

GenoPri'17: International Workshop on Genome Privacy and Security

Erman Ayday, Muhammad Naveed, and Haixu Tang



OVER the past several decades, genome sequencing technologies have evolved from slow and expensive systems that were limited in access to a select few scientists and forensics investigators to high-throughput, relatively low-cost tools that are available to consumers. A consequence of such technical progress is that genomics has become one of the next major challenges for privacy and security [1] because (i) genetic diseases can be unveiled, (ii) the propensity to develop specific diseases (such as Alzheimer's) can be revealed, (iii) a volunteer, accepting to have his genomic code made public, can leak substantial information about his ethnic heritage and the genomic data of his relatives (possibly against their will) [2], [3], and (iv) complex privacy issues can arise if DNA analysis is used for criminal investigations and medical purposes [4].

As genomics is increasingly integrated into healthcare [5] and "recreational" services (e.g., ancestry testing) [6], the risk of DNA data leakage is serious for both individuals and their relatives. Failure to adequately protect such information could lead to a serious backlash, impeding genomic research, that could affect the well-being of our society as a whole. This prompts the need for research and innovation in all aspects of genome privacy and security.

To foster research in understanding and addressing privacy and security issues in genomics, we organized the 4th International Workshop on Genome Privacy and Security (GenoPri) in 2017. This workshop aimed to bring together a highly interdisciplinary community involved in all aspects of genome privacy and security research. This workshop built on its three predecessors, GenoPri'14, GenoPri'15, and GenoPri'16 which were collocated with the Privacy Enhancing Technologies Symposium (PETS), IEEE Symposium on Security and Privacy, and American Medical Informatics Association Annual Fall Symposium (AMIA), respectively.

Following the success of GenoPri'16, expanding its reach to medical community, GenoPri'17 was collocated with the GA4GH Security Working Group (SWG) and the American

Society of Human Genetics (ASHG) Annual Meeting. The number of participants during the day exceeded 75 and the workshop included technical sessions (featuring peer-reviewed papers), a panel, and two invited talks.

The themes of the technical sessions were (i) risks of genomic data sharing—attacks and threats, (ii) future in pharmacogenomics—protection techniques for the genomic pipeline, and (iii) privacy-preserving solutions for processing genomic data. In total, eight technical papers were presented during these sessions.

The panel, moderated by Paul Flicek—EMBL-EBI, focused on "current challenges in genome privacy and security" and the participants were Dixie Baker (Martin, Blanck, and Associates), Knox Carey (Genecloud), Kristin Lauter (Microsoft), and Laura Lyman Rodriguez (National Human Genome Research). Finally, the two invited talks were delivered by Bartha Knoppers (McGill University) and Kristin Lauter (Microsoft).

The four papers selected to appear in this special section of the *IEEE/ACM Transactions on Computational Biology and Bioinformatics* were drawn from the set of accepted peer-reviewed technical papers. The papers that were accepted cover prominent perspectives of genomic privacy, including secure genomic analytics, privacy-preserving data sharing, and user studies. We have ordered the papers to illustrate the progression of privacy and security with respect to genomic data.

The first paper, "End-to-End Security for Local and Remote Human Genetic Data Applications at the EGA" describes how. The European Genome-phenome archive (EGA) provides secured access (considering storage, access, transfer, and data usage) to large genomic datasets for many users across the world. The authors also touch on points of integration with emerging standards in the GA4GH.

The second paper also provides practical solutions as the first one. The paper, "MedCo: Enabling Secure and Privacy-Preserving Exploration of Distributed Clinical and Genomic Data" describes an implementation of the UnLynx protocols (for privacy-preserving sharing of medical data across independent clinical sites) in the MedCo system between several Swiss universities.

The third paper, "Privacy-Preserving Linkage of Genomic and Clinical Data Sets," addresses a key challenge faced by the Global Alliance for Genomics and Health (GA4GH): linking datasets of the same individual. This paper reports the findings and recommendations of the GA4GH Privacy-Preserving Record Linkage Task Force for addressing this challenge.

- E. Ayday is with the Department of Electrical Engineering and Computer Science, Case Western Reserve University, Cleveland, OH 44106. E-mail: exa208@case.edu.
- M. Naveed is with the Department of Computer Science, University of Southern California, Los Angeles, CA 90089. E-mail: mnaveed@usc.edu.
- H. Tang is with the School of Informatics and Computing, Indiana University, Bloomington, IN 47405-7000. E-mail: hatang@indiana.edu.

For information on obtaining reprints of this article, please send e-mail to reprints@ieee.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TCBB.2019.2891029

The fourth paper, “On Sharing Intentions, and Personal and Interdependent Privacy Considerations for Genetic Data: A Vignette Study,” conducts a factorial vignette study to understand how ethnicity, age, genetic markers, and association of data with the individual’s name affect data sharing with different organizations.

We are highly encouraged by the increased participation and witnessed the growing of an interdisciplinary research community. The joint meeting with the GA4GH Security Working Group has been fruitful and many participants benefited from the cross-fertilization from each other. We are planning to continue the momentum in 2018 to work closely with the biomedical informatics community and push the fronts of genomic privacy technology to address the real world challenges. Thus, GenoPri’18 again will be collocated with the GA4GH Security Working Group (SWG).

REFERENCES

- [1] M. Naveed, E. Ayday, E. W. Clayton, J. Fellay, C. A. Gunter, J. P. Hubaux, B. A. Malin, and X. Wang, “Privacy in the genomic era,” *ACM Comput. Surveys*, vol. 48, no. 1, 2015, Art. no. 6.
- [2] S. S. Shringarpure, “Bustamante CD. Privacy leaks from genomic data-sharing beacons,” *Amer. J. Hum. Genet.*, vol. 97, pp. 631–646, 2015.
- [3] N. Homer, S. Szlinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, et al., “Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays,” *PLoS Genetics*, vol. 4, no. 8, 2008, Art. no. e1000167.
- [4] M. A. Jobling and P. Gill, “Encoded evidence: DNA in forensic analysis,” *Nature Rev. Genetics*, vol. 5, no. 10, pp. 739–751, Oct. 2004. PMID: 15510165.
- [5] T. Manolio, R. Chisholm, B. Ozenberger, et al., “Implementing genomic medicine in the clinic: The future is here,” *Genetics Med.*, vol. 15, pp. 258–267, 2013.
- [6] J. S. Roberts J. Ostergren, “Direct-to-consumer genetic testing and personal genomics: A review of recent empirical evidence,” *Current Genetic Med. Rep.*, vol. 1, no. 3, pp. 182–200, 2013.

Erman Ayday received the MS and PhD degrees (under the supervision of Dr. Faramarz Fekri) from the School of Electrical and Computer Engineering (ECE), Georgia Institute of Technology, Atlanta, GA, in 2007 and 2011, respectively. He is an assistant professor of computer science at Case Western Reserve University, Cleveland, OH and Bilkent University, Ankara, Turkey. Before that, he was a post-doctoral researcher at EPFL, Switzerland, working with Prof. Jean-Pierre Hubaux. His research interests include privacy-enhancing technologies (including big data and genomic privacy), applied cryptography and data security, trust and reputation management, and inference from big data. He is the recipient of the Distinguished Student Paper Award at IEEE S&P 2015, the 2010 Outstanding Research Award from CSIP at Georgia Tech, and the 2011 ECE Graduate Research Assistant Excellence Award from Georgia Tech. His other various accomplishments include several patents, research grants, and the H2020 Marie Curie individual fellowship. He has published more than 50 peer-reviewed papers in prestigious venues including ACM CCS, IEEE S&P, and the *IEEE Transactions on Dependable and Secure Computing*. He has been also serving in the program committee of many conferences including ACM CCS, NDSS, and AsiaCCS.

Muhammad Naveed received the PhD degree from UIUC, in May 2016. He is an assistant professor of computer science at the University of Southern California. He works on applied cryptography and systems security. His work has had a significant impact on Android security and has helped companies such as Google, Samsung, Facebook, and Amazon, improving security for millions of Android users. He is the recipient of the Google PhD Fellowship in Security, the Sohaib and Sara Abbasi Fellowship, the C.W. Gear Outstanding Graduate Student Award from the UIUC Department of Computer Science, the Ross J. Martin Outstanding Research Achievement Award from the UIUC College of Engineering, and the best paper award at the NYU CSAW Security Research Competition.

Haixu Tang received the PhD degree in computational biology from the Shanghai Institute of Biochemistry, He is a professor in the Department of Computer Science, and the director of the Data Science Program in the School of Informatics, Computing, and Engineering at Indiana University, Bloomington. He worked as a postDoc researcher in the University of Southern California (1990-2001) and the University of California, San Diego (2001-2004), before joining Indiana University, in 2004. He has been working on bioinformatics for over 20 years. His current research interests including computational mass spectrometry, multi-omic data integration, and privacy-preserving techniques for bioinformatics. He is a recipient of the NSF CAREER Award, in 2007. He is currently an associate editor of *BMC Bioinformatics* and serves as an editorial board member of *IEEE/ACM Transactions on Computational Biology and Bioinformatics*.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.