© IEEE Security & Privacy 11:1, Jan/Feb 2013, http://dx.doi.org/10.1109/MSP.2012.138

Going Bright: Wiretapping without Weakening Communications Infrastructure

Steven M. Bellovin | Columbia University
Matt Blaze and Sandy Clark | University of Pennsylvania
Susan Landau | Privacy Ink

Mobile IP-based communications and changes in technologies have been a subject of concern for law enforcement, which seeks to extend current wiretap design requirements for digital voice networks. Such an extension would create considerable security risks as well as seriously harm innovation. Exploitation of naturally occurring bugs in the platforms being used by targets may be a better alternative.

For law enforcement wiretaps, this is the best of times and the worst of times. Tracking suspects through transactional data vastly simplifies investigators' efforts. Yet accessing communications content through traditional means could be getting harder. Because of peer-to-peer communication methods, encryption, and service providers located outside the US, law enforcement says its ability to execute legally authorized wiretaps is becoming increasingly problematic. The US Federal Bureau of Investigation (FBI) claims its wiretapping capability is "going dark" (http://judiciary.house.gov/hearings/hear_02172011.html).

Law enforcement's preferred solution? Since 2010, the FBI has advocated expanding the scope of the Communications Assistance for Law Enforcement Act (CALEA), a 1994 law that requires that switches in digital telephone networks be built wiretap enabled. The FBI wants to extend such requirements to IP-based communications.

CALEA and the Internet

CALEA was controversial because it introduced new security risks into the voice telephone network; indeed, there have been several publicly known cases of telephone switches being compromised through their wiretap interfaces. This article is primarily focused on the issues associated with CALEA if it were to be extended to emerging Internet-based services.

There are several possible policy options for wiretapping as these trends continue. These include maintaining the status quo, which would increasingly limit content wiretaps to (decreasingly relevant) switched telephone networks.

Law enforcement could increasingly rely on (non-content) communications records, which can reveal a great deal of information about a target's location, contacts, movement, and so on.[1] The legal and privacy implications of widespread use of communications records by law enforcement are a matter of some controversy, however, and at scale, it's difficult to ensure that information about innocent third parties won't find its way into law enforcement databases along with the records of suspects.

But there is yet another possibility. As the CALEA approach has become less viable (and more dangerous to emerging infrastructure), targeted interception approaches—ones that don't entail the risks and costs of nationally mandated wiretap interfaces—have become increasingly practical. One approach is to leverage the fact that targets' communications devices in modern networks are virtually always built on complex software platforms. Continuing technical access to authorized wiretaps can be achieved—without expanding CALEA—by exploiting naturally occurring weaknesses in subjects' devices, enabling law enforcement to install surreptitious interception software at a target

 Copublished by the IEEE Computer and Reliability Societies 1540-7993/13/$31.00 © 2013 IEEE

endpoint as required. Many such weaknesses are *0-day vulnerabilities*, ones that might be completely unknown to others and for which no vendor fix exists. (Conceptually, the bug is discovered on day zero and reported and patched sometime later.)

Communication devices in modern networks are essentially always built on complex software platforms. Due to the inexact nature of software development, all complex programs contain inadvertent vulnerabilities. Without requiring any explicit wiretap support in the network or any compromise of nontargeted devices, law enforcement can exploit software vulnerabilities on end devices to facilitate interception. The US law enforcement community can fund a laboratory to develop targeted interception tools that take advantage of such vulnerabilities, an idea proposed in the 1996 National Research Council report on cryptography.[2] (Note, however, that the FBI has a role in crime prevention, but it isn't tasked with securing communications or communications infrastructure.[3]) Such an approach isn't without its own policy concerns and risks, yet it's far more protective of national communications security and privacy than other proposed alternatives, including and especially CALEA-type design mandates.

Some work in this direction is already in progress by law enforcement. As has been reported elsewhere,[4,5] the FBI has established a Domestic Communications Assistance Center (DCAC) to tackle the technical side of the "going dark" problem. In 2012, the FBI requested US$15 million to fund this lab. We believe that approaches such as expanding DCAC's efforts—and not expanding CALEA's scope—are effectively the only path to facilitating legally authorized wiretapping that doesn't also undermine the security of the US communications infrastructure.

We conclude that

- any past success network-based interception schemes such as CALEA may have enjoyed in the telephony domain won't translate to similar success for Internet-based services;
- many emerging communications services are inherently interceptable by passive means;
- requiring additional centralized interception capabilities will be unnecessarily redundant and will introduce increasingly more serious security risks to infrastructure while being increasingly less effective in producing useful evidence for law enforcement;
- law enforcement development of a sufficiently broad range of targeted passive and endpoint-based interception tools to meet ongoing wiretap needs is technically and economically feasible;
- law enforcement's use of passive interception and targeted vulnerability exploitation tools creates fewer security risks for nontargets and critical infrastructure than do design mandates for wiretap interfaces; and
- moving forward, targeted exploitation solutions are likely to be the only viable approaches for providing law enforcement with reliable interception capabilities against modern platforms, even if wiretap interfaces in infrastructure were mandated.

In particular, it is critical for national security that communications software and systems be designed to be as secure as possible against attack. Deliberate backdoors—whether by way of CALEA or through hidden "lawful intercept" access features included by software vendors—inherently make systems more vulnerable; worse yet, all users, not just wiretap targets, suffer the increased exposure. However, the absence of explicit lawful intercept backdoors need not preclude law enforcement access when it's required, as we'll discuss later.

Note that our discussion is US focused: CALEA is a US law. However, the 1994 US solution of building wiretapping capabilities into switches was rapidly taken up in many other parts of the world under the generic name "lawful intercept." The security risks inherent in extending CALEA to the Internet are security risks facing any nation contemplating similar approaches to a CALEA-type regime for IP-based networks. Thus, while our context is local, our analysis is global in its applicability.

Wiretapping: The Present Situation

By requiring that communications providers include wiretapping capabilities within switching mechanisms, CALEA was a surprising development on the regulatory front.

For a quarter of a century, the process under which authorized wiretaps were done in the US was straightforward. Two laws, the 1968 Title III of the Omnibus Crime Control and Safe Streets Act (for criminal investigations) and the 1978 Foreign Intelligence Surveillance Act (FISA; for foreign intelligence cases), governed wiretap order applications. Once a judge granted an order, the wiretap could be installed.

The divestiture of AT&T meant that instead of a single monopoly handling both telephones and service, many more product and service providers emerged, along with increasing innovation in communication technologies. Law enforcement found itself thwarted in carrying out some legally authorized wiretaps. (Because we focus on possible extensions to CALEA and the harm they represent, we don't discuss the much richer surveillance capabilities now available to law enforcement—the plethora of communications, the fact that these frequently reveal location, and so forth—that provide a different situation than when Title III and FISA were passed.) Their solution was CALEA.

CALEA was implemented through an interface standard developed by the Telephone Industry Association in consultation with law enforcement.[6] This standard pleased no one: civil liberties groups wanted greater privacy protections than the standard provided, industry wanted greater clarity on the standard's technical requirements, and law enforcement wanted greater surveillance capabilities than were included. Several lawsuits and court rulings ensued, but by 2002, the requirements for CALEA compliance finally solidified. Although nobody was fully satisfied, CALEA became the dominant mechanism for implementing telephone wiretaps.

The FBI soon raised a new concern: Voice over IP (VoIP). IP-based communications are often peer to peer, and the CALEA model of tapping at the switch doesn't easily fit in with that. The Federal Communications Commission (FCC) and a federal appeals court cut this Gordian knot by deciding that CALEA would apply to facilities-based broadband, systems with wired lines (or wireless channels) to the end user. These communications systems are centralized, just like the public switched telephone network (PSTN), and applying CALEA-type solutions isn't especially technically difficult.

Innovation rarely pauses in technology. Because of a combination of increasing levels of peer-to-peer communications and encryption, along with such changes as overseas communications providers offering services in the US (creating difficulties when a wiretap is needed), law enforcement is again facing difficulties. What CALEA extensions the FBI actually seeks remain unclear. There have been various news reports since autumn 2010, but as of this writing, no bill has been produced.

CALEA worked for a reason that no one fully articulated: circuit-switched telephones (and cellphones) were the primary mode of communication. That era is now ending. Efforts to extend CALEA-type controls to the nearly infinite number of communications devices and applications cannot be effective.

The CALEA "Solution"

CALEA was intended to address a specific and rather unique set of technological circumstances brought about by incremental advances in voice telephone technology. Prior to CALEA, there was neither a mandate requiring telephone companies to design technology facilitating wiretapping nor a standard telephone wiretapping interface. Instead, wiretaps relied on local loops, the pairs of wires between a local telephone office and its subscribers. Law-enforcement agencies developed local-loop tapping technology, which it deployed by connecting to subscriber wire pairs. Sometimes law enforcement deployed taps with assistance from the carrier; sometimes it did it on its own.

> Tapping technology was simple and largely unchanging because telephone subscriber loop technology was, at least until the 1990s, simple and relatively unchanging. Carrying analog voice and signaling, the telephone local loop remained essentially the same for half a century. Tapping a telephone was a relatively simple matter of gaining physical access to the target's pair of wires and recording the electrical signals and voice audio the wires carried.

By the 1990s, two new subscriber loop technologies had emerged that weren't directly compatible with traditional analog wiretapping techniques. ISDN employed a pair of wires between the telephone central office and the subscriber but used digital signals and digitally encoded audio, which can require far more sophisticated technology for third-party interception. The other new technology was wireless cellular, in which the local loop was replaced with a two-way radio link, allowing the subscriber to move freely about the coverage area.

It's important to note that while ISDN and cellular services might have radically altered the local loop between the telephone company and the subscriber, these technologies did relatively little to alter telephony's centralized architecture. The basic service for both ISDN and cellular was and is voice calls linked to the PSTN. Subscribers still obtain their service from a single one of relatively few providers, which are themselves highly regulated by local franchises or hold federal licenses for part of the limited wireless spectrum.

Current Internet service architectures are far more complex than the telephone networks of the 1990s. Link technologies, including cable, fiber optics, DSL, and several forms of cellular wireless, were widely implemented. VoIP services, of which there are various varieties, have added a third local loop technology; it adds the challenge of separating the infrastructure provider from the physical plant provider, greatly complicating the wiretapping effort.[7]

Currently relatively few entities have had to comply with the current CALEA voice wiretap interface mandates. Those that do provide a common basic service:

voice calls. Compared with typical Internet infrastructure, switches for voice calls—the primary devices required to have CALEA interfaces—are very expensive, amortized over long periods, and relatively slow to change. This means that the costs, innovation burden, and security risks associated with implementing CALEA for voice telephony, while not trivial, are both somewhat calculable and relatively manageable. Yet even in the domain of voice telephony, CALEA is far from a win-win solution for wiretapping. Still, for IP-based communications, CALEA represents a lose-lose situation.

CALEA Insecurities

CALEA requires that a deliberate security weakness—the wiretap interface and control system—be architected into the switches of a communications network. In 2000, the Internet Engineering Task Force observed that, "Experience shows that if a vulnerability exists in a security system, it is likely that someone will take advantage of it sooner or later."[8] That situation has come to pass for CALEA-type interfaces.

The story of the 10-month interception of the most senior officials of the Greek government in 2004 to 2005 using a CALEA-type interface that had been surreptitiously turned on is well known.[9] Less well known is the 10-year wiretapping of 6,000 Italians that occurred through Telecom Italia,[10] the targets of which included political figures, judges, referees, and celebrities. The US has not been immune. Examinations by the National Security Agency (NSA) of CALEA-compliant switches to be sold to the Department of Defense found vulnerabilities in the CALEA implementation in every single switch examined.[11]

CALEA-like interfaces are, by definition, designed for surreptitious eavesdropping. They're intentional backdoors and thus both easier to exploit and more damaging when penetrated. The recent massive increase in cyberexploitation—theft of data from governments and companies around the world—adds to the concern about the vulnerabilities created through CALEA-type architectures. Furthermore, there are subtle but essential differences between the architecture of the PSTN and those of contemporary and emerging Internet-based services. These make generalized wiretap interfaces for Internet communications far more technically difficult, complex, and economically burdensome than they are in traditional telephony.

It's certainly true that unauthorized remote wiretaps can be implemented by using other forms of remote access including craft interfaces, which are used to test installations, check reported faults, and so on. But such remote accesses are much more difficult to conduct surreptitiously because, unlike CALEA, they're deliberately designed to be logged and to trigger other, semiautomatic changes within the system. In contrast, CALEA interfaces are specifically intended for surreptitious wiretapping. By design, indicia of such taps are carefully restricted and invisible outside the CALEA control console. They thus provide a more attractive attack surface for exploitation by criminals and foreign intelligence services.

It's far from clear that CALEA for telephony has successfully balanced law enforcement requirements for surveillance access with the broader goal of preventing illicit access to critical infrastructure by criminals and foreign governments. But even if we assume that CALEA for telephony has been on the whole a success, the conditions that might have made it so aren't present in the Internet services for which the government seeks to apply the same approach. Communications infrastructure lasts a long time. Given increasing cyberexploitation efforts, switch longevity makes the security concerns even more trenchant.

The CALEA Problem

Internet-based services have very different technical and economic properties from traditional telephony. These make the CALEA approach far less attractive for the Internet while simultaneously introducing considerably more risk.

There are several reasons that lawful intercept mechanisms in all communications software are an exceedingly bad idea. The most obvious is the risk: the more code in an application, the more likely it is to have bugs. By definition, lawful intercept code is an engineered—though nominally controlled—vulnerability; a flaw in it exposes the precise sort of access many attackers will want. Even if the intercept code does not itself offer vulnerabilities, its mere existence simplifies the attackers' efforts (witness what happened in Athens and Italy).

The Internet services the government seeks to tap are provided by a large number of entities operating on ordinary computers that are architecturally, effectively ordinary end points. In contrast, on the PSTN, telephony services are provided by large, centralized switching systems operated by a small number of carriers. This architectural difference underlies the vastly greater pace of technology development in Internet services compared with the telephone network.

On the Internet, any node with sufficient bandwidth can act as a service provider. This has led to innovations that a centralized, slow-moving company wouldn't do. The Web itself was invented at a physics laboratory, not by ISPs. Skype, which provides telephone-style voice service Internet connections, doesn't even use central servers; it's a distributed, peer-to-peer network.

To tap Internet applications in the manner of CALEA, then, requires wiretap interfaces in many

widely distributed nodes, rather than a few centralized ones. It imposes design constraints on a large number of service providers around the world, rather than a few domestic phone companies. Requiring lawful intercept interfaces in all Internet communications services and software is simply untenable.

Especially early in their life cycles, Internet-based services tend to be lightweight, inexpensive, rapidly changing, and far more reliant on general-purpose software platforms than are the slow-moving of traditional telephony. Thus the highly diverse and dynamic nature of Internet-based services makes the implementation of any kind of standardized wiretap interface considerably more architecturally disruptive than it has been in the PSTN's switched voice telephony environment. A wiretap interface would have to be integrated over a wide range of often quickly deployed and poorly debugged services and then reimplemented every time a new service is introduced or a software architecture changes. Th is would prove an expensive burden on the small start-ups that drive online innovation.

Worse, many new services, especially those that rely on a peer-to-peer architecture for routing content between users, simply can't be intercepted via the centralized CALEA model. For these services, no design mandate, short of outlawing the decentralized routing scheme on which the Internet is built, can reliably capture all the traffic law enforcement might seek to intercept. Mandating centralized wiretap capabilities in these services would not only be disruptive to innovation but would also fail to deliver meaningful benefit to law enforcement.

Finally, expanding the number of CALEA-like interfaces in the network would create great insecurity. The vulnerabilities in every CALEA-compliant switch tested by the NSA show how hard it is to get the interception technology correct. Those switches were designed by large service providers working over a relatively long period of time. The difficulty of debugging and testing software to make Internet services secure is a largely unsolved problem, especially at the pace of "Internet time"; requirements for wiretap interfaces would make securing new services significantly more difficult.

Wiretapping by Compromising the Target

Suppose that the FBI were to use vulnerability-based solutions for its targeted Internet intercepts. How would this be done? What is necessary to enable it to happen?

Modern computing and communications devices suffer from an essentially unlimited number of security vulnerabilities. Furthermore, as the widespread proliferation of botnets and other criminal exploitation tools demonstrates, it's easy to exploit these vulnerabilities and gain control over an unwitting user's entire platform and virtually impossible for end users to defend against such attacks. Law enforcement can (and, to a certain extent, already does) exploit this.

However, there are additional requirements for law enforcement exploitation tools beyond those employed by criminals who compromise computers to create botnets or steal private data. Cybercriminal tools generally focus on targets of opportunity, but law enforcement will have specific targets on which to focus. This will require specialized interception tools that work well above the "probabilistic" standard of typical criminal exploits. In particular, tools must have a very high chance of successfully compromising the target without risk of alerting the target. Furthermore, the compromise tool can't risk disrupting a target's computer environment—or anyone else's. Finally, investigators must be able to rapidly determine whether their tools have successfully compromised their target's hardware, must be able to manage it during the intercept period, and must be able to "clean up" once a wiretap has ended.

There are four primary components to any law enforcement tool that exploits target endpoint vulnerabilities: selection or discovery of an appropriate underlying vulnerability, installation mechanisms, mechanisms for obtaining access to the communications being targeted, and ways to send captured data back to the responsible investigators. All are situation-dependent. Developing usable, specialized attack tools to accomplish these tasks would be the core mission of a vulnerability exploitation lab. These tools would generally need to be developed and tested by the government well in advance of their use against any particular target.

Consider a hypothetical example: a wiretap target is using an encrypted communication system we'll call CommApp. If CommApp itself is known to have a remotely exploitable vulnerability (one in which carefully formed messages sent to it over the network can compromise the application), the government can use this directly to install its wiretapping code in the application. In this case, the government's compromise tool for CommApp would have to craft an attack message and deliver it over the network when CommApp is running.

If there are no exploitable vulnerabilities in CommApp itself, the problem becomes twofold: system penetration and application penetration. Both of these are classic problems well known to the security community. For the former, the FBI would have to exploit a vulnerability in any other application used by the target. There is a core of very complex—and hence inherently likely to be vulnerable—software used in most platforms; for computers, this core includes Web browsers, email applications, word processors, spreadsheets, PDF and photo viewing interfaces, and so on. An appropriate vulnerability would be one in which opening a specially crafted file with the vulnerable application allows attack code to be installed in the target's platform. Such vulnerabilities are very common in complex applications. A penetration, then, would involve arranging for the target's computer to open an attack file with the vulnerable application, either through automated means over the network or by subterfuge.[12]

In the (very) rare cases where no remote exploitation is possible, a "black bag job"—a legally authorized surreptitious physical break-in—might be performed to install the exploit code directly on the target's device. This has been done in the past.[13,14]

Once the system is penetrated by running the FBI's code, the exploitation must gain access to the intended communication. In our example, CommApp itself could be modified. The simplest modification would be one that leaked the cryptographic keys, but there are more complex modifications, such as capturing the plaintext voice, that would also work. An alternative approach would be to employ generic modules to capture microphone input, speaker output, and so on.

The central problem in our hypothetical example is surreptitiously exfiltrating the captured content back to the FBI. For content such as text messages, the volume of data is typically low enough that any excess traffic won't disturb a broadband connection. Voice is more difficult, especially on cellphones, which have relatively limited battery and transmission capacity. Sending the captured traffic at low speed over time can avoid a noticeable spike in traffic volume. Alternatively, the exploit code might disable encryption or weaken or leak the session encryption keys, allowing intercepted content to be captured in real time by conventional interception techniques without consuming extra bandwidth.

Maintaining an exploitation development capability involves four major ongoing tasks:

- developing and maintaining a library of penetrations techniques for major operating systems and applications;
- developing and maintaining input and output capture techniques;
- analyzing popular communication applications for specific bugs; and
- as required, developing custom exploits for specific platforms.

There will also be significant operational and legal tasks as well.

Using the tools developed to execute an intercept against a target requires three steps: analyzing the target's network usage to determine the platform and applications he or she is using, compromising the platform to deliver an appropriate exploit, and monitoring the captured messages from the exploit and target. Depending on the tools used, these steps may be augmented by conventional data wiretapping techniques.

Compromising the target's platform is practical because modern software systems are—and will continue to be—inherently vulnerable to attack. New exploitable vulnerabilities in widely used software are discovered at a steady rate, literally daily.

Another aspect of modern communication tools works in our favor. A vulnerability in a commonly used communication tool is likely to be effective against many targets, while lightly used communication tools are less likely to be robust (fewer users means less likelihood of discovering security flaws and, typically, fewer vendor resources to discover the vulnerabilities), and thus vulnerabilities in these will be easier and cheaper to discover. Of course, some targets will use communications systems for which penetration is very difficult or expensive under our proposed scheme, but the same situation is also true today.

Several databases track and attempt to catalog the various characteristics of newly discovered vulnerabilities. One of the most comprehensive is the Common Vulnerabilities Enumeration (CVE) database, which provides a weekly listing of newly published vulnerabilities ranked by severity. For the week of 9 July 2012, for example, it reported 45 newly disclosed vulnerabilities, of which 14 were ranked high severity and 31 medium severity. The CVE is an authoritative repository of publicly disclosed vulnerabilities, but is not always as up to date as other databases, such as Bugtraq. The Bugtraq database has the added feature that, if available, proof-of-concept exploit code is included along with vulnerability characteristics.

In addition, several private companies and individual researchers actively search for exploitable vulnerabilities, often selling them along with exploit code. Although there is an active black market for the sale of private, nondisclosed exploitable vulnerabilities,[17] several commercial firms, such as Vupen (www.vupen.com/english/services/solutions-gov.php), VulnerabilityLab (www.vulnerability-lab.com),

Table 1. Exploitable vulnerabilities discovered from March to mid-July 2012.

Month	Vul-Labs	Microsoft V.R.	Vupen	Bugtraq	ZDI
July	15	2	6	17	14
June	32	2	25	5	39
May	31	1	39	2	0
April	37	2	38	6	20
March	9	1	41	11	13

ZDI (http://dvlabs.tippingpoint.com/advisories/disclosure-policy), and Secunia (https://secunia.com/community/advisories), provide subscription services that make available varying levels of access information about 0-day vulnerabilities to their clients.

These groups discover and release a steady stream of new vulnerabilities in widely used software platforms. Table 1 lists the numbers of remotely exploitable vulnerabilities discovered each month from several of these commercial vulnerability research groups for the period of 1 March through mid-July 2012. (The fact that a new vulnerability is found is usually published immediately. Public disclosure of the details usually occurs a few weeks later, typically to Bugtraq [www.securityfocus.com/archive/1] and Full-disclosure [http://seclists.org/fulldisclosure].)

For law enforcement to rely on this rich supply of vulnerabilities to support its wiretapping needs, it must be economical to develop "law enforcement–grade" tools that exploit them. A rough estimate suggests that the costs of operating a law enforcement exploitation laboratory wouldn't be prohibitive, especially compared with the total costs of surveillance mandates in infrastructure. To create an exploitation tool, the government must first discover (or purchase) an exploitable vulnerability. A lab must then "weaponize" the vulnerability to reliably install wiretap code in the target platforms against which it is used. The tools would have to be extensively tested to ensure that that they don't do collateral damage to other parties.

Note that a federal vulnerability laboratory would likely have additional responsibilities beyond just discovering and developing exploits. Federal law enforcement would likely be in the best position to discover the simplest way to install legally authorized wiretaps; state and local law enforcement lack such depth of expertise. The costs of supporting state and local government intercepts (chiefly educational and consulting) will likely be borne by the federal government. However, these costs are relatively small compared with the actual exploitation development activities and can be estimated by the number of state and local wiretap investigations and investigators. (The National Technical Investigators Association includes essentially all investigators who participate in intercept work; it has 4,000 members. This provides a rough upper bound on the laboratory's teaching responsibilities.)

The bulk of the cost of developing law enforcement–grade wiretap tools against any particular platform is thus the cost of discovering an appropriate vulnerability plus the cost of building reliable systems for exploiting it. Both vulnerability discovery and exploitation tool development have evolved into commodities traded on commercial and underground markets, which allows us to approximately project the cost to law enforcement of conducting these activities. Several vulnerability exploitation products are marketed explicitly as surveillance tools for law enforcement and government.

An upper bound on the cost of vulnerability discovery can be estimated straightforwardly from currently existing markets that traffic in 0-day exploits. The government could either purchase "fresh" 0-day vulnerabilities from the market or discover them internally, as budget, resources, and policy permit.

The expected costs of developing these vulnerabilities into viable law enforcement wiretap tools are more difficult to estimate precisely but can be bounded as lying between the known costs of developing typical research and/or criminal exploit tools (at the low end) and the reported costs of developing elaborate national intelligence and "cyberwar" tools (at the high end). For the most part, law enforcement's needs are likely to lie close to the lower bound and should be comparable in sophistication to commercial penetration testing and criminal exploit tools. Commercial penetration testing products, such as Metasploit (www.rapid7.com/products/metasploit-pro.jsp) and Core Impact (www.coresecurity.com/content/core-impact-overview), give estimates for the low end of this cost spectrum. Note that the "payload" of such tools—the code that actually performs the content intercepts—although probably much larger and more complex than the vulnerability exploitation code, is likely to remain reasonably constant over time. Only the exploitation code itself would likely need to be updated or customized frequently.

Policy Concerns

Expanding the scale of law enforcement exploitation of target platform vulnerabilities naturally raises policy concerns. While our focus here is on technical issues, we briefly discuss the policy concerns raised by this approach. We anticipate a fuller treatment of these in policy and legal venues.

If law enforcement purchases vulnerabilities rather than discovering them in-house, a basic issue is whether government participation in the vulnerabilities market is appropriate. Law enforcement demand might help skew incentives against disclosing patches to the software vendors themselves, and some have argued that the process increases the amount of software left unpatched.[16,17] However, because the FBI's purchase can rarely be exclusive, it isn't clear its purchasing a vulnerability would actually change things. From repressive nation-states to well-funded criminal organizations, any number of bad actors are interested in, and capable of paying for, such vulnerabilities, and the market for 0-day vulnerabilities will exist regardless of law enforcement's participation in it. Because law enforcement's needs are likely to be at the lower end of the scale of commercial penetration testing and criminal exploits, the government's participation in the vulnerabilities market is unlikely to change pricing. These low-end vulnerabilities are priced accordingly and usually aren't available for exclusive purchase.

Once developed, an exploit tool will remain useful for law enforcement until the underlying vulnerability is discovered, disclosed, and patched in the target platforms. This period of viability can actually be expected to be quite long. A recent study of 0-day vulnerabilities exploited by malware found that the average time between initial use and public disclosure of a vulnerability was 312 days; it was only sometime later that a vulnerability ceased to be exploitable.[19]

An additional concern is whether law enforcement's participation in the 0-day market supports a shady business whose very existence is contrary to good public policy. This is, of course, the type of issue with which law enforcement often wrestles (a closely related example is that successful investigations often require the use of paid informers in criminal organizations). While law enforcement's participation doesn't create a market that wouldn't otherwise exist, it does have the potential to make these markets more active and robust, possibly increasing the availability of marketed exploits to criminals.

We emphasize that by no means do we suggest that software be deliberately made or left insecure in order to facilitate law enforcement exploits. Indeed, we firmly believe that those who find vulnerabilities should disclose details to the vendor so that they can be fixed as quickly as possible. That said, serious vulnerabilities do and almost certainly will continue to exist in virtually all platforms and applications of interest. We regret this, but the fact remains that exploitable vulnerabilities do exist. Taking advantage of them is far preferable to introducing new vulnerabilities into other applications or infrastructure, as the CALEA approach does.

A related issue arising from law enforcement use of unpublished vulnerabilities (whether discovered internally or purchased) is whether the government should be reporting exploitable vulnerabilities and having them fixed, rather than quietly exploiting them. This question is especially acute for vulnerabilities in common platforms. Perhaps the FBI should be sharing discovered weaknesses with software vendors so that they might patch them and prevent criminal exploitation. On the other hand, given the vast number of potential exploits that naturally occur, law enforcement's choice to use any given vulnerability rather than report it is arguably unlikely to have a major practical impact.

These are legitimate—and difficult—policy questions. We take no position here as to whether law enforcement should purchase 0-day vulnerability information from commercial markets or discover them through in-house research, nor precisely how it should weigh the "report or exploit" question. This is, however, an issue of relative risk; we note that even in the worst case, the overall harm done by law enforcement's discovery and use of vulnerabilities would be far smaller than the harm caused from weakening the infrastructure via wiretap mandates in software and systems. However, to ensure that conflicts between public disclosure and law enforcement silence are properly weighed, it would be appropriate to have technical and policy overseers examining these decisions as they're made.

One important issue is that discovering 0-day vulnerabilities and developing tools that exploit them gives law enforcement more technical capability than it has had in the past. The use of such tools to perform content wiretaps will, of course, require a wiretap order, and thus be legally controlled. However, it's also possible that law enforcement might wish to use these tools in other circumstances, for example, in accessing stored data. What rules should govern this? In 2011, the US Court of Appeals ruled that the court cannot require the plaintiff to reveal his or her encryption key because the state would have access to all the suspect's files and had not specified which ones were of interest.[19] Analyzing the right set of legal responses to this situation is out of scope for this article, but as "Time works changes [and] brings into existence new conditions and purposes,"[20] we note that the extensive use of vulnerabilities/0-day tools could raise new legal issues.

Any shift from carrier-based interception (such as

Practical Concerns

The use of vulnerabilities to enable legally authorized wiretaps raises questions for a variety of communities. For example, the policy community must examine under what circumstances law enforcement's participation in the vulnerabilities market is appropriate. If law enforcement becomes aware that a vulnerability it uses could create serious harm to multiple users or a critical infrastructure, what should its course of action be? What are the national security implications of law enforcement's participation in the vulnerabilities market? Technologists face the issue of "do no harm": an installed vulnerability shouldn't act against anyone but the target (and for the target, the action should be limited to wiretapping the target's communications and not causing other disruption on his or her device).

Another issue is that the vulnerability shouldn't "escape" the target's machine,[1] which might enable the use of the vulnerability by other, nefarious actors. There are additional questions for researchers: What would the sorts of vulnerabilities that law enforcement want to use cost? How would law enforcement's participation in the vulnerabilities market change costs? What's the benefit of using this wiretap data as opposed to the easier-to-obtain stored communications records? All of this returns us to the policy community's issues: online social networks aren't classic communications providers under the law, but they serve many of the same functions. What should their legal responsibilities be?[2] And in an age of ubiquitous online presence, should laws regarding law enforcement's access to communications' transactional data be updated?

References

1. D. Sanger, *Confront and Conceal: Obama's Secret Wars and Surprising Use of American Power*, Crown, 2012.
2. S. Landau, "Testimony before the House of Representatives Committee on the Judiciary, Subcommittee on Crime, Terrorism, and Homeland Security," *Going Dark: Lawful Electronic Surveillance in the Face of New Technologies*, 17 Feb. 2011.

with CALEA) to direct exploitation by law enforcement can make it more difficult to detect extralegal abuse of the interception tools by rogue investigators or agencies. With CALEA, a third party—the telecom carrier—is always involved in provisioning intercepts; exploitation tools, in contrast, can be used unilaterally and potentially without the knowledge of any independent party. We note that this isn't an issue unique to computer interception; many law enforcement capabilities, from deadly weapons to access to sensitive databases, are potentially subject to misuse. Developing robust technical and procedural mechanisms to audit and control the use of, and the data collected by, interception tools must be a central requirement for their expanded use.

A crucial issue, with both legal and technical implications, is the reliability of data gathered by intercept tools. Although a detailed examination of the issue is beyond our scope here, we do note that judges must be convinced that such tools are reliable and trustworthy: such tools must capture exactly the traffic authorized, no more and no less. A tool that misses some traffic might miss exculpatory evidence; a tool that captures too much could lead to confusion over who, precisely, made incriminating comments, and may violate the warrant's limits. Mechanisms that affect third parties' computers, intentionally or accidentally, are especially problematic from this perspective. This is a different issue from minimization, which ensures that the wiretap captures only the subject of an order and only when he or she is engaged in criminal activity. Minimization will also need to be conducted, as it is for any wiretap.

The FBI has dealt with related concerns in deploying the Computer and Internet Protocol Address Verifier (CIPAV),[12] a program that "calls home," meaning that it informs the FBI of a target machine's addressing and protocol data and information such as current IP address, MAC address, open ports, and so on. CIPAV is employed to enable surveillance of the targeted machine.

CIPAV details aren't public, but thanks to documents obtained under the Freedom of Information Act, some information on how the FBI handles the legal aspects of surveilling a target's machine is available.[12] Installing CIPAV requires accessing the target's computer, so first law enforcement seeks a search warrant to install CIPAV on the target's machine. Once it has the IP address and any other information necessary for conducting the surveillance, law enforcement returns to court to obtain a pen register/trap-and-trace order (https://www.eff.org/node/58430). Such a carefully constructed approach might be an appropriate model for law enforcement's use of targeted exploitation tools generally.

Developing a large, well-funded vulnerability exploitation laboratory potentially represents a significant increase in the FBI's technological capabilities. But in a world that's rapidly converting to fully IP-based communications, such capabilities will likely become increasingly important in supporting legally authorized surveillance. Given that law enforcement and intelligence agencies are already using such techniques at a small scale today, it's critical that judges, magistrates, and policymakers be given meaningful technical context for

evaluating the impact of the intercept technologies that they're asked to authorize.

Finally, we note that we've focused here on the collection of content; we don't address the issue of data collected by pen registers and trap-and-trace devices. (Pen registers capture dialing, routing, addressing, and signaling information from the target, while trap-and-trace devices capture the information on communications received by the target.) If the communications architecture shares pen register and trap-and-trace data with a service provider, then the information is obtainable from the service provider; however, many highly decentralized architectures, including peer to peer, do not create such records.

When the basic model for voice communications was circuit-switched connections, CALEA was technically feasible even if its security might have been poor. The increasing diversity of local-loop technologies in the 1990s, even with the introduction of ISDN and wireless, still involved the same service (voice telephony) in a landscape that remained both highly regulated and relatively slow moving. The fact that voice communications were circuit-switched meant that you could make a plausible argument for CALEA's approach of shifting wiretaps out of the local loop.

But that argument is no longer applicable in the Internet context. We not only have increased diversity of the local loop (which can be IP-based, DSL, ISDN, or wireless), but we've also increased diversity of the services themselves (voice, email, IM, VoIP, and so on) and of the carrier/ISP infrastructures implementing them. From a situation that had a limited set of service providers providing centralized communications, we've moved to a world with a nearly infinite set of application providers offering highly decentralized ones. The conditions that might have briefly favored the CALEA approach increasingly no longer exist, and they're highly unlikely to return.

By placing wiretapping infrastructure costs on telecommunications carriers, CALEA functioned as a cost-shifting mechanism for the government. But the economic impact of the changes in telecommunications means that the externalities of the CALEA approach—particularly the costs to innovation and security—are now rapidly going up, even while the effectiveness of the CALEA approach is rapidly diminishing.

CALEA effectively imposed a hidden wiretapping tax. Funding a laboratory (such as DCAC) at a level that enables law enforcement to reliably conduct legally authorized surveillance is a much more efficient use of scarce resources and shifts the costs back to the model that existed before CALEA. Passive interception and targeted vulnerability exploitation tools can provide law enforcement with capabilities that give investigators what they need without simultaneously increasing the insecurity of the telecommunications infrastructure.

This latter point is critical. If legally authorized wiretaps are a tool that government occasionally needs, law enforcement will seek viable paths for conducting them. We can either mandate artificially introduced vulnerabilities across all our communications platforms (the CALEA approach), or law enforcement can take advantage of capabilities—the weaknesses that unavoidably occur in complex software systems—that are already there. The latter is ultimately preferable. Software vulnerabilities exist whether law enforcement uses them against its targets or not. By focusing on discovering and exploiting preexisting weaknesses in targets' platforms and ending the business of introducing weaknesses into the communications fabric, law enforcement effectively promotes a national infrastructure that doesn't preclude legally authorized wiretapping but that doesn't create new opportunities for criminal exploitation. This turns us away from the vulnerabilities introduced by the CALEA approach and toward a model where law enforcement supports securing the communications infrastructure, a win for both law enforcement and the broader society. ■

Acknowledgments

We're grateful for the discussions and suggestions made by Herbert Lin, Marty Stansell-Gamm, Lee Tien, Marcus Thomas, and an unnamed reviewer. Their help does not in any way imply endorsement of the ideas presented in this article.

References

1. E. Lichtblau, "Police Are Using Phone Tracking as a Routine Tool," *The New York Times*, 1 Apr. 2012, p. A1.
2. K. Dam and H. Lin, *Cryptography's Role in Securing the Information Society*, Nat'l Academy Press, 1996.
3. S. Landau, *Surveillance or Security? The Risks Posed by New Wiretapping Technologies*, MIT Press, 2011.
4. V. Caproni, "Going Dark: Lawful Electronic Surveillance in the Face of New Technologies," Subcommittee on Crime, Terrorism, and Homeland Security, Committee on Judiciary, 17 Feb. 2011; http://judiciary.house.gov/hearings/hear_02172011.html.
5. D. McCullagh, "FBI Quietly Forms Secretive Net-Surveillance," CNET, 22 May 2012; http://news.cnet.com/8301-1009_3-57439734-83/fbi-quietly-forms-secretive-net-surveillance-unit.
6. A. Gidari, "Designing the Right Wiretap Solution: Setting Standards under CALEA," *IEEE Security & Privacy*, vol. 4, no. 3, 2006, pp. 29–36.
7. S. Bellovin et al., "Security Implications of Applying the Communications Assistance to Law Enforcement Act to

Voice over IP," 2006; http://privacyink.org/pdf/CALEAVOIPreport.pdf.
8. Network Working Group, *IETF Policy on Wiretapping*, IETF RFC 2804, May 2000.
9. V. Prevelakis and D. Spinellis, "The Athens Affair," *IEEE Spectrum*, July 2007, pp. 18–25.
10. P. Colaprico, "Da Telecom dossier sui Ds, Mancini parla dei politici," (in Italian), *La Repubblica*, 26 Jan. 2007.
11. S. Landau, "The Large Immortal Machine and the Ticking Time Bomb," *J. Telecommunications and High Technology Law*, vol. 11, no. 1, 2013, pp. 1–43.
12. J. Lynch, "New FBI Documents Provide Details on Government's Surveillance Software," Electronic Frontier Foundation, 29 Apr. 2011; https://www.eff.org/deeplinks/2011/04/new-fbi-documents-show-depth-government#footnote2_01mhuxa.
13. G. Anastasia, "Big Brother and the Bookie," *Mother Jones*, Jan./Feb. 2002.
14. *United States v. Nicodemo S. Scarfo,* et al.
15. Criminal Action No. 00-404 (NHP), US District Court for the District of New Jersey.
16. A. Greenberg, "Shopping for Zero-Days: A Price List for Hackers' Secret Software Exploits," *Forbes*, 23 Mar. 2012; www.forbes.com/sites/andygreenberg/2012/03/23/shopping-for-zero-days-an-price-list-for-hackers-secret-software-exploits/.
17. M. Hofmann and T. Timm, "'Zero-Day' Exploit Sales Should Be Key Point in Cybersecurity Debate," Electronic Frontier Foundation, 29 Mar. 2012; https://www.eff.org/deeplinks/2012/03/zero-day-exploit-sales-should-be-key-point-cybersecurity-debate.
18. B. Schneier, "The Vulnerabilities Market and the Future of Security," *Forbes*, 30 May 2012; www.forbes.com/sites/bruceschneier/2012/05/30/the-vulnerabilities-market-and-the-future-of-security/.
19. L. Bilge and T. Dimitras, "Before We Knew It: An Empirical Study of Zero-Day Attacks in the Real World," *Proc. 2012 ACM Conf. on Computer and Communications Security*, ACM, 2012.
20. Grand Jury Subpoena Duces Tecum, 25 Mar. 2011, United States of America, Plaintiff-Appellee, versus John Doe, Defendant-Appellant, for Judges Understanding Capabilities of the Technical Tools in the United States Court of Appeals for the Eleventh Circuit, nos. 11-12268 & 11-15421, DC Docket no. 3:11-mc-00041-LAC.
21. *Weems v. United States*, US Reports, vol. 217, 1910, p. 349.

Steven M. Bellovin is a professor of computer science at Columbia University; he's currently on leave and serving as chief technologist of the Federal Trade Commission. (His work on this article was completed before he joined the FTC; opinions expressed are personal and not those of the FTC.) He works on security, privacy, and related public policy issues. Bellovin has a PhD in computer science from the University of North Carolina at Chapel Hill. He is a member of the ACM. Contact him at smb@cs.columbia.edu.

Matt Blaze directs the Distributed Systems Laboratory at the University of Pennsylvania, where his research focuses on security, privacy, and scale in computing and communications systems. Blaze has a PhD in computer science from Princeton University. Contact him at blaze@cis.upenn.edu.

Sandy Clark is a PhD candidate in the Distributed Systems Lab at the University of Pennsylvania and is advised by Matt Blaze and coadvised by Jonathan Smith. Her research focuses on understanding the mechanisms involved in the computer security arms race. Contact her at clarks@cis.upenn.edu.

Susan Landau works in cybersecurity, privacy, and public policy and is currently a Guggenheim Fellow. She was previously a Distinguished Engineer at Sun Microsystems and a faculty member at the University of Massachusetts Amherst and Wesleyan University. Landau has a PhD from MIT and is a member of AAAS, ACM, AMS, and IEEE. Contact her at susan.landau@privacyink.org.