Written Testimony of Edward W. Felten
Professor of Computer Science and Public Affairs, Princeton University

United States Senate, Committee on the Judiciary
Hearing on
Continued Oversight of the Foreign Intelligence Surveillance Act
October 2, 2013

Chairman Leahy, Ranking Member Grassley, and members of the Committee, I thank you for the opportunity to testify about technical issues related to surveillance.

My name is Edward W. Felten. I am a Professor of Computer Science and Public Affairs at Princeton University. I also serve as the founding Director of the Center for Information Technology Policy, an interdisciplinary research and teaching center at Princeton that focuses on public policy issues relating to computers and the Internet. My primary field is computer science, and my main research areas are computer security and privacy, and Internet technologies.

Throughout my career, I have worked to help policymakers respond effectively to technological change. In 2011-12 I served as the first Chief Technologist at the Federal Trade Commission. I have testified several times at Senate and House hearings. I am a member of the National Academy of Engineering and the American Academy of Arts and Sciences.

Today, I will provide an overview of the tools and methods that computing technology can bring to the broad collection and analysis of metadata. I am not an expert on the law and I offer no opinion on the legal status of any program. Nor do I presume to say how best to balance the legitimate goals of conducting foreign intelligence surveillance against the legitimate goals of protecting privacy and promoting civil liberties. I hope that my testimony will help you appreciate the power of metadata and control its use appropriately, consistent with the need for effective foreign intelligence.

Metadata can now yield startling insights about individuals and groups, particularly when collected in large quantities across the population. It is no longer safe to assume that this "summary" or "non-content" information is less revealing or less sensitive than the content it describes. Just by using new technologies such as smart phones and social media, we leave rich and revealing trails of metadata as we move through daily life. Many details of our lives can be gleaned by examining those trails. Taken together, a group's metadata can reveal intricacies of social, political, and religious associations. Metadata is naturally organized in a way that lends itself to analysis, and a growing set of computing tools can turn these trails into penetrating insights. Given limited analytical resources, analyzing metadata is often a far more powerful analytical strategy than investigating content: It can yield far more insight with the same amount of effort.

Advances in technology have transformed the role and importance of metadata. When focused on intelligence targets, metadata collection can be a valuable tool. At the same time, unfocused collection of metadata on the American population gives government access to many of the same sensitive facts about the lives of ordinary Americans that have traditionally been protected by limits on content collection. Metadata might once have seemed much less informative than content, but this gap has narrowed dramatically and will continue to close.

Today's hearing is a vital step in a process that must continue. Technical expertise is essential for effective oversight of these technologically complex programs, and I would respectfully urge you to consider how best to integrate technical expertise into the oversight system. The United States has the world's strongest and deepest community of technical experts. This community is eager to contribute constructively to the national discussion.

The NSA Is Collecting Massive Amounts of Telephony Metadata

On June 5, 2013, *The Guardian* disclosed an order issued by the Foreign Intelligence Surveillance Court ("FISC") pursuant to Section 215 of the Patriot Act (the "Verizon Order").[1] This order compelled Verizon to produce to the NSA on "an ongoing daily basis . . . all *call detail records* or 'telephony metadata' created by Verizon for communications (i) between the United States and abroad; or (ii) wholly within the United States, including local telephone calls." Director of National Intelligence (DNI) James R. Clapper subsequently acknowledged the authenticity of the Verizon Order.[2] Officials also acknowledged that the NSA's acquisition of call detail records extends to the country's three largest phone companies: Verizon, AT&T, and Sprint[3]. Because these companies provide at least one end of the vast majority of calls in this country, these statements suggest that the NSA is maintaining a record of the metadata associated with nearly every telephone call originating or terminating in the United States.

This is a large volume of data. Assuming that there are approximately 3 billion calls made every day in the United States, and that each call record takes approximately 50

[1] Secondary Order, *In re Application of the FBI for an Order Requiring the Production of Tangible Things from Verizon Bus. Network Servs., Inc. on Behalf of MCI Commc'n Servs., Inc. d/b/a Verizon Bus. Servs.*, No. BR 13-80 at 2 (FISA Ct. Apr. 25, 2013), *available at* http://bit.ly/11FY393.

[2] James R. Clapper, *DNI Statement on Recent Unauthorized Disclosures of Classified Information*, Office of the Director of National Intelligence (June 6, 2013), http://1.usa.gov/13jwuFc.

[3] *See* Siobhan Gorman et al., *U.S. Collects Vast Data Trove*, WALL ST. J., June 7, 2013, http://on.wsj.com/11uDoue ("The arrangement with Verizon, AT&T and Sprint, the country's three largest phone companies means, that every time the majority of Americans makes a call, NSA gets a record of the location, the number called, the time of the call and the length of the conversation, according to people familiar with the matter. . . . AT&T has 107.3 million wireless customers and 31.2 million landline customers. Verizon has 98.9 million wireless customers and 22.2 million landline customers while Sprint has 55 million customers in total.").

bytes to store, the mass call tracking program collects about 140 gigabytes of data every day, or about 50 terabytes of data each year. Assuming that a page of text takes two kilobytes of storage, the program collects the equivalent of about 70 million pages of information every day, or about 25 billion pages every year.

The Verizon Order requires the production of "call detail records" or "telephony metadata." According to the order itself, that term encompasses, among other things, the originating and terminating telephone number and the time and duration of any call. Call detail records also typically include information about the location of the parties to the call.[4]

Although this latter definition of "call detail information" includes data identifying the location where calls are made or received, I will not address mobile phone location information in this testimony. While I understand that senior intelligence officials have asserted that they have the legal authority under Section 215 to collect mobile phone location information, they have stated that the NSA is not collecting phone location information "under this program."[5]

The information acquired from Verizon also includes "session identifying information"—*e.g.*, originating and terminating telephone number, International Mobile Subscriber Identity (IMSI) number, and International Mobile station Equipment Identity (IMEI) number. These are unique numbers that identify the user or device that is making or receiving a call. Although people who want to evade surveillance can make it difficult to connect these numbers to their individual identities, for the vast majority of ordinary Americans these numbers can be connected to the specific identity of a person.

The information acquired from Verizon also includes the "trunk identifier" of telephone calls. This provides information about how a call was routed through the phone network, which naturally reveals information about the location of the parties. For example, even if the NSA never obtains cell site location information about a call,[6], trunk identifier

---

[4] *See* 47 C.F.R. § 64.2003 (2012) (defining "call detail information" as "[a]ny information that pertains to the transmission of specific telephone calls, including, for outbound calls, the number called, and the time, location, or duration of any call and, for inbound calls, the number from which the call was placed and the time, location, or duration of any call").

[5] *See* Siobhan Gorman & Julian E. Barnes, *Officials: NSA Doesn't Collect Cellphone-Location Records*, WALL ST. J., June 16, 2013, http://on.wsj.com/13MnSsp; Pema Levy, *NSA FISA Metadata Surveillance: Is The Government Using Cell Phones To Gather Location Data?*, INT'L BUS. TIMES, Aug. 2, 2013, http://bit.ly/18WKXOV.

[6] Cell site location information ("CSLI") reflects the cell tower and antenna sector a phone is connected to when communicating with a wireless carrier's network. Most carriers log and retain CSLI for the start and end of each call made or received by a phone, and some carriers log CSLI for text messages and data connections as well. Wireless carriers can also obtain CSLI by "pinging" a phone whenever it is turned on, even if it is not engaged in an active call. The precision of CSLI varies according to several factors, and "[f]or a typical user, over time, some of that data will inevitably reveal locational precision approaching that of GPS." *The Electronic*

information revealing that a domestic call was carried by a cable from Hawaii to the mainland United States will reveal that the caller was in the state of Hawaii at the time the call was placed.

Although officials have stated that the orders issued under the telephony metadata program do not compel the production of customers' names, it would be easy for the NSA to correlate many telephone numbers with subscriber names using publicly available sources. I understand that federal agencies also have available a number of legal tools to compel service providers to produce their customer's information, including their names, without probably cause or judicial preclearance.[7]

Metadata Is Easy to Analyze

Telephony metadata is easy to aggregate and analyze because it is, by its nature, *structured data*. Telephone numbers are standardized, and are expressed in a predictable format: in the United States, a three digit area code, followed by a three digit central office exchange code, and then a four digit subscriber number. Likewise, the time and date information associated with the beginning and end of each call will be stored in a predictable, standardized format.

By contrast, the contents of calls are unstructured. Some people speak English, others Spanish, French, Mandarin, or Arabic. Some speak using street slang or a pidgin dialect, which can be difficult for others to understand. Conversations lack a common structure: Some people get straight to the point, others engage in lengthy small talk. Speakers have different accents, and exhibit verbal stutters and disfluencies. Although automated transcription of speech has advanced, it is still a difficult and error-prone process.

The structured nature of metadata makes it easy to analyze massive datasets using sophisticated data-mining and link-analysis programs. That analysis is greatly facilitated by technological advances over the past decades in computing, electronic data storage, and digital data mining. Those advances have radically increased our ability to collect, store, and analyze personal communications, including metadata.

Further, the massive increases in electronic storage permit us to maintain, cheaply and efficiently, vast amounts of data. The ability to preserve data on this scale is, by itself, an unprecedented development—making possible the maintenance of a digital history that was not previously within the easy reach of any individual, corporation, or government.

Communications Privacy Act (ECPA), Part 2: Geolocation Privacy and Surveillance: Hearing Before the Subcomm. on Crime, Terrorism, Homeland Sec. & Investigations of the H. Comm. On the Judiciary, 113th Cong. (2013) (statement of Matt Blaze, Associate Professor, University of Pennsylvania), http://1.usa.gov/1awvgOa.

[7] *See* 18 U.S.C. § 2709 (national security letter); 18 U.S.C. § 2703(c), (d) (court order for records concerning electronic communication service).

This newfound data storage capacity has led to new ways of exploiting the digital record. Sophisticated computing tools permit the analysis of large datasets to identify embedded patterns and relationships, including personal details, habits, and behaviors. As a result, individual pieces of data that previously carried less potential to expose private information may now, in the aggregate, reveal sensitive details about our everyday lives—details that we had no intent or expectation of sharing.

IBM's Analyst's Notebook and Pen-Link are two such computing tools. Both are widely used by law enforcement and intelligence agencies for this purpose.[8]

IBM's Analyst's Notebook product is a multi-purpose intelligence analysis tool that includes specific telephony metadata analysis features, which are "routinely" used to analyze large amounts of telephony metadata.[9] IBM even offers training courses entirely focused on using Analyst's Notebook to analyze telephone call records.[10]

Pen-Link is a tool that is purpose-built for processing and analyzing surveillance data. It is capable of importing subscriber Call Detail Record ("CDR") data from the proprietary formats used by the major telephone companies,[11] it can import and export call data to several federal surveillance databases,[12] as well as interact with commercial providers of public records databases such as LexisNexis. Pen-Link can perform automated "call

---

[8] *Public Safety & Law Enforcement Operations*, International Business Machines (last visited Aug. 22, 2013), http://ibm.co/1avGItq ("IBM® i2® solutions help law enforcers to turn huge volumes of crime data into actionable insights by delivering tools for tactical lead generation, intelligence analysis, crime analysis and predictive analysis."); *see also Defense and National Security Operations,* International Business Machines (last visited Aug. 22, 2013), http://ibm.co/18nateN ("IBM i2 solutions for military and national security organizations have been used across the world to process and analyze the vast quantities of information that they collect, to generate actionable intelligence and to share insights that help identify, predict and prevent hostile threats."); *see also* Pen-Link, *Unique Features of Pen-Link v8* at 16 (April 17, 2008), http://bit.ly/153ee9g ("Many U.S. Federal Law Enforcement and Intelligence agencies have acquired agency-wide site license contracts for the use of Pen-Link in their operations throughout the United States…Pen-Link systems are also becoming more frequently used by U.S. intelligence efforts operating in several other countries.").

[9] *Case Studies: Edith Cowan University, IBM i2 Solutions Help University Researchers Catch a Group of Would-Be Hackers*, International Business Machines (Mar. 27, 2013), http://ibm.co/13J2036 ("Analyzing this volume of data is nothing new to many law enforcement users who routinely analyze tens of thousands of telephone records using IBM® i2® Analyst's Notebook®.").

[10] *Course Description: Telephone Analysis Using i2 Analyst's Notebook*, International Business Machines (last visited Aug. 22, 2013), http://ibm.co/1d5QlB8 ("This intermediate hands-on 3-day workshop focuses on the techniques of utilizing i2 Analyst's Notebook to conduct telephone toll analysis…Learn to import volumes of call detail records from various phone carriers, analyze those records and identify clusters and patterns in the data. Using both association and temporal charts, discover how to use different layouts and more advanced tools to analyze telephonic data quickly and effectively.").

[11] *See* Pen-Link, *Unique Features of Pen-Link v8* at 4 (Apr. 17, 2008), http://bit.ly/153ee9g (describing the capability to import 170 different data formats, used by phone companies to provide call detail records).

[12] *Id.* at 4.

pattern analysis," which "automatically identifies instances where particular sequences of calls occur, when they occur, how often they occur, and between which numbers and names."[13] As the company notes in its own marketing materials, this feature "would help the analyst determine how many times Joe paged Steve, then Steve called Barbara, then Steve called Joe back."[14]

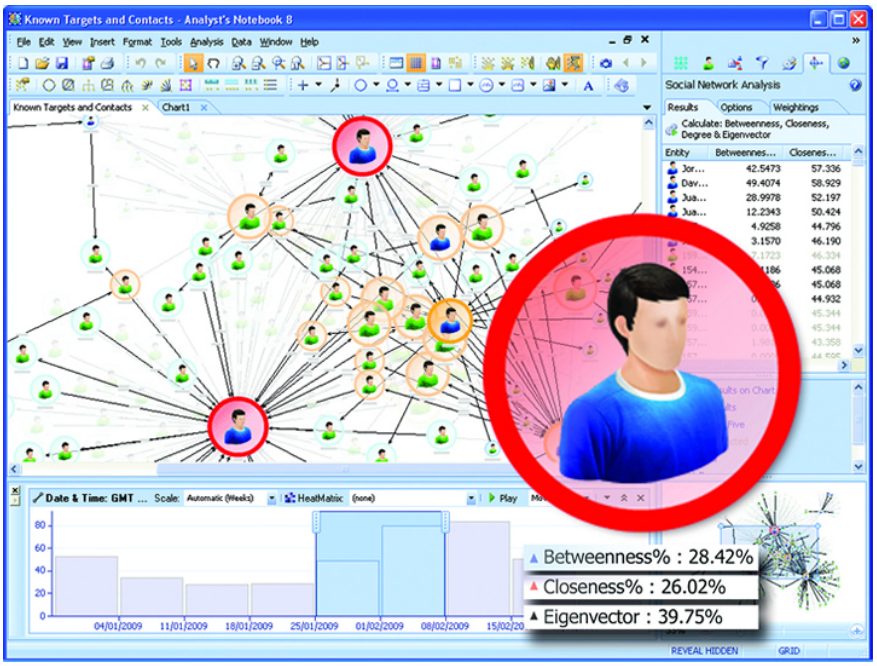

Figure 1: Screenshot of IBM's Analyst's Notebook.[15]

The contents of calls are far more difficult to analyze in an automated fashion due to their unstructured nature. The NSA would first have to transcribe the calls and then determine which parts of the conversation are interesting and relevant. Assuming that a call is transcribed correctly, the NSA must still try to determine the meaning of the conversation: When a surveillance target is recorded saying "the package will be delivered next week," are they talking about an order they placed from an online retailer, a shipment of drugs being sent through the mail, or a terrorist attack? Automatically parsing and interpreting such information, even with today's most sophisticated computing tools, is exceptionally difficult. To do so in an automated way, transcribing and data-mining the contents of hundreds of millions of telephone calls per day is an even more difficult task.

It is not surprising, then, that intelligence and law enforcement agencies often turn first to metadata. Examining metadata is generally more cost-effective than analyzing content. Of course, the NSA will likely still have analysts listen to every call made by the

[13] *Id.* at 7.
[14] *Id.*
[15] Image taken from *Data Analysis and Visualization for Effective Intelligence Analysis*, International Business Machines (last visited Aug. 22, 2013), http://ibm.co/16qT3hw.

highest-value surveillance targets, but the resources available to the NSA do not permit it to do this for all of the calls of 300 million Americans.

Americans Inevitably Create Metadata That Can Reveal Sensitive Details of Their Lives

Over the last three decades, and especially with the widespread adoption of mobile phones in the past decade, our reliance on telecommunications has significantly increased. Mobile phones are today ubiquitous, and their use necessarily requires reliance on a service provider to transmit telephone calls, text messages, and other data. These communications inevitably produce telephony metadata, which is created whenever a person places a call. There is no practical way to prevent the creation of telephony metadata, or to erase it after the fact. The only reliable way to avoid creating such metadata is to avoid telephonic communication altogether.

As a general matter, it is practically impossible for individuals to avoid leaving a metadata trail when engaging in real-time communications, such as telephone calls or Internet voice chats.

After decades of research (much of it supported by the U.S. government), there now exist many tools that individuals and organizations can use to protect the confidentiality of their communications content. Smartphone applications are available that let individuals make encrypted telephone calls and send secure text messages.[16] Freely available software can be used to encrypt email messages and instant messages sent between computers, which can frustrate surveillance efforts traditionally performed by intercepting communications as they are transmitted over the Internet.

However, most of these secure communication technologies protect only the content of the conversation and do not protect the metadata. Government agents that intercept an encrypted email may not know what was said, but they will be able to learn the email address that sent the message and the address that received it as well as the size of the message and when it was sent. Likewise, Internet metadata can reveal the parties making an encrypted audio call and the time and duration of the call, even if the voice contents of the call are beyond the reach of a wiretap.

Some security technologies are specifically designed to hide metadata trails, but those technologies do not work quickly enough to allow real-time communication. The general technique for hiding the origin and destination information for an Internet communication involves sending data through a series of intermediaries before it reaches the destination, thus making it more difficult for an entity such as a government agency to learn both the source and destination of the communication. (Such

[16] Somini Sengupta, *Digital Tools to Curb Snooping,* N.Y. TIMES, July 17, 2013, http://nyti.ms/12JKz1s (describing RedPhone and Silent Circle).

information is conventionally encrypted so that the intermediaries cannot capture it; and a series of intermediaries is used so that no one intermediary knows the identities of both endpoints.)

The most popular and well-studied of these metadata hiding systems is The Tor Project, which was originally created by the U.S. Naval Research Lab, and has since received significant funding from the State Department. One important and widely acknowledged limitation of Tor is the noticeable delay introduced by using the tool. Web browsing conducted through Tor is much slower than through a direct connection to the site, as all data must be sent through a series of Tor relays, located in different parts of the world. These volunteer-run relays are oversubscribed—that is, the demands on the few relays from hundreds of thousands of Tor users are greater than the relays can supply, leading to slowdowns due to "traffic jams" at the relays.

Browsing the web using Tor can be painfully slow, in some cases requiring several seconds or longer to load a page. Real-time audio and video communications require a connection with minimal delay, which Tor cannot deliver. Internet telephony and video conferencing services are simply unusable over metadata-protecting systems like Tor.

As a result, although individuals can use security technologies to protect the contents of their communications, there are significant technical barriers that make it difficult, if not impossible, to hide communications metadata, particularly for real-time communications services such as Internet telephony and video conferencing.

Telephony Metadata Reveals Content

Telephony metadata can be extremely revealing, both at the level of individual calls and, especially, in the aggregate.

Although this metadata might, on first impression, seem to be little more than "information concerning the numbers dialed,"[17] analysis of telephony metadata often reveals information that could traditionally only be obtained by examining the contents of communications. That is, metadata is often a proxy for content.

In the simplest example, certain telephone numbers are used for a single purpose, such that any contact reveals basic and often sensitive information about the caller. Examples include support hotlines for victims of domestic violence[18] and rape.[19] Similarly,

[17] Administration White Paper, *Bulk Collection of Telephony Metadata Under Section 215 of the USA Patriot Act* 15 (Aug. 9, 2013), http://huff.to/1ey9ua5.
[18] *National Domestic Violence Hotline*, The Hotline (last visited Aug. 22, 2013), http://www.thehotline.org.
[19] *National Sexual Assault Hotline*, RAINN: Rape, Abuse & Incest National Network (last visited Aug. 22, 2013), http://www.rainn.org/get-help/national-sexual-assault-hotline.

numerous hotlines exist for people considering suicide,[20] including specific services for first responders,[21] veterans,[22] and gay and lesbian teenagers.[23] Hotlines exist for sufferers of various forms of addiction, such as alcohol,[24] drugs, and gambling.[25]

Similarly, inspectors general at practically every federal agency—including the NSA[26]—have hotlines through which misconduct, waste, and fraud can be reported, while numerous state tax agencies have dedicated hotlines for reporting tax fraud.[27] Hotlines have also been established to report hate crimes,[28] arson,[29] illegal firearms[30] and child abuse.[31] In all these cases, the metadata alone conveys a great deal about the content of the call, even without any further information.

The phone records indicating that someone called a sexual assault hotline or a tax fraud reporting hotline will of course not reveal the exact words that were spoken during those calls, but phone records indicating a 30-minute call to one of these numbers will still reveal information that virtually everyone would consider extremely private.

In some cases, metadata is even more sensitive than the contents of a communication. For example, wireless telephone carriers permit subscribers to donate to certain charities by sending a text message from their mobile phones. These systems require the subscriber to send a specific text message to a special number, which will then cause the wireless carrier to add that donation to the subscriber's monthly telephone bill. For example, by sending the word HAITI to 90999, a wireless subscriber can donate $10 to the American Red Cross.

---

[20] *District of Columbia/Washington D.C. Suicide & Crisis Hotlines,* National Suicide Hotlines (last visited Aug. 22, 2013), http://www.suicidehotlines.com/distcolum.html.
[21] *Get Help Now! Contact us to Get Confidential Help via Phone or Email,* Safe Call Now (last visited Aug. 22, 2013), http://safecallnow.org.
[22] *About the Veterans Crisis Line*, Veterans Crisis Line (last visited Aug. 22, 2013), http://www.veteranscrisisline.net/About/AboutVeteransCrisisLine.aspx.
[23] *We Provide Crisis Intervention and Suicide Prevention for LGBTQ Youth,* The Trevor Project (last visited Aug. 22, 2013), thttp://www.thetrevorproject.org
[24] *Alcohol Addiction Helpline,* Alcohol Hotline (last visited Aug. 22, 2013), http://www.alcoholhotline.com.
[25] *What is Problem Gambling*?, National Council on Problem Gambling (last visited Aug. 22, 2013), http://bit.ly/cyosu.
[26] Barton Gellman, *NSA Statements to the Post*, WASH. POST, Aug. 15, 2013, http://wapo.st/15LliAB.
[27] *Report Tax Fraud – Tax Fraud Hotline*, North Carolina Department of Revenue (last visited Aug. 22, 2013), http://www.dor.state.nc.us/taxes/reportfraud.html.
[28] *Report Hate Crimes*, LAMBDA GLBT Community Services (last visited Aug. 22, 2013), http://www.lambda.org/hatecr2.htm.
[29] *ATF Hotlines – Arson Hotline*, Bureau of Alcohol, Tobacco, Firearms and Explosives (last visited Aug. 22, 2013), http://www.atf.gov/contact/hotlines/index.html.
[30] *ATF Hotlines – Report Illegal Firearms Activity*, Bureau of Alcohol, Tobacco, Firearms and Explosives (last visited Aug. 22, 2013), http://www.atf.gov/contact/hotlines/index.html.
[31] *Childhelp National Child Abuse Hotline*, Childhelp (last visited Aug. 22, 2013), http://www.childhelp.org/pages/hotline-home.

Such text message donation services have proven to be extremely popular. Today, wireless subscribers can use text messages to donate to churches,[32] to support breast cancer research,[33] and to support organizations such as Planned Parenthood.[34] Similarly, after a policy change in 2012 by the Federal Election Commission, political candidates such as Barack Obama and Mitt Romney were able to raise money directly via text message.[35]

In all these cases, the most significant information—the recipient of the donation—is captured in the metadata, while the content of the message itself is less important. The metadata alone reveals the fact that the sender was donating money to their church, to Planned Parenthood, or to a particular political campaign.

Metadata can expose an extraordinary amount about our habits and activities. Calling patterns can reveal when we are awake and asleep; our religion, if a person regularly makes no calls on the Sabbath, or makes a large number of calls on Christmas Day; our work habits and our social attitudes; the number of friends we have; and even our civil and political affiliations.

Aggregated Telephony Metadata Reveals Our Relationships

When call metadata is aggregated and mined for information across time, it can be an even richer repository of personal and associational details.

Metadata can identify our closest relationships. Two people in an intimate relationship may regularly call each other, often late in the evening. If those calls become less frequent or end altogether, metadata will tell us that the relationship has likely ended as well—and it will tell us when a new relationship gets underway. More generally, someone you speak to once a year is less likely to be a close friend than someone you talk to once a week. Analysis of metadata on this scale can reveal the network of individuals with whom we communicate—commonly called a *social graph*.

Metadata also reveals the structure and activities of organizations. By building a social graph that maps all of an organization's telephone calls over time, one could obtain a set of contacts that includes a substantial portion of the organization's membership, donors, political supporters, and so on. Analysis of the metadata belonging to these individual callers, by moving one "hop" further out, could help to classify each one, eventually yielding a detailed breakdown of the organization's associational relationships.

---

[32] *Several Ways to Give*, The Simple Church (2013), http://bit.ly/1508Mgw; *Other Ways to Give*, North Point Church (last visited Aug. 22, 2013), http://bit.ly/16S3IkO.
[33] *Donate by Text*, Susan G. Komen for the Cure (last visited Aug. 22, 2013), http://sgk.mn/19AjGP7.
[34] *Help Support a New Future for Illinois Women and Families*, Planned Parenthood of Illinois (last visited Aug. 22, 2013), http://bit.ly/1bXI2TX.
[35] Dan Eggen, *Text to 'GIVE' to Obama: President's Campaign Launches Cellphone Donation Drive,* WASH. POST, Aug. 23, 2012, http://bit.ly/16ibjCZ.

Even our relative power and social status can be determined by calling patterns. As *The Economist* observed in 2010, "People at the top of the office or social pecking order often receive quick callbacks, do not worry about calling other people late at night and tend to get more calls at times when social events are most often organized (sic), such as Friday afternoons."[36]

At times, by placing multiple calls in context, metadata analysis can even reveal patterns and sensitive information that would not be discoverable by intercepting the content of an individual communication.

For example, although metadata revealing a single telephone call to a bookie may suggest that the caller is placing a bet, analysis of metadata *over time* could reveal that someone has a gambling problem, particularly if the call records also reveal a series of calls to payday loan services.

With a database of telephony metadata reaching back five years, many of these kinds of patterns will emerge once the collected phone records are subjected to even the most basic analytic techniques.

In short, aggregated telephony metadata allows the NSA to construct social graphs and to study their evolution and communications patterns over days, weeks, months, or even years. Metadata analysis can reveal the rise and fall of intimate relationships, the diagnosis of a life-threatening disease, the telltale signs of a corporate merger or acquisition, or the social dynamics of a group of associates.

Data-Mining Across Many Individuals Is More Revealing

Advances in the area of "Big Data" over the past few decades have enabled researchers to observe even deeper patterns by mining large pools of metadata that span many telephone subscribers.

Researchers have studied databases of call records to analyze the communications reciprocity in relationships,[37] the differences in calling patterns between mobile and landline subscribers,[38] and the social affinity and social groups of callers.[39]

---

[36] *Mining Social Networks: Untangling the Social Web,* ECONOMIST, Sep. 2, 2010, http://econ.st/9iH1P7.

[37] Lauri Kovanen, Jari Saramaki & Kimmo Kaski, *Reciprocity of Mobile Phone Calls*, Dynamics of Socio-Economic Systems (Feb. 3, 2010), http://arxiv.org/pdf/1002.0763.pdf.

[38] Heath Hohwald, Enrique Frias-Martinez & Nuria Oliver, *User Modeling for Telecommunication Applications: Experiences and Practical Implications* 8, (Data Mining and User Modeling Group, Telefonica Research, 2013), http://bit.ly/1d7WkUU ("Interestingly, Monday is the day with most calls for landline users, while Friday is the day with most calls for mobile users. . . Mobile users spend less time on the phone than landline users.").

Researchers have discovered that individuals have unique calling patterns, regardless of which telephone they are using,[40] they have figured out how to predict the kind of device that is making the calls (a telephone or a fax machine),[41] developed algorithms capable of predicting whether the phone line is used by a business or for personal use,[42] identified callers by social group (workers, commuters, and students) based on their calling patterns,[43] and even estimated the personality traits of individual subscribers.[44]

The work of these researchers suggests that the power of metadata analysis and its potential impact on the privacy of individuals increases with the scale of the data collected and analyzed. Just as multiple calls by the same person reveal more than a single call, so too does a database containing calling data about millions of people reveal more information about the individuals contained within it than a database with calling data about just one person.

The privacy impact of collecting all communications metadata about a single person for long periods of time is qualitatively different than doing so over a period of a few days. Similarly, the privacy impact of assembling the call records of every American is vastly greater than the impact of collecting data about a single person or even groups of people. Mass collection not only allows the NSA to learn information about more people, but it also gives the NSA the ability to learn new, previously private facts about innocent Americans that it could not have learned simply by collecting the information about a few, specific individuals.

Technical Expertise Bolsters Oversight and Public Understanding

Some of the frustration voiced by the Foreign Intelligence Surveillance Court in its declassified opinions seems to stem from the Court's discovery that the NSA had not disclosed significant technical information in earlier proceedings. One need not

---

[39] Sara Motahari, Ole J. Mengshoel, Phyllis Reuther, Sandeep Appala, Luca Zoia & Jay Shah, *The Impact of Social Affinity on Phone Calling Patterns: Categorizing Social Ties from Call Data Records*, The 6th SNA-KDD Workshop (Aug. 12, 2012), http://b.gatech.edu/1d6i4RY.
[40] Corrina Cortes, Daryl Pregibon & Chris Volinsky, *Communities of Interest*, AT&T Shannon Research Labs, http://www.research.att.com/~volinsky/papers/portugal.ps.
[41] Haim Kaplan, Maria Strauss & Mario Szegedy, *Just the Fax – Differentiating Voice and Fax Phone Lines Using Call Billing Data*, AT&T Labs, http://bit.ly/19Aa8Ua.
[42] Corinna Cortes & Daryl Pregibon, *Giga-Mining*, AT&T Labs-Research, http://bit.ly/153pMcI.
[43] Richard A. Becker, Ramon Caceres, Karrie Hanson, Ji Meng Loh, Simon Urbanek, Alexander Varshavsky & Chris Volinsky, *Clustering Anonymized Mobile Call Detail Records to Find Usage Groups*, AT&T Labs-Research, http://soc.att.com/16jmKdz.
[44] Rodrigo de Oliveira, Alexandros Karatzoglou, Pedro Concejero, Ana Armenta & Nuria Oliver, *Towards a Psychographic User Model from Mobile Phone Usage*, CHI 2011 Work-in-Progress (May 7–12, 2011), http://bit.ly/1f51mOy; *see also* Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic & Alex (Sandy) Pentland, *Predicting People Personality Using Novel Mobile Phone-Based Metrics*. Social Computing, Behavioral-Cultural Modeling and Prediction (2013), http://bit.ly/1867vWU.

postulate bad faith on the NSA's part to explain how this could have happened. Technologists within the NSA surely knew how their program operated, but this knowledge had to pass through intermediaries, some of them less attuned to the significance of certain technical details, before reaching the Court. A good faith effort to simplify the technical explanation for the Court's benefit could have led to the omission of information that the Court later found highly relevant. And the Court, without access to technical advice, was not able to ask the sort of probing technical question that might have elicited the missing information.

In order to ensure strong oversight of these complex programs, the overseers must have independent access to robust technical expertise. Fortunately, the United States has the world's strongest pool of experts in these areas. I look forward to your questions today and, more broadly, to continued constructive engagement between oversight officials and technical experts.