PROTECTING PRIVACY OF CRITICAL DATA IN APPLICATIONS

THE ROLE OF TECHNOLOGICAL POLITICS ON PRIVATE USER DATA

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
Harshil Pareek

December 09, 2022

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

Benjamin Laugelli, Department of Engineering and Society
Mark Sherriff, Department of Computer Science
Brianna Morrison, Department of Computer Science

Introduction

The notion of protecting user data has been the subject of controversy since the emergence of the internet. Specifically, the emergence of social media platforms, namely Facebook, has prompted users to question how their personal information is being managed on an administrative standpoint. From a simple Google search on recipes to seeing the same results on another webpage as an marketing advertisement, or receiving multiple fraudulent calls throughout the day and incurring costs on said calls, the lives of those that regularly access the world of the internet through technological means such as a phone or computer are never kept at the root of privacy and inclusion. Rather, it is these same corporations that created a marveling technology that also dismisses privacy laws and holds users with unlimited power and control.

During my internship over the summer, I observed firsthand how secure packets of data must be to ensure information cannot reach other parties, as this was also in a preliminary and highly insecure fashion in the grand scheme of things. Breaching this security protocol could result in the release of confidential user information, similar to the controversy corporations face in protecting user data. For the technical aspect of this project, I will use my experiences to investigate potential solutions to preventing confidential user information from being publicly accessed.

With more of a public eye on this pressing issue, more strides have to be made in order for users to truly feel safe regarding the use of their information on the internet. For the STS research aspect of this project, I will investigate the framework of actor-network theory to uncover the breach Facebook encountered with Cambridge America in violating their code of conduct in openly sharing sensitive user information for monetary gain.

Technical Project

This internship experience took place in the summer of 2022 with a technology company called SAIC that specializes in providing government services and other software products. The task at hand was to develop an API that takes in specified user data to generate a JWT (JSON Web Token) to grant a user designated permissions. This could entail a user to add specific data to the database, access an endpoint only accessible to them, or other features. With regards to the database, only administrative roles and the specific user can modify, delete, or create an entity based on their token permissions, but this leaves security implications and potential breaches, which is analogous to the socio-technical issue demonstrated in the STS Project. The databases were made locally using H2 and on a server using PostgreSQL.

Several C.S. courses I have taken helped prepare me for frameworks experienced in a professional setting. CS 3240 provided a collaborative sprint style project course that gave an early insight into frameworks and designing a web application from scratch, and CS 4750 went over the structure to a relational database which came in handy for my experience over the summer. Several soft skills that were encountered during this experience was largely communication through non-verbal and verbal means, along with understanding constructive criticism and moving forward. Through weekly meetings with fellow team members and supervisors, a professional means of communication was encouraged and developed over the course of the experience.

From my experiences with SAIC, some of the challenges met were encrypting user data and ensuring its privacy within the application. When a contractor decided to delete their entry from the database, we were met with a dilemma on how to effectively remove the traces of their existence to ensure their information will not be leaked. After much deliberation, we decided to

implement a machine learning model known as Machine Unlearning, a process that removes all instances of the entry from every aspect of the application. In *Towards Making Systems Forget with Machine Unlearning,* a process is specified that, "forgetting a data item now requires recomputing only a small number of terms" (Cao, Yang, 2015) by applying a deletion process with Machine Unlearning. With this system, our application was able to protect critical user information by removing them from existence.

Much of the curriculum in CS courses is overly technical, which is relevant for those seeking careers in specific fields but unnecessary for those pursuing more general positions in Software Engineering/Development. For students to do well in technical interviews, there should be more emphasis on algorithm and data structure courses, which may be solved by adding more classes and spreading them out throughout the 8 semesters as a C.S. student. The concept of Facilitation supports the argument that a student engages with material best when it is retaught over several years, as they only remember the material for three months (Shanley, Martin, 2022). This way, a student in the interview process that is also attending relevant courses that help solve technical questions can apply their current understanding of the course material to do well on interviews and have an incentive to study more diligently.

Some other strategies to improve the course structure for CS courses include making examinations occur online and more aligned with career aspirations and goals, as well as allowing students to be more independent with their work (McGee 2014). This could mean assignments aren't due by a set time and students can work on it for as long as they need, so long as they finish them all by a certain point, similar to the structure set forth in CS 4102. Another notable suggestion is incorporating co-op programs for students. Pair programming as a strategy could be made more effective through the incorporation of principles associated with cooperative

learning (Mentz, 2008). Giving students opportunities to showcase their acquired skills in a professional setting can provide more of a unique experience and overall development towards becoming a well-rounded developer.

STS Project

The implications of data breaches conducted by Facebook and Cambridge Analytica have greatly impacted all social media platforms on the internet today. To offer context, Facebook holds the claim that every user's personal information is promised to be kept private and away from marketing advertisers. In 2013, a third-party application hosted on Facebook called 'thisisyourdigitallife' was created which collected profile information of over 300,000 users that downloaded the app, which consequently led to the full breach of over 80 million users (Kozlowska, 2018). This influx of data was then sold to Cambridge Analytica, a firm that uses data to determine various behavioral traits and political affiliations. After Facebook uncovered these sequence of events, they demanded that the data be deleted but could not ensure that copies were not created, which resulted in the mega-corporation hiding this information from the public. Researchers found that 74% of Facebook users were unaware that Facebook maintained a list of their interests and traits, and 51% of them were uncomfortable with the exploitation of their data (Hitlin, Raine, Olmstead, 2019). They came to the conclusion that Facebook's "multicultural affinity" in aligning with users' cultures is a breach of privacy and against their own code of conduct. The clear invasion of privacy that Facebook holds in their algorithm underscores the importance of maintaining adequate communication between users in order for the network to succeed. It should be evident enough that user data should not be exploited and given to any third-party source, but most companies like Facebook stand to gain no money without

outsourcing their clientele data for advertisement revenue and other sources of income. Thus, the abuse of power that organizations, or in this case Facebook, hold restrists confidentiality and raises concern for the influx of new customers. By further assessing company policies related to privacy and analyzing technical methods on how user data is outsourced, I can deduce key insights into how several developmental mistakes led to worldwide societal consequences, thus showing the effects of violating the protection of user information. This case is most similar to the framework of Actor Network Theory, in which Facebook's network failed with respect to Cambridge Analytica by failing to privatize sensitive user information that Facebook prides themselves in protecting. I argue that the primary actors can still fulfill their financial means without compromising user confidentiality, which will in turn rebrand the company into an organization that values their customer and excels in generating unbiased algorithm technology.

Actor network theory emphasizes the ways in which technology and social networks are intertwined and how they influence each other. The theory argues that actors, both human and non-human, play a crucial role in shaping technology and social networks. These actors can include people, organizations, institutions, and even technology itself. Actor network theory seeks to understand how these actors interact and influence each other in order to better understand the complex networks that emerge from their interactions (Cressman, 2009). In order to support my argument that Facebook's network failed with respect to Cambridge Analytica, I will find additional sources to support the actor-network theory and the challenges faced with operating a social network of that stature while also complying with regulations. Additionally, I will cite scholarly articles that assess how critical user data can be rerouted to avoid being reached by a biased network like Cambridge Analytica for monetary gain.

Conclusion

The deliverable for the technical problem will be reconstructing course curriculum to best represent students who wish to jumpstart their careers with early internship experience or a full-time position with a strong background in Software Development, while also noting the importance of web security and protecting sensitive information. The STS research problem will investigate the social dilemna with selling private user information without consent and aiming to protect users with their data through more encryption and a process to erase their data immediately. The combined results should aim to address the issue regarding loosely protecting confidential information while also bringing about a new curriculum that promises early career development integrated into coursework.

Word Count: 1,636

Sources

Shanley, N., Martin, F., Hite, N., Perez-Quinones, M., Ahlgrim-Delzell, L., Pugalee, D., & Hart, E. (2022). Teaching Programming Online: Design, Facilitation and Assessment Strategies and Recommendations for High School Teachers. TechTrends : for leaders in education & training, 66(3), 483–494. https://doi.org/10.1007/s11528-022-00724-x

McGee, P. (2014). Blended Course Design. International Journal of Mobile and Blended Learning, 6(1), 33–55. https://doi.org/10.4018/ijmbl.2014010103

Wilson, Greg, et al. “Best Practices for Scientific Computing.” PLoS Biology, vol. 12, no. 1, 7 Jan. 2014, p. e1001745, journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001745, 10.1371/journal.pbio.1001745.

Kozlowska, Iga. “Facebook and Data Privacy in the Age of Cambridge Analytica - the Henry M. Jackson School of International Studies.” The Henry M. Jackson School of International Studies, 30 Apr. 2018, jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/.

E. Mentz, J. L. van der Walt & L. Goosen (2008) The effect of incorporating cooperative learning principles in pair programming for student teachers, Computer Science Education, 18:4, 247-260, DOI: 10.1080/08993400802461396

Y. Cao and J. Yang, "Towards Making Systems Forget with Machine Unlearning," 2015 IEEE Symposium on Security and Privacy, 2015, pp. 463-480, doi: 10.1109/SP.2015.35.

Cressman, D. (2009). A Brief Overview of Actor-Network Theory: Punctualization, Heterogeneous Engineering & Translation.

Hitlin, P., & Rainie, L. (2019, January 16). Facebook Algorithms and Personal Data. Pew Research Center: Internet, Science & Tech; Pew Research Center: Internet, Science &

Tech.
https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personal-data/

Social Network Theory, Cambridge Analytica and what it means for design. (2020, April 24). Uxdesign.cc.
https://uxdesign.cc/social-network-theory-cambridge-analytica-and-what-it-means-for-design-b241a495a1fd