Northwestern Journal of Technology and Intellectual Property

Volume 12 | Issue 1 Article 1

2014

Lawful Hacking: Using Existing Vulnerabilities for Wiretapping on the Internet

Steven M. Bellovin
Columbia University

Matt Blaze
University of Pennsylvania

Sandy Clark
University of Pennsylvania

Susan Landau
privacyink.org, susan.landau@privacyink.org

N O R T H W E S T E R N
JOURNAL OF TECHNOLOGY
AND
INTELLECTUAL PROPERTY

Lawful Hacking: Using Existing Vulnerabilities for Wiretapping on the Internet

Steven M. Bellovin, Matt Blaze, Sandy Clark, & Susan Landau

April 2014 VOL. 12, NO. 1

Lawful Hacking: Using Existing Vulnerabilities for Wiretapping on the Internet

By Steven M. Bellovin[*], Matt Blaze[†], Sandy Clark[§], & Susan Landau[‡]

For years, legal wiretapping was straightforward: the officer doing the intercept connected a tape recorder or the like to a single pair of wires. By the 1990s, however, the changing structure of telecommunications—there was no longer just "Ma Bell" to talk to—and new technologies such as ISDN and cellular telephony made executing a wiretap more complicated for law enforcement. Simple technologies would no longer suffice. In response, Congress passed the Communications Assistance for Law Enforcement Act (CALEA)[1], which mandated a standardized lawful intercept interface on all local phone switches. Since its passage, technology has continued to progress, and in the face of new forms of communication—Skype, voice chat during multiplayer online games, instant messaging, etc.—law enforcement is again experiencing problems. The FBI has called this "Going Dark": their loss of access to suspects' communication. [2] According to news reports, law enforcement wants changes to the wiretap laws to require a CALEA-like interface in Internet software.[3]

CALEA, though, has its own issues: it is complex software specifically intended to create a security hole—eavesdropping capability—in the already-complex environment of a phone switch. It has unfortunately made wiretapping easier for everyone, not just law enforcement. Congress failed to heed experts' warnings of the danger posed by this mandated vulnerability, and time has proven the experts right. The so-called "Athens Affair," where someone used the built-in lawful intercept mechanism to listen to the cell phone calls of high Greek officials, including the Prime Minister,[4] is but one example. In an earlier work, we showed why extending CALEA to the Internet would create very serious problems, including the security problems it has visited on the phone system.[5]

---

[*] Steven M. Bellovin is a professor of computer science at Columbia University.

[†] Matt Blaze is an associate professor of computer science at the University of Pennsylvania.

[§] Sandy Clark is a Ph.D. student in computer science at the University of Pennsylvania.

[‡] Susan Landau was a 2012 Guggenheim Fellow; she is now at privacyink.org.

[1] Pub. L. No. 103-414, 108 Stat. 4279 (codified at 47 U.S.C. §§ 1001–1010 (2006)).

[2] *Going Dark: Lawful Electronic Surveillance in the Face of New Technologies: Hearing Before the Subcomm. On Crime, Terrorism, and Homeland Sec. of the H. Comm. on the Judiciary*, 112th Cong. 10 (2011) (prepared statement of Valerie Caproni, General Counsel, Federal Bureau of Investigation), *available at* http://judiciary.house.gov/hearings/printers/112th/112-59_64581.PDF.

[3] Declan McCullagh, *'Dark' Motive: FBI Seeks Signs of Carrier Roadblocks to Surveillance*, CNET (Nov. 5, 2012, 1:03 PM), http://news.cnet.com/8301-13578_3-57545353-38/dark-motive-fbi-seeks-signs-of-carrier-roadblocks-to-surveillance/.

[4] Vassilis Prevelakis & Diomidis Spinellis, *The Athens Affair*, IEEE SPECTRUM, July 2007, at 27, *available at* http://spectrum.ieee.org/telecom/security/the-athens-affair/0.

[5] Steven M. Bellovin, Matt Blaze, Sandy Clark & Susan Landau, *Going Bright: Wiretapping Without Weakening Communications Infrastructure*, IEEE SECURITY & PRIVACY, Jan/Feb 2013, at 64–66, *available at* https://www.cs.columbia.edu/~smb/papers/GoingBright.pdf.

In this paper, we explore the viability and implications of an alternative method for addressing law enforcements need to access communications: legalized hacking of target devices through existing vulnerabilities in end-user software and platforms. The FBI already uses this approach on a small scale; we expect that its use will increase, especially as centralized wiretapping capabilities become less viable.

Relying on vulnerabilities and hacking poses a large set of legal and policy questions, some practical and some normative. Among these are:

(1) *Will it create disincentives to patching?*
(2) *Will there be a negative effect on innovation? (Lessons from the so-called "Crypto Wars" of the 1990s, and in particular the debate over export controls on cryptography, are instructive here.)*
(3) *Will law enforcement's participation in vulnerabilities purchasing skew the market?*
(4) *Do local and even state law enforcement agencies have the technical sophistication to develop and use exploits? If not, how should this be handled? A larger FBI role?*
(5) *Should law enforcement even be participating in a market where many of the sellers and other buyers are themselves criminals?*
(6) *What happens if these tools are captured and repurposed by miscreants?*
(7) *Should we sanction otherwise illegal network activity to aid law enforcement?*
(8) *Is the probability of success from such an approach too low for it to be useful?*

As we will show, these issues are indeed challenging. We regard the issues raised by using vulnerabilities as, on balance, preferable to adding more complexity and insecurity to online systems.

I. INTRODUCTION

¶1 For several years, the FBI has warned that newer communications technologies have hindered its ability to conduct electronic surveillance.[6] Valerie Caproni, General Counsel of the FBI, said in Congressional testimony:

[6] *See, e.g.*, *Going Dark: Lawful Electronic Surveillance in the Face of New Technologies*, *supra* note 2

> Methods of accessing communications networks have similarly grown in variety and complexity. Recent innovations in hand-held devices have changed the ways in which consumers access networks and network-based services. One result of this change is a transformation of communications services from a straight-forward relationship between a customer and a single CALEA-covered provider (*e.g.* customer to telephone company) to a complex environment in which a customer may use several access methods to maintain simultaneous interactions with multiple providers, some of whom may be based overseas or are otherwise outside the scope of CALEA.
>
> As a result, although the government may obtain a court order authorizing the collection of certain communications, it often serves that order on a provider who does not have an obligation under CALEA to be prepared to execute it.[7]

¶2 The FBI's solution is "legislation that will assure that when we get the appropriate court order . . . companies . . . served . . . have the capability and the capacity to respond."[8]

¶3 While on the one hand this request is predictable given past precedent, it is rather remarkable given current national cybersecurity concerns and in light of stark evidence of the significant harm caused by CALEA. The request to expand CALEA to IP-based communications places the needs of the Electronic Surveillance Unit above all else, including the security risks that arise when building wiretapping capabilities into communications infrastructure and applications, other government agencies who face increased risk from hackers and nation states who may exploit this new vulnerability, and the national need for innovation which drives economic prosperity. Rather than examine the issue in terms of social good—which the FBI already does each time it prioritizes certain types of investigations (terrorism cases, drug cases, etc.) or decides whether to conduct a particular investigation—the FBI has thrown down a gauntlet that ignores long-term national interest.

¶4 The FBI's preferred solution—"requiring that social-networking Web sites and providers of VoIP, instant messaging, and Web e-mail alter their code to ensure their products are wiretap-friendly"[9]—will create security risks in our already-fragile Internet infrastructure, leaving the nation more vulnerable to espionage and our critical infrastructure more open to attack, and hinder innovation.[10] Securing communications infrastructure is a national priority. By weakening communications infrastructure and

---

(prepared statement of Valerie Caproni, General Counsel, Federal Bureau of Investigation). The FBI is the law-enforcement agency with the greatest role for setting policy on wiretapping.

[7] *Id*. at 14.

[8] See Oversight of the Federal Bureau of Investigation: Hearing Before the S. Comm. on the Judiciary, 112th Congress (2012) (statement of Robert S. Mueller, III, Director, Federal Bureau of Investigation); *see also* Declan McCullagh, *FBI 'Looking at' Law Making Web Sites Wiretap-Ready, Director Says*, CNET (May 18, 2012, 1:17 PM), http://news.cnet.com/8301-1009_3-57437391-83/fbi-looking-at-law-making-web-sites-wiretap-ready-director-says/.

[9] Declan McCullagh, *FBI: We Need Wiretap-Ready Web Sites—N*ow, CNET (May 4, 2012, 9:34 AM), http://news.cnet.com/8301-1009_3-57428067-83/fbi-we-need-wiretap-ready-web-sites-now/.

[10] Sometimes, such a solution directly benefits the U.S. military. One NSA program—Commercial Solutions for Classified—uses products from government research "layered" with private-sector products to produce communication tools with high security. *See* Fred Roeper & Neal Ziring, Presentation at RSA Conference 2012, Building Robust Security Solutions Using Layering and Independence 2–6 (2012), *available at* http://www.rsaconference.com/writable/presentations/file_upload/star-401.pdf. However, this protection does not extend to the vast majority of civilian computers.

applications, the FBI's proposal would mostly give aid to the enemy. Surely that is neither what the FBI intends nor what sound national priorities dictate.

¶5 The problem is created by technology. Over the course of the last three decades, we have moved from a circuit-switched centralized communications network—the Public Switched Telephone Network (PSTN)—run by a monopoly provider, to a circuit-switched centralized communications network run by multiple providers, to an Internet-Protocol (IP) based decentralized network run by thousands of providers. The first change, from the monopoly provider to multiple providers, gave rise to the need for the Communications Assistance for Law Enforcement Act (CALEA). This simplified law enforcement's efforts to manage wiretaps with multiple, though relatively few, providers. However, in certain situations, such as when peer-to-peer communications or communications encrypted end-to-end are used, legally authorized wiretaps may be impeded. Even if law enforcement does not currently have a serious problem in conducting authorized wiretaps, with time it will. Thus, there is a serious question of what is to be done. In proposing controls on peer-to-peer networks and on the use of encryption, [11] the FBI has floated highly flawed solutions.[12]

¶6 We propose an alternative to the FBI's proposal: Instead of building wiretapping capabilities into communications infrastructure and applications, government wiretappers can behave like the bad guys. That is, they can exploit the rich supply of security vulnerabilities already existing in virtually every operating system and application to obtain access to communications of the targets of wiretap orders.[13]

¶7 We are not advocating the creation of *new* security holes,[14] but rather observing that exploiting *those that already exist* represents a viable—and significantly better—alternative to the FBI's proposals for mandating infrastructure insecurity. Put simply, the choice is between formalizing (and thereby constraining) the ability of law enforcement to occasionally use existing security vulnerabilities—something the FBI and other law enforcement agencies already do when necessary without much public or legal scrutiny—or living with those vulnerabilities *and* intentionally and systematically creating a set of predictable new vulnerabilities that despite best efforts will be exploitable by *everyone*.

¶8 Using vulnerabilities to create exploits and wiretap targets, however, raises ethical issues. Once an exploit for a particular security vulnerability leaves the lab, it may be used for other purposes and cause great damage. Any proposal to use vulnerabilities to enable wiretaps must minimize such risks.

¶9 In a previous work, we discussed the technical feasibility of relying on the vulnerability approach; [15] here we focus on the legal and policy issues posed by this

---

[11] *See* Charlie Savage, *U.S. is Working to Ease Wiretaps on the Internet*, N.Y. TIMES, Sept. 27, 2010, at A1.

[12] *Id.* Six months after the New York Times reported the FBI was seeking additional capabilities for Internet wiretapping, FBI General Counsel Valerie Caproni testified, "Congressman, the Administration is still working on what the solution would be, and we hope to have something that we can work with Congress on in the near future." *See Going Dark: Lawful Electronic Surveillance in the Face of New Technologies, supra* note 2, at 40. As of this writing, no bill has been proposed.

[13] *See* Bellovin, Blaze, Clark & Landau*, supra* note 5, at 62–63.

[14] That is far from the case. Some of the authors have devoted much of their professional careers to preventing or coping with security holes and the problems they cause.

[15] *See* Bellovin, Blaze, Clark, & Landau*, supra* note 5, at 66–68.

approach. In particular, we examine the tension between the use of naturally occurring software vulnerabilities to legitimately aid law enforcement investigations and the abuse of the same vulnerabilities by criminals. We propose that law enforcement adopt strict guidelines requiring immediate disclosure to the vendor any vulnerabilities as soon they are discovered. As we will discuss, such guidelines would allow law enforcement to fully support crime prevention, and—because of the natural lag of the software lifecycle—still allow law enforcement to build a sufficiently rich toolkit to conduct investigations in practice.

¶10 The discussion in this paper is limited to use of vulnerabilities for *communications intercepts*, rather than generic "remote search." While the two concepts have much in common, including the use of vulnerabilities to achieve access, there are distinct differences in both the technical and legal aspects.[16]

¶11 Section II first discusses how CALEA fit into the communications environment at the time, and then its disjunction with newly evolving communication systems. We then examine the reasons for and risks of extending CALEA to IP-based communications. The continued existence of vulnerabilities, fundamental to our proposal, is discussed in Section III. In Section IV, we discuss their use for wiretapping. Using exploits to enable wiretapping raises a number of troubling questions. As the Stuxnet cyberattack amply demonstrates, even carefully tailored exploits can extend past their intended target.[17] Therefore, law enforcement's use of vulnerabilities requires careful consideration of how to limit the proliferation, which we discuss in Section V. Section VI considers whether law enforcement use of vulnerabilities should influence norms around vulnerability reporting. In Section VII, we discuss how to implement vulnerability reporting. We conclude our argument in Section VIII.

II. CALEA: THE CHANGE IN WIRETAP ARCHITECTURE

¶12 The Communications Assistance for Law Enforcement Act (CALEA) was born of a certain time and certain place. It was a law created with the expectation of multiple, but relatively few, communications providers, and of a telephone network not substantially removed from the world of the Public Switched Telephone Network (PSTN) of the 1950s to 1980s. It was anticipated that both the technical and business structure of communications networks would remain centralized. The impact of the more fundamental changes that were percolating at the time of CALEA's passage—IP-based communications and enormous numbers of services—were not anticipated at the time. In this section, we discuss the problems CALEA was intended to address and those it was

[16] "Remote search" is the capability to search the contents of a computer's files via a surreptitious Internet connection. The investigator obtains access, presumably by hacking in, and runs assorted programs; in contrast, more usual searches involve seizing the computer and bringing it to a forensics lab. *See, e.g.*, Susan W. Brenner, *Fourth Amendment Future: Remote Computer Searches and the Use of Virtual Force*, 81 MISS. L.J. 1229, *available at* http://www.olemiss.edu/depts/ncjrl/pdf/2011%20Symposium/14-%20Brenner_FINAL.pdf; *EU to Search Out Cyber Criminals*, BBC NEWS, http://news.bbc.co.uk/2/hi/technology/7758127.stm (last updated Dec. 1, 2008).

[17] *See generally* Nicolas Falliere, Liam O Murchu, & Eric Chien, *W.32 Stuxnet Dossier*, SYMANTEC (Feb. 2011), http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf [hereinafter *Stuxnet Dossier*]. Stuxnet was apparently developed and launched by intelligence or cyberwarfare agencies; as such, its design is likely quite different from a law enforcement exploit.

not intended to address, briefly mention the security risks created by these solutions,[18] and the patchwork of solutions that have emerged to cover IP-based voice communications. We conclude by describing the impact of these changes on wiretapping and CALEA.

A. History of CALEA

¶13 CALEA had its roots in the nascent switch to digital transport of voice over the phone network's local loops in the early 1990s. ISDN was touted as the next wave of telephony, since it could provide what was, for the time, very high-speed data over a switched line.[19] For all ISDN's advantages, however, it was not possible to tap ISDN lines with the traditional "two alligator clips and a tape recorder."[20] Furthermore, cellular telephony was growing rapidly; because the communication was wireless and mobile, cellular communications also could not be tapped with two alligator clips and a tape recorder. While specialized interception gear could have been developed, the FBI instead proposed in 1992 what was originally known as the Digital Telephony Bill, a standardized interface for wiretaps.[21] The bill was opposed by the telecommunications industry and civil-liberties organizations.[22] After considerable debate over the scope of coverage,[23] the current form of CALEA was passed, specifically excluding "information services."[24]

---

[18] Many countries around the world have similar laws. *See, e.g.*, Regulation of Investigatory Powers Act, 2000 c. 23, § 12 (Eng.), *available at* http://www.legislation.gov.uk/ukpga/2000/23/part/I/chapter/I/crossheading/interception-capability-and-costs. Our comments apply equally to all such laws.

[19] ISDN—Integrated Services Digital Network—was defined in Maurizio Dècina & Eric L. Scace, *CCITT Recommendations on the ISDN: A Review*, 4 IEEE J. ON SELECTED AREAS IN COMMS. 320, 320–25 (1986). In its most common form, it provided so-called 2B+D service: two 64 Kbps "bearer" channels, and a 16 Kbps data channel for signaling, e.g., call setup and teardown. *Id.* The two bearer channels could be combined into a single 128 Kbps link for pure data; this is more than twice as fast as any single-line analog phone modem can ever provide. For a variety of reasons, it never caught on in the United States as a common service.

[20] In the analog telephony era, wiretapping was very straightforward. It was almost as easy as plugging in a new extension phone, though some additional circuitry was needed or the target was not able to dial new calls or even hang up on a call. A law enforcement agent literally connected a pair of wires to the phone line going to the suspect's location; this connection could be done in the phone company's central office, at any point along the phone cable from the central office to the target, or, in the case of multiple occupancy buildings, in some utility space in the building. When the phone company started running digital signals to neighborhoods via "Subscriber Loop Carriers" (*see, e.g.*, Voyager[TNO], *The Subscriber Loop Carrier (Slick)*, PHRACK 8:52, Jan. 26, 1998 at article 11, http://www.phrack.com/issues.html?issue=52&id=11), the tap could be done in the same way, albeit from the neighborhood Remote Terminal onwards. Generally, a "loop extender" is employed to route the intercepted conversations back to a suitable facility. *See* Micah Sherr, Eric Cronin, Sandy Clark & Matt Blaze, *Signaling Vulnerabilities in Wiretapping Systems*, IEEE SECURITY & PRIVACY, Nov./Dec. 2005, at 13 vol. 3, no. 6 (2005): 13–25, http://www.crypto.com/papers/wiretap.pdf.

[21] *File 1—May '92 Version of FBI Digital Telephony Proposal*, COMPUTER UNDERGROUND DIG. (July 5, 1992), http://cu-digest.org/CUDS4/cud429.txt; *see also* WHITFIELD DIFFIE & SUSAN LANDAU, PRIVACY ON THE LINE: THE POLITICS OF WIRETAPPING AND ENCRYPTION 205–06 (Updated & Expanded ed. 2007).

[22] *See, e.g.*, Steven Levy, *Battle of the Clipper Chip*, N.Y. TIMES, June 12, 1994, http://www.nytimes.com/1994/06/12/magazine/battle-of-the-clipper-chip.html.

[23] In 1992, the FBI proposed legislation that would have "allowed the technical design mandates on *any* provider of *any* electronic communications, *including* the Internet." Corrected Petition for Rehearing En Banc at 12, Am. Council on Educ. v FCC, No. 15-0504 (D.C. Cir. July 28, 2006), *available at*

¶14 CALEA was intended to apply only to telephony. More precisely, CALEA was intended to apply only to "local exchange service," i.e., local phone service but not long distance carriers.[25] Then-FBI Director Louis Freeh made clear in his 1994 Congressional testimony that the Internet was not covered:

> Mr. FREEH. . . . We are really talking about phone-to-phone conversations which travel over a telecommunications network in whole or part. That is the arena of criminal opportunity that we are discussing.
>
> Senator PRESSLER. What other portions of the information superhighway could people communicate with the new technology that there is not now a means of listening in or following?
>
> Mr. FREEH. From what I understand, and again, I am probably the worst person in this room to answer the question, communications between private computers, PC-PC communications, not utilizing a telecommunications common net, would be one vast arena, the Internet system, many of the private communications systems which are evolving. Those we are not going to be on by the design of this legislation.
>
> Senator PRESSLER. Are you seeking to be able to access those communications also in some other legislation?
>
> Mr. FREEH. No, we are not. We are satisfied with this bill. I think it delimits the most important area and also makes for the consensus, which I think it pretty much has at this point. [26]

¶15 This consensus was reflected in the law, which defined a "telecommunications carrier" to include "a person or entity engaged in providing wire or electronic communication switching or transmission service to the extent that the Commission finds that such service is a replacement for a substantial portion of the local telephone exchange service and that it is in the public interest to deem such a person or entity to be a telecommunications carrier for purposes of this subchapter."[27]

¶16 More recently, CALEA coverage has been extended to "last mile" service: the link between a residence or business and its ISP. Although controversial because of Freeh's testimony and the exclusion of information services in CALEA, the FCC and the courts have held that this class of link is not included in the information services exclusion.[28]

---

https://www.cdt.org/wiretap/calea/20060731calearehearing.pdf. The proposal was "rejected out of hand". *Id.* (quoting *Digital Telephony and Law Enforcement Access to Advanced Telecommunications Technologies and Services: J. Hearings on H.R. 4922 and S. 2375 Before the Subcomm. on Tech. and the Law of the S. Comm. on the Judiciary & Subcomm. on Civil and Constitutional Rights of the H. Comm. on the Judiciary*, 103rd Cong. 49 (1994)).

[24] 47 U.S.C. § 1001(8)(C)(i) (2006).

[25] *Digital Telephony and Law Enforcement Access to Advanced Telecommunications Technologies and Services, supra* note 23, at 136.

[26] *Id.* at 202.

[27] *See* 47 U.S.C. § 1001(8)(B)(ii) (2006).

[28] Am. Council on Educ. v. FCC, 451 F.3d 226, 230 (D.C. Cir. 2006).

More precisely, the FCC made that ruling, and, relying on *Chevron* deference, the Court of Appeals upheld the FCC's ruling. [29]

¶17 Though important, this change to CALEA is of less concern to law enforcement than is the fate of the traditional telephone network. It is going away, and far faster than anyone had forecast. Already, more than 35% of American households do not have landline phone service, and about 16% more who have landlines never or almost never receive calls on them.[30] Indeed, the working assumption in the Federal Communications Commission (FCC) is that the PSTN will effectively cease to exist by 2018.[31]

B. Wiretap Consequences of Splitting Services and Infrastructure

¶18 It might be tempting to say that the coming end of the PSTN vindicates the FBI's vision when it proposed CALEA. The actual situation, though, is far more complex; the decoupling of services from the physical link has destroyed the chokepoint at which CALEA could be applied. This does not appear to have been anticipated at the time of CALEA's passage.

¶19 A paradigmatic case in which the decoupling presents serious wiretapping problems is when communication occurs through use of Voice over Internet Protocol (VoIP). A VoIP phone provider can be located far from its subscribers; indeed, it could be in another, possibly unfriendly, country. Furthermore, the "signaling path"—the set of links that carry the call setup messages—can differ from the "voice path"—the links that carry the actual conversation.[32] (Tapping the last mile connection is likely fruitless, since VoIP connections are often encrypted.)

¶20 This is best explained by a diagram. Figure 1 shows a plausible setup for a VoIP call from Alice to Bob.[33] Alice's and Bob's phones are each connected to their own ISPs, Net 1 and Net 4. They each subscribe to their own VoIP provider, which in turn is connected to their ISPs. The signaling messages—that is, the messages used to set up the call, indicate ringing, etc.—go from Alice's phone, through her ISP to VoIP Provider 1's ISP, to her phone company. It then contacts VoIP Provider 2, via its ISP; VoIP Provider 2 sends a message through Net 4 to Bob's phone. The actual voice path, however, goes directly from Net 1 to Net 4; neither Net 2, Net 3, nor the VoIP providers even carry the actual conversation. As noted, any or all of the messages may be encrypted.

---

[29] *See id.* at 231 (citing Chevron U.S.A. Inc. v. Natural Res. Def. Council, Inc., 467 U.S. 837, 842–43 (1984)).

[30] STEPHEN J. BLUMBERG & JULIAN V. LUKE, NAT'L CTR. FOR HEALTH STATISTICS, WIRELESS SUBSTITUTION: EARLY RELEASE OF ESTIMATES FROM THE NATIONAL HEALTH INTERVIEW SURVEY, JANUARY-JUNE 2012 1 (Dec. 2012), *available at* http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201212.pdf.

[31] TECHNICAL ADVISORY COUNCIL, FEDERAL COMMS. COMMISSION, SUMMARY OF MEETING (Sept. 27, 2011), *available at* http://transition.fcc.gov/oet/tac/tacdocs/tac-meeting-summary-9-27-11-final.docx.

[32] *See* STEVEN BELLOVIN, MATT BLAZE, ERNEST BRICKELL, CLINTON BROOKS, VINTON CERF, WHITFIELD DIFFIE, SUSAN LANDAU, JON PETERSON & JOHN TREICHLER, SECURITY IMPLICATIONS OF APPLYING THE COMMUNICATIONS ASSISTANCE TO LAW ENFORCEMENT ACT TO VOICE OVER IP 2–7 (2006), *available at* https://www.cs.columbia.edu/~smb/papers/CALEAVOIPreport.pdf (demonstrating a VoIP network in Figure 1 on pg. 4).

[33] This figure is adapted from *id.* at 4.

¶21 In this setup, where can a tap be placed? On any of the ISPs? Law enforcement has no a priori information where Alice and Bob will be—their current IP addresses—prior to their setting up a call, so law enforcement cannot serve the ISPs with a wiretap order. To make matters worse, the ISPs have nothing to do with the VoIP call, nor can they read the encrypted traffic. How about at one of the VoIP providers? They do not see the voice traffic. And, of course, they may be in a different jurisdiction (for example, Skype was originally hosted in Luxembourg). This is a scenario that has no points amenable to a CALEA-like solution.

¶22 Other services are more complex still. Consider the new phone service being offered by Republic Wireless, which uses a combination of IP and PSTN networks to make a call. The service is intended to operate primarily over WiFi networks and the Internet; however, it can switch to Sprint's 3G cellular network as needed.[34] Where could a CALEA tap be placed? A tap could certainly be placed on the Internet-facing side of Republic's facilities,[35] but that would miss Sprint calls. Conversely, there could be one on Sprint's network, but that would miss calls made via VoIP. It is of course possible to place taps on both networks, but the protocols are very different. Since the ordinary signaling mechanisms are not used, special code would be needed to hand off not the call and the information necessary to carry out the tap.[36] Pen registers would be even more involved because the types of information easily recorded—phone numbers versus IP addresses—would vary.

¶23 Apart from reasonably straightforward (though structurally different) PSTN replacements, a large variety of other communications schemes have gained popularity. Email and text messages are two obvious examples, though even these pose challenges for law enforcement due to issues of personal jurisdiction and lack of real-time access to content. Skype is perhaps the most extreme case. Its architecture, which an FCC report calls "over the top,"[37] has no central switches. Even apart from questions of jurisdiction, there are *no* locations where a CALEA-style interface could be provided. Everything is done peer-to-peer; ordinary Skype users forward signaling traffic for each other.[38]

---

[34] Walt Mossberg, *For $19, an Unlimited Phone Plan, Some Flaws*, WALL ST. J., Feb. 19, 2013, http://allthingsd.com/20130219/for-19-an-unlimited-phone-plan-some-flaws/.

[35] Tapping the customer's own Internet connection would not suffice, since the customer is likely to use multiple WiFi networks that such a tap would miss. Also, while Republic Wireless is a U.S. company, there is no reason why a similar service could not be offered by an offshore company over which U.S. courts have no jurisdiction.

[36] As of this writing, the Republic Wireless network cannot do handoffs of an in-progress call from a WiFi network to Sprint or vice-versa. According to Mossberg, *supra* note 34, that feature is planned for the near future.

[37] CRITICAL LEGACY TRANSITION WORKING GROUP, SUN-SETTING THE PSTN (2011), *available at* http://transition.fcc.gov/oet/tac/tacdocs/meeting92711/Sun-Setting_the_PSTN_Paper_V03.docx.

[38] It is unclear how true this still is. Skype has long used a "supernode," a well-connected user computer that carries considerably more traffic. Of late, Microsoft—the current owner of Skype—has been deploying dedicated supernodes in its own data centers. *See* Dan Goodin, *Skype Replaces P2P Supernodes with Linux Boxes Hosted by Microsoft (Updated)*, ARS TECHNICA (May 1, 2012, 12:23 PM), http://arstechnica.com/business/2012/05/skype-replaces-p2p-supernodes-with-linux-boxes-hosted-by-microsoft/. There have been some allegations that the replacement was done precisely to permit surveillance. *See*, *e.g.*, John D. Sutter, *Can Skype 'Wiretap' Video C*alls*?*, CNN, http://www.cnn.com/2012/07/24/tech/web/skype-surveillance (last updated July 24, 2012, 4:30 PM). However, these are disputed by Mary Branscombe, who insists the changes in architecture are about "improving performance and not appropriating bandwidth." *Forget the Conspiracy Theories: Skype's Supernodes Belong in the Cloud*, ZDNET (July 27, 2012, 1:52 PM), *available at*

Because of this, there are no trusted elements that could serve as wiretap nodes, at least for pen register orders. Furthermore, calls are always encrypted end-to-end.[39]

¶24 It is useful to contrast the Skype architecture with the conventional client-server architecture shown in Figure 1. In the conventional configuration, the VoIP providers run servers to which the individual phones—the clients—connect. These are architecturally different roles; when setting up calls, phones talk only to their associated servers and the servers talk to the clients and to each other. It is not possible for Alice's phone to contact VoIP Provider 2 directly; they have no business relationship, and therefore cannot set up a direct network link.[40] In a peer-to-peer setup such as that used by Skype, there are *no* servers, i.e., no architecturally distinguished roles.[41] Rather, *every* computer or device running a Skype client can participate in the signaling. Alice's phone (somehow) finds another Skype client and asks it to connect to Bob. This node finds another, which finds another, etc., until Bob's phone is located.[42] At that point, Alice's and Bob's phones exchange signaling messages and set up the voice path. This voice path is in principle direct, though for various reasons, including the existence of firewalls, other Skype nodes may relay the (encrypted) voice packets. The lack of central servers, other than for user registration and enhanced services such as calling out to PSTN numbers, dramatically cuts the operational costs and allows Skype to offer free or extremely cheap phone calls.[43]

¶25 All that said, one of Snowden's revelations was that the NSA can indeed intercept Skype calls.[44] No technical details have been disclosed; all we know is that the NSA can

---

http://www.zdnet.com/forget-the-conspiracy-theories-skypes-supernodes-belong-in-the-cloud-7000001720/. The one-time principal architect of Skype, Matthew Kaufman, has explained that the change was done to accommodate the switch from always-on desktops to battery-powered mobile devices. *See* Zack Whittaker, *Skype Ditched Peer-to-Peer Supernodes for Scalability, not Surveillance*, ZDNET (June 24, 2013, 4:02 PM), http://www.zdnet.com/skype-ditched-peer-to-peer-supernodes-for-scalability-not-surveillance-7000017215/. Microsoft has applied for a patent on mechanisms for eavesdropping on VoIP networks, and some commentators have alleged that this technology will be incorporated into Skype. *See, e.g.*, Jaikumar Vijayan, *Microsoft Seeks Patent for Spy Tech for Skype*, COMPUTERWORLD (June 28, 2011, 5:06 PM), https://www.computerworld.com/s/article/9218002/Microsoft_seeks_patent_for_spy_tech_for_Skype.

[39] For a good, albeit dated—and paid for by Skype—review of the encryption architecture, see TOM BERSON, ANAGRAM LABS., SKYPE SECURITY EVALUATION (Oct. 18, 2005), http://www.anagram.com/berson/abskyeval.html.

[40] This is not a technical limitation per se; however, VoIP Provider 2 knows nothing of Alice's phone, and hence is not willing to believe any assertions about its phone number, the person who uses it, etc. More importantly, because of the lack of a business relationship, it will not provide service to Alice's phone since it will not be paid for its efforts.

[41] This is not strictly true. The Skype servers, however, are involved only in registering new users and providing them with cryptographic credentials. They are not involved in call setup, let alone being in the voice path. *See What Are P2P Communications?*, SKYPE, https://support.skype.com/en/faq/fa10983/what-are-p2p-communications (last visited Nov. 11, 2013).

[42] How the call eventually reaches Bob's phone is a rather complex technical matter, and not relevant here. Let it suffice to say that Skype nodes regularly exchange enough navigational messages that it can be done.

[43] The lack of central servers was a deliberate architectural choice, designed to evade legal constraints. Architecturally, Skype was based on the Kazaa file-sharing network, which was in turn designed to operate without vulnerable nodes that could be targeted by copyright infringement lawsuits. For information about the history and technology of Skype, see generally Doug Aamoth, *A Brief History of Skype*, Time (May 10, 2011), http://techland.time.com/2011/05/10/a-brief-history-of-skype/.

[44] *See* Glenn Greenwald, Ewen MacAskill, Laura Poitras, Spencer Ackerman & Dominic Rushe,

intercept audio and video, with complete metadata. It remains unclear if the solution is one that is usable by ordinary law enforcement, or if it relies on techniques (such as advanced cryptanalysis) that rely on the intelligence community's capabilities.[45]

¶26 Text messaging has also changed. Originally, it was a simple protocol for mobile phones. Recently, a number of variant implementations have been developed. Some provide a better experience in some fashion (for example, Apple's iMessage will send copies of inbound messages to all of a user's devices, including tablets and Mac computers as well as phones); others do things like provide phone-like text messaging for non-phone devices such as tablets.[46]

¶27 Non-traditional text messaging applications have already proven problematic. According to one report, attributed to a Drug Enforcement Administration memo, the encryption used by Apple's iMessage has already stymied wiretap orders.[47] There are even instant messaging applications designed not just to encrypt traffic, but to provide "repudiation," the ability to deny that you sent certain traffic.[48]

¶28 Further, many non-obvious communications mechanisms can serve for direct communications as well. In one well-known case, General David Petraeus and Paula Broadwell sent each other messages by creating and saving draft email messages in a shared Gmail account.[49] Additionally, many multiplayer games include text or even real-time voice communications between players; while nominally intended to lend realism to the game—soldiers in the same unit in action games can talk to each other and fighters on

---

Microsoft Handed the NSA Access to Encrypted Messages, THE GUARDIAN, July 11, 2013, http://www.guardian.co.uk/world/2013/jul/11/microsoft-nsa-collaboration-user-data/print.

[45] Microsoft claims that in 2012 it produced "no content" to law enforcement from Skype calls. *See* Brad Smith, *Microsoft Releases 2012 Law Enforcement Requests Report*, MICROSOFT ON THE ISSUES (Mar. 21, 2013, 6:00 AM), https://blogs.technet.com/b/microsoft_on_the_issues/archive/2013/03/21/microsoft-releases-2012-law enforcement-requests-report.aspx. The reports themselves are available at *Law Enforcement Requests Report*, MICROSOFT, https://www.microsoft.com/about/corporatecitizenship/en-us/reporting/transparency/ (last visited Oct. 4, 2013).

[46] There are many such applications currently available and new ones are constantly appearing. *See, e.g.*, Tanya Menoni, *6 Free iPhone & iPod Touch Texting Apps*, ABOUT.COM, http://ipod.about.com/od/iphoneappsreviews/tp/4-Ways-To-Text-With-The-Ipod-Touch.htm (last visited Sept. 21, 2013).

[47] *See* Declan McCullagh & Jennifer Van Grove, *Apple's iMessage Encryption Trips up Feds' Surveillance*, CNET NEWS (Apr. 4, 2013, 4:00 AM), http://news.cnet.com/8301-13578_3-57577887-38/apples-imessage-encryption-trips-up-feds-surveillance/. Because the design of the protocol has not been published, it has not been possible for outside experts to assess this claim. Some have asserted, based on certain externally visible characteristics (like the ability to do a password reset and still see old messages), that the messages must be stored unencrypted on Apple's servers. *See*, *e.g.*, Julian Sanchez, *Untappable Apple or DEA Disinformation*?, CATO INSTITUTE (Apr. 4, 2013, 5:24 PM), http://www.cato.org/blog/untappable-apple-or-dea-disinformation. If that is true, a court order under the Stored Communications Act, 18 U.S.C. §§ 2701–2712 (2006), would provide law enforcement with the content, albeit perhaps not in real-time.

[48] *See* Nikita Borisov, Ian Goldberg & Eric Brewer, *Off-the-Record Communication, or, Why Not to Use PGP*, PROC. 2004 ACM WORKSHOP ON PRIVACY ELECTRONIC SOC'Y 77, 77–78 (2004). Note that "repudiation" (derived from its more cryptographic common counterpart, "nonrepudiation") is used here as a computer scientist would use it—it refers to certain cryptographic properties: in terms of the encryption mechanisms used, it is not possible to show mathematically that a given person has sent certain messages. Concepts that a lawyer might rely on, e.g., circumstantial evidence or eyewitness testimony to the contrary, are not part of this mathematical model. Software to add repudiation to several IM programs is available at https://otr.cypherpunks.ca/.

[49] *See* Max Fisher, *Here's the E-Mail Trick Petraeus and Broadwell Used to Communicate*, WASH. POST, Nov. 12, 2012, http://www.washingtonpost.com/blogs/worldviews/wp/2012/11/12/heres-the-e-mail-trick-petraeus-and-broadwell-used-to-communicate/.

opposing sides can yell challenges or insults—such applications can also be used for surreptitious communications. Given that the Internet *is* a communications network, this raises the specter that *all* programs can be considered communications systems.

C. New Technologies: Going Dark or Going Bright?

¶29 Collectively, the changes in telephony, the rise of new communications technology, and (to some extent) the increasing use of encryption, have been called the "Going Dark" problem because law enforcement has been unable to keep up with these changes and is losing access to criminals' communications. Technology works both ways, however; others have rightly claimed that modern developments have actually *increased* the practical ability of law enforcement to monitor criminals' behavior via assorted forms of metadata analysis; these analyses do not require warrants[50] So, how serious is the Going Dark problem? How has the balance changed?

¶30 A firm, quantitative answer to the former question is probably not possible. We cannot determine how many tap attempts would fail because law enforcement has said that it does not seek wiretap orders for calls it cannot intercept.[51] Furthermore, the situation is not static since both criminals and police adapt their tactics in response to each other's capabilities and tactics. Consider cellular telephony. Under the Omnibus Crime Control and Safe Streets Act, the Administrative Office of the U.S. Courts (AO) reports annually on all Title III wiretaps.[52] The reports include the offense under investigation, the names of the prosecuting attorney and authorizing judge, the number of intercepts conducted and number of incriminating intercepts, the cost of the surveillance, etc.[53] In 2000, the report began listing how many wiretaps were of portable devices; in that year, they comprised 719 out of a total 1,190 Title III wiretaps.[54] By 2009, it was 2,276 out of 2,376, or 96%.[55] This, of course, mirrors the trend of society as a whole; as noted, a majority of Americans rely on mobile phones for most of their incoming calls.[56]

¶31 Reliance on mobile phones provides a partial answer to the question of gaining and losing capabilities as a result of modern communication systems. Because mobile phones are far more likely to capture the target's conversations—rather than those of a spouse or business associate—mobile phone taps are more valuable than wireline taps. Furthermore, mobile data can include information on a person's location, which means

---

[50] The claim is that the existence and availability of other information, such as location data, commercial data dossiers, and readily available contact information has given law enforcement far more than technology has taken away. *See, e.g.*, SUSAN LANDAU, SURVEILLANCE OR SECURITY: THE RISKS POSED BY NEW WIRETAPPING TECHNOLOGIES, 99–101 (2011), and Peter Swire & Kenesa Ahmad, *Encryption and Globalization*, 13 COLUM. SCI. & TECH. L. REV. 416, 463–64 (2012).

[51] Personal comments to Susan Landau; *see also Going Dark: Lawful Electronic Surveillance in the Face of New Technologies*, *supra* note 2, at 12 (prepared statement of Valerie Caproni, General Counsel, Federal Bureau of Investigation).

[52] The reports are available at *Wiretap Reports Archive*, U.S. CTS., http://www.uscourts.gov/Statistics/WiretapReports/WiretapReports_Archive.aspx (last visited Feb. 25, 2013).

[53] See the list of text and appendix tables in, for example, ADMIN. OFFICE OF THE U.S. COURTS, 2011 WIRETAP REPORT 3–4 (June 2012).

[54] Admin. Office of the U.S. Courts, 2000 Wiretap Report 30 (Apr. 2001).

[55] Admin. Office of the U.S. Courts, 2009 Wiretap Report 32 (Apr. 2010).

[56] *See* Blumberg & Luke, *supra* note 30.

that 96% of wiretapped communications provide law enforcement with extremely valuable location information. The same is true of many Internet connections, whether fixed or mobile.[57] In other words, the prevalence of immediate communications—texting, cellular calls, and the like—and centralized services—for example, Gmail and Facebook—has vastly simplified law enforcement's ability to both track suspects and access their communications.

¶32 Another way to assess the overall risk of communications that law enforcement cannot monitor is to look at the net effect of prior threats: how much has the police's ability to monitor communications been affected by prior technological changes, such as encryption? The issue has long been a concern, so much so that in 1993, the government announced the so-called "Clipper Chip"—an encryption device designed to enable the government to read otherwise encrypted traffic.[58] The AO wiretap reports now include data on how often encryption has been encountered.[59] The data are interesting. The total between 2001-2011 is eighty-seven; of these, only one was the subject of a federal wiretap order. Moreover, the AO noted that law enforcement was able to decrypt all of the wiretapped communications. [60]

¶33 There is not a lack of communications products that provide end-to-end encryption, such as RIM's Blackberries, Skype, etc. While there are smart criminals who do use—

[57] A technology known as "IP geolocation" can be used to determine where an Internet user is located. It is frequently used to enforce geographic restrictions on access to content. *See, e.g.*, *Terms of Use Agreement*, MLB.COM, http://mlb.mlb.com/mlb/official_info/about_mlb_com/terms_of_use.jsp#4I (last visited Sept. 24, 2013) ("Due to the foregoing blackout restrictions, you may be required to authorized MLBAM to access your location data . . . ."). While many IP geolocation services provide fairly coarse resolution, some companies have done a far better job of geolocation by combining IP address information with outside data, such as search queries, purchase delivery records, etc.

[58] *See* John Markoff, *Electronics Plan Aims to Balance Government Access with Privacy*, N.Y. TIMES, Apr. 16, 1993, http://www.nytimes.com/1993/04/16/us/electronics-plan-aims-to-balance-government-access-with-privacy.html; *see also* Matt Blaze, *Notes on Key Escrow Meeting with NSA*, RISKS DIG. (Feb. 8, 1994, 4:04 PM), http://catless.ncl.ac.uk/Risks/15.48.html#subj1 ("They indicated that the thinking was not that criminals would use key escrowed crypto, but that they should not field a system that criminals could easily use against them. The existence of key escrow would deter them from using crypto in the first place. The FBI representative said that they expect to catch '~only the stupid criminals~' through the escrow system.").

[59] As a result of Public Law 106-197, since 2000 the AO has reported the annual total of state and federal wiretap orders encountering encryption. *See* Pub. L. No. 106-197, § 2, 114 Stat. 246 (codified at 18 U.S.C. § 2519(2)(b)(iv) (2006)).

[60] *See* ADMIN. OFFICE OF THE U.S. COURTS, 2001 WIRETAP REPORT 5 (May 2002) (reporting sixteen wiretaps encountering encryption in 2001); ADMIN. OFFICE OF THE U.S. COURTS, 2002 WIRETAP REPORT 5 (Apr. 2003) (reporting sixteen wiretaps encountering encryption in 2002 and an additional eighteen in 2001); ADMIN. OFFICE OF THE U.S. COURTS, 2003 WIRETAP REPORT 5 (Apr. 2004) (reporting one wiretap encountered encryption in 2003); ADMIN. OFFICE OF THE U.S. COURTS, 2004 WIRETAP REPORT 5 (Apr. 2005) (reporting two wiretaps encountered encryption in 2004); ADMIN. OFFICE OF THE U.S. COURTS, 2005 WIRETAP REPORT 5 (Apr. 2006) (reporting thirteen wiretaps encountered encryption in 2005); ADMIN. OFFICE OF THE U.S. COURTS, 2006 WIRETAP REPORT 5 (Apr. 2007) (reporting no wiretaps encountered encryption in 2006); ADMIN. OFFICE OF THE U.S. COURTS, 2007 WIRETAP REPORT 5 (Apr. 2008) (reporting no wiretaps encountered encryption in 2007); ADMIN. OFFICE OF THE U.S. COURTS, 2008 WIRETAP REPORT 5 (Apr. 2009) (reporting two wiretaps encountered encryption in 2008); ADMIN. OFFICE OF THE U.S. COURTS, 2009 WIRETAP REPORT 5 (Apr. 2010) (reporting one wiretap encountered encryption in 2009); ADMIN. OFFICE OF THE U.S. COURTS, 2010 WIRETAP REPORT 9 (reporting six wiretaps encountered encryption in 2010); ADMIN. OFFICE OF THE U.S. COURTS, 2011 WIRETAP REPORT 5 (June 2012) (reporting twelve wiretaps encountered encryption in 2011). All but one these were state wiretaps (the one federal wiretap that encountered encryption occurred in 2004).

and even build—their own encrypted communications networks,[61] the AO numbers demonstrate that criminals against whom Title III wiretaps are used typically do not do so. Instead, they tend to use simple solutions: Commercial Off-The-Shelf (COTS) equipment and communications in the cloud (e.g., Gmail and Facebook). Few use the peer-to-peer communication channels that pose problems for law enforcement wiretaps.[62] The implication for law enforcement use of vulnerabilities for performing Title III wiretaps is simple: law enforcement will not need to go that route very often.

¶34 Put another way, criminals are like other people: few use cutting edge or experimental devices to communicate. Instead, they stick with COTS products. If nothing else, COTS products are generally easier to use and work better, a definite advantage. Furthermore, understanding of the fine details of new technologies, such as encryption, is limited. The distinction between end-to-end encryption and client-to-server encryption is not understood by most people, criminals included. Similarly, the question of whether the encryption is going to the right party is often not even asked. Good software usually performs the proper checks,[63] but even production code has had serious errors.[64]

¶35 From this perspective, the most serious threat to legally authorized wiretapping is exemplified by the Skype architecture. Virtually all email services feature (at most) encryption from the client to the mail server; the messages reside in plaintext on the mail providers' disks.[65] By contrast, Skype provides transparent end-to-end encryption from the sender to the receiver; there is no middle man that sees the communication "in the clear." Skype is gaining an increasing share of the international telephony market.[66] Even with Skype, however, investigators are not completely shut out. Though the content is

---

[61] *See, e.g.*, Spencer Ackerman, *Radio Zeta: How Mexico's Drug Cartels Stay Networked*, WIRED (Dec. 27, 2011, 3:41 PM), http://www.wired.com/dangerroom/2011/12/cartel-radio-mexico/.

[62] *See* sources cited *supra* note 61.

[63] The best example is how web browsers use encryption. When a browser connects via HTTPS, the web server sends its "certificate" to the browser. A full explanation of certificates is out of scope here; what is important is that they contain a cryptographically protected association between the website's name and a unique cryptographic key. Browsers verify that the name of the website contacted actually appears in the certificate; thus, you will not end up with an encrypted connection to EvilHackerDudez.org when you are trying to log in to your bank.

[64] Generally speaking, encryption on the Internet requires use of a "Public Key Infrastructure". *See, e.g.,* RUSS HOUSLEY, TIM POLK, PLANNING FOR PKI: BEST PRACTICES GUIDE FOR DEPLOYING PUBLIC KEY INFRASTRUCTURE (2001). Web connections and many other sorts of traffic are protected using the "Secure Socket Layer". *See, .e.g.,* ERIC RESCORLA, SSL AND TLS: DESIGNING AND BUILDING SECURE SYSTEMS (2001). For a discussion of applications that do some checks incorrectly, see Sascha Fahl, Marian Harbach, Thomas Muders, Matthew Smith, Lars Baumgärtner & Bernd Freisleben, *Why Eve and Mallory Love Android: An Analysis of Android SSL (In)Security*," PROC. 2012 ACM CONF. ON COMPUTER AND COMM. SECURITY 50 (2012).

[65] Although probably technically feasible (though difficult, given the need to comply with industry standards), it is highly unlikely that providers, such as Google's Gmail and Microsoft's Hotmail, will switch to end-to-end encryption. There is little consumer demand, it is difficult, and Google at least relies on being able to scan messages in order to display appropriate ads. It cannot do so if the messages are encrypted.

[66] *See The Bell Tolls for Telcos?*, TELEGEOGRAPHY (Feb. 13, 2013), http://www.telegeography.com/products/commsupdate/articles/2013/02/13/the-bell-tolls-for-telcos/ ("TeleGeography estimates that cross-border Skype-to-Skype voice and video traffic grew 44% in 2012 . . . .").

encrypted, Skype leaks the IP addresses of its users.[67] This provides the equivalent of pen register data and often location information as well.[68]

¶36 Technological changes will also play a role in law-enforcement's ability to wiretap. However, it is difficult at this point to make confident predictions about the future direction of technology. The two popular trends, cloud computing and peer-to-peer networking, have opposite effects on law enforcement's ability to monitor communications.

¶37 Cloud computing moves more and more storage and computation to distant, network-connected servers. Today's email scenario is an old but telling example: all of a target's email passes through easily monitored remote servers. These servers tend to have stringent backup regimens and log everything, out of operational necessity. Even deletion operations are less than permanent;[69] preservation of data is paramount, even under extreme circumstances.[70] In theory, cloud storage could be encrypted; in practice, because of users' desire to be able to search their email messages and the lack of customer demand, there has been little, if any, real-world deployment.[71] In fact, in order to better serve ads, the Facebook and Google business models rely on the cloud data being unencrypted.

¶38 The second trend, peer-to-peer, is decentralized, with no convenient points for wiretaps or content monitoring. Rather than clients and servers, computers, phones, and other gadgets talk to each other. Consider today's email architecture, where messages from Alice to Bob flow from her phone to her ISP's outbound mail server to Bob's ISP's inbound mail server to Bob's computer. Must it be done that way, or can Alice's phone talk directly to Bob's computers? Indeed, in some scenarios even ISPs disappear; in a technology known as "mesh networking," computers ask other peer computers to relay their traffic.[72] One very active area of development for mesh networks is car-to-car traffic for automotive safety and congestion control;[73] this could end up denying law

---

[67] *See* Joel Schectman, *Skype Knew of Security Flaw Since November 2010, Researchers Say*, WALL ST. J., May 1, 2012, http://blogs.wsj.com/cio/2012/05/01/skype-knew-of-security-flaw-since-november-2010-researchers-say/.

[68] *See supra* note 57.

[69] *See, e.g.*, *Microsoft Services Agreement*, WINDOWS, http://windows.microsoft.com/en-us/windows-live/microsoft-services-agreement (last updated Aug. 27, 2012) (stating in Section 4.3: "please note that while content you have deleted or that is associated with a closed account may not be accessible to you, it may still remain on our systems for a period of time"). Other providers have similar provisions out of technical necessity.

[70] In 2010, a software problem caused thousands of Microsoft's Hotmail users to lose their entire mailboxes. Although it took several days, Microsoft was able to retrieve and restore the data from backup media. *See* Sebastian Anthony, *Hotmail Users Lose Entire Email Inboxes, Microsoft Restores Them 5 Days Later,* SWITCHED (Jan. 3, 2011, 6:50 AM), http://downloadsquad.switched.com/2011/01/03/hotmail-users-lose-entire-email-inboxes-microsoft-restores-them/.

[71] Encrypted storage and encrypted search are active research areas. However, except under special circumstances (e.g., a structured database, as opposed to email), encrypted remote search remains much more expensive than the plaintext equivalent and is likely to remain that way.

[72] *See, e.g.*, Rafe Needleman, *Unbreakable: Mesh Networks are in your Smartphone's Future,* CNET (July 13, 2013, 5:00 AM), http://www.cnet.com/8301-30976_1-57471447-10348864/unbreakable-mesh-networks-are-in-your-smartphones-future/.

[73] *See* Jon Brodkin, *Wireless Mesh Networks at 65MPH—Linking Cars to Prevent Crashes*, ARS TECHNICA (Jan. 9, 2013, 6:50 PM), http://arstechnica.com/information-technology/2013/01/wireless-mesh-networks-at-65mph-linking-cars-to-prevent-crashes/.

enforcement access to location data from cellular networks, because the phones would be talking to other phones in a peer-to-peer fashion rather than registering with phone company-run cell towers.

¶39 In a cloud world, monitoring will be easier; in a peer-to-peer world, it will be harder. It is quite possible that both trends will continue, with different applications and different markets opting for one solution over the other.

D. The TPWG's Tracking Preferences Expression Standard

¶40 CALEA II, the extension of CALEA to cover all communications applications, poses three serious problems: (1) it hinders innovation by restricting communications application developers to certain topological and trust models, (2) it imposes a financial tax on software, and (3) it creates security holes (and hence increases the risk of computer crime, cyberepionage, and cyberterrorism). This last point seems to be mentioned least in debates, although arguably it is the most important since it cannot be addressed by perfect (or at least very, very good) software development practices, reuse of standard CALEA compliance libraries, or both.

¶41 An implicit assumption behind CALEA-style laws is that there is a "good" place where intercepts can take place. Such a place would be run by trustworthy people who are not implicated in the investigation,[74] and be located where the tap cannot be detected. More or less of necessity, this translates to relying on a centralized facility, preferably one run by a large, accountable company. This worked well for the telephone taps, where all lines were connected to a phone switch run by a conventional phone company. By contrast, consider a Skype-like architecture with transmissions over a mesh network. There are *no* large companies involved in either the call setup or data paths; rather, both use effectively random links. Furthermore, there may be little or no logging present; not only is the path used for one call probably not the path used for another, there will be no logs to show what paths were used. This means little or no accountability for any parties who leak information, and no assurance whatsoever that anyone will be able to complete the tap.

¶42 The fact that a peer-to-peer service is not facilities-based—it does not rely on provider-owned equipment—also means there may be no parties to whom the law applies. For example, CALEA requires that "a telecommunications carrier shall ensure that its equipment, facilities, or services . . . enable[e] the government . . . to intercept . . . all wire and electronic communications carried by the carrier . . . concurrently with their transmission to or from the subscriber's equipment."[75] Based on the definition of telecommunications carrier provided in the statute, however, there are no carriers in some peer-to-peer architectures: "The term 'telecommunications carrier', means a person or entity engaged in the transmission or switching of wire or electronic communications as a

[74] 18 U.S.C. § 2511(2)(a)(ii) (2006) ("No provider of wire or electronic communication service, officer, employee, or agent thereof . . . shall disclose the existence of any interception or surveillance or the device used to accomplish the interception or surveillance with respect to which the person has been furnished a court order or certification under this chapter . . . . Any such disclosure, shall render such person liable for the civil damages provided for in section 2520.") Damages after the fact are one thing, but law enforcement would much rather the tap were not disclosed in the first place.

[75] 47 U.S.C. § 1002(a) (2006).

common carrier for hire . . . ."[76] or "a person or entity engaged in providing wire or electronic communication switching or transmission service to the extent that the Commission finds that such service is a replacement for a substantial portion of the local telephone exchange service."[77] In a peer-to-peer network, there is no such thing as "local" service; a "peer" need not be geographically close to any of the parties. Similarly, there may be no "manufacturer of telecommunications transmission or switching equipment" who can be compelled to "make available to the telecommunications carriers using its equipment, facilities, or services such features or modifications as are necessary to permit such carriers to comply with the capability requirements";[78] the peer nodes and any commercial entities involved in the service operation (and there need not be any) may be located outside of U.S. jurisdiction.[79]

¶43 To sum up, the laws assume a trustworthy, disinterested intermediary within the court's jurisdiction. But as the net moves towards a more decentralized architecture, such third parties simply do not exist. Current technological trends pose a serious (and probably insurmountable) philosophical challenge to CALEA-style laws.

¶44 If CALEA were to be extended to cover IP-based communications, the law would have to specify which part of the service is responsible for supplying wiretap capability. As noted earlier, peer-to-peer networking is one plausible path for the technical future. Imposing requirements that effectively block this approach would have a very serious effect on innovation. Peer-to-peer communications have enabled some important applications such as BitTorrent, which is used by NASA for sharing satellite images, by various computer companies for sharing large files (e.g., open source operating systems), by gaming companies for sharing updates, and even by content providers such as CBS and Warner Bros. for delivering programming.[80]

¶45 There is a second burden on innovation: the extra cost, both in development effort and development time, to include wiretap interfaces in early versions of software is prohibitive. At first blush, CALEA compliance seems simple since the only information that is needed is dialed-out and dialing-in phone numbers and voice. At that level, it is simple; nevertheless, the document defining the standard interface to a CALEA-compatible switch is more than 200 pages long.[81] Imagine, then, the standards necessary to cover interception of email, web pages, social networking status updates, instant messaging (for which there are several incompatible protocols), images, video downloads, video calls, video conference calls, file transfer layered on top of any of

---

[76] *Id.* § 1001(8)(A).

[77] *Id.* § 1001(8)(B)(ii).

[78] *Id.* § 1005(b).

[79] A service without any operators does not imply that no one profits. The original KaZaA filesharing service was ad-supported. *See* Ryan Naraine, *Spyware Trail Leads to Kazaa, Big Advertisers,* EWEEK (Mar. 21, 2006), http://www.eweek.com/c/a/Security/Spyware-Trail-Leads-to-Kazaa-Big-Advertisers/; *see also Universal Music Australia Pty Ltd. v. Sharman License Holdings Ltd.* (2005) 65 IPR 289 (Austl.); BRIAN BASKIN ET AL., COMBATING SPYWARE IN THE ENTERPRISE 9–11 (Tony Piltzecker et al. eds. 2006). It is unreasonable and probably infeasible to impose wiretap requirements on advertisers because the chain of indirection from the software developer to the advertisers is too long and tenuous. *See, e.g.*, Kate Kaye, *The Purchase-to-Ad Data Trail: From Your Wallet to the World*, AD AGE (Mar. 18, 2013), http://adage.com/article/dataworks/purchase-targeted-ads-data-s/240300/.

[80] *See, e.g.*, Brad King, *Warner Bros. to Distribute Films Using Bit Torrent*, MIT TECH. REV. (May 9, 2006), http://www.technologyreview.com/view/405794/warner-bros-to-distribute-films-using-bit-torrent/.

[81] *See* TELECOMMS. INDUS/ ASS'N, TR-45 LAWFULLY AUTHORIZED ELECTRONIC SURVEILLANCE J-STD-025 REV. A (May 31, 2000), *available at* http://cryptome.org/espy/TR45-jstd025a.pdf.

these, games that have voice or instant messaging functions included, and more. It is simply not a feasible approach. Nor are these improbable uses of the Internet; all of them are used very regularly by millions of people.

¶46 Applying CALEA to Internet applications and infrastructure will be a "tax" on software developers. The much lower barriers to entry (relative to traditional telephone networks currently covered by CALEA) provided by the open architecture of the Internet have bred many startups. These are small and agile; they are often the proverbial "two guys in a garage." Many will fail; even the eventual successes often start slowly. Regardless, they are essential to the Internet's success. Skype started small, yet it is now one of the largest international phone carriers.[82] Another example is Facebook, which was started by an undergraduate in his dorm room. Indeed, the Web began as an information distribution system at a European physics lab.[83] It is hard to say at what point an experiment has become large enough to be a "service" worthy of being wiretap-friendly; it is clear, though, that requiring such functionality to be built in from the start is a non-trivial economic burden and a brake on innovation. By contrast, the PSTN is primarily composed of large, established companies who buy essentially all of their equipment from other large, established companies.[84]

¶47 The most serious problem with CALEA, however, is that it has created a new class of vulnerabilities. By definition, a wiretap interface is a security hole because it allows an outside party to listen to what is normally a private conversation. It is supposed to be controlled, in that only authorized parties should have access. Restricting access to such facilities is far more difficult than it would appear; the history of such mechanisms is not encouraging.

¶48 The risks are not theoretical. In the 2004 to 2005 "Athens Affair," new code was injected into the phone switch that used the lawful intercept mechanisms to eavesdrop on about 100 mobile phones, including the Prime Minister's.[85] In a similar but less publicized incident in Italy between 1996 and 2006, about 6,000 people were the target of improper wiretaps, apparently due to corrupt insiders who sought financial gain. Again, the lawful intercept mechanism was abused.[86]

---

[82] *See supra* note 66.

[83] *See From a 1997 Hand-Out for the General Public,* TEN YEARS PUB. DOMAIN FOR THE ORIGINAL WEB SOFTWARE, http://tenyears-www.web.cern.ch/tenyears-www/Story/WelcomeStory.html (last visited Nov. 12, 2013).

[84] Even for such companies, the expense of adding CALEA facilities was non-trivial. The statute, 47 U.S.C. §§ 1007–1009 (2006), authorized $500 million "to pay telecommunications carriers for all reasonable costs directly associated with the modifications performed by carriers in connection with equipment, facilities, and services installed or deployed on or before January 1, 1995, to establish the capabilities necessary to comply with section 1002 of this title." The funding was approved in the Omnibus Consolidated Appropriations Act, which provided for funding through a combination of money supplied by various intelligence agencies and $60 million in direct funding. Omnibus Consolidated Appropriations Act, Pub. L. No. 104-208, 110 Stat. 3009 (1996). An additional $12 million was provided through unspent Department of Justice funds. More than 95% of the money was actually spent; about $40 million was rescinded by Congress in 2007. *See* U.S. DEP'T OF JUSTICE OFFICE OF INSPECTOR GEN., IMPLEMENTATION OF THE COMMUNICATIONS ASSISTANCE FOR LAW ENFORCEMENT ACT BY THE FEDERAL BUREAU OF INVESTIGATION ii–iii (Mar. 2008), *available at* http://www.justice.gov/oig/reports/FBI/a0820/final.pdf.

[85] *See* Prevelakis & Spinellis, *supra* note 4.

[86] *See* Piero Colaprico, Giuseppe d'Avanzo & Emilio Randacio, *'Da Telecom Dossier sui Ds' Mancini Parla dei Politici*, LA REPUBBLICA (Jan. 26, 2007),

¶49 The U.S. is at risk, too. Phone switches are already large, extremely complex computer systems;[87] as such, they are *inherently* at risk. An NSA evaluation of CALEA-compliant phone switches found vulnerabilities in every single one examined.[88] It is not known publicly if any American phone switches have been penetrated; however, news reports do suggest foreign interest in American use of surveillance technology to determine who America's surveillance targets are.[89]

¶50 There is one more aspect of security that has to be taken into account: who the enemies are. As has been widely reported in the press, various countries have created or are creating cyberespionage and cyberwarfare units.[90] These are highly skilled and well-equipped groups, easily capable of finding and exploiting subtle flaws in systems. To use an easy analogy, comparing the capabilities of such units to those of garden-variety hackers is like comparing the fighting power of modern infantrymen to that of a comparable-sized group of drug gang members. When considering the security of any Internet-connected systems that might attract the hostile gaze of foreign powers, this must be taken into account.

¶51 Communications systems fall into this category and have done so for many, many years. Even apart from their purely military significance, American economic interests have long been targeted by other nations. For example, in the early 1970s the Soviets reportedly used high-tech electronic eavesdropping devices to listen to the phone calls of American grain negotiators.[91] These days the attempts at economic espionage come not

---

http://www.repubblica.it/2006/12/sezioni/cronaca/sismi-mancini-8/dossier-ds/dossier-ds.html.

[87] W. Keister, R. W. Ketchledge & H. E. Vaughan, *No. 1 ESS: System Organization and Objectives*, 43 BELL SYS. TECHNICAL J. 1831, 1832 (1964) (calling the development of the 1ESS switch "the largest development project ever undertaken by Bell Laboratories for the Bell System."); Ben Chelf, *Code Complexity for Embedded Software Makers Sure Has Changed*, EMBEDDED (Jan. 22, 2009), http://www.embedded.com/electronics-blogs/industry-comment/4026959/Code-complexity-for-embedded-software-makers-sure-has-changed (speaking of "extreme software development projects (e.g., AT&T's phone switch)"); BRUCE STERLING, THE HACKER CRACKDOWN: LAW AND DISORDER ON THE ELETRONIC FRONTIER 37 (1992) (noting that the System 7 "signal transfer point"—a minor piece of phone switching equipment—is comprised of 10 million lines of source code). The best references that discuss the complexity phone switch software are proprietary documents (for example, 64 AT&T TECHNICAL J., no. 6, part 2, a special issue devoted to the 5ESS phone switch). One of the authors of this paper worked in the software engineering research department of the AT&T 5ESS phone switch development organization and saw the complexity first-hand.

[88] *See* Susan Landau, *The Large Immortal Machine and the Ticking Time Bomb*, 11 J. TELECOMM. & HIGH TECH. L. 1 (2013).

[89] *See* Kenneth Corbin, *'Aurora' Cyber Attackers were Really Running Counter-Intelligence*, CIO (Apr. 22, 2013), http://www.cio.com/article/732122/_Aurora_Cyber_Attackers_Were_Really_Running_Counter_Intelligence?taxonomyId=3089.

[90] For a discussion of exploits sponsored by the Chinese government, see MANDIANT, APT1: EXPOSING ONE OF CHINA'S CYBER ESPIONAGE UNITS, *available at* http://intelreport.mandiant.com/Mandiant_APT1_Report.pdf (last viewed Mar. 31, 2013) and David Sanger, David Barboza & Nicole Perlroth, *Chinese Army Unit is Seen as Tied to Hacking Against U.S.*, N.Y. TIMES, Feb. 18, 2013, http://www.nytimes.com/2013/02/19/technology/chinas-army-is-seen-as-tied-to-hacking-against-us.html. For a discussion of exploits being conducted by the Israeli government, see, for example, William Broad, John Markoff & David Sanger, *Israeli Test on Worm is Considered Crucial in Iran Nuclear Delay*, N.Y. TIMES, Jan. 15, 2011, http://www.nytimes.com/2011/01/16/world/middleeast/16stuxnet.html. These are just two examples of many such efforts.

[91] DAVID KAHN, KAHN ON CODES: SECRETS OF THE NEW CRYPTOLOGY 193 (1983).

just from Russia, but also from China, France, Germany, Israel, Japan, South Korea, India, Indonesia, and Iran.[92]

¶52 In 2000, the Internet Engineering Task Force, the engineering group that develops Internet communications standards through its "Requests for Comment" (RFCs) documents, concluded that "adding a requirement for wiretapping will make affected protocol designs considerably more complex. Experience has shown that complexity almost inevitably jeopardizes the security of communications . . . ; there are also obvious risks raised by having to protect the access to the wiretap. This is in conflict with the goal of freedom from security loopholes."[93] The security vulnerabilities that a wiretap introduces into a communications system are a serious problem, yet the problem apparently gets little attention from law enforcement in its efforts to expand CALEA to IP-based communications.

[92] Information on France, Germany, Israel, Japan, and South Korea can be found in INTERAGENCY OPSEC SUPPORT STAFF,

INTELLIGENCE THREAT HANDBOOK 5-5, 5-6 (1996), while information on China, India, Indonesia, and Iran can be found in OFFICE OF THE NAT'L COUNTERINTELLIGENCE EXEC., ANNUAL REPORT TO THE CONGRESS ON FOREIGN ECONOMIC COLLECTION AND INDUSTRIAL ESPIONAGE, FY07 2, 9–13 (Sept. 10, 2008), *available at* http://www.ncix.gov/publications/reports/fecie_all/fecie_2007/FECIE_2007.pdf. The US has a policy of not conducting economic espionage; in response to the recent NSA leaks, this was recently stated quite explicitly: "It is not an authorized foreign intelligence or counterintelligence purpose to collect such information to afford a competitive advantage to U.S. companies and U.S. business sectors commercially." A footnote goes on to say, "Certain economic purposes, such as identifying trade or sanctions violations or government influence or direction, shall not constitute competitive advantage." Directive on Signal Intelligence Activity, 2014 DAILY COMP. PRES. DOC. 31 (Jan. 17, 2014).

[93] NETWORK WORKING GRP., IETF POLICY ON WIRETAPPING 2 (May 2000), *available at* http://tools.ietf.org/html/rfc2804. One of the authors of this paper was on the Internet Architecture Board at the time and helped write the document.

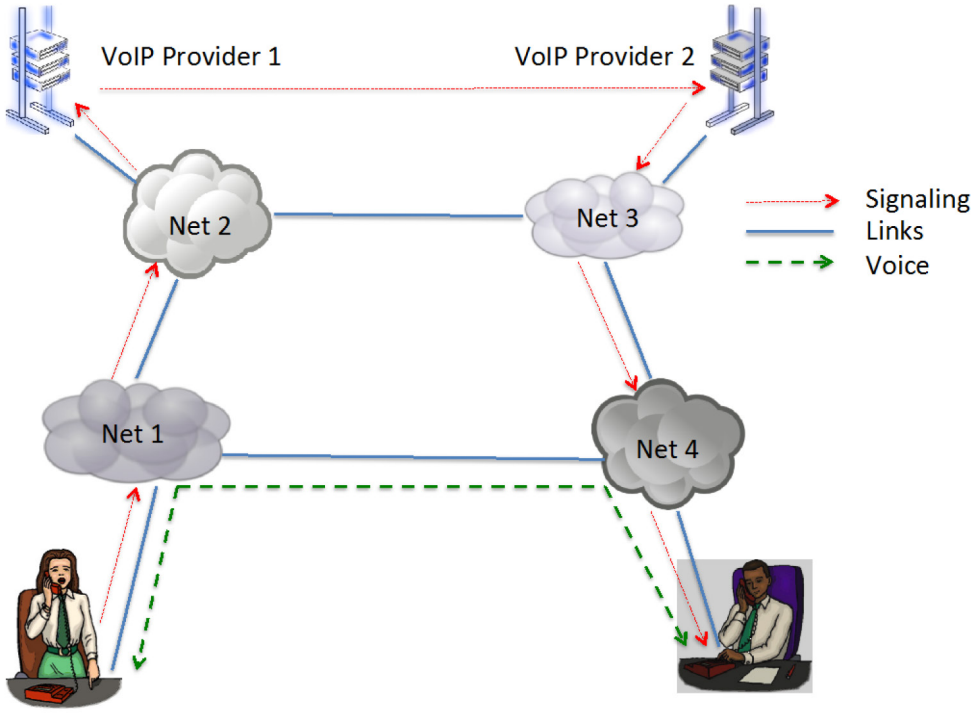

Figure 1: A Voice over IP (VoIP), showing physical links, the signaling path, and the voice path.

III. THE VULNERABILITY OPTION

¶53 We have argued extending CALEA to IP-based communications presents intolerable security risks and explained how modern communications systems are likely to impede wiretapping efforts. Given that, how might law enforcement wiretap modern communications? In this section, we describe the vulnerability option: how they can resolve the wiretap problem, why vulnerabilities exist, and why the vulnerability solution must, in fact, always be part of the law enforcement wiretap toolkit. We begin with a definition of terms.

A. Definition of Terms

¶54 We need to define a few commonly used technical terms in order to present the mechanics of employing a vulnerability for accessing a target system.[94]

> **Vulnerability**: A vulnerability is a weakness in a system that can potentially be manipulated by an unauthorized entity to allow exposure of some aspect of the

[94] Many of these terms are defined in R. SHIREY, INTERNET SECURITY GLOSSARY, VERSION 2 (Aug. 2007), *available at* http://tools.ietf.org/pdf/rfc4949.pdf. Others are common terminology in the hacker and security communities, but have yet to be defined in any authoritative work.

system. Vulnerabilities can be bugs (defects) in the code, such as a "buffer overflow"[95] or a "use-after-free instance,"[96] or misconfigurations, such as not changing a default password or running open, unused services.[97] Another common type of vulnerability results from not correctly limiting input text (this is also known as not sanitizing input), e.g., "SQL injection."[98] Alternatively, a vulnerability can be as simple as using a birth date of a loved one as a password. A vulnerability can be **exploited** by an attacker. A special instance of vulnerability is the:

Zero-day (or 0-day vulnerability): A zero-day is a vulnerability discovered and exploited prior to public awareness or disclosure to the vendor. Zero-days are frequently sold in the vulnerabilities market. The vendor and the public often only become aware of a zero-day after a system compromise.

Exploit: An exploit is the means used to gain unauthorized access to a system. This can be a software program, or a set of commands or actions. Exploits are usually classified by the vulnerability of which they take advantage and whether they require local (hands-on) access to the target system or can be executed remotely or through a web page or email message (drive-by).[99] The type of result obtained from running the exploit depends on the **payload** (rootkit, key-logger, etc.). The payload is chosen when the exploit is run or **launched**. An exploit demonstrates the use of the vulnerability in actual practice.

---

[95] A buffer overflow is caused by a program accepting more input than memory has been allocated for. Conceptually, imagine a clerk writing down someone's name, but the name as given is so long that it doesn't fit in the box on a form and spills over into the "Official Use Only" section of the form. A buffer overflow error was a central part of the Internet Worm of 1988, which resulted in the first case ever brought under the Computer Fraud and Abuse Act, 18 U.S.C. § 1030 (2006). *See* United States v. Morris, 928 F.2d 504 (2nd Cir. 1991). In some programming languages, e.g., Java, such overflows are detected automatically by the system; programmers using older languages, such as C, can use safe programming techniques that avoid the problem. A variety of tools can be used to detect potentially unsafe areas of programs. These have become increasingly common in the last 10 years, to very good effect.

[96] Programs can request storage space, then release ("free") it when they are done; after that, the space is available for other uses. A use-after-free bug involves carefully crafted accesses to memory no longer allocated for its original purpose; if some other section of the program is now reusing that storage, this section of the program may be confused by the improper reuse.

[97] A service is a mechanism by which programs listen for and act on requests from other programs; often, these services are available to any other computer that can contact this one via the Internet. The best analogy is to room numbers in a building. The building itself has a single address (the computer analog is the IP address), but the mailroom is in room 25, the information counter is in room 80, and so on. When one computer tries to contact another, it must specify the second computer's address (i.e., the building) and the service (i.e., the room number). Secure computer systems generally "listen" on very few ports, since each one represents a potential external vulnerability. (To continue our analogy, a building that does not need a mailroom will not have one that might somehow be abused.) Suppose, for example, that a computer that is not intended to act as a web server is in fact running web server code. A flaw in that web server can result in system penetration; the simplest fix is to turn off the web service since it is unneeded on that computer. See *CERT Advisory CA-2001-19 "Code Red" Worm Exploiting Buffer Overlow in IIS Indexing Service DLL*, CERT (July 19, 2001), *available at* http://www.cert.org/advisories/CA-2001-19.html, for an example of problems caused by open, unneeded services.

[98] In some contexts, parts of the input to a program can be interpreted as programming commands rather than as data. SQL injection attacks—in variant forms, they date back to at least the 1970s—occur when programmers do not filter input properly to delete such commands.

[99] A drive-by download is an attack perpetrated simply visiting a malicious or infected website. No further action by the user is necessary for the attack to succeed. Such attacks *always* result from underlying flaws in the web browser.

Payload: The payload of an exploit is the code that is executed on the target system giving the attacker the desired access. Payloads can be single action, such as surreptitiously creating a new user account on the system that allows future access, or multi action, such as opening a remote connection to the attacker's server and executing a stream of commands. The payload generally must be customized to the specific system architecture of the target.

Dropper: A dropper is a malware component or malicious program that installs the payload on the target system. A dropper can be single stage, a program that executes on the target system as a direct result of a successful exploit and carries a hidden instance of the payload, or it can be multi-stage, executing on the target system, but downloading files (including the payload) from a remote server.

Man-in-the-Middle attack: A Man-in-the-Middle attack is a method of gaining access to target information in which an active attacker interrupts the connection between the target and another resource and surreptitiously inserts itself as an intermediary. This is typically done between a target and a trusted resource, such as a bank or email server. To the target the attacker pretends to be the bank, while to the bank the attacker pretends to be the target. Any authentication credentials required (e.g., passwords or certificates) are **spoofed** by the attacker, so that each side believes they are communicating with the other. But because all communications are being transmitted through the attacker, the attacker is able to read and modify any messages it wishes to.

Spoofing: In the context of network security, a spoofing attack is a situation in which one person or program successfully masquerades as another by falsifying data and thereby gaining an illegitimate advantage.[100]

B. How Vulnerabilities Help

¶55 Our claim is that pre-existing vulnerabilities in software make extending CALEA unnecessary.[101] To understand the scenarios in which these vulnerabilities might be used, it is necessary to give a simplified description of the structure of modern computer operating systems.[102] Systems are described in terms of "layers"; each layer provides some services to the layer above it, and requests services of the layer below it. Often, a combination of hardware and software enforces the boundary between layers, ensuring that only certain requests can be made of the lower layer.

¶56 The lowest layer we will mention is the hardware: CPU chips such as Intel's Pentium series, devices such as network interfaces and hard drives, USB ports, etc. For our purposes, we will assume that this layer is error-free and secure. While not strictly true, attacks at this level are generally more feasible for the greater capabilities of national security purposes than for law enforcement.[103]

---

[100] SHIREY, *supra* note 94, at 187, 290 (defining "spoofing" as equivalent to "masquerade attack," which in turn is defined as "[a] type of threat action whereby an unauthorized entity gains access to a system or performs a malicious act by illegitimately posing as an authorized entity").

[101] Some of this material appeared in different form in Bellovin, Blaze, Clark & Landau, *supra* note 5.

[102] These days, smartphones are built the same way, so there is no need to discuss them separately.

[103] We will not discuss attacks like eavesdropping on encrypted WiFi signals. In principle, though, there might be exploitable vulnerabilities in the target's WiFi access point or router. These devices, though, are

¶57 The next layer is generally called the "kernel." The kernel protects itself against corruption (with aid from the hardware), and is also the only component that directly communicates with external hardware such as the network. When a program needs to read or write from the network or a disk drive, it cannot do so directly; instead, it asks the kernel to perform the action for it. A consequence of this is that the kernel has to enforce "file permissions": which users of the computer own which file, who can read or write them, etc. That in turn implies that there must be some strong separation between programs run by different users; again, the kernel enforces this.

¶58 The last layer of interest is the "user level" or "application level." Virtually all programs of interest—web browsers, mailers, document editors and viewers, and so on—run at user level. Running programs are typically associated with some user. The user may be a physical individual; however, all modern systems have a large number of helper processes, sometimes known as "daemons," running as some flavor of system pseudo-user. These handle such applications as the audio system, indexing files, insertion of USB devices, and more. A quick check of a modern Apple Mac showed no fewer than 10 different pseudo-users active on the machine.

¶59 All modern operating systems have a feature known as a "sandbox." A sandbox is a way of enforcing security by allowing a program to run with fewer privileges than the user who invoked it.[104] Sandboxes are frequently used for programs perceived as exceptionally vulnerable to security holes, such as PDF viewers and web browsers.

¶60 Vulnerabilities—and hence exploits of use to law enforcement—can occur at any layer, but the capabilities available to the exploit are different at different layers. While we defer details until Section IV, we note that for an exploit to work, more code is needed than just something that targets the vulnerability. In particular, to perform a wiretap—that is, to acquire the contents of a communication—the actual data sent or received has to be captured. This can be done in a particular application (e.g., Skype or a game with a voice communications feature), or it could be done at kernel level by tampering with a "device driver,"[105] in which case data from any application could be captured. A kernel exploit is well-positioned to modify device drivers; however, for complex technical reasons such an attack would find it more difficult to read and write files, export captured data via the network, etc.[106]

¶61 Most initial penetrations take place at application level.[107] The mechanisms vary widely, including infected attachments in email, malware on web pages, poor implementations of network protocols, and users downloading and voluntarily executing

---

just computers and can be hacked like any other computers.

[104] *See* SHIREY, *supra* note 94.

[105] A *device driver* is a special part of the kernel that communicates with input/output devices such as disks, audio ports, network interfaces, etc. *See, e.g.*, ANDREW S. TANENBAUM AND ALBERT S. WOODHULL, OPERATING SYSTEMS DESIGN AND IMPLEMENTATION 231–33 (3d ed. 2006).

[106] Even a brief explanation of this is well beyond the scope of this paper. The primary problems are the nature of I/O APIs—they are generally designed to copy essential parameters from application level—and the difficulty of waiting for an I/O operation to complete without a "process context." *See, e.g.*, TANENBAUM & WOODHULL, *supra* note 105.

[107] It is generally believed that since kernels do almost no processing of network packet contents (as opposed to their "headers"), they are therefore much less vulnerable to attacks. This is more generally true, too. Having a virus-infected attachment in an email message is harmless; by contrast, clicking on it causes the attachment to be processed and thus causes damage.

booby-trapped programs under a misapprehension as to the programs' purpose, provenance, and good intent.[108] The results are the same: some program the user had not intended is being run with the user's file access rights.

¶62 Under certain circumstances, this insecurity is sufficient for law enforcement purposes. For example, it generally provides adequate means for intercepting email. It may also suffice for looking at the transcript files kept by some instant messaging programs.

¶63 On the other hand, if the program penetrated is not used for the actual communications of interest, these application-level exploits alone will not suffice. Consider that on most modern platforms, users—and hence the programs they run—do not have the ability to tamper with the kernel or system-owned files; note that most applications, including Skype, are system-owned. Accordingly, if a law enforcement penetration for the purpose of eavesdropping is executed at user level, a second exploit known as a "local privilege escalation"[109] attack is needed. This second attack gives the program elevated privileges and hence the ability to change device drivers, modify files, etc.[110] While the two exploits are generally independent, frequently both are necessary; this complicates the attack.

¶64 There is one special case worth mentioning. Some daemons run with full system privileges; if these have faulty implementations of network protocols, only a single attack is needed. This is a venerable technique, going back to the first Internet worm.[111] While modern system designs try to avoid daemons with full privileges,[112] in some situations this is unavoidable.

¶65 Historically, some applications have been considerably more vulnerable to user level attacks than others; these applications include web browsers and PDF viewers. As noted, modern operating systems often run these programs in sandboxes to prevent theft of or damage to user files. Sandboxes may also deny the confined program the ability to run other system commands that may be utilized for privilege escalation. Accordingly, a third exploit may be necessary to escape from the sandbox; subsequently, privilege escalation is used as before.

---

[108] A significant percentage of software downloaded via peer-to-peer networks contains malware. *See, e.g.*, Michal Kryczka et al., TorrentGuard: Stopping Scam and Malware Distribution in the BitTorrent Ecosystem 1 (2012), http://arxiv.org/pdf/1105.3671v3.pdf; Andrew D. Berns & Eunjin (EJ) Jung, Searching for Malware in BitTorrent 4 (2008), *available at* http://www.cs.uwlax.edu/~aberns/UICS-08-05.pdf. Note that much of this is "key generation or activation utility[ies]"; i.e., tools used to steal software. *Id.*

[109] For more detail on privilege escalation, including an example, see GREG HOGLUND & GARY MCGRAW, EXPLOITING SOFTWARE: HOW TO BREAK CODE 151–53 (2004). For an additional example of a local privilege escalation attack as a proof-of-concept, see Posting of Stefan Kanthak, *Defense in Depth – the Microsoft Way (Part 11): Privilege Escalation for Dummies*, SECURITY FOCUS, http://www.securityfocus.com/archive/1/528955/30/90/threaded. "Local" indicates that the attacker must already have the ability to run code on the targeted system; it cannot be done by a "remote" attacker, i.e., one who can only make network connections to the machine.

[110] On Windows, the privileged user is known as "Administrator." On Unix-like systems, including MacOS and Linux, it is known as "root."

[111] *See, e.g.*, EUGENE SPAFFORD, THE INTERNET WORM PROGRAM: AN ANALYSIS 4–6 (Dec. 1988), *available at* http://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1701&context=cstech; Jon A. Rochlis & Mark W. Eichin, *With Microscope and Tweezers: The Worm from MIT's Perspective*, 32 COMM. ACM 689 (June 1989).

[112] The design principle is known as "least privilege." *See* SHIREY, *supra* note 94.

¶66 To summarize, there are many different points for initial attack, and all have their limitations. System privileges are needed to modify applications or device drivers and can be obtained via either a direct kernel attack, an attack on a system-level daemon, or via privilege escalation following an application level penetration.

C. Why Vulnerabilities Will Always Exist

¶67 We are suggesting use of pre-existing vulnerabilities for lawful access to communications. To understand why this is plausible, it is important to know a fundamental tenet of software engineering: bugs happen. In his classic *The Mythical Man-Month*, Frederick Brooks explained why:

> First, one must perform perfectly. The computer resembles the magic of legend in this respect, too. If one character, one pause, of the incantation is not strictly in proper form, the magic doesn't work. Human beings are not accustomed to being perfect, and few areas of human activity demand it. Adjusting to the requirement for perfection is, I think, the most difficult part of learning to program. [113]

¶68 Because computers, of course, are dumb—they do exactly what they are told to do— programming has to be absolutely precise and correct. If a computer is told to do something stupid, it does it, while a human being would notice there is a problem. A person told to walk 50 meters then turn left would realize that there was an obstacle present, and prefer the path 52 meters down rather than walking into a tree trunk. A computer would not, unless it had been specifically programmed to check for an impediment in its path. If it has not been programmed that way—if there is virtually any imperfection in code—a bug will result. The circumstances which might cause that bug to become apparent may be rare, but it would nonetheless be a bug.[114] If this bug should happen to be in a security-critical section of code, the result may be a vulnerability.

¶69 A National Research Council study described the situation this way:

> [A]n overwhelming majority of security vulnerabilities are caused by "buggy" code. At least a third of the Computer Emergency Response Team (CERT) advisories since 1997, for example, concern inadequately checked input leading to character string overflows (a problem peculiar to C programming language handling of character strings). Moreover, less than 15 percent of all CERT advisories described problems that could have been fixed or avoided by proper use of cryptography. [115]

¶70 It would seem that bugs should be easy to eliminate: test the program and fix any problems that show up. Alas, bugs can be fiendishly hard to find, and complex programs simply have too many possible branches or execution paths to be able to test them all.[116]

---

[113] FREDERICK P. BROOKS JR., THE MYTHICAL MAN-MONTH 8 (Anniversary ed. 1995).

[114] In one classic incident, a single missing hyphen in a program contributed to the loss of the Mariner 1 space probe. *See Mariner 1*, NASA, http://nssdc.gsfc.nasa.gov/nmc/spacecraftDisplay.do?id=MARIN1 (last visited Sept. 26, 2013).

[115] TRUST IN CYBERSPACE 110 (Fred B. Schneider ed., 1999).

[116] The single capability that gives a computer most of its power is the ability to do things conditionally.

¶71 Brooks includes a diagram on bugs comparing the predicted and actual rate of bugs in complex code.[117] The projection assumed a slow start, a rapid increase in the debugging rate, and a leveling off that suggested the last bugs had been found. Instead, the rate never leveled off, and the total number of bugs found was significantly higher than had been forecast.[118] Brooks himself suggests that testing takes about half of total development time.[119] However, even this is not enough: "Testing shows the presence, not the absence of bugs."[120]

¶72 We will not recount the myriad techniques other than testing that have been tried in an effort to eliminate bugs; let it suffice to say there have been many. These include formal mathematical methods, better programming and debugging tools, different organizational and procedural schemes, improved programming languages, and more. Many of these ideas have helped, but none have proved a panacea. The ability to produce error-free code is the Holy Grail of systems development: heavily desired but unattainable.[121]

¶73 When we are dealing with computer security, though, the question is somewhat different than whether the program has bugs. Rather, the proper question is whether the security-sensitive parts of the system have bugs. When formulated this way, there would seem to be an obvious solution: divide a complex system up into security-sensitive and security-insensitive pieces; bugs in the latter, though annoying, would not result in disaster. Such an approach would also improve the correctness of the security-critical components. The bug rate in code increases more than linearly with the size of the program; therefore, a program that is twice as large has more than twice as many bugs. Perhaps the security-sensitive section, which is by definition smaller, would thereby have far fewer bugs than the system as a whole.

That is, it can test a condition—is this number greater than zero? does this string of characters contain an apostrophe? is there room on the page for another line?—and continue along one program path or another, depending on the result of the test. In principle, each conditional operation can double the number of possible execution paths. (The reality is not quite that bad, because not all tests are independent.) This means that a program with just 20 conditionals may have more than 2^{20}—over 1,000,000—possible paths through it; one with 40 conditionals (a very tiny number for a realistic program) may have more than 1,000,000,000,000. Exhaustive testing is not possible under these circumstances.

[117] *See* BROOKS, *supra* note 113, at 92. The diagram is a previously unpublished one by John Harr.

[118] Neither the graph nor the text make it clear whether the graph ended because the project was finished or simply because it was a snapshot of a single year's experience and did not look at the entire project. The graph, presented at the 1969 Spring Joint Computer Conference, shows one year of experience building the #1 ESS; the programming undoubtedly took longer. *See* PHIL LAPSLEY, EXPLODING THE PHONE 233–38 (2013). The switch itself is described in Keister, Ketchledge & Vaughan, *supra* note 87. New versions of the code were unlikely to have fewer bugs; rather, the bug rate *increases* after some point. BROOKS, *supra* note 113, at 53–54.

[119] *See* BROOKS, *supra* note 113, at 10, 17 (explaining the complexity of the model).

[120] SOFTWARE ENGINEERING TECHNIQUES: REPORT ON A CONFERENCE SPONSORED BY THE NATO SCIENCE COMMITTEE, ROME, ITALY, 27TH TO 31ST OCTOBER 1969 16 (1970) (quoting E. W. Dijkstra).

[121] Operational errors are common, too. *See, e.g.*, Barton Gellman, *NSA Broke Privacy Rules Thousands of Times Per Year, Audit Finds*, WASH. POST, Aug. 16, 2013, http://www.washingtonpost.com/world/national-security/nsa-broke-privacy-rules-thousands-of-times-per-year-audit-finds/2013/08/15/3310e554-05ca-11e3-a07f-49ddc7417125_story_1.html ("One in 10 incidents is attributed to a typographical error in which an analyst enters an incorrect query and retrieves data about U.S phone calls or e-mails."). Another bug confused the country and city codes for Cairo, Egypt (20 2) with the area code for Washington, D.C. (202). *Id.* These sorts of errors led to literally thousands of incidents of improper collection of surveillance data.

¶74 This approach has been at the heart of most secure system designs for more than fifty years. It was set out mostly clearly in the so-called "Orange Book," the 1985 Department of Defense criteria for secure operating system design.[122] The Orange Book prescribed something called a "Trusted Computing Base," the security-essential portions of a system:

> The heart of a trusted computer system is the Trusted Computing Base (TCB) which contains all of the elements of the system responsible for supporting the security policy and supporting the isolation of objects (code and data) on which the protection is based. The bounds of the TCB equate to the "security perimeter" referenced in some computer security literature. In the interest of understandable and maintainable protection, a TCB should be as simple as possible consistent with the functions it has to perform.[123]

¶75 This dream has proved elusive for two very different reasons. First, modern TCBs are themselves extremely large, significantly bigger than the entirety of the 1970s and 1980s vintage systems. Although modern software is far more reliable, that does not translate into absolute reliability. It is worth noting that one of today's complex applications is tens of times larger than entire systems from the 1980s, when the Orange Book was written; this complexity, as we have noted, leaves them very vulnerable to attack. Today's operating systems are also vastly larger. Second, the notion of the TCB is less clear than it once was. More and more serious security incidents target components that fit no one's definition of "trusted," but the attacks are effective nevertheless. For example, in 1988 the very first Internet worm exploited holes outside what would likely have been considered part of the TCB.[124] In essence, although not by intent, it was a denial of service attack: it consumed most of the capacity of the infected machines. This happened at the user level; the affected programs were not part of the TCB.[125] Put another way, trying to break up the system into trusted and untrusted parts does not work as well as had been hoped; bugs anywhere can be and have been exploited by malware.

¶76 We conclude that for the foreseeable future, computer systems will continue to have exploitable, useful holes. The distinction between flaws in the TCB and flaws

---

[122] DEP'T OF DEF., DEPARTMENT OF DEFENSE TRUSTED COMPUTER SYSTEM EVALUATION CRITERIA (1985) *available at* http://csrc.nist.gov/publications/secpubs/rainbow/std001.txt. The nickname comes from the color of its cover; it is part of a series of publications known collectively as "The Rainbow Series."

[123] *Id.* at 65.

[124] The worm tried by various means to find and attack other computers. If it ever succeeded, it sent a copy of itself over to those computers and started executing there as well; meanwhile, the first copy continued to scan for other targets. There was no check to make sure that a given computer was infected only once; this meant that vulnerable systems were running very many copies of the worm, sufficiently many that legitimate programs were crowded out. Furthermore, the Internet itself was clogged by the attack traffic. Finally, since one of the vulnerable services was email, many sites turned off their mail systems in an attempt to protect themselves; this, however, hindered coordination of attempts to combat the worm since many people knew no other way to reach their colleagues at other sites. *See* SPAFFORD, *supra* note 111, and Rochlis & Eichin, *supra* note 111, for more details on the worm's behavior and structure.

[125] This is not strictly true. For technical reasons, one of the programs that were successfully attacked did run with elevated privileges; however, neither the penetration nor the excess resource consumption by it were related to those privileges. It ran as privileged (and hence by definition as part of the TCB) because the importance of avoiding excess privilege was not as well understood in the general community at that time as it is today.

outside of it is important. Non-TCB programs—frequently known as "user mode" or "application mode" programs—have the privileges of the user who runs them, whereas TCB programs are generally all-powerful and have access to more files and the ability to change them.[126]

D. Why the Vulnerability Solution Must Exist Anyway

¶77 Considering lawful intercept purely as an economic question, it is tempting to ask which is a cheaper solution: a vulnerability-based approach or a CALEA-like law. The question, however, is not that simple. Even apart from our overriding theme—that applying CALEA to Internet software creates many very serious risks to both security and innovation—and apart from the cost-shifting issue (with CALEA-like solutions, the bulk of the cost is not carried by law enforcement), there is a further, more fundamental issue: a vulnerability-based intercept capability must exist regardless of any extension of CALEA The question, then, is not which costs less, but whether the incremental cost of CALEA is justifiable given that the vulnerability-based approach must be pursued in any case.

¶78 No matter what a CALEA-like law says, there will always be important situations where CALEA interfaces will not help law enforcement conduct surveillance. Often, these will be extremely important, urgent situations involving national security, counterterrorism, or major drug gangs.[127] Those criminals involved in national security and counterterrorism are more likely than common criminals to use non-American or even custom-written communications software and procedures.[128] Other situations in which a new law will not help include situations with people who use older software that has not been upgraded to include a lawful intercept feature, and more generally situations with any communications application that automatically provides end-to-end encryption capability.[129]

¶79 In situations like these, where the case is important and built-in lawful intercept mechanisms are not available, using vulnerabilities becomes an attractive alternative. The alternative to using vulnerabilities—a so-called "black bag job" or a covert search—is far

---

[126] This stark dichotomy between all-powerful and relatively powerless code is generally seen by the computer security and operating system communities as a bad idea. Many schemes have been proposed to create intermediate levels of privilege; few, if any, have caught on *and* been more than minimally effective at protecting the system. There has been more success of late with sandboxes.

[127] The Mexican Zeta drug gang uses a home-built, encrypted radio network. *See* Michael Weissenstein, *Mexico's Cartels Build Own National Radio System*, YAHOO! NEWS (Dec. 27, 2011), http://news.yahoo.com/mexicos-cartels-build-own-national-radio-system-200251816.html.

[128] The Russian sleeper agent ring arrested in 2010 used special programs for *steganography*, a way of concealing the very existence of messages. *See* Noah Shachtman, *FBI: Spies Hid Secret Messages on Public Websites*, WIRED (June 29, 2010, 1:11 PM), http://www.wired.com/dangerroom/2010/06/alleged-spies-hid-secret-messages-on-public-websites/.

[129] Even the current CALEA statute states: "A telecommunications carrier shall not be responsible for decrypting, or ensuring the government's ability to decrypt, any communication encrypted by a subscriber or customer, unless the encryption was provided by the carrier and the carrier possesses the information necessary to decrypt the communication." 47 U.S.C. § 1002(b)(3) (2006). The "information necessary to decrypt the communications" is typically a cryptographic key. If end-users do their own key management, the provider is unlikely to have the keys.

riskier.[130] Electronically placing a vulnerability on a machine does not put a law-enforcement agent at risk; conducting a black-bag job or a covert search certainly does.

¶80 As with so much other high technology, using vulnerabilities for eavesdropping has a relatively high start-up cost, whereas continued use does not. Apart from the obvious drop in the cost per interception, the operational software is likely to improve over time. That is, as the developers have more time and gain more experience, the overall package will improve. It will provide more functionality, higher efficiency, and stronger resistance to detection. The actual exploits used will, as noted, change over time; however, the exploits are likely to be usable in many more interceptions than in a CALEA-based world, which will also drive down the cost of each interception. In other words, and to a much greater degree than in a CALEA-based approach, using vulnerabilities will improve law enforcement's abilities in all cases, especially the most critical ones.

IV. VULNERABILITY MECHANICS

¶81 In this section, we examine the potential use of vulnerabilities. We begin by exploring warrant issues for using exploits to wiretap. We discuss how vulnerabilities may be exploited, and consider minimization in this environment and what tools and procedures are available that law enforcement authorities might use or modify to gain access. We also discuss the vulnerability and exploit markets. Finally, we discuss what steps would be needed for productizing an exploit specifically for lawful access by law enforcement.

A. Warrant Issues

¶82 Obviously, any use of vulnerabilities for wiretapping requires proper authorization. However, because of the technologies involved, the process for obtaining proper authorization may be somewhat more involved than for conventional wiretaps.

¶83 One issue is that there are two distinct steps: exploiting the vulnerability, i.e., hacking the target's machine with proper permission, and actually carrying out the desired interception. Arguably, two different court orders should be obtained. Documents released under the Freedom of Information Act show the FBI has used such a two-step process to obtain information in at least one situation. The FBI first sought a search warrant to install Computer and Internal Protocol Address Verifier (CIPAV) on the target's machine, which sends address and protocol information from the target's machine to the FBI.[131] Having obtained the IP address and other relevant information by

[130] Such searches are performed when necessary. *See, e.g.*, Schactman, *supra* note 128.

[131] *See* Jennifer Lynch, *New FBI Documents Provide Details on Government's Surveillance Spyware*, ELECTRONIC FRONTIER FOUND. (Apr. 29, 2011), https://www.eff.org/deeplinks/2011/04/new-fbi-documents-show-depth-government. CIPAV is a current FBI software package analogous to what we are proposing here. Its capabilities, as described in an affidavit for a search warrant, include collecting the target machine's IP address, MAC address, operating system type and version, browser type and version, "certain registry-type information," last URL visited, etc. *See* Affidavit for State of Washington, County of King, In the Matter of the Search of any Computer Accessing Electronic Message(s) Directed to the Administrator(s) of MySpace Account "Timberlinebombinfo" and Opening Message(s) Delivered to that Account by the Government (No. MJ07-5114), at 3, *available at* http://politechbot.com/docs/fbi.cipav.sanders.affidavit.071607.pdf.

conducting surveillance, the FBI then sought a pen register/trap-and-trace order from the court; however, this is not always done. In *In Re Warrant to Search a Target Computer at Premises Unknown*, the FBI submitted a single Rule 41 warrant application, covering all activities: finding the target, installing their own software, gathering addresses, taking pictures, etc.[132]

¶84 Another issue that can cause complications is the need for ``technical reconnaissance'' to identify the proper target machine.[133] This may involve listening to other conversations, which would presumably require its own authorization.

¶85 Finally, the design of this sort of tap presents some opportunities for minimization by technical means, prior to the usual minimization that is required by law.[134] Arguably, this should be specified in the warrant as well.[135]

B. Architecture

¶86 How should a law enforcement exploit software platform be designed? The special legal requirements, the technical quirks involved in exploitation, the speed with which technology changes, the lifetime of a vulnerability, the need for non-proliferation, and even budgetary constraints all suggest that any framework of tools developed for surveillance must be easily configurable and readily adaptable. This in turn suggests that a highly modular architecture is needed for a vulnerability-based communications intercept vehicle.[136]

¶87 The particular components to be used against any given target will vary widely. Consider the choice of initial exploit. For a target with an older (and unpatched) system, an older and publicly-known exploit might be sufficient, but for wiretapping someone using a newer operating system, or one that is fully patched, an old vulnerability will not suffice, forcing the use of a newer one. Further, another target, not using the common application targeted by either of the previous two, might require yet a third vulnerability.

---

[132] No. H-13-234M, 2013 WL 1729765 (S.D. Tex. Apr. 22, 2013). Mark Eckenwiler, formerly a top Justice Department authority on surveillance, has indicated that intrusions needed to execute pen register orders can be performed solely on the lesser pen register standard. *See* Jennifer Valentino-DeVries & Danny Yadron, *FBI Taps Hacker Tactics to Spy on Suspects*, WALL ST. J., Aug. 3, 2013, http://online.wsj.com/article/SB10001424127887323997004578641993388259674.html.

[133] *See infra* Section IV.D.

[134] Minimization is as defined in the wiretap statute, 18 U.S.C. § 2518(5) (2006) ("Every order and extension thereof shall contain a provision that the authorization to intercept . . . shall be conducted in such a way as to minimize the interception of communications not otherwise subject to interception under this chapter, and must terminate upon attainment of the authorized objective . . . .").

[135] *See infra* Section IV.C.

[136] Designing systems to use modules is standard software engineering practice. By definition, modules communicate via well-defined interfaces, allowing easy substitution of different versions. *See, e.g.*, D.L. Parnas, *On the Criteria to be used in Decomposing Systems into Modules*, 15 COMM. ACM 1053, 1053–54 (1972). A good example of a modular framework is a picture editor. Many different file formats—JPEG, TIFF, PNG, etc.—can be imported into a picture editor. The editing is done in the same way, regardless of the input format; then, the new version can be stored in any of these formats. In other words, the file format input/output routine is a separate module. The same is true for vulnerability-based surveillance. With a well-designed framework, execution of a wiretap could be as simple as choosing a wiretap module, an exploit, and warrant information, entering the target information, and pressing "Go." The system would then build the payload for automatic installation. New exploits or new warrant information would be separate modules; the rest of the program would not be affected.

Any of these exploited weaknesses could potentially be closed on the targets' systems at any time, which could require the use of yet another vulnerability.[137]

¶88 There are other considerations as well. If only voice communications are to be picked up, there is no need to include a module providing keystroke-logging capability in the payload. Indeed, the less code that is included, the less the risk of the tap being discovered. Perhaps more important, code that is not included cannot be repurposed by someone else, thus aiding in non-proliferation.[138] Beyond that, selective inclusion aids in warrant compliance, by limiting what is collected to what the court's order permits. This is discussed in more detail below.[139]

¶89 A modular framework can also be extremely cost-effective relative to other designs. By design modules are plug-and-play—no matter how different they may be on the inside, the way the modules communicate with the framework is standardized. The design makes it easy to have many different people develop exploits for the same framework, and straightforward for people to use new ones. When an exploit becomes obsolete, only the module containing that exploit needs to be rewritten or replaced. Pre-configured warrant modules provide assurance to law enforcement that exploits will collect the communications they need,[140] and assurance to the judge that the exploit and payload will behave as specified in the warrant. If the investigation changes and a new warrant module is needed, the exploit executable only needs to be recompiled with the new module and reinstalled.

C. Technical Aspects of Minimization

¶90 The wiretap statute specifies that: "Every order and extension thereof . . . shall be conducted in such a way as to minimize the interception of communications not otherwise subject to interception under this chapter . . . ."[141] While this is normally a matter for judges to rule on, a properly designed intercept package can carry out some of this task. This provides greater privacy for individuals not targeted by the warrant. More subtly, by automatically eliminating a lot of the extraneous content, it eases the task of humans charged with minimization and thus likely reduces their error rate.[142]

¶91 A warrant must clearly specify what communications may and may not be collected:

> Each order authorizing or approving the interception of any wire, oral, or electronic communication under this chapter shall specify—
>
> (a) the identity of the person, if known, whose communications are to be intercepted; . . .

---

[137] See discussion of the lifetime of these components, *infra* Section IV.E.

[138] *See infra* Section V.

[139] *See infra* Section IV.C.

[140] *See id.*

[141] 18 U.S.C. § 2518(5) (2006).

[142] While we do not suggest or think that a program can perform full minimization, it can certainly carry out mechanical aspects, e.g., excluding services and perhaps users not covered by the warrant.

> (c) a particular description of the type of communication sought to be intercepted[143]

¶92 Intercepts that collect more than is authorized are legally problematic, to say the least.[144]

¶93 A modular architecture greatly simplifies the execution of the warrant. Modules for common warrant specifications would contain pre-configured values, such as types of data to collect or ignore, specified ports to listen on, and time limits. The framework would compile these values into a properly tailored exploit executable automatically, without the need for any special configuration by the law enforcement technicians.[145]

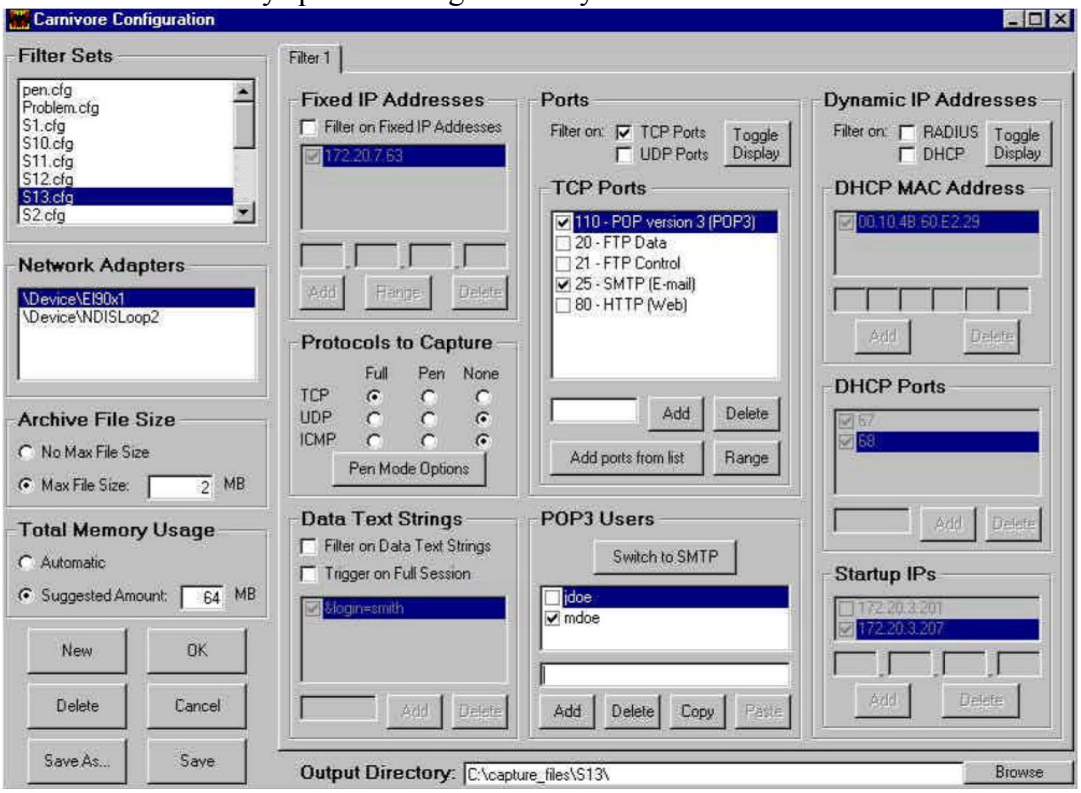

Figure 2: A sample warrant configuration screen from Carnivore. This filter is set up to intercept all inbound (POP) and outbound (SMTP) email from user mode.

---

[143] 18 U.S.C. § 2518(4).

[144] According to documents obtained by the Electronic Privacy Information Center under FOIA, when the FBI's UBL unit (Usama bin Laden unit) was conducting FISA surveillance, "The software was turned on and did not work correctly. The FBI software not only picked up the E-Mails under the electronic surveillance of the FBI's target, [redacted] but also picked up E-Mails on non-covered targets. The FBI technical person was apparently so upset that he destroyed all the E-Mail take, including the take on [redacted] is under the impression that no one from the FBI [redacted] was present to supervise the FBI technical person at the time." Memorandum from [redacted] to Spike (Marion) Bowman (Apr. 5, 2000), *available at* http://www.epic.org/privacy/carnivore/fisa.html.

[145] "Compilation" is the process of turning human-readable "source code," written in a language like C or C++, into the string of bytes that are actually understood by the underlying hardware. At compilation time, it is possible to select which sections of the program should be included in the eventual module. A classic treatment of how compilers work can be found in ALFRED V. AHO, MONICA S. LAM, RAVI SETHI & JEFFERY D. ULLMAN, COMPILERS: PRINCIPLES, TECHNIQUES, AND TOOLS (2nd ed. 2007).

¶94 The warrant configuration screen[146] from the (now obsolete) Carnivore wiretapping system[147] provides a useful example. It has options for full content and pen register capture, fields for identifying which protocols should be captured, which IP addresses or users should have their data monitored, and so on. A similar scheme should be used here, with a crucial difference: modules not selected would not be included in the payload installed on the target's machine.

¶95 Other information can also be used for minimization. Assume, for example, that police know from other means that their suspect uses only one of the user profiles (i.e., logins) on a shared computer.[148] The intercept module, if properly configured, would operate only when that user is logged in. Similar filters could be used for communications applications like Skype that have their own logins.

D. Technical Reconnaissance

¶96 The reconnaissance phase—learning enough about the target to install the necessary monitoring software—is essential to a successful compromise of a device. Because exploits must be exquisitely tailored to particular versions and patch levels, using the wrong exploit frequently results in failures, and can even raise alerts or cause suspicious crashes. There are a number of widely used, readily available tools. Many of the best tools are even available in a free, ready-to-use downloadable toolbox; for example, the Backtrack-Linux Penetration Testing Distribution.[149]

¶97 The most common first step is to check publicly available information. DNS[150] and Whois[151] lookups are used to find Internet domain and IP information. Simple use of

---

[146] This image is taken from Figure C-16 of STEPHEN P. SMITH, HENRY H. PERRITT, JR., HAROLD KRENT, STEPHEN MENCIK, J. ALLEN CRIDER, MENGFEN SHYONG & LARRY L. REYNOLDS, IIT RES. INST., INDEPENDENT REVIEW OF THE CARNIVORE SYSTEM: FINAL REPORT C-17 (2000) (aspect ratio adjusted), *available at* http://www.epic.org/privacy/carnivore/carniv_final.pdf.

[147] Carnivore was later renamed as the DCS 1000, and has since been retired in favor of commercial solutions. The apparent abandonment of the package is discussed in the 2002 and 2003 FBI reports to Congress. FED. BUREAU OF INVESTIGATION & U.S. DEP'T. OF JUST., CARNIVORE/DCS-1000 REPORT TO CONGRESS 3 (Feb. 24, 2003), *available at* https://epic.org/privacy/carnivore/2002_report.pdf; FED. BUREAU OF INVESTIGATION & U.S. DEP'T. OF JUST., CARNIVORE/DCS-1000 REPORT TO CONGRESS 4 (Dec. 18, 2003), *available at* https://epic.org/privacy/carnivore/2003_report.pdf.

[148] This is sometimes the case. *See, e.g.*, State of Ohio v. Nicholas J. Castagnola, Nos. CR 10 07 1951 (B) & CR 10 08 2244, slip op. at 11–14 (Mar. 29, 2013).

[149] The Backtrack Linux Penetration Testing Distribution is an open-source, ready-to-use linux operating system specifically customized and configured for security analysts and penetration testers. It can be installed onto a computer or booted live from a disk or thumbdrive. It contains a comprehensive set of tools for network and system scanning, vulnerability detection, exploitation, privilege escalation and forensics. There are also tutorials and How-To's available and a large user and contributor community. *See BackTrack Linux*, BACK|TRACK-LINUX.ORG, http://www.backtrack-linux.org (last visited Nov. 12, 2013).

[150] The DNS—the Domain Name System—is used to convert human-friendly names such as www.fbi.gov to the number IP address understood by low-level Internet hardware. Information in the DNS is especially useful when trying to break into organizations rather than individual users' computers. *See, e.g.*, WILLIAM CHESWICK, STEVEN M. BELLOVIN & AVIEL D. RUBIN, FIREWALLS AND INTERNET SECURITY 31–33 (2d ed. 2003).

[151] Whois is a public database lookup service provided by the Internet name registrars that provides information about the ownership of domain names, address blocks, etc. For more information, see Simone Carletti, *Understanding the WHOIS Protocol*, SIMONE CARLETTI'S BLOG (Mar. 27, 2012, 12:13 PM), http://www.simonecarletti.com/blog/2012/03/whois-protocol/, which gives examples of Whois output.

search engines and scouring social media sites often provide some information about the target's operating system, cell phone platform, service provider, and commonly used applications. With the appropriate legal process, e.g., a subpoena or court order under 18 U.S.C. § 2703(d) (2006), some of this information may also be available from the service provider.

¶98 If the investigators have access to some emails from the target, a great deal of information may be found by studying the headers. An examination of some of our test emails showed such lines as:

> Mime-Version: 1.0 (Mac OS X Mail 6.2 \(1499\))
>
> X-Mailer: Apple Mail (2.1499)

and

> X-Mailer: iPhone Mail (10B146).

which are rather clear indicators of which operating system is in use.

¶99 To remotely access a machine, an attacker generally needs to know the IP and/or MAC addresses of the machine,[152] the operating system (including exact version and patch level), what services are running on the machine, which communications ports are open,[153] what applications are installed, and whether the system contains any known vulnerabilities. This process of discovery is referred to as "Mapping" and "Enumeration."[154]

¶100 Mapping can be of the system or of the network (or both). Network mapping can be WiFi or Ethernet, and can refer to finding hidden networks, or to enumerating all the devices and their addresses connected to a particular network. Mapping the target device or system requires finding the so-called "MAC address," a hardware address transmitted when speaking over Ethernet, WiFi, or Bluetooth networks. If the target of a tap is using a smartphone at a public hotspot, detecting that person's MAC address could, for example, reveal what brand of phone is being used.

---

[152] IP and MAC addresses are networking concepts. MAC addresses are generally hard-wired in a computer's communications hardware, though sophisticated users can change them. IP addresses are often transient, but tend to remain the same for a given computer in a given location. While IP addresses are typically assigned by the network administrator of the site at which the computer is located, MAC addresses are assigned by the manufacturer and therefore indicate the computer type and model. *See, e.g.*, ANDREW TANENBAUM, COMPUTER NETWORKS (4th ed. 2003).

[153] On networked computer systems, services offered are assigned to particular (and generally standardized) "port numbers," a more or less arbitrary value between 1 and 65535. Port enumeration is the process of seeing what ports, and hence what services, are available on a given system. Using open ports for intrasystem communication, rather than more secure alternatives, was one of the items cited in the FTC complaint against HTC. *See* Complaint at 3–4, *In re* HTC America, Inc., No. C-4406 (F.T.C. June 25, 2013).

[154] "Mapping" is standard networking terminology for discovery of the computers on a network and the topology of the network itself; the word is even part of the name "NMAP." *See infra* note 155. "Enumeration" is defined in *Network Enumerators*, SECURITY WIZARDRY, http://www.securitywizardry.com/index.php/products/scanning-products/network-enumerators.html (last visited Jan. 6, 2014), though to some extent it is just a technical computer science term for learning a set of things, as opposed to "brute force" which is trying all possibilities to find one secret.

¶101 Another way to ascertain the system version is to perform "OS fingerprinting." OS fingerprinting involves looking for subtle differences in the network protocol implementations of different operating systems, and in particular the response of the system being examined to various probes. NMAP, a freely available popular network security tool, is most commonly used. In addition to OS fingerprinting, NMAP provides open service and open port identification and limited vulnerability scanning.[155]

¶102 The final step in the information-gathering phase is to scan the target system to see if it has common vulnerabilities.[156]

E. Finding Vulnerabilities

¶103 Once the target has been adequately identified and scanned, a suitable vulnerability must be identified. The primary criterion, of course, is compatibility with the user's operating system; another crucial criterion is mode of delivery. Some exploits, for example, can be delivered by email messages; others require the user visiting a particular web page, or opening a file containing a specific, vulnerable application. Email delivery is easiest because it does not require the user to take any particular action, but apart from the fact that it might be noticed there is always the risk that a spam filter will catch it.[157] Another class of exploits requires being on the same local network[158] as the victim, or on an interconnected network if there are no intervening firewalls.[159] Even infected USB flash drives have been used; indeed, the Stuxnet attack on the Iranian nuclear centrifuge plant is believed to have started that way.[160]

---

[155] GORDON "FYODOR" LYON, NMAP NETWORK SCANNING: OFFICIAL NMAP PROJECT GUIDE TO NETWORK DISCOVERY AND SECURITY SCANNING xxi–xxii, 205 (2008).

[156] There are a number of widely-used vulnerability scanning systems. Nessus (available from http://www.tenable.com/products/nessus) is the most widely used one; it can scan for thousands of vulnerabilities and plug-ins, and even provides detailed mobile device information like serial numbers, model, version, and last connection timestamps. *See* TENABLE NETWORK SEC., NESSUS: THE WORLD'S MOST TRUSTED VULNERABILITY SCANNER (2013), *available at* https://static.tenable.com/datasheets/nessus-datasheet.pdf. Another popular vulnerability scanning system is Nexpose (available from https://www.rapid7.com/products/nexpose/).

[157] Sending email messages crafted to appear genuine to a particular target is known as "spear-phishing." In skilled hands, spear-phishing is extremely effective. Press reports suggest that is one of the primary schemes used by cyberespionage units. *See, e.g.*, Jaikumar Vijayan, *DHS Warns of Spear-phishing Campaign Against Energy Companies*, COMPUTERWORLD (Apr. 5, 2013, 4:03 PM), https://www.computerworld.com/s/article/9238190/DHS_warns_of_spear_phishing_campaign_against_energy_companies.

[158] A LAN (Local Area Network) is generally a high-speed network that covers a relatively small area. Typical LANs include most home networks, WiFi hotspots, or, in an enterprise, a single department. LANs are interconnected to each other or to WANs (Wide Area Network) by *routers*. *See*, *e.g.,* ANDREW TANENBAUM & DAVID WETHERALL, COMPUTER NETWORKS (5th ed. 2010).

[159] Most home routers are technically known as Network Address Translators (NATs). For these purposes, NATs serve the same purpose as firewalls; these attacks cannot be launched at a target that is behind a NAT. *See* Geoff Houston, *Anatomy: A Look Inside Network Address Translators*, INTERNET PROTOCOL J., Sept. 2004, *available at* http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_7-3/ipj_7-3.pdf.

[160] *See Stuxnet Dossier, supra* note 17, at 3. It is unclear how the infected flash drive was introduced. *See, e.g.*, James Bamford, *The Secret War*, WIRED (June 12, 2013, 9:00 PM), http://www.wired.com/threatlevel/2013/06/general-keith-alexander-cyberwar/all/.

¶104 Many exploits are publicly announced,[161] and are often available in easy-to-launch pre-packaged scripts. The Metasploit Project hosts the largest database of these scripted, publicly available exploits (called "modules").[162] These modules can be utilized by a number of different exploitation applications, such as the Metasploit Framework and Core Impact Pro.[163] The National Institute of Standards and Technology (NIST) National Vulnerability Database (NVD) lists all known vulnerabilities, including what versions of what systems are affected and references to more information (but no exploit information). Information about the exploit, including an executable script or some proof-of-concept source code, is often published on one of a number of well-regarded websites and public mailing lists.[164]

¶105 Another group of exploits is privately held exploits; these include the zero-days described above,[165] as well as exploits for sale by professional security vulnerability researchers. We discuss these in detail in Section G.

¶106 Sometimes, no publicly available vulnerabilities will be usable, and the option of purchasing one from the vulnerabilities market will be undesirable or unavailable. In that case, law enforcement agents—more likely, a central "Vulnerability Lab"—must find one.[166] While this issue is out of scope here, we note there are many commonly available tools regularly used for finding vulnerabilities by software vendors trying to protect their products and by attackers.

¶107 Finally, in the rare case where directly compromising a target platform through an exploit is not possible, a technique known as a "Man-in-the-Middle" (MitM) attack might be used.[167] Such attacks involve interrupting the communications path between the target and some site the target is trying to access; the attack tool then intercepts communications intended for that resource. A successful MitM attack might be another way to launch an

---

[161] The US Computer Emergency Readiness Team (US-CERT) maintains a frequently updated list of vulnerabilities. Security researchers and privately owned research laboratories such as Vulnerability Lab and Immunity, Inc. announce vulnerabilities on websites and Twitter when they are discovered. Verified vulnerabilities are collected, categorized, and enumerated in the comprehensible, searchable NIST NVD database. *See National Vulnerability Database*, DEP'T OF HOMELAND SEC., http://web.nvd.nist.gov/view/vuln/search (last visited Feb. 5, 2014)

[162] Each of the exploits in the database consists of a specific vulnerability packaged into a module, which can be loaded into an attack application, such as the Metasploit Framework, to run. Because of the popularity of the Metasploit Framework, many exploits sold are available as Metasploit modules. *See, e.g.*, *Metasploit Exploit*, EXPLOIT HUB, https://exploithub.com/product-type/metasploit-exploit.html (last visited Sept. 24, 2013).

[163] The Metasploit Framework, available from http://www.metasploit.com, is the most widely used exploitation application available today. It is available in both free and commercial versions and has a wide developer base. *See* METASPLOIT, http://www.metasploit.com (last visited Sept. 24, 2013). Core Impact Pro can be purchased from http://www.coresecurity.com.

[164] There are many such mailing lists. Perhaps the best-known one is BugTraq, http://www.securityfocus.com/archive/1.

[165] *See supra* Section II.A.

[166] The FBI already operates the Domestic Communications Assistance Center, which apparently does at least some of this. *See, e.g.*, *Going Dark: Lawful Electronic Surveillance in the Face of New Technologies, supra* note 2, at 7 (2011) (statement of Valerie Caproni, General Counsel, Federal Bureau of Investigation); Declan McCullagh, *FBI Quietly Forms Secretive Net-Sur*veillance, CNET (May 22, 2012, 11:44 PM), http://news.cnet.com/ 8301-1009_3-57439734-83/fbi-quietly-forms-secretive-net-surveillance-unit.

[167] MitM attacks can be used at any time. However, they are almost always harder to do, since they require interfering with the traffic of exactly one user who may be at an unknown location. They are also more detectable than other attacks, although only by very sophisticated users.

attack; alternatively, it could permit acquisition of passwords and account information that would provide law enforcement with access to other useful resources.[168]

F. Exploits and Productizing

¶108 While off-the-shelf exploits may be available to law enforcement on the black market, law enforcement does not require their functionality, which is installing general purpose remote-access malware to send spam, steal bank account numbers, etc. Rather, they wish to gather specific items of data authorized by the warrant, and to do so in a form suitable for presentation in court. In addition, access to a target system by a law enforcement agent must take care to preserve evidence and chain of custody.[169] This implies due attention to precise logging of exactly what was done, when, and by whom. Consequently, off-the-shelf exploits (as opposed to vulnerabilities) are by themselves not likely to be particularly useful to law enforcement, except as a starting point or perhaps under exigent circumstances.[170] What law enforcement needs are specialized eavesdropping products, products that use exploits to produce legally acceptable communications intercepts, and do so as simply and as cheaply as possible while still complying with all legal requirements.

¶109 The three functional components of a law enforcement eavesdropping product—the exploit (which provides access to the system), the eavesdropping code, and the supporting infrastructure—all have different characteristics and lifetimes. Exploits have the shortest lifetime due to their specificity, installation characteristics, vendor patches, etc. Accordingly, a good methodology for use of exploits is the dropper/payload model, where the eavesdropping product is composed of two principal parts: a dropper and a specially encrypted payload *that is specifically encrypted for the particular target*. (This payload includes the second and third components.) A *penetrator* is used as the dropper, which is the initially injected code that exploits the actual vulnerability and thus gains access to the target system. Once access is acquired, the penetrator decrypts the payload. The payload is encrypted as a security measure to ensure the penetration code cannot easily be detected or reused by criminals; it also ensures that the payload targets the correct system. A payload is specifically encrypted for a particular target by using target-specific information like serial numbers, the MAC address, IP address, etc., as the key to encrypt and decrypt the payload.[171] The penetrator picks this information up, which would have been acquired during earlier technical reconnaissance, at payload installation

---

[168] Depending on the provisions of the original warrant, it may be necessary to seek a modification. In particular, a warrant permitting interception of communications does not grant the right to search stored email archives; that would require an order under the Stored Communications Act, 18 U.S.C. §§ 2701–2712 (2006).

[169] *S*ee Timothy M. O'Shea & James Darnell, *Admissibility of Forensic Cell Phone Evidence*, U.S. ATT'YS' BULL., Nov. 2011, at 47–49, *available at* http://www.justice.gov/usao/eousa/foia_reading_room/usab5906.pdf; *see also* U.S. DEP'T OF JUSTICE, ELECTRONIC SURVEILLANCE MANUAL PROCEDURES AND CASE LAW FORMS 27–31 (June 2005), *available at* http://www.justice.gov/criminal/foia/docs/elec-sur-manual.pdf (discussing sealing intercepts to protect their integrity).

[170] *See infra* Section IV.G.

[171] Encryption is accomplished through the use of an algorithm, which may be public, and a key, which is a piece of secret data. If the encryption algorithm is strong, it should be effectively impossible to decrypt the file without knowledge of the key.

time. This method protects untargeted machines from compromise: if the code is executed on the wrong machine, decryption will fail.

¶110 The payload itself should be designed to provide the access specified in the warrant with minimal changes to the target system. Those changes that are necessary should be logged and time-stamped as to provide documentation that vital evidence was neither altered nor destroyed. If the warrant includes provisions for recording communications, the payload should also contain provisions for minimization, including the ability to turn recording on and off and the length and time of communications recorded. Payloads do not change very much over time; while they may need to adapt to different major versions of operating systems, they generally rely on features not likely to change very often. Further, payloads that have already been installed are rarely disabled by vendor patches.

¶111 The supporting infrastructure (which is also part of the payload) has an intermediate lifetime. Some of the infrastructure, such as the code to set up encrypted channels to the investigators, is straightforward and not particularly tied to unusual law enforcement needs; this code will be quite long lived. The command-and-control subsystem—the mechanism with which investigators control the tap, turn recording on and off, etc.—is similarly straightforward, although the fine details will be specific to the application. Much of this code will be virtually the same even across different operating systems. On the other hand, the concealment mechanisms—the code that hides the existence of the payload from the computer's owner and specialists who may be hired to "sweep" the computer for bugs—is likely to be highly dependent on the operating system, including the particular version, and will change fairly frequently.

¶112 It is a good idea for the payload to have a self-destruct option, perhaps the time limit set by the warrant, after which the law enforcement software restores the target system to its pre-exploit state, erases itself, and removes all evidence of its presence.[172] This not only helps prevent proliferation, it may be necessary to comply with the legal requirements for time limits on wiretap orders.[173]

¶113 A good example of how non-proliferation might work in practice is demonstrated in a variant of Stuxnet[174] called Gauss. Discovered in August 2012, Gauss appears to be an espionage tool. [175] It uses a known vulnerability and shares some code with other known malware in its dropper, but even after several months of intense analysis, the behavior of its payload remain unknown. Gauss uses cryptographic methods and tools, and only installs and runs on machines specifically targeted by Gauss's developers; on

---

[172] Fritz Hohl, *Time Limited Blackbox Security: Protecting Mobile Agents from Malicious Hosts*, *in* MOBILE AGENTS AND SECURITY 90, 97–107 (Giovanni Vigna ed., 1998), *available at* http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.8427.

[173] *See* 18 U.S.C. § 2518(4)(e) (2006) ("Each order authorizing or approving the interception of any wire, oral, or electronic communication under this chapter shall specify . . . the period of time during which such interception is authorized . . . .").

[174] *See Stuxnet Dossier*, *supra* note 17.

[175] Dan Goodin, *Nation-Sponsored Malware with Stuxnet Ties has Mystery Warhead*, ARS TECHNICA (Aug. 9, 2012, 1:23 PM), http://arstechnica.com/security/2012/08/nation-sponsored-malware-has-mystery-warhead/; KAPERSKY LAB GLOBAL RESEARCH & ANALYSIS TEAM, GAUSS: ABNORMAL DISTRIBUTION, at 21, *available at* https://www.securelist.com/en/downloads/vlpdfs/kaspersky-lab-gauss.pdf, which provides proof of concept despite being an intelligence effort rather than a law enforcement one. The program collects a number of data items, but some of the code is encrypted with a target-specific string. This feature helps prevent proliferation.

non-targeted machines it remains encrypted and inert. Gauss also sets up a secure method to send data to its command and control centers. *Ars Technica* reports that "The setup suggests that the command servers handled massive amounts of traffic,"[176] indicating that this technique could send large amounts of data, not just a communications tap.

G. The Vulnerabilities Market

¶114 One simple way for law enforcement to obtain useful vulnerabilities is to buy them. With the availability of openly published vulnerability information and free exploitation tools, one might question why we discuss purchasing vulnerabilities or exploits from researchers at all. The answer is the improved security of target systems. As software developers and vendors have improved the quality of their software and incorporated defenses such as firewalls and anti-virus packages, vulnerabilities have become harder to find and to exploit. Software companies have also generally accelerated the rate at which they release security patches after critical vulnerabilities have been announced. This can result in a well-patched and well-maintained system more difficult to compromise. Additionally, as stated above, exploits must be carefully tailored to the individual target machine. This means it requires more skill to develop a working exploit, making new effective exploits a valuable commodity for their creator. A technically savvy target, someone who is conscientious about maintaining their system with up-to-date security patches, is also likely to be careful about not installing software from unverified sources, to use encryption, to not open links from email, and likely does not access questionable websites, and so may not be vulnerable to the easy public exploits. If law enforcement wishes to use a zero-day or lesser-known vulnerability to exploit a target, it must either have the appropriate vulnerability and exploit already on the shelf, or else it must purchase one on the open market. The market itself is a relatively recent phenomenon.

¶115 Finally, there may sometimes be a need to tap a particular suspect as quickly as possible. If there are no suitable off-the-shelf exploits available to the investigators and no time to find a new one, purchasing one may be the best option.[177]

¶116 The overt vulnerabilities marketplace had its start in 2004 when Mozilla launched the first successful bug-bounty program.[178] This program, still in effect today, pays security researchers for original vulnerabilities they discover.[179] Many other companies have followed suit with their own bug-bounty programs. Product developers, however, are not the only groups that are interested in obtaining information regarding software

---

[176] Dan Goodin, *Puzzle Box: The Quest to Crack the World's Most Mysterious Malware Warhead,* ARS TECHNICA (Mar. 14, 2013, 8:00 AM), http://arstechnica.com/security/2013/03/the-worlds-most-mysterious-potentially-destructive-malware-is-not-stuxnet/.

[177] That an exploit has been purchased instead of being developed in-house does not change the need to report it promptly. However, under urgent conditions some delay may be appropriate. *See infra* Section VII.B.

[178] *See* Press Release, Mozilla Found., Mozilla Foundation Announces Security Bug Bounty Program (Aug. 2, 2004), *available at* https://www.mozilla.org/en-US/press/mozilla-2004-08-02.html. For further examples of bug bounties, see Kim Zetter, *With Millions Paid in Hacker Bug Bounties, Is the Internet Any Safer?*, WIRED MAGAZINE (Nov. 8, 2012, 6:30 AM), http://www.wired.com/threatlevel/2012/11/bug-bounties/all/ (listing prices, total paid out, and launch date for several bug bounty programs).

[179] *See Bug Bounty Program*, MOZILLA, https://www.mozilla.org/security/bug-bounty.html (last updated May 22, 2013).

vulnerabilities. Governments and computer security service providers such as iDefense and ZDI also pay for vulnerability information, particularly if the details on how to use it have not been made public (zero-days).[180]

¶117 The overt and black markets in vulnerabilities, exploits, and zero-days have expanded in recent years.[181] Many legitimate security research firms have made finding vulnerabilities and developing exploits for sale part of their business model.[182] Companies and individuals sell information about privately discovered vulnerabilities, often with a proof-of-concept or full-blown exploit code, to groups of subscribers and to individuals. The prices of and amount of detail about the vulnerabilities made public varies. Some companies (e.g., Vulnerability-Lab) and researchers publicly announce that a vulnerability has been discovered in a particular product, but reserve actual details for their customers.[183] Other companies, such as Endgame, keep even the knowledge of the existence of the vulnerability private.[184] Prices range from $20 to $250,000,[185] with exclusive access to a critical zero-day generally the most expensive. Recent news reports suggest that national governments, in particular intelligence and military agencies, have become major buyers.[186]

¶118 Companies such as Vupen, Revuln, and Vulnerability-Lab sell subscription services that provide exclusive detailed information on disclosed or private critical

---

[180] In Feb 2006, iDefense, a vulnerability research company owned by VeriSign, Inc., offered a $10,000 prize for a 'previously unknown' Microsoft security vulnerability. One of the requirements for winning the prize was that the vulnerability be submitted exclusively to iDefense. *See* Brian Krebs, *Wanted: Critical Windows Flaw … Reward: $10,000*, SECURITY FIX (Feb. 16, 2006, 1:40 PM). http://blog.washingtonpost.com/securityfix/2006/02/wanted_critical_windows_flaw_r.html.

Similarly, it states in the frequently asked questions for Tipping Point's Zero Day Initiative that once a vulnerability has been assigned to TippingPoint, it cannot be distributed—or even discussed—elsewhere until a patch is available from the vendor. *See Frequently Asked Questions,* ZERO DAY INITIATIVE, http://www.zerodayinitiative.com/about/faq/#17.0 (last visited Oct. 7, 2013).

[181] Presumably, if criminals were the only ones interested in purchasing vulnerabilities, the market would still exist, but it would be underground. Similar markets do exist for other forms of criminal software, such as bots, credit card number loggers, etc.

[182] Some prominent examples include: Vupen Security, Vulnerability-Laboratory, Immunity, Inc., Netragard, NSS Labs, Inc., and Raytheon.

[183] Vulnerability Lab posts announcements of vulnerabilities discovered both on its website, http://www.vulnerability-lab.com, and on Twitter, https://twitter.com/vuln_lab.

[184] VUPEN Vulnerability Research Team, *Google Chrome Pwned by VUPEN aka Sandbox/ASLR/DEP Bypass,* VUPEN SECURITY (May 9, 2011, 5:35 PM), http://www.vupen.com/demos/VUPEN_Pwning_Chrome.php ("For security reasons, the exploit code and technical details of the underlying vulnerabilities will not be publicly disclosed. They are available to our customers as part of our vulnerability research services."); *Vulnerability Feeds*, REVULN, http://revuln.com/services.htm#vulnfeeds (last visited Nov. 13, 2013) (explaining that Revuln sells access to its 0-day Feed, which provides "[i]nformation about undisclosed and unpatched security vulnerabilities found by [their] team in third party hardware and software products of various vendors. The vulnerabilities included in [their] 0-day feed remain undisclosed by ReVuln unless either the vulnerability is discovered and reported by a third party or the vendor publicly or privately patches the issue.").

[185] Exploits currently offered for public sale from a wide variety of independent researchers can be purchased from http://exploithub.com. Further examples of exploits offered for public sale can be found in Andy Greenberg, *Meet the Hackers Who Sell Spies the Tools to Crack Your PC (And Get Paid Six-Figure Fees)*, FORBES (Mar. 21, 2012), http://www.forbes.com/sites/andygreenberg/2012/03/21/meet-the-hackers-who-sell-spies-the-tools-to-crack-your-pc-and-get-paid-six-figure-fees/.

[186] *See* Nicole Perlroth & David E. Sanger, *Nations Buying as Hackers Sell Flaws in Computer Code*, N.Y. TIMES, July 13, 2013, https://www.nytimes.com/2013/07/14/world/europe/nations-buying-as-hackers-sell-computer-flaws.html.

vulnerabilities to governments, law enforcement authorities, and corporations.[187] Annual subscriptions can run as high as $100,000 a year.[188] These companies also sell working exploits and offer special targeted exploit development for additional fees; exploit prices range from $5,000 to $250,000. The most valuable are those zero-days that can be used for cyber warfare. For example, the Endgame Systems pricelist includes a twenty-five exploit package for $2.5 million.[189] Zero-days and exploits can also be purchased from exploit brokers such as Netragard or private brokers who bid on exploits from sellers and negotiate with buyers on behalf of individual exploit developers.[190]

¶119 The FBI has already used vulnerabilities to download exploits and extract information from various targets machines. But if law enforcement uses vulnerabilities and exploits to conduct wiretaps when other methods fail[191] (and as an alternative to CALEA-style taps in the intellectual property world), it will face a difference in scale in the use of such techniques—and thus a difference in kind. That raises not just technical questions, but complex ethical and legal concerns as well. In the sections that follow, we turn to those.

---

[187] *See, e.g.*, VUPEN SECURITY, VUPEN THREAT PROTECTION PROGRAM, *available at* http://wikileaks.org/spyfiles/files/0/279_VUPEN-THREAD-EXPLOITS.pdf (last visited Nov. 16, 2013).

[188] *See* Perlroth & Sanger, *supra* note 186.

[189] Michael Riley & Ashlee Vance, *Cyber Weapons: The New Arms Race*, BLOOMBERG BUSINESSWEEK MAG. (July 20, 2011), http://www.businessweek.com/magazine/cyber-weapons-the-new-arms-race-07212011.html#p4 (quoting David Baker, the vice-president for services at the security firm IOActive, as saying, "'Endgame is a well-known broker of zero days between the community and the government.' By 'community,' he means hackers—'Some of the big zero days have ended up in government hands via Endgame . . . .'").

[190] A number of reports have been published recently documenting the vulnerabilities market and the brokers who negotiate between buyers and sellers. *See The Digital Arms Trade*, THE ECONOMIST (Mar. 30, 2013), http://www.economist.com/news/business/21574478-market-software-helps-hackers-penetrate-computer-systems-digital-arms-trade; *Zero Day Exploit Acquisition Program*, NETRAGARD, http://www.netragard.com/zero-day-exploit-acquisition-program; Andy Greenberg, *Shopping For Zero-Days: A Price-List for Hackers' Secret Software Exploits*, FORBES (Mar. 23, 2013, 9:43 AM), http://www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/.

[191] The FBI has said very little about its use of vulnerabilities, let alone why it uses them. Examination of available evidence suggests that their primary reason is when they do not know where the target system is; see, for example, Kevin Poulsen, *FBI Admits it Controlled Tor Servers Behind Mass Malware Attack*, WIRED (Sept. 13, 2013, 4:17 PM), http://www.wired.com/threatlevel/2013/09/freedom-hosting-fbi/, which discusses how the FBI used malware to identify child porn viewers who had used Tor. Also note that the FBI would not talk to the press about it, but did talk in court when they had to. See *In Re* Warrant to Search a Target Computer at Premises Unknown, No. H-13-234M, 2013 WL 1729765 (S.D. Tex. Apr. 22, 2013) for an example of such a case.

V. PREVENTING PROLIFERATION

¶120 As should already be clear, the use of an exploit to download a wiretap is far more complex than simply placing two alligator clips on a wire.[192] But what is a far more serious impediment to using exploits is that the exploits employed in the installation of the wiretap may spread beyond the targeted device. Given that possibility, does the government even have the moral right to use vulnerabilities in its efforts to combat crime and protect national security? We consider this issue, and then examine techniques to prevent proliferation of the exploit beyond the intended target.

A. Public Policy Concerns in Deploying Exploits to Wiretap

¶121 We start with some assumptions. First, there is probable cause that the suspect is committing a serious crime and using the targeted communications device to do so. Second, other means of investigation have been tried and have not netted the requisite information. Third, a wiretap order has been authorized, but the target is using a communications device that prevents the standard methods of interception from working. Is it moral to use an exploit to intercept the communication when there is some risk, however small—but perhaps larger than anticipated— that the exploit may escape the device and be used elsewhere, causing great harm?

¶122 The problem of potentially doing harm in the process of doing good is a well-known problem in philosophy known as "the doctrine of double effect," in which one pursues a moral action that has a consequence of causing harm. The philosopher Phillipa Foot argued that the distinctions should be between direct intention and oblique action, between avoidance of harm and activities to help,[193] and between duties and voluntary actions. She constructed a series of trenchant examples to illustrate this, including the following:

- Should a judge who is faced with an angry crowd demanding justice, frame and order the execution of an innocent person to save many others from deaths through rioting?[194]

¶123 Foot observes that the salient issue is not justice, but rather direct versus oblique effects.[195] That is the distinction between what we do (direct intention) and what we allow (oblique action). The judge should not hang an innocent man—direct effect—even if more people die as a result of the rioting that ensues.

¶124 Foot makes a distinction between negative duties—avoidance of harm—and positive duties—bringing aid,[196] as well as between duties and voluntary actions, and concludes that a critical distinction is whether one is bringing aid—a voluntary action—or performing one's duty.[197] Foot illustrates the issue with another example:

---

[192] See *supra* note 20.

[193] PHILLIPA FOOT, VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19–32 (1978).

[194] *Id.* at 23.

[195] *Id.* at 24 ("To choose to execute [an innocent man] is to choose that this evil *shall come about*, and this must therefore count as a *certainty* in weighing up the good and evil involved.").

[196] *Id.* at 25.

[197] *Id.* at 29.

- Should the driver of a runaway tram deliberately aim the tram at one man on the track to stop it or steer the other way, where five men are working and will be killed?[198]

¶125 The driver of the tram is performing a duty and has a responsibility to injure as few people as possible. The driver would be behaving morally in electing to take the track with the single individual.

¶126 In using vulnerabilities to execute wiretaps, law enforcement investigators are performing their required duty of investigating a criminal activity. Under Title III, if a wiretap order is granted this means that evidence is essentially unobtainable in other ways.[199] The duty of investigating the criminal activity may require wiretapping. If the only way to affect the wiretap is through the use of an exploit, then, following the logic presented by Foot regarding duty, this is the way to proceed. *But there must be due diligence to contain the harm*. There are several aspects to containing the harm, including fully vetting necessity and balancing it against the harm that may result and designing the exploit to prevent proliferation beyond the target.[200]

¶127 The law balances competing social goods. For example, the Fourth Amendment balances the social good to society of protecting itself against the social good of protecting individual privacy and security.[201] Law enforcement's use of vulnerabilities can be considered within the same framework of competing social goods. Use of vulnerabilities, at least without reporting them, is not unlike police use of confidential informants (CIs). CIs inform investigations even while aiding criminal activity.

¶128 A common law enforcement tactic is to use a lesser criminal to gather evidence about a higher-up criminal. Within limits, crimes (including further crimes) committed by a "flipped" individual are largely forgiven, so long as that person is providing good evidence against the real target of the investigation. As Daniel J. Castleman, chief of the Investigative Division of the Manhattan district attorney's office, explained, "With confidential informants we get the benefit of intimate knowledge of criminal schemes by criminals, and that is a very effective way to investigate crime . . . ."[202]

¶129 What happens with wiretaps implemented via exploits is ultimately not very different. In both cases law enforcement seeks to catch what it believes to be a genuinely dangerous criminal. But here it seeks to do so by the collection of wiretap evidence.

---

[198] *Id.* at 23.

[199] Recall that 18 U.S.C. § 2518(3)(c) (2006) requires that "normal investigative procedures have been tried and have failed or reasonably appear to be unlikely to succeed if tried or to be too dangerous." *But see* United States v. Smith, 893 F.2d 1573, 1582 (9th Cir. 1990) ("Although a wiretap should not be used routinely as the first step in a criminal investigation, it need not be the last resort.").

[200] There are other harms that may result from using the exploit, such as excessive collection, but these are not substantively different from concerns in "normal" wiretapping efforts. The issue of proliferation is substantively different.

[201] While the usual interpretation of the Fourth Amendment is that it centers on protecting the privacy of the individual against searches by the state, Jed Rubenfeld convincingly argues that the amendment really concerns providing security for individuals against searches by the state. *See* Jed Rubenfeld, *The End of Privacy*, 61 STAN. L. REV. 101, 120–38 (2008).

[202] Alan Feuer & Al Baker, *Officers' Arrest Put Spotlight on Police Use of Informants*, N.Y. TIMES, Jan. 27, 2008, at 26.

Installing the tap requires exploiting a vulnerability that law enforcement hopes will not be repaired before the tap is in place.

¶130 The purchase and secret use of vulnerabilities raises several similar moral dilemmas as the use of confidential informants (CIs). The history of police use of CIs is replete with instances where an informant went too far, committing or failing to stop serious criminal activity; this has even included murder.[203] With wiretaps the "too far" is of a somewhat different character, but with similar consequences: some crimes that the government could have stopped may not be prevented. By not reporting the vulnerability to the vendor and speeding its repair, law enforcement's inactivity is potentially enabling criminal activity against users of the hardware or software. It is thus useful to examine how law views the competing interests of preventing crime versus investigating criminal activity in the use of confidential informants, the closest analogy that exists in practice to the use of unreported vulnerabilities.

¶131 In *United States v. Murphy,* the Seventh Circuit considered a case in which FBI agents created fictitious cases in the Cook County Courts in order to uncover corruption within the legal system.[204] The Seventh Circuit ruled that the false cases were a legitimate investigatory tool, observing that "the phantom cases had no decent place in court. But it is no more decent to make up a phantom business deal and offer to bribe a Member of Congress. In the pursuit of crime the Government is not confined to behavior suitable for the drawing room. It may use decoys, . . . and provide the essential tools of the offense . . . . The creation of opportunities for crime is nasty but necessary business."[205]

¶132 The choice to use vulnerabilities without also simultaneously reporting them to the vendor is not precisely "the creation of opportunities for crime," but rather the choice not to pro-actively prevent crime. *Murphy* makes clear that this type of approach can be legally legitimate. Whether it is acceptable is a moral, public policy, and political question.

¶133 Department of Justice guidelines on the use of confidential informants state that a Justice Law Enforcement Agent (JLEA) is never permitted to authorize a CI to "participate in an act of violence; . . . participate in an act that constitutes obstruction of justice (e.g., perjury, witness tampering, witness intimidation, entrapment, or the fabrication, alteration, or destruction of evidence); . . . participate in an act designed to obtain information for the JLEA that would be unlawful if conducted by a law enforcement agent (e.g., breaking and entering, illegal wiretapping, illegal opening or tampering with the mail, or trespass amounting to an illegal search); or . . . initiate or instigate a plan or strategy to commit a federal, state, or local offense."[206] The guidelines do not state, however, that a CI *must* work to prevent a crime from occurring. The

[203] There are multiple such examples, including the well-known shooting of Viola Liuzzo, a white supporter of the Civil Rights movement who was shot by Ku Klux Klan members while driving from a march in Selma, Alabama, one of whom was an FBI informant. DIANE MCWHORTER, CARRY ME HOME: BIRMINGHAM, ALABAMA: THE CLIMACTIC BATTLE OF THE CIVIL RIGHTS REVOLUTION 572–73 (2001).

[204] 768 F.2d 1518, 1524 (7th Cir. 1985).

[205] *Id.* at 1529.

[206] Illegal activity must be authorized in advance for a period of up to ninety days. *See* DEP'T OF JUSTICE, DEPARTMENT OF JUSTICE GUIDELINES REGARDING THE USE OF CONFIDENTIAL INFORMANTS (Jan. 8, 2001), http://www.justice.gov/ag/readingroom/ciguidelines.htm.

analogous situation to the use of vulnerabilities would be that law enforcement is not required to let vendors know about the vulnerabilities they find and exploit.

¶134 Immediately reporting versus using for some time before reporting is a clash of competing social goods, which is what we need to weigh here. If our primary concern is preventing the proliferation of exploits, society will be better protected by reporting the vulnerability early even if that risks the ability of the criminal investigation to conduct its authorized wiretap.

¶135 As we know from other situations, whether rare diseases or the effect of cold weather on shuttle O-rings,[207] a rare side effect is more likely to appear when working with a large population sample. The danger of proliferation means each use of an exploit, even if it has previously run successfully, increases the risk that the exploit will escape the targeted device. This introduces a serious wrinkle in the use of vulnerabilities, one that law enforcement must address, and that we discuss in subsection C and section VI, *supra*.

B. Ethical Concerns of Exploiting Vulnerabilities to Wiretap

¶136 Even though wiretaps have long been accepted as a tool in law enforcement's toolbox, there is something distasteful about using an exploit to download interception capability. Undoubtedly, part of that distaste stems from the strong sense that vulnerabilities are to be patched, not exploited. But even if law enforcement were never to report the vulnerabilities it discovers or purchases, law enforcement's use of vulnerabilities would not make the vulnerability situation worse. Law enforcement does not currently report vulnerabilities to vendors. Thus, were law enforcement to use vulnerabilities and not report them to the vendors, there would be no change to the status quo ante. That said, there are still some concerns raised by law enforcement's use of vulnerabilities.

¶137 One danger of law enforcement's participation in the zero-day market is the possibility of skewing the market, either by increasing incentives against disclosure of the vulnerability or by increasing the market for vulnerabilities and thus encouraging greater participation in it. Because of the current size of the market and the relatively minimal need by law enforcement, we do not believe that this will be an issue. It is hard to know exactly under which circumstances vulnerabilities will be used since the FBI has not discussed under what technical circumstances they have encountered difficulties wiretapping, but we do believe usage will be rare.

¶138 What is the government's responsibility in cases where the operationalized vulnerability escapes the target? It is not unheard of for physical searches to go amiss; sometimes law enforcement executes a warrant on the wrong location or executes a wiretap warrant on the wrong phone line.[208] Such a search would, of course, invalidate

---

[207] Howard Berkes, *Reporting a Disaster's Cold, Hard Facts*, NPR (Jan. 28, 2006, 1:27 PM), http://www.npr.org/templates/story/story.php?storyId=5175151.

[208] *See, e.g.*, INTELLIGENCE OVERSIGHT BOARD MATTER, [REDACTED] DIVISION, FEDERAL BUREAU OF INVESTIGATION HEADQUARTERS, IOB MATTER 2005-160 (June 30, 2010), *available at* https://www.eff.org/sites/default/files/filenode/intel_oversight/IOB%202005-160.pdf. It is rare that such activity is publicly reported. *Documents Obtained by EFF Reveal FBI Patriot Act Abuses*, ELECTRONIC FRONTIER FOUNDATION (Mar. 31, 2011), https://www.eff.org/deeplinks/2011/03/documents-obtained-eff-

collection. But a wiretap exercised through an operationalized payload is a significantly different situation. Unlike an incorrectly executed wiretap warrant, which might simply collect information on the wrong party, a badly designed payload could escape its target and potentially affect a much larger group of people.

¶139 If the operationalized vulnerability were to escape its target, it might be adapted for malicious purposes by others, a second-order affect that increases the need for great care in developing the exploits. While the government may have some liability when it knocks down the wrong door in the course of exercising a search warrant,[209] with wiretap software the liability—in dollars or simply in costs to society— is not as well understood.

¶140 As a result, it is critical that the tools employed by law enforcement be trustworthy and reliable. In particular, the technical implementation must capture only what is authorized. In addition, all the usual security provisions apply: the system must employ full auditing of actions taken or system changes made,[210] each user of the system must log on individually, etc.[211] Such careful controls have not always been exercised in the past, as is evidenced by flaws discovered in the FBI's DCS 3000 wiretap system,[212] as well as poor documentation of telephone transactional data requests during FBI investigations post-September 11th.[213] This argues for not only judicial oversight, but technical oversight as well.

¶141 Finally, one might imagine a scenario in which law enforcement puts pressure on vendors not to fix vulnerabilities so as to facilitate exploits. Aside from being bad public policy, such an approach would be dangerous for both government and industry. If such pressure became publicly known, the vendor would suffer serious reputational harm. It is not inconceivable that the vendor could also be liable to customers for damages if the company knew of a serious vulnerability about which it had neither informed its customers nor patched to eliminate the vulnerability.[214]

C. Technical Solutions to Preventing Proliferation

¶142 The principle of only harming the target must govern the use of vulnerabilities by law enforcement. One means of ensuring that only the target is harmed is to employ

---

reveal-fbi-patriot-act.

[209] *Cf.* Jim Armstrong, *FBI Uses Chainsaw in Raid on Wrong Fitchburg Apartment*, CBS BOSTON (Jan. 31, 2012, 11:59 PM), http://boston.cbslocal.com/2012/01/31/fbi-uses-chainsaw-in-raid-on-wrong-fitchburg-apartment/.

[210] This was missing in the Greek wiretapping case. *See* Prevelakis & Spinellis*, supra* note 4.

[211] There are many commercial and government guides to operating secure computer systems. *See, e.g.*, *Operating Systems,* NAT'L SECURITY AGENCY, http://www.nsa.gov/ia/mitigation_guidance/security_configuration_guides/operating_systems.shtml (last updated Aug. 14, 2013).

[212] The system was previously known as Carnivore. *See* Steven M. Bellovin, Matt Blaze, David Farber, Peter Neumann & Eugene Spafford, *Comments on the Carnivore System Technical Review* (Dec. 3, 2000), http://www.crypto.com/papers/carnivore_report_comments.html.

[213] U.S. DEP'T OF JUSTICE, A REVIEW OF THE FEDERAL BUREAU OF INVESTIGATION'S USE OF EXIGENT LETTERS AND OTHER INFORMAL REQUESTS FOR TELEPHONE RECORDS 46–47, 70 (Jan. 2010), *available at* https://www.eff.org/sites/default/files/filenode/intel_oversight/IOB%202005-160.pdf.

[214] Current law (such as UCC Article 2) and the wording of end-user license agreements (EULAs) make this outcome unlikely. Note, though, that some computer worms have affected people who were not parties to these agreements: the worms' spread clogged the Internet sufficiently that other people could not use it. *See generally* Jane Chong, *We Need Strict Laws if We Want More Secure Software*, New Republic (Oct. 30, 2013), http://www.newrepublic.com/article/115402/sad-state-software-liability-law-bad-code-part-4.

technical mechanisms to restrict an exploit to a given target machine. The simplest mechanisms check various elements of their environment when they run, e.g., the machine's serial number or MAC address, and if they are on the wrong machine silently exit. Stuxnet employed this technique. [215] A more sophisticated technique is to use environmental data to construct a cryptographic key; if this data is not present, a key cannot be constructed and the data will not decrypt properly, and the code will not be comprehensible to any analyst. Gauss malware uses this technique, and has stymied top cryptanalysts for months.[216]

¶143 From one perspective, the part of the exploit that contains the vulnerability is the most important piece, since knowledge of it will let people write their own exploit code. The best defense against this is to use a dropper/payload architecture; that way, after the initial penetration there is no further need for the vulnerability and the code relying on it can be deleted.[217]

¶144 Promiscuous spread of penetration tools also increases the risk of proliferation. The more machines a piece of code is on, the more likely it is that someone will notice the code and reverse-engineer it. This would expose not just a carefully husbanded vulnerability, but also the surrounding infrastructure necessary to use it for lawful intercepts. This calculus is similar to one found in the intelligence community: if one acts on intelligence, one risks giving away the source of information, which would then be unavailable in the future.[218]

VI. REPORTING VULNERABILITIES

¶145 The CIPAV cases[219] demonstrate that the state employs vulnerabilities for searches[220]—the "can" problem—so we turn to the "may" problem: namely, may law enforcement do so?[221] We have already argued that the security risks that would be created by extending CALEA to IP-based communications make it a poor choice. In contrast, if the vulnerability being used to introduce a wiretap already exists, the issue is somewhat different, and the question instead concerns patching. If a vulnerability in a communications application or infrastructure is patched, the vulnerability cannot be exploited for a wiretap. But if the vulnerability is left unpatched, the result is that many are left open to attack. Thus the issue is not about introducing an exploit, but about when, and perhaps whether, to inform the vendor of the vulnerability.

---

[215] *See Stuxnet Dossier*, *supra* footnote 17.

[216] *See supra* notes 175–78 and accompanying text.

[217] The best analogy to a "dropper" is a lock pick. Once you've unlocked the door—i.e., once the dropper has used a vulnerability to penetrate the system—you no longer need the lock pick; you can move around freely inside the house. You can even open the door again, from the inside, to bring in new materials, i.e., the "payload."

[218] *See* DAVID KAHN, THE CODE-BREAKERS (1967). The theme that if one acts on intelligence, one risks giving away the source of the information, which will then be unavailable in the future pervades the book, but the discussion of the assassination of Admiral Isoroku Yamamoto on pgs. 595–601 is especially illustrative.

[219] *See supra* notes 131–33 and accompanying text.

[220] *See* Lynch, *supra* note **Error! Bookmark not defined.**.

[221] We are indebted to Marty Stansell-Gamm for the phrasing of the "may" versus "can" problem.

¶146 What is law enforcement's responsibility with regard to reporting? We start by examining the security risks created by using vulnerabilities, then consider that risk in the context of law enforcement's role in crime prevention.

A. Security Risks Created by Using Vulnerabilities

¶147 As we have already noted in Section V, there is a danger that even the most carefully crafted exploitation tools may not function as intended. There are at least three security concerns that must be weighed in choosing to use a vulnerability to conduct a wiretap: (i) the risk that the vulnerability's use will lead to overcollection, (ii) the danger that the penetration tools may have unintended side effects on the targeted system, and (iii) the danger that the vulnerability will accidentally escape its target device and find use elsewhere. (This latter point is discussed in Section V.C, *supra*.)

¶148 Unfortunately there is much precedent for overcollection. Recent examples include the NSA's overcollection[222] as a result of the FISA Amendments Act[223] and the FBI's use of "exigent" letters to collect communications transactional data.[224] Use of the vulnerabilities requires close scrutiny by judges to ensure that what is collected is only what is authorized to be collected. Judges will therefore need to evaluate just how intrusive a particular exploit may be, a technical as well as legal issue.

¶149 The wiretap statute requires that taps be done "with a minimum of interference" with the service being monitored.[225] If an exploit causes other harm to the target computer, such as damaging files or applications or leading to frequent crashes, use of the exploit would violate this provision. At least one court has already quashed an eavesdropping order on these grounds:

> Looking at the language of the statute, the "a minimum of interference" requirement certainly allows for *some* level of interference with customers' service in the conducting of surveillance. We need not decide precisely how much interference is permitted. "A minimum of interference" at least precludes total incapacitation of a service while interception is in progress. Put another way, eavesdropping is not performed with "a minimum of interference" if a service is *completely* shut down as a result of the surveillance. [226]

¶150 It is worth noting that in this case, there were no allegations of instances of the customer trying and failing to use the service; however, use of the wiretap would make the original service unavailable to the customer if requested.[227]

---

[222] A major concern was that the collection inappropriately included communications of Americans without particularized FISA warrants. *See, e.g.*, James Risen & Eric Lichtblau, *Extent of E-Mail Surveillance Renews Concern in Congress*, N. Y. TIMES, June 16, 2009, at A1.

[223] Foreign Intelligence Surveillance Act of 1978, Pub. L. No. 95-511, 92 Stat. 1783, *as amended by* Foreign Intelligence Surveillance Act of 1978 Amendments Act of 2008, Pub. L. 110-261, 122 Stat. 2436.

[224] The multiple problems included: (i) many of the exigent letters never received proper follow-up by National Security Letters, (ii) sometimes private subscriber data was given to the FBI without a written request, (iii) many of the exigent letter requests failed to specify a date, thus leading to a response that included information well outside the intended investigatory period, (iv) many of the requests were not related to an actual emergency, etc. *See* U.S. DEP'T OF JUSTICE, *supra* note 213, at 257–72.

[225] 18 U.S.C. § 2518(4) (2006).

[226] Company v. United States, 349 F.3d 1132, 1145 (9th Cir. 2002) (internal citations omitted).

[227] *Id.* at 1134–35.

¶151 Apart from legal considerations, it is worth noting that interference can lead to discovery of the tap. This has happened at least twice in what appear to have been intelligence operations. During a very sophisticated wiretap operation mounted against a Greek cellphone operator, a bug in the attacking software caused some text messages not to be delivered. The resulting error messages led to discovery of the implanted code.[228] In a better-known case, the Stuxnet virus aimed at the Iranian nuclear centrifuge plant was discovered when a computer user became suspicious and sent a computer to a Belarusian antivirus firm for analysis.[229]

B. Preventing Crime

¶152 The question of when to report vulnerabilities that are being exploited is not new for the U.S government. In particular, the National Security Agency (NSA) has faced this issue several times in its history, as we discuss below.

¶153 The NSA performs two missions for the U.S. government: the well-known mission of signals intelligence, or SIGINT, which involves "reading other people's mail,"[230] and the lesser-known mission of communications security, COMSEC, which involves protecting U.S. military and diplomatic communications.[231] In principle, it is extremely useful to house the U.S. signals intelligence mission in the same agency as the U.S. communications security mission because each is in a position to learn from the other. SIGINT's ability to penetrate certain communication channels could inform COMSEC's knowledge of potential weaknesses in our own and COMSEC's awareness of security problems in certain communications channels might inform SIGINT's knowledge of a target's potential weakness.

¶154 Reality is in fact very different. COMSEC's awareness of the need to secure certain communications channels has often been thwarted by SIGINT's desire that patching be delayed so that it can continue to exploit traffic using the vulnerability in question. How this contradictory situation is handled depends primarily on where the vulnerable communications system is operating. If the insecure communications system is being used largely in the U.S. and in smaller nations that are unlikely to harm the U.S., then patching would not hurt the SIGINT mission. In that situation, COMSEC is allowed to inform the vendor of the vulnerability. In most other instances, informing the vendor is delayed so that SIGINT can continue harvesting product. Although this was never a publicly stated NSA policy, this modus operandi was a fairly open secret.[232]

¶155 Law enforcement operates in a different domain than the military, so its considerations and values are different. The FBI's concern that it is "going dark" is in

---

[228] *See* Prevelakis & Spinellis, *supra* note 4.

[229] *See* John Borland, *A Four-Day Dive Into Stuxnet's Heart*, WIRED (Dec. 27, 2010, 8:27 PM), http://www.wired.com/threatlevel/2010/12/a-four-day-dive-into-stuxnets-heart/.

[230] Henry Stinson, the Secretary of State who shut down the "Black Chamber," the Army's signals intelligence section during and after World War I, famously said, "Gentlemen do not read each other's mail." His views changed during World War II when he was Secretary of War; the U.S. relied heavily on signals intelligence during that conflict. Though the quote is attributed to Stinson, there is some evidence that he was acting on President Hoover's orders. *See* DAVID KAHN, THE READER OF GENTLEMEN'S MAIL: HERBERT O. YARDLEY AND THE BIRTH OF AMERICAN CODEBREAKING (2004).

[231] The COMSEC mission is performed by the NSA's Information Assurance Division.

[232] Interview with redacted source, Feb. 24, 2013, on file with author Susan Landau.

regard to domestic wiretapping; law enforcement wants to exploit the vulnerabilities *exactly* when there are users in the U.S. Thus the balancing that NSA does between its SIGINT and COMSEC missions does not particularly illuminate what the state of affairs should be for the FBI. We must instead examine the issue from other vantage points.

¶156 One criterion that law enforcement should use is the likelihood of collateral damage from using vulnerabilities. By their nature some vulnerabilities are easier to exploit than others. More critically, some vulnerabilities are likely to be easier for law enforcement to exploit than for the general population of attackers to do so. Any attack that is aided by the ability to use compulsory legal process against a third party, such as an ISP, falls into this category. In these cases, failure to report the vulnerability to the vendor is less likely to have an effect on its exploitation by others.

¶157 There are also other factors that can make launching an exploit complicated, like needing knowledge of special information or material about the target. If possession of such knowledge or information is necessary for the vulnerability to be exploited, then law enforcement can be fairly confident that there is little risk in not reporting the vulnerability to the vendor.

¶158 In considering whether to report a vulnerability, law enforcement should consider how dangerous a particular vulnerability may be. Sometimes this question will be very easy to answer. If the vulnerability is in a network router or a switch, its impact is likely to be very large. Indeed, vulnerabilities in network infrastructure are fundamentally a national security risk because network devices are either ISP-grade gear, whose compromise could be used to shut down or tap a large portion of the network; enterprise gear, whose compromise could be used for targeted espionage attacks; or consumer gear, likely to be in wide use and thus the compromise could effect a large population. Without question, such vulnerabilities should be reported to the vendor immediately.

¶159 There are subtleties involved even if a vulnerability does not initially appear to be one that could create a national security risk. If the vulnerability is for an uncommon platform, it would seem that not informing the vendor of the problem is unlikely to create much risk. If the vulnerability is for an outdated version of a platform, depending on how outdated the platform is, the risk may also be relatively minor.[233] The latter is especially true for devices that are replaced frequently, e.g., smart phones. Yet it is often the case that outdated systems may be widely deployed in non-critical systems or even deployed in critical systems,[234] so that a vulnerability that exists in an outdated version of a platform may still be widely dangerous; it depends on exactly on who is using the

[233] This issue makes for an interesting insight into pirated software. The fact that a high percentage of software in China is illegally obtained has several implications for electronic surveillance. The most significant implication is probably that the versions are not only out of date—e.g., as of January 2013, 62% of Chinese Windows users had Windows XP installed, while 32% had Windows 7, *StatCounter Global Stats: Top 7 Operating Systems in China from Feb 2012 to Jan 2013*, http://gs.statcounter.com/#os-CN-monthly-201202-201301 (last visited Feb. 17, 2013)—but also that they are less secure than more modern systems. Thus, they are more open to exploitation.

[234] One example of this is Windows XP; the eleven-year-old OS is still the most common operating system in use at most government agencies. Shawn McCarthy, *8 Reasons Agency IT Will Change Course in 2013*, GCN (Nov. 16, 2012) http://gcn.com/articles/2012/11/16/8-reasons-agency-it-will-change-course-in-2013.aspx. Another is the backend system supporting voting machines in Ohio. PATRICK MCDANIEL ET AL., EVEREST: EVALUATION AND TESTING OF ELECTION-RELATED EQUIPMENT, STANDARDS, AND TESTING(Dec. 7, 2007), *available at* http://www.sos.state.oh.us/SOS/upload/everest/14-AcademicFinalEVERESTReport.pdf

platform and in what situation. This demonstrates the complexity of determining when the vendor should be told about the vulnerability.

¶160 This raises the concern of whether the FBI will actually be able make an evaluation of whether a vendor should be informed of a vulnerability. As the examples above show, the ability to discern the potential risk from any particular vulnerability ranges from relatively trivial to quite difficult. One limitation on the FBI's ability to make an evaluation is that the Domestic Communications Assistance Center (DCAC) does not have the expertise to be a cybersecurity vulnerability research center.[235] Nor should it have; that expertise lies with the NSA's Information Assurance Directorate, and duplicating the expertise is neither possible nor appropriate. Making such evaluations requires vast knowledge about systems being employed in the U.S. across a wide array of industries. Even a decade after September 11th, this information is not being tracked by the U.S. government. The FBI is certainly not in a position to know this information, or to be able to make the determination about how dangerous to the U.S. a particular vulnerability may be.

¶161 The point is that except for some obvious cases, it is usually very difficult to determine a priori whether a particular vulnerability is likely to create a serious problem. [236] It could be that some obscure, but critical part of society relies on the code with the vulnerability. It could also be that it lies in some hidden part of the nation's critical infrastructure; for example, for decades American Airlines relied on old software for planning flight operations.[237] Furthermore—and especially in an open-source world, where it may be impossible to determine all the users of a system—there is no way that law enforcement would be in a position to do a full mapping from software to users, because there is no way to tell whom they all are.

¶162 As we alluded to earlier, this is a clash of competing social goods between the security obtained by patching as quickly as possible and the security obtained by downloading the exploit to enable the wiretap to convict the criminal. Although there are no easy answers, we believe the answer is clear. In a world of great cybersecurity risk, where each day brings a new headline of the potential for attacks on critical infrastructure,[238] where the Deputy Secretary of Defense says that thefts of intellectual property "may be the most significant cyberthreat that the United States will face over the long term,"[239] public safety and national security are too critical to take risks and leave vulnerabilities unreported and unpatched. We believe that law enforcement should always err on the side of caution in deciding whether to refrain from informing a vendor of a vulnerability. Any policy short of full and immediate reporting is simply inadequate.

---

[235] See McCullagh, supra note 166.

[236] A striking example of an obviously dangerous vulnerability occurred with the February 2013 US-CERT alert concerning Java; the organization recommended disabling Java in web browsers until an adequate patch had been prepared. *Alert (TA13-032A): Oracle Java Multiple Vulnerabilities*, US-CERT (Feb. 1, 2013), https://www.us-cert.gov/ncas/alerts/TA13-032A.

[237] Robert L. Mitchell & Johanna Ambrosio*, From Build to Buy: American Airlines Changes Modernization Course Midflight,* COMPUTERWORLD (Jan. 2, 2013), https://www.computerworld.com/s/article/9234936/From_build_to_buy_American_Airlines_changes_modernization_course_midflight_.

[238] *See, e.g.*, Kim Zetter, *Researchers Uncover Holes That Open Power Stations to Hacking*, WIRED (Oct. 16, 2013, 12:00 PM), http://www.wired.com/threatlevel/2013/10/ics/.

[239] William J. Lynn III, *Defending a New Domain*, 89 FOREIGN AFF. 97, 100 (2010).

"Report immediately" is the policy that any crime-prevention agency should have, even though such an approach will occasionally hamper an investigation.[240]

¶163 Note that a report immediately policy does not foreclose exploitation of the reported vulnerability by law enforcement. Vulnerabilities reported to vendors do not result in immediate patches; the time to patch varies with each vendor's patch release schedule (once per month, or once every six weeks is common), but, since vendors often delay patches,[241] the lifetime of a vulnerability is often much longer. Research shows that the average lifetime of a zero-day exploit is 312 days.[242] Furthermore, users frequently do not patch their systems promptly, even when critical updates are available.[243]

¶164 Immediate reporting to the vendor of vulnerabilities considered critical will result in a shortened lifetime for particular operationalized exploits, but it will not prevent the use of operationalized exploits. Instead, it will create a situation in which law enforcement is both performing criminal investigations using the wiretaps enabled through the exploits, and crime prevention through reporting the exploits to the vendor. This is clearly a win/win situation.

---

[240] There are persistent rumors that government agencies have sometimes pressured vendors to leave holes unpatched. *See, e.g.*, Graeme Burton, *Microsoft Gives Zero-Day Vulnerabilities to US Security Services - Bloomberg*, COMPUTING (June 14, 2013), http://www.computing.co.uk/ctg/news/2274993/microsoft-gives-zeroday-vulnerabilities-to-us-security-services-bloomberg. This is a very dangerous path, one that should not be followed by law enforcement agencies.

[241] On the second Tuesday of every month, Microsoft issues patches both for software defects and vulnerabilities. This date is known as "Patch Tuesday." Vendors who use a 6-week "rapid-release cycle," such as Google (Chrome) and Mozilla (Firefox, Thunderbird), frequently roll their security patches into their new releases. However, not all vulnerabilities discovered are patched in the next release. *See, e.g.*, Tony Bradley, *Patch Tuesday Leaves Internet Explorer Zero Day Untouched*, PC WORLD (Apr. 9, 2013, 12:55 PM), http://www.pcworld.com/article/2033649/patch-tuesday-leaves-internet-explorer-zero-day-untouched.html; Michael Mimoso, *Oracle Leaves Fix for Java SE Zero Day Until February Patch Update*, THREATPOST (Oct. 17, 2012, 2:41 PM), http://threatpost.com/oracle-leaves-fix-java-se-zero-day-until-february-patch-update-101712/. Some vendors do issue patches considerably more rapidly; it is unclear, though, that this is always a good idea. Rapid patches often block a particular path to reach the underlying buggy code rather than repairing it. Accordingly, attackers often find new variants of the exploit without much trouble. Sometimes patches contain their own flaws. Thus, there is likely an irreducible average minimum time.

[242] Zero-day vulnerabilities average a 10-month lifespan. Leyla Bilge & Tudor Dumitras, *Before we Knew It: An Empirical Study of Zero-day Attacks in The Real World*, PROC. 2012 ACM CONF. ON COMPUTER & COMM. SECURITY 833, 834 (2012).

[243] There is a paucity of peer-reviewed research results on how soon individual users apply patches. The best studies are old and apply to enterprise servers, not individual users. *See, e.g.*, Eric Rescorla, *Security Holes... Who Cares*?, PROC. 12TH USENIX SECURITY SYMP. 75, 75 (2003); CHESWICK, BELLOVIN & RUBIN, *supra* note 151, at 74–75. Enterprises have their own needs and dynamics for patching, such as concerns about compatibility with critical local software; furthermore, all system administration is generally under the control of a centralized support group. Most wiretaps are of individuals, especially drug dealers. *See* ADMIN. OFFICE OF THE U.S. COURTS, *supra* note 53. Therefore, their behavior is likely very different. There have been a number of statements by industry consistent with our assertion. *See, e.g.*, Press Release, Skype, Survey Finds Nearly Half of Consumers Fail to Upgrade Software Regularly and One Quarter of Consumers Don't Know Why to Update Software (July 23, 2012), *available at* http://about.skype.com/press/2012/07/survey_finds_nearly_half_fail_to_upgrade.html. A recent study is useful, since it measures actual exposure of real-world web browsers. *How are Java Attacks Getting Through?*, WEBSENSE (Mar. 25, 2013, 9:01 PM), http://community.websense.com/blogs/securitylabs/archive/2013/03/25/how-are-java-attacks-getting-through.aspx. Only about 5% of users had up-to-date Java versions, despite warnings of ongoing attacks. *Id.* The best evidence, though, is empirical: the prevalence of attacks against holes for which patches are available suggests that attackers still find them useful.

¶165 It is interesting to ponder whether the policy of immediately reporting vulnerabilities could disrupt the zero-day industry. Some members of the industry, such as HP DVLabs, "will responsibly and promptly notify the appropriate product vendor of a security flaw with their product(s) or service(s)."[244] Others, such as VUPEN, which "reports all discovered vulnerabilities to the affected vendors under contract with VUPEN,"245 do not. Although it would be a great benefit to security if the inability to sell to law enforcement caused the sellers to actually change their course of action, U.S. law enforcement is unlikely to have a major impact on the zero-day market since it is an international market dominated by national security organizations.

C. A Default Obligation to Report

¶166 The tension between exploitation and reporting can be resolved if the government follows *both* paths, actively reporting and working to fix even those vulnerabilities that it uses to support wiretaps. As we noted, the reporting of vulnerabilities (to vendors and/or to the public) does not preclude exploiting them.[246] Once a vulnerability is reported, there is always a lead time before a "patch" can be engineered, and a further lead time before this patch is deployed to and installed by future wiretap targets. Because there is an effectively infinite supply of vulnerabilities in software platforms,[247] provided new vulnerabilities are found at a rate that exceeds the rate at which they are repaired, reporting vulnerabilities need not compromise the government's ability to conduct exploits. By always reporting, the government investigative mission is not placed in

---

[244] *See Disclosure Policy*, ZERO DAY INITIATIVE, http://www.zerodayinitiative.com/advisories/disclosure_policy/ (last visited Mar. 1, 2013). It goes on to say:

> The first attempt at contact will be through any appropriate contacts or formal mechanisms listed on the vendor Web site, or by sending an e-mail to security@, support@, info@, and secure@company.com with the pertinent information about the vulnerability. Simultaneous with the vendor being notified, DVLabs may distribute vulnerability protection filters to its customers' IPS devices through the Digital Vaccine service.
>
> If a vendor fails to acknowledge DVLabs initial notification within five business days, DVLabs will initiate a second formal contact by a direct telephone call to a representative for that vendor. If a vendor fails to respond after an additional five business days following the second notification, DVLabs may rely on an intermediary to try to establish contact with the vendor. If DVLabs exhausts all reasonable means in order to contact a vendor, then DVLabs may issue a public advisory disclosing its findings fifteen business days after the initial contact.

Id.

[245] *Vupen Security Research Team – Discovered Vulnerabilities in Prominent Software*, VUPEN SECURITY, http://www.vupen.com/english/research-vuln.php (last viewed Mar. 1, 2013) (emphasis added).

[246] The question of publicly disclosing vulnerabilities is at the core of a very involved debate. The two basic positions are "responsible disclosure", i.e., only to the vendor for a reasonable period (typically a few months) or "full disclosure". Without going into details, the argument for full disclosure is threefold: first, it has often been necessary to force the vendor to act; second, people have a right to know what risks they're being exposed to (think of food labeling laws and many other product disclaimers); three, it lets individuals and companies act to protect themselves until a vendor fix is available.

[247] *See* BROOKS, *supra* note 116.

conflict with its crime prevention mission. In fact, such a policy has the almost paradoxical affect that the more active the law enforcement exploitation activity becomes, the more zero-day vulnerabilities are reported to and repaired by vendors.

¶167 However, this does not mean that a law enforcement exploitation laboratory will be naturally inclined to report the fruits of its labor to vendors. From the perspective of an organization charged with developing exploits, reporting might seem an anathema to the mission, since it means that the tools it develops will become obsolete more quickly. Discovering and developing exploits costs money, and an activity that requires more output would need a larger budget.[248]

¶168 An obligation mandating that law enforcement agencies report any zero-day vulnerabilities they intend to exploit should thus be supported by a strong legal framework. Such a framework should create bright lines for what constitutes a vulnerability that must be reported, when the reporting must occur, to whom the report should be made, and which parts of the government are required to do the reporting. There are many grey areas.

¶169 First, what should constitute a reportable vulnerability? Sometimes, this will be obvious. For example, some software bugs, such as input validation errors, might allow an attacker to take control over a piece of software.[249] Such behavior is clearly an error. Once reported, the software vendor can easily repair the software to eliminate the vulnerability and "push" the patch out.[250] Other vulnerabilities are less clearly the result of specific bugs, however. Sometimes, a vulnerability results from overly powerful software features that are behaving perfectly correct as far as the software specification is concerned, but that allow an attacker to exploit them in unanticipated ways. For example, many email systems allow software to be sent as an "attachment" that is executed on the recipient's computer when the user clicks on it. If an attacker emails a user malware and the user is persuaded, however unwisely, to open it, the user's computer becomes compromised. Although it served as a vector for the malware, the email system software, strictly speaking, has behaved correctly here. The line between a "bug" and a "feature" is often quite thin.

¶170 Then there is the question of when a potential vulnerability that has been discovered becomes "reportable." Many vulnerabilities result from subtle interactions in a particular implementation,[251] and not every software bug results in an actual exploitable vulnerability. If the government is obligated to report exploitable vulnerabilities, when must it do so? An appropriate guideline would be that once the government has

---

[248] It is difficult to estimate precisely the cost of developing a particular vulnerability, but existing markets can serve as a guide here, as discussed in Section IV.

[249] *See, e.g.*, *supra* note **Error! Bookmark not defined.**.

[250] Many companies, if not most, provide automatic security updates that are simply updated via the Internet.

[251] Quite some time ago, one of the authors of this paper discovered that someone working on an important project was one of three people arrested in a hacking incident. (He eventually pled no contest. One of the other two was convicted; the third was acquitted.) An audit of the code base was performed. The team found one clear security hole, but log files showed it was an inadvertent hole coded, ironically, by one of the other auditors. There were also two independent bugs, and the comments in the code for one of the bugs did not agree with the code. Either bug alone was harmless; together, combined with a common configuration mistake, they added up to a remote exploit. There was a plausible innocent explanation for why the comments and the code did not match. It remains unclear if this was a deliberate back door or a coincidence.

developed an exploit tool, the underlying vulnerability has been confirmed to be exploitable and should promptly be reported. Note that this way of implementing the always report policy gives law enforcement investigators some lead-time in using the exploit tool. This approach provides appropriate leeway for law enforcement to do its job by exploiting these vulnerabilities, while not making them quality assurance testers for software companies.

¶171 To whom should a vulnerability report be made? In many cases, there is an obvious point of contact: a software vendor that sells and maintains the product in question, or, in the case of open-source software, the community team maintaining it. In other cases, however, the answer is less clear. Not all software is actively maintained; there may be "orphan" software without an active vendor or owner to report to.[252] Also, not all vulnerabilities result from bugs in specific software products. For example, standard communications protocols are occasionally found to have vulnerabilities,[253] and a given protocol may be used in many different products and systems. In this situation, the vulnerability would need to be reported not to a particular vendor, but to the standards body responsible for the protocol. Many standards bodies operate entirely in the open,[254] however, which can make quietly reporting a vulnerability—or hiding the fact that it has been reported by a law enforcement agency—problematic. In this situation, the choice is simple: report it openly.

¶172 Finally, there is the question of who in the government should be covered by guidelines mandating reporting. In this paper, we are concerned specifically with a law enforcement vulnerability lab. Should every U.S. government employee be included in the guidelines? Or only those developing law enforcement surveillance tools? The vast majority of government employees—even those who encounter security vulnerabilities—are not directly involved in developing wiretapping tools. For example, there are presumably system administrators in the Veterans Administration who occasionally discover security vulnerabilities in the course of their work. Should they become legally obligated to report? We propose that the reporting obligation be linked to the use of vulnerabilities for law enforcement purposes. An ordinary system administrator who discovers a vulnerability perhaps should report it, but the legal requirement should apply only to those who employ such vulnerabilities to conduct communications intercepts.

VII. EXECUTIVE AND LEGISLATIVE ENFORCEMENT

¶173 When should reporting occur—at the time of discovery or purchase of the vulnerability, or at the time of working exploit? Should there be exceptions to the

[252] Every software system has a date beyond which there will be no further patches. Microsoft, for example, lists its support plans at http://windows.microsoft.com/en-us/windows/products/lifecycle.

[253] For example, several vulnerabilities have been found that allow attacks against systems using the Secure Socket Layer (SSL) protocol, a widely used standard employed by many applications, including Web browsers, printers, and email clients, for encrypting Internet connections. *See, e.g.,* Dan Goodin, *Hackers Break SSL Encryption used by Millions of Sites*, THE REGISTER (Sept. 19, 2011), http://www.theregister.co.uk/2011/09/19/beast_exploits_paypal_ssl/.

[254] For example, all Internet Engineering Task Force (IETF) meetings and mailing lists are open to the public. See the IETF website at www.ietf.org, and in particular *The Tao of IETF: A Novice's Guide to the Internet Engineering Task Force*, IETF § 4 (Nov. 2, 2012), https://www.ietf.org/tao.html.

reporting rule in the case of an extremely important target, and how should that work? In this section, we attempt to answer these questions as well as discuss the role of oversight.

A. Enforcing Reporting

¶174 We advocate that vulnerabilities law enforcement seeks to exploit be reported by default. There are a number of ways to implement and enforce such a policy.

¶175 The simplest way to implement a default reporting policy would be guidelines that mandate reporting under certain circumstances promulgated by the administration, likely the Department of Justice.[255] However, a guidelines-only approach has inherent weaknesses. First, the guidelines would be formulated, implemented, and enforced by the very department with the most interest in creating exceptions to the rule, and that most "pays the cost" when the tools it develops and uses are neutralized. Such conflicts of interest rarely end up with the strongest possible protections for the public.

¶176 Therefore, a legislative approach may be more appropriate. Perhaps as part of the appropriations bill that funds the exploit discovery effort, Congress could mandate that any vulnerabilities the unit discovers be reported; alternatively, a reporting mandate could be added to the wiretap statute. This second approach has the advantage that it is more permanent; however, amending the Wiretap Act has proven to be a long and contentious process. Regardless, and as noted above, such legislation would need to be carefully drafted to capture a range of different circumstances.

¶177 In the absence of a legislative fix, the best solution is for the judge authorizing the use of the vulnerability to insert a reporting requirement into the warrant or order. This provision could include a return date by which the requesting agency must certify that the vendor had received appropriate notification. Apart from providing an enforcement mechanism, this approach allows for careful consideration of specific circumstances, including exceptional circumstances that might merit a delay.[256]

¶178 Finally, one might imagine that the legislature could create a tort cause of action for those harmed by a criminal exploitation of a vulnerability known to the government but not reported. This would perhaps be the most radical approach to ensuring government reporting, but it seems most unlikely. There is currently no obligation on anyone to report vulnerabilities; for Congress to suddenly create government liability for non-reporting seems improbable.[257] Our favored approach to ensure early government reporting of vulnerabilities discovered is thus a simple but unambiguous legislative mandate that the government report any zero-day vulnerabilities it seeks to exploit. We take no position here on financial liability or other remedies should it fail to do so.[258]

---

[255] For example, the reporting requirement could be added to THE ATTORNEY GENERAL'S GUIDELINES FOR DOMESTIC FBI OPERATIONS (2008), *available at* http://www.justice.gov/ag/readingroom/guidelines.pdf.

[256] Exceptional circumstances are discussed in the following subsection.

[257] Due in part to disclaimers in End User License Agreements (EULAs), there is in general no liability even for vendors or developers of insecure software. *See, e.g.*, Michael D. Scott, *Tort Liability for Vendors of Insecure Software: Has the Time Finally Come?*, 67 MD. L. REV. 425 (2008). However, the issue is a frequent topic of academic discussion and the situation could conceivably change. In some situations, a site operator can be held negligent. *See, e.g.*, *In re Heartland Payment Systems*, 851 F. Supp. 2d 1040, 1047–48 (S.D. Tex. 2012).

[258] We do not discuss or suggest remedies if the government fails to report vulnerabilities, as is urged in this paper. A radical legislative approach could be to permit damages for those harmed by the exploitation

B. Exceptions to the Reporting Rule

¶179 Although we have recommended that law enforcement report vulnerabilities upon discovery (or purchase), there may be exceptional cases when immediate reporting is not appropriate because immediate reporting of the vulnerability might lead to a target patching and preventing installation of a wiretap. In what circumstances should not reporting immediately be appropriate?

¶180 It is worth considering the principles employed in the closely related situation of emergency wiretaps. Title III includes an exception allowing wiretaps to be used without a warrant in emergency situations as long as a wiretap order is obtained within forty-eight hours.[259] The law states that an emergency situation exists when there is immediate danger of death or serious bodily injury, conspiratorial activities threatening national security, or conspiratorial activities characteristic of organized crime,[260] but practice is that warrantless wiretapping by law enforcement[261] is permitted only when there is an immediate threat to life such as kidnapping and hostage-taking situations.[262] Emergency wiretapping is not done lightly, and requires approval of someone of no rank lower than an Associate Attorney General. Once the emergency wiretap is approved (approved, not installed) law enforcement has forty-eight hours to obtain a wiretap order.[263]

¶181 Assume a situation in which, using a wiretap warrant, law enforcement downloads software to the target's machine and finds that the target is running an unusual set of programs, e.g., using the OpenBSD operating system with the Lynx web browser.[264] Law enforcement lacks suitable tools for this particular setup. To exercise the actual wiretap, law enforcement must find a vulnerability and operationalize it. Experience (with, e.g., the iPhone jailbreak efforts[265]) suggests that in most cases, this will not take too long. If the vulnerability is immediately reported as soon as it is acquired, law enforcement runs the risk that the target's device may be patched before the operationalized exploit can be used.

---

of a zero-day vulnerability that was known to the government but that the government had not reported. A more moderate approach could impose a reporting obligation on the government but disallow private recovery of damages if it fails to do so.

[259] 18 U.S.C. § 2518(7) (2006).

[260] *Id.*

[261] Note that we are discussing warrantless wiretaps for criminal investigations under Title III, not the legalities of the Bush administration's "terrorist surveillance" warrantless wiretapping program. S*ee, e.g.,* Barton Gellman, Dafna Linzer & Carol D. Leonnig, *Surveillance Net Yields Few Suspects*, WASH. POST, Feb. 5, 2006, http://www.washingtonpost.com/wp-dyn/content/article/2006/02/04/AR2006020401373.html.

[262] For a detailed discussion, see *9-7.112: Emergency Interception*, U.S . ATT'YS MANUAL, http://www.justice.gov/usao/eousa/foia_reading_room/usam/title9/7mcrm.htm#9-7.112 (last updated July 2012).

[263] 18 U.S.C. § 2518(7) (2006).

[264] OpenBSD is an open-source operating system based on Unix (available at http://www.openbsd.org/) and Lynx is a web browser (available at http://lynx.isc.org/). Because Lynx does not support graphics, it cannot have web bugs, embedded objects that track usage, making it particularly privacy protective. Both systems, which are relatively old by industry standards, continue to be developed, but neither has large market share.

[265] The best compendium of information on the history of iPhone jailbreaking is a Wikipedia page, *iOS Jailbreaking*, WIKIPEDIA, https://en.wikipedia.org/w/index.php?title=IOS_jailbreaking&oldid=589152900 (last modified Jan. 4, 2014).

¶182 As far as we know, the FBI has never reported any of the vulnerabilities used to plant CIPAV. There is thus apparently no legal requirement that currently requires law enforcement to report vulnerabilities, so we recommend a compromise. For public safety, the law should require that law enforcement report vulnerabilities to the vendor once they have been acquired or otherwise discovered, but there should also be an emergency exception similar to that of Title III. We recommend that in an emergency situation, law enforcement should have a forty-eight hour window past the usual reporting deadline in which to petition a court for a release from reporting the vulnerability until it has successfully installed a wiretap.

¶183 We expect that such a provision would rarely be invoked. First, most vulnerabilities will have been discovered and reported by law enforcement, and the tools that exploit them built and put in the arsenal for future use, well before there is any investigation that might use them. For such tools, there is no emergency—or even any investigation—to weigh against reporting at the time the vulnerability would be reported because any situations in which a vulnerability is used would come up long after the vulnerability has already been reported.

¶184 But there may be exceptional circumstances in which this pattern—vulnerabilities discovered and tools developed well in advance of their being used by law enforcement—is not followed. For example, we can imagine a very high-value organized crime investigation in which a target might be using a particular and well-hardened, non-standard platform for which no exploit tools are available in the "standard" arsenal. Law enforcement might devote targeted resources toward discovering vulnerabilities and developing tools for the specific devices used by the particular target. In such (likely very rare) situations, the investigation and target might be known at the time some vulnerability is discovered by law enforcement, and they might place a high priority on preserving their ability to exploit it during the case.

¶185 The criteria for exemption must be as stringent as the Title III exemption. If emergency wiretaps are permitted only when there is imminent danger of death (e.g., a kidnapping or hostage-taking situation) then the situation for emergency use of a vulnerability without reporting must be equally dire.

¶186 Another issue with emergency use is that the vulnerability must be such that there is a low risk of serious harm resulting from its exploitation by others against innocent persons. As we have discussed, estimating such risk is quite difficult. Given the importance of preventing crime, the decision not to report must not be made lightly. The petition not to report must include not only an argument for the importance of the interception, but also an analysis of the harm that could be caused should the vulnerability be discovered and exploited by others during the period that law enforcement is operationalizing the tool. In weighing whether to delay reporting a vulnerability, the court should consider how likely it is that the vulnerability, having been discovered, can actually be exploited, and the damage that may result from such exploitation.

C. Providing Oversight

¶187 There is potential danger that an operationalized exploit may proliferate past its intended target. Stuxnet[266] provides an interesting case in point. Although aimed at Iran, the malware spread to computers in other countries, including India and Indonesia.[267] It is unclear from the public record how this happened. It may have been due to a flaw in the code, as Sanger contends;[268] alternatively, it may have been foreseeable but unavoidable collateral damage from the means chosen to launch the attack against Iran. Either possibility, though, represents a process that may be acceptable for a military or intelligence operation but is unacceptable for law enforcement. Only the legally authorized target should be put at risk from the malware used.

¶188 Given the public policy issues raised by the use of vulnerabilities, it would be appropriate to have public accountability on the use of this technique. For example, annual reports on vulnerability use similar to the AO's Wiretap Reports, presenting such data as: How many vulnerabilities were used by law enforcement in a given year? Were they used by federal or state and local? Was the vulnerability subsequently patched by the vendor, and how quickly after being reported? Was the vulnerability used by anyone outside of law enforcement? Was the vulnerability exploited outside law enforcement during the period that law enforcement was aware of the problem but had not yet told the vendor? Did the operationalized vulnerability spread past its intended target? What damages occurred from its exploitation? Making such information open to public analysis should aid in decisions about the right balance between efficacy and public safety.[269]

D. Regulating Vulnerabilities and Exploitation Tools

¶189 As we have mentioned, even without considering its use by law enforcement, information about software vulnerabilities is inherently "dual use"—useful for both offense and defense. Related to the issue of reporting and proliferation is the question of how the law should treat information about vulnerabilities and the development of software tools that exploit them by non-law enforcement persons. Should information about vulnerabilities, and tools that exploit them, be restricted by law? How do existing statutes treat such information and tools?

¶190 The issue of how to handle such dual-use technologies is not new. The computer security community has grappled for years with the problem of discouraging illicit exploitation of newly discovered vulnerabilities by criminals while at the same time

---

[266] *See Stuxnet Dossier, supra* note 17.

[267] DAVID E. SANGER, CONFRONT AND CONCEAL: OBAMA'S SECRET WARS AND SURPRISING USE OF AMERICAN POWER 203–05 (2013).

[268] *Id*. Sanger's conclusion is somewhat controversial. *See* Steven Cherry, *Stuxnet: Leaks or Lies?*, IEEE SPECTRUM (Sept. 4, 2012), http://spectrum.ieee.org/podcast/computing/embedded-systems/stuxnet-leaks-or-lies.

[269] The same is true regarding data from the Administrative Office of the U.S. Courts' Wiretap Reports (available at http://www.uscourts.gov/Statistics/WiretapReports/WiretapReports_Archive.aspx). For example, one of the authors of the present paper used Wiretap Report data to show that FBI claims about the importance of wiretaps to solve kidnappings was incorrect. Between 1969 and 1994 wiretaps were used in only two to three kidnappings a year (out of 450 kidnappings annually). DIFFIE & LANDAU, *supra* note 21, at 211.

allowing legitimate users and researchers to learn about the latest threats, in part to develop effective defenses.[270] It is all but impossible to prevent information about vulnerabilities or software exploits that use them from getting in to the hands of criminals without hampering efforts at defense. On the one hand, information about zero-day vulnerabilities is coveted by criminals who seek unauthorized and illicit access to the computers of others. But the same zero-day information is also used, and sought out by, legitimate security researchers and computer scientists who are engaged in building defenses against attack and in analyzing the security of new and existing systems and software.

¶191 Even software tools that exploit vulnerabilities are inherently dual use. They can be used by criminals on the one hand, but are also useful to defenders and researchers. For example, computer and network system administrators routinely use tools that attempt to exploit vulnerabilities to test the security of their own systems and to verify that their defenses are effective. Researchers who discover new security vulnerabilities or attack methods often develop "proof of concept" attack software to test and demonstrate the methods they are studying. It is not unusual for software that demonstrates a new attack method to be published and otherwise made freely available by academics and other researchers. Such software is quite mainstream in the computer science research community.[271]

¶192 The software used by malicious, criminal attackers to exploit vulnerabilities can thus be very difficult to meaningfully distinguish from mainstream, legitimate security research and testing tools. It is a matter of context and intent rather than attack capabilities per se, and current law appears to reflect this.

¶193 Current wiretap law does not generally regulate inherently dual-use technology. The provision of Title III concerned with wiretapping equipment, 18 USC § 2512, generally prohibits possession and trafficking in devices that are "primarily useful" for "surreptitious interception" of communications, [272] which does not appear to apply to a

---

[270] The question of the ethics of publishing vulnerability information far antedates computers. In 1857, Alfred Hobbs, in *Rudimentary Treatise on the Construction of Door Locks,* wrote,"A commercial, and in some respects a social, doubt has been started within the last year or two, whether or not it is right to discuss so openly the security or insecurity of locks. Many well-meaning persons suppose that the discussion respecting the means for baffling the supposed safety of locks offers a premium for dishonesty, by showing others how to be dishonest. This is a fallacy. Rogues are very keen in their profession, and already know much more than we can teach them respecting their several kinds of roguery."

[271] Many security software packages that might appear to be criminal attack tools are actually designed for legitimate research and testing. For example, the *Metasploit* package (available at http://metasploit.com) is a regularly updated library of software that attempts to exploit known vulnerabilities in various operating systems and applications. Although it may appear at first glance to be aimed at criminals, it is actually intended for (and widely used by) system administrators and professional "penetration testers" to identify weaknesses that should be repaired in their systems.

[272] 18 USC § 2512(1) (2006) provides criminal penalties for any person not otherwise authorized who:

> (a) sends through the mail, or sends or carries in interstate or foreign commerce, any electronic, mechanical, or other device, knowing or having reason to know that the design of such device renders it primarily useful for the purpose of the surreptitious interception of wire, oral, or electronic communications;
>
> (b) manufactures, assembles, possesses, or sells any electronic, mechanical, or other device, knowing or having reason to know that the design of such device renders it primarily useful for the purpose of the surreptitious interception of wire, oral, or electronic communications, and that such device or any component thereof has been or will be sent through the mail or transported in interstate or foreign commerce; or

wide range of current software exploit tools developed and used by researchers. We believe this is as it should be. The security research community depends on the open availability of software tools that can test and analyze software vulnerabilities. Prohibiting such software generally would have a deleterious effect on progress in understanding how to build more secure systems, and on the ability for users to determine whether their systems are vulnerable to known attacks. In addition, we note that given that the majority of vulnerability markets are outside the U.S., and that national security agencies are heavy purchasers of these vulnerabilities,[273] regulating them is not a plausible option.

¶194 The specialized tools developed by law enforcement to collect and exfiltrate evidence from targets' computers, however, might fall more comfortably under the scope of 18 U.S.C. § 2512 (2006) as it is currently written. These tools would not be developed to aid research or test systems, but rather to accomplish a law enforcement interception goal. They would have narrowly focused features designed to make their installation surreptitious and their ongoing operation difficult to detect. They would also have features designed to identify and collect specific data, and would have no alternative use outside the surreptitious interception application for which they were developed. Such tools, unlike those used by researchers, could more easily meet section 2512's test of being "primarily useful" for "surreptitious interception," and thus would be unlawful if someone "manufactures, assembles, possesses, or sells" them except under the circumstances spelled out in that section.

VIII. CONCLUSION

¶195 Changes in telecommunications technologies led to the 1994 passage of CALEA. However, CALEA created problems because of software complexity and the fact that it introduces a security vulnerability. Due to further—and quite extraordinary—changes in the communications technologies since CALEA's passage, the law enforcement wiretapping capabilities the law engendered are now in danger of failing; to prevent this, law enforcement now seeks to expand the CALEA regime to IP-based communications. As we have discussed, the changes in communications technologies since 1994 not only undermine the present version of CALEA, they make extending the CALEA model to modern communications systems highly problematic, creating serious security risks.

(c) places in any newspaper, magazine, handbill, or other publication or disseminates by electronic means any advertisement of—

(i) any electronic, mechanical, or other device knowing the content of the advertisement and knowing or having reason to know that the design of such device renders it primarily useful for the purpose of the surreptitious interception of wire, oral, or electronic communications; or

(ii) any other electronic, mechanical, or other device, where such advertisement promotes the use of such device for the purpose of the surreptitious interception of wire, oral, or electronic communications, knowing the content of the advertisement and knowing or having reason to know that such advertisement will be sent through the mail or transported in interstate or foreign commerce . . . .

[273] Greenberg, *supra* note 185.

¶196 Nonetheless, there needs to be a way for law enforcement to execute authorized wiretaps. The solution is remarkably simple. Instead of introducing *new* vulnerabilities to communications networks and applications, law enforcement should use vulnerabilities already present in the target's communications device to wiretap in the situations where wiretapping is difficult to achieve by other means.

¶197 The exploitation of existing vulnerabilities to accomplish legally authorized wiretapping creates uncomfortable issues. Yet we believe the *technique is preferable for conducting wiretaps against targets when compared to other possible methods of wiretapping, like deliberately building vulnerabilities into the network or device, would result in less security*.

¶198 We propose specific policies to limit the potential damage of using existing vulnerabilities. First, we recommend that in order to prevent rediscovery of the vulnerability and hence proliferation of the exploit, technical defenses should be implemented. Second, we recommend that, with rare exceptions, *law enforcement should report vulnerabilities on discovery or purchase.* This means our proposal may actually have the benefit of *increasing* security generally. Finally, because the exploit may allow far greater penetrations of the target device than would be permitted by a mere wiretap, we urge guidelines to ensure that law enforcement bar use of any other information found on the computer during the exploit (unless permitted by an additional warrant).

¶199 There is a critical difference in the societal dangers entailed in the use of targeted vulnerabilities compared with the installation of global wiretapping capabilities in the infrastructure. If abused, targeted vulnerability exploitation, like wiretapping in general, has the potential to do serious harm to those subjected to it. But it is significantly more difficult—more labor intensive, more expensive, and more logistically complex—to conduct targeted exploitation operations against all members of a large population. In other words, although vulnerability exploitation is very likely to be effective against any given target, it is difficult to abuse at large scale or in an automated fashion against *everyone*. Thus our solution provides better security than extending the model of CALEA to IP-based communications would.

¶200 Vulnerability exploitation has more than a whiff of dirty play about it; who wants law enforcement to be developing and using malware to break into users' machines? We agree that this proposal is disturbing. But as long as wiretaps remain an authorized investigatory tool, law enforcement will press for ways to accomplish electronic surveillance even in the face of communications technologies that make it very difficult. We are at a crossroads where the choices are to reduce everyone's security or to enable law enforcement to do its job through a method that appears questionable but that does not actually make us less secure. In this debate, our proposal provides a clear win for both innovation and security.