NATURE | NEWS

People identified through credit-card use alone

Analysis suggests that making data anonymous is not enough to protect consumers.

Boer Deng

29 January 2015

Joe Raedle/Getty

Scrubbing names and addresses from transaction data does not render it truly anonymous.

Figuring out what data can be used to identify someone has long befuddled those tasked with keeping information private. Sometimes, the data sets they use to obscure underlying identities fail to do so. A computer-science graduate student at Carnegie Mellon University in Pittsburgh, Pennsylvania, once uncovered the medical history of then-Massachusetts governor William Weld from de-identified insurance records, for example[1].

So it is not particularly shocking that Yves-Alexandre de Montjoye, a computer-security researcher at the Massachusetts Institute of Technology (MIT) in Cambridge, and his colleagues managed to identify one individual from a sea of 'anonymized' credit-card data.

Their research, published on 29 January in *Science*[2], analysed credit-card transaction information, or 'metadata', from 1.1 million shoppers in countries that are members of the Organisation for Economic Co-operation and Development (OECD). Although names, addresses and other information directly linked to card owners had been scrubbed from the data set, de Montjoye and his colleagues could pick out 90% of individuals if they knew the date and location of just four of their credit-card transactions.

Even when researchers only had estimates of time and location of a purchase to within a few days or neighbourhood blocks, they could still identify cardholders. Women and affluent consumers were the easiest to pick out, likely because their shopping habits were more diverse, yielding transaction patterns that were comparatively more distinct and traceable.

Alex Pentland, a data-security researcher at MIT and a co-author of the study, reckons that metadata with certain properties — such as having geographical location attached — are harder to anonymize. Previously, de Montjoye had shown that mobile-phone metadata, which relays information about the location of a call, could also be used to reveal callers whose identities were thought to be unknowable[3].

Behaviour trackers

Previous analyses have demonstrated that other patterns of behaviour are unique enough for individuals to be identified using metadata. In 2006, for example, web-search data from the digital-media company AOL allowed the *New York Times* to track down an anonymous user; in 2009, researchers showed that they could find the name and even the likely political views of particular Netflix

subscribers using supposedly anonymized search data the company released for a contest[4].

This has led some to conclude that for metadata to be useful, complete anonymity is impossible. “No one has figured out how to do it,” says Pam Dixon, executive director of the World Privacy Forum, a digital-policy research group in San Diego, California. “It will only get harder to make data and metadata anonymous as more information becomes available that will tag people’s unique behaviours.”

If this is the case, strengthening security laws on protecting the release of metadata would lower the risk of people identifying consumers for nefarious purposes. The standards for protection and enforcement vary across OECD member nations, however. Even within the United States, no national law for reporting consumer data breaches exists.

Some progress has been made in updating regulations. In the wake of a lawsuit stemming from the Netflix debacle, the US Federal Trade Commission, a government consumer-protection agency, established new standards for retailers to secure their data. And this week, the US House of Representatives committee responsible for overseeing commerce met to discuss federal data breach legislation. Data custodians, customers and criminals take note.

Nature | doi:10.1038/nature.2015.16817

References

1. Sweeney, L. *Simple Demographics Often Identify People Uniquely Data Privacy Working Paper 3* (Carnegie Mellon Univ., 2000); available at http://go.nature.com/v9pvv1
2. de Montjoye, Y.-A., Radaelli, L., Singh, V. K. & Pentland, A. *Science* **347**, 536–539 (2015).
3. de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. & Blondel, V. D. *Sci. Rep.* **3**, 1376 (2013).
4. Narayanan, A. & Shmatikov, V. *Preprint at* http://arxiv.org/abs/cs/0610105 (2006).