Thesis Project Portfolio

Precision Attacks on Differential Privacy

(Technical Report)

A Deontological Analysis of Privacy Policy Presentations

(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

Jack Mingjie Liu

Spring, 2023

Department of Computer Science

Contents of Portfolio

Executive Summary

In our technologically connected world, every aspect of our lives takes place surrounded by devices connected to the internet. These gadgets are able to record our personal information which is then shared and stored by companies through digital platforms, and as a result, it is difficult to ensure that our data is private and secure. Privacy policies aim to provide transparency on how a company will utilize your data, but they are only effective if they are clear and easy to understand. In actuality, privacy policies can be overly complex and full of legalese which undermines their effectiveness. Moreover, there is another side to this issue of protecting user privacy. Even if companies are well-intentioned in protecting privacy, there exist bad actors who will engage in malicious attacks in order to steal private information. It is equally important to study whether the methods used to secure our data are sufficiently hardened against such attacks.

Differential privacy is a mathematical definition of privacy that describes the property of an algorithm to calculate aggregate statistics on a population while keeping any individual's data private. However, in 2012 a researcher named Ilya Mironov showed that special care must be taken when implementing the purely mathematical theory on concrete physical machines. My technical report explores this attack by examining the least significant bits of the output to break differential privacy guarantees when using the NumPy Python library. I found that a naïve implementation of differential privacy using the Laplacian distribution leaks information that can be used to break differential privacy over 60 percent of the time. This closely aligns with the original work by Mironov for other platforms. While some mitigation strategies have been developed for this attack, there are examples of some existing differential privacy libraries that use this naïve implementation and are thus vulnerable to malicious actors. This work underscores

how the finite precision of computers is not often considered when designing algorithms but can have dramatic impacts on security.

As mentioned above, privacy policies are used by companies to disclose their practices regarding data collection, usage, and management. However, due to the language of these documents, it is difficult for most users to understand and so they may blindly accept without fully understanding what the agreement entails. This raises the question of how data collection and data use policies are communicated to users, and how they can be made more transparent. Specifically, it is important to study the role of privacy policies in society and whether or not they are filling their intended niche. My STS Research Paper studies this under a deontological ethics framework. The main finding of my work is that some companies have made improvements towards making these policies less dense and full of legalese, but overall privacy policies are still inaccessible for a large portion of the user base. This undermines individual autonomy when it comes to deciding what they are willing to share with companies. These results came from utilizing the Flesch Reading Ease and Flesch-Kincaid Grade Level scores to quantify readability as well as comparing documents to a set of principles set forth by previous researchers. The design of these documents as well as website banners were also utilized in this analysis. In all, my work helps to provide one perspective on privacy policies to see what they do well and what can still be improved.

The work included in this portfolio represents the work I've accomplished over the past two semesters in studying the multifaceted issue of digital privacy. While I am satisfied with the work that has been done in both of my projects, there is still more to be explored. My technical project mostly replicates the results from previous researchers, but more work can be done in discovering novel attacks on differential privacy. There are several other popular mechanisms in

which differential privacy is implemented and they may be vulnerable to similar precision attacks. As for my STS component, the research focuses on a deontological perspective which puts little emphasis on the actual consequences of these policies. It would be worthwhile to perform user studies to see the impact of privacy policies on individuals and whether they serve their intended purpose. Yet another potential area of study can be investigating the actions of the companies themselves to see if they follow the word of their policies or if they engage in some deceptive practices. In the end, I hope that my work can help us take a step towards being more aware of privacy practices and being confident that our private data is secure.