in areas such as communication overhead, training optimization, trustworthy AI, and security. In this project, we will focus on improving cybersecurity mechanisms based on participant and application scenario. The security of the nodes participating in DFL and their ability to detect and prevent attacks or threats in heterogeneous scenarios are crucial for the success of this approach. We will explore architectures that treat participants differently based on their privacy restrictions, enabling a more personalized and secure approach. This is particularly relevant for application scenarios such as healthcare and connected vehicles, where protecting user privacy is essential.