3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic: Adversarial Learning and Differential Privacy
- Research axis of the 3IA: Axis 4
- Supervisor (name, affiliation, email): Melek Önen - EURECOM, Sophia Antipolis - melek.onen@eurecom.fr
- Potential co-supervisor: None
- The laboratory and/or research group: EURECOM, Sophia Antipolis

---

- Description of the topic:

The rapid integration of pervasive technologies—such as cloud computing, Internet of Things, and 5G–in all aspects of modern society result in the creation of a plethora of valuable but often sensitive (personal and/or industrial/proprietary). More and more companies are nowadays collecting huge amounts of data from a variety of sources and use machine learning tools to acquire meaningful insights and make value out of them. Among these tools, federated learning [1] emerged recently to address the communication overhead issues associated to the training of the machine learning (ML) model. The basic setting of such tools consists of multiple parties coordinating with an \textit{aggregator} whose goal is to compute a global machine learning model based on parties' inputs without leaking any information on individual parties' private data beyond the global model's parameters.

While federated learning is considered to partially address the privacy protection of the data, it is unfortunately, still exposed to various security threats such as backdoor injection attacks [2]whereby the adversary intervening the training phase tries to perturb the global model only for certain inputs with specific characteristics.

The goal of this project is to study the interplay between federated learning, adversarial learning and privacy enhancing technologies. We consider that some differentially private mechanism is used over parties data and that adversary who is able to modify (add, replace, delete, etc.) the published information. This adversary's aim is to maximize the possible damage while remaining undetected. Hence, while initially considered as a privacy enhancing technology, a differentially private mechanism can be used by the adversary as a tool to help him/her be undetected.

An initial study on this aspect was proposed in [3] from an adversarial perspective the two conflicting goals of the adversary is formulated as an optimization problem where maximizing the bias induced by the adversary is the objective function. Nevertheless, the privacy parameter is not considered as a variable in their formulation. Another line of research in [4]

identifies a connection between differential privacy and robustness against adversarial examples.

Research plan

The PhD student will first start by investigating potential attacks against federated learning and their impact on the performance and accuracy of the actual federated learning protocol. The PhD candidate will further investigate the trade-off between by investigating the trade-off between the attack and the privacy parameter and identify a threshold for which the adversary remains detected even under a certain privacy level. The ultimate goal is to develop new privacy preserving federated machine learning solutions that are additionally robust against attacks including those originating from generative adversary networks.

References

[1]B. MacMahan, E. Moore, D. Ramage, S. Hampson, B. Aguera y Arcas, *Communication-efficient learning of deep networks from decentralized data,* Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) 2017.

[2] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, V. Shmatikov, *How to backdoor federated learning?*, Proceedings of the 23rd International Conference on Artificial Intelligence and Statistics (PMLR), 2020.

[3] J. Giraldo, A. A. Cardenas, M. Kantarcioglu, J. Katz. *Adversarial Classification under Differential Privacy,* Network and Distributed Systems Security Symposium (NDSS), 2020.

[4] M. Lescuyer, V. Atlidakis, R. Geambasu, D. Hsu, S. Jana, *Certified Robustness to Adversarial Examples with Differential Privacy,* IEEE Symposium on Security and Privacy (S&P), 2019.