Associate Professorship of Coding and Cryptography
TUM Department of Electrical and Computer Engineering
Technical University of Munich

Master's Thesis

Private, Secure and Flexible Distributed Machine Learning on the Cloud

In the era of big data, running learning algorithms on single computing is becoming a bottleneck. Therefore the need of distributing the computation is inevitable. However, the deployment of distributed computing introduces new challenges that, if ignored, may outweigh the benefit of parallelism.

We consider the master-worker topology in which the master needs to run a computation on its data. The master breaks the computation into smaller tasks distributed to processing nodes referred to as workers. The workers run the tasks in parallel. The master combines the results sent back from the workers to obtain its original computation.

However, heterogeneity of the workers' computation power and/or network link properties can slow down the process for some workers. Waiting for the slowest worker is shown to outweigh the effect of parallelism. Moreover, the privacy of the data is concerned since it is shared with external workers. The leakage of the master data can be harmful or even illegal. The master also risks employing an adversarial computing node as a worker, whose goal is to corrupt the whole computation.

Recently, coding theoretic techniques have been used to speed up the computation, guarantee the privacy and security of the distributed computing paradigm under different settings. The setting of interest for this project is one where the workers have different time-varying compute powers. Thus a flexible coding technique is needed. We focus on matrix-matrix multiplication as a building block in several machine learning algorithms.

In this project, we would like to design codes that allow us to overcome all the three challenges for matrix-matrix multiplication (or more types of computation if the time allows). We start by building on the work in https://arxiv.org/abs/2004.12925. The main focus of the project is on the security of the scheme, i.e., robustness against malicious workers trying to corrupt the computation. We aim to implement the designed codes on Google cloud platform, or Amazon Web Services to test their practicality.

Prerequisites

Coding Theory
Information Theory
Linear Algebra
Probability Theory
Programming skills

Self-motivation

Contact

Marvin Xhemrishi: marvin.xhemrishi@tum.de

Rawad Bitar: rawad.bitar@tum.de

Advisors

Marvin Xhemrishi, Rawad Bitar