THE PUBLIC AND THE PRIVATE AT THE UNITED STATES BORDER WITH CYBERSPACE

John Palfrey[*]

ABSTRACT

In the twenty-first century, a state can come to know more about each of its citizens via surveillance than ever before in human history. Some states are beginning to exercise this ability. Much of this additional surveillance ability derives from enhanced access to digital information. This digital information comes in the form of bits of data that flow through both rivers and oceans of data. These rivers are full of information that passes by a given point, or series of points, in a network and can be intercepted; these oceans are also stocked with information that can be searched after the fact. These data are held in

[*] Henry N. Ess III Professor of Law and Vice Dean of Library and Information Resources at Harvard Law School and a faculty director of the Berkman Center for Internet & Society at Harvard University. This paper was commissioned and underwritten by funds from the National Center for Justice and the Rule of Law at the University of Mississippi School of Law (National Center), which is supported by a grant from the Office of Justice Programs at the United States Department of Justice. The research on Internet surveillance included in this paper was also supported by a grant from the John D. and Catherine T. MacArthur Foundation. The author gratefully thanks Hal Roberts for his research on the three levels of surveillance, development of draft sections of this paper, and shrewd editorial suggestions. Terry Fisher, Christopher Soghoian, Dr. Robert Faris, and Chris Conley offered important technical and editorial assistance. In addition, Virginia Farmer deserves thanks for excellent legal research assistance. Errors and omissions are the author's alone.

private hands as well as public. The most effective (or invasive, depending upon your vantage point) new forms of surveillance often involve a search of data held in a combination of private and public hands. Both private and public entities are increasingly encouraged to retain more data as a result of legal and market pressures. There are essentially no Fourth Amendment protections for U.S. citizens whose data is collected by a private third-party and turned over to the state. Nor are there such constitutional protections for the re-use of privately collected data by state actors. The few statutory provisions that protect citizens in these contexts are out-of-date and riddled with loop-holes. This inquiry prompts hard questions about the need to redefine the public and the private in a digital age. The meaning of the public and the private is changing, in material ways, both from the perspective of the observer and the observed. We need to rethink legal protections for citizens from state surveillance in a digital age as a result of this third-party data problem.

I. INTRODUCTION

> Technological progress poses a threat to privacy by enabling an extent of surveillance that in earlier times would have been prohibitively expensive. Whether and what kind of restrictions should, in the name of the Constitution, be placed on such surveillance when used in routine criminal enforcement are momentous issues that fortunately we need not try to resolve in this case.[1]

So punted a court in a case related to the privacy implications of police use of a Global Positioning Satellite (GPS) program. This punting was with good reason in this particular case: the court did not need to reach this larger issue to resolve the matter before them. Instead, the court simply wished to draw attention to the bigger problem on the horizon. At some point, each state that cares about the privacy interests of its

[1] United States v. Garcia, 474 F.3d 994, 998 (7th Cir. 2007).

citizens will have to grapple with this hard question. In the process, we will find that our received notions of what is public and what is private have changed in a digital age in ways that have profound implications for the future.[2]

As a nation that continues to lead in the development of new technologies, we in the United States push information technology harder and further ahead with every passing year. Much of this innovation takes place in the private sector. It is this innovation, in large measure, that makes surveillance of citizens far cheaper, just as the court said in *United States v. Garcia*.[3] The benefits for economic growth and productivity are obvious; we accept this progress as a social good. The long-term costs of this progress, in areas such as privacy, are less obvious.

As consumers and as citizens, we repeatedly trade convenience for control, handing over growing amounts of information about ourselves to others in the process. Our lives are increasingly mediated by digital technologies and described by data held in digital formats. We are racing ahead quickly with the development of new technologies while the institutions—legal and otherwise—designed to protect user privacy have lagged behind. The tradeoffs involved are rarely conscious ones.

This growing problem has its roots in the fact that, as information technologies improve in efficiency and become more integrated in everyday life, fewer and fewer citizens are likely to know what information is being collected about them and by whom. Much of this data collection is conducted not by state actors, but by private parties. Citizens have even less of a sense of how these data are being used, or might be used and re-used in the future. This issue is particularly acute in the context of

[2] The United States is not alone in this respect. "Privacy is a protean concept," according to a Canadian Supreme Court opinion, R.V. Tessling, [2004] 3 S.C.R. 432, 434 (Can.). In his opinion, Justice Binnie implies that the decision-makers are avoiding hard questions. The *Tessling* case parallels the landmark United States case, *Kyllo v. United States*, 533 U.S. 27 (2001), in important respects, though the outcomes of the two matters diverge from one another.

[3] *Garcia*, 474 F.3d at 998 ("Technological progress poses a threat to privacy by enabling an extent of surveillance that in earlier times would have been prohibitively expensive.").

young people we have surveyed, who are leaving vast tracks behind them as they lead lives substantially mediated by digital technologies.[4]

This lack of information is not just a matter of simple consumer ignorance or lack of attention on the part of citizens. Rather, the primary cause is that the technologies of surveillance in use at any moment, in this digital age, are unknowable even by experts. We have no way of detecting many kinds of prevalent surveillance; those that we can detect are often hopelessly, unknowably complex. Expert computer users are often fooled by whether the antivirus software they are running on their computer is capturing private information about them or not.[5] Virtually no one knows what information is being kept by our Internet Service Provider or e-mail host or search engine, much less what the National Security Agency (NSA) is collecting through the computers it has installed on AT&T's network.[6] And for many working Americans, there is no realistic choice but to rely upon at least some private parties—whether Google, Microsoft or Yahoo! for search, or AT&T, Verizon, or Comcast for data, and so forth—for digital communications services.

The most important aspect of this story of surveillance—and how little citizens know about it—is the change in the interaction between the public and the private. A growing number of private firms are collecting a growing amount of data about individuals. In effect, each individual has a growing "digital dossier," in the words of Professor Daniel Solove, of information held about them in a distributed set of places.[7] The

[4] *See* JOHN PALFREY & URS GASSER, BORN DIGITAL: UNDERSTANDING THE FIRST GENERATION OF DIGITAL NATIVES 17-82 (2008).

[5] The author has seen this dynamic repeatedly in the context of a related research project, StopBadware.org, which provides forums for novices and experts to work together to understand how certain forms of online surveillance work. *See* http://www.stopbadware.org/.

[6] *See* Ryan Singel, *Whistle-blower Outs NSA Spy Room*, WIRED, Apr. 7, 2006, *available at* http://www.wired.com/science/discoveries/news/2006/04/70619 (last visited Sept. 30, 2008).

[7] *See* DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 2-7 (2004).

rules for how and when the government can access and aggregate these data are few and unclear. Likewise, there are no protections for the re-use of these data among law enforcement authorities once the information is collected, regardless, in fact, of the legality of the original mode of accessing the data. The Fourth Amendment, as currently understood, reaches neither.[8] While some statutes provide limited protection to citizens, these protections are spotty and cover only specific types of information.[9] This fast-growing "third-party data problem" and its implications are the central concerns of this paper.[10]

This paper builds upon our collective understanding of this emerging third-party data problem in the context of current surveillance and data retention practices. The purpose of reviewing these practices in detail is to highlight some of their largest implications for the privacy of individuals. I begin by describing the changes that advances in digital technologies are bringing about for the practice of state surveillance of citizens with the intention of setting those changes into the context of the conception of the public and the private. The trend I underscore here is the greater interaction between the public sector (those in government who wish to learn more about certain citizens) and private actors who either collect information about citizens or who develop the technology to enable monitoring of citizen activity. As a "surveillance-ready" network, the Internet permits surveillance of citizens in more ways than ever before in human history. This surveillance can take place at the network, server, and client levels of the Internet's architecture. At the same time, citizens are unlikely

[8] *See, e.g.*, Smith v. Maryland, 442 U.S. 735, 743-44 (1979); United States v. Miller, 425 U.S. 435, 444 (1976).

[9] *See* Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083, 1085-86 (2002).

[10] Others have taken up aspects of this problem, too, from various perspectives. *See* Solove, *supra* note 9; Stephen E. Henderson, *Beyond the (Current) Fourth Amendment: Protecting Third-Party Information, Third Parties, and the Rest of Us Too*, 34 PEPP. L. REV. 975 (2007); CHRISTOPHER SLOBOGIN, PRIVACY AT RISK: THE NEW GOVERNMENT SURVEILLANCE AND THE FOURTH AMENDMENT (Univ. of Chicago Press 2007); Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 MINN. L. REV. 1 (2008).

to be in a position to keep up with the changes in these technologies and the potential surveillance practices that they permit. In the final section of the paper, I argue that the combination of these factors places increased pressure on the complex—some say unstable—legal doctrine of the "reasonable expectation of privacy."

II. THE DIGITAL DIFFERENCE

Digital technologies, including but not limited to the public Internet, make several crucial differences when it comes to surveillance. One is the rate of change in technology development and usage practices—and the corresponding difficulty for individuals to keep up with these changes.[11] A second is the scale on which activities occur and the potential area of impact of one's actions.[12] Together, these differences brought about by the rise of digital technologies place a great deal of pressure on courts' interpretation of crucial legal doctrines, such as the Fourth Amendment of the United States Constitution.

The rate of change in technology developments, and associated usage patterns, is faster in many respects than it has been in the past. The time is shortened between major changes in societies and markets; the speed of getting to market with a new technology service or product for an entrepreneur is often faster; the time that it takes for an idea or expression to reach many people can be much lower; and so forth. Consider that it was *centuries* between the invention of the first moveable type in China and the printing date of Gutenberg's first bibles, and yet further centuries before those bibles reached wide circulation. In contrast, digital technologies have reached over a billion people within a few *decades* of their invention.[13] The ability to

[11] *See generally* JAMES GLEICK, FASTER 83-93 (1999).

[12] This assertion may seem self-evident to some readers, but the Internet today reaches most parts of the globe and can be accessed at little or no cost in public places such as libraries and schools around the world. The impact of an action that takes place in one place—say, New York in the United States—may have reverberations in a place distant in physical terms and in time.

[13] *See* Robert Darnton, *The Library in the New Age*, THE NEW YORK REVIEW OF

do more with digital technologies is outstripping our ability to process what the social implications of these changes may be.

The scale on which these activities take place, and on which they have an impact, is likewise greater than it has been in the past. The reach and effect of actions in digital space, such as publication of data to a digital network, are greater and growing. The potential impact of an idea can be greater; one can reach many more people by way of an audience than ever before; the scale of harm can be greater for a bad act; the scale of entrepreneurial innovations can become enormous very quickly; the amount and timing of capital is changing; and so forth.

The broad effect of these two factors—vastly increased speed and scale in the development and reach of information technologies—is to make it hard to apply and make rules in this field. These changes apply special pressure to existing legal doctrines, which begin to seem inadequate to new tasks of protecting individual privacy. The more technology advances, and the more people use the technology to mediate more aspects of their lives, the less adequate existing legal protections for individual privacy appear. The impact of these frequent and material changes should be to prompt frequent review of existing laws. In other cases, the answer may be to establish creative, flexible institutions to address such problems without the traditional legislative involvement, to the extent that new effective laws cannot be enacted to solve the right problems.[14]

Lawmakers and judges face a significant challenge in keeping up with this rate of change in information technologies. The implications of these shifts are many; those implications may be yet greater in a few short years if the rate of change continues. I emphasize here the implications for the way that we think of the public and the private in law broadly, with

BOOKS, Vol. 55, No. 10 (June 12, 2008) (for a general discussion of this transition in information technologies and the rate of change).

[14] *See, e.g.*, JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET AND HOW TO STOP IT 170-73, 195-99 (2008). Consider the example of StopBadware.org, a neighborhood watch project on the Internet intended to protect consumers from malicious code through a partnership of private entities, which picks up where classic regulation has left off. *Id.* at 170-73.

special application to the workings of the Fourth Amendment of the United States Constitution.[15] There are yet more profound implications when one thinks of the implications at an international legal level, since the interplay between law enforcement agencies and private parties globally makes the puzzle vastly more complex. And in many states, citizens have far less to protect them from state intrusion than the Fourth Amendment in the United States.

In the context of surveillance and data retention, these two broad changes mean that, over time, more people can come to know more about what others are doing or have done. The technology of the Internet and other digital networks is surveillance-ready, which is to say the technologies are designed in such a way that surveillance is simple to perform at multiple points in the network. There are many "points of control," as Professor Jonathan Zittrain has pointed out, at which a state may seek to block or to copy bits of information.[16] These points of control include international gateways, Internet Exchange Points (IXPs), Internet Service Providers (ISPs), public access points (such as cybercafés, schools, or public libraries), corporate workplaces, technology service providers (such as blogging hosts, for instance), and other networks, like mobile and tracking devices such as GPS systems on vehicles or handheld devices.[17]

III. TYPES OF DIGITAL SURVEILLANCE: NETWORK, SERVER, AND

[15] In this paper, I do not make a specific claim as to when a warrant should issue, nor do I cover the specifics of criminal procedure, such as Rule 41 of the United States Federal Rules of Criminal Procedure (FRCP). There has been much work done in terms of how these processes should work by other legal scholars. *See, e.g.*, Orin S. Kerr, *Search Warrants in an Era of Digital Evidence*, 75 MISS. L.J. 85-135 (2005) ("This article urges legislatures and rules committees to update the statutory rules that govern the warrant process in response to the new challenge of digital searches," including specific proposals to update Rule 41 of the United States FRCP); Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531-85 (2005); and Orin S. Kerr, *Four Models of Fourth Amendment Protection*, 60 STAN. L. REV. 503-51 (2007).

[16] *See* Jonathan Zittrain, *Internet Points of Control*, 44 B.C. L. REV. 653 (2003).

[17] *See, e.g.*, Marjorie A. Shields, Annotation, *Fourth Amendment Protections, and Equivalent State Constitutional Protections, as Applied to the Use of GPS Technology, Transponder, or the Like, to Monitor Location and Movement of Motor Vehicle, Aircraft, or Watercraft*, 5 A.L.R.6th 385 (2005).

CLIENT

When one thinks of network surveillance, the image of a man wearing headphones (and perhaps chomping on cold pizza) in a van on a stakeout outside a crook's apartment building may leap to mind. It is true that one means of listening in on the conversations of a surveillance target is to seek to intercept information flows across the network, as in a river (and perhaps to copy or store it for later retrieval). But in the digital age, this mode of "tapping" a line is only one of a wide range of types of surveillance available to a law enforcement official—and not necessarily the most efficient, given the architecture and the scale of the network. The use of many more types of data, collected by many different types of actors, makes it much easier in some instances, to reconstruct a past act or to anticipate a future—lawful or unlawful—act.

Given the proliferation of the means of surveillance on digital networks, it is extremely hard to provide a full picture, at any moment, of all possible means of surveillance of a target. Such a description is hard both on the micro level (of analyzing individual surveillance activities) and on the macro level (of setting these practices into the context of the larger technical, social, political, and legal issues in which they are nested).

One reason for the difficulty at the micro level is that Internet surveillance is a "read-only" activity. If surveillance is done well at the network or server level, it is impossible to detect through direct technical means. We cannot detect the fact of surveillance, what data is being surveilled, or who is surveilling the data. At the micro level, it is difficult to detect and verify the data, actors, and uses of the data involved in any well-planned digital surveillance activity.[18]

[18] Most of the observations in this section of this paper are derived from an ongoing research project on Internet surveillance at the Berkman Center for Internet & Society at Harvard University, of which the author is principal investigator. Key researchers include Christopher Soghoian, Hal Roberts, and Chris Conley. Others have made similar observations; nothing in this section is groundbreaking in terms of computer science research. *See generally*, David Lyon, *Cyberspace, Surveillance, and Social Control: The Hidden Face of the Internet in Asia*, *in* ASIA.COM: ASIA ENCOUNTERS THE INTERNET 67, 67-82 (Kong Chong Ho, Randy Kluver, Kenneth C.C. Yang eds., 2003);

The simplest example of this sort of difficult-to-monitor surveillance is a monitoring box, placed with care in the closet of an Internet service provider.[19] If that monitoring box is physically connected to the rest of the network with only read (not write) wires coming from the network, it is technologically incapable of writing anything back to the network. Surveillance at the client level is always theoretically detectable, but most client-side surveillance tools are designed to avoid detection and can only be thwarted in this regard by sophisticated detection techniques. This problem of detection limits what we can say definitively about a given surveillance activity at any given moment. Since we cannot know the answer with perfect clarity, we have to look at that larger field of Internet surveillance as a whole to think about the potential data subject to surveillance.

On the macro level, placing Internet surveillance into the larger technical, political, social, and legal landscapes leads to further questions. The questions are large because of the stream of new information about Internet surveillance. Within this stream of information, it is important to distinguish what represents a meaningfully new type of surveillance and what represents another example of an existing type. There is a wide and growing array of potential surveillance tools. A complete catalogue of these devices would not only be dreadfully boring to read, it would almost certainly be incomplete at any given moment in time since Internet-related surveillance is meant to be carried out undetected. All of these various tools interact with each other in complex ways; for instance, the routers that

SIMSON GARFINKEL, DATABASE NATION: THE DEATH OF PRIVACY IN THE 21ST CENTURY (2000); LAWRENCE LESSIG, CODE 2.0, 200-232 (2006); DANIEL SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE (2004).

[19] The range of types of surveillance, by the government and private industry, has been described in many scholarly papers before this one. For one overview, *see* Gattiker, U.E., Holsten, H. & Miller, J.,

User Attitudes Toward Possible Governmental and/or Organizational Surveillance, Monitoring

and/or Eavesdropping on the Internet, EICAR 2000 Best Paper Proceedings, 126-157 (2000), *available at* http://eicar.weburb.dk/past_conferences/2000/papers/Monday/Security%20Trust%20and%20E-commerce/other/Gattiker.pdf (last visited Nov. 25, 2008).

control the network are clients themselves and so are open to the same sorts of questions we ask of other clients (and vulnerable to monitoring by a range of client surveillance tools).

This paper includes a concise—in fact, simplified—picture of the technical landscape of Internet surveillance through a series of examples of data collection that might permit a state to come to know more about a surveillance target. These examples help inform (and are informed by) the technical questions of what data are being *actually* and *potentially* monitored on the Internet and whom we are trusting to access (and potentially copy and re-use) those data. This paper is meant to include enough of the major sorts of Internet surveillance to place any given activity or tool within a useful technical context.

Internet surveillance can be broken down into three primary categories: network, server, and client. The Internet is composed of clients and servers, in essence a series of devices which talk to one another through the network. Every bit of data on the Internet is traveling or residing at one or more of these locations at any given time. As such, any given Internet activity must happen at or across one (or more) of these locations. In some cases, the line between client and server is blurry or non-existent, but the class of nodes that act as servers is large enough that it is useful to distinguish them from (other) clients. In this model, anything residing on the client proper is treated as client-side Internet surveillance, including both software tools like workplace keylogging systems and hardware tools like keyboard tapping devices. Any machine that predominantly accepts requests, processes them, and returns responses is treated as a server. For simplicity, everything in between the client and the server is considered here to be the network, including the wires that the data travels and the routers that direct the traffic.

A. Network Surveillance

At the highest level, a state can practice surveillance of Internet traffic that flows across the network.[20] The point of

[20] *See, e.g.*, SOLOVE, THE DIGITAL PERSON, *supra* note 18, 165-87.

control used by the state might exist at the international gateway between two states or at some intermediate point in the network. For those states that aspire to manage the network the most extensively, this control might function roughly as a "border control." In most places, though—and especially the United States—it would not be possible to establish a perfect ring around the geographic boundaries of the state. Some states, such as China and Saudi Arabia, have sought to establish such means of network control. The colorful example of the "great firewall of China" helps in a graphic sense to envision what the Chinese have sought to erect, but the image is only partially accurate; it both overstates the effectiveness of the wall at the national border and understates the extensiveness of the surveillance regime in China. The British Government is reported to be considering a dramatic expansion to their network surveillance capabilities, in the form of a £12 billion surveillance center to monitor and store emails, web surfing records, and phone records on the nation's largest telecommunications networks.[21] Other states use intermediate mechanisms, such as Internet Exchange Points (IXPs), to force large amounts of network traffic to flow through a single point which might then be monitored.[22] In other cases, such as the examples that follow, the point of control is on the network, but does not fall precisely at an international border.

1. Network Surveillance in the United States

In the United States, the passage of the Communications Assistance for Law Enforcement Act (CALEA) in 1994 was a

[21] David Leppard, *Government will spy on every call and e-mail*, THE SUNDAY TIMES, Oct. 5, 2008, *available at* http://www.timesonline.co.uk/tol/news/uk/article4882600.ece (last visited Oct 8, 2008); Graham Tibbetts, *Internet black boxes to record every email and website visit*, THE TELEGRAPH, Nov. 6, 2008, *available at* http://www.telegraph.co.uk/news/uknews/3384743/Internet-black-boxes-to-record-every-email-and-website-visit.html (last visited Nov. 11, 2008).

[22] *See e.g.*, Alan Levin, *Creating National Internet Exchange Points in Africa*, International Telecommunications Union Presentation, June 29, 2005, *available at* http://www.itu.int/ITU-D/finance/work-cost-tariffs/events/tariff-seminars/south-africa-05/presentation_levin-ixp-en.pdf (last visited Nov. 11, 2008) (providing a description of IXPs and their general usage).

watershed event in terms of network surveillance.[23] CALEA requires that telecommunications companies build tools into their telephone networks that allow companies to respond quickly to law enforcement requests for wiretaps.[24] CALEA filled an obvious gap: the growing use of a new generation of digital telephone switches did not inherently provide the same support for wiretapping as the older tools did. In 2005, the Federal Communications Commission (FCC) extended its interpretation of the law to require that Internet Service Providers (including universities, schools, libraries, and other such non-commercial ISPs) provide wiretapping access to a range of Internet data.[25] The accessible data includes voice over IP (VoIP) Internet telephone services, such as Vonage and Skype; data about when and for how long Internet broadband subscribers connect to the Internet; and packet header data (the source and destination addresses and the port number) of all VoIP packets.[26] In order to reduce the high cost of implementing this new interpretation of CALEA, the FCC has ruled that ISPs could forward their entire data stream to an independent "Trusted Third Party" to handle the wiretapping implementation.[27] The effect of this ruling can be to expose the entire data stream of an ISP using this option to a third party.[28] The Department of Justice submitted a petition in 2007, yet to be ruled on, to include the packet header data of all Internet data, not just VoIP data, but web, email, instant message, and all other Internet traffic.[29] CALEA does not provide legal

[23] *See* 47 U.S.C. §§ 1001-10 (2008).

[24] *Id.*

[25] *See* Federal Communications Commission, ET Docket No. 04-295 (Sept. 23, 2005), published 70 Fed. Reg. 59, 664 (Oct. 13, 2005).

[26] *See* Federal Communications Commission, *FCC Requires Certain Broadband and VoIP Providers to Accommodate Wiretaps*, Aug. 5, 2005, *available at* http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-260434A1.pdf (last visited Nov. 11, 2008).

[27] *See* Federal Communications Commission, Order FCC 06-56 (2006).

[28] *Id.*

[29] *See* Susan Crawford blog, CALEA roundup: 2005-2007, *available at* http://scrawford.blogware.com/blog/_archives/2007/8/16/3162684.html (Aug. 16, 2007, 13:24 EDT).

justification for anyone to access the provided data; it only mandates that the ISP build the technical capability to respond to such requests, the legality of which is in turn determined by other rules.[30]

Reports suggest that the U.S. National Security Agency (NSA) is engaged in a much more intrusive form of network surveillance by mining the full stream of data passing through major ISP backbones. AT&T engineer Mark Klein, after reading a story in the *New York Times*, reported that he had seen strange activity in a closet in an AT&T backbone building.[31] Klein had also seen documents describing data mining equipment in that closet.[32] Such a box is likely capable of executing highly sophisticated queries on the data passing through the backbone.[33] It is capable, for instance, of searching for any traffic (local or international) matching any of a set of complex profiles and forwarding the full contents of any matching data to the agency.[34] But a so-called "black box" such as the one described by Klein, located in a closet, is the best example of equipment that is impossible to probe technically. We are left with indeterminate, circumstantial evidence about the existence and functionality of the box. Questions about what data it is actually monitoring and who actually has access to the data remain unanswered.

2. The Public and Private on International Networks: Three Cases

Network-level surveillance is enabled directly by law in some instances. In other cases, private forms of surveillance become possible because of the way that citizens seek to work around the state surveillance. Three network surveillance

[30] *Id.*

[31] *See* John Markoff & Scott Shane, *Documents Show Link Between AT&T and Agency in Eavesdropping Case*, N.Y. TIMES, Apr. 13, 2006, at A1, *available at* http://www.nytimes.com/2006/04/13/us/nationalspecial3/13nsa.html (last visited Nov. 25, 2008).

[32] *Id.* at A7.

[33] *Id.*

[34] *Id.* at A8.

examples from outside the United States are instructive in this regard.

a. Sweden: FRA and Relakks

On June 18, 2008, Swedish lawmakers passed a bill giving its National Defense Radio Establishment (Försvarets radioanstalt, or FRA) the authority to monitor, without a warrant, all phone calls and Internet traffic crossing its border.[35] The FRA had previously been responsible only for intercepting radio signals, though there are reports that it has been monitoring phone calls and emails prior to the new law going into effect in 2009.[36] The new law will allow the Swedish government to access not just the data about phone calls and Internet traffic (who talked to whom when), but the content of the traffic itself when it crosses the Swedish border.[37] All Swedish ISPs will be required to install FRA monitoring equipment within their networks.[38]

There are means by which Internet users can avoid such network-level surveillance capabilities, but these means of evading the state's surveillance mechanisms may lead to further surveillance by private parties (which, in turn, might give those data to the very state agencies that the user sought to avoid). Relakks is a privacy and filtering circumvention tool that uses virtual private network (VPN) technology to encrypt and route all of its users' traffic through Sweden.[39] The tool gives its users

---

[35] *See* Sara Sundelius, *Sweden adopts controversial law to allow secret tapping of e-mails, phone calls*, Associated Press, INT'L HERALD TRIBUNE, June 18, 2008, *available at* http://www.iht.com/articles/ap/2008/06/18/europe/EU-GEN-Sweden-E-mail-Spying.php (last visited Nov. 25, 2008).

[36] *See* Svt.se, *Ingen förundersökning mot FRA*, July 14, 2008, *available at* http://svt.se/svt/jsp/Crosslink.jsp?d=22620&a=1173950&lid=puff_1173812&lpos=rubrik (last visited Nov. 25, 2008).

[37] *See* PrivacyInternational.org, *PHR2006 - Kingdom of Sweden*, Dec. 18, 2007, *available at* http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-559487 (last visited Nov. 25, 2008).

[38] *See* DW-World.de, *Swedish Government Clears Hurdles to Pass Surveillance Bill*, June 19, 2008, *available at* http://www.dw-world.de/dw/article/0,2144,3421627,00.html (last visited Nov. 25, 2008).

[39] *See* https://www.relakks.com/ (last visited Nov. 11, 2008).

privacy from snooping conducted by their own ISPs and consequently provides defense against the network level filtering that the OpenNet Initiative has documented in countries around the world.[40] This network level filtering consists of surveilling a subset of the traffic that crosses the relatively few routers that process nearly all Internet traffic. For instance, dozens of countries filter all Internet traffic crossing the digital border by running filtering software on the small number of routers that connect them to the outside world.[41] Relakks is one of many tools that routes around this network level filtering by encrypting and routing requests through a machine in a non-filtering country. By encrypting the request until it arrives at a machine that reroutes the conversation to the requested site, the circumvention tool prevents the filtering router from determining the location or content of the site. This process of rerouting conversations to shift places on the network is known as proxying. When someone in China uses a circumvention tool hosted in Sweden to proxy-request a page from the BBC's web server, instead of talking directly from China to Britain, the connection goes from China to Sweden and then from the U.S. to Britain (and back along the same route). This form of circumvention can also be effective against other sources of filtering (for instance, school or ISP filtering of music sharing traffic) as well to "country shift" the connection to allow the user to access content that is restricted (filtered) to users connecting from Sweden or Europe. But such re-routed traffic now goes through a state where both a private party and a government agency have the authority to listen to the conversation. The net effect: those trying to evade circumvention or surveillance through Relakks may unwittingly be exposing themselves to (potentially more extensive) surveillance by the FRA.

---

[40] *See* ACCESS DENIED: THE PRACTICE AND POLICY OF GLOBAL INTERNET FILTERING (Ronald Deibert, John Palfrey, Rafal Rohozinski, & Jonathan Zittrain eds., Cambridge: MIT Press, 2008).

[41] *See id*.

b. United Kingdom: British Telecom and Phorm

A case involving the giant telecommunications firm, British Telecom (BT), illustrates a similar point. In June and July of 2007, several BT Internet customers noticed strange activity during some of their Internet sessions.[42] Some of the users noticed performance problems; others found that some sites were not loading correctly, and all noticed extraneous traffic to the same system, dns.sysip.net, which was owned by a company, 121Media, that various anti-virus tools reported to be associated with rootkit spyware.[43] One user reinstalled his system from scratch after being told by BT customer service that his computer was likely infected by some sort of malware, only to find the same suspicious activity on the freshly installed system.[44] These users came to the conclusion that BT was involved somehow with 121Media, either intentionally or through some sort of intrusion. In response to questions from the users and from reporter Chris Williams of *The Register*, BT's customer service and its press office insisted that they had nothing to do with the suspicious behavior.[45] Phorm refused to comment.[46] By the end of the month, users reported the disappearance of the suspicious traffic, and chatter about the matter died without further notice.[47]

In February, 2008, Phorm (the new name of 121Media) publicly announced a deal with BT, as well as with other major

[42] *See* Posting of Frank Rizzo to thinkbroadband.com, *available at* http://bbs.adslguide.org.uk/showflat.php?Cat=&Board=bt&Number=3047764&page=3&view=expanded&sb=7&o=0&fpart=all&vc=1 (July 2, 2007, 10:35:12 EST) (last visited Nov. 25, 2008).

[43] *See* Posting of Filippo Spike Morelli to SpikeLab.org, *available at* http://www.spikelab.org/blog/btProxyHorror.html (July 9, 2007) (last visited Nov. 25, 2008).

[44] *See* Ryan Naraine, *Spyware, Rootkit Maker Stops Distribution*, eWeek.com, May 10, 2006, *available at* http://www.eweek.com/c/a/Security/Spyware-Rootkit-Maker-Stops-Distribution/ (last visited Nov. 25, 2008).

[45] *See* Chris Williams, *ISP data deal with former 'spyware' boss triggers privacy fears*, THE REGISTER, Mar. 17, 2008, *available at* http://www.theregister.co.uk/2008/03/17/bt_phorm_lies/print.html (last visited Nov. 25, 2008).

[46] *Id.*

[47] *Id.*

British Internet service providers TalkTalk and Virgin Media, to offer "a free consumer internet feature, Webwise, which ensures fewer irrelevant adverts and additional protection against malicious websites."[48] Williams quickly connected the announcement with the 2007 reports and posted a piece accusing Phorm and BT of cooperating to monitor.[49] By March, Phorm had admitted that it had undertaken a trial of its technology involving tens of thousands of users on the BT network with BT's participation, but without the knowledge of the users.[50] Privacy advocates reacted strongly against Phorm's monitoring of user connections since the revelation of the BT deal and of the secret trial, postponing any further use of Phorm on the BT network. But as of July 2008, BT planned to begin another trial of Phorm's system within a few weeks, though this time with an opt-in mechanism.[51]

Phorm attempts to allay privacy concerns by claiming that it does not store any personally identifying information or any specific information about the browsing habits of tracked users.[52] Instead, Phorm stores a persistent and unique, but randomly generated, Phorm identification number for each user, along with a list of the advertising categories triggered by each of the user's web requests.[53] So when a user requests a page about the Ford Expedition from ford.com, Phorm stores only the

[48] *See BT PLC, TalkTalk and Virgin Media Inc confirm exclusive agreements with Phorm*, Phorm, Inc. Feb. 14, 2008, *available at* http://www.phorm.com/about/launch_agreement.php (last visited Nov. 25, 2008).

[49] *See* Chris Williams, *ISP data deal with former 'spyware' boss triggers privacy fears*, THE REGISTER, Feb. 25, 2008, *available at* http://www.theregister.co.uk/2008/02/25/phorm_isp_advertising/ (last visited Nov. 25, 2008).

[50] *See* Chris Williams, *BT admits misleading customers over Phorm experiments*, THE REGISTER, Mar. 17, 2008, *available at* http://www.theregister.co.uk/2008/03/17/bt_phorm_lies/print.html (last visited Nov. 25, 2008).

[51] *See* Philip Stafford, *BT to begin further trials of ad technology*, FT.com, July 16, 2008, *available at* http://www.ft.com/cms/s/0/34c59420-5356-11dd-8dd2-000077b07658.html (last visited Nov. 25, 2008).

[52] *See Privacy : Phorm : No Personal Information*, PHORM, INC., *available at* http://privacy.phorm.com/no_personal_info.php (last visited Oct. 20, 2008).

[53] *Id.*

fact that the user (identified only by the Phorm identification marker) requested a page indicating interest in an SUV (or perhaps even in a Ford Expedition) rather than the details of the content or the specific URL of the request.[54] Phorm uses these interest categories to serve targeted ads to customers in spots on web pages that it sells to content producers.[55] Further, Phorm installs machines inside its participating ISPs that inject its own cookie into every request, allowing it to track which user is connecting to which site and to identify each user to its targeted ad server.[56]

c. Pakistan: YouTube and AnchorFree

In February 2008, the Pakistani state told all Pakistani Internet service providers to block local access to a specific URL on the popular U.S.-based video-sharing service, YouTube.com.[57] The authorities ostensibly sought to block Pakistani citizens from viewing a video hosted on YouTube that was perceived to be too critical of Islam.[58] One of the Pakistani ISPs responded not just by blocking access to the single URL within Pakistan, but instead by blocking access to the entire YouTube.com site (and, in the process, blocking global access to YouTube.com for a short while through a network anomaly).[59]

During those few days that Pakistan was blocking YouTube.com locally, many Pakistanis continued to access YouTube.com by using tools to circumvent the filter, including Hotspot Shield, a free tool by a company called AnchorFree.[60]

---

[54] *Id.*

[55] *Id.*

[56] *See* Richard Clayton, *The Phorm "Webwise" System*, May 18, 2008, *available at* http://www.cl.cam.ac.uk/~rnc1/080518-phorm.pdf.

[57] Posting of Sally Walkerman to OpenNet Initiative Blog, *available at* http://opennet.net/blog/2008/02/pakistan%E2%80%99s-internet-has-a-bad-weekend (Feb. 25, 2008 12:10 EST).

[58] *Id.*

[59] Posting of Ethan Zuckerman to My heart's in Accra, *available at* http://ethanzuckerman.com/blog/2008/02/25 (Feb. 25, 2008 15:41 EST).

[60] Sylvie Barak, *Pakistan Becomes VPN Routing Hot-spot*, THE INQUIRER, Feb. 26, 2008, *available at* http://www.theinquirer.net/gb/inquirer/news/2008/02/26/pakistan-becomes-vpn-routing (last visited Nov. 28, 2008).

AnchorFree describes its Hotspot Shield tool as: "Hotspot Shield creates a virtual private network (VPN) between your laptop or iPhone and our Internet gateway. This impenetrable tunnel prevents snoopers and hackers from viewing your email, instant messages, credit card information or anything else you send over a wireless network."[61]

But AnchorFree is, at its core, an advertising company which tracks the web browsing habits of its users.[62] What AnchorFree's description does not say is that AnchorFree makes money from the otherwise "free" tool by displaying ads to users within web pages. By using HotSpot Shield, users give AnchorFree complete access to all data exchanged while web browsing with the tool. AnchorFree provides an option to its users to remove the inserted banner ads and implies (but never explicitly says) that it does not monitor its users' traffic, but it nonetheless still has the ability to snoop on the data at any time. Thus, while the Pakistani state is watching its citizens' Internet traffic to block content it does not like, citizens are watching the blocked content by using a tool that circumvents Pakistan's filters. However, the circumvention tool itself could function as a monitoring tool, albeit likely for a different purpose.

3. Network Surveillance Data

Three sorts of data are vulnerable to surveillance on the network: (1) routing information, (2) the actual content of the data stream, and (3) contextual signatures. All Internet data packets must include the Internet Protocol (IP) address of the ultimate recipient. The Internet works more or less like a guided game of hot potato, with each packet of data getting passed through a series of intermediate routing computers that pass the packet on to another router closer to the end destination. In order to perform this function, every router along the way must include the address of the ultimate

[61] AnchorFree Inc., *AnchorFree History*, *available at* http://anchorfree.com/downloads/hotspot-shield/ (last visited Oct. 15, 2008).

[62] AnchorFree, Inc., *How it Works*, *available at* http://anchorfree.com/advertisers-agencies/how-it-works/ (last visited Nov. 11, 2008).

destination. Most data packets, all using the TCP protocol, including all web and email traffic, also include the IP address of the sender, such that data can be sent back along the same channel.

Users can hide routing information on the network by using a proxy that forwards communications between a client and a server. Using a proxy allows a user to include the proxy's address as the recipient in the packet routing data and move the ultimate recipient data into the content of the packet proper. The proxy receives the packet, looks inside the packet content for the ultimate recipient, and forwards the packet on to the recipient. Routers between the sender and proxy that look at the packet routing data see only that the client is sending a packet to the proxy. Likewise, routers between the proxy and the recipient that look at the packet routing data see only that the proxy is sending a packet to the recipient. No single router, other than the proxy itself, sees that the sender is communicating directly with the recipient.

In addition to these routing data, the packets contain the content of the data. The content includes both protocol-specific data and the content proper of the communication. The protocol-specific data includes data about the URL requested, the referring URL, the user agent, and so on for web requests as well as data that describe the originating email server, the from email address, the date and so on for emails. The content proper of the communication includes, but is not limited to, any data submitted to web server forms, any web pages retrieved, any emails sent or received, and any videos watched or songs downloaded. Any communications of any significant size are divided into separate packets for travel across the network, making it more expensive but not impossible for routers along the way to recreate the entire content objects (whole web pages, videos, and so forth) than to snoop on the routing information of a given packet or to look for simple patterns like keywords in a single packet.

Users can hide the content of their communications on the network by encrypting their messages, preventing anyone other than the sender or recipient from monitoring the data in

between the end-points. Encryption is most effective when it is applied end-to-end, so that the entire stream of data from the client to the server is encrypted, allowing no one on the network between the sender and the receiver from reading the content. Secure web (HTTPS) and email (SMTPS, IMAPS, POPS) protocols, when properly configured, are examples of this sort of end-to-end encryption. End-to-end encryption has to be negotiated by both the client and the server; however, many servers do not support encrypted communication, or only do so for some subset of content or functionality. In these cases, a user can connect through a proxy to encrypt the data from the client to the proxy. But such encrypting proxies still have unencrypted channels that enable them to talk to servers that do not support encryption, such that the proxied content remains readable to some routers on the network.

Communications that are both encrypted and proxied can hide both the routing information for and the content of a communication from the network between the client and the proxy. But even proxied and encrypted data leaks some information about the communication between the parties: the timing, number, and size of the packets as well as the fact that the communication is proxied and encrypted give away information about the transaction. In some cases, the mere fact that the data is encrypted prompts suspicions about the users and gives rise to further surveillance. Different sorts of traffic generate different signatures of packet size and timing that can allow easy identification of the nature of the communication.[63]

4. Network Surveillance and the Problem of Trust

The number of links on the Internet is the same as the sum of the number of clients, servers, and routers, roughly speaking,

[63] *See* Charles Wright, Lucas Ballard, Scott Coulls, Fabian Monrose, & Gerald Masson, *Spot Me if You Can: Recovering Spoken Phrases in Encrypted VoIP Conversations* (forthcoming, Proceedings of IEEE Symposium on Security and Privacy, May, 2008); *see also* Charles Wright, Lucas Ballard, Fabian Monrose & Gerald Masson, *Language Identification of Encrypted VoIP Traffic: Alejandra y Roberto or Alice and Bob?* (Proceedings of the 16th USENIX Security Symposium, Boston, Aug. 2007) (for discussions of traffic analysis and data leakage).

all the computers hooked up to this global network of networks. If monitoring the network required monitoring each one of those hundreds of millions of links, network monitoring on a large scale would be prohibitively expensive for even the most well-funded law enforcement or intelligence agencies. In practice, all Internet traffic flows through a much smaller number of routers, and those routers are controlled by an even smaller number of Autonomous Systems (AS) that agree to use the Border Gateway Protocol (BGP) to direct the traffic amongst themselves. These AS's are generally commercial ISPs, though they can also be managed by companies, governments, universities and other large organizations. There are fewer than a hundred thousand of these AS's in the world. The vast majority of Internet traffic flows through a much smaller number of those AS's.[64] Traffic within a given country generally stays on AS's located within that country, with some exceptions for countries with very little Internet usage. For a combination of technical, business, and policy reasons, a disproportionate amount of global Internet traffic flows through a few very large AS's in the United States, including most traffic between Europe and Asia.[65]

This topology of the Internet has a few different implications for the actors trusted with access to Internet data. The first is that a large majority of users need to access the Internet through an AS, usually a commercial ISP. These users are therefore vulnerable to surveillance by that AS itself. The second is that there is a relatively small number of these AS's within any given country, so monitoring all of the network traffic in a given country is a relatively manageable task; the state simply needs to require a small number of AS's to monitor their networks. This applies doubly for international Internet traffic, which is controlled by an even smaller number of AS's that are disproportionately located in the United States. Thus

[64] *See* BGP: The Border Gateway Protocol Advanced Internet Routing Resources, *Global Internet Exchange Points / BGP Peering Points / IXP*, *available at* http://www.bgp4.as/internet-exchanges (last visited Oct. 9, 2008).

[65] Cooperative Association for Internet Data Analysis, *IPv4 Internet Topology Report as Internet Graph*, *available at* http://www.caida.org/research/topology/as_core_network/ (last visited on Oct. 9, 2008).

monitoring international traffic within a given country requires the participation of even fewer AS's. The United States is capable of monitoring a large portion of international Internet traffic through a few of its AS's, including even traffic flowing between two non-U.S. countries.[66]

If a citizen wishes to hide her data on the network, her most likely strategy would be to move her trust around. In the process, though, the user must trust one or more actors with his or her data. One common approach to sharing trust in this manner is to use the Tor system to defend oneself against traffic analysis. In so doing, however, a user relies to a certain extent upon others on the network who are operating parts of the Tor network and whose machines relay information to others on the network.[67] Likewise, a user who tries to get around Phorm surveillance of her local ISP connection has only a few choices of ISPs in the United Kingdom, most of which have been reported to be considering adding Phorm monitoring to their networks. The user may choose to stay on the possibly monitored local network, but use a service like Relakks to proxy and encrypt her data as it travels through the local, Phorm monitoring, ISP. The user will avoid Phorm monitoring in the process, but at the cost of trusting not only Relakks but also the Swedish ISP through which Relakks talks to the Internet *and* the Swedish government that has legislated access to the data flowing through all Swedish ISPs. Her local ISP will still have the ability to tell that she is proxying and encrypting all of her data through a third party, which fact itself might prove suspicious if queried by a law enforcement. Finally, the user's data is vulnerable to network monitoring at any point along which it travels, from her local ISP to the server's ISP to any ISPs between that carry the data. Even setting aside the monitoring possibilities on her local ISP, a large portion of her data is likely to travel outside of Europe and is there likely to flow through one of a few U.S. ISPs that process a disproportionate share of international Internet traffic. The point is that, in evading

[66] *Id.*

[67] *See* http://www.torproject.org/ (last visited Oct. 8, 2008).

traffic analysis and other forms of online surveillance, a user will have to trust someone—or a series of people—to some extent in order to communicate via the network.

B. Server Surveillance

Surveillance also takes place commonly at the server level. In this taxonomy, servers include the computers that large corporations use to provide access to online services to consumers. Far more data are likely collected each day, by more parties, at the server level than through the network surveillance examples described above. But rarely do states directly have access to these data; a law enforcement or intelligence official would have to obtain it through the corporate or other institutional data-holder.

1. The Fuss About Google

Americans have begun to worry that perhaps Google's engineers know too much about what we do. To address these worries, Google has published a series of videos on YouTube that explain the privacy implications of the data collected by its search engine. A spokeswoman on one of these videos is named Ms. Ohye, a professional, reassuring-sounding support engineer. In the video, she explains what sorts of data Google collects and, implicitly, why users should not be particularly concerned about the data collection. Ms. Ohye says:

> To improve our search results, as well as maintain security and prevent fraud, we remember some basic information about searches. Without this information, our search engine wouldn't work as well as it does or be as secure. . . . We're able to [replace 'carss' with 'cars'] because we've studied search queries in our logs and found that when people type in 'carss' they often really mean 'cars.' . . . Only your provider can directly match you with your IP address. . . . What a cookie doesn't tell Google is personal stuff about you like where you live and what your phone number is. . . . In the same way that a store keeps a receipt of your purchases, Google keeps a type of receipt of

> your visit called a log. . . . As you can see, logs don't contain any truly personal information about you.[68]

User search data does help in the way Google says it does ("improve our search results," "maintain security," and "prevent fraud"). Ms. Ohye's descriptions of the way Google collects data are true in a technical sense. But the importance of data is determined by the larger world in which it lives—by the other data to which it connects. So when Ms. Ohye asserts that a cookie does not tell Google "personal stuff about you like where you live,"[69] this is only true in the sense that the cookie that contains your driver's license number does not tell the police where you live. However, even though the cookie itself is just a random string of letters, as with a driver's license number, it can indeed be used to lookup personal information "like where you live."

For example, the cookie might connect, reasonably well, all searches performed by a single person. Many people search for their own names and addresses at some point (if for no other reason than to see their houses in Google Maps). The cookie might connect those two searches to the same (otherwise pseudonymous) person, thus potentially identifying the name and address of the person behind the random gibberish of a particular cookie. This method of identification is not perfect, but researchers have consistently shown the ability to crack the identity of individual users in these kinds of data collections with anonymous but individually unique identifiers, most dramatically on supposedly anonymized data sets released by AOL and Netflix.[70]

---

[68] Google, Inc., Google Search Privacy: Plain and Simple, *available at* http://www.youtube.com/watch?v=kLgJYBRzUXY (Aug. 8, 2007).

[69] *Id.*

[70] *See* Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, *available at* http://www.nytimes.com/2006/08/09/technology/09aol.html?ex=1312776000 (last visited Nov. 28, 2008); Arvind Narayanan & Vitaly Shmatikov, Robust De-anonymization of Large Datasets, *How To Break Anonymity of the Netflix Prize Dataset*, *available at* http://arxiv.org/abs/cs/0610105 (last visited Nov. 28, 2008). *See also* Christopher Soghoian, *The Problem of Anonymous Vanity Searches*, 3 I/S: J.L. & POL'Y FOR INFO. SOC'Y 299, 300 (2007).

In fact, it is likely that this collection of search terms, IP addresses, and cookies represents one of the largest single collections of personal data on or offline. Google may or may not choose to do the work necessary to translate its collection of search data into a database of personally identifiable data, but it does have the data and the ability to query personal data out of the collection at any time if it chooses. Even if Google's staff never chooses to access these data, an intruder or disgruntled employee may query the data himself, or a state may demand access to it, with the force of law behind its request. And even assuming perfect security, Google is still subject to many sorts of government requests, as demonstrated by the success of the recent subpoena by Viacom for Google's entire retained history of which users watched which YouTube videos.[71]

One of the key Google services, AdWords, is a contextual online advertising system that displays ads on Google and its partner search sites, major content sites like the *New York Times*, and, through the AdSense program, on a huge number of small content providers like blogs. Through AdWords, Google targets ads to relevant consumers by displaying ads based on the current interests of a user. For ads on search pages, it displays ads related to the keywords entered by the user for the given search. For example, a user who searches for "digital camera" will see ads targeted to people interested in digital cameras. For ads on content providers, the system displays ads relevant to the keywords present in the content itself. So an article on digital cameras on a blog will draw AdWords ads about digital cameras. Advertisers bid for the right to advertise on desired keywords. An advertiser who wants to advertise to users interested in digital cameras has to win a continuous action for the term "digital camera" (or one of a host of related terms). At any given time, the ads that appear with a given term are those whose owners have made the highest bids for a single click on an ad that appears alongside a search (or a page of relevant content) for the bid keyword.

---

[71] Miguel Helft, *Google Told to Turn Over User Data of YouTube*, N.Y. TIMES, July 4, 2008, *available at* http://www.nytimes.com/2008/07/04/technology/04youtube.html (last visited Nov. 28, 2008).

Google watches consumers click on ads through its AdWords system and watches the advertisers through AdWords auctions that determine the value of advertising topics. Content providers watch the value of those advertising topics to determine which sorts of content to publish. The advertisers use the Adwords system as a stateless form of market research to target consumers without knowing anything about them. Content providers watch consumers to determine which sorts of content generate the most interest. The combined effect of these activities is to create a system of surveillance that collectively monitors what topics content producers are writing about, what topics users are searching for, and what the value of those topics (or rather the users who are reading about those topics) is to advertisers.

All of the monitoring that happens within this system of surveillance happens in real time. Google adjusts the placement of ads in real time according to the current results of the auction, content providers watch the profitability of their content in real time and make adjustments to attract more ad clicking customers, and advertisers adjust their bids and update their ads in real time to attract more users. The effect of the system is continuous but stateless market research that is constantly adjusting to the current interests of users rather than the historical interests over time tracked by user profiling organizations like comScore, Phorm, and NebuAd.

2. Server Surveillance Data

A server has access only to data that is sent directly to it over the network, which consists of the client address, the time of the communication, and the protocol-specific request (in the case of an http server, the URL, referrer, user agent, and any cookies). The server also likely captures the content explicitly submitted by the user, including names, addresses, credit card numbers, health problems, sexual proclivities, and so on. All of these data, other than the client and the time of the transaction, may be encrypted while traveling over the network, giving the end server access to some data that is not available on the network. For web servers, the protocol data may include

cookies, which are often used to assign to the requesting user a pseudonymous identity that remains the same between separate requests.

3. Server Surveillance Trust

People voluntarily, if not exactly knowingly, submit vast amounts of data to Internet servers. When one pauses to think about it, one may be aware that one is submitting names, addresses, and credit card numbers to Amazon when buying a new book, personal email to Microsoft through the popular free service Hotmail, movie preferences to Netflix when renting videos, medical information when using online personal health services, and all manner of personal information when connecting to others through social networks or online dating services.[72] What determines the risk of privacy intrusions is not so much the data itself as what the actors controlling the servers do with the data, with whom they share it, and how the data are later combined with other data. So, for some users, allowing a company to collect a credit card number to execute a purchase is fine, but using that same credit card number to request data about the user's purchase history from the credit card company in order to target advertising to him or her is not fine. Likewise, it may be fine for Microsoft to collect personal emails through Hotmail, but it becomes a concern if Microsoft were to sell its users' email content to a consumer research company. A consumer may be pleased that Netflix uses video preference information for its own recommendation engine, but many users would be uncomfortable if they found out that Netflix was combining its users' video rental history with (even public) information from users' Bebo social networking pages to make

[72] In a further twist to this story, users sometimes seek to protect communications by using supposedly secure services, only to have those service providers turn over data to law enforcement officials. *See, e.g.*, http://blog.wired.com/27bstroke6/files/steroids.source.prod_affiliate.25.pdf (last visited Nov. 28, 2008). One prominent case involved the Canadian email provider Hushmail, which, according to court documents, assisted U.S. law enforcement authorities by turning over emails relating to the activities of suspected criminals in the United States. *Id*.

video recommendations.

All of these different possible uses of data represent networks of trust which the user must evaluate. The Google story provides an example of the complexity and confusion surrounding these issues of trust. Studies of users of all ages, but especially young people, suggest that few have a clear idea about how data is collected, who has access to it, and what is done with it.[73] The potential of Google's vast store of user data creates a serious risk of disclosure throughout this network of trust, regardless of Google's intentions.

Sometimes the risk is more than theoretical. Over the past several years, many users around the world have come to rely upon the Skype service for voice over Internet protocol calling and online chat. The service has often been touted as among the most secure ways for dissidents and activists to communicate.[74] However, Internet researcher Nart Villeneuve found that "[t]he full text chat messages of TOM-Skype users, along with Skype users who have communicated with TOM-Skype users, are regularly scanned for sensitive keywords, and if present, the resulting data are uploaded and stored on servers in China."[75] In other words, while Internet activists were trusting Skype with their communications, keyword monitoring of their chat in China directly resulted in the potential for state surveillance of their supposedly private communications.[76]

---

[73] PALFREY & GASSER, *supra* note 4, at 39-82.

[74] *See* Robert Amsterdam, *Technology's Threat to Human Rights and Free Speech*, (Oct. 31, 2008), *available at* http://www.robertamsterdam.com/2008/10/technologys_threat_to_human_ri.htm (last visited Nov. 25, 2008).

[75] NART VILLENEUVE, INFORMATION WARFARE MONITOR/ ONI ASIA JOINT REPORT, BREACHING TRUST: AN ANALYSIS OF SURVEILLANCE AND SECURITY PRACTICES ON CHINA'S TOM-SKYPE PLATFORM 4 (Oct. 1, 2008), *available at* http://www.infowar-monitor.net/breachingtrust.pdf (last visited Nov. 28, 2008).

[76] John Markoff, *Surveillance of Skype Messages Found in China*, N.Y. TIMES, Oct. 2, 2008, at C1, *available at* http://www.nytimes.com/2008/10/02/technology/internet/02skype.html (last visited Nov. 28, 2008).

4. Server Surveillance Implications

Google's current policy is to keep a log of each search query with a complete IP address for eighteen months and then remove a portion of the IP address to make it harder to resolve back to a specific user.[77] Data retention rules, such as the European Union Data Retention Directive, mean that a given firm may be—or perceive itself to be—compelled to retain more information about users for greater periods of time. The scrubbing of server data provides only limited protection against identification since the combination of search terms over time can still be used to resolve the data to individual users. The combination of technological change, rules and practices of data retention, and individual willingness to share private information with firms gives rise to a growing potential for data related problems at the server level.

C. Client Surveillance

1. Government Surveillance at the Client Level

Governments have various levels of control over and access to the data collected by client side software, but they also use this software to directly collect data in certain instances. The uses of client side software can be difficult to uncover, but the practice of using it is well-known. For example, United States court documents reveal that in 2007 the U.S. Drug Enforcement Agency (DEA) installed a keylogger on a suspect's machine to capture the encryption keys necessary to read the suspect's PGP (Pretty Good Privacy) encrypted email.[78] In fact, the alleged New Jersey organized crime boss, Nicholas Scarfo, was caught through the FBI's use of a keylogger on a personal computer.[79]

---

[77] Posting of Chris Soghoian to CNET news blog, *available at* http://news.cnet.com/8301-13739_3-10038963-46.html (Sept. 11, 2008, 7:40 PDT).

[78] Declan McCullagh, Feds Use Keylogger to Thwart PGP, Hushmail, CNET News Blog, (July 10, 2007), *available at* http://news.cnet.com/8301-10784_3-9741357-7.html (last visited Nov. 28, 2008).

[79] Electronic Privacy Information Center, United States v. Scarfo, Criminal No. 00-404 (D.N.J.), *available at* http://epic.org/crypto/scarfo.html (last visited Oct. 12, 2008).

Some major anti-virus companies have received and complied with court orders to ignore the presence of government agency spyware when scanning for viruses.[80] Similarly, there is evidence that the German police are aggressively pursuing the use of client side software to calls placed on the Skype Internet telephone service.[81] The software used in both cases includes features identical to those in malware, including remote installation via a trojan or virus, remote control and data reporting, and stealth features (and policies) that hide the software from the user.

A variety of companies make keyboard logging devices that are hardware, rather than software. These devices tend to take the form of small plugs that sit between the USB plug of the logged keyboard and the USB plug of the logged computer. The devices record every key pressed on the keyboard and therefore can be used to capture passwords, emails, typed documents, and any such information entered on the keyboard. These devices are much easier to install than software keyloggers, so long as the installer has physical access to the machine, and they are impossible to detect via anti-virus software. A company called KeyCarbon distributes a PCI keylogger.[82] As a PCI card, this device is attached to the motherboard inside the computer case, giving it all of the advantages of a USB keylogger (though requiring a few minutes instead of a few seconds to install) but making the device much more difficult to detect via visual inspection.[83] The only way to detect this device is physically to open the computer case and then look inside the box in the right

[80] Posting of Declan McCullagh & Anne Broache to CNET News blog, *available at* http://news.cnet.com/Will-security-firms-detect-police-spyware/2100-7348_3-6197020.html (July 17, 2007, 9:40 PDT).

[81] Posting of Louis Charbonneau to Reuters News Blog, *available at* http://www/reuters.com/article/technologynews/idusl21173920071122 (Nov. 23, 2007, 12:28 EST); *see also* posting of Kim Zetter to Wired Blog Network, *available at* http://blog.wired.com/27bstroke6/2008/01/leaked-document.html (Jan. 29, 2008, 16:10 EST).

[82] KeyCarbon Computer Security Hardware, *available at* http://www.keycarbon.com/products/keycarbon_laptop/overview/ (last visited Oct. 13, 2008).

[83] *Id.*

place.[84]

2. Market Research

ComScore Incorporated is one of several large companies that collect data about Internet users for market research. ComScore collects the entire web browsing stream, including encrypted requests, from the two million members of its "consumer panel."[85] It claims to try to avoid capturing data like usernames, passwords, and credit card numbers, but actively collects personal information that directly identifies the owner of the web browsing data.[86] ComScore touts the ability to connect this online data to a variety of other consumer databases, including supermarket purchases and automobile registrations.[87] It has admitted in the past to using its monitoring data to look up information in its panel members' online banking accounts to verify the household incomes reported in surveys.[88]

ComScore recruits these panel members through a combination of sweepstakes, network performance improvement tools, claims of anti-malware protection, and an appeal to those who seek to improve the efficiency of the Internet.[89] ComScore discloses to the panel members that the software is monitoring their web browsing activities, but it also keeps a strong separation between itself and the operations that collect the data, currently OpinionSquare and PermissionResearch, by not directly naming the tools or the organizations that operate them anywhere on comScore.com or even in its SEC annual report filing, and it has had to recreate those operations at least once to

---

[84] *Id.*

[85] comScore, Inc., comScore Methodology, *available at* http://www.comscore.com/method/method.asp (last visited Oct. 13, 2008).

[86] *Id.*

[87] *Id.*

[88] Posting of Evan Hansen to CNET News Blog, *available at* http://www.news.com/Net-privacy-and-the-myth-of-self-regulation/2010-1071_3-281580.html (Oct. 16, 2001, 4:00 PDT).

[89] *See* comScore, *supra* note 85.

evade detection by anti-spyware tools.[90] ComScore sells this vast trove of data, estimated by comScore at 28 terabytes collected per month in 2007, as market research to many of the largest companies in the world.[91] ComScore is far from alone in collecting vast amounts of data about individuals for the purposes of greasing the wheels of commerce, online and off; ChoicePoint, its parent company LexisNexis, and many others are in the business of serving as "data brokers."[92]

3. Malware and Antivirus Software

Though the furthest cry from the man in the van with headphones, malware installed on an individual's computer can perform some of the most intrusive surveillance possible using today's technology. A motley assortment of computer programs and related systems called bots, virus, worms, trojans, and other malware infect up to twenty-five percent of all computers connected to the Internet.[93] All of these malware programs are surveillance devices: they have complete access to all data entering, leaving, or residing on the client computer. Malware programs generally use their access to user data to perform a variety of illicit activities including sending spam, committing click fraud, attacking other computers, and stealing financial or other identifying information from the client. The direct impact of most of these activities on any given infected user is relatively small—outgoing spam only costs the user bandwidth and even

[90] Hoovers, United States Securities and Exchange Commission Form 10-K Annual Report for comScore, Inc., *available at* http://www.hoovers.com/free/co/secdoc.xhtml?ID=113840&ipage=5792564 (filed on Mar. 11, 2008); *see also* posting of Stefanie Olsen to CNET News Blog, *available at* http://news.cnet.com/ComScore-Spyware-or-researchware/2100-1032_3-5494004.html (Dec. 20, 2004, 4:00 PST).

[91] *See* Tom Taulli, *comScore Busts a Move*, The Motley Fool, June 28, 2007, *available at* http://www.fool.com/investing/high-growth/2007/06/28/comscore-busts-a-move.aspx (last visited Nov. 11, 2008).

[92] Martin H. Bosworth, *Lexis-Nexis Parent to Buy Choicepoint*, (Feb. 23, 2008), *available at* http://www.consumeraffairs.com/news04/2008/02/choicepoint_sale.html (last visited Nov. 30, 2008).

[93] Tim Weber, Business Editor, BBC News Online, *Criminals may overwhelm the web*, (Jan. 25, 2007), *available at* http://news.bbc.co.uk/1/hi/business/6298641.stm (last visited Nov. 28, 2008).

credit card theft is generally insured by the credit card company.

These malware programs have the potential to exploit more directly the vast trove of personal data stored on the vast number of infected machines, as demonstrated by the Sircam worm in 2001.[94] Like many worms of its era, the Sircam worm propagated by sending a copy of itself to every email address in each infected computer's address book.[95] Unlike other worms, however, Sircam tried to fool its targets into opening the infecting payload by attaching a trojan to a document picked randomly from the victim's own computer.[96]

The primary effect of the worm, other than clogging up the network, was to send the randomly chosen private files of over a million people to a list of their friends, families, colleagues, and digital acquaintances.[97] The *New York Times*, for instance, reported receiving a travel diary, a document about a father's will, and a memo from a real estate branch manager that "offers colleagues 'Objectives For The Year 2000,' the first of which is, 'Positive attitude at all times.'"[98] Others reported receiving "memos, CVs, job listings, diary entries, expense forms and complaint letters."[99] For many, the disclosure of these documents, particularly to friends and colleagues, represented serious breaches of privacy with significant personal repercussions.

Most malware, circa 2008, tries to hide itself from detection. It tends to avoid heavy-handed tactics like those of the Sircam worm. Modern malware largely restricts itself to hijacking the

---

[94] Symantec, w32sircamworm@mm_2007, *available at* http://www.symantec.com/security_response/writeup.jsp?docid=2001-071720-1640-99&tabid=2 (last visited Oct. 13, 2008).

[95] *Id.*

[96] *Id.*

[97] *Id.*

[98] Richard Meislin, Compressed Data: A Virus, Yes, but One That Brings Interesting Things, N.Y. TIMES, (July 30, 2001), *available at* http://query.nytimes.com/gst/fullpage.html?res=9C00EEDA103DF933A05754C0A9679C8B63 (last visited Nov. 28, 2008).

[99] Mark Ward, Technology Correspondent, BBC News Online, *Sircam virus steals files*, (July 24, 2001), *available at* http://news.bbc.co.uk/1/hi/sci/tech/1454155.stm (last visited Nov. 28, 2008).

user's identity to send spam or commit click fraud or to stealing profit-generating information like credit card numbers. The Sircam story, however, demonstrates that these forms of malware have the capability to breach the privacy of millions of infected users in much more serious ways than stealing credit card numbers. The collection of malware-infected computers as a whole represents the single largest surveillance device on the Internet, capable of collecting and processing the personal data of probably hundreds of millions of computers.

Viruses, too, can collect personal information. On Christmas Day 2007, the SANS Internet Storm Center received a report from a user who had found trojan malware on a new digital photo frame purchased from a large retailer, Sam's Club.[100] Within a week, two more readers of the SANS blog reported finding the same virus on other digital photo frames; and, within a couple of weeks, readers reported similar viruses on a wide array of different consumer devices that acted as USB storage devices, including MP3 players, portable hard drives, and video cameras.[101] The malware found on many of these devices was the MocMex trojan horse, a powerful piece of malware capable of hiding itself from a variety of anti-virus tools.[102]

To protect from viruses, bots, worms, and other such malware, millions of computer users install anti-virus systems on all client computers. But this anti-virus software is itself highly intrusive, operating at the lowest levels of the operating system, incurring significant performance penalties, and attempting to avoid the notice of malware (which are themselves trying to detect the anti-virus systems to disable them). The extensive access given to anti-virus software gives it the capability to do the same sorts of harm that a piece of malware

[100] Digital Hitchhikers, *available at* http://isc.sans.org/diary.html?storyid=3787 (Dec. 25, 2007, 23:24 UTC).

[101] Digital Hitchhikers, Part Four, *available at* http://isc.sans.org/diary.html?storyid=3892 (Jan. 28, 2008, 10:37 UTC).

[102] Deborah Gage, *Virus from China the Gift that Keeps on Giving*, SAN FRANCISCO CHRONICLE, Feb. 15, 2008, at C-1, *available at* http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2008/02/15/BU47V0VOH.DTL&type=printable (last visited Nov. 28, 2008).

can do, including both stealing data from and disabling the host computer. The difference between a legitimate anti-virus tool and malware is the trustworthiness of the software distributor. Most people trust that Symantec—the provider of a common type of anti-virus software—is only using its complete access to protect the user's own computer, while few people trust that a malware developer or operator will not abuse its access. Internet users have little choice but to trust someone with their data.

4. Client Surveillance Data Implications

Client side surveillance provides the most complete access to user data. As such, surveillance techniques by the state that either gain access at the client level or obtain data collected by others will yield the most detailed picture of the activities of a suspect. It is for this reason, of course, that law enforcement officials often seize laptops and turn them over to a computer forensics expert to review.

Virtually every bit of data stored, sent, received, viewed, played, or typed on a computer is vulnerable to client-side surveillance. The only data that is potentially not accessible is encrypted data that is never accessed during the period of surveillance. Everything else on a computer is available to a client surveillance tool. The primary limiting factor is simply what the tool developers choose to target with their tools. Malware mostly targets various sorts of easily monetized data, including email addresses, credit card numbers, bank account information and login credentials, and game and other software keys. The most sophisticated anti-virus tools monitor all data stored on and transmitted to the computer, checking all data for malware signatures, but not keyboard or screen activity. Market research tools like comScore typically monitor all network traffic, whether encrypted or not, but not stored data or keyboard or screen activity. Workplace and family monitoring tools usually monitor keystrokes and sometimes screenshots of the minute to minute (or even second to second) activity on the computer screen but not stored or network data.

Client side surveillance tools are often designed not to track

certain types of data in order to provide a clearer, more focused set of data to the party carrying out the surveillance. In fact, the biggest problem of surveillance on the client side is often dealing with the sheer amount of data. For example, even one screenshot a minute on a single computer can generate a daunting amount of data. This problem is magnified when applied over a large set of monitored clients. Botnets (networks of malware-infected computers controlled from a single point) only search for a limited set of data, like credit card numbers, which can easily be monetized, presumably because of the difficulty (and, therefore, unprofitability) of sifting through the vast trove of other sorts of data on infected computers. Likewise, a primary challenge for corporate anti-virus systems that must manage entire networks of clients is to manage the resulting flood of data.

Any client side program has at least the *potential* to access every sort of data that resides on or passes through the computer. In fact, even non-surveillance oriented programs (screen savers, games, chat programs, and so on) have this access to the computer once they have been installed.[103] This same access applies to most hardware devices installed on the computer as well. Any device that has access to a shared channel—such as the USB, PCI, or IDE channels—has the ability to monitor all data on that bus directly, and any device that requires an operating system driver can use that driver to access the breadth of data available on the client computer. Even devices that do not directly have the ability to access a shared bus or run drivers may have the ability to infect clients with malware.

IV. IMPLICATIONS OF THE DIGITAL DIFFERENCE

The fact that digital technologies afford this increased level of surveillance is not inherently bad. There are, in fact, many

[103] Most modern operating systems have some level of access control that is intended to prevent ordinary programs from having complete access to the data on the computer. In practice, it is almost always possible to exploit a vulnerability that breaks through the access control system to give the client complete access to the system.

good reasons to exercise this power. From the perspective of law enforcement, these multiple modes of digital surveillance can lead to new and better ways to prevent crimes and to bring criminals to justice. And though the fear of aggregating these many sets of data, held in many hands, is true in concept, the reality is that it is technically difficult to pull disparate data sets together in a way that is helpful for purposes of surveillance.[104]

The issue raised by these changes in technologies and their usage is whether individual data-sharing practices and the safeguards for civil liberties are keeping up with the changes.[105] Of the many things that have changed in the digital era, the manner in which surveillance can be carried out by the state is high on the list of the more dramatic types of changes. Compare traditional surveillance to Internet-era surveillance and the differences become plain. In traditional surveillance, the law enforcement official primarily focuses on known targets and has access largely to live communications and imperfect records of past interactions. The data that can be tracked are mostly in rivers (such as tapping a phone conversation) rather than in the form of oceans (which might later be searched).

In the digital age, law enforcement officials have many more means of surveillance than in the past. (Of course, there are reasons to celebrate this fact in the interests of public security.) Law enforcement officials have the opportunity to

---

[104] Simson L. Garfinkel, *Information of the World, Unite!*, SCI. AM., 82-5 Sept. 2008.

[105] There is interesting literature related to privacy in a digital age and how we might reconceptualize it in light of recent changes in technology, which often expands the frame of a discussion such as the one here in this paper. As one example, Julie Cohen has recently written an article on the information privacy law project and its intersection with the Fourth Amendment doctrine. Julie E. Cohen, *Privacy, Visibility, Transparency, and Exposure*, 75 U. CHI. L. REV. 181 (2008). In this article, Cohen notes that privacy and technology literature focuses on informational privacy, but that spatial privacy is a relevant concept often overlooked in recent discourse. *Id*. at 183, 189-90. Cohen suggests that the spatial dimension of privacy is important, in light of internet usage, as "online conduct and online surveillance are experienced spatially." *Id.* at 194. When online activities are conceptualized as something one does in a "networked space," this "zone of personal space" is a meaningful and useful concept for privacy discourse. *Id*. at 195. Recognizing that online activities exist in a networked space also suggests that individual behavior changes within that space (for better or worse) in response to the awareness that one is being surveilled within that space. *Id.* at 195-96.

either focus on a known target or look for patterns across a network that might give rise to suspicion. Both rivers and oceans of information are available, depending upon the period during which data are retained in any given situation. Law enforcement officials have access to past, stored communications, which may or may not have seemed relevant during the initial surveillance. Data could be collected, placed in a large database, and later sorted, at much greater speeds and at lower costs than ever before. In addition, past actions can be recreated much more effectively based on third-party data collected in the course of everyday interactions. Profiles can be sewn together across data sources to create a more comprehensive profile of an individual or a pre-determined "type" of person.

The practice of non-targeted Internet surveillance is also rendered much easier in the digital era. A law enforcement official might establish a "driftnet" to capture all traffic through critical points in the network, much as the NSA is believed to have done through the black box in a closet at AT&T. Automated filtering and processing of traffic is far cheaper and faster than ever before, such that a non-targeted search might yield patterns or point to suspicious practices. This form of searching requires less human involvement than in the past. A person is needed for coding the system, setting filters, and reviewing results—but not for reviewing everything as it goes past. Such non-targeted surveillance might enable law enforcement officials to identify persons matching a predefined profile, to identify unusual patterns not predefined, or to record networks for unknown, future searching purposes.

In the digital era, governments also collect and review information about individuals for purposes not related to the traditional mode of law enforcement. For instance, government officials collect information for the purpose of record-matching between different government databases in order to ensure that welfare recipients are not perpetrating fraud.[106] The United

[106] *See* Virginia Eubanks, *Technologies of Citizenship: Surveillance and Political Learning in the Welfare System*, *in* SURVEILLANCE AND SECURITY: TECHNOLOGICAL POLITICS AND POWER IN EVERYDAY LIFE 89, (Torin Monihan, ed., 2006). *See also* JOHN

States Census Bureau collects highly sensitive personal data—such as income, race, ethnicity, physical handicap, and so forth—as part of its American Community Survey from a small sample of the population; recipients are told that they are "required by law" (Title 13 of the United States Code) to respond.[107] The handwritten data entered by citizens are then keyed into government databases.[108]

At the same time, most people are not able to keep up with all the changes in the technology that encode these data and otherwise enable this surveillance to happen in so many different ways. Through our research with young people, for instance, we found they were almost universally surprised to learn of the extent to which information was collected about them by third parties and the extent to which that information could be monitored, stored, and later searched.[109] As the possibilities for law enforcement to practice surveillance in concert with private third parties grow each year, public awareness of these practices is not keeping pace.

V. THE PUBLIC, THE PRIVATE, AND THE REASONABLE

Two primary concerns arise out of this fast-changing state of affairs. The first is about the convergence of the public and the private. The related concern is whether citizens are able to make reasonable choices about how they lead lives mediated by these technologies and what the consequences of those choices might be with respect to what the state can come to know about them.[110]

GILLIOM, OVERSEERS OF THE POOR: SURVEILLANCE, RESISTANCE, AND THE LIMITS OF PRIVACY (2001).

[107] Census.gov, ACS Congressional Tool Kit 1, *available at* http://www.census.gov/acs/www/SBasics/congress_toolkit/ACS%20Tool%20Kit.pdf (last visited Oct. 18, 2008); *see also* Census.gov, American Community Survey (ACS), *available at* http://www.census.gov/acs/www/index.html (last visited Oct. 18, 2008).

[108] *See* Census.gov, About the Data (Methodology): Data Collection & Processing, *available at* http://www.census.gov/acs/www/AdvMeth/CollProc/CollProc1.htm (last visited Oct. 18, 2008).

[109] PALFREY & GASSER, *supra* note 4, at 39-82.

[110] There are, to be sure, other fears to which this set of facts might give rise, such as the overall suite of powers granted to law enforcement authorities to carry out network

A. Shifting Meanings of Public and Private

There are two meanings of "public" and "private" that have salience in the context of surveillance. The first is from the perspective of the person who is observing someone else. Is this observer a public actor? And is this public actor relying solely upon data collected by him or herself, or instead relying upon data collected by private actors? Second, this distinction matters from the perspective of the person who is being observed. The issues involved in each one are different from the other.

1. The Perspective of the Observer

The Fourth Amendment regulates the conduct of state actors where that conduct might violate an individual's reasonable expectation of privacy.[111] From the observer's perspective, much turns on whether the actor falls into the category of a state actor—for shorthand, here, representing the "public"—and whether that state actor carries out the surveillance directly. In this sense, the meaning of the "public" makes perfect sense: if the actor is a state actor and is carrying out the surveillance directly on a known target, then the traditional Fourth Amendment analysis follows. This paper does not dwell on this topic since there is a rich body of scholarship on the question of when and whether a warrant should issue.[112]

The issue addressed in this paper focuses on when the observer is a public actor and when the relevant data have been gathered by a *private* actor. The question asked near the threshold of a Fourth Amendment analysis does not have to do with the data involved or how those data were initially gathered. Vastly more data are now gathered through digital means, at

surveillance. Orin S. Kerr takes up these general issues and debunks several of the common myths on this related topic. Orin S. Kerr, *Internet Surveillance Law after the USA Patriot Act: The Big Brother that Isn't*, 97 NW. U. L. REV. 607 (2003).

[111] *See* Smith v. Maryland, 442 U.S. 735, 739 (1979) (citing Katz v. United States, 389 U.S. 347, 361 (1967) (Harlan, J., concurring)).

[112] *See, e.g.*, William J. Stuntz, *Warrants and Fourth Amendment Remedies*, 77 VA. L. REV. 881 (1991).

the network, server, and client levels. So long as the collection of the data is performed by the private party, and the public party obtains the information from this private party, then the protections of the Fourth Amendment do not generally attach.

In practice, United States law affords some protections against public scrutiny of data collected by private parties. In many cases, the private party will refuse to give information to a law enforcement official without a subpoena or other formal order. Some companies perceive their business model to turn on whether they are trusted providers of Internet services, and as such make public their resistance to government demands, as Google successfully did against a Department of Justice request for a large amount of data on search queries; other firms, such as AOL, Yahoo!, and Microsoft, complied without a battle in court.[113]

What is different in an Internet era is that the vast majority of useful data about an individual is increasingly held in private hands—and individuals have a harder and harder time each year avoiding placing the data there in the first place. The emphasis in terms of legal protections on whether the observer is a public actor who carries out the search directly means that the core Constitutional protection is against "unreasonable searches."

2. The Perspective of the Observed

From the perspective of the person who is potentially observed—call her the "citizen"—"public" and "private" mean something different than what it means from the perspective of the person conducting the surveillance. What matters from the citizen's perspective is whether he or she has a reasonable expectation that the activities under surveillance are taking place in public or private.

There are three key problems that derive from this perspective in the digital age. The first is that the activity might be taking place in a context that the citizen believes is

[113] Katie Hafner & Matt Richtel, *Google Resists U.S. Subpoena of Search Data*, N.Y. TIMES, Jan. 20, 2006, at A1; Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D. Cal. 2006).

"private," but where a third party is recording that activity. The recording might be taking place in a perfectly lawful, consensual way. For instance, the citizen signed up for an Internet Service Provider's package that permits access to the Internet, an email account, and a back-up service. Similarly, as researchers have shown, young people in particular consistently perceive their online audiences as more (and occasionally less) "private" than they really are.[114] For instance, they might post information to a page that they maintain on a social network site, such as MySpace or Facebook, to which only friends have access. In each of these cases, information might be shared in a certain context which feels "private" but which is plainly open to surveillance of multiple kinds. Neither the Fourth Amendment nor privacy-related statutes would protect the citizen, even if the citizen's perception of the context was "private."

Federal courts in the United States have addressed this general issue in several matters, but the law remains unsettled. In *Warshak v. United States*, the Sixth Circuit held that "individuals maintain a reasonable expectation of privacy in e-mails that are stored with, or sent or received through, a commercial ISP."[115] In a demonstration of how shaky the law is on this subject, the *Warshak* opinion was vacated on October 9, 2007, granting the government a rehearing en banc.[116] On rehearing, the constitutional question was avoided, as the court asserted that plaintiff's claim for injunction against future government action was not ripe.[117] Though the original *Warshak* opinion was vacated, the matter continues to be briefed

[114] danah boyd, *Why Youth (Heart) Social Network Sites: The Role of Networked Publics*

in Teenage Social Life, *in* YOUTH, IDENTITY, AND DIGITAL MEDIA (David Buckingham ed., 2007), *available at* http://www.mitpressjournals.org/doi/pdf/10.1162/dmal.9780262524834.119 (last visited Nov. 28, 2008). Urs Gasser and I found much the same thing in our focus groups and interviews conducted while researching privacy matters for Born Digital.

[115] Warshak v. United States, 490 F.3d 455, 473 (6th Cir. 2007); *see also* Katz v. United States, 389 U.S. 347, 351 (1967) (stating that something an individual "seeks to preserve as private" is therefore "constitutionally protected").

[116] *Warshak*, 490 F.3d at 455.

[117] Warshak v. United States, 532 F.3d 521 (6th Cir. 2008).

as of this writing, and scholars have already begun to take up the possible implications of such a ruling in depth.[118]

State courts have wrestled with this issue as well. In *State v. Reid*, the Supreme Court of New Jersey held that under a search and seizure provision of the New Jersey Constitution, a citizen has a protected privacy interest in information they provide to an ISP.[119] On the other hand, in *State v. Delp*, the Oregon Court of Appeals held that under Oregon law, the defendant did not have a privacy interest in non-content information disclosed to an ISP.[120]

The second, closely related problem is the difficulty, in a digital age, for the citizen to keep anything truly private from third parties. As the citizen's life is increasingly mediated by digital technologies, her social life, her work life, her civic life, and any other lives she leads are often led partly in digital public spaces. For those who carry BlackBerries on their hips or send text messages constantly, as many young people do, an increasing array of information about us is potentially in the hands of many people we do not know. In most cases, we also do not know what is being stored, for how long, and when and how it might be accessed or sold, by whom or to whom.

The final problem is that it is very difficult for the citizen to keep up with the pace of technological change. The rate of development of new digital technologies is very fast, such that even technology experts have little sense of what is even commercially available in fields tangentially related to their own. Few people would be knowledgeable enough about digital technologies to have an effective sense of what information they are sharing is publicly accessible and what is private.

This problem of keeping up with generally available digital technologies is particularly important in the context of the

---

[118] Tamar R. Gubins, *Warshak v. United States: The Katz for Electronic Communication*, 23 BERKELEY TECH. L. J. 723 (2008). *See also* ZITTRAIN, *supra* note 14, at 188; Orin Kerr, The Volokh Conspiracy, A Series of Posts on Warshak v. United States, *available at* http://volokh.com/posts/1182208168.shtml (last visited Oct. 22, 2008).

[119] State v. Reid, 945 A.2d 26, 33-34 (N.J. 2008).

[120] State v. Delp, 178 P.3d 259, 264 (Or. Ct. App. 2008).

Fourth Amendment. *Kyllo v. United States* holds an important place in the doctrine on Fourth Amendment law with respect to new technologies.[121] In *Kyllo*, the Court held that the use of technology to obtain any "information regarding the home's interior that could not otherwise have been obtained without physical intrusion into a constitutionally protected area, constitutes a search—at least where (as here) the technology in question is not in general public use."[122] The last part of the *Kyllo* test is particularly important given the topic of this paper. The "general public use" test is vague and hard to implement. A search conducted with technologies that would seem to be in "general public use" may not be well known to a large percentage of a given population. Therefore, the use of new information technologies to conduct a search may not merit Fourth Amendment protection, even when the person being observed might expect the information to be protected.

B. Protecting What is "Reasonable"

The "reasonable expectation of privacy" standard is tricky even for law professors to understand. It is the most debated, puzzling aspect of Fourth Amendment jurisprudence, and scholarship.[123] For ordinary people, it is only getting more difficult—in a technological age—to develop a sense of what is reasonable. The range of possible surveillance methods at the network, server, and client levels underscore this difficulty facing citizens. Moreover, citizens face a dual bind: they need to decide whom to trust with their data (often private parties), but they also have to wonder—at the point of giving up control over their data—about who else might later access those data (often public entities).

As more and more of our lives are subject to these various forms of surveillance over "public" networks, it is crucial that some of our basic communications are treated as "private" under law. Some lower courts, in interpreting the Fourth Amendment,

---

[121] Kyllo v. United States, 533 U.S. 27 (2001).

[122] *Id*. at 28.

[123] Kerr, *Four Models of Fourth Amendment Protection*, *supra* note 16, at 504.

have suggested that sometimes actions in a public place, if meant to be private, might give rise to Constitutional protection. As one court held, "what a person knowingly exposes to the public through an open door or window does not receive Fourth Amendment protection, yet what a person tries to keep private, even in a public place, may be entitled to constitutional protection."[124] But the question of when data in public places online, if ever, give rise to such protections has not been made clear.

When it comes to the third-party data problem, though, there is little risk of misinterpreting the Supreme Court's current position. In 1976, the Court made plain that citizens have no reasonable expectation of privacy in data given by them to third parties in *United States v. Miller*.[125] Courts have long held that even evidence unlawfully obtained by a private party may later be used by a state actor against a defendant in a criminal proceeding.[126] The Fourth Amendment is usually implicated if a state actor *participates* in a search by a private party, or requests that a private party initiate the search, but that is not ordinarily the case in network, server, or client surveillance on the Internet.[127] Courts have held that a government search which occurs *after* a completed private search is lawful without a warrant if it does not exceed the scope of the private search.[128] The risk of government review of those data falls on the part of the person who deposited that information.[129]

In 1979, in *Smith v. Maryland*, the Court elaborated on the

---

[124] United States. v. Davis, 326 F.3d 361, 365 (2d Cir. 2003).

[125] 425 U.S. 435, 442-43 (1976).

[126] Burdeau v. McDowell, 256 U.S. 465, 476 (1921).

[127] *See generally* 1 WAYNE R. LAFAVE, SEARCH AND SEIZURE: A TREATISE ON THE FOURTH AMENDMENT § 1.8(b) (4th ed. 2004) (detailing that the exclusionary rule applies to private searches or non-police searches where the government has requested the search, or has become involved with the search either at the beginning or at any time before the search is concluded).

[128] United States v. Jacobsen, 466 U.S. 109, 115 (1984) (noting that governmental actions pursuant to a private search of an individual's property or possessions "must be tested by the degree to which they exceed the scope of the private search").

[129] United States v. Miller, 425 U.S. 435, 443 (1976).

treatment of data given by an individual to a private third party.[130] The Court held that citizens have no reasonable expectation of privacy for the phone numbers they dial, because they voluntarily expose this information to their phone companies.[131] As a result, the Fourth Amendment protections did not attach in a case in which the government asked a phone company to install a pen register on a suspect's phone line.[132] This third-party doctrine becomes increasingly problematic as technology, and usage of it, evolves. While *Smith* may have had relatively narrow implications when originally decided, its holding might have much broader applicability in an Internet era since individuals knowingly give data to third parties when they send emails and other everyday digitally-mediated communications.[133] In spite of rapid technological change, *Smith* and *Miller* remain good law for the most part, though in some instances state courts have refused to define the Fourth Amendment privacy protections so narrowly, thus declining to follow the two cases on state law grounds.[134]

Much could have turned on the original holding in *Warshak*, if it were to become law. In the vacated *Warshak* opinion, the court noted a distinction between voluntarily submitting certain information to a third party, and exposing information to an "intermediary that merely has the ability to access the information sought by the government."[135] This distinction, if recognized by other courts, would conceivably disassociate the third-party doctrine developed in *Smith* and

[130] Smith v. Maryland, 442 U.S. 735 (1979).

[131] *Id*. at 742.

[132] *Id*. at 745-46.

[133] Ric Simmons, *From Katz to Kyllo: A Blueprint for Adapting the Fourth Amendment to Twenty-First Century Technologies*, 53 HASTINGS L.J. 1303, 1338-39 (2002).

[134] *See, e.g.*, State v. Gunwall, 720 P.2d 808, 813 (Wash. 1986) (holding that under the Washington Constitution, a citizen has a privacy interest in their phone records and thus a search warrant is necessary for the installation of a pen register); State v. Thompson, 810 P.2d 415, 418 (Utah 1991) (holding that bank customers have a right of privacy in their bank records under the state constitution). *See also generally* Francis A. Gilligan & Edward J. Imwinkelried, *Cyberspace: The Newest Challenge for Traditional Legal Doctrine*, 24 RUTGERS COMPUTER & TECH. L.J. 305, 330 (1998).

[135] Warshak v. United States, 490 F.3d 455, 470 (6th Cir. 2007).

Miller from Fourth Amendment issues regarding Internet usage. But absent a clear opinion of this sort, the third-party data issue continues to loom over commonplace activities by citizens over digital networks.

VI. TAKING UP THE "MOMENTOUS ISSUES"

Though technology may help safeguard personally identifiable information to some extent in a digital age, the answer is not likely that privacy can be enhanced through technology alone. The market can help, too, as firms compete to earn user trust, but it provides an imperfect solution at best. In addition, straight legal reform is unlikely to offer a complete answer. The amount of information about each person is growing and the ability to access it is likewise increasing. United States citizens are putting information about their lives into myriad public systems willingly.

Two proposals might help protect citizens in a digital era without unduly hampering the ability of law enforcement official to perform their job. First, courts might recognize circumstances in which citizens demonstrate an intent to keep data private when considering the extent to which those communications are deemed to be protected under law. Second, through legislation, states might extend a right to demand deletion to citizens to enable them to prompt the destruction after the fact of certain personally identifiable information held in private hands.

Courts might begin by recognizing a sliding scale that enables one to distinguish between those transactions, whether involved with public or private parties, that use digital networked technologies that give rise to some legal protection from government surveillance and those that do not. Imagine that a citizen sends an unencrypted email from her account with a commercial ISP, such as Verizon, to a friend, who uses Comcast as her ISP. The citizen believes that this email was "private." But plainly, in terms of both government and corporate ability to access it, that message is not "private." The email is recorded by her ISP; it is conveyed much as a postcard would be, across an open network; and it is recorded by her

friend's ISP. This communication is one in which the public actor would have little trouble gaining access and for which the law might offer, as today, little or no protection.

Especially if absent the protections that a *Warshak*-style holding might afford, a citizen ought to be able to manifest her intent to keep certain data communications private in a way the law might be able to recognize in various ways.[136] For instance, if that same email is encrypted from end-to-end, which is not technically difficult after setting up a basic system of communication, the message in every form—even when unencrypted, in plain text, and having traveled over public networks—might be deemed to have certain protections under law as "private" data. If that email was collected at a client on either end of the transaction, that data would receive additional protection under law. Likewise, an e-commerce transaction in which the citizen took steps to use a pseudonym or, more advanced, a zero-knowledge proof to identify herself only to a merchant (but not to someone intercepting the data) might be deemed to be "private"—as might the credit card information and other details of the transaction.[137] In other words, if a citizen expresses her intent for a certain transaction to be

[136] It is of course the case that there are certain statutory protections for personal data under United States law, such as the 1974 Privacy Act (5 U.S.C. § 552a). In particular, the Privacy Act requires that federal agencies use personal data only for its intended purpose. *See* 5 U.S.C. §§ 552a(e)(1), (3)(B) (2008). In practice, however, the effect of this statute in terms of citizen protection has been modest. There are additional statutory limitations to sharing information between private parties and the government. For a specific example, see the 1986 Stored Communications Act ("SCA"), 18 U.S.C. §§ 2701-11, under the Electronic Communications and Privacy Act ("ECPA"), which provides limitations to the sharing of information between an electronic communications service and a law enforcement agency. Under the SCA, a provider may only voluntarily disclose the contents of a communication to a law enforcement agency where (1) the service provider inadvertently obtained the information and (2) the information appears to pertain to criminal activity. 18 U.S.C. § 2702(b)(7)(A)(i)-(ii) (2008). Alternatively, a governmental entity may *require* that a provider disclose certain electronic communication information, but in most instances a warrant is necessary. *See* 18 U.S.C. § 2703 (2008). The European Commission has similarly struggled with these issues, and has sought to place strong, comprehensive protections for citizens.

[137] *See* Anna Lysyanskaya, *Cryptography: How to Keep Your Secrets Safe*, SCI. AM., Sept. 2008, *available at* http://www.sciam.com/article.cfm?id=cryptography-how-to-keep-your-secrets-safe (last visited Nov. 28, 2008).

private, using either proprietary or widely accessible technical means, the law could meet her partway. The obvious shortcomings of this intent-based approach include the subjective nature of intent generally and the extent to which an encrypted email (one related to which the intent has been manifested) is less likely to be subject to surveillance by the very nature of its protection.

Second, the law might also be amended to help a citizen demand that information about him or her held by private firms be taken out of the "public" space after the fact. Since one of the fears about government surveillance is the ability to search in oceans of data, held in private hands after the fact, a citizen might be permitted by law to demand that her account not only be closed, but all data associated with her name be deleted by the technology service provider.[138] Such a "right to demand deletion" would have demerits, to be sure.[139] The costs

[138] *See* Paul Ohm, *The Fourth Amendment Right to Delete*, 119 HARV. L. REV. F. 10 (2005), *available at* http://www.harvardlawreview.org/forum/issues/119/dec05/ohm.pdf, responding to Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 HARV. L. REV. 531 (2005) (for discussion of such a right that might be embedded in Fourth Amendment protections). *See also* Paul Ohm, *The Olmsteadian Seizure Clause: The Fourth Amendment and the Seizure of Intangible Property*, 2008 STAN. TECH. L. REV. 2 (2008).

[139] Paul Schwartz alludes to an idea similar to the right to demand deletion in his discussion of privacy-control in the internet age. Paul M. Schwartz, *Internet Privacy and the State*, 32 CONN. L. REV. 815 (2000). Schwartz notes that in light of rising privacy concerns due to rapid technological change, there is agreement that privacy means individuals should have the right to control the use of the data they disperse over the internet. *Id.* at 820. This is what he terms the "privacy-control" paradigm. *Id.* However, Schwartz ultimately rejects the privacy-control paradigm, noting that given collective action problems, bounded rationality, and informational asymmetries, informational self-determination is not a realistic goal. *Id.* at 822. He additionally notes that given public accountability and bureaucratic rationality considerations, it is not clear that privacy concerns trump other needs requiring that this information not be secluded from the public realm. *Id.* at 828. Though Schwartz critiques privacy-control, the alternative he proposes still assigns value to the idea that individuals should have some control over the information they disperse on the internet, but that this right is not absolute. *Id.* at 834. Thus Schwartz champions constitutive privacy, claiming that "information privacy norms should create shifting, multidimensional data preserves that insulate personal data from different kinds of observation by different parties. Different kinds of 'outing,' that is, revelation of otherwise fully or partially hidden aspects of one's life, should be prevented before different audiences." *Id.* at 834-35.

Lior Jacob Strahilevitz discusses generally a "right to destroy," but he does not discuss it in the context of the Fourth Amendment or privacy rights. Lior Jacob

associated with implementing such a system would be high for the private sector; law enforcement officials would inevitably lose access to potentially very useful information about criminals who invoke the law to protect evidence of their wrongdoing; and some companies might be prompted to aggregate more personal information with someone's identity in order to be able to comply with deletion orders. Any legislative fix of this sort would need to be drafted carefully to mitigate these potential harms.[140]

Other legislative steps might make sense over time, especially to the extent that the third-party data problem grows as reliance on digitally mediated technologies, using private services, grows.[141] A further legislative rule might be enacted to stop certain government means of accessing personally identifiable data held in private hands, such as the purchase of data from data brokers. Likewise, rules might limit the uses to which such aggregated data are put by government actors, such as limits on sharing data across agencies for purposes other than that for which the data was initially collected. A tort-based regime might help to regulate the sharing or disclosure of information across the public-private border to the extent that rigorous procedures are not honored during the transfer.[142] And

Strahilevitz, *The Right to Destroy*, 114 YALE L.J. 781 (2005). Rather, Strahilevitz refers to the right to destroy as a property right—one of the "sticks in the bundle of [property] rights." *Id.* at 794. Ohm and Strahilevitz accept a right to destroy/delete as pre-existing, and the conclusion follows that if such a right is a defined property interest, then the Fourth Amendment is implicated when the government interferes with an individual's right to delete.

140 There is a closely analogous right to disclosure and correction included in the EU privacy rights and in U.S. safe harbor definition related to the EU Privacy Directive.

141 Fred Cate, for instance, proposes a selection of statutory protections to fill the gap in privacy protections affected by the *Smith* and *Miller* decisions in the internet age. Fred H. Cate, *Government Data Mining: The Need for a Legal Framework*, 43 HARV. C.R.-C.L. L. REV. 435 (2008). Cate proposes (among other things) the authorization of data mining programs that promote accountability, oversight, and compliance with legal requirements. *Id.* at 487. He suggests that such programs should be strictly limited, both in terms of who can use data mining systems, and in terms of the quantity and quality of information revealed. *Id.*

142 One critique of the current law/tort-regime is that there is usually no suppression remedy in this sort of legal action. Though the ECPA may regulate government access to electronic communication information collected by a third party, Fourth Amendment

still others have argued that the problems of informational privacy in a digital age are best solved through the courts.[143]

The market, too, might help citizens, though the limits of a purely market-based solution are quickly evident. Some firms do compete on the basis of establishing "best practices" for protecting consumer data; Google signaled their leadership on privacy in resisting the Department of Justice's demand for search data. In the international context, a series of technology companies is working on a code of conduct to govern when they will turn data over to state actors.[144] Best practices for businesses and easy-to-understand icons for citizens might provide a greater sense to citizens of the risks that they run by using certain services. These icons might make clear, for instance, how long data collected by a certain firm are retained and how the company responds to government requests for information. But citizens would need to be savvy enough to interpret the relevant market signals. And the core third-party data problem—the government's ability to access these data held in "public" places—would be little changed. If we leave it to the market alone, no meaningful legal barrier would be erected between the private sector players and the public agencies.

remedies are not available for victims of an ECPA violation. Though government entities are liable to suit under the ECPA, it has been held that the statute does not afford a suppression remedy. *See* Tucker v. Waddell, 83 F.3d 688 (4th Cir. 1996); *see also* United States v. Kennedy, 81 F. Supp. 2d 1103 (D. Kan. 2000).

[143] Ric Simmons, for instance, asserts that in *Katz*, the court delineated a results-based test rather than a method-based rule for decision in Fourth Amendment cases. Simmons, *supra* note 133. According to Simmons, the Katz court intended that Fourth Amendment protections attach to particular kinds of information collected, as opposed to particular search methods used. *Id.* In decisions like *Smith*, courts have strayed from the *Katz* results-based test and adopted more of a methods-based test. If the court returns to a results-based test, it is plausible to expect that Fourth Amendment protections attach to information dispersed over the Internet that an individual has intended to keep private.

[144] *See* http://cyber.law.harvard.edu/research/principles; *see also* Jonathan Zittrain & John Palfrey, *Reluctant Gatekeepers: Corporate Ethics on a Filtered Internet*, *in* ACCESS DENIED: THE PRACTICE AND POLICY OF GLOBAL INTERNET FILTERING 103-22 (Ronald Deibert, John Palfrey, Rafal Rohizinski, & Jonathan Zittrain eds., 2008).

VII. CONCLUSION

Ultimately, we as citizens need to decide how much access to the increasingly rich records of our lives we want to grant our law enforcement officials. At present, in practice, we as citizens would need to limit what we turn over to private parties if we wish to limit what the state can come to know about us. As the court said in *Garcia*, "Whether and what kind of restrictions should, in the name of the Constitution, be placed on such surveillance when used in routine criminal enforcement are momentous issues. . . ."[145] These momentous issues play out at the United States border with cyberspace, and at international borders, many times every day.

Given what we know about how people are leading their lives mediated by digital technologies, and given what we know about how much surveillance can happen in converged public and private spaces online, it is time that we take up these issues directly and re-evaluate the inevitable trade-offs we are making every day. Central to the challenge that we face is an information asymmetry: while we have an all-too-graphic idea of the high costs of not ensuring our security today, we know too little about what others will be able to come to know about us in the future. What we do know is that the third-party data problem grows with each passing year. We know that private entities collect much data about us. We know that the law affords citizens little protection from abuse by either public or private parties. And we know that the rules that we set today to govern the collection, sharing, and access of data will have long-term consequences, both for security and for individual privacy. Based upon these facts, it is time to rethink whether the scope of the Fourth Amendment is sufficient to protect individual privacy from intrusion by the state, especially with respect to data initially collected by private parties. More fundamentally, it is also time to consider whether the public-private distinction, as it has developed over the past century and a half, makes sense in a digital age.

[145] United States v. Garcia, 474 F.3d 994, 998 (7th Cir. 2007).