Internship project description
Running untrusted code in federated learning

Jan Ramon

04/2024

1 Motivation and context

Over the last decades, there has been an increasing interest in exploiting data. On the other hand, recently there has also been an increasing awareness of the risks of collecting sensitive data centrally, given the frequency of data leaks, hacking or abuse. INRIA's Magnet team is interested in decentralized privacy-preserving machine learning where the sensitive data remains with the data owners, and machine learning is performed collaboratively by these data owners by participating in collaborative algorithms which (through the use of differential privacy and/or encryption) generate the desired statistical models but prevents sensitive data from being revealed. Important ongoing research projects in the team include the TIP, TRUMPET and FLUTE projects. This internship fits into the larger research program including these projects.

In the Horizon Europe projects TRUMPET and FLUTE we will develop a platform for privacy-preserving federated learning on medical data. With such platform, multiple hospitals owning patient data will collaborate on their joint data to learn together a statistical model (which is more accurate than what they can learn with their own, smaller dataset), however they will do so without revealing any of their data. The medical researcher hence will not see the data directly but can interact with it by asking queries. As the answer to queries can reveal sensitive information, we will protect the answer using differential privacy (or later similar notions). A group of medical researchers have now listed what kind of queries they want to ask.

2 Problem

As the data does not travel to the researcher, the researcher needs to send its queries to the data owner to execute them on the private data and return only the query results (aggregated over all data owners using multi-party computation or similar techniques). The vast majority of libraries of machine learning algorithms nowadays is written in Python. However, naively sending python code is insecure as it could contain malicious code fragments which are non-trivial to spot due to the many ways in which Python allows overloading fundamental class operations.

3 Objectives

The objective of this project is to develop a “domain-specific language (DSL)”, i.e., a subset of python which is so restricted that it is safe to execute by a data owner but at the same time sufficiently rich to satisfy most machine learning needs. The data owner can then reject any query sent by an untrusted data user which can not be verified to be safe.

4 Plan

Here is a tentative work plan:

- Getting familiar with Python dunder methods and other overloadable class properties, and with meta-interpreter basics (2 weeks)
- Listing machine learning requirements and privacy requirements (1 week)
- Designing a secure solution (3 weeks)
- Implementing a first prototype containing basic functionality (4 weeks)
- Integrating in a machine learning use case and the surrounding knowledge discovery cycle (4 weeks)
- Performing experiments and extending the DSL as needed by the use case (4 weeks)
- Optionally, providing a formal security proof (6 weeks)
- Completion of the internship report (2 weeks)

The timing (here shown on a scale of 26 weeks) can be adapted according to the personal preferences of the student or the requirements of his school or type of project. The optional item can be attempted by strong students or in longer projects.

5 Environment

The project will be conducted in the INRIA MAGNET team. The student will collaborate and interact with various other collaborators in the team. It is possible the student will be asked to present progress to the team or to partners in ongoing projects. An application for ZRR access will need to be made to the FSD.

6 Requirements

We expect the student has a strong knowledge of basic computer science concepts, e.g., data structures and algorithms, and of statistics / machine learning. Knowledge of Python is required.