Risking Communications Security: Potential Hazards of the "Protect America Act"

Steven M. Bellovin, Columbia University
Matt Blaze, University of Pennsylvania
Whitfield Diffie, Sun Microsystems
Susan Landau, Sun Microsystems
Peter G. Neumann, SRI International
Jennifer Rexford, Princeton University

October 22, 2007

Abstract

1 Introduction

The Protect America Act passed in August 2007 changes U.S. law to allow warrantless foreign-intelligence wiretapping from within the U.S. of any communications believed to include one party located outside the United States. U.S. systems for foreign intelligence surveillance located outside the United States minimize access to the traffic of U.S. persons by virtue of their location. The new law does not — and could lead to surveillance on a unprecedented scale that will unavoidably pick up some purely domestic communications. The civil-liberties concern is whether the new law puts Americans at risk of spurious — and invasive — surveillance by their own government. The security concern is whether the new law puts Americans at risk of illegitimate surveillance by others. We focus on security. If the system is to work, it is important that the surveillance architecture not decrease the security of the U.S. communications networks.

The choice of architecture matters; minor changes can have significant effects, particularly with regard to limiting the scope of inadverdent interception. In attempting to collect communications with one end outside the United States, the new law allows the development of a system that will probably pick up many purely domestic communications. How will the collection system determine that communications have one end outside the United States? How will the surveillance be secured?

We examine security risks posed by the new law and put forth recommendations to address them. We begin by presenting background, first legal and policy, and then technical. Next we

examine the difficulties in monitoring international Internet traffic. We follow with a general discussion of risks in communications surveillance systems and then an analysis of those we fear may result from implementing the Protect America Act. We conclude with a set of recommendations regarding design and implementation.

2 Background

2.1 Legal and Policy Issues

Prior to the Protect America Act, U.S. wiretapping law was essentially governed by Title III of the Omnibus Safe Streets and Crime Control Act of 1968[1], which regulated the procedure for wiretaps in criminal investigations, and FISA, 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, email account, etc.) on which the tapping was to occur [1, pp. 323-332, 338-341]. Law enforcement obtained a warrant and communicated this information to the communications provider, which installed the tap.

The Need for Oversight

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 [17] and 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 United States knew, another branch of the government can provide the objectivity necessary for such an investigation. Public knowledge also matters. When the FISA court was dissatisfied with the administration in 2002, it declassified its opinion, helping to shape the later debate on the PATRIOT Act renewal and other administration requests for changes in the wiretap laws.

In the United States, government officials learned the hard way that oversight was critical if surveillance technologies were to be kept within legal bounds. During the Watergate era, a special Senate Committee (the Senate Select Committee to Study Governmental Operation with respect to Intelligence Activities) investigated thirty-five years of government electronic surveillance in the United States, uncovering many abuses. These included wiretaps on Congressional staff and Supreme Court Justices as well as the wiretapping of Martin Luther King for many years[2] and 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

[1] 18 U.S.C. 2510 et seq.

[2] One incident involved wiretapping during the 1964 Democratic Party convention when President Lyndon Johnson was supporting the seating of one delegation from Mississippi and King, was supporting another; the tapping enabled Johnson to learn about King's strategy and counter it.

Strike for Peace. It was clear that the "national-security" grounds for many of the wiretaps were not justified. The requirements governing FISA wiretapping were lifted almost verbatim from the carefully crafted recommendations of the Senate committee report [2, pp. 292-330]. 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; see [3, pp.280-285]).

The U.S. Foreign Intelligence Surveillance Act (FISA), first adopted in 1978, governs electronic surveillance of communications within the United States for foreign intelligence purposes. It permitted surveillance with warrants in three basic cases:

- any person in the United States communicating via wire
- a U.S. person [3] in the United States whether communicating via wire or radio
- any person in the United States communicating via radio with people all of whom are in the United States [4].

The law was also clear in its exception: no warrant was required to intercept radio communications between persons in the U.S. and persons abroad unless the government was intentionally targeting a particular known U.S. person who was in the U.S. This exception was viewed as a temporary one; the Senate Judiciary Committee Report on the FISA legislation makes it clear that interception of radio communications was to be considered separately [4, p. 34]. But separate legislation never came to pass, and so the warrantless exception continued.

The U.S. is a major communications hub in our communication-centered world, giving NSA significant opportunities for access to traffic. There are numerous reasons for U.S. centrality in the world's communication's systems. One is cost: the U.S. is the world's leading economy and fiber optic cables — which is how modern wired communications travel — have been built to the United States. With their economies of scale, these cables enable U.S. providers to underbid regional carriers. For example, much of South America transits its traffic through Miami. Another 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 over AT&T lines. A third reason is the Internet. Many of the servers that are the very reason for communication — yahoomail, hotmail, gmail, etc. — are in the U.S. (though 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 foreign communications that travel by radio. Thus the exemption allowing warrantless interception became increasingly less applicable. In recent years the National Security Agency (NSA), the U.S. signals intelligence agency, pressed to have the exemption updated. While many in the field agreed that there was plausibly a problem as a result of fiber-optic cables, the Protect America Act (PAA)[5], passed in August 2007, was an entirely different matter.

[3] U.S. citizens, permanent residents, and U.S. corporations, per 50 U.S.C. §1801 (i).

[4] 50 U.S.C. §1801 (f); the rules are, in fact, even more complicated, but this is sufficient for our purposes.

[5] P.L. 110-55.

At issue was the dropping of the warrant requirement for communications (in any medium) of U.S. 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 U.S. and the question is on what basis communications would be collected.

Surveillance Excesses: Then and Now

Some might argue that the excesses of surveillance in the nineteen-sixties and seventies 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 over the FBI's abuse of the National Security Letters, "administrative" subpoenas that are issued with *no* judicial oversight and that require the recipient to turn over certain records. The IG concluded that FBI agents may have violated the law 3,000 times since 2003 in their collection of telephone and financial records of U.S. citizens and foreign nationals [15].

2.2 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 here. 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 may be kept briefly or for very long periods.

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

Increasingly communications are IP-based. The Internet is the interconnection of many networks (this is the origin of the term) and the connections occur in various ways. For the largest networks, they occur at peering connections: interconnections between administratively separate domains. International communications enter the United States by satellite, terrestrial microwave, older copper cable, and newer fiber-optic cable. There are about twenty-five cableheads in the United States. At the cablehead, the incoming signals are split in several ways. First, optical and digital cross-connects are used to send various channels 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, etc.

[6] Protect America Act, §105(a), 2007.

2.3 A Likely Architecture

Since information about the design of the NSA surveillance architecture is not public, it is impossible to know exactly what this architecture might be. However, a current court case gives hints. In late 2005 and spring 2006 the *New York Times* and *USA Today* revealed post-September 11th warrantless wiretapping by the NSA. Shortly afterwards, 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 architecture for NSA surveillance 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 of the AT&T San Francisco office in which the electronic surveillance took place [5]. 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) [6] and J. Scott Marcus (a designer of large-scale IP-based data networks, former CTO at GTE Internetworking and at Genuity, and former senior advisor for Internet Technology at the Federal Communications Commission) [7].

Optical fiber carrying inter-ISP peering traffic associated with AT&T's Common Backbone [7, p. 15] was "split," dividing the signal so that 50% went to each output fiber (the weakened signal on each output fiber still contained sufficient information to allow reading the communications) [7, pp. 12-14]. One of the output fibers was diverted to the secure room; the other carried the communications on to the AT&T switching equipment. The secure room contained traffic analyzers and logic servers made by Narus Inc.; Narus states that such devices are capable of doing real-time data collection (recording data for consideration) and capture at high data rates. Certain traffic was selected and sent over a dedicated line to a "central location." The setup in the San Francisco office was one of many throughout the country, including in Seattle, San Jose, Los Angeles, and San Diego [6, p. 7]. 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," [7, p. 24] 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" [7, pp. 24-25].

2.4 Call Detail Records

Modern telecommunications allow the construction of smooth-running organizations that span the globe; telecommunications are the nervous systems of these organizations. To listen to an organization's communications is to read its mind; following just the pattern of its communications, Call Detail Records, is a long step in this direction. CDRs contain very complete call traffic data (calling and called numbers and location, time of day, call duration, etc.) and provide a window into the past. CDR data is used internally by the phone companies for billing, engineering, marketing, and fraud detection. Unlike a wiretap or pen register, which provide, respectively, access to the content or number being dialed in real time, a CDR database contains a wealth of data on past calls. Thus an interested government agency need not have the proper legal authorization or technology in place before a call is made but may search the call detail database later, once a suspect has been identified. For international calls and some purely domestic calls, two CDRs exist

[7] United States Second District Court for Northern California, Case 3:06-cv-0672-vrw.

for each communication, one from the origination point — which may be an interface to another company — and one from the termination.

Although historically transactional information has been viewed as much less deserving of privacy protection than call content, in fact access to CDRs can be a major privacy risk. Cortes et al. showed, for example, that, even though the calling number had changed, it was possible to identify an individual caller from a 300-terabyte CDR database by simply looking at called number patterns [8]. George Danezis relates a story in which Intel Corporation researchers studying ambient Bluetooth activity to improve ad-hoc routing protocols issued its 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[9, pp. 7-8].

The "reasonably believed to be located outside the United States" aspect of the PAA arguably changes the rules on using Call Detail Records (CDRs). CDRs are records of such transactional information as calling and called numbers for phone calls, IP addresses and user URI in the case of VoIP, SMTP headers for email, etc., time and date of communications, etc. They can be surprisingly revelatory of relationships and organizational structure (although this data does not always reveal where the parties to a communication are physically located). CDRs can, in particular, be used for targeting further surveillance, i.e., 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.

3 Difficulties in Monitoring International Internet Traffic

Monitoring international traffic requires an effective way to identify whether the communication starts or ends outside the United States. This is a surprisingly difficult problem to solve on today's Internet. Perhaps even more surprisingly, this is not an easy task on the 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" [10, p. 36]. (Or, to put it another way, CDR is an amazingly effective guide to to communications activity, but the data can't always provide real-time answers to the location of a call.) NSA has worked on the problem, and the agency even has a patent for using time latency to determine a communication's location (U.S. Patent # 6,947,978: Method for geolocating logical network addresses).

Monitoring international traffic requires either (i) limiting monitoring to links that carry only international traffic or (ii) filtering out any traffic transferred between two domestic hosts. The first approach seems easy if monitoring is limited to the cableheads terminating links connecting the U.S. 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 serious technical problems.

While 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 U.S., due to the vagaries of Internet routing[8]. As such, monitoring the

[8]This is partially a result of a 1940s AT&T master plan that made the U.S., Canada, and most of the Caribbean one

links at the U.S. borders, with the goal of warrantless tapping of international traffic, could lead to unintentional tapping of domestic traffic. Because these links operate at very high speed, it is difficult to analyze the 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 — 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 (i) both directions of the traffic are visible and (ii) the link speeds are typically small enough for detailed data collection. But monitoring near the domestic end-point would almost certainly capture a large amount of traffic exchanged with other U.S.-based hosts. To identify and filter the domestic traffic, 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 geo-location services (such as Quova, www.quova.com). Although geo-location mapping services are often accurate to small 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. Geo-location services apply techniques to limit these errors, but the techniques are necessarily imperfect.

Even if the geo-location services are accurate, the source and destination addresess 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 may 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 may correspond to the relay node, rather than one of the communicating end-points, complicating the efforts to determine whether both end-points are domestic. In addition, some users apply anonymization tools like Tor (The onion router) that intentionally hide the source and destination addresses from packet sniffers. Whether the traffic traverses a relay or an anonymizer, the monitor may 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. For example, an IP address may correspond to a NAT box. 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 may 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 [11]. Mapping from the host to a particular user is difficult if the machine is shared among many people, as in a cyber-cafe or an academic lab. In addition, mobile hosts such as laptops or PDAs acquire new IP addresses frequently (see e.g.,[12]).

integrated country, with no cableheads, or even international gateways, between them.

Even if the communicating end-points can be appropriately identified, determining what application they are running is difficult. In the simplest case, applications are easily discernable from numerical identifiers (i.e., "port numbers") in the 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. (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 U.S. to other countries. In addition, a malicious party trying to avoid detection might intentionally pad or jitter the packets to evade detection, adding further complexity to an already difficult problem. Finally, some applications, like Skype, encrypt the data, making it difficult to extract meaningful information about the content of the communication between the end hosts.

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

4 Risks

Surveillance technology is an "architected security breach" [16, p. 418] into a communications network and thus a risky business to embark upon. Two situations illuminate different reasons for our concern.

For almost a year beginning in April 2004, over one hundred phones belonging to members of the Greek government, including the prime minister, ministers of defense, foreign affairs, justice and public order, and opposition members in the Greek parliament, were wiretapped through surreptitious software that turned official built-in tapping capabilities —capabilities to be invoked *only* with legal authorization — to the advantage of as yet unknown parties. What is known is that private communications at the highest levels of the Greek government were wiretapped for ten months [13].

The United States has also experienced difficulties with communications surveillance systems. Under the Communications Assistance for Law Enforcement Act, the FBI was responsible for determining technical specifications for wiretapping built into switches of digital telephone networks, and DCS 3000 was designed to meet those requirements. Recently released FBI documents reveal serious problems in the system's implementation [9]. The system's auditing system was primitive, surprising for a system intended for collection of evidence. The system has no unprivileged userids, relied on passwords rather than token-based or biometric authentication, and even used an outdated hashing algorithm[10]. Most seriously, the system relied on a single shared login, rather

[9] See http://www.eff.org/flag/061708CKK/

[10] MD5 appears in a 2007 "system security plan,"[14, p. 32] several years after Chinese researchers found serious

than a login per authorized user. The ability to audit user behavior depended entirely on people 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. Insider attacks are a real risk. Recall that the most damaging spy in FBI history, Robert Hanssen, abused his authorized access to the internal FBI computer systems to steal information and to track progress of the investigation aimed at him.

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

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

- The apparent removal of the protective role provided by the communication carriers in all previous interception efforts within the U.S. communication system.
- The placement of the system properly within the U.S. rather than at borders.
- The likelihood that the system will be built out of pieces previously used abroad. This runs the risk that opponents will already be familiar with the equipment via intelligence-sharing agreements, capture of equipment, etc.
- The use of Call Detail Records, 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:

- Capture of the system to enable spying on U.S. traffic.
- Defeat of the system by making use of information learned from foreign examples to defeat its selection and filtering strategies.
- Spoofing of the system by similar means.

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

Risk 1: Risk of exploitation by opponents: A system that is accessing domestic communications necessarily poses a greater direct risk to the communications of Americans than a surveillance

problems with this already weak hashing algorithm.

system fielded overseas. Who controls the filter? Is NSA designing sufficiently robust mechanisms to assure complete control? Engineering economy makes the reuse of systems previously fielded abroad likely and thus likely to be familiar to both allies and opponents. Communication security equipment is often not shared with allies, explicity to avoid foreign familiarity with its operation. Is there a risk that knowledge of the surveillance system acquired by studying eqipment outside the U.S. will be applied to defeating or subverting similar equipment deployed within the U.S.?

Even prior to the PAA, U.S. communications were vulnerable to surveillance, but building SIGINT systems is expensive. The system designed as a result of the PAA must not simplify foreign powers' ability to gain access to U.S. communications. Can the wiretapping of communications of U.S. persons be done without increasing the risk that their communications will be exploited by others who are not authorized to do so?

Risk 2: Removal of safeguards by communications carriers: Previous approaches to foreign intelligence surveillance of U.S. persons went through the communications carriers, who combine technical expertise regarding communications with responsibility for their customers' security and privacy. What risks are introduced by leaving a single entity in charge of selection and retention decisions? "Two-person control" — control by two authorized parties who know how a system should work — is as applicable to organizations as to individuals. The process apparently embodied in the AT&T San Francisco in which communications are diverted to an NSA safe room and then collected according to rules determined by the intelligence agency provides no recourse in cases where "mistakes were made."

Risk 3: Lack of inherent technical minimization of traffic: Intercepting at switches creates unnecessary risks because the switches handle domestic as well as foreign communications. This risk, different from risk 1, feeds into that risk; potential overcollection of purely domestic traffic increases the value of targeting the U.S. access and collection system.

Risk 4: Domestic traffic penetrating too deeply into the NSA collection system: Collection outside the U.S. makes it easier to filter out "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 effectively? This is more of a privacy risk than a security one, though insider attacks means that it is also a security risk.

Risk 5: Call Detail Record 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 an old system, since the protection mechanisms and system architecture were never designed for such 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.

There are also ways in which the Protect America Act enables an architecture that may reduce risk. Being able to place equipment on U.S. 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 U.S. communications is hard to assess — but inspection authority over the installation and information about the 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.

Note that we have not enumerated all possible future scenarios. In particular, the security risks will be 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 radio-frequency ID (RFID) tags and sensors will use networks for communication and control. While many of these devices will be on local-area networks, others will make use of the Internet. [16, pp. 433-434] Any future surveillance architectures must take such growth and directions into account.

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 U.S. communications, a target of great value. The nation may build for its opponents something that would be too expensive for them to build for themselves: a system that allows them to see the intelligence interests of the U.S., a system that may 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 U.S. communications by unauthorized parties nor permit abuse by authorized ones.

5 Recommendations

The change from a system that wiretaps particular lines upon 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 the warrantless wiretapping is, it will 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 U.S. 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 U.S. government agents. How should the U.S. mitigate the risks?

Minimization matters. Allowing collection of calls on U.S. territory necessarily entails greater access to the communications of U.S. persons. An architecture that minimizes the collection of communications lowers the risk of exploitation by outsiders and exposure to insider attacks. Traffic should be collected at international cableheads rather than at tandem switches or backbone routers, which also carry purely domestic traffic. Surveilling at the cableheads will help minimize collection but it is not sufficient in and of itself. Intercepted traffic should be studied (by geo-location and any other available techniques) to determine whether it comes from non-targeted U.S. persons and if so, discarded before any further processing is done. It should be fundamental to the design of the system 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. "Two-person control" is applicable to organizations as well as individuals.

Oversight matters. The new system is likely to operate differently from previous wiretapping regimes, and likely to be using new technologies for purposes of targeting wiretaps. There should be appropriate oversight by publicly accountable bodies. While the details of problems may 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 functioning of the system. Of particular concern is the real-time use of CDR for targeting content, which must neither be abused by the U.S. 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 the 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?"

Security of U.S. communications has always been fundamental to U.S. 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 may be greater than the gain. To prevent greater threats to U.S. 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 may be fundamentally unacceptable.

References

[1] Solove, Daniel and Marc Rotenberg, *Information Privacy Law,* Aspen Publishers, 2003.

[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, April 23, 1976.

[3] Diffie, Whitfield and Susan Landau, *Privacy on the Line: The Politics of Wiretapping and Encryption, updated and expanded edition,* MIT Press, 2007.

[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, November 15, 1977.

[5] Exhibit A, *Tash Hepting et al. v. AT&T Corporation et al.,* United States Second District Court for Northern California, Case 3:06-cv-0672-vrw, June 8, 2006.

[6] Klein, Mark, 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, June 8, 2006.

[7] Marcus, J. Scott, 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, June 8, 2006.

[8] Cortes, Corinna, Daryl Pregibon and Chris Volinsky, "Computational Methods for Dynamic Graphs," AT&T Shannon Labs, January 9. 2004.

[9] Danezis, George, "Introducing Traffic Analysis: Attacks, Defences and Public Policy Issues."

[10] Schneider, Fred, *Trust in Cyberspace,* Computer Science and Telecommunications Board, National Research Council, 1999,

[11] Clayton, Richard, *Anonymity and Traceability in Cyberspace,* University of Cambridge Computer Lab, Technical Report Number 653, November 2005.

[12] Bellovin, Steve, Matt Blaze, Ernie Brickell, Clinton Brooks, Vint Cerf, Whitfield Diffie, Susan Landau, Jon Peterson, John Treichler, "Security Implications of Applying the Communications Assistance to Law Enforcement Act to Voice over IP," http://www.itaa.org/news/docs/CALEAVOIPreport.pdf, 2006.

[13] Prevelakis, Vassilis and Diomidis Spinellis, "The Athens Affair," *IEEE Spectrum,* July 2007, pp. 18-25.

[14] Information Assurance Section, Federal Bureau of Investigation, "Controlled Interface 100 (CI-100) System Security Plan (SSP) DCS-3000 to EDMS," April 16, 2007.

[15] Inspector General, Federal Bureau of Investigation, *A Review of the Federal Bureau of Investigation's Use of the National Security Letters,* March 2007.

[16] Landau, Susan, "National Security on the Line," *Journal on Telecommunications and High Technology Law,* Vol. 4, No. 2, (2006).

[17] 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 September 10, 2002, S. Hrg. 107-947, One Hundred Seventh Congress, Second Session.

[18] United States Foreign Intelligence Surveillance Court, *In Re All Matters Submitted to the Foreign Intelligence Surveillance Court, Memorandum Opinion,* May 17 2002.