Implementing a Zero Trust Architecture System using Multi-Factor Authentication
(Technical Topic)

How have developments in network technologies during the 2000s increased control over user data by e-commerce companies
(STS Topic)

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Joseph Padraic Bannon
November 30, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISORS

MC Forelle, Department of Engineering and Society

Rosanne Vrugtman, Department of Computer Science

Introduction

Identity has undergone a significant shift in relevance and application during the development of the internet. During the late 1990s, internet services incorporated the ability for users to personalize their websites. This change gave e-commerce companies a business advantage over competing services, a process highlighted by O'Reilly media (O'Reilly, 2007). As a part of this change, these companies to collected information on their users related to the content they viewed or purchases they made. The value of user data increased exponentially when companies developed methods to track users over multiple websites and create a common identity per user, according to the *Hastings Science and Technology Law Journal* (Geronimo, 2017). Furthermore, a paper from the *Journal of Big Data* illustrated how developments in data tracking have vastly improved the ability of companies to recommend advertisements and products as well as provide services free of charge (Faroukhi et al., 2020). This fundamental shift in internet services is referred to as "Web 2.0" and serves as the social foundation for the technologies I will be examining in my research topic.

However, the shift in the business model of e-commerce resulted in negative effects on users. E-commerce companies have increased the amount of data collected and the sensitivity of the data collected, as cited by the *German Law Journal* (Quinn & Malgieri, 2021). The *International Data Privacy Law Journal* has noted how the escalation of data collection has also raised ethical issues concerning user privacy and informed consent (Altman et al., 2018). Another paper from Georgia Institute of Technology identifies issues with the leaking of users' personal data through advertisements (Meng et al, 2016), while researchers from the New York University demonstrated the effects of AI based on user data making life altering decisions (Stoyanovich et al., 2020). Additionally, Professor Chen from University of Pennsylvania Law

School described how e-commerce have lobbied against regulation protecting consumer data and explains the short falls of existing legislation (Chen, 2004). In general, Web 2.0 and the business model of collecting user data has created networks of control in which e-commerce companies exercise power over users and government regulation.

Initially, this paper will cover my technical topic, zero-trust architecture. The *Computer & Security Journal* defines zero-trust to mean that there is no implicit trust granted to user accounts and user identity must be verified from data they provide (Buck et al., 2021). Zero-trust architecture relates to the changes in online identity and collection of user data because it is built on collecting data from users to verify their identity. The main idea is that existing authentication systems based on network security can't keep up with the variety of user devices and cloud services in use. As a result, zero trust is based on identity instead of network location. During my summer internship, I implemented a zero-trust architecture using multi-factor authentication and collecting data on potential risk factors (i.e., time of day, location etc) of a user login. I will cover the technical details from this implementation and skills I gained through this experience.

The STS topic for this paper will focus on the shift in control of data from users to e-commerce companies. In the current state of online data collection, users have no influence over what companies do with their data. Understanding and educating users on the landscape of e-commerce data collection is vital to solving ethical concerns relating to informed consent and privacy. The goal of this paper is to identify changes that can be made to how e-commerce companies collect user data to preserve users' privacy and informed consent while still providing benefit to the e-commerce companies.

Technical Topic

The technical topic for this paper is zero-trust architecture. Zero-trust architecture is meant to replace traditional security based on network perimeters with security based on continuous verification of user identity. The shift in security methodology is necessitated by the diversification of devices and resources in a network. For example, the majority of enterprise level networks use cloud services and allow users to connect on or off premise using Bring Your Own Device (BYOD) policy. The need for users to connect to networks from outside the traditional network security perimeter exposes weaknesses in security because the network perimeter becomes a single point of failure. Zero-trust architecture addresses this concern by not granting users in the network perimeter implicit trust. Instead, users are authenticated at a granular level, meaning that every authentication takes into account the login context. According to a paper in the *MDPI Journal*, a user's login context can include their identity, device information, the resource they are accessing, etc (Sarkar et al., 2022). Furthermore, the network is segmented into components and users must reauthenticate when traversing the network and after a specific period of time in the network.

A central component of zero trust architecture is using multi-factor authentication (MFA), which is defined by NIST as using multiple independent methods of verifying a user's identity (Rose et al., 2020, p. 4). According to researchers from the National Computer Technology Center in Thailand, multi-factor authentication can be implemented by combining a password, second hardware factor (ie a token or phone), biometrics or geolocation (Vorakulpipat, 2021). Researchers from Princeton University also note that the number of online services available to users, as well as the increased use of social logins (for example logging in with Google), has led to users duplicating passwords for multiple sites (Lee & Lee, 2017). This is a massive vulnerability flaw as attackers only need to compromise one password to gain access to a majority of users' online identities. The advantage of using MFA, is that attackers must compromise every factor of authentication to gain access to a system, which eliminates passwords as a single point of failure, according to the *MDPI Journal* (Ometov et al., 2018).

During my summer internship, I worked on an implementation of a multi-factor authentication system for a state government. My role in the project was working on documentation for a proof-of-concept model of the MFA implementation as well as a project kick off presentation for the client. I gained experience with interpreting multi-factor authentication security policy and translating customers' needs into the technical implementation. For example, we used Symantec VIP hosted on Azure cloud to enable contextual based authentication. The contextual based authentication generated a risk score based on the user's login context. Data from the login context included geographic location, time of day, the devices used to login and other previous user behaviors. This allowed for a granular user access policy where logins with higher risk scores were prompted for additional authentication, even if they had a correct password. These features allowed for the client to facilitate customized login solutions that complied with government regulations as well as serve customers with varying types of devices. Through my experience, I learned the importance of flexibility and robustness in an identity and access management system.

STS Topic

The STS topic for this paper focuses on the changes in e-commerce data collection due to development of network protocols. To address this topic, I will use Actor Network Theory by Bruno Latour (1992) to create networks of e-commerce companies, users, government regulation and network technology. In particular, I am interested in how the introduction of network technologies has shifted control of user data to e-commerce companies. A paper in the *Journal of Internet Commerce* notes that during the 2000s, technologies such as cookies, flash cookies and web beacons were developed to provide a seamless experience on the web (Sipior et al., 2011). However as data collection became more of a focus, these technologies were developed to surreptitiously collect user data that is then aggregated by multiple companies over multiple websites (Geronimo, 2017). Furthermore, e-commerce companies obfuscate the data they collect while users don't know how to prevent their data from being collected. While the majority of users understand HTTP cookies enough to block or delete them, other tracking technologies have been developed to circumvent these tactics. For example, flash cookies will respawn if deleted by a user with their knowledge and blocking them disables the user from watching flash video.

In this way, the lack of knowledge about tracking technologies creates an imbalance of power between users and e-commerce companies. As outlined by Sarah West in *Business and Society Journal*, the collection of data creates, "power that is weighted toward actors [companies] that can make sense of the data" (2019, p. 23). Moreover, a study in the *International Journal of Service Industry Management* shows that when users are aware of this power imbalance, "[they] will attempt to regain power by using privacy-enhancing technologies" (Wirtz et al., 2007, p. 328). The response from e-commerce companies is to justify the system of data collection power by painting them as enabling online services such as social media. Users are also hindered from opting out of the data collection system by the design of web services. Practical examples are outlined by researchers from Cornell University and include web browsers preventing users from controlling their cookie settings and flash cookies being necessary in order to play flash video (Millett et al., 2001; Sipior et al., 2011). Furthermore, Professor Chen from University of Pennsylvania Law School shows how e-commerce companies use lobbyists to weaken or prevent regulation from affecting data collections at a governmental level (Chen, 2004). The book *Networks of Control*, also asserts shows that little action is taken to protect user privacy because most users are unaware of the data collection system and how it affects them personally (Christl & Spiekermann, 2016). Another glaring example of the lack of accountability of e-commerce company is provided by researchers from the Georgia Institute of Technology and New York University, who demonstrate how personal data can be leaked to third parties through target advertisements to users (Meng et al., 2016). More broadly, this data disproportionately affects marginalized groups and by increasing existing social and economic inequality.

To analyze the system of data collection, I will use Actor Network Theory (ANT) by Bruno Latour (1992) as my STS framework. Actor Network Theory is a sociotechnical framework that builds network of relationships between actors and posits actors only exist in relationship to one another. ANT then uses these relationships to describe how actors generate power and control over other actors. I will apply Actor Network Theory to understand how e-commerce companies use network technology and lobbying against regulation to generate control over user’s and their data. Specifically, I will be looking at how the addition of network technology from the early 2000s affected this network and shifted power towards e-commerce companies. A critique of Actor Network Theory is that it is only descriptive, but I will be using it to suggest action to solve power imbalances in data collection. The goal of this research is to identify methods to collect user data to preserve users' privacy and consent while still providing benefits to the e-commerce companies.

Methodology

The research question is, “How have developments in network technologies increased control over user data by e-commerce companies?” The primary methodology used for is research is literature review. Literature review will be sourced from the academic articles about the effects of network technology on the data collection in the e-commerce industry. I will scope my research on how cookies and related technology was implemented after 2004, which is when O’Reilly media specifies “Web 2.0” began (O’Reilly, 2007). Additionally, I will be combining this research with articles on the impacts of data collection on user privacy, informed consent and automated decision making. I will analyze this research by establishing a baseline of methods used to collect user data before the advent of web tracking technology. Then, after web tracking technology is introduced to the data collection network, I will analyze the effects on e-commerce companies, users and government regulation. The justification for this research is to first study how the addition of network tracking technology created power imbalances in favor of e-commerce companies. Then, study the response from other actors to this imbalance, for example passing legislation at a governmental level or users modifying their online behavior to limit how much of their data can be collected.

Conclusion

In conclusion, the evolution of network technology has had profound effects on how identity is managed on the internet. In my technical topic, I researched how zero trust architecture uses identity as a means of security by using multiple factors of authentication to verify identity and eliminating implicit trust based on network perimeters. Moreover, I elaborated on my experience and what I learned combining multiple technologies to implement this architecture. In my STS topic, I explored how the development of network protocols changed how e-commerce companies collect user data. Furthermore, I plan to use Actor Network Theory to identify how the interactions between actors generates control over data collection. The purpose of this paper is to understand the methods used by ecommerce companies in data collection and how they create an imbalance over control of user data. This work is meant to inform other researchers by illustrating how the introduction of specific technologies have contributed to the power imbalance between users and e-commerce companies. The goal of this research is to contribute to a data collection framework in that can be used by e-commerce companies to improve their product without violating rights of users.

Bibliography

Altman, M., Wood, A., O'Brien, D. R., & Gasser, U. (2018). Practical approaches to big data privacy over time. *International Data Privacy Law*, *8*(1), 29–51. https://doi.org/10.1093/idpl/ipx027

Buck, C., Olenberger, C., Schweizer, A., Völter, F., & Eymann, T. (2021). Never trust, always verify: A multivocal literature review on current knowledge and research gaps of zero-trust. *Computers & Security*, *110*, 102436. https://doi.org/10.1016/j.cose.2021.102436

Chen, C. (2004). United States and European Union Approaches to Internet Jurisdiction and Their Impact on E-Commerce Comment. *University of Pennsylvania Journal of International Economic Law*, *25*(1), 423–454.

Christl, W., & Spiekermann, S. (2016). *Networks of control: A report on corporate surveillance, digital tracking, big data & privacy*. Facultas.

Faroukhi, A. Z., El Alaoui, I., Gahi, Y., & Amine, A. (2020). Big data monetization throughout Big Data Value Chain: A comprehensive review. *Journal of Big Data*, *7*(1), 3. https://doi.org/10.1186/s40537-019-0281-5

Geronimo, M. (2017). *Online Browsing: Can, Should, and May Companies Combine Online and Offline Data to Learn About You? 9*, 23.

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (ed.), *Shaping Technology / Building Society: Studies in Sociotechnical Change* (pp. 225-258). The MIT Press.

Lee, W.-H., & Lee, R. B. (2017). Implicit Smartphone User Authentication with Sensors and Contextual Machine Learning. *2017 47th Annual IEEE/IFIP International*

Conference on Dependable Systems and Networks (DSN), 297–308. https://doi.org/10.1109/DSN.2017.24

Meng, W., Ding, R., Chung, S. P., Han, S., & Lee, W. (2016). The Price of Free: Privacy Leakage in Personalized Mobile In-App Ads. *Proceedings 2016 Network and Distributed System Security Symposium*. Network and Distributed System Security Symposium, San Diego, CA. https://doi.org/10.14722/ndss.2016.23353

Millett, L. I., Friedman, B., & Felten, E. (2001). Cookies and Web browser design: Toward realizing informed consent online. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 46–52. https://doi.org/10.1145/365024.365034

Ometov, A., Bezzateev, S., Mäkitalo, N., Andreev, S., Mikkonen, T., & Koucheryavy, Y. (2018). Multi-Factor Authentication: A Survey. *Cryptography*, *2*(1), 1. https://doi.org/10.3390/cryptography2010001

O'Reilly, T. (2007). *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software* (SSRN Scholarly Paper No. 1008839). https://papers.ssrn.com/abstract=1008839

Quinn, P., & Malgieri, G. (2021). The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework. *German Law Journal*, *22*(8), 1583–1612. https://doi.org/10.1017/glj.2021.79

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero Trust Architecture*. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-207

Sarkar, S., Choudhary, G., Shandilya, S. K., Hussain, A., & Kim, H. (2022). Security of Zero Trust Networks in Cloud Computing: A Comparative Review. *Sustainability*, *14*(18), 11213. https://doi.org/10.3390/su141811213

Sipior, Janice C., Ward, Burke T., & Mendoza, Ruben A. (2011). Online Privacy Concerns Associated with Cookies, Flash Cookies, and Web Beacons. *Journal of Internet Commerce*, *10*(1), 1–16. https://doi.org/10.1080/15332861.2011.558454

Stoyanovich, J., Howe, B., & Jagadish, H. V. (2020). Responsible data management. *Proceedings of the VLDB Endowment*, *13*(12), 3474–3488. https://doi.org/10.14778/3415478.3415570

Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, *34*(1), 134–153. https://doi.org/10.1016/j.clsr.2017.05.015

Vorakulpipat, C., Pichetjamroen, S., & Rattanalerdnusorn, E. (2021). Usable comprehensive-factor authentication for a secure time attendance system. *PeerJ Computer Science*, *7*, e678. https://doi.org/10.7717/peerj-cs.678

West, S. M. (2019). Data Capitalism: Redefining the Logics of Surveillance and Privacy. *Business & Society*, *58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Wirtz, J., Lwin, M. O., & Williams, J. D. (2007). Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management*, *18*(4), 326–348. https://doi.org/10.1108/09564230710778128