January 2008

Risking Communications Security: Potential Hazards of the Protect America Act

Steven M. Bellovin
Columbia University

Matthew A. Blaze
University of Pennsylvania, blaze@cis.upenn.edu

Whitefield Diffee
Sun Microsystems

Susan Landau
Sun Microsystems

Peter G. Neumann
SRI International

See next page for additional authors

Risking Communications Security: Potential Hazards of the Protect America Act

Abstract
A new US law allows warrantless wiretapping whenever one end of the communication is believed to be outside national borders. This creates serious security risks: danger of exploitation of the system by unauthorized users, danger of criminal misuse by trusted insiders, and danger of misuse by government agents.

Keywords
protect America act, US wiretap law, civil liberties, surveillance, wiretapping

Author(s)
Steven M. Bellovin, Matthew A. Blaze, Whitefield Diffee, Susan Landau, Peter G. Neumann, and Jennifer Rexford

Risking Communications Security: Potential Hazards of the Protect America Act

A new US law allows warrantless wiretapping whenever one end of the communication is believed to be outside national borders. This creates serious security risks: danger of exploitation of the system by unauthorized users, danger of criminal misuse by trusted insiders, and danger of misuse by government agents.

Steven M. Bellovin
Columbia University

Matt Blaze
University of Pennsylvania

Whitfield Diffie and Susan Landau
Sun Microsystems

Peter G. Neumann
SRI International

Jennifer Rexford
Princeton University

In August 2007, United States' wiretapping law changed: the new Protect America Act permits warrantless foreign-intelligence wiretapping from within the US of any communications believed to include a party located outside it. US systems for foreign intelligence surveillance located outside the United States minimize access to the traffic of US persons by virtue of their location. The new act could lead to surveillance on an unprecedented scale that will unavoidably intercept some purely domestic communications. A civil liberties concern is whether the act puts Americans at risk of spurious—and invasive—surveillance by their own government, whereas the security concern is whether the new law puts Americans at risk of illegitimate surveillance by others.

Building surveillance technologies into communication networks is risky. The Greeks learned this lesson the hard way; two years ago, they discovered that legally installed wiretapping software in a cell-phone network had been surreptitiously enabled by parties unknown, resulting in the wiretapping of more than 100 senior members of the government for almost a year.[1] Things are not much better in Italy, where a number of Telecom Italia employees have been arrested for illegal wiretapping (with attempts at blackmail).[2]

In this article, we focus on security, not civil liberties. If the intercept system is to work, it is important that the surveillance architecture not decrease the security of the US communications networks. Although we are writing about a US law and its consequences for the security of US communications, the examples of Greece and Italy make clear that the same issues occur internationally.

Background

The combination of data sources may make this surveillance more powerful—and create more risk—than was intended. We start with background on legal and policy issues, then technical concerns; this extensive background is necessary because architecture matters a lot, and in subtle ways.

Legal and policy issues

US wiretap law has a long and complex history. (See the sidebar, "US wiretap law," for a summary; other work has more details.[3]) Briefly, there are different standards and procedures for criminal versus national security wiretaps; in the latter case, so-called Foreign Intelligence Surveillance Act (FISA) warrants can be issued for specific circumstances:

- any person in the US communicating via wire (the word "wire" includes fiber optics);
- a US person (including US citizens, permanent residents, and US corporations)[4] in the US whether communicating via wire or radio; and
- any person in the US communicating via radio with people, all of whom are also in the US[5] (the rules are, in fact, even more complicated, but this is sufficient for our purposes).

Warrants are generally not needed to intercept radio communications.

The Protect America Act (Public Law No. 110-55) dropped the warrant requirement for communications (over any medium) of US persons located in the United States with persons "reasonably believed to be located outside the United States."[6] Modern communications technology—mobile phones, WiFi, and the Internet—often make it difficult to discern whether communication is from a location inside or outside the US, so the question is on what basis communications will be collected. In other words, there is an important distinction between the requirements of the law and what can be done with available technology.

Much of the motivation for changes to FISA arises from the geography of the world's communications infrastructure combined with recent changes to telecommunications technology. The US is a major hub in our communication-centered world, giving the National Security Agency (NSA), which is the US signals intelligence agency, significant opportunities for access to transit traffic.

There are numerous reasons for US centrality in the world's communications systems. One is cost: the US is the world's leading economy, and fiber optic cables—how modern wired communications travel—have been built installed between the US and overseas. With their economies of scale, these cables enable US providers to underbid regional carriers—for example, much of South America transits its traffic through Miami. Another reason is politics, which can lead to strange communication paths. For many years, communications could not travel directly between Taiwan and the People's Republic of China: calls traveled by way of Sacramento via AT&T lines. A third reason is the Internet. Many servers that are the very reason for communication—for example, Yahoo Mail, Hotmail, and Gmail—are in the US (although this is an ever-decreasing percentage of the world's mail servers, especially as China comes online).

At the time that FISA was written, communications satellites (radio) had revolutionized international communications. In subsequent decades, there was a major shift to fiber optic cables with a decreasing percentage of intercontinental communications traveling by radio. Thus the exemption allowing warrantless radio interception became increasingly less applicable. In recent years, the NSA has pressed to have the exemption updated. While many in the field agree that there is plausibly a problem resulting from the introduction of fiber optic cables, the Protect America Act went considerably further.

Collection

Signals intelligence is organized into a seven-step process: access, collection, processing, exploitation, analysis (intelligence analysis), reporting, and dissemination. The first three are of particular concern. Access is what happens at a radio, a fiber splitter, a tap on a wire, or a tap in a telephone switch. Collection is the process of recording signals for consideration. Recorded signals can be kept briefly or for very long periods.

Processing is shorthand for selecting the information you want (and filtering out what you don't). As in any learning process, if you can find information at all, you often have too much of it and must extract what interests you from what doesn't. This is where the choice of architecture is significant, both in terms of minimizing data collection and in determining how the combination of data sources is used. We return to this point later.

Increasingly, communications are IP-based. The Internet is the interconnection of many networks, and these connections occur in various ways. For the largest networks, these form at peering connections: interconnections between administratively separate domains (such as ISPs).

International communications enter the US by satellite, terrestrial microwave, older copper cable, and newer fiber optic cable. There are roughly 25 cable heads in the US. (This is an estimate based on Telegeography's "Global Communications Cable and Satellite Map 2002," which shows four cable heads on the Atlantic Coast and five on the Pacific. There are at least an additional five each coming terrestrially from Canada and Mexico.) At the cable head, incoming signals split in several ways. First, the signals are sent via multiplexors and demultiplexors to the proper carriers (since most transoceanic fibers are owned by consortia of communications companies). Each carrier's channels are further subdivided: voice signals are sent (perhaps via other gear) to phone switches, Internet signals to routers, and so on.

A likely architecture

The NSA has not disclosed its surveillance architecture, so it is impossible to know exactly how its system works. However, a current court case gives hints. In late 2005 and spring 2006, *The New York Times* and *USA Today* revealed that the NSA had been wiretapping without warrants post-9/11. Shortly afterward, civil liberties groups and individuals sued AT&T over the "illegal spying of telephone and Internet communications." Affidavits filed in *Tash Hepting et al. v. AT&T Corporation et al.*,[7] describe the NSA surveillance architecture at the AT&T switching office in San Francisco. AT&T has acknowledged the authenticity of the documents describing the layout and configuration for the secure room in its San Francisco office, which is where electronic surveillance took place.[8] Our discussion is based on these documents and on affidavits submitted by two expert witnesses, Mark Klein (a technician in the AT&T San Francisco office)[9] and J. Scott Marcus (a designer of large-scale

US wiretap law

Through the decades, electronic surveillance has been effective in both capturing and convicting criminals and spies, as well as in denying them and terrorists full use of modern communications out of the fear of being eavesdropped, tracked, and caught. Although rarely mentioned, the latter is a particularly important side effect of electronic surveillance. But while there is no question that electronic surveillance can be extremely effective, there has always been tension between national security and privacy.

Prior to the Protect America Act, United States wiretapping law was essentially governed by Title III of the Omnibus Crime Control and Safe Streets Act of 1968 (P.L. 90-351), which regulated the procedure for wiretaps in criminal investigations, and the Foreign Intelligence Surveillance Act (FISA; P.L. 95-511), which did the same for foreign intelligence surveillance. These laws and their later derivatives laid out clear and specific procedures for obtaining wiretap warrants, which, with very minor exceptions, specified the particular line (or particular IP address or email account) on which the tapping was to occur.[1] Law enforcement obtained a warrant and sent this information to the communications provider, which installed the tap.

The US learned the hard way that oversight was critical if surveillance technologies were to be kept within legal bounds. During the Watergate era, the Senate Select Committee to Study Governmental Operations with respect to Intelligence Activities investigated 35 years of government electronic surveillance in the US, uncovering many abuses. These included wiretaps on congressional staff, Supreme Court Justices, Martin Luther King Jr. (in his case, sometimes for purely political purposes), as well as government investigations of such decidedly non-violent groups as the American Friends Service Committee, the National Association for the Advancement of Colored People, and the Women's Strike for Peace. It was clear that the "national security" rationale for many of the wiretaps was not justified. FISA was passed in response to these issues; the requirements governing FISA wiretapping were lifted almost verbatim from the carefully crafted recommendations of the Senate committee report.[2] Some of these safeguards delimiting government surveillance were removed by the USA Patriot Act (arguably the most important change the Patriot Act made in wiretapping law was modifying the requirement that foreign intelligence be the "primary" purpose of a FISA tap to a "significant" purpose[3]).

The law was also clear in its exception: no warrant was required to intercept radio communications between persons in the US and persons abroad unless the government was intentionally targeting a particular known US person who was in the US. This exception was viewed as a temporary one; the Senate Judiciary Committee Report on the FISA legislation makes clear

IP-based data networks, former CTO at GTE Internetworking and at Genuity, and former senior advisor for Internet Technology at the US Federal Communications Commission).[10]

Optical fiber carrying the inter-ISP peering traffic associated with AT&T's Common Backbone[11] was "split," dividing the signal so that 50 percent went to each output fiber (the weakened signal on each output fiber still contained sufficient information to allow reading the communications).[12] One of the output fibers was diverted to the secure room; the other carried communications on to AT&T's switching equipment. The secure room contained Narus traffic analyzers and logic servers; Narus states that such devices are capable of real-time data collection (recording data for consideration) and capture at 10G bits per second (bps). Certain traffic was selected and sent over a dedicated line to a "central location." The San Francisco office set up was one of many throughout the country, including in Seattle, San Jose, Los Angeles, and San Diego.[13] According to Marcus's affidavit, the diverted traffic "represented all, or substantially all, of AT&T's peering traffic in the San Francisco Bay area,"[14] and thus, "the designers of the ... configuration made no attempt, in terms of location or position of the fiber split, to exclude data sources comprised primarily of domestic data."[15]

Call detail records

Modern telecommunications allow the construction of smooth-running organizations that span the globe; telecommunications are the nervous systems of these organizations. The "reasonably believed to be located outside the United States" aspect of the Protect America Act arguably changes the rules on the government's use of call detail records (CDRs), which can be surprisingly revelatory of relationships and organizational structure (although this data does not always reveal where communicating parties are physically located). It appears that the US government has real-time access to CDRs without need for a court order.

CDRs are essentially the raw data for traditional phone bills. Phone companies build and maintain comprehensive databases of such information, which contain complete call traffic data: records of such transactional information as calling and called numbers for phone calls, IP addresses and user URIs in the case of voice-over IP (VoIP); SMTP headers for emails; location, time, and date of communications; call duration; and related information. To listen to an organization's communications is to read its mind, and following just the pattern of its communications is a large step in this direction. CDRs provide a window into the past. Phone companies use such data for billing, engineering, marketing, and fraud detec-

that interception of radio communications was to be considered separately.[4] But separate legislation never came to pass, and the exception continued.

In 2002, Attorney General John Ashcroft proposed changing FISA procedures. The FISA Court, whose job it is to review FISA wiretapping warrant applications, was not pleased with this, in part because of mistakes that had occurred in earlier FISA applications. The court issued a report criticizing the proposal[5] and the US Federal Bureau of Investigation's (FBI) mishandling of the wall between foreign intelligence cases and criminal investigations: "In virtually every instance, the government's misstatements and omissions in FISA applications and violations of the Court's orders involved information sharing and unauthorized disseminations to criminal investigators and prosecutors." An extremely important check on government abuse is oversight. As the founders of the US knew, another branch of government can provide the objectivity necessary for such an investigation. Public knowledge also matters. When the FISA court was dissatisfied with the Bush Administration's actions in 2002, it declassified its opinion,[5] helping to shape the later debate on the USA Patriot Act renewal and other administration requests for changes in the wiretap laws.

Some might argue that the excesses of surveillance in the 1960s and '70s were long ago, occurring during a period of domestic unrest and international tension. But government excesses in this realm continue. A recent report by the FBI Inspector General, for example, sharply criticized the bureau regarding its abuse of National Security Letters, "administrative" subpoenas that are issued with *no* judicial oversight and that require the recipient to turn over certain records. The Inspector General concluded that FBI agents might have violated the law 3000 times since 2003 in their collection of telephone and financial records of US citizens and foreign nationals.[6]

References

1. D. Solove and M. Rotenberg, *Information Privacy Law*, Aspen Publishers, 2003, pp. 323–341.
2. United States Senate, Select Committee to Study Governmental Operations with Respect to Intelligence Activities (1976), *Intelligence Activities and the Rights of Americans, Final Report: Book II*, Report 94-755, Ninety-Fourth Congress, Second Session, 23 Apr. 1976, pp. 292–330.
3. W. Diffie and S. Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption*, updated and expanded edition, MIT Press, 2007, pp. 280–285.
4. United States Senate, Committee of the Judiciary (1977), *Legislative History P.L. 95-511 Foreign Intelligence Surveillance Act, Report 95-604*, Ninety-Fifth Congress, First Session, 15 Nov. 1977, p. 34.
5. United States Foreign Intelligence Surveillance Court, Memorandum Opinion (as Corrected and Amended,) May 17, 2002, in United States Senate, Committee on the Judiciary, 2002, *The USA PATRIOT Act in Practice: Shedding Light on the FISA Process*, Hearing on 10 Sept. 2002, S. Hrg. 107–947, One Hundred Seventh Congress, Second Session.
6. Inspector General, US Federal Bureau of Investigation, A Review of the Federal Bureau of Investigation's Use of the National Security Letters, Mar. 2007.

tion. Unlike a wiretap or pen register, which provide, respectively, real-time access only to the content or number currently being dialed, a CDR database contains a wealth of data on previous communications. Thus, an interested government agency doesn't need to have the proper legal authorization or technology in place before a call is made but may search the call detail database during the communication to determine a "community of interest"—the network of people with whom the suspect is in contact—as well as later, once a new target has been identified. For international and some purely domestic calls, two CDRs exist for each communication, one from the origination point—which could be an interface to another company—and one from the termination.

Although transactional information has historically been viewed as much less deserving of privacy protection than call content, in fact, access to CDRs can be a major privacy risk. Corinna Cortes and her colleagues at AT&T Shannon Labs showed, for example, that, even though the calling number had changed, it was possible to identify an individual caller from a 300-Tbyte CDR database by simply looking at patterns of called numbers.[16] While at Katholieke Universiteit Leuven, George Danezis related a story in which Intel researchers studying ambient Bluetooth activity to improve ad-hoc routing protocols issued staff members Bluetooth devices. One of the discoveries was that a pair of researchers were meeting nightly, a relationship that had not been previously known to the other lab members.[17]

CDRs can be used for targeting more detailed surveillance, such as wiretapping. The more tightly coupled CDR and content collection are, the more likely it is that, without regard to the intentions of the parties involved, content wiretapping will occur as a result of CDR information.

Difficulties in monitoring international Internet traffic

Monitoring international traffic requires an effective way to identify whether communication starts or ends outside the US. This is a surprisingly difficult problem to solve on today's Internet. Perhaps even more surprisingly, this is not an easy task on a telephone network either. According to a 1998 National Academies study, "the underlying telephone network is unable to provide [caller ID] information with high assurance of authenticity."[18] (Or, to put it another way, although CDR is an amazingly effective guide to communications activity, the data can't always provide real-time answers about a call's location.) NSA has worked on the problem, and the agency even has a patent for using time latency to determine a communication's location

(US Patent # 6,947,978: Method for geolocating logical network addresses).

International traffic monitoring requires either limiting monitoring to links that carry only international traffic or filtering out any traffic transferred between two domestic hosts. The first approach seems easy if monitoring is limited to cable heads connecting the US to other countries. The second approach also seems easy, by mapping the IP addresses of the sending and receiving hosts to their geographic locations. However, both approaches have limitations.

Even if the traffic does not traverse a relay or anonymizer, real-time association of an IP address with a particular person of interest is difficult.

Although most traffic on international links travels to or from a foreign host, a small amount of *domestic* traffic traverses these links as well—for example, some domestic traffic travels through Canada and then back to the US due to the vagaries of Internet routing. (This is partially a result of a 1940s AT&T master plan that made the US, Canada, and most of the Caribbean one integrated country, with no cable heads, or even international gateways, between them.) As such, monitoring links at the US border, with the goal of warrantless tapping of international traffic, could lead to unintentional tapping of domestic traffic. Because these links operate at a very high speed, it is difficult to analyze measurement data as they are collected. Furthermore, Internet traffic does not necessarily follow symmetric paths—the traffic from host A to host B does *not* necessarily traverse the same links as the traffic from B to A—so monitoring both ends of a conversation sometimes requires combining data collected from multiple locations, making this type of monitoring difficult in practice.

Monitoring very close to the sending or receiving host ensures that both directions of the traffic are visible and that the link speeds are typically small enough for detailed data collection. But monitoring near the domestic endpoint would almost certainly capture a large amount of traffic exchanged with other US-based hosts. To identify and filter the domestic traffic, a signals intelligence agency such as the NSA could map the remote host's IP address to a country using registries that identify the institution that owns the IP address block. The problem is that these registries are notoriously incomplete and out of date. Instead, the NSA could use existing IP geolocation services (such as Quova, www.quova.com). Although such services are often accurate to a few tens of miles, errors of hundreds of miles or more are not uncommon. As such, a host might easily look as though it resides on the opposite side of the border with another country, such as Mexico or Canada.

Even if geolocation services are accurate, the source and destination addresses in the IP packet do not necessarily correspond to communicating hosts. Some VoIP services, such as Skype, routinely use relay nodes to enable calls between two hosts that could not otherwise communicate, due to a firewall or a Network Address Translator (NAT), a device that enables multiple hosts on a private network to access the Internet using a single public IP address. A relay node is a third machine that might reside in the same country as one, or both, of the other hosts, or in yet a third country. Depending on where traffic is monitored, the source or destination address can correspond to the relay node, rather than one of the communicating endpoints, complicating efforts to determine whether both endpoints are domestic. In addition, some users apply anonymization tools like Tor (The Onion Router) that intentionally hide source and destination addresses from packet sniffers. Whether traffic traverses a relay or an anonymizer, the monitor could capture erroneous IP addresses that do not correspond to the ultimate source and destination of the traffic.

Even if the traffic does not traverse a relay or anonymizer, real-time association of an IP address with a particular person of interest is a difficult task. (We should note that EU member states consider email addresses, even when not associated with an IP address, personally identifiable information.) For example, an IP address might correspond to a NAT box so identifying the particular host responsible for the traffic requires access to transient information available only to the NAT box. Even in the absence of NAT boxes, the IP address of each end host can be assigned dynamically through the Dynamic Host Configuration Protocol (DHCP). Mapping the IP address to a particular host may require DHCP logs from the local site, and these logs are often incorrect.[19] Mapping from the host to a particular user is difficult if the machine is shared among many people, as at a cybercafe or an academic lab. In addition, mobile hosts such as laptops or PDAs frequently acquire new IP addresses.[20]

Even if the communicating endpoints can be appropriately identified, determining what application they are running is difficult. (Knowing the application helps determine whether the information will be useful; there is little foreign intelligence value in wiretapping a transmission of the latest Hollywood movie.) In the simplest case, applications are easily discernible from numerical identifiers (for example, port numbers) in data packets. However, some applications do not use well-known port numbers, and others intentionally use port numbers normally reserved

for other applications in order to evade detection; for example, some peer-to-peer file sharing applications use port 80, the conventional port for Web traffic. (There is active research in determining the type of traffic using other information.) Such analysis is difficult to perform in real time on high-speed links, such as the links connecting the US to other countries. In addition, a malicious party trying to avoid detection might intentionally pad or jitter packets to evade detection, adding further complexity to an already difficult problem. Finally, some applications such as Skype encrypt data, making it difficult to extract meaningful information about the content of the communication between end hosts.

The real problem is that these difficulties are intrinsic to the Internet's basic design. Additional issues arise when interworking VoIP with other telephony services, such as the public-switched telephone network. An international call might terminate in the US and then use VoIP the rest of the way (and vice versa), requiring joint analysis across two kinds of communication networks. The many difficulties in accurately distinguishing domestic and foreign communication make it unlikely that an intelligence agency could avoid tapping domestic calls.

Risks

Surveillance technology is an "architected security breach"[21] into a communications network and thus a risky business on which to embark. (Telecommunications carriers must "listen in" on communications for quality control. Intercept architectures are, however, more complicated; they have to target particular individuals without leaving a trace. Monitoring the network for quality control is much simpler because monitoring can pick up any conversation and is allowed to fail more often.) Two situations illuminate different reasons for our concern.

Let's go back to a point we brought up earlier. The Greek wiretapping case began in summer 2004, just before the Olympic Games in Athens. More than 100 cell phones belonging to the prime minister and ministers of defense, foreign affairs, justice, and public order—as well as opposition members in the Greek parliament—were wiretapped through the activation of wiretapping modules in the network's telephone exchange switches, capabilities that were supposed to be invoked *only* with legal authorization. The wiretapping capability had been provided in a system update, but because the Greek Vodaphone network had not purchased the wiretapping capability, the system lacked the management software that installed and logged wiretaps. Not only did the intruders turn on the network's wiretapping capability, they also installed a rootkit that hid any activity of their own software updates. Each time there was a communication on one of the tapped phones, a duplicate communication was sent to one of 14 cell phones in the network, all of which were prepaid, anonymous accounts. While we know private communications at the highest levels of the Greek government were wiretapped for 10 months, who did it remains unknown.[1]

The US has also experienced difficulties with communications surveillance systems. Under the Communications Assistance for Law Enforcement Act (P.L. 100-667), the US Federal Bureau of Investigation (FBI) was responsible for determining technical specifications for wiretapping built into switches of digital telephone networks. DCS 3000, an FBI suite of systems for collecting and managing data from wiretaps for criminal investigations, was designed to meet those requirements. Recently released FBI documents reveal serious problems in the system's implementation.[22] Its auditing system was primitive, surprising for a system intended for evidence collection. The system has no unprivileged user IDs, relying on passwords rather than token-based or biometric authentication, and even uses an outdated hashing algorithm (MD5 appears in a 2007 "system security plan,"[23] several years after Chinese researchers found serious problems with this already weak hashing algorithm). Most seriously, the system relied on a single shared login, rather than a login per authorized user. The system's ability to audit user behavior depended entirely on following proper processes, including using a manual log sheet to show who was using the system at a given time. Remote access—in an insecure fashion—is permitted from other DCS 3000 nodes, making the system vulnerable to insider attacks. These are a real risk: recall that the most damaging spy in FBI history, Robert Hanssen, abused his authorized access to internal FBI computer systems to steal information and track progress of the investigation aimed at him.

The problems in the DCS 3000 implementation illustrate the risks in building a communications surveillance system. We do not know whether DCS 3000 was merely poorly implemented or whether it was poorly specified. What were the requirements on the FBI system? Did they include full auditing and full user identity? What were the project's goals? Were the designers required to meet *all* requirements or goals? These are questions that should have been asked of the DCS 3000 designers—and should be asked of any builder of a communications surveillance system.

Although the NSA has extensive experience in building surveillance systems, that does not mean things cannot go wrong. When you build a system to spy on yourself, you entail an awesome risk. In designing a system to satisfy the needs of the Protect America Act, the risk is made worse by four phenomena:

- removal of a protective role provided by communi-

cation carriers in all previous interception programs within the US communication system. This protective role was the result of the specificity required in wiretap warrants.
- placing the system properly within the US rather than at US borders;
- likelihood that the system will be built out of pieces previously used abroad, which runs the risk that opponents are already familiar with the equipment via intelligence-sharing agreements or capture of equipment; and
- use of CDRs, originally built for network development purposes, in an entirely new way involving "customers" outside the phone company.

These architectural decisions facilitate three distinct types of problems:

- system capture to enable spying on US traffic;
- system defeat by using information learned from foreign examples to defeat selection and filtering strategies; and
- system spoofing by similar means.

All of these can be used not only to make the surveillance system less effective, but also to turn it into a tool for capturing communications that are not implicated in any illegal activity—endangering both security and privacy. We see several specific risks as a result.

Risk of exploitation by opponents. A system that accesses domestic communications necessarily poses a greater direct risk to the communications of Americans than a surveillance system fielded overseas. To avoid foreign familiarity with its operation, communication security equipment is not often shared with allies. However, engineering economy reuses systems previously fielded abroad; thus, both allies and opponents are likely to be familiar with US surveillance equipment. Is there a risk that knowledge of the surveillance system acquired by studying equipment outside the US will be applied to defeating or subverting similar equipment deployed within the US? Is the NSA designing sufficiently robust mechanisms to assure complete control of the filtering and selection mechanisms?

Even prior to the Protect America Act, US communications were vulnerable to surveillance, but building signals intelligence systems is expensive. The system designed as a result of the Protect America Act must not reduce foreign powers' difficulty in gaining access to US communications. Can the communications of US persons be tapped without increasing the risk that these communications will be exploited by others without authorization to do so?

Removal of safeguards by communications carriers. What risks are introduced by leaving a single entity in charge of selection and retention decisions? "Two-person control" would be prudent—for example, control by two authorized parties who understand how a system should work. In its absence, any process such as the one apparently embodied in the AT&T San Francisco switching office (in which communications are diverted to an NSA safe room and then collected according to rules determined by the intelligence agency) provides little recourse in cases where mistakes are made.

Lack of inherent technical minimization of traffic. Intercepting at switches or routers creates unnecessary risks because the switches handle domestic as well as foreign communications. This risk, although distinct from the risk of exploitation by opponents, feeds into it; potential overcollection of purely domestic traffic increases the value of targeting the US access and collection system.

Domestic traffic penetrating too deeply into the NSA collection system. Collection outside the US inherently filters out most "US-person traffic" before it gets to NSA headquarters at Fort Meade. Does the design of the expanded surveillance system eliminate domestic traffic early and as effectively? This is more of a privacy risk than a security one, although insider attacks make it a security risk as well.

CDR information. CDR systems were originally intended to be used by telephone company employees for determining customer usage patterns and thus anticipating future needs. It is a truism in the security field that problems frequently occur when new uses are found for old systems, given that the protection mechanisms and system architecture were never designed for the extended uses. Will new vulnerabilities be created when copies of the CDR data are sent to law enforcement or intelligence agencies? It is impossible to give a definitive answer, but the past history of such changes does not leave us sanguine.

Risk reduction

There are also ways in which the Protect America Act enables an architecture that could reduce risk. Being able to place equipment on US soil reduces the need to place equipment abroad. Beyond the direct security risks to equipment, which could be alleviated by high-quality shielding and tamper resistance, there is an intrinsic risk. When intercept capability is installed in other countries' communication systems, the privilege must be paid for—often by sharing information. Host countries might demand not only a share of the intelligence take—whether this could ever pose a

threat to US communications is hard to assess—but also inspection authority over the installation and information about techniques. Intercept facilities hosted by foreign governments are expected not to spy on the host countries themselves. However, the charge that the surveillance facilities are doing so is often made, and the host countries quite reasonably insist on taking measures aimed at preventing this.

New security risks

Security risks are exacerbated by the direction of the Internet's development. The Internet is currently a network with only millions of devices connected to it, but the world is rapidly moving to a situation in which billions of small, resource-limited devices such as RFID tags and sensors will use networks for communication and control. While many of these devices will be on local area networks, many will use the Internet.[24] Any future surveillance architectures must take such growth and directions into account.

Implicit in the FISA update was the need to protect the US against non-state actors, who have indeed shown themselves to be adept at using the Internet to communicate. Some of the tools provided for by the Protect America Act could in fact aid in the disruption of various nefarious plots. But building surveillance technology into a communications infrastructure creates risk of penetration by trusted insiders, foreign powers, and non-state actors (with trusted insiders being the greatest threat). Disrupting attacks by non-state actors could be a short-term gain, but surveillance architectures rarely go away. The dangers created by the Protect America Act present a long-term risk. (This is exemplified by the exploit in the Greek wiretapping case, which relied on an earlier software version that included wiretapping capabilities but not the auditing system.)

The Protect America Act, a law enacted in haste, holds the possibility of a vast increase in the number of Americans whose communications and communication patterns will be studied. The surveillance provides access to US communications, a target of great value. The US could build for its opponents something that would be too expensive for them to build for themselves: a system that lets them see the US's intelligence interests, a system that could tell them how to thwart those interests, and a system that might be turned to intercept the communications of American citizens and institutions. It is critical that the new surveillance system neither enable exploitation of US communications by unauthorized parties nor permit abuse by authorized ones.

Recommendations

The change from a system that taps particular lines on receipt of a wiretap order specifying those lines to one that sorts through transactional data in real time and selects communications of interest is massive. Where interception occurs and how the data sources (CDRs, traffic, other information) are combined and used will not only affect how powerful a tool warrantless wiretapping is, but will also affect how likely the system is to pick up purely domestic communications. In building a communications surveillance system itself—and saving its enemies the effort—the US government is creating three distinct serious security risks: danger of exploitation of the system by unauthorized users, danger of criminal misuse by trusted insiders, and danger of misuse by US government agents. How should the US mitigate these risks?

Minimization matters

Allowing collection of calls on US territory necessarily entails greater access to the communications of US persons. An architecture that minimizes collection of communications lowers the risk of exploitation by outsiders and exposure to insider attacks. Traffic should be collected at international cable heads rather than at tandem switches or backbone routers, which also carry purely domestic traffic. Although interception at the cable heads will help minimize collection, it is not sufficient in and of itself. Intercepted traffic should be studied (by geolocation and any other available techniques) to determine whether it comes from non-targeted US persons and, if so, discard it before any further processing is done. It should be fundamental to the system's design that the combination of interception location and selection methods minimizes the collection of purely domestic traffic.

Architecture matters

Using real-time transactional information to intercept high-volume traffic makes architectural choices critical. Robust auditing and logging systems must be part of the system design. Communication providers, who have technical expertise and decades of experience protecting the security and privacy of their customers' communications, should have an active role in both design and operation. Thus, "two-person control" is appropriate for this situation.

It is critical that the new surveillance system neither enable exploitation of US communications by unauthorized parties nor permit abuse by authorized ones.

Oversight matters

The new system is likely to operate differently from

previous wiretapping regimes and will likely use new technologies for purposes of targeting wiretaps. There should be appropriate oversight by publicly accountable bodies. While the details might remain classified, there should be a publicly known system for handling situations when mistakes are made. To assure independence, the overseeing authority should be as far removed from the intercepting authority as practical. To guarantee that electronic surveillance is effective and free of abuse *and* that minimization is in place and working appropriately, it is necessary that there be frequent, detailed reports on the system's functioning. Of particular concern is the real-time use of CDR for targeting content, which must neither be abused by the US government nor allowed to fall into unauthorized hands. For full oversight, such review should be done by a branch of government different from the one conducting the surveillance. We recommend frequent ex post facto review of CDR-based real-time targeting. The oversight mechanism must include outside reviewers who regularly ask, "What has gone wrong lately—regardless of whether you recovered—that you have not yet told us about?"

US communications security has always been fundamental to national security. The surveillance architecture implied by the Protect America Act will, by its very nature, capture some purely domestic communications, risking the very national security that the act is supposed to protect. In an age so dependent on communication, the loss could well be greater than the gain. To prevent greater threats to US national security, it is imperative that proper security—including minimization, robust control, and oversight—be built into the system from the start. If security cannot be assured, then any surveillance performed using that system will be inherently fraught with risks that are fundamentally unacceptable. □

References

1. V. Prevelakis and D. Spinellis, "The Athens Affair," *IEEE Spectrum,* July 2007, pp. 18–25.
2. P. Kiefer, "Phone Taps in Italy Spur Rush toward Encryption," *New York Times*, 29 Apr. 2007; www.nytimes.com/2007/04/29/technology/29cnd-encrypt.html?ex=1335499200&en=aa06d98a600afc6f&ei=5088&partner=rssnyt&emc=rss.
3. W. Diffie and S. Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption,* updated and expanded edition, MIT Press, 2007.
4. *US Code*, Title 50, section 1801(i), 1978.
5. *US Code*, Title 50, section 1801(f), 1978.
6. Protect America Act, section 105(a) 2007.
7. United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 Jan. 2006.
8. Exhibit A in *Tash Hepting et al. v. AT&T Corporation et al.*, United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006.
9. M. Klein, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006.
10. J.S. Marcus, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006.
11. J.S. Marcus, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006, p. 15.
12. J.S. Marcus, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006, pp. 12–14.
13. M. Klein, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006, p. 7.
14. J.S. Marcus, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006, p. 24.
15. J.S. Marcus, affidavit in *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3: 06-cv-0672-vrw, 8 June 2006, pp. 24–25.
16. C. Cortes, D. Pregibon, and C. Volinsky, "Computational Methods for Dynamic Graphs," AT&T Shannon Labs, 9 Jan. 2004.
17. G. Danezis, "Introducing Traffic Analysis: Attacks, Defences and Public Policy Issues"; http://research.microsoft.com/users/gdane/papers/TAIntro.pdf.
18. F. Schneider (ed.), *Trust in Cyberspace,* Computer Science and Telecommunications Board, National Research Council, 1999, p. 36.
19. R. Clayton, *Anonymity and Traceability in Cyberspace,* Univ. of Cambridge Computer Lab, tech. report 653, Nov. 2005.
20. S. Bellovin et al., "Security Implications of Applying the Communications Assistance to Law Enforcement Act to Voice-over IP," 2006; www.itaa.org/news/docs/CALEAVOIPreport.pdf.
21. S. Landau, "National Security on the Line," *J. Telecommunications and High Technology Law,* vol. 4, no. 2, 2006, p. 418.
22. *EFF v. Department of Justice*, Civil Action No. 06-1708-CKK (D.D.C.) (filed 3 Oct. 2006); www.eff.org/issues/foia/061708CKK.
23. Information Assurance Section, US Federal Bureau of Investigation, "Controlled Interface 100 (CI-100) System Security Plan (SSP) DCS-3000 to EDMS," 16 April 2007.

24. S. Landau, "National Security on the Line," *J. Telecommunications and High Technology Law*, vol. 4, no. 2, 2006, pp. 433–434.

Steven M. Bellovin *is a professor of computer science at Columbia University. His technical interests include network security, privacy, and the social implications of computers. Bellovin has a PhD in computer science from the University of North Carolina at Chapel Hill. Contact him at smb@cs.columbia.edu.*

Matt Blaze *is an associate professor of computer and information sciences and director of the Trusted Network Eavesdropping and Countermeasures project at the University of Pennsylvania. His research interests include secure systems, cryptology and cryptographic protocols, and large-scale systems. Blaze has a PhD in computer science from Princeton University. He is a member of the ACM, IACR, and the IEEE, and is a director of the Usenix association. Contact him at mab@crypto.com.*

Whitfield Diffie *is chief security officer at Sun Microsystems. His technical interests include cryptography, network security, and signals intelligence. Diffie has a doctorate in technical sciences from the Swiss Federal Institute of Technology. Contact him at whitfield.diffie@sun.com.*

Susan Landau *is a distinguished engineer at Sun Microsystems, where she works on security, cryptography, and public policy, including surveillance issues, digital rights management, and identity management. She is coauthor (with Whitfield Diffie) of* Privacy on the Line: the Politics of Wiretapping and Encryption, *updated and expanded edition (MIT Press, 2007). Landau has a PhD from MIT, an MS from Cornell University, and a BA from Princeton University. She is an AAAS fellow and an ACM distinguished engineer. Contact her at susan.landau@sun.com.*

Peter G. Neumann *is principal scientist in the Computer Science Lab at SRI International, where he is concerned with computer systems and network trustworthiness, security, reliability, survivability, safety, and many risk-related issues such as voting system integrity, crypto policy, social implications, and human needs, including privacy. He is the author of* Computer-Related Risks *(Addison-Wesley, 1995). Neumann has doctorates from Harvard and Darmstadt. He is a fellow of the ACM, the IEEE, and AAAS. He moderates the ACM Risks Forum (comp.risks, www.risks.org). Contact him at neumann@csl.sri.com.*

Jennifer Rexford *is a professor of computer science at Princeton University, where she works on Internet measurement, routing protocols, and network management. She is coauthor (with Balachander Krishnamurthy) of* Web Protocols and Practice: HTTP/1.1, Networking Protocols, Caching, and Traffic Measurement *(Addison-Wesley, 2001). Rexford has a PhD and MSE from the University of Michigan, and her BSE degree from Princeton University. Contact her at jrex@cs.princeton.edu.*