Seed money

Fully Decentralized Authority For Privacy---Conscious Medical Data Sharing

Name of PI

Prof. Jean-Pierre Hubaux

Allotted Budget

70’000 CHF

Covered period

Oct. 2016 - Sept. 2017

Description of goals

The use of genomic data in clinical practices has been gaining momentum over the past decade. For rare diseases and the diagnosis and treatment of cancer, tremendous advances have been made, while using genomics and other *omics* data for disease risk-assessment in healthy individuals is a more recent and future-looking approach. In most cases, genomic medicine tools remain in the realm of research, and only some of them are crossing over into clinical application, where they have the potential to markedly alter the clinical care of patients.

One of the major problems slowing down this adoption is that clinical and applied research on rare diseases cannot reach its full potential as it often suffers from inadequate (in terms of size and match) control population and replication. This procedure is heavily hindered by the fact that institutions do not share genetic and clinical data due to ethical and security regulations/concerns. In some countries, governmental policies even prohibit sharing of individual---level data across national borders.

As a consequence, preserving privacy in genomic/medical data analyses while enabling scientific discovery via cross-institutional collaborations remains a big challenge that needs to be solved for realizing the vision of personalized medicine.

Most of the attempts of secure medical data sharing systems that exist today are based on the centralized approach (e.g. Genomics England [1]) where data of patients from different stakeholders are pseudonymized and stored on a single and centralized repository where researchers can obtain aggregated or individual information under different level of registered access. Although such an approach seems to be the simplest and most intuitive one, it is heavily based on the trust between the data providers and the central repository (i.e., data providers need to completely trust the centralized repository to securely store and manage the data of their patients). Moreover, because such a repository (or server) is responsible for the security, availability and functionality of the whole system, it represents a *weakest link* security, which could be easily exploited by attackers. For example, a breach caused by an external (hacker) or internal (insider) attack could compromise all patients’ data as it recently happened for Anthem in the U.S. [2].

In this project, we want to overturn such a paradigm.

Our goal is to explore and propose a novel, fully decentralized and peer---to---peer architecture that does not rely on the trust on a single entity, but provides *strongest-link* security by decentralizing the data and the computations over different medical institutions.

We will develop and combine new and well-known privacy-enhancing technologies based on (i) consensus protocols to decentralize trust, (ii) homomorphic encryption to allow computation on encrypted data and preservation of patients’ privacy and (iii) verifiable computation to ensure the correctness of the secure analyses. In particular, we will make use of a new concept of collective authority (or “cothority”), recently introduced in [3], i.e., a group of independent servers (e.g., if we consider the Swiss case, each server could be owned by a different cantonal hospital) to split the trust and yet allow efficient and distributed computations on encrypted data.

Through our system, a researcher will be able to obtain aggregated information in a privacy-preserving way from other institutions without the need of moving the genomic or clinical data around, thus being compliant with institutional or governmental regulations. Moreover, all medical institutions in the system will always be able to keep control on their own data, hence ensuring the privacy of their patients. Security and privacy of the data will be guaranteed as long as at least one institution is not compromised.

The provided seed money would make it possible to carry out performance investigations and the implementation of a first prototype. It would be used to remunerate 1 post-doc for 4 months (4*10’000 CHF) and 1 PhD student for 7 months (7*6’000 CHF), for a total budget of 82’000 CHF (we assume there are no overheads). This research investigation will leverage on our substantial know-how developed over the last 5 years on the topic of genome privacy and security [4].

Jean Louis Raisaro and Jean---Pierre Hubaux, EPFL

References

[1] Genomics England. http://www.genomicsengland.co.uk

[2] “Anthem cyber-attack,” https://www.anthemfacts.com

[3] Syta, Ewa, et al. "Keeping Authorities "Honest or Bust" with Decentralized Witness Cosigning." 2016 IEEE Symposium on Security and Privacy. IEEE, 2016.

[4] http://lca.epfl.ch/projects/genomic-privacy/

Milestones achieved

Through our system (MedCo), a biomed researcher is able to obtain aggregate information in a secure and privacy-preserving way from other institutions, without the need of moving the genomic or clinical data around, thus being compliant with institutional or governmental regulations. Moreover, all medical institutions in the system are always able to keep control on their own data, hence ensuring the privacy of their patients. Security and privacy of the data is guaranteed as long as at least one institution is not compromised.

This work has been extremely useful to obtain the PHRT project “Data Protection for Personalized Health” (3 mio CHF over 3 years).

The work is described in the following paper:

Jean Louis Raisaro et al.
MedCo: Enabling Privacy-Conscious Exploration of Distributed Clinical and Genomic Data GenoPri 2017
http://www.genopri.org/uploads/3/9/9/9/39999711/genopri17_paper_6.pdf