Alternative approaches for privacy-preserving in federated learning

Contract type: Fixed-term contract
Level of qualifications required: Graduate degree or equivalent
Function: PhD Position

About the research centre or Inria department

Inria is a national research institute dedicated to digital sciences that promotes scientific excellence and transfer. Inria employs 2,400 collaborators organised in research project teams, usually in collaboration with its academic partners. This agility allows its scientists, from the best universities in the world, to meet the challenges of computer science and mathematics, either through multidisciplinarity or with industrial partners.

A precursor to the creation of Deep Tech companies, Inria has also supported the creation of more than 150 start-ups from its research teams. Inria effectively faces the challenges of the digital transformation of science, society and the economy.

Assignment

Assignment: Federated learning (FL) enables a large number of IoT devices (mobiles, sensors) to cooperatively learn a global machine learning model while keeping the devices’ data locally ([MN+]17, [LS20]). For example, Google has applied FL in their application Gboard to predict the next word the users would type on their smartphones ([HM+18]). FL can help to mitigate privacy concerns, as the raw data is kept locally by the users and never needs to be sent elsewhere. However, maintaining the data locally does not provide itself formal privacy guarantees.

Many attacks have shown the vulnerability of federated learning systems: the adversary can reconstruct private data points (e.g., images and private features) ([LH19, GBM20, DXN+22]), infer the membership of the data instance ([MDCST19, ZHL17]) and reconstruct the local model of the user ([NX20]) just by eavesdropping the exchanged messages. As a result, differentially private (DP) algorithms ([HRZ18, BGTT18]) have been proposed for FL to protect privacy by injecting random noise into the transmitted messages. DP ensures that if the user changes one training sample, the adversary does not observe much difference in the exchanged messages and then may not confidently draw any conclusions about the presence or absence of a specific data sample. Therefore, attacks are less efficient ([B19]). However, the noise typically deteriorates the performance of the model.

Alternatively, some methods that were initially designed to improve model generalization have been empirically shown to be effective against privacy attacks as well, as the resulting model memorizes less the training samples. For example, in the centralized training scenario, pruning the neural network ([PT11]) can mitigate the privacy leakage from membership inference ([WWH+21] and model inversion ([HRZ+20]) attacks. Mixing up training data samples ([CDL18]) may also help to defend against adversarial attacks ([PKX20]). Besides, methods which exploit other sources of randomness, like batch sampling ([HT19]) and mixing up the average weights in decentralized learning ([XZD21]), can amplify the DP guarantees. However, how to adapt and combine these techniques in federated system where the devices may exhibit different computation/memory capacities and data distributions, as well as have different privacy requirements, is still an open problem.

Research Goal

The goal of this PhD is to propose new privacy-preserving methods for FL which do not necessarily add synthetic noise to updates (as DP does), but rather rely on different approaches, like parameter pruning or quantization, or exploit other sources of randomness, like the batch sampling procedure or the “mixup” procedure. Pruning and quantization can lead to models with better generalization that are in turn less vulnerable to attacks, but also reduce the computation/communication requirements of FL training and then potentially its energy consumption. Because of devices’ heterogeneity (in terms of computation, memory, data, and privacy requirements), we need to design device-aware strategies, e.g., to select the level of model compression through pruning or quantization. The candidate will also study how these alternative techniques may be combined with more traditional DP approaches, potentially leading to improved accuracy-utility trade-offs against FL privacy attacks.

Reference:


Main activities

Research

Skills

The candidate should have good programming skills and previous experience with PyTorch or TensorFlow. He/She should also be knowledgeable on machine learning and have good analytical skills. We expect the candidate to be fluent in English.

Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

Remuneration

Gross Salary per month: 2051€ brut per month (year 1 & 2) and 2158€ brut per month (year 3)

positions are accessible to people with disabilities.