Module 8: Security Issues in Data Pipelines and the Cloud

Overview

In Module 8 we discuss security factors related to data pipelines and to the Cloud. We describe different tasks of the pipeline that can be attacked focusing on poisoning data, algorithms, and the project's outcomes. We further introduce the concept of multi-tenancy in the cloud and the possible resulting threats. Finally, we discuss common best practices to secure the data pipelines.

Objectives

Upon completion of this module, students will be able to:

1. Describe security factors related to data pipelines and to the Cloud
2. Identify the result of attacks in different data and tasks of a pipeline
3. Describe the concept of multi-tenancy in the cloud
4. Identify the new threats posed by multi-tenancy
5. Apply best practices to secure the data pipelines

Readings (2 hours)

- **Security Threats and Defensive Approaches in Machine Learning System Under Big Data Environment, C. Hongsong, et. al., Wireless Personal Communications, 117:3505–3525, 2021 (https://canvas.vt.edu/courses/176740/files/28995288?wrap=1)**
- **Cloud and Big Data Security System's Review Principles: A Decisive Investigation, K. Mishra, et. al., Wireless Personal Communications, June 2022 (https://canvas.vt.edu/courses/176740/files/28995289?wrap=1)**

Watch (1 hour)

- **Lecture 8 Video (https://canvas.vt.edu/media_objects_iframe/m-6dEZtC6NgWYputTdWbKHCbKjRv1oYVL9?type=video?type=video)**

- **Lecture 8 Slides (https://canvas.vt.edu/courses/176740/files/28995315?wrap=1)**

Assignment

- No Lab and no Assignment. Please work toward your Project.

Recitation (1 hour)

The recording of the optional synchronous ZOOM session for this lecture will be linked here.