Privacy-Preserving Genomics on a Large Scale

Oleksandr Tkachenko
TU Darmstadt

29th Crypto Day, 6/7 September 2018

In this talk, we present privacy-preserving solutions for Genome-Wide Association Studies (GWAS) based on Secure Multi-Party Computation (SMPC). Using SMPC, we protect the privacy of patients when medical institutes collaborate for computing statistics on genomic data in a distributed fashion. Previous solutions for this task lack efficiency and/or use inadequate algorithms that are of limited practical value. Concretely, we optimize and implement multiple algorithms for the χ^2-, G-, and P-test in the ABY framework (Demmler, Schneider & Zohner (2015b), NDSS'15) and evaluate them in a distributed GWAS scenario. Statistical tests generally require advanced mathematical operations. For operations that cannot be calculated in integer arithmetic, we make use of the existing IEEE 754 floating point arithmetic implementation in ABY (Demmler, Dessouky, Koushanfar, Sadeghi, Schneider & Zeitouni (2015a), CCS'15). To improve performance, we extend the mixed-protocol capabilities of ABY by optimizing and implementing the integer to floating point conversion protocols of Aliasgari, Blanton, Zhang & Steele (2013) (NDSS'13), which may be of independent interest. Furthermore, we consider extended contingency tables for the χ^2- and G-test that use codeword counts instead of counts for only two alleles, thereby allowing for advanced, realistic analyses. Finally, we consider an outsourcing scenario where two non-colluding semi-trusted third parties process secret-shared input data from multiple institutes. Our extensive evaluation shows, compared to the prior art of Constable, Tang, Wang, Jiang & Chapin (2015) (BMC Medical Informatics and Decision Making'15), an improved runtime efficiency of the χ^2-test by up to factor 37×. We additionally demonstrate practicality in scenarios with millions of participants and hundreds of collaborating institutes. The results of this work were published by Tkachenko, Weinert, Schneider & Hamacher (2018) at ASIACCS'18.

References

Mehrdad Aliasgari, Marina Blanton, Yihua Zhang & Aaron Steele (2013). Secure Computation on Floating Point Numbers. In *NDSS*.

Scott D Constable, Yuzhe Tang, Shuang Wang, Xiaoqian Jiang & Steve Chapin (2015). Privacy-preserving GWAS analysis on federated genomic datasets. In *BMC medical informatics and decision making*.

Daniel Demmler, Ghada Dessouky, Farinaz Koushanfar, Ahmad-

crypto day matters 29 (2018)
Beiträge vom 29. Krypto-Tag, 06.-07.09.2018, Bosch, Renningen.
doi:10.18420/cdm-2018-29-25

Reza Sadeghi, Thomas Schneider & Shaza Zeitouni (2015a). Automated Synthesis of Optimized Circuits for Secure Computation. In *CCS*.

Daniel Demmler, Thomas Schneider & Michael Zohner (2015b). ABY – A Framework for Efficient Mixed-Protocol Secure Two-Party Computation. In *NDSS*.

Oleksandr Tkachenko, Christian Weinert, Thomas Schneider & Kay Hamacher (2018). Large-Scale Privacy-Preserving Statistical Computations for Distributed Genome-Wide Association Studies. In *ASIACCS*.