http://www.wsj.com/articles/metadata-can-expose-persons-identity-even-when-name-isnt-1422558349

U.S.

Metadata Can Expose Person's Identity Even Without Name

New Analytic Formula Identifies People Without Names, Account Numbers

By **ROBERT LEE HOTZ**

Updated Jan. 29, 2015 7:55 p.m. ET

Your shopping habits can expose who you are even when you are just one of a million nameless customers in a database of anonymous credit-card records, according to a new study that shows how so-called metadata can be used to circumvent privacy protections in commercial and government databases.

Researchers at the Massachusetts Institute of Technology, writing Thursday in the journal Science, analyzed anonymous credit-card transactions by 1.1 million people. Using a new analytic formula, they needed only four bits of secondary information—metadata such as location or timing—to identify the unique individual purchasing patterns of 90% of the people involved, even when the data were scrubbed of any names, account numbers or other obvious identifiers.

"We are introducing a way to find what you need to identify an individual—how much data makes you stand out in the crowd," said MIT data analyst Yves-Alexandre de Montjoye, who led the study. "This touches on the fundamental limit of anonymizing data."

Researchers at MIT analyzed anonymous credit-card transactions by 1.1 million people and could identify 90% of the people involved using just four pieces of secondary information such as location or timing. *PHOTO: THE RECORD OF BERGEN COUNTY/ASSOCIATED PRESS*

Researchers drew on records of purchases over a period of three months by shoppers at 10,000 stores, provided by an unnamed bank in an undisclosed country. Each transaction was time-stamped with the day of purchase and linked to a shop.

Even with so little to go on, they could readily identify a person's unique purchasing pattern. "We did everything you would need to do to find a person in the data, but we did not try to attach a specific name to it," Mr. de Montjoye said.

After

WSJ.D

WSJ.D is the Journal's home for tech news, analysis and product reviews.

- Alibaba Sales Disappoint; Firm Rebuts China Claims (http://www.wsj.com/articles/alibabas-revenue-rises-on-mobile-growth-1422533361)
- Can Apple Keep Up Its Growth or Did Customers Just Want Big Phones? (http://www.wsj.com/articles/can-apple-keep-up-its-growth-spurt-1422493028)
- Samsung Sold Fewer Smartphones and Profit Fell 27% (http://www.wsj.com/articles/samsung-electronics-fourth-quarter-profit-falls-27-1422489917)
- Microsoft's New Outlook for iOS and Android Is Worth a Try (http://blogs.wsj.com/personal-technology/2015/01/29/microsofts-new-outlook-for-ios-and-android-is-worth-a-try/)

isolating a purchasing pattern, researchers said, an analyst could find the name of the person in question by matching their activity against other publicly available information such as profiles on Linkedin and Facebook, Twitter messages that contain time and location information, and social-media "check-in" apps such as Foursquare.

The finding is the latest indication that we expose more about ourselves than we may realize through the patterns of our digital transactions, from smartphone-app usage to mobile calling data. About 60% of payments in the U.S. are made with credit cards, and mobile payments run at about $1 billion a year, the researchers note.

The new technique is likely to be of interest to the many research firms, advertisers, retailers and trade associations that build and buy extensive data bases to track customers and better target advertising.

The MIT research has shown "it is very, very, very difficult to remove any ability to identify people in these data sets, especially financial data," said Joseph Hall, chief technologist at the Center for Democracy & Technology, a nonprofit that studies privacy and data issues. He wasn't involved in the project. "Data brokers who buy and collect very large quantities of information like this have the ability to take thousands of data points and pin those on individuals."

The finding also will add to debates over the bulk collection of personal data by government surveillance programs, which cross-match cellphone metadata and electronic databases, including credit-card information, several data-privacy experts said.

"We think of metadata as being not as important as content, but it turns out to be remarkably revelatory," said cybersecurity analyst Susan Landau at Worcester Polytechnic Institute in Massachusetts, who wasn't involved in the project. "Little bits of data combined with the data we shed in other places really create portraits."

Last November, for example, ride-share company Uber disclosed it had combined its customer records of late-night trips in major cities with local crime reports to calculate the likelihood that its weekend riders were visiting prostitutes.

In the same vein, researchers at the U.K.'s Cambridge University reported in 2013 that the pattern of "likes" posted by people on Facebook unintentionally exposed their political and religious views, drug use, divorce and sexuality. Earlier this month, psychologists at the University of California in Riverside reported that those "likes" were a more-accurate measure of someone's personality than the assessment of their close friends.

In the MIT study, researchers could tell women and men apart just by how they lingered at different shops. They also could pick out people in higher income brackets. The method can be applied to almost any data set that records behavior, they said.

"We have these unique patterns that identify us," said MIT computer scientist Alex Pentland, who worked on the study. "They show up in any data where there is diverse behavior that changes over time. Your pattern will be different and I could identify you," he said.

Write to Robert Lee Hotz at sciencejournal@wsj.com