Sweeney PIT-UN Final Report 2020

Final Report

Project Title: Technology Science Research Collaboration

Appendix (Investigation Plans)

Students who have used Investigation Plans produced by PI Sweeney have gone on to conduct research that was published on Technology Science. See summaries below.

"Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps" by Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney.

The authors tested 110 popular, free Android and iOS apps to look for apps that shared personal, behavioral, and location data with third parties. 73% of Android apps shared personal information such as email address with third parties, and 47% of iOS apps shared geo-coordinates and other location data with third parties. 93% of Android apps tested connected to a mysterious domain, safemovedm.com, likely due to a background process of the Android phone. They show that a significant proportion of apps share data from user inputs such as personal

information or search terms with third parties without Android or iOS requiring a notification to the user. http://techscience.org/a/2015103001/

"Venmo'ed: Sharing Your Payment Data With the World" by Aran Khanna.
The Venmo app allows people to pay each other online. The student created an extension that visualizes information Venmo makes publicly available. The author analyzed the transactions of 350,000 Venmo users and found that 74% had at least 5 public transactions, with 21% averaging a public transaction more than once a week. His extension can identify relationships between users, including how much time they spend together. It can also identify members of private social organizations, attendees of private events, and even users' food purchases. http://techscience.org/a/2015102901/

"De-anonymizing South Korean Resident Registration Numbers Shared in Prescription Data" by Latanya Sweeney and Ji Su Yoo.
South Korea's national identifier, the Resident Registration Number (RRN) includes encoded demographic information and a checksum with a publicly-known pattern. The authors conducted two de-anonymization experiments on 23,163 encrypted RRNs from prescription data of South Koreans. They demonstrate the data's vulnerability to de-anonymization by revealing all 23,163 unencrypted RRNs in both experiments. http://techscience.org/a/2015092901/

"Identity as a Service: Iceland's Kennitala and the Convergence of Identifier and Authenticator in Online Third Party Applications" by Gili Vidan (an HKS student).
Iceland's national identifier, the Kennitala (KT), is computed from one's date of birth and some random digits. The author found five Icelandic subjects online and was able to guess and verify their KT using a dating app. This experiment suggests that KT registry may be reverse-engineered and expose personal data on services that rely on the KT for authentication to imposters. http://techscience.org/a/2015092902/

"Only You, Your Doctor, and Many Others May Know" by Latanya Sweeney.
Washington State is one of 33 states that share or sell anonymized health records. The author conducted an example re-identification study by showing how newspaper stories about hospital visits in Washington State leads to identifying the matching health record 43% of the time. This study resulted in Washington State increasing the anonymization

protocols of the health records including limiting fields used for the re-identification study. http://techscience.org/a/2015092903/

"Defeating ISIS on Twitter" by Batsheva Moriarty.

The author evaluated 1.5 million tweets from 1,500 ISIS-affiliated Twitter accounts to determine if they were humans or bots. She compared ISIS tweets to a control group of 700,000 non-ISIS Arabic tweets. ISIS tweets exhibited unique, un-unified tweet, retweet, and favoriting patterns suggesting that the accounts are controlled by humans. http://techscience.org/a/2015092904/

"Finding Fraudulent Websites Using Twitter Streams" by Daniel Rothchild.

The author developed a monitoring program that searches Twitter in real time for tweets with potentially suspicious links. The program found more than 70,000 suspicious tweets in 24 hours, with 56% of the tested links appearing fraudulent. http://techscience.org/a/2015092905/ Note: Most fraud cases at the FTC result from receiving a significant number of consumer complaints, and often the time delay allows fraudsters to relocate before action occurs. The FTC now uses this kind of approach to detect fraud in real-time, while the fraud occurs.

"Facebook's Privacy Incident Response: a study of geolocation sharing on Facebook Messenger" by Aran Khanna.

In 2012, a media outlet reported that Facebook Messenger shared personal geolocations by default. In 2015, my demonstration displayed Facebook's shared data on a map; it was downloaded over 85,000 times. After 9 days of news coverage, Facebook released an update that requires a user's permission to share geolocations. http://techscience.org/a/2015081101/

"Did You Really Agree to That? The Evolution of Facebook's Privacy Policy" by Jennifer Shore and Jill Steinman.

The authors examined changes to Facebook's Privacy Policy from 2005 to 2015 using the relevant parts of the 2008 Patient Privacy Rights (PPR) framework. They found that Facebook's score declined by 2015 in 22 of 33 measures of privacy protection and transparency on a 5-point scale. The measures included the extent of internet monitoring, informing users about what is shared with third parties, clearly identifying data used for

profiling, and giving users choices in privacy settings.
http://techscience.org/a/2015081102/

"Price Discrimination in The Princeton Review's Online SAT Tutoring Service" by Keyon Vafa, Christian Haigh, Alvin Leung, and Noah Yonack.

The authors tested whether customers are seeing the same price for SAT tutoring on The Princeton Review's website. They searched the website from 33,000 ZIP codes across the US. They found three different prices depending on the ZIP code input seemingly on a regional basis.
http://techscience.org/a/2015090102/

"Unintended Consequences of Geographic Targeting" by Jeff Larson, Surya Mattu, and Julia Angwin of ProPublica.

The authors analyzed the price variations for an online SAT tutoring service offered by The Princeton Review. Their analysis showed that Asians were 1.8 times as likely to be quoted a higher price than non-Asians. People who live in high-income ZIP codes were twice as likely to be quoted a higher price than lower income residents.
http://techscience.org/a/2015090103/

"Larger Issuers, Larger Premium Increases: Health insurance issuer competition post-ACA" by Eugene Wang and Grace Gee.

Health insurance plans on 34 state exchanges are studied for pricing changes from 2014 to 2015. The largest insurance company in each state on average increased their rates 75% more than smaller insurers in the same state. The largest insurance companies do not appear to be paying for higher medical costs per premium dollar versus smaller insurers in the reported experience period of 2013.
http://techscience.org/a/2015081104/
Note: The finding caught the attention of the American Hospital Association, the Congressional Subcommittee on Regulatory Reform, Commercial and Antitrust Law, and the Centers for Medicare and Medicaid, which sent an advisory requesting further justification for rate increases.

"Who's Paying More to Tour These United States? Price Differences in International Travel Bookings" by Michael Rose and Mohammed Rahman.

The authors tested whether customers from around the world see the same price online when searching for U.S. hotel rooms and rental cars.

They simulated connecting online from 30 countries around the world to travel site Kayak.com. Simulated customers in five locations, including Hong Kong and Australia, were quoted hotel prices significantly above the global average. Prices shown to domestic customers in the U.S. were slightly below the average. http://techscience.org/a/2015081105/

"The Model Minority? Not on Airbnb.com: A Hedonic Pricing Model to Quantify Racial Bias against Asian Americans" by John Gilheany, David Wang, and Stephen Xi.

The authors tested if Asians receive lower prices on Airbnb's vacation rental website. They identified 101 White and Asian hosts on Airbnb in Oakland and Berkeley in April 2015. They found that on average Asian hosts earn $90 less per week or 20% less than White hosts for similar rentals. http://techscience.org/a/2015090104/
Note: These results led to further inquiries by advocacy groups and Airbnb adopting a new platform that recommends prices.

"No More Secrets: Gmail and Facebook can determine your political values" by Melissa Hammer.

The author created separate Facebook and Gmail accounts based on political preference for Democrats or Republicans. On Facebook, the two profiles received different suggestions while on Gmail similar ads appeared. http://techscience.org/a/2015090105/

Finally, here is the publication of student work from a student-produced Investigation Plan.

"Deepfake Bot Submissions to Federal Public Comment Websites Cannot Be Distinguished from Human Submissions" by Max Weiss.

The author created a computer program (a bot) that generated and submitted 1,001 deepfake comments regarding a Medicaid reform waiver to a federal public comment website, stopping submission when these comments comprised more than half of all submitted comments. He then formally withdrew the bot comments. When humans were asked to classify a subset of the deepfake comments as human or bot submissions, the results were no better than would have been gotten by random guessing. His work demonstrated that Federal public comment websites currently are unable to detect Deepfake Text once submitted, but technological reforms (e.g., CAPTCHAs) can be implemented to help

prevent massive numbers of submissions by bots. This work just published last December, but it has already inspired many improvements at public comment websites.
https://techscience.org/a/2019121801/