Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science

Proposal for Thesis Research in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Title: Protecting User Data in Large-Scale Web Services

Submitted by: Frank Wang
32 Vassar Street, #32-G978
Cambridge, MA 02139

(Signature of Author)

Date of Submission: April 25, 2018
Expected Date of Completion: July 2018
Laboratory: Computer Science and Artificial Intelligence Laboratory

Brief Statement of the Problem:

Web services like Google, Facebook, and Dropbox are now an essential part of peoples lives. Users willingly provide their data to these services because these services deliver substantial value in return through their centralization and analysis of data, such product recommendations and ability to easily share information. To provide this value, these services collect, store, and analyze large amounts of their users sensitive data. However, once the user provides her information to the web service, she loses control over how the application manipulates that data. For example, a user cannot control where the application forwards her data. Even if the service wanted to allow users to define access controls, it is unclear how these access controls should be expressed and enforced. Not only is it difficult to develop these secure access control mechanisms, but it is also difficult to ensure these mechanisms are practical. My thesis addresses these concerns.

1 Introduction

Web services like Google, Facebook, and Dropbox are now an essential part of peoples lives. Users willingly provide their data to these services because these services deliver substantial value in return through their centralization and analysis of data, such product recommendations and ability to easily share information. For example, users are willing to share their data with Facebook to learn about the social lives of their friends as well as share their own social lives more easily. Similarly, users are willing to provide their data to Amazon to discover better product recommendations. To provide value to users, these services collect, store, and analyze large amounts of their users' sensitive data. However, once the user provides her information to the web service, she *loses control* over how the application manipulates that data. For example, a user cannot control where the application forwards her data. Even if the service wanted to allow users to define access controls, it is unclear how these access controls should be expressed and enforced. Not only is it difficult to develop these secure access control mechanisms, but it is also difficult to ensure these mechanisms are *practical*. This thesis addresses these concerns. More specifically, it focuses on *building practical, secure mechanisms for protecting user data in large-scale, distributed web services.*

In this thesis, I will describe three systems that address a variety of concerns around data leakage in web applications. The first two systems focus on protecting user data against server-side leakage. Splinter leverages a recent cryptographic primitive, function secret sharing, to practically execute these queries without revealing sensitive information to the servers. The next system, Riverbed, provides practical information flow control for distributed systems without requiring developers to label state or write code in special languages. The final system, Veil, focuses on client-side leakage. Veil allows web page developers to enforce stronger private browsing semantics without browser support.

2 Splinter

3 Proposed Timeline

References