Thesis Project Portfolio

Tools to Enhance Internal Cybersecurity Awareness

(Technical Report)

The Role of Data Fragmentation in Data Breaches

(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

Andrew Li

Spring, 2023

Department of Computer Science

Contents of Portfolio

Executive Summary

Big data underlies the products, services, and innovations that our digital economy is sustained by. With the new opportunities this massive amount of information brings, there are also commensurate risks and inefficiencies. The most problematic of these issues revolve around cybersecurity, where a single failure can put the sensitive personal data of millions of people at risk at once. Protecting large quantities of data while leveraging their use is a key challenge for companies today. This challenge is especially salient for financial institutions that must safeguard their customers' financial information and utilize comprehensive data collection to prevent breaches, prevent identity theft, and provide better services. A major obstacle in the effective storage and use of data is the fact that big data naturally trends towards disorganization and fragmentation as databases become increasingly complex and diverse. Data fragmentation plagues businesses and institutions in a fraught cybersecurity environment where data breaches are common. To better understand the vulnerabilities and opportunities that big data entails, it is important to investigate the role data fragmentation plays in modern data breaches as well as methods that use data to prevent and mitigate future cyberattacks.

The technical report outlines my summer project at Capital One to build and extend a universal search feature into an internal cybersecurity knowledge base. The purpose of this project was to aid in Capital One's cyber resiliency efforts by expanding the accessibility and utility of the large amounts of information Capital One creates to monitor and analyze the cyber environment. In particular, the project aimed to increase knowledge and familiarity with cybersecurity risks and vulnerabilities among non-technical associates within Capital One. To do this, the project built a new search application into an existing internal site. The project focused on user-friendliness and simplicity while maintaining powerful advanced features for associates

with more familiarity with the database. The result was a unified portal that connected the disparate data sources in Capital One into a single access point that could be searched and filtered easily.

The research paper investigates two data breaches—Capital One in 2019 and Equifax in 2017—and the role data fragmentation played or did not play in it. I perform this analysis using two in-depth case studies focusing on the various entities involved in the data breaches. By basing the case studies on Actor Network Theory, I find that data fragmentation is linked to proliferation in the number of institutions, organizations, and companies involved in the storage and use of data and software. The increasing complexity of these entities as well as their relationships to each other lead to data fragmentation that invites room for error as well as the shifting of blame when cybersecurity incidents do happen. While the trends towards fragmentation of both information and entities is not necessarily a bad thing, the challenges it brings must be addressed by both private and public reform.

These two projects were a challenging but rewarding exercise. Together, they demonstrate the possibility for unification to help reverse some of the pernicious side-effects of our reliance on data. Though they were limited in scope, I hope that they provide a starting point for further research and development that can identify and create solutions for the complex issue of cybersecurity.