First Steps Toward Compression-aware Algorithms
(Technical Report)

Analysis of Target Corporation's 2013 Data Breach Via a Deontological Framework
(STS Research Paper)

An Undergraduate Thesis Portfolio

Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia, Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Yonathan Fisseha

May 1, 2020

Socio-technical Synthesis

Efficient Multi-party Computation for Private Data Sharing

Data privacy is at the intersection of my STS and technical research. As the world is becoming increasingly interconnected with the growth of the World Wide Web, more private data is also shared between individuals and corporations. The technical research recognizes the dilemma of this need to share data while also maintaining some level of privacy. It formalizes this notion by meditating on why people often want to share data– to derive new knowledge from the aggregated data– and offers a technical solution that balances these two needs. On the other hand, the STS research explores the ethical complexities of private data storage and computation. It considers the particularly egregious case of Target's 2014 data breach from a moral theory perspective and shows that data breaches have a moral dimension worth exploring. The concern for the security of private data and recognition of the need to share data thus forms the foundation of both projects.

The technical research explores the application of compression-aware algorithms to multi-party computing to provide efficient and secure means to compute on aggregated data. It considers a specific, but generally applicable, use case of measuring similarities between strings in a private manner. More specifically, the privacy model used is one where no information about the two strings is leaked except the result of the similarity measure. To accomplish this privacy model, I rely on multi-party computing protocols that provide a generalized framework to privately compute an arbitrary function on a given set of inputs. Much like other algorithms performed via a multi-party computing protocol, string similarity measures are computationally expensive and slow. The technical research aims to provide an efficient means to compute string similarity algorithms by first compressing the strings and then by designing a compression-aware version of the string similarity measuring algorithm. Towards this goal, this research provides the experimental study necessary to design these compression-aware algorithms.

The STS research is similarly focused on understanding privacy and practical security.

More specifically, it analyzes the 2014 breach of Target Corporation's computer systems from an ethical perspective in an effort to understand the ethical responsibilities of parties that compute on and store private user data. The analysis is based on deontological ethics and uses the code of ethics created by the Association for Computing Machinery (ACM) to provide a rigorous normative judgment on the specific case. I argue that while it is difficult and perhaps impossible, to equally distribute moral blame for the data breach that resulted from the attack, it is possible, and important, to hold Target, as a collective, morally responsible for the attack. The research descriptively shows that Target has violated at least three rules from ACM's code of ethics by analyzing the technical and organization flaws that were exploited leading to the system breach and customer data leak. The goal of this analysis is to show that entities that process and store private user data can and must be held to higher standards of moral responsibility beyond their industry's technical and legal bare minimum requirements.

Working on these projects simultaneously has enabled me to build a better context around both my technical and STS research. The STS research directly influenced my technical research by giving me a better understanding of the actors involved in multi-party computing scenarios. I studied inter-organizational data sharing, businesses' need to aggregate data and compute on aggregated data across organizational boundaries, and the privacy concerns that arise from this. Consequently, I made various decisions to adjust the technical project such that it addresses business-to-business data sharing as opposed to data sharing among individuals. The technical project was helpful in contextually understanding the technical challenges of data sharing in a privacy-conscience manner. This made the analysis in the STS research more robust and the moral judgment more just because I understood the practical boundaries of security in the technical domain.

Table of Contents