COMMENTARY

Transactional information is remarkably revelatory

Susan Landau[a,1]

When Edward Snowden revealed that the United States government had been collecting domestic communications metadata in bulk, the administration responded that there was no great concern. The data were only how long, when, and which number called which, not what participants said. Two days after the secret Foreign Intelligence Surveillance Court order ordering the collection was made public, President Obama said,

> When it comes to telephone calls, nobody is listening to your telephone calls. That's not what this program is about. As was indicated, what the intelligence community is doing is looking at phone numbers and durations of calls. They are not looking at people's names, and they're not looking at content. (1)

Revealing What You Do

However, as "Privacy properties of telephone metadata" by Mayer et al. (2) demonstrates, communications metadata provides much more private information than the President's statement indicated. Traffic analysis, studying who talks to whom and when, has long been the backbone of signals intelligence. There are many examples showing the value of these call detail records (CDRs) outside the intelligence context. Service providers use CDRs to determine who uses what products, and thus predict where and what type of services will be needed. In 2003 Cortes, Pregibon, and Volinsky showed how to use calling patterns to determine which new financial accounts are likely to be fraudulent (3). During the Iraq insurgency, the US military used CDRs to recognize burner phones as they came online.

As a result of always being with their owners, mobile phones provide detailed maps of users' lives and have become a valuable tool in computational social science research (4). Data from phones—specifically where people are—is being used to map people's movements during a disaster (5), to learn how to halt the spread of infectious diseases (6), and so forth. CDRs can also answer more mundane questions. In 2012, the phone company Orange made 5 mo of Cote d'Ivoire mobile phone metadata available as part of a research challenge (7); researchers showed that calling patterns even help in planning bus routes (8).

This, then, is the underlying context for the PNAS paper by Mayer et al. (2). Communications transactional information was long believed to be personally revelatory, but proof was lacking. The paper's authors rather convincingly supply that proof. Mayer et al. show that from CDRs, it can be determined that someone is suffering from a multiple sclerosis relapse, having cardiac arrhythmia problems, seeking to buy an automatic rifle, intending to start a marijuana-growing venture, or having an abortion.

Protecting users' privacy should be paramount in studies such as the one by Mayer et al. (2). The authors carefully informed their volunteer participants of information being collected, enabling participants to remove data even after collection.

Partially as a result of the Snowden disclosures, communications surveillance law is in flux (the other reason for the flux is the shift to IP-based communications). Thus, this paper (2) addressing the privacy content of CDRs comes at an important time and is likely to have a significant impact.

Content, Metadata, and the Law

In discussing communications surveillance, the specific situation matters. Mayer et al. (2) focus on the United States, and I will as well.

US jurisprudence treats communications metadata—dialing, routing, addressing, and signaling (DRAS)—differently from communications content. In 1967, in *Katz v. United States*, the Supreme Court established that even when speaking in so visible a space as a public phone booth, a person is entitled to Fourth Amendment protections against unreasonable search (9). The 1968 Omnibus Crime Control Act ("Wiretap Act") consequently established a warrant procedure for government wiretapping in criminal investigations (10). In 1979, in *Smith v. Maryland*, the Court ruled that dialing information was not deserving of the same stringent protections (11). Thus, under *Katz v. United States*

[a]Department of Social Science and Policy Studies, Worcester Polytechnic Institute, Worcester, MA 01609

Author contributions: S.L. wrote the paper.
The author declares no conflict of interest.

See companion article on page 5536.

[1]Email: susan.landau@privacyink.org.

and *Smith v. Maryland* wiretapping requires a "super warrant" (so called because requirements for obtaining wiretap warrants are more stringent than for a regular search warrant), whereas DRAS can be collected under lesser court orders. Orders are targeted; wiretap warrants must specify the subject of the order, whereas collection of DRAS must be tied to a particular, ongoing investigation.

The National Security Agency's (NSA) collection of communications metadata was bulk, meaning the majority of records were not targets of an investigation. Because such acquisition was not particularized, the legal issues are substantively different from the *Smith v. Maryland* case (11). As Mayer et al. (2) note, there have been conflicting court decisions on the legality of bulk collection of domestic metadata.

The court order on bulk collection raised concern over how many subscriber records of nonsuspects were being accessed by the NSA. Here I take exception to one statement in the excellent paper by Mayer et al. (2). The authors state, "[O]ur results strongly suggest that until 2013, analysts had legal authority to access telephone records for the majority of the entire US population" (2). That statement may be factually accurate but it is misleading. NSA policies apparently constrained the agency's use of the bulk communications data. [There were various mechanisms, including normally going two "hops" from the target and use of a "defeat list" that eliminated high-volume numbers from such chaining (12, 13). Mayer et al. (2) calculate 25,000 accesses of the records from a single target, so the 300 accesses of the database in 2012 (14) should lead to accessing records of under 7.5 million Americans.] That said, the authors' (2) misstatement is minor and has no impact on the conclusions of this important and valuable work.

Legal Changes Are Brewing

We are at a complex juncture regarding use of communications metadata. In the wake of the Snowden disclosures, President Obama ordered an advisory committee review of the NSA's activities, which recommended that bulk telephony metadata collection continue but with data stored at private providers or a third party (15). An independent government agency, the Privacy and Civil Liberties Oversight Board, concluded that the bulk metadata collection program provided little value while having serious implications for privacy and civil liberties; it recommended ending the program (16).

A National Academies study committee examined technical alternatives to accomplish the goals of bulk collection (disclosure: I served on this committee) (17). Because bulk collection involves storage before someone is designated a target, the collection provides a view into the past. Thus, it enables, for example, finding "alternate identifiers" (e.g., names, phone numbers) that were used before targets were subjects of investigation. The committee concluded there was no technical substitute for some aspects of bulk collection.

Policy and law have changed following the Snowden disclosures. In January 2014, the President limited results of a metadata query to two hops (18). In June 2015, Section 215 of the *USA PATRIOT Act*, used to authorize bulk collection, was replaced by the *USA FREEDOM Act*, which permits government access of CDRs from service providers.

The Snowden disclosures are likely to have contrary impacts when it comes to metadata. One impact has been to create greater controls on bulk collection and use, but another may be to increase use of CDRs.

After the disclosures, US technology companies, shocked by the extent of NSA collection against their customers, began to secure their internal workings (19) and their products. This security included providing end-to-end encryption for an increasing number of communications modalities (e.g., email, WhatsApp, and so forth) as well as securing devices by default; but this complicates law-enforcement investigations. The companies' efforts drew a response from the FBI, which has been seeking legislation controlling encryption's deployment. A number of the FBI's expected allies argued against the FBI position. Ex-officials from the Departments of Defense and Homeland Security, and from the NSA, believe that current threats make widely available, uncompromised encryption a necessity (20, 21). Still, all is not lost for law enforcement.

Mayer et al. rather convincingly supply that proof. They show that from CDRs, it can be determined that someone is suffering from a multiple sclerosis relapse, having cardiac arrhythmia problems, seeking to buy an automatic rifle, intending to start a marijuana-growing venture, or having an abortion.

Since the 1990s, the NSA has increasingly faced encrypted communications, yet "From that time until now, NSA has had better 'sigint' [signals intelligence] than any time in history," according to former NSA Director Mike McConnell (22). NSA former General Counsel Stewart Baker explained, "Metadata absolutely tells you everything about somebody's life. If you have enough metadata you don't really need content" (23). If content is encrypted—and increasingly it will be—the government may push for greater access to metadata.

Although the government increasingly recognizes the value of metadata, new technology complicates collection. Beginning with *Katz v. United States* and *Smith v. Maryland*, US law drew a sharp distinction between strong protections afforded to communications content and weaker ones for communications metadata. In recent work, colleagues and I have shown that the logic behind this breaks down for IP communications (24). Much of the law is inapplicable, and fixing the situation will likely require more than piecemeal change through the courts. Instead, new surveillance law will need to take new communication technologies into account.

One issue will likely be data aggregation, the ability to amass large amounts of data for analysis. This is already on the Supreme Court's radar. In 2012, Supreme Court Justice Alito asked, "Is it possible to imagine a case in which a constable secreted himself somewhere in a coach and remained there for a period of time in order to monitor the movements of the coach's owner?" in considering the data accumulated by a GPS device over 28 days (25). In 2014, the Supreme Court observed that, "A person might carry in his pocket a slip of paper reminding him to call Mr. Jones; he would not carry a record of all his communications with Mr. Jones for the past several months, as would routinely be kept on a phone" (26).

The paper by Mayer, Mutchler, and Mitchell (2) is important. It establishes, beyond any doubt, that telephone metadata is remarkably revelatory. In revising communications surveillance law, Congress will have to decide new balances between law-enforcement needs and privacy. Legislation will have to take the privacy results of this paper into account.

1 White House (2013) Statement by the President, June 7, 2013. Available at https://www.whitehouse.gov/the-press-office/2013/06/07/statement-president. Accessed April 9, 2016.
2 Mayer J, Mutchler P, Mitchell JC (2016) Evaluating the privacy properties of telephone metadata. *Proc Natl Acad Sci USA* 113:5536–5541.
3 Cortes C, Pregibon D, Volinsky C (2003) Computational methods for dynamic graphs. *J Comput Graph Stat* 12(4):950–970.
4 Blondel V, Decuyper A, Krings G (2015) A survey of results on mobile phone datasets analysis. *EPJ Data Science* 4:10.
5 Bengtsson L, Lu X, Thorson A, Garfield R, von Schreeb J (2011) Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Med* 8(8):e1001083.
6 Wesolowski A, et al. (2014) Commentary: Containing the Ebola outbreak—The potential and challenge of mobile network data. *PLoS Curr*, 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.
7 Blondel V, et al. (2013) Mobile Phone Data for Development: Analysis of Mobile Phone Datasets for the Development of Ivory Coast. *Selected Contributions to the D4D Challenge Sponsored by Orange, May 1–3*. Available at https://perso.uclouvain.be/vincent.blondel/netmob/2013/D4D-book.pdf. Accessed April 14, 2016.
8 Berlingerio M, et al. (2013) AllAboard: A system for exploring urban mobility and optimizing public transport using cellphone data. *Proceedings of NetMob, May 1, 2013* (MIT Media Lab, Cambridge, MA).
9 Katz v. United States, 389 US 347 (1967).
10 Omnibus Crime Control Act of 1968, Pub. L. No. 90-357, §§ 2510-2520, 82 Stat. 197, 211-225 (1968); current version at 18 U.S.C. §§ 2510-2530 (2003).
11 Smith v. Maryland, 442 US 735 (1979).
12 US Foreign Intelligence Court (2007) Docket PRTT[XX], Declaration of [XXX] Chief, Special Foreign Intelligence Surveillance Act Oversight and Processing, Oversight and Compliance, Signals Intelligence Directorate, Natl. Sec. Agency, Classified January 8, 2007. Available at www.clearinghouse.net/chDocs/public/NS-DC-0064-0006.pdf. Accessed April 10, 2016.
13 NSA (2011) OVSC1205 Special Training on FISA (Analytical) and OVSC1206 Special Training on FISA (Technical), Module 4: Access, Sharing, Dissemination, and Retention under the BR and PR/TT FISC Orders, Version 28 (Final), October 17, 2011. Available at https://www.aclu.org/files/assets/NSA%20Course%20Materials%20-%20Module%204.pdf. Accessed April 10, 2016.
14 Administration White Paper "Bulk Collection of Telephony Metadata under Section 215 of the USA PATRIOT Act," August 9, 2013, p. 4. Available at perma.cc/8RJN-EDB7. Accessed April 11, 2016.
15 President's Review Group on Intelligence and Communications Technologies (2013) *Liberty and Privacy in a Changing World*. Available at https://www.whitehouse.gov/sites/default/files/docs/2013-12-12_rg_final_report.pdf. Accessed April 11, 2016.
16 Privacy and Civil Liberties Oversight Board (2014) *Report on the Telephone Records Program Conducted under Section 215 of the USA PATRIOT Act and on the Operations of the Foreign Intelligence Surveillance Court*. Available at https://www.pclob.gov/library/215-Report_on_the_Telephone_Records_Program.pdf. Accessed April 11, 2016.
17 Committee on Responding to Section 5(d) of Presidential Policy Directive 28: The Feasibility of Software to Provide Alternatives to Bulk Signals Intelligence Collections, National Research Council (2015), *Bulk Collection of Signals Intelligence: Technical Alternatives* (National Academies Press, Washington, DC).
18 White House (2014) FACT SHEET: The Administration's Proposal for Ending the Section 215 Bulk Telephony Metadata Program. March 27. Available at https://www.whitehouse.gov/the-press-office/2014/03/27/fact-sheet-administration-s-proposal-ending-section-215-bulk-telephony-m. Accessed April 11, 2016.
19 Sanger D, Perlroth N (2014) Internet giants erect barriers to spy agencies. *New York Times*, Section A, pp 1.
20 McConnell M, Chertoff M, Lynn W (2015) Why the fear over ubiquitous data encryption is overblown. *Washington Post*. Available at https://www.washingtonpost.com/opinions/the-need-for-ubiquitous-data-encryption/2015/07/28/3d145952-324e-11e5-8353-1215475949f4_story.html. Accessed July 28, 2015.
21 Thiessen M (2016) Gen. Michael Hayden on Apple, the FBI, and data encryption. Available at https://www.aei.org/publication/gen-michael-hayden-on-apple-the-fbi-and-data-encryption/. Accessed March 23, 2016.
22 Nakashima E (2015) Former national security officials urge government to embrace rise of encryption. *Washington Post*. Available at https://www.washingtonpost.com/world/national-security/former-national-security-officials-urge-government-to-embrace-rise-of-encryption/2015/12/15/3164eae6-a27d-11e5-9c4e-be37f66848bb_story.html. Accessed April 17, 2016.
23 Rusbridger A (2013) The Snowden leaks and the public. *New York Rev Books*, Available at www.nybooks.com/articles/2013/11/21/snowden-leaks-and-public/. Accessed April 17, 2016.
24 Bellovin S, Blaze M, Landau S, Pell S (2016) It's too complicated: The technological implications of IP-based communications on content/non-content distinctions and the third-party doctrine. *Harv J Law and Tech*, in press.
25 Jones v. United States, 615 F. 3d 544, Justice Alito, concurring, p.3 (2012).
26 Riley v. California, 573 US 2473 (2014).