Senator Klobuchar Questions for the Record: Privacy, Technology, and the Law Subcommittee of the Judiciary Committee hearing on "Protecting Americans' Private Information from Hostile Foreign Powers"

Questions for Prof. Susan Landau, Professor of Cybersecurity Law and Policy at Tufts Fletcher School

Health data raises significant concerns because of its sensitivity. Many entities that collect this kind of data maintain that it is anonymized, but that is unlikely to resolve privacy concerns when the data can easily be re-associated to specific people.

- Is data anonymization sufficient to protect sensitive health data?
- How does the availability of data from publicly available sources contribute to the likelihood that health data can be re-associated with a particular person?

Response:

Sensitive health data under the control of health providers is protected under HIPAA; as I am not a health care expert, I cannot comment on how well this data is anonymized. But I will note that over a decade ago, it became clear how easy it was to reidentify people whose records were "anonymized" by comparing the public data to information stored in other databases.

The classic example is the 1997 reidentification of Governor William Weld from a database from the Group Insurance Commission of records of hospital visits. Names had been omitted from the database, but gender, zip codes, and birthdates were there. Cross checking with Cambridge public voter records revealed the governor's information, including his diagnosis and prescription.[1] By a decade later, such types of so-called "anonymized" data had become increasingly easy to re-identify due to multiple forms of data collection (and publication of same).[2]

[1] D.C. Barth-Jones, The 'Re-Identification' of Governor William Weld's Medical Information: a critical re-examination of health data identification risks and privacy protections, then and now, Available from SSRN: http://ssrn.com/abstract=2076397.

[2] See, e.g., Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization,* UCLA Law Review (57:6), 2010, 1701.

There is health care data is collected outside of HIPAA. Call detail records can reveal information about a person's health situation, such as relapsing-remitting MS or cardiac arrythmia.[3] Location information that can reveal visits to medical facilities, drug treatment centers, abortion clinics, etc. Detailed location information about an individual can be gleaned from the GPS data on their phone, which users may supply to apps. It can also be determined from other sensors on their device, include accelerometers, gyroscopes, and magnetometers; again, this information may be available to apps—and thus to data brokers in order to enable the online advertising system. Other types of health information can also be discerned through monitoring the user's activity: accelerometers and gyroscopes can give away whether they are walking, biking, etc.

This data is readily associated with an individual. It does not fall under HIPAA, and, except for "unfair or deceptive acts or practices," there are no federal controls on its use at present. Thus, the information above could be shared with health insurance companies, for example.

[3] Jonathan Mayer, Patrick Mutchler, and John C. Mitchell, *Evaluating the privacy properties of telephone metadata*, PNAS 113 (20), 5540.

Senator Marsha Blackburn Questions
Senate Judiciary Subcommittee on Privacy, Technology, and the Law Hearing
"Protecting Americans' Private Data from Hostile Foreign Powers"
Wednesday, September 14, 2022
4:00 PM

Question for Professor Landau and Mr. Pottinger: I understand that foreign countries and companies under state ownership or control can use click through ads to gain access to platform user data. For example, in places like China, where Twitter is blocked, the CCP could use ads on the platform to collect data on users evading the block through virtual private networks.

A. *In your experience, is this a typical practice that happens at global tech platforms?*

I'm not able to answer the question; I do not know the extent to which Chinese companies place ads on sites used in foreign countries nor do I know the extent to which Chinese users employ VPNs to gain access outside of China.

B. *What types of data can be collected through such a process and what kinds of harms could it cause?*

In order to target ads, the online advertising system accumulates data about users: their age, religion, income, activities, interests, etc., and then sells ads based on these population "segments"—e.g., urban mothers aged 35-45 with one or two small children. A user receiving an ad targeted at a particular segment will reveal that the user is likely to have these characteristics. But the ad provides more information than that. The device may provide an ad ID; on iPhones this can be done only by opting into such collection; on Androids, the user must explicitly opt out of the ad ID collection. This ad ID, while ostensibly pseudononymous, effectively allows identifying the user. Various harms could then ensue, since the CCP-generated ads may be placed on sites that users are not permitted to access.