Practical Solutions for Protecting Individual Genomic Privacy

Jean Louis Raisaro[1], Zhicong Huang[1], Mathias Humbert[1], Erman Ayday[1], Paul J McLaren[2], Jean-Pierre Hubaux[1], Amalio Telenti[3], Jacques Fellay[2].

1. School of Comp. and Comm. Sciences, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

2. School of Life Science, Ecole Polytechnique Fédérale de Lausanne, Switzerland.

3. Institute of Microbiology, University Hospital and University of Lausanne, Lausanne, Switzerland

The increasing availability of genomic data has major implications for personal privacy. The issues raised by genomic privacy reside at the crossroads of medicine, computer science, legislation and public policy. We here describe the design and development of new privacy enhancing technologies that aim to find the optimal balance between usability and privacy of genomic data in clinical care and in biomedical research.

First, we propose a privacy-preserving algorithm for genetic risk testing in clinical care that uses homomorphic encryption and proxy re-encryption. After genomic data (e.g. sets of variants from whole genome sequencing) are generated by a certified institution, they are encrypted and stored at a centralized "storage and processing unit" (SPU). Our architecture, while preserving data privacy, enables a medical unit to retrieve the encrypted information from the SPU and to use it for individualized care. We deployed this solution in a pilot pharmacogenomics study of 180 patients participating in the Swiss HIV Cohort Study. Retrieval and processing of encrypted genotypes, for a test using 50 markers (SNPs), take less than 1 second on commodity hardware. An interim analysis showed this to be acceptable by clinicians as both usability and privacy of genomic data are preserved.

Second, we developed a system to protect the privacy of mapped short reads (e.g. bam files). Millions of sequencing reads from individual genomes are stored at a SPU in encrypted form. Our scheme allows a medical unit to privately retrieve a subset of the reads without revealing the nature of the request to the SPU. In addition, the SPU can mask particular parts of the retrieved reads if they are not in the requested range or not consented by the patient (e.g., regions revealing sensitive diseases).

Finally, we provide a method, based on graphical models and belief propagation, to estimate the erosion of genomic privacy of an individual when genomic data of some of his/her relatives are publicly available. We showed how a target genome can be reconstructed by relying on Mendel's laws and linkage disequilibrium. As a result of this inference attack, we proposed different possible definitions of genomic privacy metrics.