Thesis Project Portfolio

Cybersecurity: Protecting Genomic Data by Improving the Security Hygiene of DNA Processing Programs

(Technical Report)

A Sociotechnical Analysis of Consumer Genetic Testing on The Understanding of Privacy

(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Stephanie M. Skahen

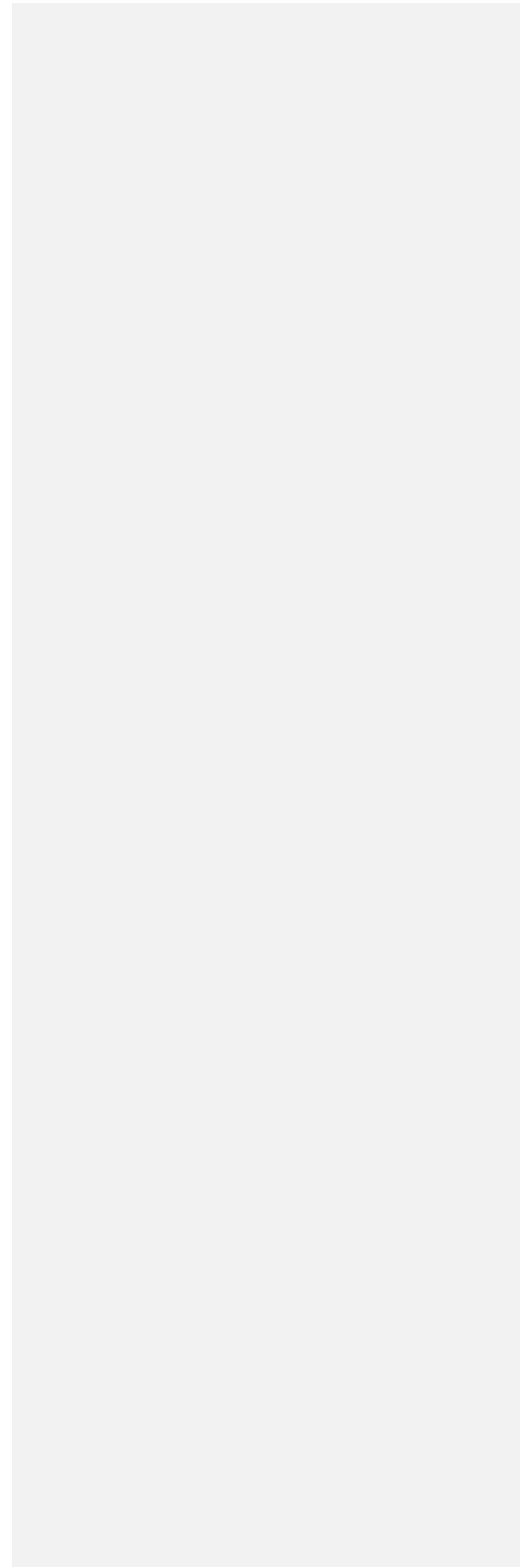Spring, 2023

Department of Computer Science

Table of Contents

Both projects included in this thesis are about the privacy of Direct to Consumer (DTC) DNA testing. The sociotechnical report discusses how the perception of what information is and should be considered private. It outlines whether genetic information is still considered private to the individual or not. The technical report offers a solution to ensuring that genetic information is protected from data breaches. One aspect focuses on if the information is considered private, while the other secures that information and keeps it private.

There is a high volume of genetic information from DTC DNA testing sites such as 23andMe and Ancestry.com that is processed, stored, accessed, and shared on the internet. Without proper protections and security practices, the risk that genetic information can be stolen and used with ill intent increases. Security can be improved by creating policies and mandates in line with security hygiene practices taught in CS 3710 and other cyber security courses at UVA. For the initial hygiene status of DTC synthesis and storage companies, I conducted a literature review and analysis of policies and research into their vulnerabilities. Based on preliminary research, it appears that until recently, there were no proper policies in place to prevent companies from selling or sharing the anonymized data they collect. The databases containing this information and processing software contain "messy" code and opening doors for possible breaches. Based on my analysis I propose that these companies adopt new cybersecurity practices including: removing buffer overflow opportunities, and creating tighter identity access protocols. Follow on analysis should be done on the effectiveness of these practices on large scale databases.

With the increase of consumer genetic testing, storage, and tracing the privacy concerns surrounding individuals' genetic information become important. This prompts the question: how

could the widespread documentation and storage of genomic data and the sociotechnical systems associated with it, impact public understanding and perception of privacy in the United States? This is answered using a discourse analysis and is analyzed through the framework of Co-Production. The purpose of this research is to determine the extent of which Americans believe their genetic information is kept private and whether it matters to individuals within American society.

Working on both of these projects simultaneously allowed me to view the problem of privacy from two different perspectives. I was analyzing what privacy meant to an individual and what it takes to give up some control over one's own information, while also identifying what role corporations and software companies play in ensuring the privacy of individuals' information. DTC genetic testing users release their information for research or family connection with the expectation that the companies have done their job in protecting their information from unauthorized actors. I was able to identify the relationship between a users privacy and the company receiving their data, while also providing a framework through which to further protect their data and give users more control over their information.