JADS Jheronimus Academy of Data Science

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

DSCE DATA SCIENCE CENTER EINDHOVEN

Multi-Owner Data Science Meets Privacy!

Running Machine Learning Algorithms on top of Multiparty Computation

The Problem

We consider a setting where multiple data owners want to collaborate by learning from their joint data set. How can those data owners be certain that their data is handled prudently (w.r.t data privacy) by all collaborators, and is only used for achieving the specific learning goal?

Secure Multiparty Computation

Secure Multiparty Computation (MPC) is a well-known technique from the field of cryptography. In MPC, a number of parties can jointly perform a computation on data (provided by some or all the parties), such that *each party learns nothing beyond what can be deduced from the output of the computation*.

Algorithmic Challenges

The “costs” in terms of computational and communication complexity of arithmetic operations (addition, multiplication, comparison, …) in MPC are very different from ordinary (non-secure) computation, hence various design choices behind a particular algorithm might need to be revisited when the algorithm is to be run as a multiparty computation.

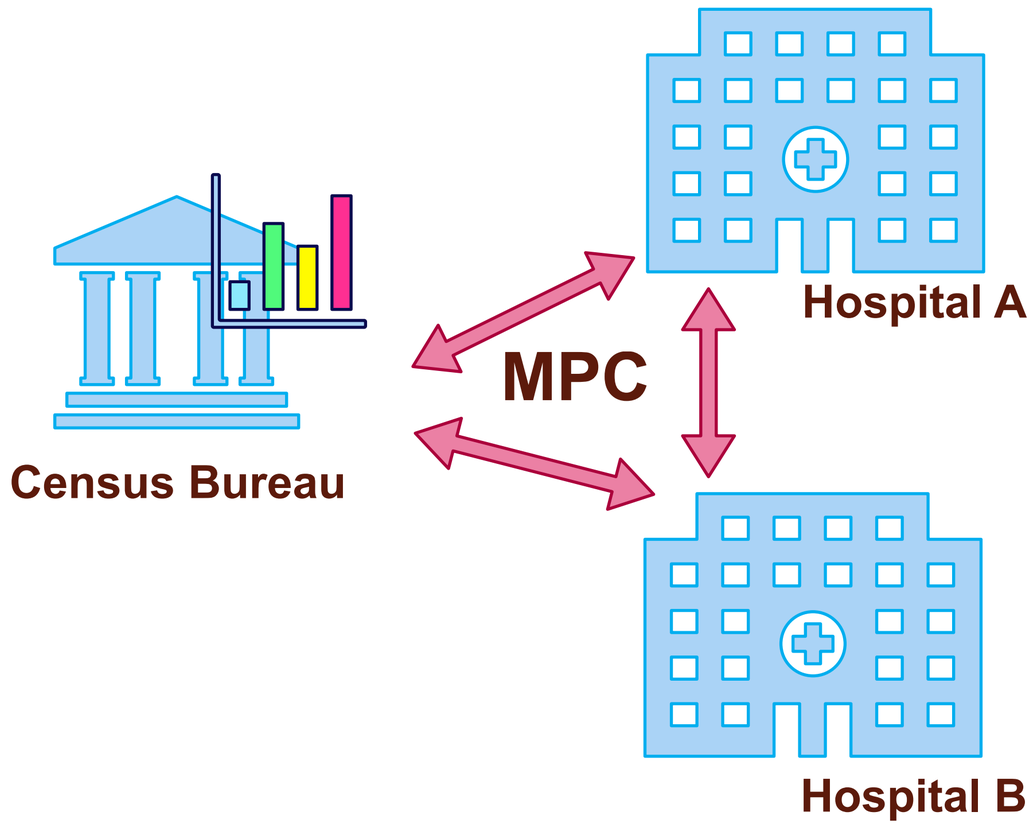
“Is this related to Blockchain?”

Yes and no. Just like “Blockchain”, Multiparty Computation is an instance of modern cryptography involving multiple parties. However, the aims are very different:

- A *Blockchain* is a distributed ledger with immutability and data-ordering guarantees, and uses some consensus mechanism for adding content to the ledger.
- In Multi-Party Computation, the goal is to jointly compute on “blinded” data with privacy and non-malleability guarantees

Example

Consider a scenario where the Census Bureau (e.g., CBS) wants to learn a statistic over medical records of several hospitals



- Census Bureau will only learn pre-agreed statistic, and *nothing beyond this*.
- Data is never replicated, thus a data breach at the Census Bureau will not expose the medical records: **no single point of failure w.r.t. data privacy**
- Hospital A cannot “see” data from Hospital B and *vice versa*

Results

Python package for MPC: **MPyC**

MPyC implementations of:

- Ridge Regression
- Multilayer Perceptron (for Image Classification)
- Convolutional Neural Networks
- Decision Trees

SODA Scalable Oblivious Data Analytics

PRIVBLEDGE

Niek J. Bouman (Data Mining / Cryptography, TU Eindhoven)
joint work with Frank Blom, Niels de Vreede, Berry Schoenmakers (TU Eindhoven)