Research Statement - Brian Jay Tang

Summary

My research aims to improve the current standards of privacy by leveraging machine learning (ML) and natural language processing (NLP) to help users easily understand and control the collection and sharing of their data. The tools I have developed can be leveraged by privacy professionals and privacy regulators to automatically monitor data flows and policy violations. I will discuss (1) why I chose to focus on privacy research and (2) how my past projects will enable me to conduct impactful research.

Motivation

Growing up, I would happily sign up and use free services without understanding that these services made money by collecting and selling personal information and products created from user data. Now, thousands of products leverage ML models trained on user data. Through my undergraduate research projects, I've learned that ML models are easily attacked, struggle with stereotypes and demographic fairness, and can leak user data. Recently, the use of ML has also sparked ethical debates about surveillance and power imbalances. Ultimately, data privacy lies at the heart of all these issues. Users have the right to understand and control their data, especially when this data can be used to discriminate or surveil people. Even with recent laws aimed at improving privacy control and transparency, it is unrealistic to expect users to track down hundreds of data processors and request data deletion. Likewise, transparency is unhelpful for users who must digest the thousands of privacy policies encountered every year. I, like millions of other users, felt hopeless knowing that my data has been sold and shared across thousands of data processors. Though now, as a researcher and a 2nd year Ph.D. student at the University of Michigan, I have dedicated my research to creating usable privacy-preserving technologies that scale to meet the challenge of data privacy.

Past Research

Through my previous research, I explored new properties and tradeoffs in face recognition and adversarial ML [1, 2], as well as the demographic fairness of such systems [3]. I helped create and evaluate NLP tools to automatically detect unsanctioned data collection activities [4–6]. I also designed and evaluated two user-centric privacy-enhancing tools for mobile devices and social robots [7, 8]. I intend to leverage my breadth of experience in ML, human-computer interaction, and security & privacy research by focusing on creating ML systems that improve users' understanding and control of their privacy.

Research Plans

For the remainder of my Ph.D., I intend to focus the depth of my research on creating usable privacy tools to aid users, privacy professionals, and regulators. For instance, I plan to create a tool to extract 3rd party data flows from privacy policies. Users could use this tool to understand what entities control their data and automatically request deletion of their data across all 3rd party platforms for a specific data type. Another project of particular interest to me is creating a tool capable of performing a dynamic privacy analysis for websites and apps to determine whether data collection is in accordance with stated privacy policies. I also plan to create a personalized privacy assistant which can analyze the privacy policies of any product/service and provide the user with notifications or suggestions for preserving their privacy.

References

[1] V. Chandrasekaran, C. Gao, **Brian Tang**, K. Fawaz, S. Jha, and S. Banerjee, "Face-off: Adversarial face obfuscation," in *21st Privacy Enhancing Technologies Symposium*, 2021. [Online]. Available: `https://arxiv.org/abs/2003.08861`.

[2] V. Chandrasekaran, **Brian Tang**, N. Papernot, K. Fawaz, S. Jha, and X. Wu, "Rearchitecting classification frameworks for increased robustness," 2020. arXiv: `1905.10900`. [Online]. Available: `https://arxiv.org/abs/1905.10900`.

[3] H. Rosenberg, **Brian Tang**, K. Fawaz, and S. Jha, "Fairness properties of face recognition and obfuscation systems," 2021. arXiv: `2108.02707`. [Online]. Available: `https://arxiv.org/abs/2108.02707`.

[4] D. Bui, **Brian Tang**, and K. G. Shin, “Do opt-outs really opt me out,” in *In Review: 29th ACM Conference on Computer and Communications Security 2022*, 2022.

[5] ——, “Detection of inconsistencies in privacy practices of browser extensions,” in *Conditional Accept: 43rd IEEE Symposium on Security and Privacy 2022*, 2022.

[6] ——, “Automatic detection of cookie consent violations,” in *In Review: 32nd USENIX Security Symposium 2023*, 2023.

[7] **Brian Tang** and K. G. Shin, “Real-time protection of mobile device screen information from shoulder surfing,” in *In Review: 32nd USENIX Security Symposium 2023*, 2023.

[8] **Brian Tang**, D. Sullivan, B. Cagiltay, V. Chandrasekaran, K. Fawaz, and B. Mutlu, “Confidant: A privacy controller for social robots,” in *17th ACM/IEEE International Conference on Human-Robot Interaction*, 2022. [Online]. Available: `https://arxiv.org/abs/2201.02712`.