Harvard Journal of Law & Technology
Volume 22, Number 1 Fall 2008

AMENDING THE ECPA TO ENABLE A CULTURE OF CYBERSECURITY RESEARCH

Aaron J. Burstein[*]

TABLE OF CONTENTS

I. INTRODUCTION 168

II. THE UNIQUE PROMISE OF TECHNICAL RESEARCH IN IMPROVING CYBERSECURITY 172
A. Defining Cybersecurity 172
B. Defending Against Known Threats: The Inadequacy of Prevention 175
C. The Limits of Deterrence 179
D. Adapting to Evolving Threats Through Detection and Resilience: The Case for Focusing on Technical Research 181

III. HOW COMMUNICATIONS PRIVACY LAW LIMITS CYBERSECURITY RESEARCH 184
A. Communications Privacy Law 184
1. Wiretap Act 185
2. Stored Communications Act 188
3. Pen/Trap Statute 191
4. State Laws 193
5. Gaps 194
B. Institutions 198

IV. COPING WITH THE DEARTH OF CYBERSECURITY DATA 200
A. Scientific Goals of Data Sharing 200
B. Data Needs: A Picture of the Ideal 201

---

* TRUST and ACCURATE Research Fellow, Samuelson Law, Technology & Public Policy Clinic and Berkeley Center for Law & Technology, University of California, Berkeley School of Law (Boalt Hall), aburstein@law.berkeley.edu. I am deeply grateful to Deirdre Mulligan for sharing many critical ideas during the writing of this Article, and for reviewing earlier drafts. I thank Kevin Bankston, Kimberly Claffy, Joseph Lorenzo Hall, Paul Ohm, Vern Paxson, and participants in the 2007 Intellectual Property Scholars Conference and the 2008 Privacy Law Scholars Conference for helpful comments. I also thank 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.

C. Public Releases 203
1. Non-Content Data 203
2. Communications Contents 205
D. Private Access 207

V. A PRIVACY-PRESERVING FRAMEWORK FOR CYBERSECURITY RESEARCH 208
A. Requirements for a Cybersecurity Research Exception to the ECPA 209
B. Institutions 214
C. Creating New Threats? 219

VI. CONCLUSION 221

I. INTRODUCTION

Computer and network security (together, "cybersecurity") have become matters of major economic, social, and national security importance. Computer networks have joined other systems like transportation, energy, defense, and health care that are critical to the functioning of the national economy.[1] Indeed, computer networks are the "nervous system" that ties together and controls these other components of our national infrastructure.[2] Increasingly sophisticated network attacks, however, constantly threaten this infrastructure and the activities that rely on it. These attacks do not simply damage an isolated machine, or disrupt an individual's or single enterprise's access to the Internet. Instead, modern attacks threaten to target infrastructure that is integral to the economy, national defense, and daily life.[3]

Although society has benefited from innovative applications that connect people and devices via the Internet,[4] malicious parties have taken advantage of the Internet's connectivity by exploiting technological and human vulnerabilities to perpetrate attacks for personal,

---

1. *See* PRESIDENT'S CRITICAL INFRASTRUCTURE PROT. BD., NATIONAL STRATEGY TO SECURE CYBERSPACE vii (2003) [hereinafter PCIPB], *available at* http://www.whitehouse.gov/pcipb/cyberspace_strategy.pdf; *see also* U.S. GOV'T ACCOUNTABILITY OFFICE, CRITICAL INFRASTRUCTURE: CHALLENGES REMAIN IN PROTECTING KEY SECTORS 8 (2007) [hereinafter GAO, PROTECTING KEY SECTORS], *available at* http://www.gao.gov/new.items/d07626t.pdf. 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. PCIPB, *supra* note 1, at 1.

3. *See* COMPUTER SCI. & TELECOMMS. BD., NAT'L ACAD. OF SCIS., TOWARD A SAFER AND MORE SECURE CYBERSPACE vii (Seymour E. Goodman & Herbert S. Lin eds., 2007) [hereinafter CSTB, MORE SECURE CYBERSPACE], *available at* http://books.nap.edu/openbook.php? record_id=11925.

4. *See* Jonathan L. Zittrain, *The Generative Internet*, 119 HARV. L. REV. 1974, 1980 (2006) ("Generativity denotes a technology's overall capacity to produce unprompted change driven by large, varied, and uncoordinated audiences. The grid of PCs connected by the Internet has developed in such a way that it is consummately generative.").

financial, and political gain.[5] The FBI estimated in 2005 that cybercrime costs the United States $67.2 billion annually.[6]

But the risks of insecurity go beyond financial damage. For example, Estonia endured a massive flood of Internet traffic in 2007, which crippled networks within the country, leading to a shutdown of banks and other services.[7] In 2003, the "Slammer" worm spread rapidly across the Internet, shutting down South Korea's "entire Internet system" and disrupting ATM transactions in the United States.[8] The following year, the "Witty" worm deleted random data from the hard drives of the hosts it infected worldwide.[9] As networked devices — not only personal computers but cell phones, appliances, and even the materials in buildings — become pervasive,[10] the potential for harm from successful attacks will continue to grow. Although the United States has not suffered major Internet physical infrastructure outages as a result of cyberattacks, attempts to defeat the defenses of critical information systems are relentless.[11]

Understanding how to detect and defend against such attacks is an active research area within computer science,[12] and technical research[13] in this area is, in turn, a central element of national cybersecurity policy.[14] The era of network-wide attacks began in 1988, when the "Internet Worm," a program that replicated itself from one networked computer to another without human intervention, quickly spread to an estimated five to ten percent of computers connected to

---

5. *See* COMPUTER SCI. & TELECOMMS. BD., NAT'L ACAD. OF SCIS., CYBERSECURITY TODAY AND TOMORROW: PAY NOW OR PAY LATER 4 n.9 (2002) [hereinafter CSTB, CYBERSECURITY TODAY AND TOMORROW], *available 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.").

6. U.S. GOV'T ACCOUNTABILITY OFFICE, CYBERCRIME: PUBLIC AND PRIVATE ENTITIES FACE CHALLENGES IN ADDRESSING CYBER THREATS 15 (2007) [hereinafter GAO, ADDRESSING CYBER THREATS], *available at* http://www.gao.gov/new.items/d07705.pdf.

7. *See* John Schwartz, *When Computers Attack*, N.Y. TIMES, June 24, 2007, at WK1.

8. *Internet Worm Strikes*, HERALD SUN (Melbourne), Jan. 28, 2003, at News 10.

9. Colleen Shannon & David Moore, *The Spread of the Witty Worm*, IEEE SECURITY & PRIVACY, July/Aug. 2004, at 46.

10. CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 20.

11. The Department of Defense has reported, for example, that it experiences approximately forty successful cyberattacks per month and tens of thousands of close calls per year. Bob Brewin, *Successful Cyberattacks Against DOD Drop*, FCW.COM, Mar. 29, 2007, http://www.fcw.com/online/news/98089-1.html.

12. *See, e.g.*, Yinglian Xie et al., *Forensic Analysis for Epidemic Attacks in Federated Networks*, 2006 PROC. IEEE INT'L CONF. ON NETWORK PROTOCOLS 43, *available at* http://www.ieee-icnp.org/2006/papers/s2a1.pdf.

13. This article uses "technical research" interchangeably with "cybersecurity research" to mean research performed using the methods of computer science or engineering, to distinguish it from approaches to studying cybersecurity based in social science, law, and policy.

14. *See* PCIPB, *supra* note 1, at xi.

the Internet.[15] The Worm exploited flaws in individual computers, traversing their networks without regard to organizational boundaries, and quickly spread from one organization's network to another. The response to the Worm also crossed institutional boundaries, with researchers and administrators sharing alerts and suggestions for mitigation with their peers at other organizations.[16] This informal coordination of defenses helped to stop the Internet Worm relatively quickly, and computer security experts who studied the Worm recommended creating a formal organization to coordinate information sharing about vulnerabilities and malicious activity. Given the complexity of the Internet and the diversity of malicious activity connected with it, understanding what information to share and how to analyze it remains a difficult scientific problem.

Unfortunately, current U.S. law adds to the difficulty. Communications privacy laws — specifically the Electronic Communications Privacy Act ("ECPA") — impede the sharing of Internet data with cybersecurity researchers.[17] The ECPA currently prohibits many instances of the acquisition, use, and disclosure of e-mails, Internet usage histories, instant messaging conversations, and other forms of electronic communications, without providing a research exception.[18] The central argument of this Article is that the ECPA should be amended to include a cybersecurity research exception and that a properly crafted and administered exception would pose little risk to communications privacy.

---

15. *See generally* Eugene H. Spafford, *A Failure to Learn from the Past*, 2003 PROC. COMPUTER SECURITY APPLICATIONS CONF. 217, *available at* http://www.acsac.org/2003/papers/classic-spafford.pdf. Professor Spafford defines a worm as "a program that can run independently and can propagate a fully working version of itself to other machines." *Id.* at 218. This ability to run independently distinguishes a worm from a computer virus, which "is a piece of code that adds itself to other programs, including operating systems. . . . [I]t requires that its 'host' program be run to activate it." *Id.*; *see also* 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 ("CFAA"), 18 U.S.C. § 1030).

16. *See* Spafford, *supra* note 15, at 227 (crediting the "'old-boy' network" of researchers and system administrators with quickly stopping the Internet Worm). For more of this history, see Zittrain, *supra* note 4, at 2003–07.

17. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended in scattered sections of 18 U.S.C.).

18. The Digital Millennium Copyright Act, ("DMCA") on the other hand, contains an encryption research exception to its anti-circumvention provisions. 17 U.S.C. § 1201(g) (2006). Though this exception has been criticized as being too narrow, *see* Joseph P. Liu, *The DMCA and the Regulation of Scientific Research*, 18 BERKELEY TECH. L.J. 501, 509–12 (2003), it at least illustrates Congress's recognition that it should strike some balance between the benefits of research and the risk that the research exception will be abused. *See* H.R. REP. NO. 105-551, pt. 2, at 27 (1998) ("The goals of this legislation would be poorly served if these provisions had the undesirable and unintended consequence of chilling legitimate research activities in the area of encryption."); *see also* David Nimmer, *Appreciating Legislative History: The Sweet and Sour Spots of the DMCA's Commentary*, 23 CARDOZO L. REV. 909, 950 (2002) (discussing the legislative history of 17 U.S.C. § 1201(g)).

Sharing cybersecurity data in this manner would entail some risk. Allowing easier access to communications data increases the chance that the data will be misused. Data sharing, of course, can threaten more than communications privacy. The firms that control communications data are often reluctant to share it out of concern that their customers will react negatively, or that the data will expose sensitive information.[19]

The result is that much technical cybersecurity research is bound to the data available from the researcher's own institution, which in most cases is quite limited. Organizations seek to make their own information systems as secure as they can within resource constraints, even if the defenses they employ end up harming cybersecurity overall. As two cybersecurity researchers have put it:

> 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.[20]

Add to this the fact that the current culture of security encourages individuals and institutions to view security as an expense rather than a necessary means of avoiding lost time, money, and information, and the depth of the cybersecurity problem becomes apparent.[21] Given the need to coordinate responses on a wide scale to combat network threats, it is appropriate to consider how law might support system-level cybersecurity research and responses while protecting privacy.

Both Congress and the Executive Branch have recently become aware of the need to integrate privacy into cybersecurity policy.[22] In particular, the guiding national cybersecurity policy document, the

---

19. *See, e.g.*, Gonzales v. Google, Inc., 234 F.R.D. 674, 684–85 (N.D. Cal. 2006) (explaining Google's argument for refusing to disclose search query data to the government on the ground that the data is trade secrets, and noting that "[b]y declaration, Google represents that it does not share this information with third parties and it has security procedures to maintain the confidentiality of this information").

20. Adam Slagell & William Yurcik, *Sharing Computer Network Logs for Security and Privacy: A Motivation for New Methodologies of Anonymization*, 2005 PROC. INT'L CONF. ON SECURITY & PRIVACY FOR EMERGING AREAS IN COMMC'N NETWORKS 80.

21. *See* CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 88.

22. *See* Cyber Security Research and Development Act, Pub. L. No. 107-305, § 4, 116 Stat. 2367, 2368–70 (2002) (codified at 15 U.S.C § 7403 (2006)) (appropriating millions for the National Science Foundation to grant to cybersecurity researchers). The statute also indicates that increased information sharing is needed. *Id.* § 1(4)(B), 116 Stat. at 2367 (codified at 15 U.S.C. § 7401(4)(B)) ("Computer security technology and systems implementation lack . . . adequate coordination across Federal and State government agencies and among government, academia, and industry . . . .").

National Strategy to Secure Cyberspace, recognizes that a new approach is necessary to encourage firms with data to share it with researchers who can put it to use.[23] The *National Strategy* also recognizes that cybersecurity responses must protect privacy and civil liberties.[24] But anyone searching this document for details about how to reconcile security and privacy will be disappointed. Communications privacy law, in particular, is an example of the law's failure to coordinate cybersecurity research and practice.

This Article argues that the ECPA's barriers to cybersecurity research are substantial and that addressing them forthrightly best serves the interests of research and privacy. The argument proceeds in four parts. Part II explains how the economic and technical components of cybersecurity render market- and law enforcement-based efforts to improve cybersecurity inadequate. Improving cybersecurity depends critically on continued research, but cybersecurity research currently faces a dearth of realistic, usable data to study modern-day threats. Part III argues that communications privacy law and norms contribute significantly to this shortage. The ECPA, in particular, reinforces the existing cultural resistance to cooperation among cybersecurity researchers by making data sharing among these researchers legally risky. Part IV demonstrates that the dearth of usable data is a serious impediment to research. Increasing cybersecurity researchers' access to such data would significantly aid their research. Part V presents a variety of measures — legal, institutional, and technological — that are necessary to improve communications data sharing with cybersecurity researchers while protecting individual privacy interests in the data. The Article argues that Congress should create a cybersecurity research exception to the ECPA granting formal permission to share communications data for research purposes, subject to strict institutional controls. This change would help confer legitimacy on the use of communications data in research, which, in turn, could shape norms that favor sharing.

II. THE UNIQUE PROMISE OF TECHNICAL RESEARCH IN IMPROVING CYBERSECURITY

A. Defining Cybersecurity

To avoid the possibility that "cybersecurity" will become too malleable a term in this Article, I will provide a definition. Elements of

23. *See* PCIPB, *supra* note 1, at 22.

24. *Id.* at 14–15; *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*.").

cybersecurity familiar to computer scientists include the following: a computer or network system's resistance to becoming unavailable or unusable due to unauthorized uses; resistance to attacks that corrupt data stored on the system and cause information to leak out of the system; and a guarantee that data can be restored after an attack.[25] A somewhat 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;[26] cybersecurity applies these activities to networked computer systems.[27]

Applying either definition to real systems — a necessary step in any discussion of whether a technology or policy is likely to improve

---

25. *See* CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 3; *see also* COMPUTER SCI. & TELECOMMS. BD., NAT'L ACAD. OF SCIS., TRUST IN CYBERSPACE 14 (Fred B. Schneider ed., 1999) [hereinafter CSTB, TRUST IN CYBERSPACE], *available at* http://books.nap.edu/openbook.php?record_id=6161 (defining "security" to mean that a system "resists potentially correlated events (attacks) that can compromise the secrecy, integrity, or availability of data and services").

26. *See, e.g.*, 17 U.S.C. § 1201(e) (2006) ("[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* note 25, at app. K (defining "vulnerability" and "threat").

27. I chose to use the term "cybersecurity," rather than "computer and network security" or "information security" in this article for two reasons. First, it is less cumbersome than "computer and network security" and second, it denotes something more specific than "information security." According to one academic security expert, "cybersecurity" is equivalent to "computer and network security." *See* Matt Bishop, *What Is Computer Security?*, IEEE SECURITY & PRIVACY, Jan./Feb. 2003, at 67, 67 ("Computer and network security, or cybersecurity, are critical issues."). The Department of Homeland Security adds the gloss that cybersecurity pertains to deliberate, malicious attacks on networked information systems, *see* U.S. DEP'T OF HOMELAND SEC., NATIONAL INFRASTRUCTURE PROTECTION PLAN 103 (2006), *available at* http://www.dhs.gov/xlibrary/assets/NIPP_Plan.pdf (defining "cybersecurity" to mean "[t]he prevention of damage to, unauthorized use of, or exploitation of, and, if needed, the restoration of electronic information and communications systems and the information contained therein to ensure confidentiality, integrity, and availability"), leaving the inference that "information security" might refer to a broader category of threats, including unintentional errors.

Others might take issue with this choice of terms. Computer security expert Ed Felten recently conjectured that "cybersecurity" has come to replace the equivalent term "information security" as policymakers and the military have begun attempting to exert more influence over computer and network security. *See* Posting of Ed Felten to Freedom to Tinker, What's the Cyber in Cyber-Security?, http://www.freedom-to-tinker.com/?p=1319 (July 24, 2008, 06:01 EDT). Helen Nissenbaum has argued that "cyber-security" and "computer (and network) security" are terms that should be kept separate. According to Professor Nissenbaum, "computer security" reflects the technical and scientific study of vulnerabilities and attacks on computers and networks. Helen Nissenbaum, *Where Computer Security Meets National Security*, 7 ETHICS AND INFO. TECH. 61, 63 (2005). "Cyber-security," on the other hand, is more closely linked to notions of national security. In addition to the technical matters of computer and network security, this outlook emphasizes the consequences of successful attacks, such as the potential to disable elements of critical infrastructure. *Id.* at 64.

I believe that some focus on the consequences of successful attacks is necessary to understand the risks of insecurity, and this potential for harm is part of my justification for recommending a research exception to federal electronic communications privacy laws.

or harm security — raises the tricky question of how to know whether a particular system is secure. Some within the cybersecurity community have adopted an absolutist perspective: security is a "binary and negative property" that defines "secure . . . as the opposite of being *insecure*."[28] Establishing that a system is secure requires proving that it is free from vulnerabilities. The rationale behind this view is that attackers have the motivation and resources to learn from failed attempts to attack a system; thus, a system with any vulnerability must be viewed as insecure.[29]

However, a more flexible view of security is gaining support among cybersecurity experts. This view holds that the complexity of modern information systems makes it practically, if not theoretically, impossible to prove that a system is vulnerability-free.[30] The question then becomes how to measure security, and this question remains unanswered. A number of metrics offer ways to order the likelihood or severity of identified threats, but none applies to all systems in all contexts.[31] Risk-oriented metrics are gaining favor among cybersecurity experts as a way to compare the security of different systems and evaluate the effectiveness of security policies and technologies.[32]

Given the difficulty of applying a formal definition of security to real information systems, it is not surprising that no technical approach has addressed cybersecurity vulnerabilities, nor does any single approach seem likely to do so.[33] Instead, improving cybersecurity requires a holistic approach that incorporates policy and technology simultaneously. A high-level taxonomy includes four approaches: prevention, deterrence, detection and recovery, and resilience.[34] These approaches are interdependent; progress or setbacks made using one approach can inform activities under the others. The remainder of this Part sketches the strengths and weaknesses of each approach and relationships among them. It also argues that the detection and resilience approaches would clearly benefit from policy reforms.

---

28. CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 133.

29. *See id.* at 45–46.

30. *See id.* at 133 (citing examples of the few domains of problems in which computer scientists have developed formal proofs of security).

31. *See id.* at 135 ("[T]he search for an overall cybersecurity metric — one that would be applicable to all systems and in all environments — is a largely fruitless quest.").

32. *See, e.g.*, Steven M. Bellovin et al., A Clean-Slate Design for the Next-Generation Secure Internet 4 (2005) (unpublished manuscript), *available at* http://www.cs.columbia.edu/~smb/papers/ngsi.pdf (stating that "there has been (to some extent) a [sic] evolution in thinking, from security as an absolute all-or-nothing objective to an approach based on acceptable insecurity and security as risk management").

33. *See* CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 72 (stating that "the simple reality that there is no silver bullet, or even a small number of silver bullets, that will solve 'the cybersecurity problem'").

34. This taxonomy is given in Bellovin et al., *supra* note 32, at 25–26, which attributes it to computer scientist Adrian Perrig.

B. Defending Against Known Threats: The Inadequacy of Prevention

The Internet is far less secure than we know how to make it.[35] Known engineering and management practices can reduce the number of vulnerabilities that technology firms introduce into their products.[36] For many categories of hardware and software, products from different sources differ in their security characteristics, implying that users have some choice about the level of security in the technologies they use. In short, many attacks succeed because technology firms, individual users, and organizations fail to take steps to prevent them.

But preventing all cyberattacks is technically and economically infeasible. On the technical side, it is practically impossible to find all potential vulnerabilities in systems as complex as modern computers.[37] It is also difficult to separate vulnerabilities in individual computers from threats to the Internet.[38] Networks allow attackers to exploit vulnerabilities on individual computers, and individual computers serve as launch pads for network-wide attacks. Threats constantly evolve to exploit newly discovered vulnerabilities, which are sometimes revealed by defenses. Software patches, for example, may offer clues for attacking unpatched systems[39] or introduce new vulnerabilities.[40] It is also difficult to isolate a malicious host from the rest of the Internet, which means that all users are susceptible to at-

35. 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 . . . .").

36. Software developers fail to follow practices that could make their programs more secure. Consider the class of vulnerability known as the "buffer overflow," which is a programming error that may allow an attacker to execute arbitary commands on a remote computer. *See* Sandeep Grover, *Buffer Overflow Attacks and Their Countermeasures*, LINUX J., Mar. 10, 2003, http://www.linuxjournal.com/article/6701. Certain programming languages are not susceptible to buffer overflows, while programming practices and automated tools can greatly reduce the number of such errors in languages that are susceptible. CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 59. Yet approximately half of vulnerabilities entered into a comprehensive national database are attributed to buffer overflows. *Id.* at 59–60.

37. *See* CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 105 ("Model checking, code and program analysis, formal verification, and other 'semantics-based' techniques are becoming practical only for modestly sized real-system software components.").

38. *See* Bellovin et al., *supra* note 32, at 3 ("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.").

39. *See generally* Ashish Arora et al., *Impact of Vulnerability Disclosure and Patch Availability — An Empirical Analysis*, WORKSHOP ON ECON. INFO. SECURITY, May 13, 2004, http://www.dtc.umn.edu/weis2004/telang.pdf (examining how the timing and content of vulnerability disclosure and patch releases affects attackers' ability to derive exploits from this information).

40. CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 60 ("[O]ften patching introduces additional security flaws.").

tacks launched by exploiting the "weakest link" in the network.[41] This dynamic makes prevention a Sisyphean task.

The distributed denial of service ("DDoS") attack in Estonia illustrates these technical difficulties.[42] Botnets — networks of individual computers that have been compromised, have malicious software installed on them, and are centrally controlled by a remote attacker (a "botmaster")[43] — were suspected to be at least a partial cause of the attack.[44] To set up a botnet, attackers exploit vulnerabilities on individual computers to install software that will later respond to the botmaster's commands.[45] The sources of the vulnerabilities are numerous and include operating systems, web browsers, and common applications.[46] The resulting network might contain over a million computers that the botmaster can direct to send spam, send malicious software, or, as was the case in Estonia, conduct a DDoS attack.[47]

The economic dimension of improving cybersecurity also presents challenges. One of the major findings of economic studies of cybersecurity is that individuals and firms underinvest in security be-

---

41. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 7 ("The overall security of a system is only as strong as its weakest link.")

42. *See* Schwartz, *supra* note 7. The flood of traffic in a DDoS attack can consume enough of the target's system resources to render the system unavailable for its intended uses. Depending on the attacker's plans, the response of the target, and the responses of other Internet infrastructure operators, a DDoS attack can last for hours or longer. *See, e.g.*, CLAY WILSON, CONG. RESEARCH SERV., BOTNETS, CYBERCRIME, AND CYBERTERRORISM: VULNERABILITIES AND POLICY ISSUES FOR CONGRESS 7 (2008), *available at* http://www.fas.org/sgp/crs/terror/RL32114.pdf (noting that the series of DDoS attacks against Estonian targets lasted for weeks, causing the repeated shutdown of some websites for several hours at a time or longer).

43. *See* Mark Allman et al., *Fighting Coordinated Attackers with Cross-Organizational Information Sharing*, 2006 REC. WORKSHOP ON HOT TOPICS IN NETWORKS: HOTNETS V 121, *available at* http://www.read.cs.ucla.edu/hotnets5/program.pdf (describing botnets as "armies of enslaved hosts . . . controlled by a single person or small group"); *see also* Moheeb Abu Rajab et al., *A Multifaceted Approach to Understanding the Botnet Phenomenon*, 2006 PROC. ACM SIGCOMM CONF. ON INTERNET MEASUREMENT 41, *available at* http://www.cs.jhu.edu/~fabian/papers/botnets.pdf.

44. Robert Vamosi, *Cyberattack in Estonia — What It Really Means*, CNET NEWS, May 29, 2007, http://news.cnet.com/Cyberattack-in-Estonia-what-it-really-means/2008-7349_3-6186751.html.

45. 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. *See* John Markoff, *Attack of the Zombie Computers Is a Growing Threat, Experts Say*, N.Y. TIMES, Jan. 7, 2007, § 1, at 1.

46. Niels Provos et al., *The Ghost in the Browser: Analysis of Web-based Malware*, WORKSHOP ON HOT TOPICS IN UNDERSTANDING BOTNETS (HOTBOTS '07), Apr. 10, 2007, http://www.usenix.org/event/hotbots07/tech/full_papers/provos/provos.pdf. 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 Cyber Security Bulletin SB07-190 — Vulnerability Summary for the Week of July 2, 2007, http://www.us-cert.gov/cas/bulletins/SB07-190.html (last visited Dec. 19, 2008).

47. *See* Rajab et al., *supra* note 43, at 6; Robert Lemos, *Dutch Bot-net Suspects Infected 1.5 Million PCs, Officials Say*, SECURITYFOCUS, Oct. 20, 2005, http://www.securityfocus.com/brief/19.

cause it is an externality.[48] That is, the security practices of one person can affect the security of others. An externality may be positive, as is the case when a bank uses technology to give customers secure access to their accounts, reducing the chance that an attacker will intercept a bank customer's account information and use it to incur fraudulent charges against online merchants. But externalities can also be negative, as is the case when a vulnerable software product goes unpatched and provides a means to attack other users on the network.[49]

Actions that affect cybersecurity present a mixture of negative and positive externalities.[50] A tilt toward negative externalities may be seen by examining the incentives of the three major categories of actors in cybersecurity: users, technology producers, 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; learning about security imposes a cost from which an individual might not see a benefit.[51] Secure software is typically less convenient to use[52] and from the user's perspective is, at best, only as functional as comparable insecure software.[53] For organizations, quantifying the return on an investment is also difficult, because security successes do not result in an observable positive payoff, and security improvements often spill over to the benefit of other users — including competitors — on the network.[54] As a result, individual as well as large institutional users "tend to un-

---

48. *See* Ross Anderson & Tyler Moore, *The Economics of Information Security*, 314 SCIENCE 610, 610 (2006). For more background on network externalities, see generally Michael L. Katz & Carl Shapiro, *Systems Competition and Network Effects*, 8 J. ECON. PERSP. 93 (1994); Mark A. Lemley & David McGowan, *Legal Implications of Network Economic Effects*, 86 CAL. L. REV. 479, 483–84 (1998); Peter S. Menell, *Tailoring Legal Protection for Computer Software*, 39 STAN. L. REV. 1329 (1987).

49. Harold Demsetz defined positive and negative externalities together:

> [T]he concept [of externality] includes external costs [and] external benefits, . . . pecuniary as well as nonpecuniary . . . . No harmful or beneficial effect is external to the world. Some person or persons always suffer or enjoy these effects. What converts a harmful or beneficial effect into an externality is that the cost of bringing the effect to bear on the decisions of one or more interacting persons is too high to make it worthwhile . . . .

Harold Demsetz, *Toward a Theory of Property Rights*, 57 AM. ECON. REV. 347, 348 (1967).

50. *See* Anderson & Moore, *supra* note 48, at 611.

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

52. *See* CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 7 (noting that cybersecurity measures "interfere[] with daily work").

53. *See id.* at 8 ("[A] secure system doesn't allow users to do any more than an insecure system . . . .").

54. *See id.* at 8 ("[B]ecause serious cyberattacks are rare, the payoff from security investments is uncertain (and in many cases, it is society rather than any individual firm that will capture the benefit of improved security)."); *id.* 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.").

derinvest in security."[55] Indeed, the federal government has failed to be a cybersecurity role model. The U.S. Department of Defense used to publish "The Orange Book,"[56] 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 businesses and individuals would in turn adopt the same practices. Instead, the 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."[57] Like private-sector networks and computers, those owned by the government are highly vulnerable to attack.[58]

The second group of cybersecurity actors — technology producers — responds to these users' tendencies to prefer functionality to security. Building a secure information system requires firms to direct at least some of their engineering efforts toward security rather than toward features that most users more immediately desire.[59]

By contrast, economics favors the third group of cybersecurity actors, the attackers. Since the "overall security of a system is only as strong as its weakest link,"[60] the resources required to defend a system are generally far greater than those necessary to attack it. The weakest link might be technological — a software vulnerability, for example — or it might result from human action, such as a system user giving his password to a person he erroneously believes to have a legitimate need for it. Whatever may cause a breach, the weakest link principle implies that defending a system is much more costly than attacking it.[61] Finally, attackers are highly motivated by financial,

---

55. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 8–9; *see also* Anderson & Moore, *supra* note 48, at 612 ("Although a rational consumer might well spend $20 to prevent a virus from trashing his hard disk, he might not do so just to prevent an attack on someone else.").

56. Formally, this was known as the *Trusted Computer System Evaluation Criteria*. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 9 n.14. The final version of the Orange Book was published on December 26, 1985. The full text is available at http://nsi.org/Library/Compsec/orangebo.txt.

57. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 9.

58. *See, e.g.*, Caron Carlson, *GAO Slams IRS Network Security*, EWEEK.COM, Mar. 27, 2006, http://www.eweek.com/c/a/Government-IT/GAO-Slams-IRS-Network-Security/; Ellen Messmer, *GAO Slams FBI Network Security*, PC WORLD, May 25, 2007, http://www.pcworld.com/article/132250/gao_slams_fbi_network_security.html.

59. *See* Anderson & Moore, *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).

60. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 7.

61. 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." Ross Anderson, *Why Information Security Is Hard — An Economic Perspective*, 2001 PROC. OF THE COMPUTER SECURITY APPLICATIONS CONF. 358, 364, *available at* http://www.cl.cam.ac.uk/~rja14/Papers/econ.pdf.

political, or personal interests. Attackers have long acted to gain prestige,[62] but in recent years financial incentives have become increasingly compelling.[63] And, more recently, attacks have revealed political motives.[64]

C. The Limits of Deterrence

A second element of cybersecurity, deterrence, also faces significant limitations in improving security. Deterrence might take the form of proactive regulation, in which the government sets security standards and punishes those who fail to live up to them. It might also take the form of laws that provide penalties for individuals who commit specific bad acts.

Direct regulation of the information technology sector has been conspicuously absent from the government's approach to improving cybersecurity.[65] On matters of cybersecurity, the government has followed the non-regulatory approach that has marked the course of information technology development over the past few decades. Presently, this shows little sign of changing.[66] This outlook may vary in response to political change as well as the possibility that an attack would prompt more far-reaching regulation,[67] but such an approach would mark a major shift in the government's approach.

---

62. At a relatively early stage the House of Representatives recognized that financial gains might not be the only motivation for 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." H.R. REP. NO. 99-647, at 63 (1986).

63. *See, e.g.*, S. REP. NO. 99-541, at 36 (1986), *reprinted in* 1986 U.S.C.C.A.N. 3555, 3590 (1986) (noting enhanced criminal penalties for violations of 18 U.S.C. § 2701 where the perpetrator acts for private financial gain); GAO, ADDRESSING CYBER THREATS, *supra* note 6, 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.").

64. For example, the recent military conflict between Russia and Georgia was preceded by cyberattacks against Georgian government sites. *See* John Markoff, *Before the Gunfire, Cyberattacks*, N.Y. TIMES, Aug. 12, 2008, at A1. The source of the attacks is not known, and the Russian government denied any involvement. *Id.*

65. However, Congress has enacted legislation creating cybersecurity standards for particular economic sectors, such as financial institutions and health care providers, rather than setting standards that apply directly to technology producers. A full examination of these laws is beyond the scope of this article. *See generally* Aaron J. Burstein, How a Framework for Information Security Law Could Improve Information Security 12 (Jan. 2008) (unpublished manuscript), *available at* http://www.thei3p.org/docs/publications/whitepaper-infoseclaw.pdf ("[I]t is not surprising that there is no generally applicable set of information security regulations for private organizations in the United States. Instead, the primary means of regulating firms' information security practices is through sector-specific statutes and regulations that prohibit disclosures of certain kinds of information.").

66. PCIPB, *supra* note 1, at 15 ("[F]ederal regulation will not become a primary means of securing cyberspace.").

67. Jonathan Zittrain has argued forcefully that cybersecurity might be the "fulcrum" that spurs extensive regulation of technology. *See* Zittrain, *supra* note 4, at 2003.

By contrast, Congress and law enforcement agencies have given considerable attention to deterring individual bad actors by punishing various forms of cybercrime. Early in the era of networked information systems, Congress responded to malicious attacks committed over networks by defining expansive new crimes.[68] Since then, Congress has updated communications privacy laws to give law enforcement officials easier access to communications and records that facilitate cybercrime prosecutions.[69] Finally, federal agencies continue to make law enforcement a high priority in cybersecurity policy.[70]

---

68. *See, e.g.*, Computer Fraud and Abuse Act of 1986 ("CFAA"), Pub. L. No. 99-474, 100 Stat. 1213 (codified as amended at 18 U.S.C. § 1030 (2006)); S. REP. NO. 104-357, at 6–14 (1996) (explaining that the "CFAA" amendments were intended to facilitate prosecutions); S. REP. NO. 99-432, at 2 (1986) ("The proliferation of computers and computer data has spread before the nation's criminals a vast array of property that, in many cases, is wholly unprotected against crime."); Orin S. Kerr, *Cybercrime's Scope: Interpreting "Access" and "Authorization" in Computer Misuse Statutes*, 78 N.Y.U. L. REV. 1596, 1615–24 (2003).

69. For example, the USA PATRIOT Act created a "computer trespasser" exception to Title III of the Omnibus Crime Control and Safe Streets Act of 1968 ("Wiretap Act"), Pub. L. No. 90-351, §§ 801–804, 82 Stat. 197, 211–25, which allows law enforcement officials to intercept electronic communications being routed to a specific computer if the owner gives his or her authorization. *See* United and Strengthening America by Providing Appropriate Tools Required to Intercept and Obstruct Terrorism (USA PATRIOT) Act of 2001, Pub. L. No. 107-56, § 217, 115 Stat. 272, 291 (codified as amended at 18 U.S.C. § 2511(2) (2006)). Generally, the Wiretap Act prohibits anyone from intentionally intercepting electronic communications. It is discussed more extensively in Part III of this article.

The vast quantities of data that Internet service providers and many websites keep are also available to law enforcement, with barriers that range from moderate to low. *See generally* Orin S. Kerr, *A User's Guide to the Stored Communications Act, and a Legislator's Guide to Amending It*, 72 GEO. WASH. L. REV. 1208 (2004) (discussing statutory limits on the government's ability to compel providers to disclose information in their possession regarding their customers); Daniel J. Solove, *Reconstructing Electronic Surveillance Law*, 72 GEO. WASH. L. REV. 1264 (2004) (discussing the interplay between law enforcement and surveillance laws). Still, some members of Congress and the Department of Justice would argue that additional surveillance powers are necessary. A bill recently introduced in Congress would require Internet service providers to retain data in a manner consistent with regulations issued by the Department of Justice. *See* Internet Stopping Adults Facilitating the Exploitation of Today's Youth (SAFETY) Act of 2007, H.R. 837, 110th Cong. § 6 (2007), *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_bills&docid=f:h837ih.txt.pdf.

70. The nexus between cybersecurity and law enforcement is evident in the USA PATRIOT Act's "computer trespasser" exception to the Wiretap Act as well as the data retention requirement proposed in the Internet SAFETY Act. *See* USA PATRIOT Act § 217; Internet SAFETY Act § 6. The computer trespasser exception exempts law enforcement officials from the Wiretap Act's warrant requirements, provided the owner of the computer under attack authorizes the interception of communications. 18 U.S.C. § 2511(2). The Internet SAFETY Act's data retention requirement, though packaged with a concern for combating child pornography, would appear to make data available for cybersecurity-related investigations. *See* Internet SAFETY Act § 6 (leaving the issue of any limits to use of retained data to Attorney General's regulations). Finally, the Identity Theft Enforcement and Restitution Act of 2007 would further broaden the CFAA by lowering the damage threshold for defining an offense, and adding offenses for cyber-extortion and conspiracy to commit cybercrimes. S. 2168, 110th Cong. §§ 5–7, *available at* http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=110_cong_bills&docid=f:s2168is.txt.pdf.

Despite these favorable legal conditions, criminal prosecutions of cybercrime face a number of practical limitations. The potential presence of evidence and suspects in countries that are reluctant to cooperate with investigators can contribute to a high level of difficulty and expense in cybercrime prosecutions. Agency resources are limited, and enforcement has slanted heavily toward crimes such as copyright infringement, targeted computer break-ins, and financial fraud.[71] Law enforcement officials also face the challenge of understanding rapidly changing cybersecurity threats. These resource constraints strongly suggest that deterrence is an incomplete solution in cybersecurity policy.

D. Adapting to Evolving Threats Through Detection and Resilience: The Case for Focusing on Technical Research

The related approaches of detection and resilience brighten this gloomy picture. From a scientific standpoint, new methods of detecting network-based attacks address some of the technical and economic difficulties of prevention.[72] A major contributor to the brittleness of prevention is that it is impossible to identify all threats to a system in advance. Indeed, certain classes of attacks, such as DDoS attacks, become evident only when one has a view of traffic flowing across large portions of the Internet. No single organization or user, with the possible exception of Internet backbone providers,[73] is likely to have such a broad view of the network.[74]

Attack detection research is exploring technologies that integrate the views from many organizations to identify these types of malicious activity.[75] These methods are becoming increasingly sophisticated at detecting malicious traffic "on the fly," without depending on matching a traffic pattern with one that is known to be malicious.[76] In addition, by depending on centralized analysis of network traffic, the detection approach mitigates the weakest link problem associated with prevention. Unlike other approaches that depend on making all computers on the network secure, this approach isolates infected systems in order to contain attacks.[77] Such "graceful degradation" of a network

71. *See, e.g.*, GAO, ADDRESSING CYBER THREATS, *supra* note 6, at 15–18.

72. *See supra* Part II.B.

73. *See* FTC, BROADBAND CONNECTIVITY COMPETITION POLICY 30 (2007) ("Generally, individual backbone networks are made up of a multiplicity of redundant, high-speed, high-capacity, long-haul, fiber-optic transmission lines that join at hubs or points of interconnection across the globe."), *available at* http://www.ftc.gov/reports/broadband/v070000report.pdf.

74. *See* Allman et al., *supra* note 43, at 1.

75. *See, e.g.*, Xie et al., *supra* note 12.

76. *See* Allman et al., *supra* note 43, at 6.

77. *See* CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 200.

is central to the concept of resilience.[78] The detection approach, therefore, enhances the resilience of a network.

Currently, the promise of these combined approaches largely remains a prospect for improving cybersecurity at some point in the future. These methods are mostly the subject of technical research and have not been widely deployed. The lack of data from actual networks remains a major obstacle to these avenues of research. Organizations possessing such data, such as ISPs and universities, are reluctant to share it because of concerns about users' rights and expectations under communications privacy laws. In addition, there are only a few institutions that enable data sources to share data with researchers under the strictly controlled conditions necessary for advancing cybersecurity research over the long term.

Relative to other scientific fields, cybersecurity research is in 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 this data. An individual's privacy interests in her medical records are strongly held for a number of reasons. The information in medical records is strongly connected to the core of personhood, and personal autonomy dictates that individuals should have control over this information.[79] There is also a utilitarian justification for medical privacy. If individuals do not believe that their medical information will be kept confidential, they may be less likely to provide truthful, complete information in the first place.[80] As a result, their health care might suffer and widespread refusal to provide truthful medical information could adversely impact the system as a whole. A variety of federal and state laws address medical information confidentiality.[81] The laws that offer this protection also con-

78. Bellovin et al., *supra* note 38, at 26 (defining resilience as "maintain[ing] a certain level of availability or performance even in the face of active attacks").

79. *See generally* 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, 1441 (2002).

80. *See* JANLORI GOLDMAN & DEIRDRE K. MULLIGAN, PRIVACY AND HEALTH INFORMATION SYSTEMS: A GUIDE TO PROTECTING PATIENT CONFIDENTIALITY 9 (1996).

81. At the federal level, the Health Insurance Portability and Accountability Act of 1996 ("HIPAA") establishes the framework for the protection of personal health information ("PHI"). Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.). The standards for protecting PHI are spelled out in the HIPAA "Privacy Rule," which the U.S. Department of Health and Human Services enacted pursuant to HIPAA. Standards for Privacy of Individually Identifiable Health Information, 45 C.F.R. §§ 160–164 (2007). *See generally* HIPAA § 264, 110 Stat. at 2033 (codified at 42 U.S.C. § 1320d-2 (2000)) (granting the Secretary of Health and Human Services authority in certain circumstances to issue regulations addressing privacy rights in PHI). For an overview of state constitutional, statutory, and common law protection for personal health information, see James G. Hodge, Jr., *The Intersection of Federal Health Information Privacy and State Administrative Law: The Protection of Individual Health Data and Workers' Compensation*, 51 ADMIN. L. REV. 117, 128–32 (1999).

tain a number of exceptions.[82] These exceptions are based on a number of public or common needs, including the social value of medical research.[83]

Cybersecurity research does not receive such explicit legal support. There are no exceptions to communications privacy laws that would afford researchers greater access to electronic communications, despite the potential benefits of expanding access. 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."[84] Other leading researchers have argued that greater access to real network traffic datasets would "cause a paradigmatic shift in computer security research."[85] Some researchers have also adapted their approaches to reflect their inability to obtain data that provides a sufficiently broad view of network events.[86] Institutions for collecting network data for research purposes and coordinating researchers' access to the data are basically non-existent. One academic Internet research group[87] has made an extensive effort to establish a repository of this 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."[88] Other researchers have called for network data repositories to serve cybersecurity researchers in the model of a "Cyber-Center for Disease Control," and they note that this data sharing endeavor "raises potentially immense policy issues concerning privacy."[89]

Congress has recognized the central role of sharing data in advancing research, resting its appropriation for cybersecurity research in part on a finding that "[f]ederal investment in computer and network security research and development must be significantly in-

---

82. *See, e.g.*, HIPAA Privacy Rule, 45 C.F.R. § 164.512(b) (authorizing disclosures of PHI for public health activities); CAL. CIV. CODE §§ 56.10(b)–(c) (West 2008) (authorizing disclosure of PHI for various purposes, including research).

83. *See* Gostin & Hodge, *supra* note 79.

84. Ruzena Bajcsy et al., *Cyber Defense Technology Networking and Evaluation*, COMMC'NS ACM, Mar. 2004, at 58, 58.

85. Phillip Porras & Vitaly Shmatikov, *Large-Scale Collection and Sanitization of Network Security Data: Risks and Challenges*, 2006 PROC. NEW SECURITY PARADIGMS WORKSHOP 57, *available at* http://www.cs.utexas.edu/~shmat/shmat_nspw06.pdf.

86. *See* Xie et al., *supra* note 12.

87. Cooperative Association for Internet Data Analysis ("CAIDA"), http://www.caida.org/home/ (last visited Dec. 19, 2008).

88. COOP. ASS'N FOR INTERNET DATA ANALYSIS, TOWARD COMMUNITY-ORIENTED NETWORK MEASUREMENT INFRASTRUCTURE: PROJECT SUMMARY 1 n.1 (2005) [hereinafter CAIDA], http://www.caida.org/funding/cri2005/nsfcri_2005.pdf.

89. Stuart Staniford, Vern Paxson & Nicholas Weaver, *How to 0wn the Internet in Your Spare Time*, 2002 PROC. USENIX SECURITY 149, 162–63, *available at* http://www.usenix.org/publications/library/proceedings/sec02/full_papers/staniford/staniford.pdf.

creased to . . . better coordinate information sharing and collaboration among industry, government, and academic research projects."[90] The National Science Foundation and Department of Homeland Security have also identified increasing the availability of network data as a critical priority.[91] Still, these agencies have not squarely addressed the threats to privacy arising from increased network data sharing among researchers. The remainder of this Article clarifies and attempts to resolve these privacy issues.

III. HOW COMMUNICATIONS PRIVACY LAW LIMITS CYBERSECURITY RESEARCH

Cybersecurity researchers have identified the dearth of accessible cybersecurity data as a problem.[92] Communications privacy laws begin to explain this dearth of data, but neither the explanation for nor the solution to this problem stops with an examination of these laws. Even if privacy laws did not prohibit sources of data, such as Internet service providers ("ISPs"), from disclosing data to cybersecurity researchers, a variety of institutional factors also inhibit data sharing. This Part identifies precisely how communications privacy laws, privacy norms, and the outlooks of organizations that handle communications data combine to inhibit data sharing.

A. Communications Privacy Law

The ECPA's approach to cybersecurity 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.[93] These single firms may, under some circumstances, disclose relevant data to law enforcement agencies to assist in prosecutions.[94] The threats described

90. Cyber Security Research and Development Act, Pub. L. No. 107-305, § 2(5)(C), 116 Stat. 2367, 2368–70 (2002) (codified at 15 U.S.C § 7401 (2006)).

91. *See* CAIDA, *supra* note 88.

92. *See* Bajcsy et al., *supra* note 84.

93. For example, the provider exception to the Wiretap Act, 18 U.S.C. § 2511(2)(a)(i) (2006), authorizes providers of electronic communications services to intercept and disclose communications to law enforcement officials when it is necessary to protect that provider's rights or property. This exception is limited to protection of a service provider's own rights or property; it does not cover disclosures of communications that pertain to threats to other providers' rights or property. *See* GINA STEVENS & CHARLES DOYLE, CONG. RESEARCH SERV., PRIVACY: AN OVERVIEW OF FEDERAL STATUTES GOVERNING WIRETAPPING AND ELECTRONIC EAVESDROPPING 15 (2001), *available at* http://fas.org/sgp/crs/intel/98-326.pdf ("The exemption [18 U.S.C. § 2511(2)(a)(i)] . . . lets the telephone company protect *itself* against fraud . . . ." (emphasis added)).

94. 18 U.S.C. § 2511(2)(a)(i); *id.* § 2702(b)(5) (permitting the provider of an electronic communications service to disclose the contents of a stored communication to protect its "rights or property"); *id.* § 2702(c)(3) (permitting the provider of an electronic communica-

in Part II 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 but not research is at odds with cybersecurity policy priorities.

The ECPA is a notoriously complex set of statutes,[95] which this Article not attempt to describe fully here. Instead, the Article sketches the types of data that the ECPA regulates and emphasizes how the ECPA's structure relates to the institutional and economic hurdles that prevent 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;[96] the Stored Communications Act ("SCA"), which regulates disclosure of electronic communications contents as well as addressing information;[97] and the Pen/Trap statute, which regulates the real-time collection of communications addressing information.[98]

1. Wiretap Act

The single-firm view of data privacy originated with the enactment of Title III of the Omnibus Crime Control and Safe Streets Act of 1968 ("Wiretap Act").[99] As amended by the ECPA, the Wiretap

---

tions service to disclose records or information pertaining to a customer in order to protect the provider's "rights or property").

95. *See, e.g.*, United States v. Smith, 155 F.3d 1051, 1055 (9th Cir. 1998) (commenting that "the intersection of the Wiretap Act and the Stored Communications Act is a complex, often convoluted, area of the law" (citations omitted)); Steve Jackson Games, Inc. v. U.S. Secret Serv., 36 F.3d 457, 462 (5th Cir. 1994) (noting that the Wiretap Act "is famous (if not infamous) for its lack of clarity"); *In re* Application for Pen Register & Trap/Trace Device with Cell Site Location Auth., 396 F. Supp. 2d 747, 753 (S.D. Tex. 2005) (noting, in the context of an application to install a pen register device, that "rigorous attention must be paid to statutory definitions 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, 820–24 (2003).

96. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, §§ 101–111, 100 Stat. 1848, 1848–59 (addressing "interception of communications and related matters" and amending 18 U.S.C. §§ 2510–21).

97. *Id.* §§ 201–202, 100 Stat. at 1860–68 (creating a new chapter in Title 18 to regulate "store wire and electronic communications and transactional records access"); 18 U.S.C. § 2702(a)(1)–(2) (prohibiting disclosure of the "contents" of electronic communications by an electronic communications service ("ECS") or a remote computing service ("RCS")); *id.* § 2702(a)(3) (prohibiting an ECS or RCS from divulging "a record or other information pertaining to a subscriber to or customer of such service").

98. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, §§ 301–302, 100 Stat. 1848, 1868–72 (creating a new chapter in Title 18 to regulate "pen registers and trap and trace devices").

99. Pub. L. No. 90-351, §§ 801–804, 82 Stat. 197, 211–25 (codified as amended at 18 U.S.C. §§ 2510–2522 (2006)). For the historical context of the Wiretap Act, see Deirdre K. Mulligan, *Reasonable Expectations in Electronic Communications: A Critical Perspective on the Electronic Communications Privacy Act*, 72 GEO. WASH. L. REV. 1557, 1561 n.26 (2004).

Act prohibits anyone from intentionally intercepting electronic communications, such as e-mail.[100] This prohibition applies to the government, individuals, and private firms, such as the ISPs, that provide individuals and businesses with Internet access.[101] The Wiretap Act's breadth reflects an "overriding congressional concern" with protecting communications contents from eavesdropping.[102] Part of the rationale for the ECPA was to extend statutory privacy protection to communications whose constitutional protection was unclear and which at that time fell outside the Wiretap Act's scope.[103]

Though the interception prohibition applies to a broad set of interceptors,[104] it is not absolute. The Wiretap Act permits interceptions to proceed under search warrants,[105] and also under a few other statutory exceptions. The exceptions that are most relevant to providing cybersecurity researchers with access to network data are (1) the presence of consent[106] and (2) the "provider exception," which allows an employee of an "electronic communication service"[107] to "intercept, disclose, or use" 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."[108]

The provider exception affords cybersecurity researchers with access to communications contents. This access is limited, however, and particularly so in cases where researchers from multiple organizations seek to share data. 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.[109] The statute limits the *monitoring* of wire communications to "mechanical or service

---

100. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (codified as amended in scattered sections of 18 U.S.C.). The basic prohibition on intercepting electronic communications is found at 18 U.S.C. § 2511(1).

101. *See* 18 U.S.C. § 2511 (creating a blanket prohibition on interceptions, subject to specific exceptions).

102. Gelbard v. United States, 408 U.S. 41, 48 (1972) (citing S. REP. NO. 90-1097, at 66 (1968)).

103. *See* S. REP. NO. 99-541, at 3 (1986) (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), to suggest that electronic copies do not receive constitutional protection). In *Miller*, the Supreme Court held that the Fourth Amendment did not protect a bank's customer from having his electronic bank records disclosed to law enforcement officials. *Miller*, 425 U.S. at 440.

104. Some commentators nevertheless criticize the Wiretap Act based on its narrow definition of "interception" and its exclusion of certain kinds of surveillance altogether. *See, e.g.*, Solove, *supra* note 69, at 1280–82.

105. *See* 18 U.S.C. §§ 2515–2517.

106. *Id.* § 2511(2)(c). This Article defers discussion of consent to Part III.A.5.

107. *Id.* § 2510(15) (defining this term to mean "any service which provides to users thereof the ability to send or receive wire or electronic communications").

108. *Id.* § 2511(2)(a)(i).

109. U.S. DEP'T OF JUSTICE, SEARCHING AND SEIZING COMPUTERS AND OBTAINING ELECTRONIC EVIDENCE IN CRIMINAL INVESTIGATIONS 125–29 (2002), *available at* http://www.usdoj.gov/criminal/cybercrime/s&smanual2002.pdf.

quality control checks."[110] It is unclear at what point a service provider's interceptions would constitute "monitoring" and thus be subject to this limitation.[111] Given the lack of case law clarifying the provider exception clause, it is understandable that service providers may be reluctant to allow staff researchers to intercept communications.

The second clause of the provider exception also limits the exception's usefulness in providing cybersecurity researchers with access to network data. Permitting interceptions to "protect the rights or property" of the provider does not allow "unlimited" interceptions;[112] rather, there must be a "substantial nexus" between the monitoring and the threat to the provider's rights or property.[113] Courts have not fully elaborated the kinds of threats that would allow a provider to intercept communications under this exception, but monitoring the network for employee fraud, at least, is within the scope of the exception.[114] This finding is similar to cases involving wire communications.[115] However, the ECPA's legislative history also suggests that the limits on monitoring electronic communications are looser than they are for wire communications.[116]

Still, 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.[117] 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. That is, since 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 sub-

---

110. 18 U.S.C. § 2511(2)(a)(i).

111. *See* U.S. DEP'T OF JUSTICE, *supra* note 109, 125–29 (stating that "[t]his language permits providers to intercept, use, or disclose communications in the ordinary course of business *when the interception is unavoidable*" (emphasis added)).

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

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

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

115. *See, e.g.*, United States v. Pervaz, 118 F.3d 1, 5 (1st Cir. 1997) (approving use of the provider exception by a phone company conducting its own investigation into theft of service or fraud); United States v. Villanueva, 32 F. Supp. 2d 635, 639 (S.D.N.Y. 1998) (same).

116. *See* H.R. REP. NO. 99-647, at 47 (1986) ("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 . . . do not involve humans listening in on voice conversations. Accordingly, they are not prohibited.").

117. Richard Salgado, *Legal Issues*, *in* KNOW YOUR ENEMY: LEARNING ABOUT SECURITY THREATS 225, 230–31 (Honeynet Project ed., 2d ed. 2004), *available at* http://www.honeynet.org/book/Chp8.pdf.

scribers probably are not covered by this part of the provider exception.

Whether the provider exception applies to the disclosure of the data, as opposed to its mere 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 intercepts electronic communications contents under the provider exception, disclosing the contents to outside researchers might stretch the requirement of protecting the original service provider's rights or property. One possible way to satisfy the substantial nexus requirement would be for a provider to bring in an outside expert to monitor traffic relating to a threat to the provider's network. An outside researcher's interest, however, lies in developing new methods, which may or may not be effective in detecting threats against network equipment or services. The connection may therefore be too attenuated to satisfy the requirement. Instead, the Wiretap Act allows disclosure to law enforcement officials.[118]

2. Stored Communications Act

A more permissive statutory scheme, the SCA, applies to accessing communications that are in storage, rather than in transit from source to destination.[119] The SCA distinguishes between the contents of a communication and non-content information.[120] Contents of an electronic communication refers to "any information concerning the substance, purport, or meaning of that communication."[121] Non-content information includes records pertaining to a subscriber or user of an electronic communications service.[122] Logs of IP addresses that a user has reached, as well as the "to" and "from" fields in e-mail records are also considered non-content records.[123] Whether other records, such as textual Web addresses ("URLs") that contain search engine queries, are content or non-content records is still a matter of debate.[124]

---

118. *See Villanueva*, 32 F. Supp. 2d at 639. Note, however, that law enforcement officials may not direct the employees of a service provider to monitor a network unless they have a warrant.

119. 18 U.S.C. §§ 2701–2712 (2006).

120. *See id.* §§ 2702–2703 (setting forth different voluntary and required disclosure rules for contents of a communication and non-contentnon-content records).

121. *Id.* § 2510(8).

122. *Id.* § 2702(a)(3).

123. *Cf.* United States v. Forrester, 495 F.3d 1041, 1048–49 (9th Cir. 2007) (concluding that "to" and "from" fields are non-content records in the context of the Pen/Trap statute).

124. *See* Brief of Amici Curiae Law Professors Requesting Additional Briefing If This Court Addresses Google's ECPA Defense at 4, Gonzales v. Google, Inc., 234 F.R.D. 674 (N.D. Cal. 2006) (No. 5-06-mc-80006-JW), *available at* http://www.cdt.org/security/20060224law-profs-amicus.pdf (asserting that there is no case law on the ECPA's applica-

In contrast to the Wiretap Act, the SCA permits nearly unrestricted use of communications contents and records *within* a service provider.[125] Thus, the SCA does not present an obstacle to cybersecurity researchers obtaining data controlled by the organizations that employ them.

Disclosures of this data to persons outside the service provider, however, may be regulated by the SCA.[126] The extent of regulation depends on two factors: whether the provider discloses communications contents or non-content records, and whether the recipient is a governmental entity.[127]

The SCA prohibits the voluntary disclosure of communications contents by a service provider to any other person, subject to the exceptions discussed below.[128] The restrictions on voluntarily disclosing non-content records are far looser due to two limitations in the SCA. The first limitation is that only an entity that provides service "to the public" is covered by this part of the SCA.[129] The second is that, even if a service provider falls under these voluntary non-content record disclosure restrictions, it may provide records to any recipient other than a governmental entity.[130]

For cybersecurity researchers, the definition of a "governmental entity" is critical because many researchers are employed by state universities or national labs. The ECPA does not define a governmental entity, nor does a definition appear in Title 18 of the United States Code, but courts have interpreted the phrase to include an extremely broad array of government agencies. The Seventh Circuit, for exam-

---

tion to URLs or search queries stored by a company that provides electronic communications service).

125. *See* 18 U.S.C. § 2702(b)(4) (allowing disclosure of the contents of a communication "to a person employed or authorized or whose facilities are used to forward such communication to its destination"). The SCA applies only to "electronic communication services" ("ECS") and "remote computing services" ("RCS"). *See id.* § 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" 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 divided into 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 of the main discussion, I am concerned only with an ECS and use this term interchangeably with "service provider," unless otherwise noted.

126. The SCA focuses on the identity of the persons who disclose and receive data; it essentially ignores the terms under which the data exchange takes place. That is, as far as the SCA is concerned, it does not matter whether data is exchanged as part of a commercial transaction or as part of an informal, non-commercial relationship. *See infra* note 138.

127. 18 U.S.C. § 2702(b)–(c).

128. *Id*. § 2702(a). 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 Kerr, *supra* note 69.

129. 18 U.S.C. § 2702(a)(3).

130. *Id.* § 2702(c)(6).

ple, has stated that the use of "governmental entity" in the ECPA "is considerably broader than 'the federal government'" and serves to "distinguish the public from the private sector."[131] The court did specify that the ECPA used the term in order to "attach[] a price tag" to the use of government power to compel private parties to produce information.[132] But in the SCA, and the ECPA as a whole, it is far from clear that a public sector entity must have compulsory powers to be a governmental entity. Had Congress intended to limit disclosures of non-content information only to public sector entities that have compulsory powers, it could have used narrower language, such as "investigative or law enforcement entities."[133]

This statutory structure creates an odd result for cybersecurity researchers. Put simply, a service provider may share non-content data with a researcher from a private university, but sharing the same data with a researcher from a public university raises a serious question under the SCA. Though this result may be consistent with the purpose of extending Fourth Amendment protections against government intrusions into the realm of electronic communications,[134] it does little to protect individual privacy. On one hand, government cybersecurity researchers are unlikely to *use* this communications data differently than would a researcher within the service provider firm or a researcher at another private firm. On the other hand, the SCA permits voluntary disclosure of records to protect the "rights or property" of the provider. Disclosures for this purpose are more likely to result in invasive investigative practices. 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.

Like the Wiretap Act, the SCA contains a "provider exception" that permits some disclosure of communications contents when neces-

131. Ameritech Corp. v. McCann, 403 F.3d 908, 912 (7th Cir. 2005).

132. *See id.* at 912–13 (citing administrative grand jury, and trial subpoenas as examples of the government's compulsory powers).

133. *See* 18 U.S.C. § 2510(7) (defining an "[i]nvestigative or law enforcement officer" as an individual "who is empowered by law to conduct investigations of or to make arrests for offenses enumerated" in the ECPA).

134. *See* U.S. CONST. amend. IV. The Senate 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.

S. REP. NO. 99-541, at 1–2 (1986). The Senate Report continues: "Most importantly, the law must advance with the technology to ensure the continued vitality of the fourth amendment. Privacy cannot be left to depend solely on physical protection, or it will gradually erode as technology advances." *Id.* at 5.

sary to protect the "rights or property" of the provider.[135] There are few, if any, cases interpreting this exception. Still, the similarity between the provider exception in the SCA and in the Wiretap Act suggests a similar purpose and scope: to allow service providers to monitor their systems for threats to their own rights or property.[136] As with the analogous exception in the Wiretap Act, the SCA's provider exception envisions that individual firms are well positioned to detect threats against them and that disclosure to law enforcement agencies is the appropriate way to handle such threats.[137] The SCA does not restrict the use of communications within an electronic communication services firm; rather, the SCA focuses solely on disclosure, whether voluntary, compelled, or resulting from some kind of breach in a service provider's access controls.[138]

3. Pen/Trap Statute

The third and final title of the ECPA is the Pen/Trap statute, which is the non-content counterpart to the Wiretap Act.[139] The statute's name refers to devices that collect incoming addressing information (trap and trace devices) and outgoing addressing information (pen registers).[140]

---

135. 18 U.S.C. § 2702(b)(5).

136. *See* Kerr, *supra* note 69, at 1221 n.91.

137. *See* U.S. DEP'T OF JUSTICE, *supra* note 109, at 225 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 non-content records to government agents "if such disclosure protects the [Provider]'s rights and property").

138. *See* 18 U.S.C. §§ 2702–2710 (defining conditions and process for voluntary and compelled disclosure of communications contents and non-content records). The SCA also 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." *Id.* § 2701(a). This provision holds liable the person who obtains unauthorized access to a stored communication, rather than the communications service provider. Indeed, the provider of the electronic communications service may access stored communications. *Id.* § 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) (finding that individuals on a website's list of eligible users authorized them to give consent for use of the website on behalf of the website's owner); *In re* DoubleClick Inc. Privacy Litig., 154 F. Supp. 2d 497, 514 (S.D.N.Y. 2001) (finding that website owners utilizing DoubleClick's targeted advertising service consented to DoubleClick's interception by use of such service).

139. 18 U.S.C. §§ 3121–3127.

140. *Id.* § 3127(3)–(4) (defining "pen register" and "trap and trace" device, respectively). Because the use of both types of devices is regulated by this statute, it is sometimes called the "Pen/Trap statute." *See also* U.S. DEP'T OF JUSTICE, *supra* note 109, 112–14.

The Pen/Trap statute regulates the real-time collection of communications addressing information.[141] 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.[142] As stated above, addressing information includes essentially all non-content information about a particular communication, such as IP addresses[143] and the "to" and "from" fields in e-mail messages.[144] It is unclear whether uniform resource locators ("URLs") — the addresses that most Internet users use to connect to websites — are addressing information or contents.[145]

The Pen/Trap statute follows the Wiretap Act's approach; it applies to all persons but creates exceptions for a service provider's internal use and for limited government access to addressing information. The statute permits a service provider to collect addressing information in the ordinary course of business.[146] 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."[147]

The Pen/Trap statute does not offer a clear path for giving cybersecurity researchers access to real-time non-content data. The statute authorizes service providers to install pen registers to protect their users from abuse or to protect the provider's rights or property.[148] Like the provider exceptions in the Wiretap Act and the SCA, the service provider's own security is the trigger for 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.[149]

---

141. *See* 18 U.S.C. § 3121(a) (prohibiting any person from installing a pen register or a trap and trace device).

142. *Id*. §§ 3121–3123. *See also* Brown v. Waddell, 50 F.3d 285, 287 (4th Cir. 1995).

143. *See generally In re* Application of the United States for an Order Authorizing the Use of a Pen Register and Trap on [xxx] Internet Service Account/User Name [xxxxxxxxxxx@xxx.com], 396 F. Supp. 2d 45, 48–49 (D. Mass. 2005) (regarding IP addresses as addressing information, not content).

144. *See* Solove, *supra* note 69, at 1287 (concluding that addressing information includes the "To:" and "From:" fields in an e-mail message).

145. Indeed, classification of URLs might depend on the particular URL. *See In re* Application of the United States, 396 F. Supp. 2d at 48–49 (requiring a trap and trace order to list data that the recipient Internet service provider would be prohibited from disclosing because the URLs in the list that contain search terms "would reveal content").

146. *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").

147. *Id.* §§ 3122–3123.

148. *Id.* § 3121(b)(1).

149. *See id.* § 3121(b)(1) (permitting service providers to install a pen register or tap and trace device "relating to the protection of rights or property of [the] provider, or to the protection of users of that service from abuse of service or unlawful use of service"); *id.*

The Pen/Trap statute therefore provides little guidance about whether, and under what conditions, it is permissible to disclose addressing information to cybersecurity researchers. One possible standard is that any recorded addressing information becomes a non-content record subject to the disclosure provisions of the SCA.[150] In that case, a service provider might voluntarily disclose the addressing information to law enforcement officials to protect its "rights or property."[151] But this restriction probably would not permit disclosure to cybersecurity researchers affiliated with a governmental entity. Internal research uses of the data would be permissible, for cybersecurity purposes or otherwise, by analogy to the internal use of non-content records under the SCA.[152]

Alternatively, the Pen/Trap statute, by failing to prohibit disclosure, could be read to authorize 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 statute's exceptions. However, these exceptions are triggered by concerns far broader than provider security. Any collection of addressing information "relating to the operation, maintenance, and testing"[153] would suffice to authorize disclosure. This reading of the statute would effectively gut the non-content provisions of the SCA. The creation of the Pen/Trap statute and the SCA through the same act of Congress makes this interpretation unlikely.[154]

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.[155] Though most of these statutes offer ap-

§ 3121(b)(2) (permitting providers to install pen/trap devices "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").

150. *See supra* Part III.A.2 (noting that a service provider's records of its customers' Internet usage are likely within the SCA's definition of non-content records).

151. 18 U.S.C. § 2702(c)(3). Note that a service provider that does not provide service to the public may disclose non-content 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).

152. *See supra* Part III.A.2.

153. 18 U.S.C. § 3121(b)(1).

154. *Cf.* United Sav. Ass'n of Tex. v. Timbers of Inwood Forest Assocs., 484 U.S. 365, 371 (1988) ("Statutory construction, however, is a holistic endeavor. A provision that may seem ambiguous in isolation is often clarified by the remainder of the statutory scheme . . . .").

155. *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 Prose-*

proximately the same level of protection as the Wiretap Act for communications,[156] some are more protective.[157] California, for example, requires that all parties to a communication consent to its interception,[158] whereas the Wiretap Act provides a one-party consent rule.[159]

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.[160] State communications privacy laws have the greatest impact on the question of defining researchers' access to cybersecurity data when the state laws are more restrictive than federal law. This Article discusses ways to address the complicating effect of state privacy law on cybersecurity research in Part V.

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 it 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 this is not the only use that providers make of this data. This situation leaves a large gap between industry and academic norms and users' understanding of data retention and use. Google's announcement in March 2007 that it would limit its retention of individuals' search histories to eighteen months illustrated this gap.[161] This announcement seemed to serve as public notice of how exten-

cution, 28 SEATTLE U. L. REV. 955, 965 & n.58 (2005) (noting that all states but Vermont have adopted a statutory equivalent of the Wiretap Act).

156. *Id.* at 965 & n.59.

157. Several state courts and at least one federal court have found that state wiretap statutes must be at least as protective as the Wiretap Act. *See id.* at 966 & n.65 (citing Commonwealth v. Vitello, 327 N.E.2d 819, 834 (Mass. 1975), People v. Conklin, 522 P.2d 1049, 1056 (Cal. 1974), and United States v. Mora, 821 F.2d 860 (1st Cir. 1987)).

158. CAL. PENAL CODE § 632(a) (West 2008) (defining an offense for intercepting "intentionally and without the consent of *all* parties to a confidential communication" (emphasis added)).

159. 18 U.S.C. § 2511(2)(c)–(d) (2006).

160. Gostin & Hodge, *supra* note 79, at 1465–66 (discussing the effects of the lack of federal preemption in the context of health information disclosure rules upon public health and medical research).

161. *See* Posting of Peter Fleischer to Official Google Blog, How Long Should Google Remember Searches?, http://googleblog.blogspot.com/2007/06/how-long-should-google-remember.html (June 11, 2007, 22:08 PDT).

sively Google retains data,[162] yet it is unclear whether users in general gained from the announcement a better understanding of the company's data retention and usage practices.[163]

Two recent cases illustrate the ECPA's lack of controls on retention and internal use. When the Recording Industry Association of America ("RIAA") in July 2002 began suing 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 customers assigned a particular network address on a particular day and time.[164] Although Verizon fought the subpoena on a number of grounds, Verizon did not argue that it did not have the information that the RIAA sought.[165] Retaining this information is consistent with Verizon's current privacy policy.[166] The point of the Verizon example is simply to illustrate that ISPs retain, for at least

---

162. *See* Maija Palmer, *EU Probes Google Grip on Data*, FIN. TIMES (LONDON), May 25, 2007, at 13 (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"); Adam Cohen, *What Google Should Roll Out Next: A Privacy Upgrade*, N.Y. TIMES, Nov. 28, 2005, at A18 (criticizing Google for its privacy policies, including data retention); Victoria Shannon, *Footprints in the Sand*, INT'L HERALD TRIB. (PARIS), Mar. 22, 2007, at 21 (suggesting that Google could have used the occasion to better educate users in protecting their privacy).

163. An analysis based on survey data collected after Google's announcement found that a significant percentage of Internet users falsely believe that "[i]f a website has a privacy policy, it means that the site cannot use information to analyze your online activities." CHRIS JAY HOOFNAGLE & JENNIFER KING, RESEARCH REPORT: WHAT CALIFORNIANS UNDERSTAND ABOUT PRIVACY ONLINE 16 (2008), http://groups.ischool.berkeley.edu/samuelsonclinic/files/online_report_final.pdf. Moreover, this report finds that a majority of users who rarely or never shop online wrongly believed that the above statement was true or did not know whether it was true or false. *Id.* The report specifically notes that these users' misunderstanding extends to their use of Internet search engines, and that they are "using the internet while profoundly misunderstanding the rules of the road." *Id.*

164. *See In re* Verizon Internet Servs., Inc., 240 F. Supp. 2d 24, 28 (D.D.C. 2003).

165. *See id.* at 28–29. The district court ordered Verizon to comply with the subpoena, *id.* at 45, but Verizon asked the district court to stay its order pending an appeal of the court's interpretation of the statutory subpoena provision. The district court refused, *In re* Verizon Internet Services, Inc, 257 F. Supp. 2d 244, 247 (D.D.C. 2003), *administrative stay vacated by* Recording Indus. Ass'n of America, Inc. v. Verizon Internet Servs., Inc., Nos. 03-7015, 03-7053, 2003 WL 21384617, at *1 (D.C. Cir. June 4, 2003), and Verizon produced the names of four of its subscribers while the appeal was pending. Electronic Privacy Information Center, RIAA v. Verizon, http://epic.org/privacy/copyright/verizon/ (last visited Dec. 19, 2008). Other ISPs conspicuously failed to raise the argument that they did not have the subscriber information that the RIAA sought. *See, e.g.*, Charter Communications' Motion to Quash Subpoena Served by Recording Industry Association of America, *In re* Charter Commc'ns, Inc., No. 4:03MC00273CEJ (E.D. Mo. Oct. 3, 2003), *available at* http://www.eff.org/IP/P2P/20031003_motion_to_quash.pdf (declining to argue that Charter did not have the information necessary to comply with the RIAA's subpoena for personal identifying information linked to an IP address).

166. *See* Verizon Online — Policies — Privacy Policy, http://www.verizon.net/policies/vzcom/privacy_popup.asp (last visited Dec. 19, 2008) (stating that "Verizon does not sell or disclose individually-identifiable information obtained online, or information about you or your account or service, to anyone outside of Verizon or its authorized vendors, contractors and agents unless . . . disclosure is required by law").

several months, sufficient information to link an IP address to an individual subscriber.

A second example involves search engine data retention. In August 2005, as part of the defense of the Child Online Protection Act,[167] the U.S. Department of Justice 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.[168] Google did not deny that the queries were available, though it moved to quash the subpoena on other grounds.[169] Moreover, although Google's memorandum indicated that the company performs some analysis of its search queries, the memorandum did not specify the kinds of analyses.[170]

The ECPA's consent provisions operate in a similar way. While they provide broad leeway for cybersecurity research within a single firm, they do not necessarily grant such leeway for disclosure to outside cybersecurity researchers. The typical means of securing consent is via the provider's terms of service, which often include, or incorporate by reference, a privacy policy.[171] Consent provisions may be in the middle of extensive terms of service agreements posted online;[172] courts do not require specific acknowledgement of a consent provision. Courts have held, for example, that establishing the invalidity of consent to an interception under the Wiretap Act requires proof that

---

167. Child Online Protection Act, Pub. L. No. 105-277, 112 Stat. 2681 (1998) (codified at 47 U.S.C. § 231 (2000)). The Department of Justice's defense of the Act is found in *ACLU v. Gonzales*, 478 F. Supp. 2d 775 (E.D. Pa. 2007).

168. Gonzales v. Google, Inc., 234 F.R.D. 674, 679 (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. *Id.*

169. Google's Opposition to the Government's Motion to Compel at 10–13, *Google*, 234 F.R.D. 674 (No. 5:06-mc-80006-JW), 2006 WL 728287.

170. *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 [reachable URLs and search queries] with third parties." *Id.* at 11 n.2.

Google was not exceptional for search engines at the time of this case. The other search engines that received subpoenas in 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. *See Google*, 234 F.R.D. 674, 679 (N.D. Cal. 2006).

171. *See, e.g.*, Amazon.com, Conditions of Use, http://www.amazon.com/gp/help/customer/display.html?nodeId=508088 (last visited Dec. 19, 2008) ("Please review our Privacy Notice, which also governs your visit to Amazon.com, to understand our practices.").

172. *See, e.g.*, eBay Privacy Policy, http://pages.ebay.com/help/policies/privacy-policy.html?_trksid=m40 (last visited Dec. 19, 2008) ("By accepting the Privacy Policy and the User Agreement in registration, you expressly consent to our collection, storage, use and disclosure of your personal information as described in this Privacy Policy.").

the party seeking consent acted primarily out of motivation to commit a tort or crime.[173]

It is unclear to which activities consent extends. Communication service providers, regardless of whether or not they offer service to the public, could use privacy policies to make data available for research. But, at the same time, privacy policies may leave customers wondering what the policies allow, if they read them at all.[174] The privacy policies of prominent ISPs and e-mail providers, which are subject to the SCA, contain broad clauses that affect a user's consent to share his or her communications information with an ambiguous group of "affiliates."[175]

Major ISPs obtain user consent to collect information about Internet usage for network performance engineering and research.[176] One ISP "store[s] e-mail messages and video mail messages [sent and received by its users] on computer systems for a period of time."[177] In the academic environment, the University of California, Berkeley collects and stores transactional records pertaining to communications between users of Berkeley's network and outside Internet addresses.[178] Berkeley stores "raw" data identifiable to specific IP addresses for one month at most, unless "a privacy filter is applied to the

---

173. *In re* DoubleClick Inc. Privacy Litig., 154 F. Supp. 2d 497, 514–15 (S.D.N.Y. 2001) (citing cases from the D.C. Circuit and the First Circuit).

174. A recent study that examined consumer beliefs about electronic commerce in general, rather than relationships with communications service providers in particular, found that "[c]onsumers do not understand the nature and legality of information-collection techniques." JOSEPH TUROW, DEIRDRE K. MULLIGAN & CHRIS JAY HOOFNAGLE, RESEARCH REPORT: CONSUMERS FUNDAMENTALLY MISUNDERSTAND THE ONLINE ADVERTISING MARKETPLACE 1 (2007), http://groups.ischool.berkeley.edu/samuelsonclinic/files/annenberg_samuelson_advertising.pdf.

175. *See, e.g.*, Microsoft Online Privacy Statement, http://privacy.microsoft.com/en-us/fullnotice.aspx (last visited Dec. 19, 2008).

176. *See* Anestis Karasaridis, et al., *Wide-Scale Botnet Detection and Characterization*, WORKSHOP ON HOT TOPICS IN UNDERSTANDING BOTNETS (HOTBOTS '07), Apr. 10, 2007, http://www.usenix.org/events/hotbots07/tech/full_papers/karasaridis/karasaridis.pdf (stating that the research was "performed as part of the product evolution for AT&T Internet Protect"); AT&T, Internet Protect, http://www.business.att.com/enterprise/Service/business-continuity-enterprise/threat-management-enterprise/internet-protect-enterprise/ (last visited Dec. 19, 2008) (explaining that the Internet Protect service involves real-time analysis of 2.5 petabytes per day of traffic on AT&T's backbone network). It is also common for online service providers, such as free e-mail services and search engines, to obtain user consent to collect and use information for research purposes. *See, e.g.*, Yahoo! Privacy, http://info.yahoo.com/privacy/us/yahoo/details.html (last visited Dec. 19, 2008) (noting that "Yahoo! automatically receives and records information from your computer and browser" and uses this information to "conduct research"); Privacy Policy — Google Privacy Center, http://www.google.com/intl/en/privacypolicy.html (last visited Dec. 19, 2008) (stating that Google processes personal information, including information obtained from users' connections to a Google site, for research).

177. Comcast High-Speed Internet Privacy Information, http://www.comcast.net/privacy/ (last visited Dec. 19, 2008).

178. *See* Cliff Frost, CNS Data Collection and Retention, http://cns.berkeley.edu/dept/CNS%20Data%20Collection%20and%20Retention.doc (setting forth "Netflow Data" retention policy).

data."[179] The university may store anonymized network usage data indefinitely.[180] Appropriate staff may review this data "to understand the volume and characteristics . . . of the traffic flowing through various points in the network."[181] 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."[182]

B. Institutions

Institutional forces also contribute to the dearth of cybersecurity data. Relevant data is widely scattered among public and private actors. There is no overarching organizational mechanism — least of all the government — to encourage or compel those actors to disclose data to cybersecurity researchers.[183] 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 unwilling to do so for a variety of reasons. Among these reasons are that disclosure creates a risk of customer backlash, the firm fears 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.[184] In summary,

---

179. *Id.*

180. *Id.*

181. *Id.*

182. UNIVERSITY OF CALIFORNIA OFFICE OF THE PRESIDENT, ELECTRONIC COMMUNICATIONS POLICY 15 (2005), http://www.ucop.edu/ucophome/coordrev/policy/PP081805ECP.pdf.

183. This was not always the case. Until April 1995, the NSF operated the Internet's "backbone" — the networking equipment that connects separate institutions over long distances. During this time the NSF regularly provided network data to researchers. *See* CAIDA, *supra* note 88, at 1; *see also infra* Part IV for a discussion of the current data needs of cybersecurity researchers.

184. *See* Mark Allman & Vern Paxson, *Issues and Etiquette Concerning Use of Shared Measurement Data*, 2007 PROC. ACM SIGCOMM CONF. ON INTERNET MEASUREMENT 135, 136, *available at* http://www.icir.org/mallman/papers/etiquette-imc07.pdf ("Releasing data is fraught with potential problems . . . includ[ing] potentially compromising the privacy of users, exposing activity that might embarrass the institution, . . . enabl[ing] an attacker to more effectively mount an attack, and exposing aspects of the network's operation to possible competitors."); KIMBERLY CLAFFY, COOP. ASS'N FOR INTERNET DATA ANALYSIS, TEN THINGS LAWYERS SHOULD KNOW ABOUT THE INTERNET 2 (2008), http://www.caida.org/publications/papers/2008/lawyers_top_ten/lawyers_top_ten.pdf ("Even for data that is legal to share, there are overwhelming counter incentives to sharing any data at all in the competitive environment we have chosen . . . .").

One can infer that the availability of data within an organization would attract research talent by comparing the amount of network data available to researchers at AT&T, *see* Karasaridis et al., *supra* note 176, with the data available to outside researchers, *see* CAIDA, *supra* note 88, at 3 ("For years it has been virtually impossible for researchers to get access to passive (sniffed) data from Internet backbone links due to privacy reasons . . . . As of March 2005 there is no available data on Internet backbone links, and so researchers can no longer analyze Internet backbone workloads.").

there are few institutional forces that promote sharing of cybersecurity-relevant data, and there are few incentives for network service providers to promote the concept that sharing data in support of cybersecurity provides a public benefit.

References to data collection in privacy policies are illustrative. Instead of defining "research" explicitly, these policies tend to couch the sharing and collection of communications data in terms of the benefits of improved service and more tailored solicitations from business partners.[185] A provider could state that it requires consent from its users to share their communications-related data in order to advance cybersecurity research. Privacy policies, however, tend to obfuscate rather than clarify the provider's actual data retention and handling practices. The benefit that could arise from the research facilitated by this kind of data sharing may be 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 been an effective means for cybersecurity researchers to obtain access to data.[186]

Cybersecurity research policies at universities — potentially promising sources of network data — are also difficult to penetrate. Universities tend to offer strong privacy protection to their faculty and students.[187] Interviews I conducted with a number of university researchers revealed that they face significant challenges in obtaining access to data from their own institutions. These challenges are even more severe when researchers also wish to retain these datasets.

---

185. The privacy policies of online service providers with substantial research operations — identified in *infra* note 188 — seem most relevant here. AT&T does not mention research. *See* AT&T Privacy Policy, http://www.att.com/gen/privacy-policy?pid=7666#104 (last visited Dec. 19, 2008). The others mention research once without specifying what data researchers will use, or how they will use it. *See* Privacy Policy — Google Privacy Center, http://www.google.com/intl/en/privacypolicy.html (last visited Dec. 19, 2008) (stating that Google uses personal information for "[a]uditing, research and analysis in order to maintain, protect and improve our services"); Microsoft Online Privacy Statement, http://www.microsoft.com/info/privacy/fullnotice.mspx (last visited Dec. 19, 2008) ("Microsoft collects and uses your personal information to operate and improve its sites and services. These uses may include . . . performing research and analysis aimed at improving our products, services and technologies . . . ."); Yahoo! Privacy, http://info.yahoo.com/privacy/us/yahoo/details.html (last visited Dec. 19, 2008) ("Yahoo! uses information for the following general purposes: to . . . conduct research . . . .").

186. Cybersecurity researchers have stated that greater access to real network traffic datasets would "cause a paradigmatic shift in computer security research." Porras & Shmatikov, *supra* note 85, at 1. 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." Slagell & Yurick, *supra* note 20, at 1.

187. For example, a number of universities recently announced that they would limit their cooperation with requests to disclose personally identifying information about their students in connection with the recording industry's investigations into alleged copyright infringement.

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.[188] In a world in which access to network data is highly constrained, a firm that offers its researchers access to network data could be a much more attractive place to work. This consideration might make firms reluctant to share data, even if it is legally permissible for them to do so.

IV. COPING WITH THE DEARTH OF CYBERSECURITY DATA

A. 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 advance certain scientific goals. These goals provide background for the descriptions of available cybersecurity data in Sections B and C of this Part, and for evaluating the legal and institutional proposal in Part V.

First, cybersecurity researchers have called for making access to cybersecurity data as broad as possible.[189] Broad access to data would remove the element of luck that is sometimes involved in obtaining data. This condition would also allow many researchers to examine the same dataset, aiding efforts to make experimental computer science results reproducible by multiple researchers.[190]

Second, the condition of utility counsels that cybersecurity data should be made available in as "raw" a form as possible.[191] Scrambling or anonymizing data degrades its usefulness to researchers, and in some cases this kind of processing can render data unfit for a specific research use.

Third, cybersecurity researchers advocate an extended period of data availability to allow different researchers to use the same data.[192]

---

188. For example, AT&T, Google, Microsoft, and Yahoo! all have large research divisions. *See* AT&T Labs Research, http://www.research.att.com/ (last visited Dec. 19, 2008); About Google Research, http://research.google.com/about (last visited Dec. 19, 2008); Microsoft Research Overview, http://research.microsoft.com/aboutmsr/overview/ (last visited Dec. 19, 2008); Yahoo! Research, http://research.yahoo.com/ (last visited Dec. 19, 2008).

189. *See, e.g.*, Vitaly Shmatikov, Threats to Anonymized Datasets 4 (Sept. 27, 2005) (unpublished presentation), *available at* http://www.cyber.st.dhs.gov/public/PREDICT/Vitaly-athreats1.pdf.

190. *See* Bajcsy et al., *supra* note 84, at 61 ("The lack of open, objective, and repeatable validation of cyber defense technologies has been a significant factor hindering wide-scale adoption of next-generation solutions.").

191. *See* Shmatikov, *supra* note 189, at 4.

192. *See, e.g.*, Ruoming Pang et al., *The Devil and Packet Trace Anonymization*, ACM SIGCOMM COMPUTER COMMC'N REV., Jan. 2006, at 29, *available at* http://www.icir.org/enterprise-tracing/devil-ccr-jan06.pdf (describing process of releasing anonymized datasets on the Internet).

Permanent, public datasets would not only facilitate the evaluation of published research but would also allow cybersecurity researchers to examine network trends over time.

A fourth criterion for cybersecurity data is that it should reflect Internet traffic's many different applications, protocols, and dynamics.[193] Different research questions require different kinds of data, which in turn implicate different legal and policy questions. Some data raises difficult questions about protecting individual privacy, while other data creates security risks for the firms that provide them.[194] Finally, cybersecurity researchers recognize both the need to protect the privacy of individuals whose activities are represented in communications data, and the potential for shared data to aid an attacker who targets the data source.[195] 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 answer the question of how to anonymize selected aspects of data.[196]

B. Data Needs: A Picture of the Ideal

To develop a more concrete picture of cybersecurity research approaches and data needs, consider again the cyberattack against Estonia, which was discussed in Part II. This attack was an example of a DDoS attack: traffic from many hosts on the Internet flooded network 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 of the Internet. The network equipment that routes traffic to a destination does not examine whether that traffic is malicious, or whether the recipient's network is too clogged to accept more data. Rather, the computer sending

193. *See* Allman & Paxson, *supra* note 184, at 135 (noting that "there is major benefit in sharing datasets in order to gain broader, more representative insight into the highly diverse nature of Internet traffic and dynamics"); *see generally* Ruoming Pang et al., *A First Look at Modern Enterprise Traffic*, 2005 PROC. ACM SIGCOMM CONF. ON INTERNET MEASUREMENT 15, *available at* http://www.icir.org/enterprise-tracing/first-look-imc05.pdf (describing characteristics of Internet traffic along dimensions of origin, applications in use, protocols, timing, and network load).

194. *See* Douglas Maughan, PREDICT Overview (Sept. 27, 2005) (unpublished presentation), 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).

195. *See* Pang et al., *supra* note 192, at 18 (noting that attackers might use network datasets to construct a "map" of computers on a network and use this information to attack the network).

196. *See* Allman & Paxson, *supra note* 184, at 136 ("[R]esearchers have developed a number of anonymization techniques to scrub data for release. While useful, these techniques do not — and cannot — provide guaranteed protection against information leakage. . . . [U]ltimately the choice about what to release, how to obscure the data, and to whom to release the data, are *policy decisions*." (internal citations omitted)).

a message will keep sending it until the receiver confirms that it has received the full, uncorrupted message, even if this delivery takes a long time. In other words, a DDoS attack exploits the Internet's basic delivery guarantee and its lack of performance and accountability guarantees.[197]

Cybersecurity researchers study DDoS attacks from a number of different angles. Some have focused on real-time detection of attacks,[198] while others have focused on analyzing attacks after they occur.[199] In order to validate either approach, researchers need to correlate data from the many different sources that direct traffic to the attack target to determine whether a detection algorithm correctly distinguishes attack traffic from innocuous communications. Real data can also help algorithms or their creators learn to reduce false positives, which can create so much noise that network operators end up missing real attacks.[200]

Also consider the "Witty" worm discussed in Part I. Early warnings about such attacks would allow a network operator to take steps to stop the worm from spreading to uninfected machines on its network.[201] 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 caused. Both objectives remain topics of active research, and both require large volumes of electronic communications data from many separately controlled organizations in order to effectively validate reconstruction and repair methods.[202]

Both communications contents and addressing data are valuable to worm researchers in particular and cybersecurity researchers in general.[203] Ideally, researchers would have access to addressing data from multiple entities, such as ISPs, in order to test these methods.[204]

---

197. As described in Part II, 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.

198. *See* Xie et al., *supra* note 12, at 43.

199. *See* Allman et al., *supra* note 43, at 121.

200. *See* Heather LaRoi, *Prof Aims to Improve Internet Security*, WIS. STATE J., Jan. 26, 2008, at D1 ("The problem is if you have hundreds of false positives and you have to weed through every one, the chance of you missing a real one is greatly increased.").

201. *See* Xie et al., *supra* note 12, at 52–53.

202. *See id.*

203. As discussed in Part III, addressing information receives less protection than contents under the ECPA. As discussed later in this Part, however, some important areas of cybersecurity research would greatly benefit from access to communications contents.

204. *See* Xie et al., *supra* note 12, at 44.

C. Public Releases

What kinds of data are actually available to study these problems? ISP data is not available to cybersecurity researchers,[205] unless the researchers happen to work for an ISP.[206] Publicly available data falls roughly into the fundamental ECPA categories, non-content data and communications contents. There is far more publicly available non-content data, but even this data retains significant limitations on its utility for cybersecurity researchers.

1. Non-Content Data

The most significant public release of non-content data occurred in 2006, when researchers affiliated with Lawrence Berkeley National Laboratory ("LBNL") and a non-profit research institute placed approximately eleven gigabytes of anonymized data on the Internet.[207] In doing so, the researchers noted that "[s]haring of network measurement data . . . has been repeatedly identified as critical for solid networking research."[208] This set of "packet traces" included mainly source and destination addresses[209] and thus did not contain communications contents.[210] Moreover, the researchers anonymized the addresses of LBNL users as well as the sites they visited.[211]

While this release was a significant advance in the amount of data available for cybersecurity research, the researchers themselves noted several limitations. First, when publicly releasing the data, the researchers took pains to remove traffic that revealed too much about the laboratory's network layout and could be used to attack that network.[212] 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.[213] Second, developing the anonymization algorithm that the researchers applied to the data was itself a difficult problem. They believed the anonymization to be difficult to reverse, but other

---

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

206. *See* Karasaridis, et al., *supra* note 176, at 2 (reporting AT&T researchers' results from "billions of flow records" that appear to have been obtained from AT&T's networks).

207. LBNL/ICSI Enterprise Tracing Project — Trace File Download, http://www.icir.org/enterprise-tracing/download.html (last visited Dec. 19, 2008).

208. Pang et al., *supra* note 192.

209. *See id.* at 17–20 (describing the information contained in the packet traces).

210. *See* 18 U.S.C. § 3121(c) (2006) (limiting Pen/Trap interceptions to "dialing, routing, addressing, and signaling information" and specifically distinguishing such information from the contents of communications).

211. LBNL/ICSI Enterprise Tracing Project — Project Overview, http://www.icir.org/enterprise-tracing/index.html (last visited Dec. 19, 2008).

212. Pang et al., *supra* note 192, at 17–18.

213. *Id.* at 7.

researchers subsequently published a paper demonstrating an attack on the anonymization scheme.[214] 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.[215]

A second way to obtain non-content network data is from one of the few network data collection organizations operating today.[216] 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.[217] Other non-commercial data collection organizations work on the same model of providing only high-level statistics and analysis, rather than raw data.[218] Cybersecurity researchers, however, frequently need data that provides insight into the behavior of individual computers on the Internet, rather than aggregated statistics. Network data collection organizations, whose perspectives are limited to their own machines, also tend to have limited views of the Internet.[219] Many of these organizations also have business policies that create research barriers. Some organizations do not offer third parties access to raw data,[220] while others offer a few specific types of datasets that do not contain the data necessary for a particular research

---

214. Scott Coulls et al., *Taming the Devil: Techniques of Evaluating Anonymized Network Data*, 2008 PROC. NETWORK & DISTRIBUTED SYS. SECURITY SYMP. 125, *available at* http://www.isoc.org/isoc/conferences/ndss/08/papers/08_taming_the_devil.pdf. Two authors of the original paper describing the anonymization scheme criticized the publication of an unauthorized attack on the scheme as a breach of scholarly etiquette that would make future public releases of data less likely. *See* Allman & Paxson, *supra* note 184, at 138. More broadly, the susceptibility of anonymized data to reidentification presents an obstacle to public releases of datasets. *See, e.g.*, Schneier on Security, Anonymity and the Netflix Dataset, http://www.schneier.com/blog/archives/2007/12/anonymity_and_t_2.html (Dec. 18, 2007, 05:53 PST) (commenting on a paper that reported an algorithm that could uniquely identify 99% of anonymized Netflix movie reviews from eight such reviews and other publicly available data).

215. Pang et al., *supra* note 192, at 21–22.

216. For a detailed overview, see Slagell & Yurcik, *supra* note 20, at 82–85.

217. *See* Symantec DeepSight Threat Management System, https://tms.symantec.com/ (last visited Dec. 19, 2008); Slagell & Yurcik, *supra* note 20, at 83 (discussing DeepSight).

218. *See* Slagell & Yurcik, *supra* note 20, at 83 (discussing Internet Storm Center). The Internet Storm Center collects and analyzes traffic logs from a large number of users for signs of large-scale malicious activity. *See* SANS Internet Storm Center, About the Internet Storm Center, http://isc.sans.org/about.html (last visited Dec. 19, 2008). According to its website, 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.*

219. *See* Pang et al., *supra* note 193, at 1 (noting that "[i]t has long been established that the wide-area Internet traffic seen at different sites varies a great deal from one site to another and also over time, such that studying a single site *cannot* be representative" (internal citations omitted) (emphasis in original)).

220. *See, e.g.*, SANS Internet Storm Center, *supra* note 218 (describing the organization's activities, including collecting Internet traffic logs from contributors, analyzing them, and releasing high-level reports).

use.[221] For example, DatCat, which catalogs but does not store network datasets available on the Internet, lists datasets that may focus on specific applications (such as Skype) or network services (such as DNS).[222] Though these datasets may be useful for some research projects, DatCat does not provide the infrastructure for collecting additional data or fully understanding the conditions under which the listed data was collected.

2. Communications Contents

In light of the ECPA's restrictions on the disclosure of communications contents, it is unsurprising that cybersecurity researchers can tap few data sources for communications contents. The few available datasets were released under circumstances that are unlikely to reappear frequently, and the aims of the releasing institutions were not closely related to cybersecurity. Nevertheless, these releases illustrate the difficulties that public disclosure of communications contents encounters. Part V argues that improving cybersecurity researchers' access to such data will require a combination of legal reform and institutional response.

Though AOL intended to help researchers when it published to the Internet a dataset of 20 million search queries from more than 650,000 users in August 2006, the uproar surrounding the release dealt a setback to efforts to promote more data sharing.[223] AOL made a crude attempt to anonymize the data, but it quickly became apparent that the company had not done enough.[224] Within days of the release, journalists from the *New York Times* reported that they had determined the identity of one woman whose queries were released and published an interview with her.[225] Though some researchers welcomed the release, the broader public reaction was highly critical of AOL's choice. Several AOL users sued AOL under a number of privacy-related theories, including violations of the Stored Communica-

221. *See* Slagell & Yurick, *supra* note 20, at 2–3.

222. *See* DatCat, Recently Contributed Collections and Publications, http://imdc.datcat.org/RecentCollections (last visited Dec. 19, 2008).

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

224. AOL did not release the identities of the users whose queries were contained in the sample, and it obfuscated the identifier for each of those users. The company did not, however, delete or obfuscate the contents of the queries themselves. Search queries can contain significant substance, including the searcher's identity. *See* Eytan Adar, User 4XXXXX9: Anonymizing Query Logs (May 8, 2007) (unpublished manuscript), *available at* http://www.cond.org/anonlogs.pdf.

225. Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES, Aug. 9, 2006, at A1.

tions Act.[226] Top AOL management quickly denied that the release was an official act of the company, and three employees involved in releasing the data left AOL.[227] In the end, even the researchers whom AOL intended to help by releasing this dataset were reluctant to use it.[228]

Despite the tremendous public outcry, the data itself was of limited use to cybersecurity researchers. Such researchers prefer logs of search queries that are linked to search *results*, which usually gives them a better sense of whether the queries led to malicious software sites or played a role in coordinating attacks.[229] Obtaining this information in a comprehensive fashion often requires access to a search engine index,[230] which is a closely guarded secret of search engine companies. In a recent paper, researchers from Google reported results obtained using Google's search index.[231] One conclusion that can be drawn from this study is that these strong proprietary data sources help their owners to attract research talent and maintain the prestige of company research divisions. This in turn provides a strong incentive not to share information with other organizations.

A set of publicly available e-mails from the accounts of former Enron employees comprises a second major source of communications contents. The Federal Energy Regulatory Commission ("FERC") released these e-mails as part of its investigation into Enron's activities in western states' energy markets between 2000 and 2001.[232] The dataset contains approximately a half-million e-mails from 150 Enron users.[233] This is a considerable amount of e-mail, but it is relatively small when compared to the volume of e-mail that passes through a large enterprise's mail server in a single day. For research that involves scanning a realistic mixture of messages, this dataset is inadequate. In addition, the Enron dataset does not contain attachments,[234] making it less useful to researchers interested in scanning e-mail attachments for viruses or other malicious code. Despite

---

226. Complaint at 2–3, 12, Ramkissoon v. AOL LLC, No. 4:06-cv-05866-SBA (N.D. Cal. Sept. 22, 2006) (on file with the *Harvard Journal of Law & Technology*) (alleging, among other things, violations of 18 U.S.C. § 2702).

227. *See* Tom Zeller, Jr., *AOL Acts on Release of Data*, N.Y. TIMES, Aug. 22, 2006, at C1.

228. Katie Hafner, *Researchers Yearn To Use AOL Logs, but They Hesitate*, N.Y. TIMES, Aug. 23, 2006, at C1.

229. *See, e.g.*, Provos et al., *supra* note 46, at 2.

230. *See, e.g.*, *id.* (describing how Google researchers used the search engine index to catalog malware threats).

231. *Id.*

232. Enron Email Dataset, http://www.cs.cmu.edu/~enron/ (last modified Apr. 4, 2005). *See* Federal Energy Regulatory Commission, Information Released in Enron Investigation, http://www.ferc.gov/industries/electric/indus-act/wec/enron/info-release.asp (last visited Dec. 19, 2008).

233. *Id.*

234. *Id.*

its limitations, the Enron e-mail dataset is in wide use among cybersecurity researchers seeking to understand such threats, because it is the best available source of data.[235]

The circumstances surrounding the release of the Enron e-mails were unusual. The FERC released the e-mails to provide insight into the culture of a company whose implosion was a singular event in U.S. corporate history. These circumstances overrode many of the concerns about individual privacy that would normally attend a public airing of the contents of the e-mail servers of a major U.S. corporation.[236]

D. Private Access

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

Even so, there are problems with this approach. First, it does not scale well. Researchers' relationships with data sources outside their own institutions develop over many years.[237] The need to build trust presents a significant barrier for researchers who are entering cybersecurity research or expanding into a new area. Sources typically provide data on the condition that the researcher will not distribute it to any other researcher, thus thwarting the goal of making public datasets part of cybersecurity research. Access to data also depends on the continuing cooperation of the data source; personnel changes or professional disagreements can limit a researcher's access.

The second problem with relying on trust relationships is that they often severely limit the details that researchers may publish about the data that they use. Despite some exceptions,[238] researchers are often circumspect about their sources.[239] This lack of detail can make

---

235. This assertion is based on interviews conducted with cybersecurity researchers who prefer to remain anonymous when discussing common practices for obtaining access to data for research.

236. Attachments were removed to protect privacy, and some e-mails were redacted following former employees' requests. Enron Email Dataset, *supra* note 232.

237. *E.g.*, CAIDA, *supra* note 88, at 3.

238. *See, e.g.*, Berkeley Email User Mobility Traces, http://www.cs.berkeley.edu/~czerwin/traces/ (last visited Dec. 19, 2008) (containing released anonymized records of e-mail account activity information identified as being from the University of California, Berkeley Electrical Engineering and Computer Science department's e-mail server).

239. *See, e.g.*, Vyas Sekar et al., *A Multi-Resolution Approach for Worm Detection and Containment*, 2006 PROC. INT'L CONF. ON DEPENDABLE SYS. & NETWORKS 189, 191,

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, may have less trouble identifying the sources of their data. However, those companies may also put a lower premium on publishing results, especially when sensitive or confidential data underlies the research.

V. A PRIVACY-PRESERVING FRAMEWORK FOR CYBERSECURITY RESEARCH

Communications privacy law lacks a policy apparatus to provide cybersecurity researchers with access to communications data. As noted in Part II, research in other scientific fields strikes a different balance between individual privacy interests and the social interest in research. Medical research provides a particularly instructive model. The HIPAA Privacy Rule permits health care providers to disclose patient records to researchers without individual consent, assuming that the proposed research meets certain substantive requirements[240] and undergoes proper institutional review.[241] To pass this rule's substantive test, applicants for the consent waiver must demonstrate that the research would not be feasible without the data.[242] To pass the procedural test, applicants must convince an institutional review board that the disclosure would not adversely affect the privacy interests of the individuals whose data is involved and that data confidentiality, control, and destruction measures are in place.[243] These provisions do not differ based on the government affiliation of the recipient of the data. The HIPAA Privacy Rule also does not preempt state laws that are more protective of privacy.[244] Finally, federal law provides a shield that researchers may invoke to refuse to disclose under subpoena the data that they have obtained.[245]

---

available at http://research.microsoft.com/users/yxie/papers/dsn06.pdf ("We us[ed] a week-long packet-header trace collected . . . at the border router of a university department . . . .").

240. 45 C.F.R. § 164.512(i) (2007).

241. *See infra* Part V.B.

242. 45 C.F.R. § 164.512(i)(2)(ii)(C).

243. Gostin & Hodge, *supra* note 79, at 1473 (citing 45 C.F.R. § 164.512(i)(2)(ii)).

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

245. 42 U.S.C. § 241(d) (2000) grants the Secretary of Health and Human Services discretion to designate certain data exempt from further disclosure. This exemption is quite powerful:

> The secretary may authorize 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.

The public health rationale for research exceptions to medical privacy has begun to apply with increasing force to cybersecurity research. There are close parallels between the spread of infectious diseases and the spread of some types of Internet-based attacks. Both involve large numbers of systems 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."[246] Cybersecurity researchers have also begun to call for an institutional solution, a "Cyber-Center for Disease Control."[247] A top priority in the "Cyber CDC" proposal is to "develop robust communication mechanisms for gathering and coordinating 'field information.'"[248]

The example of medical research does not translate flawlessly to cybersecurity research. It does, however, supply an example of a functioning, 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 cybersecurity research. Section A argues that similar parameters can define a similar exception to be carved out in the laws governing cybersecurity research. Section B then argues that institutional support is necessary to make the exception workable. Finally, Section C addresses the concern that a cybersecurity research exception would create new threats to privacy and security.

A. Requirements for a Cybersecurity Research Exception to the ECPA

Any expansion of access to data for cybersecurity research will necessarily be in tension with certain existing private and public rights. The government is uniquely placed to fund cybersecurity research, but its presence in the field is a major impediment to obtaining the data that cybersecurity researchers seek. The data exists in abundance, but is mostly controlled by private entities that do not have the incentives to conduct research that serves the cybersecurity interests of the larger Internet community. Communications privacy law imposes few limitations on either internal use of data within the private sector or commercially advantageous disclosures to private parties,

---

Id. The author is grateful to Chris Hoofnagle for making him aware of this provision.

246. *See, e.g.*, Xie et al., *supra* note 12, at 43; Kostas G. Anagnostakis et al., A Cooperative Immunization System for an Untrusting Internet 2, 4 (2004) (unpublished manuscript), *available at* http://www1.cs.columbia.edu/~angelos/Papers/icon03-worm.pdf; Staniford et al., *supra* note 89, at 13.

247. This idea was first suggested by Staniford et al., *supra* note 89, at 15–18. Others have echoed the call, including at least one legal scholar. *See* Neal Kumar Katyal, *Digital Architecture as Crime Control*, 112 YALE L.J. 2261, 2286 (2003) (calling for a "Center for Digital Disease Control").

248. Staniford et al., *supra* note 89, at 15–16.

but disclosures to governmental entities engaged in research are forbidden.[249] Entities covered by the ECPA can disclose data to law enforcement officials to provide evidence of criminal activity against a communicating party, but again may not provide data to a governmental entity engaged in research. However, any legal change that promotes disclosure to certain governmental entities without limiting the acceptable grounds for disclosure would strip away many of the ECPA's privacy protections. Finally, data protected under the ECPA is only protected while in the possession of the original collector. Once in the possession of an entity not covered by the ECPA, it is no longer protected from voluntary or compelled disclosure to the government.

The basic outline of an ECPA exception for cybersecurity research is simple: cybersecurity researchers must have access to electronic communications data — both content and non-content information. This is true even of information that the ECPA would otherwise forbid them from holding without the consent of the individuals whose communications are among those that the researchers obtain.[250]

First, a cybersecurity research exception should extend to all titles of the ECPA, including the prohibition on real-time interception of communications contents.[251] Allowing researchers to use communications would not mark a significant normative or practical shift from the ECPA's current protections. One normative foundation for the anti-interception rules of the Wiretap Act was to add protection against commonly understood impositions on privacy; at the time these provisions were enacted, catching a conversation as it occurred was likely to be the only opportunity for interception.[252] 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. These changes undermine the rationale for privileging real-time interception, and the ECPA therefore should not be allowed to continue to bind the hands of researchers.

---

249. See *supra* Part III.A.2 for the rules regarding disclosure in the SCA.

250. This exception would, of course, remain subject to the Fourth Amendment's prohibitions on unreasonable searches and seizures. Certain applications of the exception might raise questions under the Fourth Amendment — for example, allowing state university researchers to intercept the full contents of communications on a commercial ISP's network — but those scenarios would likely be rare. This narrow category of potential Fourth Amendment issues is not further discussed here, but it should be noted that the exception proposed above would present significant benefits to cybersecurity research even if researchers and research organizations steered clear of all applications that implicate the Fourth Amendment.

251. *See* 18 U.S.C. § 2511 (2006) (prohibiting such interceptions generally).

252. *See* H.R. REP. NO. 99-647, at 17 (1986) (stating that when the Wiretap Act was passed in 1968, "the contents of a traditional telephone call disappeared once the words transmitted were spoken and there were no records kept").

Another 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.[253] Law enforcement officials, for example, need to examine such details personally to form a criminal profile.[254] Preventing this kind of privacy invasion remains a strong justification 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, such as e-mail, to test the effectiveness of many types of cyberdefenses. Part of evaluating these programs is determining whether they will work with a realistic volume and variety of communications.[255] Moreover, since the primary utility of real-time data in cybersecurity research is in testing the performance of defense techniques under real-world conditions, research interceptions would involve data only to the extent necessary to fix bugs in cybersecurity applications or to establish that those applications are correctly identifying attacks. This real-time data is necessary to the research in developing such cybersecurity applications, as simulated data would have insufficient scientific validity. Thus, the risk of later unintended uses of intercepted communications is minimal.[256]

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.[257] 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

253. *See, e.g.*, Katz v. United States, 389 U.S. 347, 351–53 (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).

254. Modern technology has vastly increased the ability of law enforcement officials to conduct individualized surveillance by using stored communications records. *See generally* Daniel J. Solove, *Digital Dossiers and the Dissipation of Fourth Amendment Privacy*, 75 S. CAL. L. REV. 1083 (2002).

255. *See, e.g.*, Vinod Yegneswaran, Paul Barford & Somesh Jha, *Global Intrusion Detection in the DOMINO Overlay System*, PROC. NETWORK & DISTRIBUTED SYS. SECURITY SYMP., Feb. 5, 2004, http://www.isoc.org/isoc/conferences/ndss/04/proceedings/Papers/Yegneswaran.pdf.

256. See *infra* Part V.C for a discussion of measure to prevent researchers from abusing their access rights.

257. 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) (2006).

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. An IRB at the source institution would be required to approve any disclosure of data to a cybersecurity researcher, and an IRB at the institution of each researcher must approve the researcher's protocol. This would give institutions power to punish infractions by researchers — by suspending their research activities, for example.

Requiring ex ante IRB approval would prevent the cybersecurity research exception from becoming an ex post justification for a data use or disclosure that the ECPA otherwise would have prohibited. Given that cybersecurity researchers are likely to seek large quantities of sensitive data, granting ex post protection under the research exception would pose an unacceptable risk to individual privacy interests.[258] Presenting a research proposal to an IRB would impose a certain amount of discipline on researchers, preventing them from feeling entitled to request or disclose data as a matter of course.[259] The review process would help maintain accountability by generating a record that institutions and government regulators could examine.

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 commercial ISPs and public e-mail services, are covered by the exception, there is little gained by extending the cybersecurity research exception to the SCA. Excluding some providers, especially based on a factor as outdated as offering service to the public, would make the exception more administratively burdensome and would reduce its effectiveness by casting doubt as to which providers are allowed to share data with researchers.

Fifth, the exception should prohibit any researcher who receives data under the exception from redistributing it in a manner not ap-

---

258. This is in contrast to other security research exceptions, such as the encryption research and security testing provisions of the Digital Millennium Copyright Act ("DMCA"), 17 U.S.C. § 1201(g), (j) (2006). The difference between the DMCA exceptions and the proposed ECPA exception is that in communications interception cases, unauthorized disclosures or uses of communications data can destroy the autonomy interests that underlie communications privacy rights. By contrast, copyright holders can undo at least some of the damage of infringement and unauthorized circumvention by obtaining injunctions that require infringers and circumventors to cease distributing infringing copies and circumvention tools. *See, e.g.*, Universal City Studios, Inc. v. Corley, 273 F.3d 429 (2d Cir. 2001) (granting injunction to prevent dissemination of a circumvention tool).

259. A certain amount of serendipity that might come from researchers finding unexpected uses in a dataset would necessarily be lost, though amendments to IRB submissions could add flexibility.

proved by the source and recipient IRB.[260] The justification for this limitation is threefold. 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. Also, the justification 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 source's networks or systems. A researcher who receives this 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. Finally, a data source might wish to keep data away from researchers employed by a competitor. In that case, the source is best situated to assess the competitive risk involved in disclosure, and an IRB approval requirement for sources ensures that this proprietary information is protected.

Sixth, the exception should grant data obtained by a researcher the same level of protection from compelled disclosure as the data would have in the hands of the original source. This would mean that intercepted communications contents would be available only to law enforcement officers presenting the appropriate warrant,[261] and that stored communications records could be released only to a party that has obtained the appropriate court order or subpoena.[262] This latter requirement would prevent researchers from falling outside the purview of the SCA. Currently, the SCA's disclosure provisions apply to data when it is in the possession of certain entities but do not apply to the data after it is disclosed. In addition, the exception should prohibit researchers from making a voluntary disclosure of data they receive, even if the ECPA would allow it.

Seventh, basic subscriber information held by ISPs about their customers should not be covered under the exception. As explained in Part IV, the interests of cybersecurity researchers in real data lie in what the data reveals about network traffic flows and the spread of malicious code across networks. None of the resulting analysis depends on a researcher's 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 separat-

260. This limitation would allow a researcher to distribute data to students or collaborators named in the protocol approved by the IRB. It would also allow public distribution of data, if the release were part of the approved protocol.

261. *See* 18 U.S.C. § 2518.

262. In other words, if the original data source were an electronic communication service or remote computing service, a researcher who obtained data from such a source would be subject to compelled disclosure provisions that apply to that type of entity. *See* 18 U.S.C. § 2703 (setting forth requirements for obtaining stored communications contents and non-content records).

ing individual identities from their network traffic,[263] 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.[264] Creating an exception to state laws that add protections to a relatively weak federal regime of statutory privacy protection is not something which should be considered lightly, but the research exception might prove unworkable otherwise. Data that is relevant to a cybersecurity research question might necessarily come from many different states.[265] Though the problem of differing state regulations arises in many information collection contexts, the effect of differing laws on Internet-based data collection would be more acute because nearly every communication crosses state lines. Enacting a cybersecurity research exemption to the ECPA, though it might override some state laws, would create a clear national standard for disclosure to researchers.

B. Institutions

The call to create a cybersecurity research exception to the ECPA prompts two further questions: whether it would be administrable and whether it would be effective in reversing the strong institutional forces that currently oppose data sharing. These questions are considered in turn.

Federal law provides a broadly applicable structure, the institutional review board, for reviewing research. IRBs provide a starting point for administering the ECPA research exception. 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.[266]

Federal rules for IRBs are administered by the Department of Health and Human Services through the "Common Rule."[267] The

263. See *supra* Part IV.C.2 for a discussion of AOL's release of "anonymized" search engine queries. Other content that might become available under the cybersecurity research exception, such as e-mail, would carry its own link between individuals and records.

264. See the discussion of state law found in *supra* Part III.A.4.

265. *See, e.g.*, Staniford et al., *supra* note 89, at 15–18 (proposing a decentralized, widely distributed set of network "sensors" to collect information about network-based cybersecurity threats).

266. *See generally* Philip Hamburger, *The New Censorship: Institutional Review Boards*, 2004 SUP. CT. REV. 271, 272–73 (recounting a history of IRBs). 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*, 101 NW. U. L. REV. 687 (2007).

267. 45 C.F.R. § 46 (2007). *See* Lawrence O. Gostin, James G. Hodge, Jr., & Lauren Marks, *The Nationalization of Health Information Privacy Protections*, 8 CONN. INS. L.J. 283, 311 (2001) (referring to 45 C.F.R. § 46 as "the Common Rule").

Common Rule defines "research,"[268] "institution,"[269] and "human subject."[270] 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,[271] and the Rule permits joint review of multi-institutional research proposals.[272] The IRB composition requirement would help to ensure examination of proposals involving cybersecurity data from a number of disciplinary angles. 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 be frequent. Finally, the Common Rule provides standards for IRB approval of projects seeking approval of a waiver of research subject consent.[273]

The Common Rule, however, provides little guidance for protecting the privacy of research subjects.[274] 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.[275] Both are necessary to assess the risk of privacy and security violations in the event of a breach of confidentiality, whether accidental or intentional.

---

268. 45 C.F.R. § 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.").

269. *Id.* § 46.102(b) ("Institution means any public or private entity or agency (including federal, state, and other agencies).").

270. *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.").

271. *Id.* § 46.107.

272. *Id.* § 46.114.

273. *Id.* § 46.117(c).

274. *See* Gostin & Hodge, *supra* note 79, at 1472 (noting that the Common Rule "conditions IRB approval of government-sponsored research on whether 'there are adequate provisions to protect the privacy of subjects'") (quoting 45 C.F.R. § 46.111(a)(7)). It should be noted that IRBs are conservative. Without specific guidance for reviewing a novel use of data, an IRB might reject or demand a significant scaling back of the research protocol in order to be seen as providing strict protection for the users of the network. *See* Carpenter, *supra* note 266, at 696 (noting that IRBs "are much less adept at identifying substantial *nonphysical* risks" than physical risks and end up making decisions "on the basis of worst-case scenarios" (citation omitted) (internal quotation marks omitted)). The possibility of losing the institution's federal funding or subjecting the institution to a large fine for ethical lapses also presents IRBs with a large incentive to be conservative. *See generally id*.

275. *See* 45 C.F.R. § 164.512(i)(2)(ii)(A) (enumerating requirements for a consent waiver, including adequate safeguards against "improper use and disclosure" of personal health information, an adequate plan to destroy information that could be used to identify individuals, and adequate assurance data obtained will not be reused or redistributed in a manner not specified in the protocol).

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,[276] so it is unrealistic to expect complete severance of network activity data from the identity of the individual whose activity the data represents. The prospects for anonymization also vary based on the kind of data in question: e-mail may be all but impossible to anonymize, while some other 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 destroying linkages between data points that must remain available for a proposed use. Finally, the IRB should consider a cybersecurity research proposal's plan for transporting data to the recipient, containing the data while in use, and ensuring the destruction of usable copies once researchers have finished using the data.

The Department of Homeland Security ("DHS") 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),[277] is intended to coordinate the process of giving cybersecurity researchers access to "network operational data."[278]

PREDICT implements 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 provides for three types of data handlers: Data Providers, which are the original sources of network data; Data Hosts, which store datasets once they have been approved for use; and Researchers. A fourth entity, the Coordinating Center, is administered by a nonprofit corporation under contract with DHS.[279]

The Coordinating Center serves many of the functions that an IRB would serve under the ECPA exception.[280] The Center approves 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.[281] The Coordinating Center also acts an initial

276. *See* Pang et al., *supra* note 192, at 29–30; *see also* David E. Bakken et al., *Data Obfuscation: Anonymity and Desensitization of Usable Data Sets*, IEEE SECURITY & PRIVACY, Nov./Dec. 2004, at 34, 34–35.

277. PREDICT > Home, https://www.predict.org/ (last visited Dec. 19, 2008).

278. RTI International, PREDICT Portal Overview, https://www.predict.org/Portals/0/files/Documentation/MANUAL%20OF%20OPERATIONS/PREDICT_Overview_final.pdf [hereinafter PREDICT Portal Overview].

279. *Id.*

280. *See id.*; *see also supra* Part V.A.

281. Data Providers designate the sensitivity of the data that they provide and control the conditions under which data may be released. The Coordinating Center establishes catego-

gatekeeper for deciding whether a researcher may access PREDICT data.[282]

Most importantly, the Coordinating Center helps to orchestrate the review boards that process applications for uses of PREDICT data.[283] Each use of PREDICT data requires separate approval from the board.[284] Each board is to be composed of at least one representative each from DHS, the Coordinating Center, the Data Host, the Data Provider, and the members of the "[c]yber-defense research community."[285] The Data Provider can reject any proposed use of its data.[286]

Despite these safeguards, it is unclear whether PREDICT satisfies the ECPA's prohibitions on voluntary disclosure of non-content records to governmental entities. Though DHS itself may not host PREDICT data as currently organized, other governmental entities, such as state universities, may ultimately house data. In those cases, PREDICT would have to ensure that the communications service providers — such as commercial ISPs — do not provide data to these hosts.[287] PREDICT might also need to ensure that any such data that providers contribute does not end up in the control of government-affiliated researchers. Alternatively, PREDICT would need to bar such researchers altogether. This is not a fault of PREDICT's design; rather, it is what the ECPA demands. Nonetheless, the effect is to con-

---

ries of dataset types and requires Data Providers to comply with anonymization and other data sanitization requirements for data in a given category. *See* RTI International, Memorandum of Agreement: PCC and Data Provider, https://www.predict.org/Portals/0/files/Documentation/MOAs/PREDICT_MOA_PCC_Data_Providers_final.pdf [hereinafter Data Provider MOA].

282. *See* PREDICT Portal Overview, *supra* note 278, at 2.

283. *See* Data Provider MOA, *supra* note 281, at 3 (noting that a review board "in conjunction with the PCC and the Data Provider, reviews and approves or rejects applications for requested Data").

284. *See* RTI International, Memorandum of Agreement: PCC and Researcher/User 4, https://www.predict.org/Portals/0/files/Documentation/MOAs/PREDICT_MOA_PCC_Researcher_final.pdf (stating that "PCC hereby grants to Researcher/User, on behalf of Data Provider and/or Data Host, a right to use the Data solely for the purposes described in the Researcher/User's approved application").

285. *See* RTI International, Memorandum of Agreement: PCC and Data Host 4, https://www.predict.org/Portals/0/files/Documentation/MOAs/PREDICT_MOA_PCC_Data_Host_final.pdf.

286. *Id.*

287. Among the datasets likely to be provided to PREDICT are a national laboratory dataset containing "anonymized *contents*," several university datasets, and data not related to the statutory definition of an electronic communication. Vern Paxson, LBNL/ICSI Enterprise Traces 3 (Sept. 27, 2005) (unpublished presentation), *available at* http://www.cyber.st.dhs.gov/public/PREDICT/PREDICT.Sep05A.pdf (discussing a dataset containing "anonymized *contents*"). *See generally* Univ. of Mich. et al., Virtual Center for Network and Security Data (Sept. 27, 2005) (unpublished presentation), *available at* http://www.cyber.st.dhs.gov/public/PREDICT/DHS-anon-workshop-overview-09272005.pdf; Tom Vest, PCH/PREDICT Update: Routing Topology and Network Quality Data Collection and Hosting (Sept. 27, 2005) (unpublished presentation), *available at* http://www.cyber.st.dhs.gov/public/PREDICT/DHS050927v1.pdf (discussing routing table data). None of these sources come from a service provider to the public, illustrating possible apprehensions about ECPA violations.

tinue to divide the cybersecurity research community into governmental and non-governmental entities. Instead, the more relevant issue is whether a given researcher is employed by a law enforcement agency.

Whether institutions similar to PREDICT would arise under the ECPA exception proposed herein 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 review process would not only simplify administration of the law 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 a given researcher is not a law enforcement or intelligence officer.[288] 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.[289] Third, the ECPA research exception would allow new types of data sharing institutions to evolve, with particular privileges for those that entirely avoid the involvement of law enforcement agents. 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 dearth of data? 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.[290] 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. This process would attach a measure of legitimacy to research and therefore help to change the cultural reluctance of cybersecurity data providers to share data sources.[291]

---

288. See the definition of "law enforcement officer," *supra* note 133.

289. *See* CLAFFY, *supra* note 184, at 21–22 (noting that the ECPA, among other laws, has crippled researchers' access to network data).

290. As discussed in Part III.B, under current communications privacy law and norms, firms are not forthcoming about their uses of communications data in research.

291. The rational for the DMCA's encryption research exception is instructive. In this case, Congress created an exception to the DMCA's anti-circumvention rules in order to preserve the legality of encryption research undertaken for socially beneficial purposes. *See* H.R. REP. NO. 105-551, pt. 2, at 27 (1998) ("The goals of this legislation would be poorly served if these provisions had the undesirable and unintended consequence of chilling le-

Congressional approval of a research exception would help companies reconsider their acclimation to the current environment, which discourages organizations from sharing how much communications data they store, what the data shows, and how they use it.

Second, the protections inherent in the exception might help to address organizations' concerns. 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 about a network's operational structure, which could embarrass the organization or threaten its security. The legal and institutional structures proposed herein 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. Second, companies, as a condition of releasing data, could demand that recipients convince an IRB that their actions are exclusively related to cybersecurity. Extensions of the IRB structure to address misuse of data might 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 operate the Internet to reevaluate their own interests in cybersecurity. Specifically, the fundamental economic difficulties of cybersecurity might begin to shift if organizations with a common interest in cybersecurity — and with legal protection for providing data to researchers — develop institutional controls that manage the risk of disclosure to truly adversarial recipients. Whether these changes in conditions will be sufficient to reverse a culture that disfavors cybersecurity research remains to be seen.

C. Creating New Threats?

A further question about the cybersecurity research exception is whether it would create new risks to individual privacy or the security of data providers. One threat might arise from law enforcement officials or others who seek the data from cybersecurity researchers rather than the original source. 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. Par-

---

gitimate research activities in the area of encryption."); *see also* Liu, *supra* note 18, at 506–09 (discussing the legislative history of and rationale for the exception). In addition, the DMCA not only sets forth an acceptable purpose for circumventing technical protection measures but also provides courts with guidance to determine whether the exception applies in individual cases. Thus, legislation was helpful in delineating legitimate research from illegitimate circumvention as well as providing institutional structure to help make application of the exception consistent with the intended, legitimate purpose.

ties seeking communications data would have to obtain it 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 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, admittedly, be significant risks under the ECPA cybersecurity research exception. A central tenet of computer security is that it is perilous to ignore the skill and motivation of an adversary.[292] This outlook applies both to insiders, who hold some authorization to have access to data or other system resources, and to outsiders.[293] Though insider threats pose particularly daunting challenges, computer scientists are learning to manage such risks through combinations of technology and organizational policy.[294] The cybersecurity research exception would use both of these approaches to manage insider risks.

First, IRBs would review research protocols to ensure that they contain adequate technical measures to maintain the confidentiality of datasets.[295] Researchers and IRBs would have flexibility in determining which measures are appropriate given 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.[296]

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 was disclosed under the exception. The duration of this suspension might depend upon whether his ac-

---

292. COMPUTER SCI. AND TELECOMMS. BD., NAT'L ACAD. OF SCIS., SUMMARY OF DISCUSSIONS AT A PLANNING MEETING ON CYBER-SECURITY AND THE INSIDER THREAT TO CLASSIFIED INFORMATION 10 (2001), *available at* http://books.nap.edu/openbook.php?record_id=10197&page=10.

293. *See, e.g.*, CSTB, MORE SECURE CYBERSPACE, *supra* note 3, at 215–19 (discussing insider threats).

294. *See id.* at 188–91 (discussing how authentication and access control technologies, forensic measures, and personnel management practices such as job rotation and distributing sensitive information on a need-to-know basis can help manage the insider threat).

295. *See supra* Part V (discussing how the HIPAA Privacy Rule provides a model for including technical safeguards in research protocols that involve personal information).

296. For an example of a security plan that contains these elements and others, see David Wagner, Security Plan for Source Code Review Teams, http://www.sos.ca.gov/elections/voting_systems/ttbr/source_code_security_plan.pdf.

tions were intentional or accidental, as well as upon 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 researchers who cannot establish that they are trustworthy or who do not have a legitimate need to use sensitive cybersecurity related data.

A special case of insider data disclosure occurs when research results are published. One of the reasons that cybersecurity researchers seek increased and more formalized access to data is so that they can identify what data they used during an experiment, and discuss whether particular features of the data brought especially noteworthy results. The IRBs associated with the ECPA exception could allow data sources to decide whether researchers could reveal the source of data in publications.[297] The question of how to decide which details of a dataset should be published is difficult to answer. The approach that PREDICT takes is reasonable: require proposed publications to be reviewed 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.[298] The data source should be represented during this review but should not be allowed to veto publication.[299]

Ultimately, no combination of technological, legal, and institutional controls could eliminate these insider risks. But this is no different from other cybersecurity risks.[300] The proper comparison for privacy and security risks under the cybersecurity research exception is not to a world without any research-related disclosures of communications data, but rather to a world in which we continue down the current, non-cooperative path of cybersecurity research.

VI. CONCLUSION

Privacy regulations have limited the potential of cybersecurity research. This Article shows that considering communications privacy objectives sooner rather than later could substantially advance the national interest in improving cybersecurity.

With the exception of PREDICT, the federal government has taken little action to reconcile cybersecurity with privacy. Federal cybersecurity policy has failed to distinguish between the separate interests

297. PREDICT takes this approach. *See* Data Provider MOA, *supra* note 281, at 3.

298. *See id.* at 3 (describing the Publication Review Board).

299. *See id.* (giving the Publication Review Board the power to veto publication).

300. CSTB, CYBERSECURITY TODAY AND TOMORROW, *supra* note 5, at 7 ("The best is the enemy of the good. Risk management is an essential element of any realistic strategy for dealing with security issues." (internal citation omitted)).

in information sharing of law enforcement, the private sector, and researchers. The types of data sharing that serve these interests vary considerably with context, as do the privacy interests involved. Separating these interests is overdue and is a major objective of this Article.

The Department of Homeland Security, in its role as the leader in cybersecurity policy, has done little work to separate cybersecurity and privacy, and its initiatives may be suffering as a result. A recent Government Accountability Office 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."[301]

The Department of Justice, in its role as 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 such data retention.

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 dearth of cybersecurity research data. 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 the sharing of 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 cybersecurity research over the long term.

301. GAO, PROTECTING KEY SECTORS, *supra* note 1, at 14.