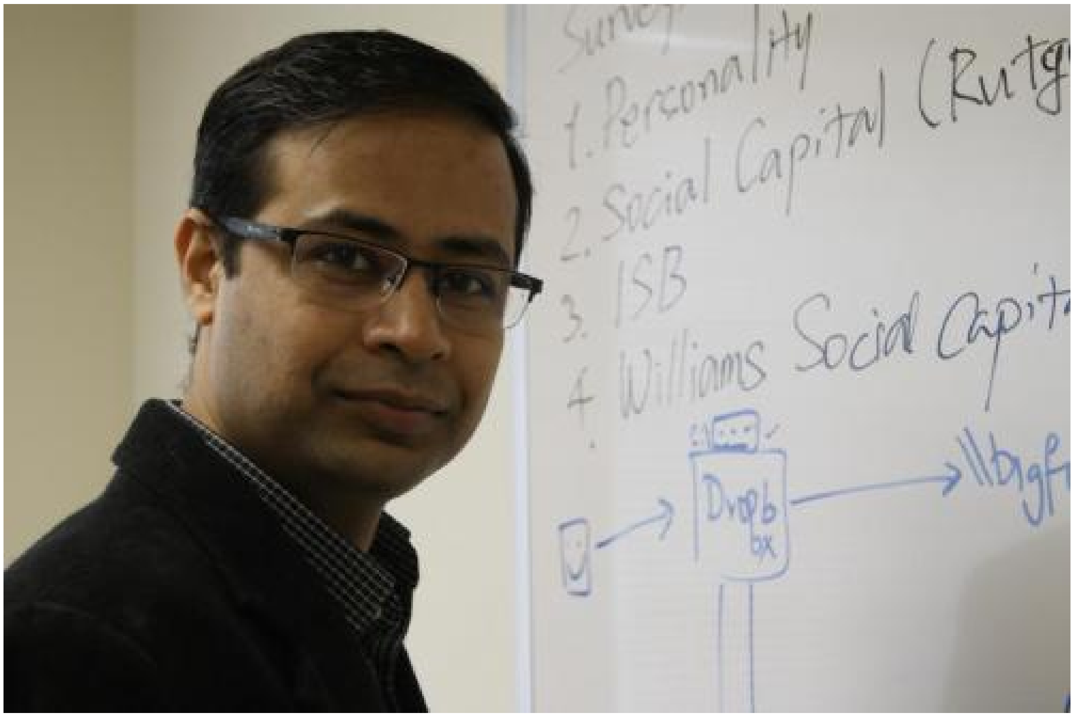
PHYS.ORG

Research reveals we may need a new definition for privacy

March 6 2015, by Carol Peters



Vivek Singh's research paper on credit card metadata dispels the belief that anonymous personal data is truly private.

If you still believe your personal credit information is truly private, newly released research by a Rutgers professor may lead you to reconsider.

Rutgers Today spoke with Vivek Singh, assistant professor of Library and Information Science in the School of Communication and Information, about his paper, "Unique in the Shopping Mall: On the Reidentifiability of Credit Card Metadata." Published Jan. 30, 2015 in *Science*, the study focused on the question: "Out of an anonymous set of credit card data from millions of people, how easily can you find one person?" The main finding: Information gleaned from just four transactions can uniquely identify a person most of the time (90 percent).

The anonymous set of credit card data from 1.1 million transactions considered in the study did not include any names, account numbers or addresses – anything that would be considered an easy identifier. Removing such obvious identifiers is required by the U.S. Privacy Act and the European Union Data Privacy Directive. However, this study shows that it is possible to reidentify users quite easily even after removing information such as names and account numbers.

The paper, as Singh explains, shows why "we need to rethink the ideas we have about privacy."

Singh began working on this research with his colleagues while he was a postdoctoral fellow at MIT Labs. He also is currently a visiting professor at MIT.

What are the implications of your finding that it's relatively easy to identify a person who thinks he's anonymous?

Singh: Recently, large-scale data about financial transactions, health records and taxi trips have been made public for research, which can help fight diseases, or yield better urban planning. They did not have any direct personal information like name, social security number or address, and hence the individuals in these datasets were assumed to not be reidentifiable. This research shows that this expectation is not true in practice.

This finding has large implications for what we trust are private sources of information about us. It is relatively easy for anyone, with just a bit of information, to find out very private details about our lives.

We therefore need to redefine our current definition of privacy. We stress that our research does not prove that we all have any lesser privacy than before or that privacy is gone. But our research does show that we do need to rethink how we measure and define it.

Please explain how movement patterns are linked to your ability to reidentify people in your study.

Singh: Our research reveals that people's movements and behavior patterns uniquely identify them, and thus human behavior becomes very unique very quickly.

You can think of it this way: Let's say you are standing in a line in front of a shop on Monday and you see that there are 100 other people shopping on the same day at the same shop. When you go to another shop on Tuesday, you will probably only see 30 people from the previous shop. On Wednesday you will only see three of the same people, and by the time you visit the fourth shop on Thursday, you are probably the only one from the original 100 you saw at the shop on Monday. Thus human behavior naturally becomes more unique over space and time. This uniqueness is what makes us reidentifiable.

How does your research reveal that credit card information is not as private as we may think?

Singh: The key idea is that if someone were to acquire simple information, such as the date, location, time and the dollar amount of just four credit card transactions, they could reidentify 90 percent of users in sets of data that do not include any names, account numbers, home addresses, phone numbers or other obvious identifiers.

When we say reidentify we have this in mind: If you know four shops where a person went on some days, 90 percent of the time he was the only one that visited these four shops on these days. Therefore, a person is very likely to be reidentified if just four points are known.

Such external points of information (the date, location, time and the dollar amount of just four credit card transactions) could be discovered by anyone reading a tweet such as "Enjoying coffee that cost just a buck in Central Park, 28 Jan 2015," or by combining different sources of partial information.

How does your study show the link between social media and privacy?

Singh: The study does not focus on social media, but it does make us realize the secondary effects of sharing information across social media. Just sharing information about four such transactions on social media, e.g. the location and time of shopping for shoes on sale on a Sunday, could be used by somebody else to identify a person in a large database of users.

Social media makes it easier for users to share information about their movements, which makes it even easier for their private information to be found and revealed.

Here is an example. Recently, a separate study used a set of data kept by a taxi company in Manhattan to reveal the names of celebrities who used the service, what they paid and how much they tipped the driver.

How can this situation be remedied to increase privacy?

Singh: There are many good and positive reasons that scientists may want to share their sets of data publicly. For one thing, sharing information such as medical records can help eradicate disease. Taxi records could be useful to city planners, for example, to show traffic patterns. In some cases scientists would like to release their data in order to make their findings more generally known or to allow a whole community of researchers to work on them. However, our research has shown that doing this comes with the risk of private information being compromised.

Scientists need to rethink the way such sets of data can be made anonymous so as to allow for analysis and scientific research while still maintaining the privacy of individuals. For example, a user's information could be decentralized (split up and shared in different locations), released in small increments over time rather than as a complete set of data, or a known amount of noise (irrelevant or incorrect information) could be added to the data so it makes individuals indistinguishable from others.

More information: "Unique in the shopping mall: On the reidentifiability of credit card metadata." *Science* 30 January 2015: Vol. 347 no. 6221 pp. 536-539 DOI: 10.1126/science.1256297

Provided by Rutgers University

Citation: Research reveals we may need a new definition for privacy (2015, March 6) retrieved 26 May 2024 from https://phys.org/news/2015-03-reveals-definition-privacy.html