PUBLIC RELEASE: 19-JAN-2017

Your 'anonmyized' web browsing history may not be anonymous

PRINCETON UNIVERSITY, ENGINEERING SCHOOL

IMAGE: ARVIND NARAYANAN, ASSISTANT PROFESSOR OF COMPUTER SCIENCE, PRINCETON UNIVERSITY. **view more ›**

CREDIT: FRANK WOJCIECHOWSKI

Raising further questions about privacy on the internet, researchers from Princeton and Stanford universities have released a study showing that a specific person's online behavior can be identified by linking anonymous web browsing histories with social media profiles.

"We show that browsing histories can be linked to social media profiles such as Twitter, Facebook or Reddit accounts," the researchers wrote in a paper scheduled for presentation at the 2017 World Wide Web Conference Perth, Australia, in April.

"It is already known that some companies, such as Google and Facebook, track users online and know their identities," said Arvind Narayanan, an assistant professor of computer science at Princeton and one of the authors of the research article. But those companies, which consumers choose to create accounts with, disclose their tracking. The new research shows that anyone with access to browsing histories -- a great number of companies and organizations --can identify many users by analyzing public information from social media accounts, Narayanan said.

"Users may assume they are anonymous when they are browsing a news or a health website, but our work adds to the list of ways in which tracking companies may be able to learn their identities," said Narayanan, an affiliated faculty member at Princeton's Center for Information Technology Policy.

Narayanan noted that the Federal Communications Commission recently adopted privacy rules for internet service providers that allow them to store and use consumer information only when it is "not reasonably linkable" to individual users.

"Our results suggest that pseudonymous browsing histories fail this test," the researchers wrote.

In the article, the authors note that online advertising companies build browsing histories of users with tracking programs embedded on webpages. Some advertisers attach identities to these profiles, but most

promise that the web browsing information is not linked to anyone's identity. The researchers wanted to know if it were possible to de-anonymize web browsing and identify a user even if the web browsing history did not include identities.

They decided to limit themselves to publicly available information. Social media profiles, particularly those that include links to outside webpages, offered the strongest possibility. The researchers created an algorithm to compare anonymous web browsing histories with links appearing in people's public social media accounts, called "feeds."

"Each person's browsing history is unique and contains tell-tale signs of their identity," said Sharad Goel, an assistant professor at Stanford and an author of the study.

The programs were able to find patterns among the different groups of data and use those patterns to identify users. The researchers note that the method is not perfect, and it requires a social media feed that includes a number of links to outside sites. However, they said that "given a history with 30 links originating from Twitter, we can deduce the corresponding Twitter profile more than 50 percent of the time."

The researchers had even greater success in an experiment they ran involving 374 volunteers who submitted web browsing information. The researchers were able to identify more than 70 percent of those users by comparing their web browsing data to hundreds of millions of public social media feeds. (The number of original participants in the study was higher, but some users were eliminated because of technical problems in processing their information.)

Yves-Alexandre de Montjoye, an assistant professor at Imperial College London, said the research shows how "easy it is to build a full-scale 'de-anonymizationer' that needs nothing more than what's available to anyone who knows how to code."

"All the evidence we have seen piling up over the years showing the strong limits of data anonymization, including this study, really emphasizes the need to rethink our approach to privacy and data protection in the age of big data," said de Montjoye, who was not involved in the project.

Besides Narayanan, the researchers involved in the project included: Jessica Su, Ansh Shukla, and Sharad Goel of Stanford. Support for the project was provided in part by the National Science Foundation. The researchers thanked Twitter for supporting the project by providing access to the Gnip search API.

###

Media Contact

John Sullivan
js29@princeton.edu
609-258-4597

@eprinceton