Title

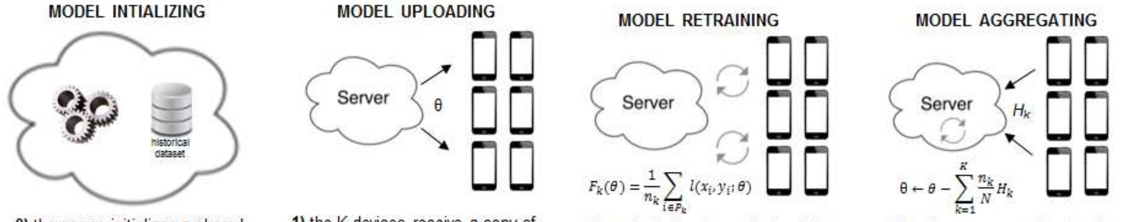
Post-doc : Development of algorithms for decentralized, resilient federated learning.

Location and unit: CEA LIST / SID

Located in Saclay, south of Paris, CEA LIST (http://www-list.cea.fr/) is a scientific and technological research center dedicated to the development of software, embedded systems and sensors for applications such as defense, security, energy, nuclear power, the environment and health. CEA LIST is part of the dynamic and stimulating ecosystem of the University of Paris-Saclay - the largest French scientific center with 60,000 students. It has more than 700 researchers focusing on intelligent digital systems, centered around artificial intelligence, the factory of the future, innovative instrumentation, cyber-physical systems and digital health. Within this institute, the SID (Data Intelligence Service) works on algorithms and methodologies for artificial intelligence and signal processing. The laboratory's research and technological advances are guided by a variety of applications, for which the specificities and constraints on the data or the execution environment require a fine design of AIs and their integration as the unitary bricks of complex systems.

Project description

The postdoctoral fellow will join the Carnot FANTASTYC project, which puts together researchers on distributed ledger technology, privacy and machine learning with the aim of developing software assets for decentralized, privacy-preserving and resilient federated learning.


MODEL INTIALIZING
historical dataset
MODEL UPLOADING
Server
θ
MODEL RETRAINING
Server
$F_k(\theta) = \frac{1}{n_k} \sum_{i \in P_k} l(x_i, y_i; \theta)$
MODEL AGGREGATING
Server
H_k
$\theta \leftarrow \theta - \sum_{k=1}^{K} \frac{n_k}{N} H_k$

0) the server initializes a shared model (randomly or using historical data)

1) the K devices receive a copy of the shared model parameters θ

2) each device k uses its local data to compute at edge a model update *H_k minimizing a cost function F_k*

3) updates are sent back to the server where they are averaged to improve the shared model

Fig-1 : Federated learning main steps

Federated learning (FL) is a machine learning setting (see Fig. 1) in which many clients (e.g. mobile devices or whole organizations) collaboratively train a model under the orchestration of a central server (e.g. service provider), while keeping the training data decentralized (communicating only the model parameters) [1]. Hence, in traditional FL, a central server orchestrates the training process and receives the contributions of all clients and thus represents a single point of failure and/or a communication bottleneck. Against this background, the first objective of this fellowship is to envisage a fully decentralized efficient version of the FL, replacing communication with the server by peer-to-peer communication between individual clients on some communication graph. Note that in this peer-to-peer setting there is no longer a global state of the model, but the process can be designed such that all local models converge to the desired global solution, i.e., the individual models gradually reach consensus. On doing that, the successful applicant is expected to tackle some of the open challenges that involve passing to fully decentralized learning, including: (1) the design, specification and implementation of efficient decentralised learning protocols; (2) the evaluation of communication and computational costs of protocols on different network topologies possibly leading to the design of new resource-aware distributed learning protocols; and (3) tackling the compromise between generic and personalized models depending on the evaluated non-IID of data distributions available to individual clients (e.g. different models for clusters of participants). For the design and implementation of the distributed framework the post-doc is expected to

collaborate with other CEA labs involved in the project that will be providing a privacy-preserving distributed ledger technology infrastructure. The other focus of this position will be the study of the robustness of distributed FL against the presence of malicious participants (i.e. Byzantine and backdoor attacks) [2,3].
The application domain envisaged in the project is personalized privacy-preserving health monitoring.

[1] Google AI blog: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[2] Blanco-Justicia, A., Domingo-Ferrer, J., Martínez, S., Sánchez, D., Flanagan, A., & Tan, K. E. (2020). Achieving Security and Privacy in Federated Learning Systems: Survey, Research Challenges and Future Directions. *arXiv preprint arXiv:2012.06810*.

[3] Peva Blanchard, El Mahdi El Mhamdi, Rachid Guerraoui, Julien Stainer: Machine Learning with Adversaries: Byzantine Tolerant Gradient Descent. NIPS 2017: 119-129

Qualifications

1. PhD in machine learning from an accredited university
2. Excellent communication skills, both verbal and written in English
3. Communication skills in French
4. Experience in federated learning is a plus
5. Experience in distributed systems and/or robust attacks is a plus

Salary & benefits

The salary will depend on the applicant’s profile and experience. The position comes with various social benefits (e.g. health insurance). This position is open for one year, renewable once.

Application and contacts

To apply send an updated CV and a motivation letter to :

Aurélien Mayoue (aurelien.mayoue@cea.fr)

Cédric Gouy-Pailler (cedric.gouy-pailler@cea.fr)

Stéphane Gazut (stephane.gazut@cea.fr)

The position is open immediately (September 2021). Review of applications will begin as soon as applications are received and continue until the position is filled.