Tutorial: Federated Learning x Security in Network Managements

Yann Busnel, Léo Lavaur

Federated Learning × Security in Network Managements

Yann Busnel Leo Lavaur

Tutorial duration: 3.5 hours

Abstract

Federated learning (FL) is a machine learning (ML) paradigm that enables distributed agents to learn collaborative models without sharing data. In the context of network security, FL promises to improve the detection and mitigation of attacks, notably by virtually extending the local dataset of each participant. However, one of the major challenges of this recent technology is the heterogeneity of the data used by the participants. Indeed, some participants with very different monitoring contexts could penalize the global model. Furthermore, identifying malicious contributions is made more difficult in heterogeneous environments.

In this tutorial, we will first present the fundamentals of federated learning, then focus on its use in network monitoring, and more specifically, in collaborative intrusion detection (Federated Learning-based Intrusion Detection System—FIDS). Secondly, we will address some of the open research questions in this context [LPBA22], before focusing on the problem of training data heterogeneity. Finally, we will discuss the security of FL architectures, and more specifically, the problem of poisoning attacks. All these parts will be illustrated by hands-on exercises, guided step by step throughout the tutorial.

1 Speakers

Yann Busnel

Full professor, Dean of Research and Innovation, at IMT Nord Europe, Lille, France

Bio. Yann Busnel has joined IMT Nord Europe as Dean of Research and Innovation from June 2023. After more than 15 years of experience as a faculty, including 6 years as a full professor at IMT Atlantique and member of the IRISA laboratory, he now oversees all research and innovation activities in line with Institut Mines-Télécom's strategy. He contributes closely to the definition of strategic orientations and to the operational management of the institution. He also represents the Executive Board internally and externally with its partners in the research environment.

He holds an Habilitation to Supervise Research and a PhD respectively from the École Normale Supérieure de Rennes and the University of Rennes. After starting his career in Italy (La Sapienza Universita di Roma), he worked at the University of Nantes, then at ENSAI in Rennes, before joining IMT. As a professor at IMT, he was previously head of the Rennes campus of IMT Atlantique, in charge of education and research purposes.

His research topics are mainly related to Models for large-scale distributed systems and networks, with application in Data stream analysis, Cybersecurity, Massive health data and Artificial Intelligence. Recently, his areas of application range from (i) cybersecurity and dependability to (3) the analysis of medical data, in the context of pharmacovigilance or genomic sequence analysis, and (3) the self-organized coordination of fleets of drones. He is co-head of the national network of research on Distributed Systems and Networks (GDR RSD and GDR Security) . He has published over 100 articles in peer-reviewed journals and conferences. He has also coordinated several national and international collaborative research projects.

Leo Lavaur

PhD Candidate, at IMT Atlantique, Rennes, France

Bio. Léo Lavaur (Graduate Student Member, IEEE) received the engineering degree in information security from the National Engineering School, South Britany (ENSIBS), Vannes, France, in 2020. He is currently pursuing the Ph.D. degree in cybersecurity with the Engineering School, IMT Atlantique and the Chair on Cybersecurity in Critical Networked Infrastructures (Cyber CNI), Rennes, France. During his studies, he also worked in industry with Orange Cyberdefense as a part-time Employee for three years, where he worked on application security, and Wi-Fi rogue access-point detection and location.

He now studies the collaboration in security systems, and how to share data without compromising security. His current research focuses on the challenges of using federated learning as a framework for collaborative intrusion detection systems. In particular, he works on the detections of malicious contributions in heterogeneous environments, as well as on the creation of datasets for evaluating FIDS in heterogeneous settings.

2 Audience and background requirement

This tutorial is open to anyone (MSc to Faculty) with a basic knowledge of machine learning (particularly neural networks) and Python programming.

3 Tutorial content

This tutorial is structured as an alternation between lectures and practical exercises. Lectures will take about 20 to 30 minutes each, with the exercises filling the rest. This program is divided in three parts:

(i) fundamentals of FL,

(ii) FL for collaborative security,

(iii) security of FL architectures.

Fundamentals of FL. First, we will introduce the audience to FL and some of its applications, from general framework to the context of network management. In the practical part, learners will be introduced to Flower, an open-source FL framework, and to existing datasets for network security. The objective is to lay the foundations for the rest of the tutorial, and to make sure that everyone has the necessary knowledge to follow.

FL for collaborative security. In this part, we will focus on the use of FL in the context of network security, and more specifically, in collaborative intrusion detection (Federated Learning-based Intrusion Detection System—FIDS). The lecture will give an overview of the challenges of collaborative security in FL, with a focus on the heterogeneity between clients. In the practical exercise, we will implement common models on standard IDS datasets and observe these challenge through small experiments.

Security in collaborative FL. Finally, the last lecture will address some challenges of running FIDS in terms of security. Depending on the exposure of the federation (open or closed entrance, trustworthiness of the participants, *etc.*), such collaborative architectures can be vulnerable to various types of attacks. The lecture will especially focus on the problem of data poisoning, and how it can be mitigated. In particular, we will talk about the difficulty of detecting such attacks in heterogeneous environments, with leads from the literature, and mention some of our current research on the topic. The practical exercise will be a hands-on experiment on data poisoning, where attendees will be able to observe the impact of such attacks on the models, and how existing works try to mitigate them.

4 Materials and requirements

4.1 Hand-out

The following material will be provided to the attendees *after* the tutorial:

- the slides decks of the lectures (three parts),
- commented Jupyter notebooks (also three) for the different practical exercises, with the necessary code and data.

4.2 Hands-on

For the practical exercises, attendees will need to have a working Python environment, with a recent version of Python and a set of common libraries (NumPy, Pandas, Matplotlib, TensorFlow, Flower, ...). Actual dependencies will be provided before the conference. However, attendees must be warned beforehand that issues with their environment will not be addressed during the tutorial.

Alternatively, the Jupyter notebooks will be available on Google Colab, so that attendees can run them without installing anything on their machine.

Depending on NoF's organizing committee, access to a locally hosted JupyterHub should be considered to avoid depending on the quality of the Internet connection (see infrastructure requirements below).

4.3 Infrastructure requirements

Multiple options are possible for the practical exercises:

a) either the attendees will run the notebooks on their own machine, in which case they will need a working Python environment (see above),

b) either the notebooks will be hosted on Google Colab, in which case attendees will need a Google account and a web browser; this option is more convenient for the attendees, but requires a good and stable Internet connection,

c) or the notebooks will be locally hosted on a JupyterHub (or alike) instance, in which case attendees will also need a web browser; this options requires an instance that can accommodate all attendees runing ML tasks, ideally with GPU support.

References

[LPBA22] Léo Lavaur, Marc-Oliver Pahl, Yann Busnel, and Fabien Autrel. The evolution of federated learning-based intrusion detection and mitigation: A survey. *IEEE Transactions on Network and Service Management*, 19(3):2309–2332, 2022.