Privacy-preserved Distributed Analysis for Sensitive Datasets

The analysis of large datasets has brought new insights that have significantly influenced our daily life. Sweden have significantly contributed to data collection and analysis of the genetic architecture of medical and mental disorders [1,2,3]. The ultimate aim of such genetic studies is to understand how actions and interactions of genetic and environmental risk factors affect individuals, and to target prevention and treatment in a precision medicine context [5]. However, accessing data, even in an unidentified form, is regulated by national laws and requires strict data protection steps.

The conventional methods for data analysis proven to be effective but at the same time have raised serious concerns about privacy, integrity, and security. The General Data Protection Regulation (GDPR) is the artifact of the growing concerns. The legal framework of GDPR set the guidelines for data protection and privacy. However, technically privacy preserved data analysis is an area that still requires significant efforts [4]. The established machine learning models are rigid and require a global view of the data i.e. single site contains the complete dataset. One of the expectations from the privacy-preserved data analysis is that it will not require a global view of the data. Organizations keep their sensitive data within their secure premises yet have means to run the large-scale analysis. Different efforts have been made to accomplish this task using integrity checks, secure communication protocols and identify a class of models that can execute in distributed settings.

In this project, the task is to design and implement a framework that allows the privacy-preserved distributed data analysis platform for sensitive biological data. For this, we will explore the blockchain technology [6]. So far, transactional systems used blockchain as a successful model for data integrity and reduce the need for the trust. The technology has a lot of potentials and we think by using blockchain technology we can move one step further in the direction of the privacy-preserving data analysis.

Eventually, the methods and tools developed in this project will be used for providing a secure large-scale robust distributed computing platform for genetic analysis of medical and mental disorders.

Contact Persons:
Salman.Toor@it.uu.se , Behrang.Mahjani@mssm.edu

References:

[1]. A study on genetic and environmental factors in autism: https://ki.se/meb/saga
[2]. A study of genetic and environmental factors in obsessive-compulsive disorder, Tourette syndrome and tics: https://ki.se/meb/egos-studien
[3]. Nina Roswall Sven Sandin Hans-Olov Adami Elisabete Weiderpass, Cohort Profile: The Swedish Women's Lifestyle and Health cohort, International Journal of Epidemiology, Volume 46, Issue 2, 1 April 2017
[4]. Mahsa Shabani, Pascal Borry, Rules for processing genetic data for research purposes in view of the new EU General Data Protection Regulation, European Journal of Human Geneticsvolume 26, pages149–156 (2018)
[5]. Euan A. Ashley,Towards precision medicine, Nature Reviews Genetics volume 17, pages 507–522 (2016)
[6]. Guy Z., Oz N., Alex S. P. Decentralizing Privacy: Using Blockchain to Protect Personal Data. 2015. IEEE CS Security and Privacy Workshops. DOI: 10.1109/SPW.2015.27.