Toward a Culture of Cybersecurity Research

Aaron J. Burstein[*]

Draft as of July 31, 2007

Abstract

Improving cybersecurity is an urgent national priority. Research offers great promise of addressing cybersecurity threats in the short and long term. Progress in cybersecurity research, however, is beset by a lack of access data from communications networks. Legally and informally protected individual privacy interests have contributed to the lack of data, as have the institutional interests of organizations that control these data. A modest research exception to federal communications privacy law would remove many of the legal barriers to sharing data with cybersecurity researchers. This reform would also counter many of the non-legal objections that network providers have to sharing data.

[*] TRUST and ACCURATE Research Fellow, Samuelson Law, Technology & Public Policy Clinic and Berkeley Center for Law & Technology, University of California School of Law (Boalt Hall), `aburstein@law.berkeley.edu`. I am deeply grateful to Deirdre Mulligan for her mentorship, for sharing many critical ideas during the writing of this Article, and for reviewing earlier drafts. I am also grateful to the many cybersecurity researchers who shared their insights into the practical aspects of obtaining data to conduct their research, but who have preferred to remain anonymous. This work was supported in part by TRUST (The Team for Research in Ubiquitous Secure Technology), which receives support from the National Science Foundation (NSF award number CCF-0424422). The views expressed in this Article are mine and do not purport to represent the views of the NSF, or any other person.

Contents

1 Introduction

Computer and network security (together, "cybersecurity") have become matters of major economic, social, and national security importance. Present-day attacks on computer systems and networks do not simply damage an isolated machine, or disrupt an individual's or single enterprise's access to the Internet. Instead, modern attacks target *infrastructure* that is integral to the economy, national defense, and daily life. Computer networks have joined food, water, transportation, and energy as critical resources for the functioning of the national economy.[1] Indeed, computer networks are the "nervous system" of our national infrastructure.[2]

Ever since the first computer "worm" traversed the Internet,[3] it has been apparent that attacks can spread rapidly across organizational boundaries. Just as society has benefited from the nearly infinite connections of devices and people through the Internet,[4] so have malicious parties taken advantage of the Internet's connectivity to "launder" connections that might lead back to them.[5] As networking becomes pervasive—integral not only to computers and cell phones but also to appliances and even building materials—the potential harm from attacks that spread via networks from user to user, or enterprise to enterprise, will continue to grow.[6]

[1] *See* President's Critical Infrastructure Protection Bd., *National Strategy to Secure Cyberspace* vii (Feb. 2003) [hereinafter *National Strategy to Secure Cyberspace*]. For an analysis of computer networks as infrastructure, see Brett M. Frischmann *An Economic Theory of Infrastructure and Commons Management*, 89 Minn. L. Rev. 917 (2005).

[2] *National Strategy to Secure Cyberspace, supra* note 1, at 1.

[3] *See* United States v. Morris, 928 F.2d 504 (2d Cir. 1991) (upholding conviction of the worm's author under the Computer Fraud and Abuse Act, 18 U.S.C. § 1030). A worm is a computer program that replicate itself. US-CERT, Security of the Internet, http://www.us-cert.gov/reading_room/tocencyc.html (last visited Feb. 12, 2007).

[4] *See* Jonathan Zittrain, *The Generative Internet*, 119 Harv. L. Rev. 1974, 1980 (2006), (describing the Internet's "consummately generative" combination of computers and networks).

[5] *See* Computer Science and Telecommunications Board (CSTB), *Cybersecurity Today and Tomorrow: Pay Now or Pay Later* 4 n.9 (2002), *at* http://books.nap.edu/html/cybersecurity/ ("Tracing attacks is generally difficult, because serious attackers are likely to launder their connections to the target. That is, an attacker will compromise some intermediate targets whose vulnerabilities are easy to find and exploit, and use them to launch more serious attacks on the ultimate intended target.") [hereinafter *Cybersecurity Today and Tomorrow*].

[6] Computer Science and Telecommunications Board (CSTB), *Toward a Safer and More Secure Cyberspace* (Seymour E. Goodman and Herbert S. Lin, eds.) (June 26, 2007) 17-18,

Another long-standing lesson of Internet-based attacks is that sharing information about vulnerabilities and malicious activity can help prevent or mitigate harm.[7] In particular, disclsoures of network data among the separate organizations that constitute the Internet can help detect threats that no single organization could identify. Sharing network data also gives individuals and organizations time to react once a threat has emerged.

Developing defenses against cross-organizational threats, including methods to share information about these threats, is therefore critical to cybersecurity. But attacks change rapidly and often display tremendous (though destructive) ingenuity; keeping pace with them through research is a constant challenge.[8] Basic research is therefore likely to remain an important part of improving cybersecurity. Providing researchers with data about real-world networks, software vulnerabilities, and security practices from diverse organizations is essential to supporting this research.[9]

Sharing cybersecurity data, however, carries considerable risks. Many types of data relevant to cybersecurity research relate to Internet use, and these data generally receive at least some privacy protection under the Electronic Communications Privacy Act (ECPA). Even when data are not protected by the ECPA, the firms that control them are reluctant to share data with outsiders out of concern that the firm's users will react negatively, or that disclosing the data will harm the firm's own security or comeptitive position. The result is that researchers face a dearth of relevant data, and are often unable to obtain data from outside their own institutions.

These legal and practical difficulties are symptoms of a deeper problem in cybersecurity. Our current cybersecurity culture simply does not encourage coordination against threats. The Internet's diversity of ownership and interests contributes to this problem. As two cybersecurity researchers have put it:[10]

at http://books.nap.edu/openbook.php?record_id=11925&page=17.

[7] For a review of how an informal network of systems administrators spread information to aid defense against the worm at issue in *Morris*, see Zittrain, *The Generative Internet*, *supra* note 4, at 2003-07 (2006).

[8] *See, e.g.,*, Yinglian Xie et al., *Forensic Analysis for Epidemic Attacks in Federated Networks* 43, 43 (2006), in *Proceedings of the IEEE International Conference on Network Protocols (ICNP)* (2006) [hereinafter Xie et al., *Epidemic Attacks*].

[9] *See National Strategy to Secure Cyberspace*, *supra* note 1, at 8.

[10] Adam Slagell and William Yurick, *Sharing Computer Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization* 1 [hereinafter Slagell & Yurick, *Sharing Network Logs.*

> It is typical in the current security culture for each autonomous organization . . . to locally optimize network management and security protection. . . . There is a culture of pushing attackers away from oneself without any consideration of the poor overall security resulting from this lack of coordination between organizations.

Add to this the fact that the current cybersecurity culture encourages individuals and institutions to view security as an expense without any offsetting benefit, and the depth of the cybersecurity problem becomes apparent.[11]

Both Congress and the Executive Branch have recently taken an interest in overcoming the lack of coordination in sharing cybersecurity information.[12] In particular, the guiding cybersecurity policy document, the *National Strategy to Secure Cyberspace*, recognizes that a new approach is necessary to encourage firms with data to share them with researchers. The *National Strategy to Secure Cyberspace* also recognizes that cybersecurity responses must protect privacy and civil liberties.[13] Anyone who searches this document for details about how to reconcile these privacy and security objectives, however, will be disappointed.[14]

In this Article I argue that leaving the privacy conversation for another day serves neither privacy nor cybersecurity. My argument proceeds in four parts. First, the economic and technical components of cybersecurity render market-based and law enforcement-based efforts to improve cybersecurity inadequate. Improving cybersecurity depends critically on continued research. Second, I argue that communications privacy law and norms contribute significantly to the cybersecurity data dearth. The ECPA, in particular, has a distinctive model of security—one that strongly reinforces the current cybersecurity culture's disposition against coordination. A close examination

[11] *See* CSTB, *Toward a Safer and More Secure Cyberspace*, *supra* note 6, at 15.

[12] *See* Cyber Security Research and Development Act of 2002, Pub. L. No. 107-305, which appropriated $377 million for the National Science Foundation to grant to cybersecurity researchers. *Id.* §§ 4(a)(3), 4(b)(7). A finding in the law is that increased information sharing is needed.

[13] *National Strategy to Secure Cyberspace*, *supra* note 1, at 14. *See also id.* at 54 ("Cybersecurity and personal privacy need not be opposing goals. Cyberspace security programs must strengthen, not weaken, such protections. The federal government will continue to regularly meet with privacy advocates to discuss cybersecurity and the implementation of this *Strategy*.").

[14] *See id.*

of the ECPA's language and structure shows that the current law does litte to advance individual privacy interests by prohibiting disclosure of data to cybersecurity researchers. In addition, by making data sharing with cybersecurity researchers a legally risky proposition, the ECPA also reinforces institutions' many economic reasons to countenance the data dearth. Third, the dearth of usable data is, in fact, a serious impediment to research. Increasing cybersecurity researchers' access to data would significantly aid this research. Fourth, a variety of measures are necessary to ease the cybersecurity data dearth. The first step is to create a cybersecurity research exception to the ECPA. The provisions that I outline for this exception would leave the baseline levels of individual privacy protection and law enforcement access to communications essentially unchanged. This legislatively enacted change would update ECPA's security model to one that recognizes the need for cybersecurity coordination, and it would confer legitimacy on the use of communications data in research. This legitimacy could, in turn, create incentives for firms to share data with cybersecurity researchers, rather than keeping it behind institutional walls. Institutional controls are integral to the exception that I propose and could serve to address the security and competition concerns that accompany data disclosure to cybersecurity researchers.

2 Cyber-Insecurity

To avoid, at the outset, the possibility that "security" will become too malleable a term in this Article, I will attempt a definition. A somewhat formal definition includes: a computer or network system's resistance to becoming unavailable or unusable; resistance to attacks that corrupt data stored on the system; and resistance to attacks that cause information to leak out of the system.[15] A more functional definition emphasizes that security involves a process of identifying and remedying the vulnerabilities of a system within the context of a specified set of threats posed by an adversary;[16] cybersecurity

[15] CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 3. *See also* Computer Science and Telecommunications Board (CSTB), *Trust in Cyberspace* 14 (Fred B. Schneider, ed.) (1999), http://bob.nap.edu/html/trust/trust-1.htm#Page%2014, (defining "security" to mean that a system "resists potentially correlated events (attacks) that can compromise the secrecy, integrity, or availability of data and services").

[16] *See, e.g.,* 17 U.S.C. § 1201(e) ("[T]he term 'information security' means activities carried out in order to identify and address the vulnerabilities of a . . . computer, computer system, or computer network."). *See also* CSTB, *Trust in Cyberspace*, *supra*

applies these activities to networked computer systems.

2.1 Threats

The toll of cybersecurity failures is high and growing. The FBI estimated in 2005 that cybercrime costs the United States $67.2 billion annually.[17] But the risks of insecurity go beyond far beyond financial damage. As numerous academic and government reports have noted, networks connect national defense facilities; electrical power generation facilities; physical infrastructure such as airports, highways, bridges, and dams; and distribution systems for food and drinking water.[18] Major physical infrastructure in the United States has not failed as a result of cyber attacks, but attempts to defeat the defenses of government systems are relentless. The Department of Defense has reported, for example, that it experiences approximately 40 successful cyber attacks per month and tens of thousands of close calls per year.[19]

Incidents from abroad give a fuller sense of the potential for cybersecurity breaches to bring serious disruptions to a national economy. In May 2007 Estonia endured a massive flood of Internet traffic, which crippled networks within the country, leading to a shutdown of banks and other services.[20] In 2003, the "Slammer" worm spread rapidly across the Internet, shut down South Korea's "entire Internet system" and disrupted ATM transactions in the United States.[21] Another worm in 2004 deleted random data from the hard drives of the hosts it infected worldwide.[22]

These examples may well understate the possibility for more serious disruptions. Specifically, none of them appear to involve the backing of a sovereign foreign power. A recent GAO report, however, noted that "Chi-

note 15, App. K (defining "vulnerability," "threat," and "exploit").

[17] U.S. Govt. Accountability Office (GAO), *Cybercrime: Public and Private Entities Face Challenges in Addressing Cyber Threats* 15 (June 2007) [hereinafter GAO, *Cybercrime*].

[18] GAO, *Critical Infrastructure: Challenges Remain in Protecting Key Sectors* 8, Mar. 20, 2007 [hereinafter GAO, *Critical Infrastructure*]; *National Strategy to Secure Cyberspace*, *supra* note 1, at 1; CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 1.

[19] Bob Brewin, *Successful cyberattacks against DOD drop*, FCM.COM, Mar. 29, 2007, http://www.fcw.com/article98089-03-29-07-Web&printLayout.

[20] John Schwartz, *Bit Wars: When Computers Attack*, N.Y. TIMES, June 24, 2007.

[21] *Internet Worm Strikes*, Herald Sun (Melbourne, Australia), Jan. 28, 2003.

[22] Colleen Shannon and David Moore, *The Spread of the Witty Worm*, 2 IEEE SECURITY & PRIVACY 46 (July-Aug. 2004).

nese strategists are writing about exploiting the vulnerabilities created by the U.S. military's reliance on technologies and attacking key civilian targets."[23] There is a healthy and continuing debate over whether the potential for "cyberterrorism" has been overstated.[24] Some argue that the absence of truly catastrophic failures of the Internet due to malicious activity provide a false sense of confidence in the current level of cybersecurity, while others argue that predictions of a "digital Pearl Harbor" are overblown and might be attributable to attempts by governments to gain surveillance power, or by private firms to sell products or services. These differences often reduce to disagreements about the likely motivations of attackers or about how best to allocate finite resources to spend on security;[25] there is widespread agreement that the Internet and the computers connected to it face serious security risks.

2.2 Causes

2.2.1 Vulnerable Technologies

The examples given above also illustrate how small events—a vulnerability in one program that allows a computer to be compromised, for example—can lead to large-scale attacks. Examining the means for these attacks in a bit more detail will show the importance of sharing data from many organizations in order to detect and prevent attacks.

Consider the attack on the Internet in Estonia.[26] The means for carry-

[23] GAO, *Cybercrime*, *supra* note 17, at 15. This statement was based on the U.S.-China Economic and Security Review Commission's 2006 report to Congress.

[24] *See, e.g.,* John D. Podesta and Raj Goyle, 23 YALE L. & POL'Y REV. 509 (2005) (stating that "[c]yberterrorism poses a significant danger that requires a strong and unequivocal response, but such a response need not sacrifice important constitutional safeguards").. *Cf.* CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5 at 6 ("[A] successful cyber-attack launched on the air traffic control system in coordination with airliner hijackings could result in a much more catastrophic disaster scenario than was seen on September 11, 2001.").

[25] *See* Eric Rescorla, *The Internet Is Too Secure Already*, presentation USENIX Security (2003).

[26] This type of attack is known as a "distributed denial of service" (DDoS). The flood of traffic in a DDoS attack might not be malicious, in the sense that the computers participating in the attack seek to destroy data or bypass a password or other form of authentication. Instead, the volume of traffic alone can consume enough of the attack victim's resources to render it useless. Depending on the attacker's plans, the response of

ing out this attack was a botnet[27]—a network of individual computers that have been compromised, have malicious software installed on them, and are controlled by a remote, central attacker,[28] To set up a botnet, attackers scan networks for computers that have known vulnerabilities. As the attackers find vulnerable computers, they install software that they will later use during an attack as well as software that allows the host to be remotely controlled. The sources of the vulnerabilities are legion and include operating systems, Web browsers, and common applications.[29] The resulting network might contain thousands[30] or even millions of "zombies" or "drones."[31] As a result, network traffic indicative of a botnet-based attack is usually spread across many different networks. This can make it extremely difficult for any single network operator to pick out attack traffic from the innocuous background. Moreover, the wide distribution of botnet traffic makes it difficult for researchers to obtain data about these networks, since researchers would need to coordinate with many different organizations to gain a complete picture of a botnet-based attack. To add to these difficulties, the legal obligation of a potential source of network data to keep these data private might differ based on the identity of the provider, as well as the recipient. As I explain below, no institutions have arisen to solve this coordination problem.

The portfolio of botnet operators is more diverse than the "denial of service attack" that Estonia experienced. For example, botnets serve as distri-

the victim, and the responses of other Internet infrastructure operators, a DDoS attack can last for hours or longer.

[27] CNet.com, http://news.com.com/Cyberattack+in+Estonia–what+it+really+means/2008-7349_3-6186751.html.

[28] *See* Mark Allman, Ethan Blanton, Vern Paxson, and Scott Shenker, *Fighting Coordinated Attackers with Cross-Organizational Information Sharing* 1 (2006) (describing botnets as "armies of enslaved hosts . . . controlled by a single person or small group") [hereinafter Allman et al., *Fighting Coordinated Attackers. See also* Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose & Andreas Terzis, *A Multifaceted Approach to Understanding the Botnet Phenomenon* 1 (2006) [hereinafter Rajab et al., *Understanding Botnets.*

[29] Niels Provos et al., *The Ghost In The Browser: Analysis of Web-based Malware* 1 *in Proceedings of the First Workshop on Hot Topics in Botnets (HotBots '07)* (2007). The typical weekly software vulnerability report issued by the United States Computer Emergency Readiness Team (US-CERT) contains dozens of reported vulnerabilities. *See, e.g.,* US-CERT, *Vulnerability Summary for the Week of July 2, 2007*, http://www.us-cert.gov/cas/bulletins/SB07-190.html (last updated July 9, 2007).

[30] Rajab, *Understanding Botnets*, *supra* note 28, at 1.

[31] Robert Lemos, *Dutch bot-net suspects infected 1.5 million PCs, officials say*, SecurityFocus, Oct. 20, 2006, http://www.securityfocus.com/brief/19.

bution networks and storage systems for information that is used to commit fraud and identity theft.[32] Botnets have also become a tool of choice for sending spam.[33] In high volumes, spam is more than annoyance; it consumes enough network resources to make the network unavailable for legitimate uses. In both of these uses, the wide distribution of computers in the botnet work to the attackers' advantage and against attack identification and analysis. In the case of spam, the ability of a botnet to send unwanted messages from thousands of computers makes tracing the messages to the sender extremely difficult. In the case of identity theft, distribution on a botnet allows personal financial data to survive even if a few hosts are removed from the botnet. This contributes to a more efficient and reliable black market for financial information.[34] Traces of network activity would be highly valuable to researchers seeking to understand these uses of botnets, but these data are generally unavailable.

Infections of personal computers with botnet-related software have become widespread, reaching perhaps one-quarter of PCs connected to the Internet.[35] Frequently the malicious activity on an infected computer is not perceptible to its owner, even when the computer is participating in an attack. The software that infects individual bots frequently takes steps to hide its tracks from anti-virus software and other forms of forensic detection, such as the inspection of system logs. A botnet controller can quickly scan large numbers of computers and, because the network is centrally controlled, does not need to coordinate his activities with other perpetrators. These factors make setting up a botnet inexpensive and reduce or eliminate the need for human co-conspirators. Moreover, the wide distribution of hosts in a botnet makes malicious activity difficult to trace to the perpetrator. All of these properties make botnets attractive tools for committing crimes, and help to explain why they have emerged as a tool of choice among criminals operating

[32] FBI, *Over 1 Million Potential Victims of Botnet Cyber Crime*, June 13, 2007, *at* http://www.fbi.gov/pressrel/pressrel07/botnet061307.htm [hereinafter "Bot Roast Press Release"].

[33] *See* Robert Lemos, *Bot Nets Likely Behind Jump in Spam*, The Register, Oct. 31, 2006, http://www.theregister.co.uk/2006/10/31/botnet_spam_surge/ (reporting that "[t]here is strong evidence that bot nets—networks of compromised PCs—are behind the recent jump in spam").

[34] John Markoff, *Attack of the Zombie Computers Is Growing Threat*, N.Y. Times A1, Jan. 7, 2007.

[35] *Criminals 'may overwhelm the web'*, http://news.bbc.co.uk/2/hi/business/6298641.stm, Jan. 25, 2007.

on the Internet.

More familiar cybersecurity threats persist and, like botnets, threaten both the security of the information infrastructure itself as well as other forms of infrastructure. Consider the example of computer worms. While some worms simply consume the resources of infected hosts,[36] others carry code that damages the hosts. For example, the “Witty” worm deleted random data from the hard drives of the hosts it infected.[37] Worms can also serve as a tool to form botnets by delivering software that enslaves individual machines to a botmaster.[38] Even if they do not damage the computers that they infect, worms can flood networks with traffic, rendering them unreachable from the outside and unusable within.[39]

The overall picture of cybersecurity threats is complex. It is difficult to separate threats to individual computers from threats to the Internet.[40] Networks allow attackers to exploit vulnerabilities on individual computers; and individual computers serve as launch pads for network-wide attacks. At a technical level, worms and attacks that use botnets can affect many organizations and political entities simultaneously. Understanding these threats, and developing robust, automated methods to defend against them, present steep challenges; and these problems remain the focus of active research.[41]

[36] *See, e.g.*, United States v. Morris, 928 F.2d 504 (2d Cir. 1991).

[37] Shannon & Moore, *The Spread of the Witty Worm*, *supra* note 22.

[38] San Jose Mercury (editorial), http://www.siliconvalley.com/mld/siliconvalley/news/editorial/16251109.htm, Dec. 18, 2006.

[39] *See, e.g.*, *Internet worm strikes*, Herald Sun (Melbourne, Australia), Jan. 28, 2003 (reporting that the SQL Slammer worm caused “South Korea’s entire Internet system” to shut down and disrupted ATM transactions in the United States).

[40] *See* Steven M. Bellovin, David D. Clark, Adrian Perrig & Dawn Song, *A Clean-Slate Design for the Next-Generation Secure Internet* 3 (2006):

> While a network purist might say that the security of the end-host is not the responsibility of the network, if we pose ‘good security’ as an overall goal for a next generation Internet, this promise must make sense to the lay audiences—the public, Congress, and so on. For us to claim good security and not acknowledge the problems faced by average people, such as zombies, phishing, spam, spyware, worms and viruses—all of which involve the end-nodes—would leave the claim of ‘good security’ vacuous to all but a small group of researchers.

[41] *See, e.g.,*, Xie et al., *Epidemic Attacks*, *supra* note 8.

2.2.2 Economic Barriers

Though "there are no silver bullets for 'fixing' cybersecurity,"[42] cyberspace is still far less secure than we know how to make it.[43] Economics also plays an important role in explaining the poor state of cybersecurity. In recent years information economists have made considerable progress in understanding incentives of technology firms to build (in)secure products and of individual users to choose to purchase insecure systems. This area of study has also shed some light on how to understand and alter attackers' incentives.[44]

As a general matter, defending against malicious attacks presents an inherent economic hurdle because the resources required to defend a system are generally far greater than those necessary to attack it. That is, the "overall security of a system is only as strong as its weakest link."[45] This "weakest link" might be technological—a software vulnerability, for example—or it might result from (non-malicious) human action. Whatever may cause a breach, the weakest link principle implies that defending a system is much more costly than attacking it. Computer scientist Ross Anderson has cast this problem in the imagery of old Westerns: "In a world in which the 'black hats' can attack anywhere but the 'white hats' have to defend everywhere, the black hats have a huge economic advantage."[46]

Still, if many of the sources of computer and network vulnerabilities are known,[47] why do technology users and producers fail to fix the problems that they know about and can remedy? One of the major findings of economic studies of cybersecurity is that market-based solutions—the actions of individuals or firms acting out of their own self-interest—meet severe difficulties because cybersecurity is an externality.[48] That is, the security practices of

[42] CSTB, *Toward a Safer and More Secure Cyberspace*, *supra* note 6, at 1-3.

[43] CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 8 ("From an operational standpoint, cybersecurity today is far worse than what known best practices can provide.").

[44] As I discuss in Part 3, examining the incentives of network service providers to provide data to, or withhold it from, cybersecurity researchers sheds considerable light on the cybersecurity data dearth.

[45] CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 7.

[46] Ross Anderson, *Why Information Security Is Hard—An Economic Perspective* 7, in *Proceedings of the First Workshop on the Economics of Information Security.*

[47] CSTB, *Trust in Cyberspace*, *supra* note 15, at 136 (discussing common programming practices that lead to security vulnerabilities).

[48] For a review of recent progress in the economics of information security, see Ross Anderson & Tyler Moore, *The Economics of Information Security*, 314 Science 610 (2006).

one person can affect the security of others. An externality may be positive. For example, a bank's use of a technology to allow customers secure access to their accounts could reduce the chance the chance that an attacker will intercept a bank customers account information and use it to incur fraudulent charges against online merchants. But externalities can be negative, too, as is the case when a vulnerable software product goes unpatched and provides a means to attack others on the network. As scholars working in this area have noted, cybersecurity is a mixture of negative and positive externalities.[49]

On balance, the negative externalities dominate the positives in cybersecurity. This may be seen by examining the incentives of the three major categories of actors in cybersecurity: users, technology suppliers, and attackers. Most users lack both the information and incentives to purchase secure technologies. It is difficult for individual users to distinguish between secure and insecure products.[50] Secure software is typically less convenient to use[51] and is, at best, only as functional as comparable insecure software.[52] Quantifying the return on an investment is also difficult, because security successes are not manifest in the form of a positive payoff, and security improvements often spill over to the benefit of other users—including competitors—on the network.[53] As a result, individual users as well as large institutional users "tend to underinvest in security."[54]

This finding complements other research that has found that computer platforms, software, and standards display network externalities: the more widely used a technology is, the more valuable it becomes to each individual user. *See* Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects*, J. ECON. PERSP. 93 (1986); Peter S. Menell, *Tailoring Legal Protection for Computer Software*, 39 STAN. L. REV. 1329 (1987); Mark A. Lemley, *Legal Implications of Network Effects*, 86 CAL. L. REV. 479 (1998).

[49] *See* Anderson & Moore, *The Economics of Information Security*, *supra* note 48, at 611.

[50] *See id.* at 610 ("Insecure software dominates the market for the simple reason that most users cannot distinguish it from secure software . . .").

[51] *See* Computer Science and Telecommunications Board, *Cybersecurity Today and Tomorrow: Pay Now or Pay Later* 7 (2002) (noting that security "interfere[s] with daily work"), *supra* note 5

[52] *See id.* at 9 ("[A] secure system doesn't allow users to do any more than an insecure system . . .").

[53] *See* CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 9 n.13 ("[A] party that makes investments to prevent its own facilities from being used as part of a DDOS attack will reap essentially no benefits from such investments, because such an attack is most likely to be launched against a different party.").

[54] *See* Anderson & Moore, *The Economics of Information Security*, *supra* note 48, at

Incentives for technology suppliers also frequently point away from building secure systems. Building a secure information system is costly; it requires firms to direct at least some of their engineering efforts toward security rather than toward features that most users would more immediately desire.[55]

Finally, attackers are highly motivated. Their incentives have long been a mixture of prestige and financial gain,[56] but in recent years financial incentives have become increasingly compelling.[57] Other malicious attacks reveal political motives. For example, the attack against networks in Estonia was allegedly launched in response to outrage over the Estonian government's removal of a monument to Soviet troops who fought in World War II.[58] These incentives are and can appeal powerfully to financial, political, or personal interests.

Conversely, the deterrents against malicious activity are weak. The Internet allows attackers to cover their tracks and operate from outside the jurisdiction that they attack, and thus is most likely to prosecute them. Where criminal investigations are possible, law enforcement officials often face stiff challenges in gathering evidence,[59] further reducing the deterrent

610; CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 9.

[55] *See* Anderson & Moore, *The Economics of Information Security*, *supra* note 48, at 610 (noting that "developers are not compensated for costly efforts to strengthen their code" because users frequently cannot tell that it is more secure than comparable products).

[56] At a relatively early stage Congress recognized that financial gains might motivate some perpetrators of computer crimes:

> In some instances, unauthorized access to wire or electronic communications is undertaken for purposes of malice or financial advantage. Other instances, however, arise from the activities of computer amateurs, often called "hackers," whose goal is primarily the access itself. Still, "hacking" cannot be dismissed as a harmless prank . . .

H.R. Rep. No. 99-647, 99th Cong., 2d Sess., 63 (1986). *See also* S. Rep. No. 99-541, 99th Cong., 2d Sess., 36 (noting enhanced criminal penalties for violations of 18 U.S.C. § 2701 where the perpetrator acts for private financial gain).

[57] *See, e.g.,* GAO, *Cybercrime*, *supra* note 17, at 15 ("The overall loss projection due to computer crime was estimated to be $67.2 billion annually for U.S. organizations, according to a 2005 FBI survey.").

[58] *Estonia Suspects Russia of Cyberattacks*, May 30, 2007, http://online.wsj.com/article/SB117942061275506340.html. Estonia's initial allegations that the Russian government was behind these attacks were later dropped. *Estonia Drops Cyberwar Theory, Claims Packets Were "Terrorism"*, June 14, 2007, http://feeds.wired.com/~r/wired27b/~3/123005736/estonia_drops_c.html.

[59] See Part 2.3.2 for a description of these challenges.

effect of criminal penalties.[60]

2.3 Government Responses

Like other conditions of market failure, the multiple levels of market failure in cybersecurity have prompted the government to intervene. Three basic approaches—regulation, law enforcement, and research—are part of the cybersecurity policy discussion.

2.3.1 Regulation

Direct government intervention has been conspicuously absent from the government's approach to improving cybersecurity. As a general matter, private-sector and Congressional opposition to technological mandates have been fairly deep. This is not to say that members of the public and private sectors have not tried and, in some cases, succeeded in passing technological mandates. Copyright is one of these exceptional areas. For example, Congress enacted in the Digital Millennium Copyright requirement for VCRs to use a specific type of technology to prevent serial copying of video cassettes.[61] Later efforts to impose more ambitious mandates, however, met with stiff resistance from consumers and the high technology industry.[62]

On matters of cybersecurity, the government has followed the non-regulatory approach that has marked the course of high technology development over the past few decades. At present, this shows little sign of changing. The *National Strategy to Secure Cyberspace* is explicit on this point: "[F]ederal regulation will not become a primary means of securing cyberspace."[63] This outlook is subject to change in response to political change as well as the

[60] *See* Gary S. Becker, *Crime and Punishment: An Economic Approach,* 76 J. POL. ECON. 169 (1968). Somewhat predictably, Congress has responded by considering legislation that would lower the threshold for defining certain acts as crimes. For example, H.R. 836 (Feb. 6, 2007), http://thomas.loc.gov/home/gpoxmlc110/h836_ih.xml. would make a series of changes to the federal Computer Fraud and Abuse Act (CFAA), 18 U.S.C. § 1030, including adding an offense for conspiracy to commit any offense defined in the Act and lowering the damage element of these offenses. H.R. 836 §§ 6, 10.

[61] *See* 17 U.S.C. § 1201(k).

[62] *See, e.g.,* the Consumer Broadband and Digital Television Promotion Act of 2002; broadcast flag.

[63] *National Strategy to Secure Cyberspace*, *supra* note 1 at 15.

possibility that an attack would prompt more far-reaching regulation,[64] but such an approach would mark a major shift in the government's approach. Indeed, the government has backed away from leading by example in matters of cybersecurity. The federal government used to publish "The Orange Book,"[65] which set security guidelines for commercially produced computers and software. The idea behind the Orange Book was that the government would buy equipment that met the Orange Book's specifications, and that this adoption would diffuse to businesses and individuals. Instead, the government abandoned the Orange Book:[66]

> [T]he government demanded secure systems, industry produced them, and then government agencies refused to buy them because they were slower and less functional than other nonsecure systems available on the open market.

Government computer and network systems join private sector systems in being highly and almost uniformly vulnerable to attack.[67]

2.3.2 Law Enforcement

Congress' initial response attacks to malicious attacks on networked computers to define new crimes[68] Federal agencies continue to make law enforcement a high priority in cybersecurity policy. Criminal prosecutions of cybercrime, however, face a number of severe limitations. Prosecuting a domestic cybercrime is difficult and expensive. Investigators must secure evidence from far-flung sources that may be unable or unwilling to cooperate without being ordered to do so or served with a search warrant.[69]

[64] One commentator has argued forcefully that cybersecurity might be the "fulcrum" that spurs extensive regulation of technology. *See generally* Zittrain, *The Generative Internet*, *supra* note 4.

[65] Formally, this was known as the "Trusted Computer System Evaluation Criteria."

[66] CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 9.

[67] *See, e.g.,* Ellen Messmer, *GAO Slams FBI Network Security*, PC WORLD, May 25, 2007, http://www.pcworld.com/article/id,132250-c,privacysecurity/article.html; Caron Carlson, *GAO Slams IRS Network Security*, EWEEK.COM, Mar. 27, 2006, http://www.eweek.com/article2/0,1895,1943488,00.asp.

[68] *See, e.g.,* the Computer Fraud and Abuse Act of 1986 (codified at 18 U.S.C. § 1030); Orin S. Kerr, *Cybercrime's Scope: Interpreting "Access" and "Authorization" in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596 (2003).

[69] Legally speaking, the barriers against voluntary disclosures of information to law enforcement officials are due to the ECPA, which I discuss at length in Part 3. I do not,

Like cybersecurity researchers, law enforcement officials face the challenge of understanding rapidly changing cybersecurity threats. Enforcement has slanted heavily toward crimes such as copyright infringement, targeted computer break-ins, and financial fraud.[70] Going higher up the technological chain might be an unappealing enforcement priority because it lacks a direct tie to specific harms, though the FBI has recently directed considerable attention to capturing botnet operators.[71]

The obstacles to prosecuting crimes involving international actors are even more daunting. Treaties address some issues, such as allowing communications data to cross national boundaries. Informal forums have also developed to allow law enforcement officials, as well as private companies, to exchange information about cybersecurity threats.[72] Though these efforts are at an early stage of development, there is widespread recognition that international law enforcement alone will not solve the cybersecurity problem.

2.3.3 Research

Government support for cybersecurity research augments the law enforcement and regulatory approaches. This approach fits the economics of cybersecurity; the market has thus far failed to deliver sufficiently secure computer systems, and neither private network owners nor government agencies have yet to devise adequate means of exchanging information about vulnerabili-

however, spend much time discussing the ECPA's provisions for compelling the disclosure of communications data. For reviews of these provisions, see Orin S. Kerr, *A User's Guide to the Stored Communications Act*, 72 Geo. Wash. L. Rev. 1208 (2004) and Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 Geo. Wash. L. Rev. 1264 (2004).

My statement in the main text is descriptive; I do not mean to suggest that law enforcement officials lack adequate laws to use in prosecutions, or that investigators do not have sufficient means to gather evidence through direct surveillance or through disclosure from third parties. Some members of Congress and the Department of Justice would argue that additional surveillance powers are necessary. A bill recently introduced into Congress, the would "Internet Stopping Adults Facilitating the Exploitation of Today's Youth Act (SAFETY) of 2007," would require Internet service providers to retain data in a manner consistent with regulations issued by the Department of Justice. *See* H.R. 837, (Feb. 6, 2007) § 6, http://thomas.loc.gov/home/gpoxmlc110/h837_ih.xml.

[70] *See, e.g.,* GAO, *Cybercrime*, *supra* note 17, at 15-18.

[71] *See supra* note 32, (describing the FBI's "Operation Bot Roast").

[72] *See, e.g.,* Alberto R. Gonzales, *Prepared Remarks at the Technet Intellectual Property Event*, June 27, 2007, http://www.usdoj.gov/ag/speeches/2007/ag_speech_070627.html (mentioning International Botnet Task Force).

ties and actual attacks. From a regulatory perspective, research represents a prospect for overcoming the general lack of private incentives to develop more secure systems. Thus, research is a long-term component of cybersecurity policy.

Congress has shown a substantial commitment to improving cybersecurity through research by appropriating $377 million for the National Science Foundation (NSF) to grant to cybersecurity researchers.[73] Similarly, federal agencies fund considerable amounts of cybersecurity research. The U.S. Department of Homeland Security, which is responsible for protecting the country against threats to the information infrastructure,[74] has a research projects agency, which has funded research relating to cybersecurity.[75] The Department of Defense funds cybersecurity research through the Defense Advanced Research Projects Agency (DARPA). Another significant source of cybersecurity research funding is the Office of Air Force Research. Even the National Security Agency has a history of transferring technology to the private sector. Finally, researchers at national laboratories conduct cybersecurity research.

Currently, a practical limitation on this research is the need for data to understand adversaries and to test new methods of cyber defense. Researchers who focus on attacks carried out over networks—such as the denial of service and worm attacks described above—have been particularly hard hit.[76] A group of leading computer and network security experts recently wrote that "[c]urrent deficiencies and impediments to evaluating network security mechanisms include . . . [a] lack of relevant and representative network data."[77]

[73] *See* Cyber Security Research and Development Act of 2002, Pub. L. No. 107-305, §§ 4(a)(3), 4(b)(7) [hereinafter "CSRDA"].

[74] DHS now has an Office of Cyber Security and Communications, which "will focus both on cybersecurity and on emergency and interoperable communications, identifying cyber vulnerabilities and threats, and helps protect against and respond to cyber-based attacks, including performing analysis on the potential consequences of a successful attack." Implementation of the Post-Katrina Emergency Management Reform Act And Other Organizational Changes (Jan. 17, 2007), *at* http://www.dhs.gov/xabout/structure/gc_1169243598416.shtm.

[75] *See* Homeland Security Advanced Research Projects Agency (HSARPA) Broad Agency Announcement (BAA) 04-17: Cyber Security Research and Development (CSRD) (last visited Feb. 22, 2007), *at* http://www.hsarpabaa.com/main/BAA0417_solicitation_notice.htm.

[76] In Part 4 I provide a explore in greater detail the link between cybersecurity research objectives and researchers' needs for network data.

[77] R. Bajcsy et al., *Cyber Defense Technology Networking and Evaluation*, 47 Comm.

Other leading researchers have argued that greater access to real network traffic datasets would "cause a paradigmatic shift in computer security research."[78] Some researchers have also adapted their research approaches to reflect their inability to obtain data that provide a sufficiently broad view of network events.[79]

Congress has recognized the central role of sharing data to advancing research, as its appropriation of cybersecurity research funding rested, in part, on a finding that "Federal investment in computer and network security research and development must be significantly increased to . . . better coordinate information sharing and collaboration among industry, government, and academic research projects."[80] The NSF and DHS have also made increasing the availability of network data a critical priority.[81]

In other words, cybersecurity research faces a "data dearth," with network data being the most scarce.[82] Institutions for collecting network data for research purposes, and coordinating researchers' access to the data, are basically non-existent. A leading academic network research group has made an extensive effort to establish a repository of these data, without success. These researchers note that "while technical measurement challenges exist, the non-technical aspects (legal, economic, privacy, ethical) quickly became, and have remained for a decade, the persistent obstacles to progress in this area."[83] Other researchers, who have called for network data repositories to serve cybersecurity researchers in the model of a "cyber Center for Disease Control," also note that this data sharing endeavor "raises potentially immense policy issues concerning privacy."[84]

ACM 59, 59 (March 2004).

[78] Phillip Porras and Vitaly Shmatikov, *Large-Scale Collection and Sanitization of Network Security Data: Risks and Challenges* 1, *in Proceedings of the New Security Paradigms Workshop* (Schloss Dagstuhl, Germany, Sept. 19-22 2006).

[79] *See* Xie et al., *Epidemic Attacks*, *supra* note 8.

[80] CSRDA, *supra* note 73, § 2(5)(C).

[81] *See* CAIDA, Toward Community-Oriented Network Measurement Infrastructure: Project Summary 1-2 (Aug. 2005), *at* http://www.caida.org/funding/cri/nsfcri_2005.pdf [hereinafter CAIDA, *Community Network Measurement*].

[82] Scientists studying software security generally are able to obtain copies of software to study, but discussing the vulnerabilities that they discover raises issues under copyright law and, occasionally, criminal law. I do not discuss this issue further in this Article but note it as a topic for further study.

[83] CAIDA, *Community Network Management supra* note 81, at 1 n.1.

[84] Stuart Staniford, Vern Paxson & Nicholas Weaver, *How to 0wn* [sic] *the Internet in Your Spare Time* 15-17, in *Proceedings of the 11th USENIX Security Symposium (Security*

3 Communications Privacy's Security Model

Cybersecurity researchers are correct to identify communications privacy concerns as central to explaining the cybersecurity data dearth, but neither the explanation nor the solution to the data dearth stops with examining these laws. Federal communications privacy law—the ECPA, primarily— prohibits or limits the acquisition, disclosure, and use of many types of network data that cybersecurity researchers would like to use.[85] But even if privacy laws do not prohibit sources of data, such as Internet service providers (ISPs), from disclosing them to cybersecurity researchers, a variety of institutional factors inhibit data sharing. This Part identifies precisely how communications privacy laws, privacy norms, and the outlooks of institutions that handle communications data fit together.

3.1 Communications Privacy Law

The ECPA has a model of cybersecurity, but it is badly outdated. This model is based on the notion that single firms are best equipped to identify and respond to threats to their own systems. They may, under some circumstances, disclose relevant data to law enforcement agencies to assist in prosecutions. The threats described in Part 2 do not fit this model; they almost always require views from multiple organizations to detect and analyze, and they spread rapidly from one organization to the next. Moreover, the provision for disclosure of communications data for law enforcement purposes, but not for research, is at odds with cybersecurity policy priorities. The ECPA's single-firm view of security is also at odds with the ECPA's core purpose of extending the Fourth Amendment's protection against government surveillance to computer networks.

The ECPA is a notoriously complex set of statutes,[86] which I do not

'02) [hereinafter Staniford et al., *How to 0wn the Internet*].

[85] *See supra* notes 77-78.

[86] *See, e.g.*, Steve Jackson Games, Inc. v. U.S. Secret Serv., 36 F.3d 457, 462 (5th Cir. 1994) (noting that the Wiretap Act is "is famous (if not infamous) for its lack of clarity") (parentheses in original); United States v. Smith, 155 F.3d 1051, 1055 (9th Cir. 1998) (stating that the *Steve Jackson Games* court "might have put the matter too mildly" and continuing: "[T]he intersection of the Wiretap Act and the Stored Communications Act is a complex, often convoluted, area of the law.") (citations omitted); In re Application of the United States, 396 F. Supp. 2d 747, 753 (noting, in the context of an application to install a pen register device, that "rigorous attention must be paid to statutory definitions

attempt to describe fully here. Instead, I sketch the types of data that the ECPA regulates and emphasize how the ECPA's structure relates to the institutional and economic hurdles that stand in the way of access to data for cybersecurity research. The ECPA consists of three titles: amendments to the Wiretap Act, which governs the "interception" of the contents of electronic communications; the Pen/Trap statute statute, which regulates the real-time collection of communications addressing information; and the Stored Communications Act (SCA), which regulates disclosure of electronic communications contents as well as addressing information.[87]

3.1.1 Wiretap Act

The origins of the single-firm view of data privacy began with the enactment in 1968 of the Wiretap Act.[88] The Wiretap Act prohibits *anyone* from intentionally intercepting electronic communications.[89] This prohibition applies to the government, private firms—such as the ISPs that provide individuals and businesses with Internet access—and individuals. The Wiretap Act's breadth reflects an "overriding congressional concern"with protecting the contents of communications from eavesdropping.[90] When it was enacted, the Wiretap Act protected a relatively large proportion of communications between remote parties and addressed the only opportunity to obtain the contents of those communications.[91] Nearly twenty years later Congress extended the Wiretap Act's blanket prohibition on intercepting conversations between two persons to "electronic communications," such as e-mail. Part

when interpreting this complex statute," i.e., the ECPA). *See also* Orin S. Kerr, *Lifting the 'Fog' of Internet Surveillance: How a Suppression Remedy Would Change Computer Crime Law*, 54 *Hastings L. J.* 805 (2003)

[87] See 18 U.S.C. §§ 2702(a)(1)-(2) (prohibiting disclosure of the "contents" of electronic communications by an electronic communications service or a remote computing service); § 2702(a)(3) (prohibiting an ECS or RCS from divulging public shall not knowingly divulge a record or other information pertaining to a subscriber to or customer of such service").

[88] Wire and Electronic Communications Interception and Interception of Oral Communications, Pub. L. No. 90-351, 82 Stat. 197 (1968) (current version at 18 U.S.C. 2510-2522).

[89] 18 U.S.C. § 2511(1)(a).

[90] Gelbard v. United States, 408 U.S. 41, 48 (1972) (citing S. Rep. No. 1097, 90th Cong., 2d Sess., 66 (1968)).

[91] *See* H.R. Rep. No. 99-647, 99th Cong., 2d Sess., 17 ("[T]he contents of a traditional telephone call disappeared once the words transmitted were spoken and there were no records kept."). *See also* S. Rep. No. 99-541, 99th Cong., 2d Sess., 18 (stating that the Wiretap Act addressed "the issue of privacy communications in comprehensive fashion").

of the rationale for the ECPA was to extend statutory privacy protection to communications, such as e-mail, whose constitutional protection was unclear and which did not fall within the Wiretap Act's scope.[92]

Thus, the interception prohibition applies to a broad set of (potential) interceptors,[93] but it is not absolute. In addition to permitting interceptions to proceed under the appropriate search warrants,[94] the Wiretap Act permits interceptions under a few statutory exceptions. One of these exceptions, the "provider exception," allows an employee of an "electronic communication service"[95] to "intercept, disclose, or use" a communications when such activity "is a necessary incident to the rendition of his service or to the protection of the rights or property of the provider of *that service*."[96]

The provider exception embodies a single-firm view of security. The first clause in the exception—"a necessary incident to the rendition of [the employee's] service"—has not been interpreted in the context of electronic communications, but courts have given this part of the provider exception narrow scope where wire communications are concerned.[97] The applicability of the second part of this exception to cybersecurity research is uncertain. There do not appear to be any cases discussing what kinds of threats to a provider's rights or property allow a provider to intercept communications under this exception. Courts discussing this exception have approved its use in the telephone context when a phone company conducts its own investigation into

[92] *See* S. Rep. No. 99-541, 99th Cong., 2d Sess., 3 (stating that "providers of electronic mail create electronic copies of private correspondence for later reference" and citing *United States v. Miller*, 425 U.S. 435 (1976), for doubt concerning the protection of e-mail under the Fourth Amendment). In *United States v. Miller*, the Supreme Court held that a bank customer had no Fourth Amendment interest in protecting his bank records against disclosure to law enforcement officials.

[93] This assertion of the breadth of the Wiretap Act is not meant to take away from criticisms of the Wiretap Act based on the Act's restrictions to the narrowness of the definition of "interception" and the exclusion of certain kinds of surveillance altogether. *See* Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 Geo. Wash. L. Rev. 1280-81 (2004).

[94] *See* 18 U.S.C. §§ 2515-2517.

[95] 18 U.S.C. § 2511(15) (defining this term to mean "any service which provides to users thereof the ability to send or receive wire or electronic communications").

[96] 18 U.S.C. § 2511(2)(a)(i) (emphasis added).

[97] *See* Dep't of Justice, *Searching and Seizing Computers and Obtaining Electronic Evidence in Criminal Investigations*, § IV(D)(3)(c), July 2002, [hereinafter DOJ Search & Seizure Manual].

theft of service or fraud.[98] The provider exception, however, does not allow "unlimited" interceptions, even in the single-firm context;[99] there must be a "substantial nexus" between the monitoring and the threat to the provider's rights or property.[100] In the context of computer networks, the provider exception is broader. Monitoring the network for employee fraud, for example, is within the scope of the exception.[101] The ECPA's legislative history also offers some support for the acceptability of looser limits on monitoring electronic communications.[102]

It is unclear how much room the "substantial nexus" requirement allows for research. One commentator has noted that "there is some tension" between the limited interpretations given to the provider exception and the use of interceptions simply to learn more about attackers' tactics.[103] Although cybersecurity researchers might, in some cases, provide information that allows their employers to protect their networks, this connection is likely to be highly attenuated in many cases. That is, researchers usually seek to develop methods of detecting malicious traffic; their results might not be immediately applicable to that purpose. Researchers who wish to monitor traffic relating to botnets or the intrusion of personal computers owned by an ISP's subscribers—not by the ISP itself—probably are not covered by this part of the provider exception.

Whether the exception applies to disclosure, as opposed to use, for research purposes is even less clear. The predicate for invoking the exception is that the "rights or property *of the provider*" are at risk. Even if a researcher

[98] *See, e.g.,* United States v. Pervaz, 118 F.3d 1, 5 (1st Cir. 1997); United States v. Villanueva, 32 F. Supp. 2d 635, 636 (S.D.N.Y. 1998).

[99] United States v. Auler, 539 F.2d 642, 646 (7th Cir. 1976).

[100] United States v. McLaren, 957 F. Supp. 215, 219 (M.D. Fla. 1997).

[101] United States v. Mullins, 992 F.2d 1472, 1478 (9th Cir. 1993).

[102] *See* H.R. Rep. No. 99-647, 99th Cong., 2d Sess., 47:

> The provider of electronic communications services may have to monitor a stream of transmissions in order properly to route, terminate, and otherwise manage the individual messages it contains. These monitoring functions, which may be necessary to the provision of an electronic communication service, do not involve humans listening in on voice conversations. Accordingly, they are not prohibited.

[103] Richard Salgado, *Legal Issues*, in KNOW YOUR ENEMEY: LEARNING ABOUT SECURITY THREATS (2d. ed.) 225, 230-31 (The Honeynet Project ed.) (2004), http://www.honeynet.org/book/Chp8.pdf.

intercepts electronic communications contents under cover of the provider exception, disclosing the contents to researchers who are not employed by the communications provider might stretch the requirement of protecting the original service provider's rights or property. Unless a provider brings in an outside expert to monitor traffic relating to a threat to the provider's network, it is difficult to find a nexus between allowing interception by an outsider and protecting the provider's property. The argument would have to be that the researcher's analysis would offer protection against a threat; like the in-house researcher, however, an outside researcher's interest is in developing new methods, which may or may not actually be effective in detecting threats against network equipment or services. Instead, the Wiretap Act allows disclosure to law enforcement officials.[104]

In summary, the Wiretap Act does not provide an exception of sufficient scope and generality to allow access the contents of communications for research purposes. Although the provider exception might allow researchers to participate in interception when there is a nexus to protecting communication provider's "rights or property," the scope of this exception has been little clarified by cases. Even if a researcher's interests coincide with a provider's immediate security interests, it is less likely that an outside researcher's interests will. As a result, the need of cybersecurity researchers to validate techniques under similar conditions will go unmet under the present Wiretap Act.

3.1.2 Stored Communications Act (SCA)

A more permissive statutory scheme, the Stored Communications Act (SCA),[105] applies to accessing communications, or records about communications, that are in storage rather than in transit from source to destination. In contrast to the Wiretap Act, the SCA permits nearly unrestricted use—including cybersecurity research, but far from limited to it—of communications contents and records within a service provider.[106] Disclosures of these data to per-

[104] *See* United States v. Villanueva, 32 F. Supp. 2d at 636. Note, however, that law enforcement official may not *direct* the employees of a service provider to monitor a network unless they have a warrant.

[105] 18 U.S.C. § 2701-2712.

[106] The SCA applies only to "electronic communication services" (ECS) and "remote computing services" (RCS). *See* 18 U.S.C. § 2510(15) (defining "electronic communication service" to mean "any service which provides to users thereof the ability to send or receive wire or electronic communications"); *id.* § 2711(2) (defining "remote computing service"

sons outside the service provider, however, may be regulated by the SCA. The extent of regulation depends on two factors: whether the recipient is a "governmental entity," and whether the provider discloses communications contents or "noncontent" records.

The SCA offers greater protection to communication contents than to noncontent information. The content of an electronic communication means "any information concerning the substance, purport, or meaning of that communication."[107] Noncontent information includes records pertaining to a subscriber or user of an electronic communications service.[108] Logs of IP addresses that a user has reached, as well as the "to" and "from" fields in e-mail records are also considered noncontent records.[109] Whether other records, such as textual Web addresses (URLs) that contain search engine queries, are content or noncontent records is still a matter of debate.[110]

The SCA prohibits the voluntary disclosure of communications contents by a service provider to any other person,[111] unless an exception—including a provider security exception—applies.[112] Like the Wiretap Act, the SCA contains a "provider exception" that permits some disclosure of communications contents when necessary to protect the "rights or property" of the provider.[113] There are few, if any, cases interpreting this exception. Still,

to mean "the provision to the public of computer storage or processing services by means of an electronic communications system"). The ECS category is further refined by services that are open to the public and those that are not. *See id.* § 2702 (regulating voluntary disclosure of communications by "an electronic communications service to the public" only). For simplicity the main discussion, I am concerned only with an ECS and use this term interchangeably with "service provider," unless otherwise noted.

[107] 18 U.S.C. § 2510(8).

[108] 18 U.S.C. § 2702(a)(3).

[109] *See* United States v. Forrester, No. 05-50410, slip op. at 13-16 (9th Cir., July 6, 2007) (so holding in the context of the Pen/Trap statute statute). *See also infra* notes 128-129

[110] *See* Brief of Amici Curiae Law Professors Requesting Additional Briefing If This Court Addresses Google's ECPA Defense 5, Gonzales v. Google, Inc., 5-06-mc-80006-JW, Feb. 24, 2006 (arguing that "[t]here is no case law on the ECPA's application to URLs or search queries stored by a company that provides electronic communications service").

[111] 18 U.S.C. § 2702(a).

[112] Because the SCA' s voluntary disclosure provisions are far more important for relating the ECPA to cybersecurity research, I do not discuss details of the SCA's compelled disclosure provisions. For an exposition and analysis of those provisions, see Orin S. Kerr, *A User's Guide to the Stored Communications Act, and a Legislator's Guide for Amending It*, 72 Geo. Wash. L. Rev. 1208 (2004) [hereinafter Kerr, *User's Guide*.

[113] 18 U.S.C. § 2702(b)(5) (creating a provider protection exception for disclosure of electronic communication contents).

the close similarity between it and the Wiretap Act's provider exception suggests a similar purpose and scope:[114] to allow service providers to monitor their systems for threats to their own "rights or property." This perspective assumes that a single firm will be able to detect relevant threats against it; and, as with the analogous exception in the Wiretap Act, the SCA's provider exception envisions disclosure to law enforcement agencies as the remedy for such threats.[115] This single-firm outlook is consistent with the SCA's lack of restrictions on data retention. The SCA does not restrict the internal use of communications by an electronic communication services firm; the SCA focuses solely on disclosure, whether voluntary, compelled, or resulting from some kind of breach in service provider's access controls.[116]

Restrictions on voluntarily disclosing noncontent records are far looser than those pertaining to communications contents, owing to two limitations in the SCA. First, only an entity that provides service "to the public" is covered at all by this part of the SCA. Second, even if a service provider falls under these voluntary noncontent record disclosure restrictions, it may provide records to any recipient other than a "governmental entity."[117] Thus, for cybersecurity researchers, the definition of a "governmental entity" is critical. There is no definition of this phrase in the ECPA or, for that mat-

[114] *See* Kerr, *User's Guide*, *supra* note 112, n.92.

[115] *See* DOJ Manual, *supra* note 97, App. G (stating, in a sample letter from a service provider to a law enforcement agency, that the provider is permitted to disclose communications contents and noncontent records to government agents "if such disclosure protects the [Provider]'s rights and property").

[116] *See* 18 U.S.C. §§ 2702-2710 (defining conditions and process for voluntary and compelled disclosure of communications contents and noncontent records). The other provision of the SCA mentioned in the main text prohibits a person from accessing, or exceeding authorized access to, an electronic communications service facility and "obtain[ing], alter[ing], or prevent[ing] authorized access to a[n] . . . electronic communication while it is in electronic storage in such system." 18 U.S.C. § 2701(a). In this case it is the person who obtains unauthorized access to a stored communication, rather than the communications service provider, that is liable for a violation of the SCA. Indeed, the provider of the electronic communications service, however, may access stored communications. 18 U.S.C. § 2701(c)(1). As one court has noted, this is a "provider exception," but its "breadth presents a striking contrast to the Wiretap Act's own, much narrower provider exception." United States v. Councilman, 418 F.3d 67, 82 (1st Cir. 2005) (en banc). Furthermore, the user of the service may authorize access to his or her stored communications. 18 U.S.C. § 2701(c)(2). Courts have required little in the way of formality to find consent on the user's part. *See, e.g.,* Konop v. Hawaiian Airlines, Inc., 302 F.3d 868, 880 (9th Cir. 2002); In re DoubleClick Inc. Privacy Litig., 154 F. Supp. 2d 497 (S.D.N.Y. 2001).

[117] 18 U.S.C. § 2702(a)(3); *id.* § 2702(c)(6).

ter, in Title 18 of the United States Code; but courts have given it broad effect. The Second Circuit, for example, suggests that the breadth of "governmental entity" is similar to that of "person" as defined in the ECPA,[118] namely "any employee, or agent of the United States or any State or political subdivision thereof, and any individual, partnership, association, joint stock company, trust, or corporation."[119] Moreover, the ECPA provides a definition of "investigative or law enforcement officer," [120] which refers specifically to a person who is empowered to investigate violations of the law or to make arrests. Putting these pieces together suggests that "governmental entity" would cover researchers at state universities and federal research laboratories. Since this describes many cybersecurity researchers, this distinction in the SCA carries considerable consequences for cybersecurity research. A service provider, such as a commercial ISP, may share noncontent data with a researcher from a private university, but sharing the same data with a researcher from a public university—a potential "governmental entity"—raises a serious question under the SCA.

This division displays a lineage rooted in the Fourth Amendment, which limits only governmental intrusions into privacy.[121] The difficulty that the

[118] *See* Organizacion JD Ltda. v. United States Dep't of Justice, 18 F.3d 91, 94-95 (2d Cir. 1994) (quoting S. Rep. No. 541, 99th Cong., 2d Sess. 43 (1986), reprinted in 1986 U.S.C.C.A.N. 3555, 3597, which states that "an aggrieved party 'may recover from any person or entity—including governmental entities—who knowingly or intentionally violated this chapter [Chapter 119 of Title 18, i.e., the Wiretap Act]"').

[119] 18 U.S.C. § 2510(6).

[120] 18 U.S.C. § 2510(7) ("'Investigative or law enforcement officer' means any officer of the United States or of a State or political subdivision thereof, who is empowered by law to conduct investigations of or to make arrests for offenses enumerated in this chapter, and any attorney authorized by law to prosecute or participate in the prosecution of such offenses.")

[121] U.S. Const. amend. IV. The House Report issued in connection with the ECPA is quite explicit about the underlying, Fourth Amendment-based model of privacy:

> When the Framers of the Constitution acted to guard against the arbitrary use of government power to maintain surveillance over citizens, there were limited methods of intrusion in the "houses, papers and effects" protected by the Fourth Amendment. During the intervening 200 years, development of new methods of communication and devices for surveillance has expanded dramatically the opportunity for such intrusions.

H.R. Rep. No. 99-647, 99th Cong., 2d Sess., 16. The House Report continues: "Privacy cannot be left to depend solely on physical protection. Additional legal protection is necessary to ensure the continued vitality of the Fourth Amendment." *Id.* at 19.

SCA's provider exception creates for cybersecurity research, however, is that the "governmental entity" category covers much more than law enforcement agencies and personnel; but, somewhat paradoxically, the SCA permits voluntary disclosure of records to protect the "rights or property" of the provider.[122] As I argued above, law enforcement officials are most likely the permissible recipients under these circumstances. Most cybersecurity researchers who are employed by governmental entities are unlikely to have law enforcement duties, which would make disclosing communications to them under the provider exception somewhat suspect. On the other hand, cybersecurity researchers are unlikely to *use* these communications data differently than would a researcher within the service provider firm. The result is that the SCA provides an exception that accommodates law enforcement but thwarts data disclosure for other uses, even though those uses may occur within the organizational boundaries of a service provider.

3.1.3 Pen/Trap Statute

The third and final title of the ECPA is the Pen/Trap statute statute,[123] which is the non-content counterpart to the Wiretap Act. The statutes name refers to devices that collect incoming addressing information (pen registers) and outgoing addressing information (trap and trace devices).[124]

The Pen/Trap statute statute regulates the real-time collection of communications addressing information.[125] The statute generally prohibits any person from installing or using a device that collects addressing information in real time, though law enforcement officers may do so if they obtain a court order.[126] As stated above,[127] addressing information includes essentially all noncontent information about a particular communication, such as IP ad-

[122] 18 U.S.C. § 2702(c)(3).

[123] 18 U.S.C. § 3121-3127.

[124] *Id.* §§ 3127(3)-4 (defining "pen register" and "trap and trace" device, respectively). Since both devices are regulated in the same way by the statute, it is often referred to as "Pen/Trap statute." *See also* DOJ Search & Seizure Manual, *supra* note 97, § IV(C)(1).

[125] 18 U.S.C. § 3121(a) (prohibiting any person from installing a "device or process which records or decodes dialing, routing, addressing, or signaling information transmitted by an instrument or facility from which a wire or electronic communication is transmitted").

[126] 18 U.S.C. § 3121-3123. *See also* Brown v. Waddell, 50 F.3d 285 (4th Cir. 1995).

[127] *See* supra note 109 and accompanying text.

dresses[128] and the "to" and "from" fields in e-mail messages.[129] It is unclear whether uniform resource locators (URLs)—the addresses that most Internet users use to connect to Web sites—are addressing information or contents.[130]

The Pen/Trap statute statute follows the Wiretap Act's approach of applying to all persons and then creating exceptions for a service provider's internal use as well as limited government access to addressing information. The Pen/Trap statute statute permits a service provider to collect addressing information in the ordinary course of business.[131] In addition, the government may obtain a court order allowing it to install a pen register by certifying that the addressing information it would obtain is "relevant to an ongoing criminal investigation."[132]

The security model of the Pen/Trap statute statute is difficult to discern. The statute authorizes service providers to install pen registers to protect their users from abuse or to protect the provider's rights or property.[133] Like the provider exceptions in the Wiretap Act and the SCA, this exception makes the security of a single firm—the service provider—the condition for triggering the exception. The Pen/Trap statute's exception, however, is concerned only with the condition for allowing a service provider to install a pen register; the statute lacks a corresponding disclosure provision.[134]

[128] *See generally* In re Application of the United States, 396 F. Supp. 2d 45, 48-49 (D. Mass. 2005) (regarding IP addresses as addressing information, not contents).

[129] *See* In re Application of United States for an Order Authorizing the Installation and Use of a Pen Register and a Trap and Trace Device on E-Mail Account, 416 F. Supp. 2d 13, 18 (D.D.C. 2006) (so holding but not specifying which fields the ISP could disclose to the government). One commentator has concluded that addressing information includes the "To: and "From:" fields in an e-mail message. Daniel J. Solove, *Reshaping the Framework: Electronic Surveillance Law*, 72 Geo. Wash. L. Rev. 1264, 1287 (2004).

[130] Indeed, classification of URLs might depend on the particular URL. *See* In re Application of the United States, 396 F. Supp. 2d 45, 48-49 (D. Mass. 2005) (reasoning that URLs that contain search terms "would reveal content" and requiring a trap-and-trace order to list data that the recipient Internet service provider would be prohibited from disclosing). *See also* 110.

[131] *See* 18 U.S.C. § 3121(b)(1) (permitting an electronic communication service to install a pen register or trap and trace device in a context "relating to the operation, maintenance, and testing of a wire or electronic communication service or to the protection of the rights or property of such provider").

[132] 18 U.S.C. § 3122-3134. Thus, the necessary showing for obtaining a pen register order is much lower than it is for obtaining a wiretap warrant.

[133] 18 U.S.C. § 3121(b)(1).

[134] *See* 18 U.S.C. § 3121(b)(1) (permitting installation of a pen/trap device to "to the protec[t] . . . the rights or property of [the] provider, or to [protect] of users of that service

In the absence of a disclosure standard, the Pen/Trap statute provides little guidance about whether, and under what conditions, it is permissible disclose addressing information to cybersecurity researchers. One possible standard is that any recorded addressing information becomes a noncontent record subject to the disclosure provisions of the SCA.[135] In that case, a service provider may voluntarily disclose the addressing information to law enforcement officials to protect its "rights or property."[136] But, as argued above, this restriction probably does not permit disclosure to cybersecurity researchers affiliated with a governmental entity.[137] Internal research use of the data, for cybersecurity purposes or otherwise, would be permissible, by analogy with internal use of noncontent records under the SCA.[138]

An alternative disclosure standard would hold that the Pen/Trap statute, by failing to prohibit disclosure, implicitly authorizes *any* disclosure of addressing information by a service provider, so long as the provider collected the information in a manner consistent with one of the Pen/Trap statute statute's exceptions permitting pen register use. However, these exceptions are triggered by concerns far broader than provider security; any collection of addressing information "relating to the operation, maintenance, and testing"[139] would suffice. This reading of the statute would effectively gut the noncontent provisions of the SCA. The creation of the Pen/Trap statute statute and the SCA through the same act of Congress makes this interpretation unlikely.

The ECPA presents real difficulties for the endeavor of disclosing data for cybersecurity research. The ECPA's support for cybersecurity research resides in provider exceptions that are most likely useful for self-defense;

from abuse of service or unlawful use of service"); *id.* § 3121(b)(2) (permitting pen/trap installation by a provider "to record the fact that a wire or electronic communication was initiated or completed in order to protect such provider, another provider furnishing service toward the completion of the wire communication, or a user of that service, from fraudulent, unlawful or abusive use of service").

[135] As discussed above, a service provider's records of its customers Internet usage are likely within the SCA's definition of noncontent records.

[136] 18 U.S.C. § 2702(c)(3). Note that a service provider that does not provide service to the public may disclose noncontent records to a governmental entity—a category that encompasses far more than law enforcement agencies—even if the disclosure would not meet the requirements of § 2702(c)(3).

[137] *See* the discussion of "governmental entity," *supra* pages 24-28.

[138] *See id.*

[139] 18 U.S.C. § 3121(b)(1).

these exceptions simply do not envision that sharing network data could advance the cause of a single firm's security. Moreover, the applicability of these exceptions to research depends on whether the research is applicable to the service provider's operational security. Finally, Government-affiliated researchers are at a distinct disadvantage relative to privately employed researchers; the distinction is particularly sharp where stored noncontent records are at issue. As I explain in Part 4, however, this category covers data that are the most likely to be useful to cybersecurity researchers.

3.1.4 State Laws

State privacy statutes and common law have the potential to complicate further the question of cybersecurity researchers' access to communications data. Most states have adopted their own versions of the federal Wiretap Act.[140] Though most of these statutes offer approximately the same level of protection as the Wiretap Act for communications,[141] some are more protective than the federal statute.[142] California, for example, requires that all parties to a communication consent to its interception,[143] where as the Wiretap Act provides a one-party consent rule.[144]

As a practical matter, state laws that deviate to the more protective side of communications privacy have the potential to raise further the costs of assembling cybersecurity datasets, or to prevent disclosure of data where federal law might allow it.[145] The greatest impact of state communications

[140] *See* Daniel R. Dinger, *Should Parents Be Allowed to Record a Child's Telephone Conversations When They Believe the Child Is in Danger?: An Examination of the Federal Wiretap Statute and the Doctrine of Vicarious Consent in the Context of a Criminal Prosecution*, 28 Seattle U. L. Rev. 955, 965-68 & n.58 (2005) (noting that Vermont is the only state that has not adopted a statutory equivalent of the Wiretap Act) [hereinafter Dinger, *Parental Wiretaps*].

[141] *Id.* at 966-67 & nn.65-67.

[142] Several state courts and at least one federal court have found that state wiretap statutes must be at least as protective. *See id.* (citing cases from Massachusetts and California, and United States v. Mora, 821 F.2d 860 (1st Cir. 1987)).

[143] Cal. Penal Code § 632(a) (defining an offense for intercepting "intentionally and without the consent of *all* parties to a confidential communication") (emphasis added).

[144] 18 U.S.C. §§ 2511(2)(c)-(d).

[145] *See* Lawrence O. Gostin & James G. Hodge, Jr., *Personal Privacy and Common Goods: A Framework for Balancing Under the National Health Information Privacy Rule*, 86 Minn. L. Rev. 1439, 1465-66 (discussing the effects upon public health and medical research of the lack of federal preemption in the context of health information disclosure

privacy laws is on the question of defining researcher's access to cybersecurity data arises when a state has a law that is more restrictive than federal law. In Part 5 I discuss ways to address the complications of state privacy law upon cybersecurity research.

3.1.5 Gaps

The gaps in the ECPA are as important as its positive protections for establishing the baseline of the current state of communications privacy in the cybersecurity research context. The ECPA leaves two significant gaps. First, retention and internal use of data by a firm that controls them are essentially unregulated. Second, courts have interpreted the ECPA's consent provisions broadly in favor of finding consent. Both of these gaps potentially mean that many users have already agreed to allow service providers to use their data for cybersecurity research, though it is unlikely this is the only use that providers make of these data. This situation leaves a large gap between industry (and academic) norms for data retention and use, on one hand, and users' understanding of those practices. An illustrative example comes from Google's recent announcement that it would limit its retention of individuals' search histories to 18 months.[146] This announcement actually seemed to serve as public notice of how extensively Google retains data.[147] Other search engines, e-mail providers, and ISPs have not followed Google's lead.

Two recent cases illustrate the point about the lack of controls on retention and internal use. When the Recording Industry Association of America (RIAA) in July 2002 began its campaign to sue users of peer-to-peer file sharing services, it issued a subpoena to Verizon Internet Services, demanding that Verizon disclose the names and other identifying information for the customer assigned a particular network address on a particular day and time.[148] Although Verizon fought the subpoena on a number of grounds,

rules) [hereinafter Gostin & Hodge, *Personal Privacy and Common Goods*].

[146] *Google Reduces Search Log Retention to 18 months*, http://googleblog.blogspot.com/2007/06/how-long-should-google-remember.html, June 12, 2007.

[147] *See* FT.COM, http://www.ft.com/cms/s/dc89ec96-0a24-11dc-93ae-000b5df10621.html, May 24, 2007 (reporting, after Google's retention policy change, that "European data protection officials have raised concerns that Google could be contravening European privacy laws by keeping data on internet searches for too long").

[148] *See* Recording Indus. Ass'n of Am. v. Verizon Internet Servs., 240 F. Supp. 2d 24, 28 (D.D.C. 2003), *remanded by* 351 F.3d 1229 (D.C. Cir. 2003).

Verizon did not argue that it did not have the information that the RIAA sought.[149] Retaining this information is consistent with Verizon's privacy policy (at least in its current form); the point of this example is simply to illustrate that ISPs retain, for at least several months, sufficient information to link an IP address to an individual subscriber.

A second example involves search engine data retention. As part of its defense[150] of the Child Online Protection Act,[151] the U.S. Department of Justice in August 2005 issued subpoenas to several major search engines, including Google. The government sought from Google "'[a]ll queries that have been entered on your company's search engine between June 1, 2005 and July 31, 2005 inclusive,"' among other things.[152] Google moved to quash this subpoena, and the grounds for its motion are instructive for building a picture of the kinds of data that Google retains. Google's arguments in opposition to enforcing the subpoena were that the search query data would not lead to admissible evidence,[153] that the search queries are trade secrets,[154] and that complying with the subpoena would impose an undue burden on Google.[155] Google admitted, however, that the queries were available.[156] Moreover, Google's memorandum indicates that Google performs some analysis of the search queries that it keeps, though it left the kinds of analysis unspecified.[157] It is unclear whether the other search engines that received subpoenas in

[149] *See generally id. See also* Charter Comms., Inc., Memorandum in Support of Motion to Quash Subpoena Served by Recording Indus. Ass'n of Am., No. 4:03MC00273CEJ (E.D. Mo., Oct. 3, 2003) (not arguing that Charter did not have the information necessary to comply with the RIAA's subpoena for personal identifying information linked to an IP address).

[150] The case is *ACLU v. Gonzales*, 478 F. Supp. 2d 775 (E.D. Pa. 2007).

[151] 47 U.S.C. § 231).

[152] Gonzales v. Google, Inc., 2006 U.S. Dist. LEXIS 13412, **7 (N.D. Cal 2006) (quoting from page 4 of the subpoena issued to Google). The government also demanded a list of all URLs reachable by queries to Google's search engine, as of July 31, 2005.

[153] Google's Opposition to the Government's Motion to Compel 6, No. 5:06-mc-80006-JW (N.D. Cal., Mar. 13, 2006).

[154] *Id.* at 10.

[155] *Id.* at 15-17.

[156] *Id.* at 16.

[157] *See id.* at 11 ("Access to Google's internal systems, and, in particular, Google's query log and index are each restricted to a small group of trusted employees with special clearance based, in part, on the length of their employment and demonstrated need for access."). Interestingly, Google noted that "[u]nequivocally, it is and has been Google's policy for years not to share any such information [reachable URLs and search queries] with third parties." *Id.* at 11 n.2.

this case—AOL, Microsoft, and Yahoo—have similar practices. All three complied with the DOJ's subpoenas without creating a public record of their data retention practices, except to the extent revealed by their compliance.[158]

Similarly, the ECPA's consent provisions provide broad leeway for cybersecurity research within a single firm, though not necessarily for disclosure to outside cybersecurity researchers. The typical means of securing consent is in the provider's terms of service, which often include, or incorporate by reference, a privacy policy. Consent provisions may be in the middle of extensive terms of service agreements posted online; courts do not require specific acknowledgment of a consent provision. Courts have held, for example, that establishing that consent to an interception under the Wiretap Act was invalid requires proving that the party seeking consent acted primarily out of motivation to commit a tort or crime.[159]

To which activities consent extends is often unclear; privacy policies of all communication service providers—whether or not they offer service to the public—serve as a wildcard for cybersecurity researchers and individual users. The privacy policies of prominent ISPs and e-mail providers, which are subject to the SCA, contain broad clauses that effect a user's consent to share his or her communications information with an ambiguous group of "partners" and "affiliates."[160]

Major ISPs obtain user consent to collect information about Internet usage for network performance engineering and, in at least one case, for research. Another ISP "store[s] e-mail messages and video mail messages [sent and received by its users] on computer systems for a period of time."[161] To

[158] *See* Gonzales v. Google, Inc., 2006 U.S. Dist. LEXIS 13412, **7 (N.D. Cal 2006) ("The subpoena required that these companies produce a designated listing of the URLs which

would be available to a user of their services. The subpoena also required the companies [AOL, Google, Microsoft, and Yahoo] to produce the text of users' search queries. AOL, Yahoo, and Microsoft appear to be producing data pursuant to the Government's request.").

[159] In re DoubleClick Inc. Privacy Litig., 154 F. Supp. 2d 497 (S.D.N.Y. 2001) (citing cases from the D.C. Circuit and First Circuit).

[160] *See, e.g.*, Microsoft Online Privacy Statement (Jan. 2006), http://privacy.microsoft.com/en-us/fullnotice.aspx. *Cf.* Panix and Privacy (Oct. 14, 2005), http://www.panix.com/panix/privacy.html ("Panix will not give out your real-world information (address, phone number, even your real name if you choose to withhold it) to anyone, at any time, for any reason short of legal obligation.").

[161] Comcast High-Speed Internet Privacy Statement (rev. Jan. 2006), *at* http://www.comcast.net/privacy/index.jsp ("Information Use and Disclosure" sec-

take an example from an academic environment, the University of California, Berkeley collects and stores transactional records pertaining to communications between users of Berkeley's network and addresses on the Internet.[162] Berkeley stores "raw" data identifiable to specific IP addresses for one month at most, unless "a privacy filter is applied to the data."[163] The university may store anonymized network usage data indefinitely.[164] Appropriate staff may review these data "to understand the volume and characteristics . . . of the traffic flowing through various points in the network."[165] In addition, university-wide policy provides that "[n]etwork traffic may be inspected to confirm malicious or unauthorized activity that may harm the campus network or devices connected to the network."[166]

3.2 Institutions

The second part of my explanation for the cybersecurity data dearth is institutional. Relevant data are widely scattered among public and private actors. There is no overarching organizational mechanism—least of all the government—to encourage, let alone compel, those actors to disclose data to cybersecurity researchers. The problem of data collection now involves "tapping into thousands of data sources effectively and sharing critical information—intelligently and to the data owners' satisfaction."[167] Single firm dynamics also contribute to the dearth. Even if it is legally permissible for a firm to disclose data to a cybersecurity researcher, the firm is often un-

tion).

[162] *See* Cliff Frost, CNS Data Collection and Retention: Current Practice 1-2, Oct. 25, 2002, *at* http://cns.berkeley.edu/dept/CNS%20Data%20Collection%20and%20Retention.doc (setting forth "Netflow Data" retention policy).

[163] *Id.*

[164] *Id.*

[165] *Id.*

[166] University of California Office of the President, Electronic Communications Policy § V.B (rev. Aug. 18, 2005), *at* http://www.ucop.edu/ucophome/policies/ec/html/pp081805ecp.html.

[167] Slagell & Yurick, *Sharing Computer Logs*, *supra* note 10, at 1. This was not always the case. Until April 1995, the NSF operated the Internet's "backbone"—the networking equipment that connects separate institutions (e.g., universities) over long distances. During this time the NSF regularly provided network data to researchers. *See* CAIDA, *Community Network Measurement*, *supra* note 81 at 1. *See also infra* Part 3.2 for a discussion of the current data needs of cybersecurity researchers.

willing to do so for a variety of reasons: disclosures create a risk of customer backlash; the firm fears further, unauthorized disclosure; assembling datasets and vetting the recipients is a cost with little prospect of reward; and internal use of data provides firms with a competitive edge in the market for research talent. In summary, there are few institutional forces to promote sharing of cybersecurity-relevant data, and there are few incentives for network service providers to promote the notion that sharing data in support of cybersecurity provides a public benefit.

A symptom of this situation is in privacy policies. Few of the privacy policies that I reviewed mention research explicitly. Instead, they tend to cast communications data collection and sharing in terms of the benefits of improved service and the chance to receive more fully informed solicitations from business partners. There is no legal reason that a provider could not state that it requires consent from its users to share their communications-related data in order to advance cyber security research. Privacy policies tend to do the opposite, however, by obfuscating the provider's actual data retention and handling practices. Perhaps the benefit that might arise from the research facilitated by this kind of data sharing is too intangible and indirect to be palatable to these services' users. At the same time, there are few indications that the widespread use of consent in communication service providers' terms of use and privacy policies has facilitated effective means for cybersecurity researchers to obtain access to data.[168]

Cybersecurity research policies at universities—potentially promising sources of network data—are similarly difficult to penetrate. Although it is difficult to generalize, universities tend to offer strong privacy protection to their faculty and students.[169] Based on unstructured interviews I conducted with

[168] Cybersecurity researchers have stated that greater access to real network traffic datasets would "cause a paradigmatic shift in computer security research." Phillip Porras and Vitaly Shmatikov, *Large-Scale Collection and Sanitization of Network Security Data: Risks and Challenges* 1, *in Proceedings of the New Security Paradigms Workshop* (Schloss Dagstuhl, Germany, Sept. 19-22 2006). But, as others have noted, "while the data needed exists, tapping into thousands of data sources effectively and sharing critical information—intelligently and to the data owners' satisfaction—is an open problem." Adam Slagell and William Yurick, *Sharing Computer Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization* 1.

[169] For example, a number of universities have recently announced that they would strictly limit their cooperation with requests to disclose personally identifying information about their students in connection with the recording industry's investigations in to alleged copyright infringement. Though this context is far different from disclosing data for cy-

researchers who work at a number of universities, I have found that they face significant challenges in obtaining access to data from their own institutions. These challenges are even more severe when researchers seek access to datasets that the researchers wish to retain.

A final element of this picture is the role that access to data plays in competition among communications service providers. Many of these firms maintain research operations.[170] In a world in which access to network data is highly constrained, the ability of a firm to offer its researchers access to it could form a critical basis for making that firm an attractive place to work. This consideration might make firms reluctant to share data, even if it is legally permissible for them to do so.

4 Coping with the Cybersecurity Data Dearth

4.1 Scientific Goals of Data Sharing

In seeking to share data for cybersecurity research, researchers act not only out of a desire to advance their own research, but also to make the data sharing scheme consistent with broader scientific goals. Some of these goals criteria are in tension with others. Nonetheless, they are worth keeping in mind, both as background for the following descriptions of available cybersecurity data, as well as for evaluating the legal and institutional proposal that I make in Part 5.

Cybersecurity researchers have called for making access to cybersecurity data as broad as possible.[171] Broad access to data would ensure that the element of luck that is sometimes involved in obtaining data is removed. This condition would also allow many researchers to examine the same dataset, allowing more analysis than any single researcher could perform. Where unrestricted access is not feasible, researchers would like the means for controlling access to the data to accommodate as many researchers as possible.

Second is the condition of utility, which counsels that cybersecurity data should be made available in as "raw" a form as possible.[172] Scrambling or

bersecurity research, it does illustrate a strong concern for students' privacy.

[170] AT&T, Google, and Microsoft all have large and prestigious research divisions.

[171] *See* Vitaly Shmatikov, *Threats to Anonymized Datasets* 4, *at* http://www.cyber.st.dhs.gov/public/PREDICT/Vitaly-athreats1.pdf (Sept. 27, 2005).

[172] *See* Vitaly Shmatikov, *Threats to Anonymized Datasets* 4, *at* http://www.cyber.st.dhs.gov/public/PREDICT/Vitaly-athreats1.pdf (Sept. 27, 2005).

anonymizing data degrades their usefulness to researchers, and in some cases this kind of processing can render data unfit for a specific research use.

Third, cybersecurity researchers hold that data should be persistently available.[173] Persistent datasets not only facilitate the evaluation of published research but also would allow cybersecurity researchers to examine network trends over time.

A fourth criterion for cybersecurity data is that it should be diverse. Different research questions require different kinds of data. Just as importantly, different kinds of data implicate different legal and policy questions. Some data raise difficult questions about protecting individual privacy, while others create security risks for the firms that provide them.[174] Finally, cybersecurity researchers strongly recognize the need to protect the privacy of individuals whose activities are represented in communications data, as well as the potential for shared data to aid an attacker who targets the data source.[175] Furthermore, cybersecurity researchers recognize that policy considerations—at the institutional level or beyond—must inform the decision of *what* data to anonymize, if any; technology can only answers the question of *how* to anonymize selected aspects of data.

4.2 Data Needs: A Picture of the Ideal

To have a more concrete picture of cybersecurity research approaches and data needs, consider again the cyber attack against Estonia, which I discussed in Part 2. This attack was an example of a DDoS attack: traffic from many hosts on the Internet flooded networks connections between Estonia and the rest of the world. Understanding this kind of attack is a high priority for researchers because it takes advantage of the basic end-to-end architecture the Internet: the network equipment that routes traffic to a destination (i.e., Estonia) does not examine whether that traffic is malicious, or whether the

[173] *See* CAIDA InfoCat (maintaining comprehensive index of publicly available network datasets); *see also* Ruoming Pang et al. *The Devil and Packet Trace Anonymization* (2006) (making anonymized datasets available on the Web).

[174] *See* Douglas Maughan, PREDICT Overview 17, (Sept. 27, 2005) *at* http://www.cyber.st.dhs.gov/public/PREDICT/PREDICT%20-%20Workshop%20-%20Sep2005%20-%20Maughan.pdf (describing different kinds of data needed for cybersecurity research).

[175] *See* Slagell & Yurick, *Sharing Network Logs*, *supra* note 10, at 1 (noting that sharing network data creates the risks of being "mishandled by friendly peers, or fall[ing] directly/indirectly into the hands of malicious cracker").

recipient's network is too clogged to accept more data. Rather, the equipment that runs the basic protocols of the Internet keeps trying to deliver data to a destination if it fails the first time. In other words, a DDoS attack exploits a vulnerability that is a basic feature of the Internet.[176] Cybersecurity researchers study DDoS attacks from a number of different angles. Some have focused on real-time detection of attacks,[177] while others have focused on analyzing attacks after they occur.[178] In order to test either approach, researchers need to correlate data from the many different sources that direct traffic to the attack target. That is, electronic communications data from many separate sources is necessary to determine whether a detection algorithm correctly distinguishes attack traffic from innocuous communications.

Or consider the "Witty" worm that I discussed in Part 2. One objective in the face of such threats is to gain an early warning about the early phases of an attack, so that a network operator can take steps to stop the worm from spreading, thus mitigating network congestion or the destruction of data.[179] Another objective is to reconstruct the path of a worm after an attack in order to understand how it behaved, as well as to repair damage that the attack might have done. Both objectives remain topics of active research, and both require large volumes of electronic communications data from many separate organizations in order to be effectively tested.[180]

Most researchers state that the contents of these communications are not of interest to them. Instead, it is the addressing information—where a packet of information comes from, and where it is going—that provides the critical information.[181] Ideally, researchers would have access to addressing data from multiple entities, such as ISPs, in order to test these methods.[182]

[176] As I described in Part 2, many DDoS attacks are launched from botnets, which tend to form because attackers can exploit software vulnerabilities to gain control of many computers. This approach is not necessary to running a DDoS attack; any network of attack computers under central command-and-control—perhaps a state power—could serve to launch a DDoS attack.

[177] Xie et al., *Epidemic Attacks*, *supra* note 8.

[178] Allman et al., *Fighting Cooridinated Attackers*, *supra* note 28.

[179] Xie et al. *Epidemic Attacks*, *supra* note 8.

[180] *See id.*

[181] As discussed in Part 3, addressing information receives less protection than contents under the ECPA. As I discuss later in this part, however, some important areas of cybersecurity research would greatly benefit from access to communications contents.

[182] *See* Xie et al. *Epidemic Attacks*, *supra* note 8, at 44.

4.3 Public Releases

What kinds of data are actually available to study these problems? ISP data are not available to cybersecurity researchers,[183] unless the researchers happen to work for an ISP.[184] Publicly available data fall roughly into the categories that the ECPA treats as fundamental: non-content data and communications contents. There is far more publicly available non-content data, but even these data impose significant limitations on their utility for cybersecurity researchers.

4.3.1 Non-Content Data

The most significant public release of non-content data occurred in 2006, when researchers affiliated with Lawrence Berkeley National Laboratory and a non-profit research institute placed approximately 11 gigabytes of anonymized data on the Web. In doing so, the researchers noted that "[s]haring of network measurement data . . . has been repeatedly identified as critical for solid networking research."[185] This set of "packet traces" included only the "packet headers," which are generally considered to be equivalent to the addressing information of the ECPA. Moreover, the dataset contains the IP addresses of the sender and recipient of each communication; the names of users or the sites visited were not released.

While this release was a significant advance in the amount of data available for cybersecurity research, the researchers themselves noted several limitations. First, by publicly releasing the data, the researchers took pains to remove traffic that revealed too much about the laboratory's network layout, and thus could be used to attack that network.[186] Though they described the kinds of traffic that they removed, they noted that a failure of other researchers to account for the removal could lead them to draw invalid conclusions from the data's characteristics.[187] Second, developing the anonymization algorithm that the researchers applied to the data was it-

[183] *See id.* (noting "the non-availability of multi-[administrative domain] traffic datasets," where an administrative domain is roughly equivalent to an ISP).

[184] *See* Anestis Karasaridis, Brian Rexroad & David Hoeflin, *Wide-Scale Botnet Detection and Characterization* 1 (AT&T researchers reporting results from "billions of flow records" that appear to have been obtained from AT&T's networks).

[185] Ruoming Pang et al., *The Devil and Anonymization* 1 (2006).

[186] Pang et al., *The Devil and Packet Trace Anonymization*, at 7.

[187] *Id.*

self a difficult problem. They believed the anonymization to be difficult to reverse, but recently several anonymization methods have been broken, allowing attackers to match publicly released network traffic data with specific IP addresses. Third, the anonymization process removed some of the structure from the data. Depending on the specific use of the data, this loss of information might lead researchers to draw invalid conclusions, or altogether prevent its use in a study.[188]

A second way to obtain non-content network data is from one of the few network data collection organizations operating today.[189] Most of these organizations' goals and methods differ significantly from those of cybersecurity researchers. Some are operated by computer security companies and do not provide raw data. Instead, they collect and analyze data, and then provide alerts and threat statistics to their subscribers.[190] Other non-commercial data collection organizations work on the same model of providing only high-level statistics and analysis, rather than raw data.[191] cybersecurity researchers, however, frequently need access to data at the level of individual computers on the Internet, rather than aggregated statistics. Network data collection organizations also tend to have limited views of the Internet, perhaps encompassing just a few machines. They might also offer a few specific types of datasets that simply do not contain the data necessary for a particular research use. [192]

4.3.2 Communications Contents

In light of the ECPA's restrictions on the disclosure of communications contents, it is unsurprising that there are fewer sources of data that cybersecurity

[188] *Id.* at 6-7.

[189] For a detailed overview, see Slagell & Yurick, *Sharing Network Logs*, *supra* note 10 at 3-6.

[190] *See* Symantec, DeepSight Threat Management System, *at* https://tms.symantec.com/ (last visited July 12, 2007); Slagell & Yurick, *Sharing Network Logs*, *supra* note 10, at 4 (discussing DeepSight).

[191] *See id.* (discussing Internet Storm Center and DShield). The Internet Storm Center collects and analyzes traffic logs from a large number of users for signs of large-scale malicious activity. *See* About the Internet Storm Center, http://www.dshield.org/about.html (last visited Feb. 21, 2007). According the its Web site, the Internet Storm Center "gathers millions of intrusion detection log entries every day, from sensors covering over 500,000 IP addresses in over 50 countries." *Id.* It is unclear what the sources of data are for these operations, but individual volunteers appear to play a significant role.

[192] *See* Slagell & Yurick, *Sharing Network Logs*, *supra* note 10, at 5.

researchers can tap for communications contents. The few datasets that are available were released under differing circumstances that have rarely been repeated. The interests that the releasing institutions sought to serve were not particularly closely related to cybersecurity; nevertheless, they illustrate the difficulties that public releases of communications contents encounter. Improving cybersecurity researchers' access to such data, I argue in Part 5, will require a combination of legal reform and institutional response.

Though AOL intended to help researchers when it released on the Web a dataset of 20 million search queries from more than 650,000 users in August 2006, the company dealt a setback to efforts to promote more data sharing.[193] The release of these search query logs quickly made headlines. Though AOL made a crude attempt to anonymize the data, it quickly became apparent that the company had not done enough.[194] Within days of the release, journalists from the *New York Times* reported that they determined the identity of one woman whose queries were released (and published an interview with her).[195] of search query logs. Though some researchers welcomed the release, public reaction was quite the opposite. Several AOL users sued AOL under a number of privacy-related theories, including a violation of the Stored Communications Act.[196] Top AOL management quickly disavowed the release as any official act of the company, and three employees involved in releasing the data quickly left AOL.[197] In the end, even the researchers whom AOL intended to help by releasing this dataset were reluctant to use it.[198]

[193] Saul Hansell, *AOL Removes Search Data on Vast Group of Web Users*, N.Y. TIMES C4, Aug. 8, 2006.

[194] AOL did not release the identities of the users whose queries were contained in the sample, and it obfuscated the IP addresses of the computers from which the queries came. AOL did not, however, delete or obfuscate the contents of the queries themselves. As discussed in *supra* note 130 and accompanying text, search queries can contain a significant amount of substance, including the searcher's identity.

[195] Michael Barbaro and Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES A1, Aug. 9, 2006.

[196] *See* http://tinyurl.com/f52oy [Washington Post] (Aug. 22, 2006); Ramkissoon v. AOL LLC, 4:06-cv-05866-SBA, (N.D. Cal., Sept. 22, 2006) (alleging a violation of, among other things, 18 U.S.C. § 2702), *available at* http://www.bermanesq.com/pdf/AOL%20Privacy-Cplt.pdf.

[197] SAN JOSE MERCURY NEWS, http://www.siliconvalley.com/mld/siliconvalley/news/editorial/15326879.htm, Aug. 22, 2006.

[198] *See* Katie Hafner, *Tempting Data, Privacy Concerns; Researchers Yearn To Use AOL Logs, But They Hesitate*, N.Y. TIMES Aug. 23, 2006 (noting that Microsoft and Excite also released search query data, but only to selected researchers).

Though the release of AOL searches created a tremendous stir, the data were actually of limited use to cybersecurity researchers. To be useful to cybersecurity researchers, the search queries have to be linked to the search *results*, which might give researchers some sense of whether the queries lead to malicious software sites or somehow play a role in coordinating attacks.[199] To obtain this information in a comprehensive fashion requires access to a search engine index,[200] which is a closely guarded secret of search engine companies. A recent paper that took this approach was authored by Google researchers, underscoring the point that one institutional force cutting against increased data sharing among cybersecurity researchers is attracting research talent and maintaining the prestige of a research division.

A second major source of communications contents is in the form of a set of e-mails from the accounts of former Enron employees. The Federal Energy Regulatory Commission (FERC) released these e-mails as part of its investigation into Enron's activities in western states energy markets in 2000-01.[201] This dataset contains approximately a half-million e-mails from 150 Enron users.[202] This is a considerable amount of e-mail, but it is relatively small compared to the volume of e-mail that passes through a large enterprise's e-mail server in a single day. For research that involves scanning a realistic mixture of messages, this volume is not realistic. In addition, the Enron dataset does not contain attachments,[203] making it less useful to researchers interested in scanning e-mail attachments for viruses or other malicious code. Despite its limitations, the Enron e-mail dataset is in wide use among cybersecurity researchers seeking to understand such threats, because it is currently the best source of data.[204]

The circumstances surrounding the release of the Enron e-mails were unusual—FERC released the e-mail to allow some insight into the culture of a company whose implosion was a singular event in U.S. corporate history—

[199] *See* Provos et al., *The Ghost in the Browser*, *supra* note 29, at 2.

[200] *See id.*

[201] *See* Federal Energy Regulatory Commission, Information Released in Enron Investigation, http://www.ferc.gov/industries/electric/indus-act/wec/enron/info-release.asp (July 7, 2005). The complete set of release e-mails is available at http://www.cs.cmu.edu/~enron/ (Apr. 4, 2005) [hereinafter *Enron E-Mails*].

[202] *Id.*

[203] *Id.*

[204] This assertion is based on interviews that I conducted with cybersecurity researchers, who have preferred to remain anonymous when discussing common practices for obtaining access to data for research purposes.

and these circumstances overrode many of the concerns about individual privacy.[205]

4.4 Private Access

The second principal approach to obtaining data for cybersecurity research is by working closely with representatives of data sources, such as ISPs and university information technology departments. These relationships require a high degree trust on the part of the data source, because the source often allows the researcher to access large amounts of raw data that the source is obligated to keep confidential. This approach offers the advantage of allowing the researcher to control how data are collected, potentially yielding high-quality datasets that are tailored for a specific use.

But there are problems with this approach. First, it does not scale well. Researchers' relationships data sources outside of their own institutions are personal and develop over the course of years.[206] The need to build trust presents a significant barrier for researchers who are entering cyber security research or expanding into a new area. Sources typically provide data on the condition that the researcher will not distribute them to any other researchers, thus thwarting the goal of making public datasets part of cyber security research. Access to data also depends on the continuing cooperation of the data source; personnel changes or professional disagreements could end a researcher's access to data.

The second problem with relying on trust relationships presents is that they can severely limit the details that researchers may publish about the data that they use. Although there are some notable exceptions,[207] researchers are often circumspect about the sources of their data.[208] This

[205] Attachments were removed to protect privacy, and some e-mails were redacted upon former employees' requests. (It is unclear whether FERC or a site that hosts the dataset performed these redactions.) *See Enron E-Mails*, *supra* note 201.

[206] *See* CAIDA, Toward Community-Oriented Network Measurement Infrastructure: Project Summary 3 (Aug. 2005), *at* http://www.caida.org/funding/cri/nsfcri_2005.pdf.

[207] See the discussion of publication of anonymized network traces above. Researchers at the University of California, Berkeley also have released anonymized records of e-mail account activity information from the Electrical Engineering and Computer Science department's e-mail server. The data are available at http://www.cs.berkeley.edu/~czerwin/traces/WMCSA-Traces.tar (last visited Feb. 21, 2007).

[208] *See, e.g.*, Vyas Sekar Yinglian Xie Michael K. Reiter Hui Zhang, *A Multi-Resolution*

lack of detail can make it difficult for researchers to evaluate published work. Researchers who work for organizations that can provide data, such as ISPs, search engines, and e-mail providers, might have less trouble identifying the sources of their data; but those companies might also put less of a premium on publishing results, especially when competitively sensitive or confidential data underlie the research.

5 A Privacy-Preserving Framework for Cybersecurity Research

Compared to other scientific fields, cybersecurity research is in a kind of limbo. In fields ranging from economics to medicine, well-developed policies support providing researchers with access to data in a way that preserves privacy interests in those data. Cybersecurity lacks this policy apparatus. Medical research provides a stark contrast. The individual privacy interests in medical records are strongly held and defensible under a number of justifications. The information in medical records can lie very close to the core of personhood; personal autonomy dictates that individuals have control over this information.[209] There is also a utilitarian justification for medical privacy. If individuals do not believe that the information that they provide to medical providers will be kept confidential, they may be less likely to provide truthful, complete information in the first place. As a result, their health care might suffer; and this might, in turn, place others at risk. A variety of federal and state laws enforce medical information confidentiality.[210]

Notwithstanding the generally held desire and legal protections for health information privacy, the laws that offer this protection also contain a number of exceptions. These exceptions are based on a number of public or common needs, including the social value of medical research.[211] In particu-

Approach for Worm Detection and Containment 3, in *Proceedings of the 2006 International Conference on Dependable Systems and Networks (DSN '06)* (2006) ("We us[ed] a week-long packet-header trace collected . . . at the border router of a university department . . .").

[209] *See* Gostin & Hodge, *Personal Privacy and Common Goods*, *supra* note 145.

[210] At the federal level, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule limits the disclosures of personal health information that health care providers may make without patient consent. At the state level, many states create a patient-physician privilege, which the patient holds.

[211] *See* Gostin & Hodge, *Personal Privacy and Common Goods*, *supra* note 145.

lar, the HIPAA Privacy Rule permits health care providers to disclose patient records without individual consent to researchers, if the proposed research meets certain substantive requirements and undergoes proper institutional review.[212] Substantively, applicant for the consent waiver must demonstrate that the research would not be feasible without the data. Procedurally, the applicant must satisfy an institutional review board that the disclosure would effect minimal harm to the privacy interests of the individuals whose data are involved, and that data confidentiality, control, and destruction measures are in place.[213] These provisions do not differ based on whether the recipient of data is affiliated with the government or not. The HIPAA Privacy Rule also does not preempt state laws that are more protective of privacy than the Rule itself.[214] Finally, federal law also provides a shield that researchers may invoke in order to refuse to disclose under subpoenas data that they have obtained.[215]

The public health rationale for research exceptions to medical privacy have begun to apply to with increasing force to cybersecurity research. On a scientific level, there are close parallels between the spread of infectious diseases and the spread of Internet-based attacks. Both involve large numbers of systems—biological or computer—that share vulnerabilities and cannot completely defend themselves. Cybersecurity researchers have made the parallels explicit in their work by referring to Internet worm outbreaks as "epidemics."[216] On an institutional level, cybersecurity researchers have be-

[212] *See infra* Part 5.2 for further discussion of these standards.

[213] *See* Gostin & Hodge, *Personal Privacy and Common Good*, *supra* note 145, at 1472-73.

[214] *See id.* at 1465 (citing 45 C.F.R. § 160.203(b)).

[215] Section 301(d) of the Public Health Service Act (codified at 42 U.S.C. § 241(d) grants the Secretary of Health and Human Services discretion to designate certain data exempt from further disclosure. This exemption is quite powerful, as it

> authorize[s] persons engaged in biomedical, behavioral, clinical, or other [health-related] research . . . to protect the privacy of individuals who are the subject of such research by withholding from all persons not connected with the conduct of such research the names or other identifying characteristics of such individuals. Persons so authorized to protect the privacy of such individuals may not be compelled in any Federal, State, or local civil, criminal, administrative, legislative, or other proceedings to identify such individuals.

I am grateful to Chris Hoofnagle for making me aware of this provision.

[216] *See* Xie et al., *Epidemic Attacks*, *supra* note 8; Kostas G. Anagnostakis et al., *A Cooperative Immunization System for an Untrusting Internet* (2004); Staniford et al.,

gun to call for a "cyber Center for Disease Control."[217] The top priority in the "cyber CDC" proposal is to "develop robust communication mechanisms for gathering and coordinating 'field information,'" though the researchers acknowledge that developing the infrastructure for this endeavor "raises potentially immense policy issues concerning privacy and access control."[218]

Though the example of the medical research and privacy context does not translate in all its details to the cybersecurity research and communications privacy context, it does supply a functioning example of a complex research exception. The major structural elements—laws and regulations that define "research" and the conditions of permissible disclosures in the context of institutions that administer the exception and access to data—are directly applicable to the cybersecurity research context. I also argue that institutional support is necessary to make the exception workable. Finally, this section addresses the concern that a cybersecurity research exception would create new threats to privacy and security.

5.1 Law

The project of expanding access to data for cybersecurity research must confront several tensions. The government is uniquely suited to fund cybersecurity research and has displayed a commitment to do so, but the presence of the government on the scene is a major impediment to cybersecurity researchers obtaining the data that they seek. These data are in abundance, but they are mostly controlled by private entities that do not have the incentives to conduct research that serves cybersecurity in general. Communications privacy law imposes few limitations on either internal use of data within the private sector[219] or commercially advantageous disclosures to private parties, but disclosures to governmental entities engaged in research are forbidden. Entities covered by the ECPA can disclose data to law enforcement officials to provide evidence of criminal activity against *that* entity; but, again, they

How to 0wn the Internet], *supra* note 84.

[217] This idea was first suggested by Staniford et al., *How to 0wn the Internet*, *supra* note 84, at 15-18. Others have joined, including at least one legal scholar. *See* Neal Kumar Katyal, *Digital Architecture as Crime Control*, 112 Yale L.J. 2261, 2286 (calling for a "Center for Digital Disease Control").

[218] Staniford et al., *How to 0wn the Internet*, *supra* note 84, at 15-16.

[219] Researchers with federal government funding are subject to the Department of Health and Human Services' regulations concerning the protection of human subjects. *See* 45 C.F.R. § 46.

may not provide data to a "governmental entity" engaged in research. However, legal change that promotes disclosure to certain governmental entities without limiting further disclosure would strip away many of the ECPA's protections. Finally, data protected under the ECPA are subject to "laundering"; once in the possession of an entity that is not covered by the ECPA, they are no longer protected from voluntary or compelled disclosure to the government, whether under the guise of law enforcement or otherwise.

The basic outline of an ECPA exception for cybersecurity research is simple: I propose to allow cybersecurity researchers to obtain access to electronic communications data that the ECPA would otherwise forbid, without the consent of the individuals whose communications are among those that the researchers obtain.

A cybersecurity research exception should apply to all titles of the ECPA, including the prohibition on real-time interception of communications contents.[220] Allowing researchers to use communications would not mark a significant normative or practical shift from the ECPA's current protections. A normative foundation for the anti-interception rules of the Wiretap Act is that they protect privacy in a reasonably thorough manner; at the time these provisions were enacted, catching a conversation as it occurred was likely to be the only opportunity for a party to intercept it. This is no longer the case, especially where electronic communications are concerned. Communications contents (and addressing information) are often stored at the direction of the service provider, the user, or both. Both forms of data become available under less restrictive provisions afterwards. A second rationale for the anti-interception rules is that the objective of an eavesdropper—whether he is a law enforcement official or not—is to learn details that a person has chosen and reasonably expects to keep private.[221] Sifting through these details, with an eye toward assembling them into a criminal profile, for example, typically requires the attention of a human being, who, in turn, may form judgments about the subject of the surveillance.[222] Preventing this kind of privacy

[220] *See* 18 U.S.C. § 2511 (prohibiting in general such interceptions).

[221] *See, e.g.,* Katz v. United States, 389 U.S. 356 (1967) (establishing that the Fourth Amendment protects a conversant's interest in privacy when he has a subjective expectation of privacy, and that expectation is objectively reasonable). *See also* H.R. Rep. No. 99-647, 99th Cong., 2d Sess., 16-17 (1986) (stating that, when the Wiretap Act was passed in 1968, "the contents of a traditional telephone call disappeared and no records were kept").

[222] Modern technology has vastly increased the ability of law enforcement officials to

invasion remains a strong foundation for the anti-interception prohibitions, but cybersecurity researchers are not interested in such uses of intercepted communications contents. Instead, they seek to use streams of electronic communications—e-mail, for example—to test the effectiveness of worm or spam detection software. Part of evaluating these programs is determining whether they will work with a real-world volume and variety of communications.[223] Moreover, since the primary utility of using real-time data in cybersecurity research is to test the performance of defense techniques under real-world conditions, research interceptions would involve recording little or no data for later analysis. Thus, the risk of later, unintended uses of intercepted communications is minimal.[224]

A second element of the research exception is that it should be available to any cybersecurity researcher, provided that the researcher is not a law enforcement agent.[225] Researchers at governmental entities, such as national laboratories and state universities, make vital contributions to cybersecurity research and have no responsibility (or power) to enforce laws. Making an ECPA cybersecurity research exception inapplicable to them would provide no safeguard against the use of communications data by law enforcement agencies, but it would severely complicate the administration of a cybersecurity research exception.

Third, protection under the research exception should be contingent upon approval by an institutional review board (IRB) *before* research activity begins. This limitation would achieve several goals. First, it would prevent the cybersecurity research exception from becoming an ex post justification for a

conduct individualized surveillance by using stored communications records. *See* Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083 (2002).

[223] *See* V. Yegneswaran, P. Barford & S. Jha, *Global Intrusion Detection in the DOMINO Overlay System* (2004).

[224] The scientific validity of using real-time data under these conditions depends on the extent to which conditions are similar to interceptions performed at other times or in other places.

[225] The ECPA's definition of "investigative or law enforcement officer" would suffice for this purpose:

> [A]ny officer of the United States or of a State or political subdivision thereof, who is empowered by law to conduct investigations of or to make arrests for offenses enumerated in this chapter, and any attorney authorized by law to prosecute or participate in the prosecution of such offenses.

18 U.S.C. § 2510(7).

data use or disclosure that the ECPA otherwise would have prohibited. Second, presenting a research proposal to an IRB would impose a certain amount of discipline on researchers; they would not feel entitled to request or disclose data as a matter of course. Third, the review process would generate a paper trail that institutions and, possibly, government regulators could use to maintain accountability under the exception. Fourth, in contrast to other security research exceptions,[226] an after-the-fact determination of whether research was actually related to cybersecurity is not practical. It would pose an unacceptable risk to individual privacy interests to permit ex post protection under the research exception, given that cybersecurity researchers are likely to seek large quantities of sensitive data.

Fourth, the exception should apply to all types of service providers. This is especially important in the context of the SCA, the voluntary disclosure provisions of which apply only to providers of services "to the public." Unless these providers—which include, for example, commercial ISPs and e-mail services—are covered by the exception, there is little to gain from extending the cybersecurity research exception to the SCA. The Wiretap Act and the Pen/Trap statute, by contrast, apply to any service provider. Excluding certain ones, especially on a factor as dated as being available to the public, would make the exception more administratively burdensome and would reduce its effectiveness.

Fifth, the exception should prohibit any researcher who receives data under the exception from redistributing them. The justification for this limitation is twofold. The first is prudential. Since explicit legal protection for cybersecurity data sharing is a new idea, taking a cautious approach, by making each disclosure of data subject to approval by the relevant IRBs, is warranted. The second reason relates to the security of data providers. Some types of data that cybersecurity researchers would like to obtain include information that could help an attacker find weaknesses in the sources networks or systems. A researcher who receives these data might not appreciate the full extent of such risks; therefore, allowing the source to maintain control over distribution of the data is necessary to protect the source. A related point is that a data source might wish to keep data away from researchers employed by a competitor. In that case, the data source is best situated to assess the competitive risk involved in disclosure.

[226] *See, e.g.,* the encryption research and security testing provisions of the Digital Millennium Copyright Act (DMCA), 17 U.S.C. §§ 1201g & (j).

Sixth, the exception should provide by default a shield from any obligation to comply with compulsory process requiring a researcher to disclose data.[227] Another threat that would arise from a cybersecurity research exception is that it would allow disclosure of data to persons and entities not covered by the ECPA's compulsory process regulations. While this is how the SCA currently works—the disclosure provisions apply to data when they are in the possession of certain kinds of entities but do not follow the data after they are disclosed—would have the practical effect of placing massive quantities of data under the control of entities that may have no interest or ability to defend against government or civil requests or demands for data.

Seventh, basic subscriber information should not be covered under the exception. As explained in Part 4, the interests of cybersecurity researchers in real data lie in what the data reveal about network traffic flows, the spread of malicious code across networks, and so forth. None of this analysis depends on a researchers' being able to identify whose name was associated with an IP address at a particular time. There is simply no scientific reason to allow the disclosure of such data under an ECPA exception. This prohibition would not be entirely effective in separating individual identities from their network traffic,[228] but it would remove an easy means for researchers to link communications records to individuals.

Finally, the research exception should preempt state laws that provide higher levels of protection than the ECPA.[229] Creating an exception to state laws that add protection to a relatively weak federal regime of statutory privacy protection is not something that I suggest lightly, but the research exception might prove unworkable otherwise. Data that are relevant to a cybersecurity research question might come from many different states.[230] Though the problem of differing state regulations arises in many information collection contexts—health information is one example— the effect of differing laws on Internet-based data collection would be more acute because nearly every communication crosses state lines. Internet companies have

[227] *Cf.* 42 U.S.C. § 241(d); *see supra* note 215.

[228] *See supra* page 42 (discussing AOL's release of "anonymized" search engine queries). Other contents that might become available under the cybersecurity research exception, such as e-mail, would carry their own link between individuals and records.

[229] *See* the discussion of state law in § 3.1.4, *supra.*

[230] *See, e.g.,* Staniford et al., *How to 0wn The Internet*, *supra* note 84 (proposing a decentralized, widely distributed set of network "sensors" to collect information about network-based cybersecurity threats).

dealt with this complication by specifying in their terms of service which state's law will govern in any dispute with a user.[231] Enacting a cybersecurity research exemption to the ECPA, though it might override some state laws, would have the considerable virtues of public ratification and creating a clear national standard for disclosure to researchers.

5.2 Institutions

The call to create a cybersecurity research exception to the ECPA prompts two further questions: (1) Is it administrable? and (2) Would it be effective in reversing the strong institutional forces that currently oppose data sharing? I consider these questions in turn.

Federal law provides a broadly applicable structure for reviewing research that provides a starting point for how to administer the ECPA research exception. Institutional review boards (IRBs) arose in the United States to prevent harm to human research subjects in health and medical experiments, but their use has expanded over time to cover all federally funded research involving human subjects.[232]

Federal rules for IRBs are administered by the Department of Health and Human Services through the "Common Rule,"[233] The Common Rule defines

[231] *See, e.g.,* Verizon, Verizon Internet Access Service Terms of Service § 19.4 (specifying that Virginia law will apply to all disputes), http://netservices.verizon.net/ (last visited July 20, 2007).

[232] For a recent (but critical) history of IRBs, see Philip Hamburger, *The New Censorship: Institutional Review Boards*, 2004 Sup. Ct. Rev. 271 (2005). For a critical view of the expansion of IRB approval requirements into the social sciences (including legal scholarship), see Dale Carpenter, *Institutional Review Boards, Regulatory Incentives, and Some Modest Proposals for Reform*, Nw. L. Rev. (forthcoming 2007). According to Professor Hamburger, the impetus for the rapid expansion of IRBs for federal, state, and private research was the revelation in the late 1960s of the Tuskegee syphilis experiment, in which medical researchers studied the "course of untreated syphilis in black men in Macon County, Georgia." Hamburger, *Institutional Review Boards* at 272 n.4. During the experiment, the researchers failed to seek informed consent, misled the subjects into believing that they were receiving free medical care, and failed to inform the subjects that penicillin was available to treat their disease. *Id.*

[233] 45 C.F.R. § 46.

"research,"[234] "institution,"[235] and "human subject."[236] The Common Rule also supports a number of features that would be desirable for administering a cybersecurity research exception. For example, the Common Rule requires IRB members to have diverse backgrounds, with representation from scientific and nonscientific disciplines,[237] and the Rule permits joint review of applications of multi-institutional research proposals.[238] The IRB composition requirement would help to ensure examination of proposals involving cybersecurity data from a number of disciplinary angles, while the cooperative research provision would facilitate the efficient review of joint proposals that, given the need of cybersecurity researchers to work with common datasets, would likely by common. Finally, the Common Rule provides standards for IRB approval of projects seeking approval of a waiver of research subject consent.[239]

The Common Rule, however, provides little guidance on protecting the privacy of research subjects.[240] As I have hinted above,[241] this level of guidance is inadequate for a cybersecurity research exception to the ECPA. To return again to the medical context, the HIPAA Privacy Rule provides an example of how to layer privacy considerations on top of the basic IRB structure. The Privacy Rule's guidelines include consideration of user privacy as well as the security of the data source; both are necessary to assess risk in

[234] 46 C.F.R. § 102(a) ("Institution means any public or private entity or agency (including federal, state, and other agencies).").

[235] *Id.* § 46.102(d) ("Research means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy, whether or not they are conducted or supported under a program which is considered research for other purposes.").

[236] *Id.* § 46.102(f) ("Human subject means a living individual about whom an investigator (whether professional or student) conducting research obtains (1) Data through intervention or interaction with the individual, or (2) Identifiable private information.").

[237] *Id.* § 46.107.

[238] *Id.* 46.114.

[239] *See* 45 C.F.R. § 46.117(c) (authorizing consent waiver from all subjects if the IRB finds minimal risk from a breach of confidentiality or a minimal risk of harm to subjects).

[240] *See* Gostin & Hodge, *Balancing Personal Privacy and Common Good*, at 1472-73 (noting that the Common Rule "conditions IRB approval of government-sponsored research on whether 'there are adequate provisions to protect the privacy of subjects' ") (citing 45 C.F.R. § 46.111(a)(7)).

[241] *See supra* Part 5.1.

the event of a breach of confidentiality, whether accidental or intentional.[242]

User privacy considerations for cybersecurity data should include, first of all, an assessment of the extent to which anonymization is practical. Data anonymization is an open research question,[243] so it is not realistic to expect that an arbitrary collection of data could be processed in such a way as to sever the link between the data and the individual or entity whose activities are captured in the data. The prospects for anonymization also form a continuum based on the kind of data in question—e-mail may be all but impossible to anonymize, while some forms of network data might have little connection to an individual user. Still, IRB members from the data source institution and the data recipient institution should weigh, on a case-by-case basis, the extent to which anonymization is possible without ruining structure in the data that is necessary for a proposed use. Finally, the IRB should consider the proposal's plan for transporting data to the recipient, containing the data when they are in use, and ensuring the destruction of usable copies once the researchers have completed their use of the data.

The Department of Homeland Security has funded a network dataset repository that provides an interesting example to examine under these principles. This repository, known as PREDICT (Protected Repository for the Defense of Infrastructure Against Cyber Threats),[244] is intended to coordinate the process of giving cyber security researchers access to "network operational data."[245] A revealing fact about PREDICT is that it is not in operation, and there is no indication when this might change.

Still, based on its organizational documents, PREDICT would implement many of the safeguards that I have argued would accompany IRB review for an ECPA exception. Under current law, however, PREDICT faces considerable limitations. In brief, PREDICT has three types of entities for handling data: Data Providers, which are the original sources of network data; Data Hosts, which store datasets once they have been approved for use; and Re-

[242] Intentional breaches could be the result of either insider or outsider activity. It is important to consider both; recently, insider threats have gained considerable attention in the cybersecurity research community. *See* CSTB, *Toward a Safer and More Secure Internet*, *supra* note 6, at 215-20.

[243] *See* Pang et al., *The Devil and Anonymization*, *supra* note 185; Bakken et al., *Data Obfuscation: Anonymity and Desensitization of Usable Data Sets*, 2 IEEE SECURITY & PRIVACY 34 (Nov./Dec. 2004).

[244] PREDICT Home Page, *at* https://www.predict.org/ (last visited Feb. 22, 2007).

[245] RTI International, PREDICT Portal Overview 1, Aug. 17, 2006, *at* https://www.predict.org/.

searchers. A fourth entity, the Coordinating Center, would be administered by a nonprofit corporation under contract with DHS.[246]

The Coordinating Center serves many of the functions that an IRB would serve under the ECPA exception that I proposed. The Center would approve datasets for use in PREDICT and has considerable flexibility to consider the sensitivity of the privacy interests in each dataset; it may treat privacy as a continuum, rather than according to the broad categories of the ECPA.[247] The Coordinating Center also acts an initial gatekeeper for deciding whether a researcher may be considered eventually to use PREDICT data.

Most importantly, the Coordinating Center orchestrates the review board that processes applications for uses of PREDICT data. Each use of PREDICT data requires separate approval from the board. The board is to be composed of official from DHS, the Coordinating Center, the Data Host, the Data Provider, and as a member of the "[c]yber-defense research committee."[248] As part of this review board, the Data Provider can reject any proposed use of its data.[249]

Despite these safeguards, PREDICT faces several legal ambiguities under the current ECPA. PREDICT is bound by the jagged edges of the ECPA discussed in Part 3, particularly the prohibitions voluntary disclosure of non-content records to governmental entities. Though DHS itself may not host PREDICT data under the current organizational documents, it appears to be possible that other forms of "governmental entity," such as a state university, will house data. In those cases, PREDICT would have to ensure that the communications service providers "to the public"—such as commercial ISPs—do not provide data to these hosts.[250] PREDICT might also need to

[246] *See* Memoranda of Agreement between PREDICT Coordinating Center (PCC) and Data Provider, Data Host, and Researcher. All three memoranda are available at https://www.predict.org/.

[247] Data Providers designate the sensitivity of the data that they provide and control the conditions under which data may be released. *See* PREDICT Overview. The Coordinating Center establishes categories of dataset types and requires Data Providers to comply with de-identification and other data "sanitization" requirements for data in a given category. PCC-Data Provider MOA at 4.

[248] *See* PCC-Data Host MOA at 4.

[249] *See* PCC-Data Host MOA at 4 ("The Data Provider representative shall have absolute veto power over any application for access to its Data.").

[250] There are signs of this limitation in the available listings of datasets that PREDICT expects to make available. According to presentations given at a September 2005 conference hosted by DHS, the only entities that were scheduled, at that time,

ensure that any such data that providers contribute do not end up in the control of government-affiliated researchers, or else bar such researchers altogether. This is not a fault of PREDICT's design; it is simply the way it must be under the current ECPA. Nonetheless, the effect is to continue to divide the cybersecurity research community along the lines of a "governmental entity," rather than the question of whether a given researcher is employed by a law enforcement agency.

Whether similar institutions would arise under the ECPA exception that I propose is a matter of conjecture, but the exception would eliminate three of the major limitations that PREDICT faces. First, the exception would make it clear that government-affiliated researchers would be allowed access to communications data acquired or disclosed after proper IRB review. This would not only produce an administrative simplification but would also expand the set of researchers that could examine a particular dataset to include those at state universities, national laboratories, or core government agencies, provided that the particular researcher is not a law enforcement officer.[251] Second, a cybersecurity research exception to the ECPA would make it clear that any data source, including a commercial ISP, could share data with eligible researchers. This condition would create the potential for cybersecurity data sharing to provide a wide view of the Internet, which researchers have previously considered unattainable because of the legal risks.[252] Third, the ECPA research exception would allow a greater diversity of data sharing institutions to evolve, including ones that entirely avoid the involvement of

to contribute data to PREDICT were most likely not services that the SCA regulates under its non-content record provisions. PREDICT expected to receive datasets from several universities, *see* Michael Bailey, Virtual Center for Network and Security Data 2, 7, *at* http://www.cyber.st.dhs.gov/public/PREDICT/DHS-anon-workshop-overview-09272005.pdf (Sept. 27, 2005); a national laboratory (including a dataset containing "anonymized *contents*"), *see* Vern Paxson, LBNL/ICSI Enterprise Traces 3, *at* http://www.cyber.st.dhs.gov/public/PREDICT/PREDICT.Sep05A.pdf (Sept. 27, 2005) (emphasis in original); and data that were not related to the statutory definition of an electronic communication. *See generally* Tom Vest, PCH/PREDICT Update: Routing Topology and Network Quality Data Collection and Hosting, *at* http://www.cyber.st.dhs.gov/public/PREDICT/DHS050927v1.pdf (Sept. 27, 2005) (discussing routing table data).

[251] *See* the definition of "law enforcement officer," *supra* note 225.

[252] Of course, whether cybersecurity data sharing on this scale would actually occur also depends on whether firms with relevant data find that institutional controls sufficiently address concerns about competitors having access to their data. I address this question below.

officials from law enforcement agencies.[253] One could imagine, for example, consortia of universities and corporate network operators arising to combine the operators' wealth of data with the universities' depth of research talent. A combination of grant money and institutional funding could sustain these efforts, allowing data to remain available over time.

Still, the question remains: Would a cybersecurity research exception to the ECPA actually alleviate the data dearth? The answer depends on how the exception would alter the elements of the current "security culture" that both derive from and add to the ECPA's current security model.[254] A definitive answer is impossible to provide, but a research exception to the ECPA shows promise along several fronts for changing this culture. First, a legislatively enacted research exception would require public debate. As I discussed in Part 5.1, this process would serve to attach a measure of legitimacy to research as a reason to relax some of the ECPA's current restrictions.[255] Congressional approval of a specific reason for analyzing communications data would stand in stark contrast to the current environment, in which institutions' desire to avoid bringing to light more information about how much communications data they store, and how they use it, acts as a deterrent to share the data with researchers.

This leads to a second institutional inhibition against data sharing that a research exception might begin to reverse. Commercial firms in particular do not want their competitors to have access to data that might reveal competitively sensitive details of their networks. Such details include everything from the activities of network users to information that network data could supply about how an operator arranges its network to enhance performance. The legal and institutional structure that I have proposed contain two safeguards against such uses of data. First, the institution providing data could refuse to allow a competitor's employees to access the data. A second, less draconian control is that the recipient's proposed use of data must satisfy an IRB as being related to cybersecurity. The proposal in this Article does not discuss possible remedies available to a data provider for misuse, aside from withdrawal of the provider's permission for a particular researcher to use the data. Extensions of the IRB structure to handle this problem might

[253] Recall that one member of the review board in PREDICT must be DHS official.

[254] *See* the discussion in Part 3.2.

[255] *See* Lawrence O. Gostin et al., *the Law and the Public's Health: A Study of Infectious Disease Law in the United States*, 99 COLUM. L. REV. 59, 93-94 (discussing conditions for legitimacy of government action in the context of controlling infectious disease) (1999).

be possible.

Third, the combination of legal clarity and institutional support that the proposal in this Article carries might encourage the autonomous, yet interconnected, entities that create the Internet to re-evaluate their own interests in cybersecurity. Specifically, the fundamental economic difficulties of cybersecurity might begin to shift if means develop to share data explicit legal protection and institutional controls that manage the risk of disclosure to truly adversarial recipients. Whether these conditions will be sufficient to reverse a culture that disfavors cybersecurity to varying degrees in the legal, economic, and technological realms remains to be seen.

5.3 Creating New Threats?

A further question about the cybersecurity research exception that I have proposed is whether it would create new threats to individual privacy or the security of data providers. One threat that might arise is from law enforcement officials or others who seek the data from cybersecurity researchers, rather than the source of the data. The exception is equipped to meet these threats; researchers who receive data under the exception would be barred from voluntarily disclosing it, and others would be unable to use subpoenas or other methods to compel disclosure of the data. In this regard, the research exception would leave current law unchanged. Parties seeking communications data would have to obtain them from the source, using the legally required procedures.

A new risk, however, would arise from the researchers who obtain data. For example, a researcher who receives Internet usage logs from a commercial ISP might to post them on the Web, notwithstanding his duty under the exception to use the data only on an isolated network. Whether the researcher does so intentionally or by mistake is largely irrelevant; the loss of privacy is the same. Or a researcher, having used a dataset to learn what kinds of botnet traffic an ISP has learned to detect, might create malicious software that evades this detection.

These "insider threats" would be significant risks under the ECPA cybersecurity research exception, but they are risks that computer scientists have learned how to manage. A central tenet of computer security is that it is perilous to ignore the "skil[l] and motivat[ion]" of an adversary.[256] This out-

[256] Computer Science and Telecommunications Board, *Summary of Discussions at a Plan-*

look applies to both insiders, who hold some authorization to have access to data or other system resources, and to outsiders.[257] There is no justification for assuming that cybersecurity researchers do not present insider risks.

The cybersecurity research exception contains several measures that would help to manage such risks. The first is technological.[258] One responsibility of the IRBs is to review a plan for maintaining the confidentiality of datasets. Researchers and IRBs would have flexibility in determining which measures are appropriate, depending on the specific data and proposed use at issue. They could, for example, mandate that the data be delivered on a separate disk, that the researchers use the data only on an isolated network, and that all communications relating to the data be encrypted.[259] A second way that the proposed cybersecurity research exception could manage risk is through a suspension of an errant researcher's ability to use or obtain data that were disclosed under the exception. The duration of this suspension might depend on whether it was intentional or accidental, as well as the quantity and sensitivity of data that the researcher leaked or misused. Third, IRBs would impose stringent controls over which researchers would be able to obtain data in the first place under the cybersecurity research exception. Furthermore, the IRB would review each proposed use of data. These controls would increase the chances of barring identifying researchers who cannot establish that they are trustworthy (perhaps because of past misdeeds), or who do not have a legitimate, security research-related need to use sensitive data.

A special case of "insider" data disclosure is that of publishing research results. One of the goals that cybersecurity researchers have in seeking increased and more formalized access to data is the ability to identify what data they used during an experiment, and to discuss whether particular features of the data brought especially noteworthy results. The IRBs that I propose to use with the ECPA exception could allow data sources to decide whether researchers could reveal the source of data in publications.[260] The

ning Meeting on Cyber-Security and the Insider Threat to Classified Information 10 (2001), *at* http://books.nap.edu/openbook.php?record_id=10197&page=10.

[257] *See* CSTB, *Toward a Safer and More Secure Cyberspace*, *supra* note 6, at 215-20 (discussing insider threats).

[258] *See id.* (explaining role of technology in managing insider risk).

[259] For an example of a security plan that contains these elements and others, see David A. Wagner, Security Plan for Source Code Review Teams, *at* http://www.sos.ca.gov/elections/voting_systems/ttbr/source_code_security_plan.pdf (last visited July 20, 2007).

[260] PREDICT takes this approach. *See* PCC-Data Provider Memorandum of Agreement.

question of how to decide which details of a dataset should be published is difficult to answer generally and in advance. The approach that PREDICT takes is a reasonable solution: require proposed publications to be approved by a review board to determine whether they comply with the conditions for access to the data and whether they would put the confidentiality of the data at risk.[261] The data source should be represented during this review but should not be allowed to veto publication.[262]

Ultimately, no combination of technological, legal, and institutional controls could eliminate these insider risks. But this is no different from other cybersecurity risks; "[t]he best is the enemy of the good. Risk management is an essential element of any realistic strategy for dealing with security issues."[263] The proper comparison for privacy and security risks under cybersecurity research exception is not a world without any research-related disclosures of communications data, but rather a world in which we continue down the current, non-cooperative path of cybersecurity research.

6 Conclusion

Cybersecurity research and policy have stumbled badly over the question of privacy. This Article has advanced the case that a modest, administrable cybersecurity research exception is a critical first step to taking the relationship between privacy and cybersecurity beyond more than a broad objective. I have argued that the ECPA's security model, which allows enshrines a notion of organizations protecting themselves through disclosure to law enforcement agencies, is badly outdated. Improving cybersecurity requires coordinated information sharing. A cybersecurity research exception would make this formally possible, and, with the forms of institutional that I have proposed, the exception would help to allay the many institutional concerns that network providers have about sharing data. Moreover, this exception would be strictly limited to research and would not expand access to communications data for law enforcement purposes. Thus, considering communications privacy objectives sooner rather than later could substantially advance the national interest in improving cybersecurity.

[261] *See* RTI Int'l, Memorandum of Agreement Between PCC and Data Provider at 2, https://www.predict.org/ (describing the Publication Review Board).

[262] *See* id.

[263] CSTB, *Cybersecurity Today and Tomorrow*, *supra* note 5, at 7.

With the exception of PREDICT, the federal government has shown little serious thought about how to reconcile cybersecurity with privacy. Going back to the *National Strategy to Secure Cyberspace*, cybersecurity policy has throughly mixed law enforcement, private sector, and research interests in information sharing into the nebulous catch-all of public-private partnerships. But the types of data sharing that serve these interests vary considerably with context, as do the privacy interests. Separating these interests is overdue and has been a major objective of this paper.

The Department of Homeland Security, in its role as the leader in cybersecurity policy, has done little work to separate them; and its initiatives may be suffering as a result. A recent GAO study found that "the private sector continues to be hesitant to provide sensitive information regarding vulnerabilities to the government as well as with other sector members due to concerns that, among other things, it might be publicly disclosed."[264]

The Department of Justice, which understandably acts out of its interest as the enforcer of cybercrime laws, has held extensive meetings with major ISPs, seeking their voluntary commitment to retain network data. At the same time, the Department is pushing legislation that would require ISPs to retain data.

Neither of these approaches—DHS's apparent reluctance to seek legal change, and the DOJ's campaign for law enforcement-oriented data retention requirements—serves the needs of cybersecurity research. Inaction is likely to perpetuate the current cybersecurity data dearth. A data retention requirement might increase the amount of data stored by ISPs and other network operators, but it would do nothing to provide legal protection for sharing communications data with cybersecurity researchers. Legal support for sharing data with cybersecurity researchers under a strictly controlled disclosure regime provides the best way to advance research and cybersecurity over the long term.

[264] GAO, *Critical Infrastructure*, *supra* note 18, at 14.