FEDERATED MULTI-ARMED BANDIT PROBLEM

CONSUMER PERCEPTION OF PRIVACY

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
Aaron Parson

October 27, 2023

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

Rider Foley, Department of Engineering and Society

Rosanne Vrugtman, Department of Computer Science

Introduction

Privacy in the age of information has become a major concern for people, organizations, and government bodies. In response to this, we must rethink the rules governing the ethical responsibilities for data during information age (Zwitter, 2014). In the information age, the collection of data has speedily moved towards digital systems. Using these systems, we have started to collect more information than we have ever before. This massive encompassing collection of data has led to what is known as the phenomena of Big Data. The underlying idea of Big Data is to collect data on everything that can be collected. This provides an understanding of the world that is more closely related to nature than to what is deemed important to collect. However, there is a need for affirming ethics of usage. Using Big data for analysis can create improper correlations and generalizations over various marginalized groups. While there has been a call to action by different groups of people, it has been difficult for policy and technology to mitigate the growing privacy concerns as our world moves more towards digital. In the U.S the Federal Trade Commission has taken control over defining how to regulate data collection and inform consumers about data collection practices (Ohlhausen, 2014). While frameworks are in place, many companies have found various ways around these regulations allowing for data collection often to be unwilling. There is also a need to understand what privacy means to consumers. This is necessary for the FTC to take into account since its entire operation for protecting consumer privacy stems from consumers' desire for privacy. Yet, it has found that there is a misalignment between consumers' preferred privacy settings and what they will actually choose in practice. To improve privacy laws, understanding how consumers define privacy, view their actions in maintaining their privacy, and their rating of privacy enhancing technologies is crucial.

Last year, I conducted Machine Learning Research on Federated Learning. Federated Learning (FL) is an emerging technology that addresses various limitations in machine learning, including data sharing, privacy concerns, and infrastructure costs, while enhancing data diversity. Removing these limits on machine learning models will allow for better performance, thereby improving the potential for what is possible in AI and analytics fields. While this technology is able to aid in the battle in maintaining consumer privacy, it is one stepping stone in a much larger solution. In this paper, I want to explore my project in Federated Learning and how it can be used to help improve consumer privacy. Moreover, I want to gain a better understanding of how to improve consumers' perception of privacy and privacy enhancing technologies and how these technologies and laws can better be applied.

Federated Learning and Privacy:

FL is a distributed system of devices that works to solve machine-learning problems (Zhang & Bai et al 2021). The structure of the network includes multiple client devices, each running smaller machine-learning models on their locally collected data. These client devices will then send their machine-learning model updates to a server that aggregates the client updates into a global model. Once the aggregation is complete, the server sends the global model changes back to client devices, which integrate those updates into their own models. From this, they start training again on their local data and repeat the process. What makes this technology privacy preserving is that the data never leaves the device that it is collected on. While FL shows promise as a new technology, certain challenges need to be solved to make it more useful. These challenges include privacy protection, data sufficiency, and statistical heterogeneity.

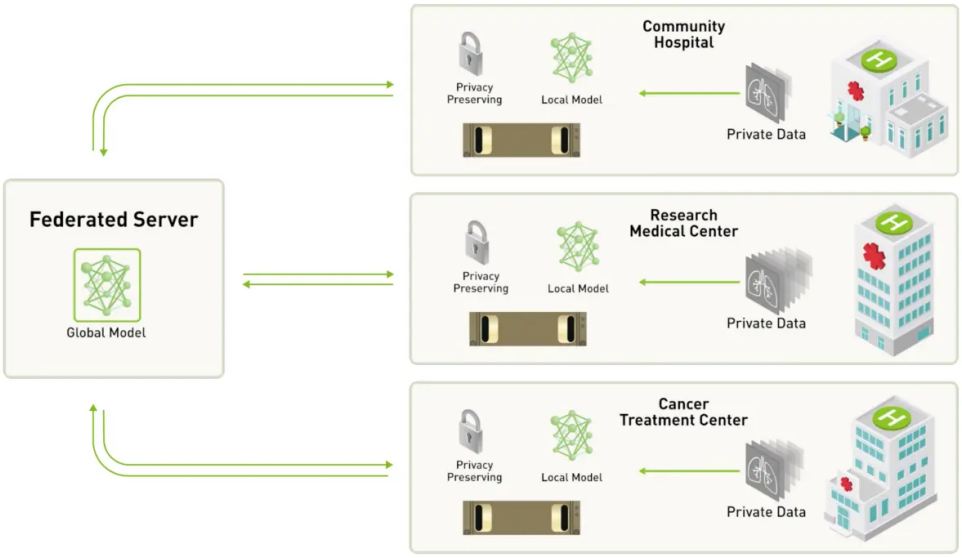

Figure 1: Example use of Federated Learning Model (Reike, 2019)

Furthermore, the problem that we are modeling in FL is the Multi-Armed Bandit problem (MAB). It is a reinforcement learning problem in which our goal is to find a slot machine out of a variable number of machines that has the true highest reward while minimizing the regret. The regret is the difference between the reward received and the true highest reward that accompanies searching for this slot machine. Placing the MAB problem in an FL context allows each of the 'players' in the problem to share information with each other, benefiting each player's ability to estimate the best arm to pull. This therefore creates a distributed recommending system in which privacy can be better protected (Shi & Shen 2021).

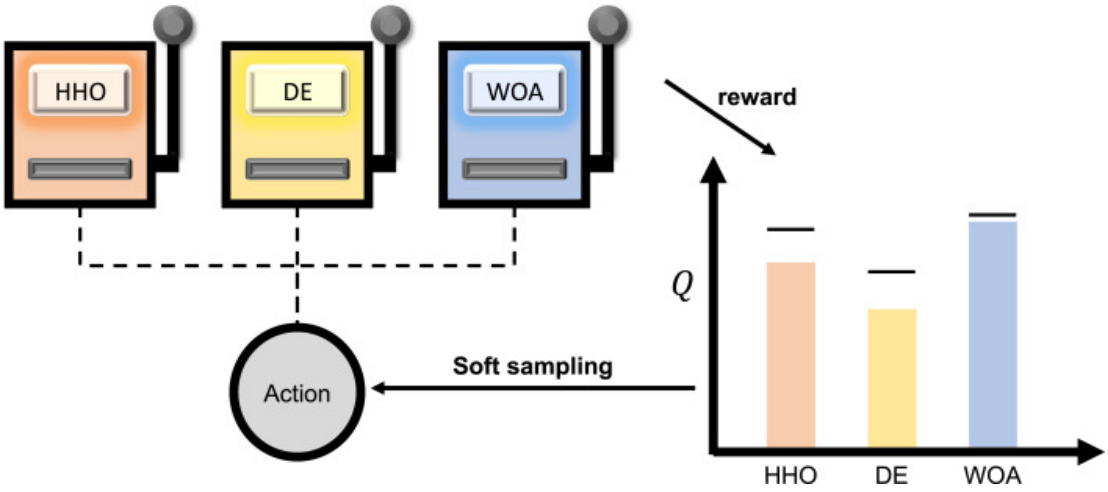

Figure 2: Example of Multi-arm Bandit simulation (Meidani & Barati Farmani, 2022)

Slot machines (left) are played and their rewards are recorded until the true highest average reward can be estimated.

The importance of Federated Learning is that it will be able to benefit various domains including advertisements, clinical trials, and finance. There are broad applications to the Federated MAB problem. This is mostly because FL rebuilds an already implemented system that has propagated through many fields. The combination of the MAB problem and FL will impact any type of system that recommends a product or caters content. FL can be a possible solution to the lag in legislation protecting customer privacy. While legislation is being created to address customer privacy problems, none has fully come to pass (Sherman, 2021). Many companies collect a multitude of data on customers, some of which customers are unaware of and unwilling to share. This can include medical information, financial history, location, and other sensitive data. Many of these companies sell data to other companies. However, this data is often used to tailor customer experience to a product or application. With the quick development of technology, legislation is unable to understand the various ways information is collected and

put protections on customer data. Since FL is privacy-preserving, this may provide a middle ground where companies can still collect information but not data that disregards client privacy.

Consumer Perception of Privacy

Utilizing the concept of technological momentum, as introduced by Hughes (1987), provides a valuable framework for comprehending how technology interfaces with people. Technological momentum characterizes the evolution and perpetuation of technology not as an independent entity but rather as a system that accumulates momentum from diverse conceptual influences during its development. In this way, the durability of the technology can highlight the trajectory of the system. One theme discussed, *soft determinism*, played a vital role in the creation of FL. Soft determinism depicts a larger entity or system that constricts a smaller entity to certain rules or guidelines. As we see more exploits of customer data, legislation has started restricting what data is allowed to be collected and shared. Regulations like the California Consumer Privacy Act, which gives customers the right to know what information is collected, and the termination of the EU-US Privacy Shield, which now restricts the flow of data between the U.S. and EU (Chalamala, 2022). This restriction of data collection required companies to develop new ways of collecting information while preserving privacy. Another theme discussed in Hughes is *transfer.* It is the reinvention of a product or system to adapt to different environmental conditions. In the digital age, the previous system of solitary global models did not take advantage of the computational power of the systems we carry around now. FL can take advantage of these smaller devices, often being able to build better models on individual devices than previous machine learning systems. Lastly, Hughes also discusses the limits of control, which can be defined as the social factors that limit the ability of technology to grow or progress.

Since FL is a network of systems, the quality of system maintenance is a social factor that can limit its abilities. The system is vulnerable to power outages, connectivity issues, and data entry errors, all which compound as the system scales.

In one sense, privacy can be defined through the lens of contextualized integrity. This is that people have the right to give or restrict access to their personal information (Goldfarb & Tucker, 2020). While seeing privacy as a right can show the agency in which we need to protect it, it fails to capture the abuse of privacy. Information has become a commodity (Smith & Shao, 2007). In the transaction of access to the system and information collection, data collectors have much more to gain (Zwitter, A. 2014).

One facet of this abuse happens in digital advertising, which creates an entire narrative and lifestyle associated with the product (Habibova, 2020). This subtly influences people's desires, aspirations, and consumer choices to conform to a certain lifestyle. Although this is not a wrongful manipulation in itself , with the evolution of the internet, the intensity of these manipulations have grown dramatically from linear to digital advertisements. While customer data is necessary for ecommerce platforms to develop usages of data has created a battle between privacy and innovation (Sarathy & Robertson, 2003). As innovation develops, the erosion of privacy will continue as companies strive to arrive at optimizing their platforms using techniques such as price differentiation (Odlyzko, 2003)

Some technologies can help bridge the gap between innovation and privacy. Some examples of these are Federated Learning, Differential Privacy, K-anonymity, and Homomorphic Encryption. Federated Learning which as described earlier is able to allow for Machine learning without data ever leaving the device and can be used with other privacy enhancing techniques (Li & Sharma et al., 2020). K-anonymity and Differential Privacy work to anonymize data while

still allowing for its use (Soria-Comas & Domingo-Ferrer, 2015). The latter ensures that individual data has no impact on the output of the data set, making it difficult for cyber criminals to obtain individual data through public output and background knowledge (Zhong, 2019)

Research and Methods:

While protecting consumer privacy will be an ongoing fight as technology advances, understanding consumer perceptions of privacy and privacy enhancing technology as well as their preference can be beneficial. Hughes describes how technology can gain or lose momentum as it progresses. This concept helps formulate the question of how consumer perception of privacy can assist and diminish the building of the momentum of various systems including social media and ecommerce platforms. Because of this, we will see how these perceptions held by consumers will also affect the patterns of technological evolution. This understanding gives insight into what aspects of privacy are viewed as important in interacting with different technology platforms. Therefore, companies understanding how consumers desire privacy can benefit their own platforms allowing for a win-win situation. The major stakeholders in this question are e-commerce and social media consumers. While this is an extremely broad group of people since there is a large number of people participating in ecommerce platforms, there is a positive effect. The understanding of consumer perceptions and desire in privacy and privacy enhancing technology requires deep interaction with a diverse group of consumers. To do this, I will construct a survey that will ask respondents to rate various views of their own privacy, the privacy of others, how they feel about privacy enhancing technology, and what improvements they want to see. These metrics can extend from scaling from 1-10, ranking of choices, and additional comments. These will then be compiled and analyzed to detect trends in

views of privacy and desires of privacy and be used in order to detect preferences. The survey will be distributed through word of mouth, emails, bulletin boards, and more.

Conclusion:

From the beginning of recorded history till 2003 only 5 billion gigabytes had ever been collected (Zwitter, 2014). Today, we collect 5 billion gigabytes of data every 10 seconds. While consumers want innovation and the comfort that comes with it, many also want to maintain their ability to control their privacy. Understanding how consumers desire their own privacy and what they feel comfortable with sharing is extremely important especially for marginalized groups of people. This project's goal is to understand this information and use it to provide insight into future regulations and legislation that can allow for companies and individuals to both benefit.

References

Bleier, A., Goldfarb, A., & Tucker, C. (2020). Consumer privacy and the future of data-based innovation and marketing. *International Journal of Research in Marketing*, *37*(3). https://doi.org/10.1016/j.ijresmar.2020.03.006

Chalamala, S. R., Kummari, N. K., Singh, A. K., Saibewar, A., & Chalavadi, K. M. (2022).Federated learning to comply with data protection regulations. CSI Transactions on ICT,10(1), 47–60. https://doi.org/10.1007/s40012-022-00351-0

Habibova, K. A. (2020, May 13). Digital Advertising and Digital Communication as a Means of Mass Manipulation. https://doi.org/10.2991/assehr.k.200509.101

Li, Z., Sharma, V., & Mohanty, S. (2020, April 2). Preserving Data Privacy via Federated Learning: Challenges and Solutions. Retrieved October 16, 2023, from ieeexplore.ieee.org website: https://ieeexplore.ieee.org/document/9055478

Meidani, K., Mirjalili, S., & Barati Farimani, A. (2022). MAB-OS: Multi-Armed Bandits Metaheuristic Optimizer Selection. *Applied Soft Computing*, *128*, 109452. https://doi.org/10.1016/j.asoc.2022.109452

Odlyzko, A. (2003b, July 27). Privacy, Economics, and Price Discrimination on the Internet. Retrieved October 15, 2023, from papers.ssrn.com website: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=429762

Ohlhausen, M. K. (2014). Privacy Challenges and Opportunities: The Role of the Federal Trade Commission. *Journal of Public Policy & Marketing*, *33*(1), 4–9. https://doi.org/10.1509/jppm.33.1.4

Rieke, N. (2019, October 13). What Is Federated Learning? | NVIDIA Blog. Retrieved October 27, 2023, from The Official NVIDIA Blog website: https://blogs.nvidia.com/blog/2019/10/13/what-is-federated-learning/

Sarathy, R., & Robertson, C. J. (2003). Strategic and Ethical Considerations in Managing Digital Privacy. *Journal of Business Ethics*, *46*(2), 111–126. https://doi.org/10.1023/a:1025001627419

Sherman, J. (n.d.). Weak US Privacy Law Hurts America's Global Standing. Retrieved October 27, 2023, from Wired website: https://www.wired.com/story/weak-us-privacy-law-hurts-americas-global-standing/

Shi, C., & Shen, C. (2021). Federated Multi-Armed Bandits. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(11), 9603–9611. https://doi.org/10.1609/aaai.v35i11.17156

Smith, R., & Shao, J. (2007). Privacy and e-commerce: a consumer-centric perspective. *Electronic Commerce Research*, *7*(2), 89–116. https://doi.org/10.1007/s10660-007-9002-9

Soria-Comas, J., & Domingo-Ferrer, J. (2015). Big Data Privacy: Challenges to Privacy Principles and Models. *Data Science and Engineering*, *1*(1), 21–28. https://doi.org/10.1007/s41019-015-0001-x

Thomas Parke Hughes. (1986). *The evolution of large technological systems*. Berlin Wzb.

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, *216*, 106775. https://doi.org/10.1016/j.knosys.2021.106775

Zhong, G. (2019). E-Commerce Consumer Privacy Protection Based on Differential Privacy. *Journal of Physics: Conference Series*, *1168*(3), 032084. https://doi.org/10.1088/1742-6596/1168/3/032084

Zwitter, A. (2014). Big Data ethics. *Big Data & Society*, *1*(2), 205395171455925. https://doi.org/10.1177/2053951714559253