Cyber Threats and Cyber Realities: Law, Policy, and Regulation in Business, the Professions and National Security

Roger Williams University School of Law

and School of Justice Studies,

June 17-20, 2013

TABLE OF CONTENTS:

STEWARDSHIP OF CYBERSPACE

DUTIES FOR INTERNET SERVICE PROVIDERS

MELISSA E. HATHAWAY
Hathaway Global Strategies LLC
JOHN E. SAVAGE
Brown University

MARCH, 2012

CYBERDIALOGUE2012

WHAT IS STEWARDSHIP IN CYBERSPACE?

Canada Centre for Global Security Studies
MUNK SCHOOL OF GLOBAL AFFAIRS
UNIVERSITY OF TORONTO

ABSTRACT

In today's interconnected world, the Internet is no longer a tool. Rather, it is a service that helps generate income and employment, provides access to business and information, enables e-learning, and facilitates government activities. It is an essential service that has been integrated into every part of our society. Our experience begins when an Internet Service Provider (ISP) uses fixed telephony (plain old telephone service), mobile-cellular telephony, or fixed fiber-optic or broadband service to connect us to the global network.[1] From that moment on, the ISP shoulders the responsibility for the instantaneous, reliable, and secure movement of our data over the Internet.

INTRODUCTION

ISPs come in many forms and sizes and go by many names: the phone company, the cable company, the wireless company, etc. They are the Internet stewards: planning and managing resources, providing reliable connectivity, and ensuring delivery for traffic and services. And while the communications infrastructure security as a whole is generally believed to be robust, recent events suggest that the networks and the platforms on which Internet users rely are becoming increasingly susceptible to operator error and malicious cyber attack. In 2012, we should therefore ask whether ISPs have additional duties to ensure the reliable delivery of an essential service.

In this article, we expose the gap between ISPs' written responsibilities and the unwritten, yet expected ones. Specifically, we define eight ISP duties:

1. Duty to provide a reliable and accessible conduit for traffic and services
2. Duty to provide authentic and authoritative routing information
3. Duty to provide authentic and authoritative naming information
4. Duty to report anonymized security incident statistics to the public
5. Duty to educate customers about threats
6. Duty to inform customers of apparent infections in their infrastructure
7. Duty to warn other ISPs of imminent danger and help in emergencies
8. Duty to avoid aiding and abetting criminal activity

1 Services include: Public-switch telephone network (dial-up); Digital Subscriber Line (DSL) (usually copper), Asymmetric Digital Subscriber Line (ADSL); broadband wireless; cable modem (cable Internet); Fiber to the Premises (FTTx) (optical fiber); Integrated Services Digital Network (ISDN) (transmission of voice, video, data, and other network services over the traditional circuits); frame relay (wide-area network); Ethernet; Asynchronous Transfer Mode (ATM); satellite Internet access; and synchronous optical networking (SONET) (using lasers over fiber).

The latter duties are helpful in calibrating threats and funding responses to them.

The Internet is radically different from the plain old telephone service (POTS) that has provided voice communication since the nineteenth century. POTS established a "circuit" or path through the telephone network that remained constant during the communication session. The telephone network operated according to a strict, regimented set of processes and technologies that provided a highly reliable service, but adapted to change slowly. It did not have an application programming interface (API) to allow for third-party access and experimenting with telecommunication services was discouraged.

The Internet operates much differently. Long messages are decomposed into packets that move from source to destination following potentially different paths through the network. This is called packet switching. The Internet also provides a simple interface to communication networks that makes it easier for third parties to create innovative communication-based products to connect and access the Internet and providers to introduce a new generation of value-added services and applications. Yet, the Internet and the communications and services that ride on it rely on the integrity of routing and naming infrastructures. These two critical functions are essential to the proper functioning of the Internet.

ROUTING

The Internet is a network of networks. Networks consist of end systems, called hosts, and intermediate systems, called routers, connected via communication channels. Information travels through a network on paths chosen by a routing process that is implemented by routers. These paths automatically change many times a day, as congestion on one network might make an alternate path more attractive, or if a network has downtime – either intentional or not – so a new path is needed until the preferred path is restored.

Unfortunately, the technology in use today to ensure the Internet is operational is based on trust; it cannot give adequate guarantees that the expected network configuration is the one in place. As a consequence, one ISP can issue an update to another, whether by accident or by design, that will send Internet traffic to the wrong destinations. This lack of trust has resulted in major disruptions of Internet routing and can enable malicious activity, such as monitoring traffic, identity theft, and disruption of commerce.

NAMING

ISPs provide naming services to both their customers and other Internet users. Domain names are human-friendly names that are translated into Internet Protocol (IP) addresses, for example www.acme.com is a domain name, and 216.27.178.28 is its IP address. People like to use domain names and routers like to use IP addresses. Therefore, a system that converts one to the other was needed: the Domain Name System (DNS).

The DNS is the "telephone directory" that does this translation for the Internet. This "telephone directory" is implemented as a hierarchical collection of "servers." The DNS system was not designed to be inherently secure. When a request comes in to translate, a series of queries and responses occur until a mapping is found for the domain name in question. When a request is made, the requestor accepts the first response that it receives, and then uses it. Imagine asking a question to a room of strangers,

and assuming that whoever answered the question first (regardless of accuracy) is truthful. This is what the DNS essentially does and this vulnerability can and often does result in end users being misdirected to fraudulent websites on the Internet.

The Domain Name System Security Extensions (DNSSEC) is a set of extensions to the underlying DNS protocol suite that was designed to address this problem, but it has not yet been widely implemented by ISPs. DNSSEC uses cryptographically signed messages to authenticate the sender, which ensures that only "authorized" entities can resolve a name to an IP address or answer the question.[2]

THE ROLE OF ISPS

Approximately twenty-five ISPs carry as much as 80 percent of all the Internet traffic.[3] They own and operate a critical infrastructure that facilitates the delivery of essential goods and services. As intermediaries and stewards of this infrastructure, they have an important role to play in fostering security.

When a new ISP connects to the Internet it implicitly agrees to certain terms concerning the transmission of packets, sharing of routing information, resolution of domain names, reporting on the status of the Internet, and handling emergencies.[4] Until now these understandings were not made explicit. But there should be an explicit duty to comply with technical aspects of Internet participation. Given the rapid rise in the Internet's complexity and the critical role the Internet has come to play in the global economy, providers should be obligated to be stewards of the global enterprise. We can no longer be one click away from an infection, disruption, or worse yet, no service.

DUTIES INCUMBENT ON ISPS

The major telecommunications providers and ISPs, collectively, have unparalleled access into global networks, which enables them, with the proper tools, to detect cyber intrusions and attacks as they are forming and transiting towards their targets. Today, some ISPs limit spam, notify customers of botnet infections, and partner with law enforcement to deny the distribution of child pornography.[5] Internationally, this collection of autonomously administered networks already adheres to common protocols, enables seamless global connectivity, and collaborates to ensure twenty-four/seven uninterrupted service. If nations worked together to define codes of conduct that all ISPs agree to follow, it would result in a more secure Internet infrastructure and service. Here are some duties to which ISP might subscribe.

2 Cryptographic signing is a digital guarantee that information has not been modified, as if it were protected by a tamper-proof seal that is broken if the content is altered.

3 Sriram Vadlamani, "The Top 25 Telecom Companies in the World, Based on Brand," Asian Correspondent.com, 12 April 2009, http://asiancorrespondent.com/515/top-25-telecom-companies-in-the-world-based-on-brand-value/. The Cooperative Association for Internet Data Analysis (CAIDA) show that the top twenty Autonomous Systems account for the majority of the IPv4 prefixes and addresses. http://as-rank.caida.org/ Also, DoubleClick AdPlanner for April 2011 show that the largest 25 of the top 1000 properties accounted for 80 percent of web traffic globally.

4 A network that is under the administrative control of one organization is called an autonomous system (AS). There are approximately 40,000 ASes operating today. For the purposes of this paper, we treat the acronym ISP as a synonym for either ISP or AS. Routing within an AS is called intradomain routing whereas routing between ASes is called interdomain routing.

5 A bot is a malicious form of software that could use your computer to send spam, host a phishing site, or steal your identity by monitoring your keystrokes. Infected computers are then controlled by third parties and can be used for cyber attacks.

1. Duty to provide reliable and accessible conduit for traffic and services

The Internet is a basic, ubiquitous, and essential communications tool for all of society. Governments around the world are adopting policies to facilitate citizen access to the Internet via a fast, reliable, and affordable Information Communications Technology (ICT) infrastructure. This vision is reflected in the Organisation for Economic Co-Operation and Development's (OECD) Internet Economy; Europe's Digital Agenda; the United States's National Broadband Plan; and in the International Telecommunications Union's (ITU) initiatives.[6] Economic progress, citizen access, and infrastructure quality are measured in terms of price, bandwidth, speed/quality of service, skills, content and language, and applications targeted to low-end users.[7] Progress is being made and these global initiatives are bringing faster broadband Internet access for every citizen to facilitate our information society needs and global e-commerce demands.

For example, Finland passed a law in 2010 legislating that every one of its citizens will have the right to access one megabit per second (Mbps) broadband connection, obligating twenty-six telecommunications companies to provide that quality of service.[8] Finland went on to amend their constitution to make broadband access a constitutional right. The United Kingdom promises to have a minimum connection of two Mbps to all homes by 2012.[9]

Government efforts to provide universal access at lower cost to consumers have been underway for decades. Telecommunications liberalization brought the promise of global income gains (economic growth) by making access to knowledge easier. The General Agreement on Tariffs and Trade (GATT) Uruguay Round (1986-1993) began the discussion among nations. It was further codified in the Marrakech Treaty in 1994, where the General Agreement on Trade in Services (GATS) principles called for the transparency of, access to, and use of public telecommunications transport networks (PTTN) and services "on reasonable and non-discriminatory terms." This included obligations for interconnection to PTTN (including private networks) as well as safeguards for public-service responsibilities (duty to warn) and to protect the technical integrity of the network (reliable service).

In 1997, the World Trade Organization (WTO) adopted a Basic Telecommunications Agreement (BTA) to liberalize facilities-based international service and to allow foreign entities to own a majority interest in facilities used to provide international voice and data service.[10] Examples of the services covered by the agreement include voice telephony, data transmission, telex, telegraph, facsimile, private leased circuit services (i.e., the sale or lease of transmission capacity), fixed and mobile satellite systems and services, cellular telephony, mobile data services, paging, and personal communications systems.

6 The International Telecommunication Union (ITU) is the United Nations' specialized agency for information and communication technologies.

7 *Measuring the Information Society*, 2011 International Telecommunications Union, Geneva, Switzerland.

8 "1 Mbit Internet Access a Universal Service in Finland from the Beginning of July,"Press Release, 29 June 2010, Finland Ministry of Transport and Communications, http://www.lvm.fi/web/en/pressreleases/-/view/1169259.

9 "Government Reveals Super-Fast Broadband Plans," BBC News, 6 June 2010, http://www.bbc.co.uk/news/technology-11922424.

10 "Report on International Communications Markets 2000 Update," prepared for Senator Ernest F. Hollings, United States Senate Committee on Commerce, Science, and Transportation, Federal Communications Commission, 4 May 2001, 3.

In addition to the basic agreement, fifty-five governments agreed to value-added services (or telecommunications for which suppliers "add value" to the customer's information by enhancing its form or content or by providing for its storage and retrieval, such as on-line data processing, on-line database storage and retrieval, electronic data interchange, e-mail, or voice mail). The World Trade Organization's director-general, Mr. Renato Ruggiero, stated that "information and knowledge, after all, are the raw material of growth and development in our globalized world."[11]

The rapid adoption of technology and growing migration of essential services to be delivered on Internet-based infrastructure demands a re-examination of whether ISPs should be classified as nondiscriminatory. That is, must they treat all customers equally in terms of service or can they "discriminate?" Can we really say that the Internet or those who provide information services over the Internet deserve a similar degree of explicit responsibility as that assigned to "telecommunications service providers?"

Internationally, most nations do not distinguish between basic services (traditional modes of communications) and enhanced services (Internet-based services). However, the United States has made that distinction. The Telecommunications Act of 1996 created separate regulatory regimes for companies providing voice telephone service, cable television service, and providers of information services (broadband). The law did not necessarily envision the convergence of voice, data, and video services and infrastructures. A year after this law was enacted, the United States agreed at the WTO to treat both value-added services (Internet) and traditional communications (voice) in a nondiscriminatory manner. The Federal Communications Commission (FCC) or Congress should clarify this contradiction. Why? Because the rapid adoption of technology and growing migration of essential services to be delivered on Internet-based infrastructure demands that broadband and other Internet-based services be classified as core telecommunications services. This obligates the providers to deliver a *reliable* service that contributes to the stability and resiliency of the global communications infrastructure.

The United States and other countries are pursuing deeper integration of critical infrastructures with Internet-based technologies, like the "smart grid," a computerized network that facilitates electricity and information flows between homes and electrical suppliers; computerized health records; public safety alerts (Voice Over Internet Protocol); and next-generation air-traffic management. However, these essential services may not be built to the same standards for which the traditional voice telephone system was built. Broadband network reliability and resiliency are vital for all services that traverse a network, including traditional communications services. Our reliance on the dependable operation of communications networks is growing. Therefore, it may be necessary to expand existing communications reliability and resilience programs, including best practices and associated outage reporting, as these services transition from traditional modes of communications to Internet-based technologies. Outage reports and other reliability data collected by regulators provide insight on the overall health of communications reliability and security of the critical infrastructure and, where necessary, enables regulators to work with individual entities or the industry as a whole to bring

11 "WTO Telecom Talks Produce Landmark Agreements," World Trade Organization, paper 16, 15 February 1997, http://www.wto.org/english/res_e/focus_e/focus16_e.pdf.

about improvements.[12]

The FCC realizes that it "needs a clear strategy for securing the vital communications networks upon which critical infrastructure and public safety communications rely."[13] Europe is already moving forward with streamlining its regulatory process as part of the Digital Agenda for Europe. Europe recognizes that compliance monitoring and enforcement of a nondiscrimination policy allows for more choices, at affordable prices, underpinned by a higher standard of service.[14]

Many nations have recognized that it is in their national economic interest to enhance access to and participation in the Internet. ISPs provide an essential citizen service – the Internet – and they also provide the conduit upon which other essential services depend (e.g., Smart Grid). Therefore, it is their duty to serve as reliable and accessible conduits to Internet traffic and services.

2. Duty to provide authentic and authoritative routing information

Interdomain routing (from ISP to ISP) occurs primarily through the Border Gateway Protocol (BGP).[15] BGP has become a standard because of its simplicity and resilience. Under BGP, each ISP announces destinations that can be reached via it and the paths that packets will take to these destinations. (Think of this as a message that says I am open for business, I can route your information, and if you send it to me, it will pass through these ISPs.) These announcements propagate to neighbours and eventually to all routers on the Internet. BGP relies on trust among the operators of gateway routers—routers between ASes—to ensure the integrity of Internet routing information. However, this trust has been compromised on a number of occasions, revealing fundamental weaknesses in this critical Internet utility and service.

When BGP vulnerabilities are exploited, Internet traffic can be misdirected and misused. For example, in February 2008, Pakistan Telecom was ordered by the Pakistan telecommunications ministry to prevent its users from viewing certain YouTube addresses. Announcements of short paths to these addresses were designed to draw traffic from within Pakistan to the provider who then proceeded to discard the traffic. Unfortunately, these announcements leaked from Pakistan and made portions of YouTube inaccessible to about two thirds of all Internet users for about two hours.[16]

On 10 April 2010, BGP users received an alert regarding a possible prefix hijack by China's largest ISP, China Telecom. For approximately fifteen minutes, this ISP generated approximately 37,000 unique prefixes that were not assigned to them.[17] This is what is typically called a prefix hijack and while the hijack had modest to minimal impact on total Internet traffic volumes, China was ten times more affected than the United States. This event underscores the

12 "Audit Report: The Department's Management of the Smart Grid Investment Grant Program," United States Department of Energy, Office of Inspector General, OAS-RA-12-04, January 2012.

13 *Connecting America: The National Broadband Plan*, The United States Federal Communications Commission, 16 March 2010.

14 "Commission Launches Public Consultation on the Application, Monitoring and Enforcement of Non-discrimination Obligations in Electronic Communications," European Commission, 28 November 2011, http://ec.europa.eu/information_society/policy/ecomm/library/public_consult/non_discrimination/index_en.htm.

15 K. Butler, T.R. Farley, P. McDanier, and J. Rexford, "A Survey of BGP Security Issues and Solutions," Proceedings of the IEEE, 98, no. 1 (January 2010): 100-122.

16 Declan McCullagh, "How Pakistan Knocked YouTube Offline." CNET News, 25 February 2008, http://news.cnet.com/8301-10784_3-9878655-7.html.

17 "Chinese ISP Hijacks the Internet," BGP.mon blog, 8 April 2010, http://bgpmon.net/blog/?p=282.

vulnerability of the BGP routing infrastructure and reminds us that an intentional criminal could store, alter, or just throw away the traffic.[18]

In the Chinese case, given the brevity of the incident and the fact that no traffic was known to have been lost, the redirection may have been an accident. However, as we learned a few years ago, it is possible for an ISP to create path announcements that can deliberately move traffic to a particular ISP where a man-in-the-middle attack can be perpetrated. In such an attack, packets can be read, modified, or destroyed.[19]

The only way to solve the BGP trust problem is to develop and administer a system that allows each step in the process to be signed and certified. Routers should be able to affirm with high confidence that each routing announcement has not been modified in transit and that the sender is authorized to make such an announcement.

Of the many proposals that have been made to meet the trust requirements, Secure BGP (S-BGP) is the most secure.[20] Unfortunately, it has not been deployed, possibly because at the time the proposal was made, it was considered to be computationally demanding and its implementation requires a global public key infrastructure (PKI). Although the situation has changed, adoption of S-BGP will be challenging due to the large number of routers now in operation globally. Through simulation and analysis, Gill and colleagues have made a convincing argument that by seeding large ISPs with S-BGP and having them provide attestations for stub ASes (85 percent of all ASes are stubs), profits will drive ISPs to adopt it.[21]

Packets can still transit from IP to IP without the DNS.[22] However, without BGP, packets can't move at all. Regulators around the world have begun discussions with industry regarding the adoption of secure routing procedures and protocols based on existing work in industry and the research we described. ISPs need a process or framework for securing BGP announcements that includes specific technical procedures and protocols. The framework, if adopted by large ISPs (even the leading ten or fifteen companies), could go a long way toward making the Internet a more reliable, secure service.[23] Protocols and infrastructure are needed for everyday use of the Internet. ISPs have a duty to provide authentic and authoritative routing information. To us, this means they should adopt S-BGP or something equivalent.

3. Duty to provide authentic and authoritative naming information

As we mentioned, the Domain Name System (DNS) is the "telephone directory" for the Internet. This directory is implemented as a hierarchical collection of "servers." There are thirteen root zone servers that contain the names of the top-level-domain (TLD) name servers associated with suffixes such as .mil, .edu, or .com. Each of these servers contains the names of subdomain name servers, such as brown.edu, which resolve or translate universal resource locaters (URLs) into IP addresses. The root zone, top-level, and subdomain name servers are authorized by the Internet Corporation for Assigned Names and

18 "Chinese BGP Hijack Putting Things into Perspective," BGP.mon blog, 21 November 2010, http://bgpmon.net/blog/?p=323.

19 Joel Hruska, "Gaping Hole Opened in Internet's Trust Based BGP Protocol," http://arstechnica.com/security/news/2008/08/inherent-security-flaw-poses-risk-to-internet-users.ars.

20 See "Secure BGP Project," http://www.ir.bbn.com/sbgp/.

21 Gill. P., Schapira, M, and Goldberg S. "Let the Market Drive Deployment: A Strategy for Transitioning to BGP Security," Proceedings of SIGCOMM 2011, 15-19 August 2011.

22 Joel Hruska, "Gaping Hole Opened in Internet's Trust-based BGP Protocol."

23 Ibid.

Numbers (ICANN) to provide name resolution. Thus, these servers are said to be authoritative.

For efficiency reasons, ISPs maintain DNS caches.[24] If a user asks for a translation that is not in the cache, the ISP finds it and inserts it into the cache. These entries have a time-stamp associated with them and are refreshed when the "time-to-live" limit is reached.

The DNS system may be designed for efficiency but not security. For example, when a computer or a DNS cache asks for the resolution of a domain name, a series of queries and responses to a root zone server, top-level-domain server, domain and subdomain server occur, in that order, until a mapping for the URL in question is found. When a request is issued at each stage of the transaction, the initiator accepts the first response that it receives to its query. This provides an opportunity for a man-in-the-middle attack in which a malicious agent can insert a response that directs the initiator to a nonauthoritative server. The DNS also provides a key function for IP applications such as VoIP. In some cases, when a user makes a call with VoIP, the user's machine will contact a DNS server to get the IP address of the called number. However, if the DNS cache is poisoned, the calls could be misdirected to somebody else who could then obtain the user's personal and confidential information.

Several dramatic abuses of the untrustworthy DNS system have occurred. Two recent examples demonstrate its vulnerability. In November 2011, the Federal Bureau of Investigation (FBI), working in cooperation with Estonian authorities and others, dismantled an international cybercrime ring that infected millions of computers worldwide by downloading a malicious piece of software (i.e., a Trojan) called DNSChanger.[25] This piece of malware changed the IP address of the DNS cache used by various computer operating systems so that instead of using a local, and presumably honest cache, it redirected the compromised machine to a compromised DNS cache. Not only could this Trojan evade the proscriptions of the recently introduced pieces of legislation of Stop Online Piracy Act (SOPA, H.R. 3261) and Protect Intellectual Property Act (PIPA, S. 968)—it also misdirected users to sites where they participated, unwittingly, in "click fraud." Clicks that appeared legitimate generated millions of dollars in income for the fraudsters. A recent report claims that DNSChanger continues to infect computers at half of the Fortune 500 companies and half of all federal agencies in the USA.[26]

A second example involves VeriSign, an American firm that operates two root servers, and three top-level domains (TLDs) namely the .com, .net, and .name domains. VeriSign announced that it had been repeatedly hacked in 2010 but that it does not believe that its DNS database servers were breached.[27] If their system was breached, trust in their management of key components of the DNS database would be seriously damaged.

The security extensions to DNS (DNSSEC) we mentioned were developed by the Internet Engineering Task Force (IETF) and are designed

24 A cache is a file that holds copies of the mappings of domain names to IP addresses.

25 "Operation Ghost Click: International Cyber Ring That Infected Millions of Computers Dismantled," FBI website, http://www.fbi.gov/news/stories/2011/november/malware_110911, 9 November 2011, accessed 5 February 2012.

26 Brian Krebs, "Half of Fortune 500s, US Govt. Still Infected with DNSChanger Trojan" Krebs on Security, February 2012, http://krebsonsecurity.com/2012/02/half-of-fortune-500s-us-govt-still-infected-with-dnschanger-trojan/.

27 VeriSign Annual 10-K Corporate Filing. See also, Joseph Menn, "VeriSign Hacked: Security Repeatedly Breached at Key Internet Operator," Reuters, 2 February 2012.

to address the vulnerabilities with DNS.[28] They rely on digital signatures to certify that the parties requesting updates to DNS mappings are authorized by a central trust anchor to make those changes. The bottom line is that DNSSEC is intended to improve data integrity on DNS connections through the authentication process. However in order for DNSSEC to work, it must be supported at every level of the DNS hierarchy, from root server to browser. A chain of trust must be established from the information producer to the information consumer. Without this unbroken chain of trust, opportunities for exploitation remain.

Today, most of the root servers are implementing DNSSEC and many of the TLDs are deploying DNSSEC. ISPs need to upgrade their systems and increase their technical knowledge to deploy DNSSEC deeper into the infrastructure. Accelerating the deployment of DNSSEC will help eliminate BGP vulnerabilities and bring a higher level of service quality to their customers. Because customers need assurance that their traditional voice, VoIP, email, video, or other service is going to get to its correct destination and maintain its integrity along the way, ISPs have a duty to provide authentic and authoritative naming information as part of their service.

4. Duty to report anonymized security incident statistics to the public

A major impediment to calibrating the scope and scale of security threats to the Internet is the paucity of public data. Some ISP customers are reluctant to have incident data concerning their enterprises or infrastructures reported out of concern for their reputations as responsible guardians of data being tarnished.[29] This lack of transparency limits the security product industry's ability to deliver products that perform with higher assurance levels. It also limits the research community's access to data that could facilitate idea creation and innovative solutions that increase security across the entire architecture.

ISPs should have a duty to report data sets, including but not limited to the (1) volume of spam in transit; (2) estimated number of compromised machines owned by customers of an ISP; (3) remediation steps proposed to customers by an ISP and actions the ISP has taken; (4) frequency, intensity, sources, and targets of distributed-denial-of-service attacks; (5) location, frequency, and duration of network outages and route disruption; and (6) the frequency, source, and target of cache-poisoning attacks, to facilitate solution development. It would also be helpful if the ISP reported event data that exceeded predetermined thresholds similar to their responsibilities when there is a disruption of communications service.

Initially it may suffice for only the largest ISPs to report such data. They have more resources at their disposal and they service the largest percentage of compromised machines.[30] Reporting incident data may either be encouraged

28 Internet Engineering Task Force overview of DNSSEC, http://www.dnssec.net/rfc.

29 Recent guidance issued by the Securities Exchange Commission (SEC) notes that all public companies have existing obligations to disclose material risks and events on their public filings (13 October 2011). A risk or event is material if it is important for the average investor to know before making an investment decision. The clarifying guidance states that "material risks can include cyber risks and material events can include cyber breaches, including the theft of intellectual property/trade secrets, penetrations which compromise operational integrity, etc." See http://www.sec.gov/divisions/corpfin/guidance/cfguidance-topic2.htm.

30 According to Michel van Eeten, and others, in 2009 60 percent of all infected machines were in the top 200 of all ISPs. ("The Role of Internet Service Providers in Botnet Mitigation: An Empirical Analysis Based on Spam Data," OECD Science, Technology and Industry Working Papers, 2010/05.)

by national or transnational authorities or prohibited by law. For example, the European Parliament and Council of Ministers reached an agreement on pan-European telecommunications reform that is being transposed into national laws.[31] Section 13(a), "Security and Integrity of Networks and Services," of the Regulatory Framework for Electronic Communications in the European Union outlines a number of duties for ISPs. Among them is the duty to "notify the competent national regulatory authority of a breach of security or loss of integrity that has had a significant impact on the operation of networks or services; and where appropriate, the national regulatory authority concerned shall inform the national regulatory authorities in other Member States and the European Network and Information Security Agency (ENISA)."[32] The directive goes on to say that the regulators can ask the ISPs to "inform the public when it determines that disclosure of the breach is in the public interest."[33] Finally, the directive requires that "once a year, the national regulatory authority concerned shall submit a summary report to the Commission and ENISA on the notifications received and the action taken in accordance with this paragraph."[34]

In the United States, by contrast, many attorneys interpret the Electronic Communications and Privacy Act of 1986, along with the Telecommunications Act of 1996, as limiting ISPs' ability to share this data.[35] Aggregate threat data is collected by commercial security firms, such as Symantec and McAfee, which make it available to customers for a fee.

Because ISP customers are reluctant to have data released about their enterprises, their cooperation may require that safeguards be put in place, including keeping data private while allowing useful statistics based on the data to be computed. Such safeguards have been the holy grail of statistics since at least the 1970s.[36] In 2006, two papers emerged that provided a basis for showing how it is possible to give highly accurate responses to queries on statistical databases while minimizing the probability of identifying individual records.[37] The authors made a key observation—that privacy comes from uncertainty. Using this observation they defined the concept of differential privacy, which is based on query functions that use random numbers to generate results.

A randomized query function is said to offer differential privacy if the probability that it produces an outcome when a single element is in the data set is within a constant multiplicative factor of the probability that it produces the same outcome when the element is not in the data set. Thus, a differentially private query function behaves approximately the same whether

31 "Regulatory Framework for Electronic Communications in the European Union," European Parliament Council, 2009. Specifically, see directive 2009/140/EC of the European parliament and of the council of 25 November 2009 that amends directives 2002/21/EC on a common regulatory framework for electronic communications networks and services, 2002/19/EC on access to, and interconnection of, electronic communications networks and associated facilities, and 2002/20/EC on the authorization of electronic communications networks and services.

32 "Regulatory Framework for Electronic Communications in the European Union," European Parliament Council, 2009, p. 55.

33 Ibid., p. 55.

34 Ibid., p. 55.

35 The Telecommunications Act of 1996. Pub. L. No. 104-104, 110 Stat. 56. Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, 100 Stat. 1848 (1986). Lawyers for the ISPs interpret the ECPA to prohibit the voluntary provision of customer data.

36 Tore Dalenius, "Towards a Methodology for Statistical Disclosure Control," *Statistik Tidskrift* [Statistical Review] 15 (1977): 429-44.

37 Cynthia Dwork, "Differential Privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming* (ICALP), 2: 1-12 and Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," in *Proceedings of the 3rd Theory of Cryptography Conference* 2006, 265-284.

the element is in the data set or not. Functions of this kind have been developed for a large number of useful queries.[38]

If ISPs assumed the duty to report anonymized statistics on security incidents to the public, it would likely lead to the emergence of a standard of care or best practices for all ISPs to follow. It would also spark the development of innovative solutions and the deployment of better capabilities for enterprise and infrastructure protection.

5. Duty to educate customers about threats

Most ISPs deploy advanced technologies that detect malicious and harmful activity. They have unique insights on the scope and scale of cyber threats and incidents affecting our homes, businesses, and infrastructures. As such, they can also play a unique role in educating their customers about the threats. Customers who are able to recognize a threat and are presented with user-friendly resources/tools are capable of enhancing their security, and as a result are better poised to protect themselves. Government and industry educational resources are emerging in every corner of the world, many of which have an ISP as a critical component of the education campaign.

For example, in December 2011 a coalition of twenty-eight service providers, network operators, and equipment suppliers in the European market began working together to make a better and safer Internet for children ("Coalition").[39]

The Coalition is a cooperative voluntary effort aimed at making it easier to report harmful content, ensuring privacy settings are age-appropriate, and offering wider options for parental control, reflecting the needs of a generation that is going online at an increasingly young age. European Commission Vice President Neelie Kroes said, "this new Coalition should provide both children and parents with transparent and consistent protection tools to make the most of the online world. The founding Coalition members are already leaders in children's safety online. Working together we will be setting the pace for the whole industry and have a great basis for fully empowering children online."[40]

In the United States, two projects have emerged worth noting. The first is a web-wide partnership entitled GetNetWise.[41] It is a public service funded and developed by Internet industry corporations and public interest organizations to help ensure that Internet users have safe, constructive, and educational or entertaining online experiences. The GetNetWise coalition wants Internet users to be just "one click away" from the videos, educational materials, and other helpful hints they need to make informed decisions about their and their family's use of the Internet. The service is facilitated by the Internet Education Foundation, a nonprofit organization dedicated to educating the public and policymakers about the potential of a decentralized global Internet to promote communications, commerce, and democracy.

The second program is the National Cyber Security Alliance (NCSA). Its sponsors include AT&T, Verizon, Microsoft, Google, McAfee, Symantec,

38 Cynthia Dwork and Adam Smith, "Differential Privacy for Statistics: What We Know and What We Want to Learn," *Journal of Privacy and Confidentiality*, 1, no. 2 (14 January 2009): 135-54.

39 Founding Coalition members are: Apple, BSkyB, BT, Dailymotion, Deutsche Telekom, Facebook, France Telecom-Orange, Google, Hyves, KPN, Liberty Global, LG Electronics, Mediaset, Microsoft, Netlog, Nintendo, Nokia, Opera Software, Research in Motion, RTL Group, Samsung, Sulake, Telefonica, TeliaSonera, Telenor Group, Tuenti, Vivendi, and Vodafone.

40 "Digital Agenda: Coalition of Top Tech and Media Companies to Make Internet Better Place for Our Kids," Press Release, 1 December 2011, European Commission. http://europa.eu/rapid/pressReleasesAction.do?reference=IP/11/1485.

41 GetNetWise, http://www.getnetwise.org/about/

Cisco, ADP, and many others. The organization's purpose is to educate and therefore empower a digital society to use the Internet safely and securely at home, work, and school, protecting the technology that individuals use, the networks they connect to, and shared digital assets. It develops and disseminates educational materials for home, classroom, and business use.

Australia commissioned a study to understand the depth of education initiatives around the world. The research report by Galexia documents more than sixty-eight different initiatives and highlights the different techniques used to educate consumers on the basics of cyber security.[42] The study notes that many of these initiatives help fight illegal and harmful online content and conduct while at the same time promoting the safer use of both the Internet and other communication technologies.

Other innovative activities include the fielding of video games to educate the public. In the United States there is a partnership between i-SAFE (a non-profit organization dedicated to educating and empowering youth (and others) and Carnegie Mellon University to safely, responsibly, and productively use Information and Communications Technologies (ICT). They are integrating an on-line game called "MySecureCyberspace" into thousands of K-12 programs across the United States.[43] Children play the game in a digital "city" and learn to secure key infrastructures and critical services. Children become aware of online security and privacy issues as they interact with the game's Carnegie Cadet characters in a virtual world. Similarly, the United Kingdom has launched an on-line virtual reality game entitled "Smokescreen" that guides teenagers through the dangers of social networking.[44] The game has over thirteen "missions" that place teenagers in situations that force them to ask themselves "what would I do if it happened to me?" What if ISPs promoted innovative educational materials like these? In a secondary educational campaign, the Ministry of Defense aired a number of television commercials alerting citizens of their responsibility for on-line security. The commercials present scenarios in which criminals, terrorists, and predators review personally posted data on YouTube, Twitter, Facebook, etc., to achieve their nefarious purposes.[45]

The cyber-security problem space is growing faster than the solution space. If ISPs undertake the duty to educate their customers about the threats, then our respective government leaders will be able to engage in a broader conversation about all of the solutions that can be brought to bear to address the problem comprehensively.

6. Duty to inform customers of apparent infections in their infrastructures

Media headlines throughout the past year have been rife with high-profile cybercrime events, confirming that insecure computers are being infected every day. Criminals have shown that they can harness bits and bytes with precision

42 *An Overview of International Cybersecurity Awareness and Educational Initiatives*: A Research Report, Australian Communications and Media Authority, May 2011. http://www.acma.gov.au/webwr/_assets/main/lib310665/galexia_report-overview_intnl_cybersecurity_awareness.pdf

43 "Cyber Education," Carnegie Mellon University, http://www.carnegiemellontoday.com/pdfs/news_pdfs/CMSecurity_Cyber-Education.pdf and the game is accessible on the web at www.mysecurecyberspace.com.

44 The game is accessible on the web at http://www.smokescreen-game.com/ and "Smokescreen —A New Resource for Promoting Saftey Online." https://blogs.glowscotland.org.uk/glowblogs/ISRU-News/2010/05/06/smokescreen/.

45 United Kingdom, Ministry of Defense Online Security Campaign. See "Personal Security Online" videos on YouTube at http://www.youtube.com/watch?v=hpKilrYDLxg; http://www.youtube.com/watch?v=-UziYBdnQhk; http://www.youtube.com/watch?v=1UyWN0uREfk; and http://www.youtube.com/watch?v=qXZSzs-P2kQ.

to deliver spam, cast phishing attacks, facilitate click fraud, and launch distributed-denial-of-service (DDoS) attacks. The increasing frequency of these events in recent years and the scale of those affected have been alarming. Some estimates suggest that, in the first quarter of 2011, almost 67,000 new malware threats were seen on the Internet every day. This means more than forty-five new viruses, worms, spyware, and other threats were being created every minute – more than double the number from January 2009. As these threats grow, security policy, technology, and procedures need to evolve even faster to stay ahead of the threats.[46] A recent Symantec report suggests that these trends will continue.[47] Between 2010 and 2011 the numbers were discouraging.

- There were 286 million unique variants of malware that exposed and potentially exfiltrated our personal, confidential, and proprietary data;
- Each data breach exposed, on average, 260,000 identities;
- There was a 93 percent increase in web-based attacks (compromised/hijacked websites where the visitor would become infected);
- The underground economy paid anywhere from $.07 to $100 for each of our stolen credit card numbers; and
- Realizing that mobile payments and mobile platforms (e.g., smart phones and iPads™) would be the newest vector of technology adoption, there was a 42 percent increase in mobile-operating-system vulnerabilities and subsequent exploitation.

46 Cybersecurity Green Paper, United States Department of Commerce, Internet Policy Task Force, June 2011, ii.

47 "Symantec Internet Security Threat Report: Trends for 2010," Volume 16, April 2011.

While consumer education is necessary, recent efforts have shifted toward having the ISPs act as the intermediary or control point for impeding the spread of infection and eradicating the malicious activity.[48]

Australian ISPs are showing the world that industry can organize and implement a consistent approach to help inform, educate, and protect their customers in relation to cyber security.[49] Thirty leading ISPs serving over 90 percent of the Australian market have opted in to providing a four-pronged security service, including: (1) a notification/management system for compromised computers, (2) a standardized information resource for end users, (3) a comprehensive resource for ISPs to access the latest threat information, and (4) a reporting mechanism to CERT Australia to facilitate a national high-level view of threat status. Australian customers are notified about suspicious activity, their ISP assists them stopping the infection, and if need be, the ISP quarantines them so that the computers cannot browse the wider web until they have been repaired. "The Australian experiment has been stunningly successful," said Michael Barrett, chief information security officer for PayPal. "We will see more countries adopting this model."[50] The Australian model is now promoted by the OECD, which found that

48 According to a recent report by the OECD, "Internet intermediaries bring together or facilitate transactions between third parties on the Internet. They give access to, host, transmit and index content, products and services originated by third parties on the Internet or provide Internet-based services to third parties." See "The Economic and Social Role of Internet Intermediaries," OECD, 2010. Available online at www.oecd.org/dataoecd/49/4/44949023.pdf.

49 "Internet Service Providers Voluntary Code of Practice for Industry Self-Regulation in the Area of Cyber Security," Internet Industry Association (Australia), 1 June 2010, http://iia.net.au/images/resources/pdf/iiacybersecuritycode_implementation_dec2010.pdf.

50 Joseph Menn, "US Starts to Tackle Hacking Curse," *Financial Times*, 12 October 2011.

ISPs represent nearly 87 percent of the total market (service) in forty nations.[51] They also recognize that peer pressure among the ISPs is an important incentive that contributes to security and opting in to an overall program.

In Japan, more than seventy Internet service providers, representing 90 percent of the customer base, have assumed the duty to inform their customers of infections. ISPs notify consumers if their machines appear to be part of a botnet infection and offer government-funded tools offered through Cyber Clean Center (CCC) to clean the computers.[52] This voluntary program has shown remarkable reduction of infection rates. From 2007 to 2011, ISPs have reduced the rate of botnet infection from about 2.5 percent of personal computers to just 0.6 percent.[53]

In the Netherlands, Dutch ISPs signed an anti-botnet pact and jointly launched an initiative to fight malware-infected computers and botnets. The effort involves fourteen ISPs and represents 98 percent of the consumer market. ISPs are sharing information to obtain better coverage and reduce response times. They have accepted the responsibility to notify their victimized users and quarantine the infections until assistance can be provided.[54]

In Germany, the German Federal Office for Information Security (BSI) has mandated that its ISPs track down infected machines and provide advice to users on how to clean their computers.[55] Telefonica has taken this initiative further. It recently launched customer protection insurance against online fraud at a cost of five euro per month. "The customer and up to six family members are covered against data misuse, fraudulent online payment practices and theft or damage of the Telefonica Germany DSL router, modem or surf stick. Telefonica claims to be the first network operator to offer a customer protection insurance."[56]

And in the United States, Comcast is a market leader and early adopter of the duty to inform and protect its customers. Through its service known as Constant Guard, Comcast proactively contacts its customers via an email "service notice" if Comcast believes one or more of its customers' computers is infected with malicious software (e.g., it is a bot). Comcast's efforts in this regard have received the attention of the Federal Communications Commission (FCC).

Service providers, network operators, and equipment suppliers are working together as part of the FCC's Communications Security, Reliability and Interoperability Council (CSRIC) to propose a set of agreed-upon voluntary practices that would constitute the framework for an opt-in implementation model for ISPs to conduct botnet remediation. [57] This initiative is modelled after the Australian iCODE Project and, if widely adopted in the United States, could make a sig-

51 "The Role of Internet Service Providers in Botnet Mitigation: An Emperical Analysis based on Spam Data." STI Working Paper, May 2010, Organisation for Economic Co-operation and Development, Directorate for Science Technology and Industry, 12-Nov-2010, p. 41.

52 "Botnets: Detection, Measurement, Disinfection and Defence," European Network and Information Security Agency (ENISA), 2011, p. 98.

53 Joseph Menn, "US Starts to Tackle Hacking Curse," *Financial Times*, 12 October 2011.

54 Gadi Evron, "Dutch ISPs Sign Anti-Botnet Treaty," Dark Reading, 29 September 2009, http://www.darkreading.com/blog/227700601/dutch-isps-sign-anti-botnet-treaty.html.

55 John Leyden, "German ISPs Team up with Gov Agency to Clean up Malware," The Register, 9 December 2009.

56 "Telefonica Germany Offers Internet Insurance," Telecom Paper, 9 February 2012, http://www.telecompaper.com/news/telefonica-germany-offers-internet-insurance.

57 CSRIC's mission is to provide recommendations to the FCC to ensure, among other things, optimal security and reliability of communications systems, including telecommunications, media, and public safety.

nificant difference in ensuring the health of their Internet backbone.

These examples show that ISPs are already assuming the duty to inform customers of apparent infections in their infrastructures principle. Some ISPs might participate strictly for business purposes—to reduce fraud, infections, and unnecessary bandwidth use. Others may engage for more altruistic purposes: they may wish to assume responsibility for the safety of the Internet and their users, perhaps at their own expense. Either way, "it is important that ISPs collectively battle this problem and protect their customers as well as prevent nuisance to the rest of the Internet," says Albert Vergeer, director of Internet for KPN, XS4ALL, and Telfort.[58]

7. Duty to warn other ISPs of imminent danger and help in emergencies

ISPs have a unique view of the malware and activity transiting their infrastructure. They also have a responsibility to provide uninterrupted service to their customers. As we see more organized and semiorganized groups disrupt services and infrastructures in support of the "cause of the day" using DDoS or similar malware, ISPs may have to adopt and practise Good Samaritan behaviour.

Good Samaritan laws more typically apply in countries in which the foundation of the legal system is English common law.[59] In many countries that use civil law (i.e., the legal system inspired by Rome) as the foundation for their legal systems, the same legal effect is more typically achieved using a principle of duty to rescue.[60] Perhaps one of the best internationally recognized of these laws is the use of the SOS.[61] When a threatened party uses SOS, it triggers a duty to assist (DTA) that marshals available resources to help victims avoid or recover from harm. Similar duties to assist exist in both domestic and international contexts, such as a nuclear accident or a pilot's Mayday call. Duncan Hollis has called for the creation of an e-SOS, a duty to assist in the case of cyber emergencies.[62] Even the North Atlantic Treaty Organization's (NATO) has article 4, which is a consultation and information-sharing arrangement that activates when a member nation perceives its territorial integrity, political independence, or security is threatened.[63] Even the Telecommunications Act of 1996 contains a Good Samaritan provision to protect ISPs from liability when they act in good faith to block or screen offensive content hosted on their systems.[64]

To effectively defend the information infrastructure requires that private and public parties identify threats quickly and mitigate their impact effectively. As at sea, the timing and scale of some cyber threats can overwhelm the

58 Gadi Evron, "Dutch ISPs Sign Anti-Botnet Treaty," Dark Reading, 29 September 2009, http://www.darkreading.com/blog/227700601/dutch-isps-sign-anti-botnet-treaty.html

59 Hyder Gulam and John Devereaux, "A Brief Primer on Good Samaritan Law for Health Care Professionals," *Australian Health Review* 31, no. 3 (2007): 478–82.

60 A duty to rescue is a concept in tort law that arises in a number of cases, describing a circumstance in which a party can be held liable for failing to come to the rescue of another party in peril.

61 SOS is not an acronym, but a specific Morse Code, represented as ". . .---. . ." It was adopted as the standard distress signal in 1912 by the London International Telegraph Convention. G.E. Wedlake, SOS: The Story of Radio-Communication (Newton Abbot, UK: David & Charles 1973).

62 Duncan Hollis, "An e-SOS for Cyberspace," Harvard International Law Journal, 52, no. 2 (Summer 2011): 37.t

63 The North Atlantic Treaty, 4 April 1949, North Atlantic Treaty Organization, http://www.nato.int/cps/en/natolive/official_texts_17120.htm.

64 The Telecommunications Act of 1996. Pub. L. No. 104-104, 110 Stat. 56. The 1996 Telecommunications Act included a "Good Samaritan" provision to protect Internet Service Providers (ISPs) from liability when they act in good faith to block or screen offensive content hosted on their systems. Id. § 230(c).

most sophisticated individuals, groups, and even states. For example, in July 2009, the United States and South Korea fell victim to a DDoS attack against thousands of computers and major government, media, and financial websites. The attacks were launched from at least five different control hosts in multiple countries, including the United States. The United States government turned to industry to determine the origin and character of the threat and asked the ISPs to shut down the operations and restore services.

In Germany, the Anti-Botnet-Advisory Center helps customers remove botnet threats and other malicious software from their computers. The centre is supported by a group of ISPs that informs affected customers of their infections and then assists with specific tools to help the customer eliminate or eradicate the infection.[65] The centre is working with Norton, Kaspersky, and Avira to provide tailored software that "cleans" customer computers of malicious software. Similarly, the Finish Communications Regulatory Authority (FICORA) directs network operators to disconnect the infected machines of its customers from the Internet until the machine is disinfected.[66]

In late January 2012, the Polish government experienced multiple attacks targeting websites under the gov.pl domain. Most of the attacks were DDoS based, attributed to Anonymous, which declared radical protests after the Polish government revealed plans to sign the ACTA treaty on 26 January. Websites of the Polish Parliament, Ministry of Foreign Affairs, and Internal Security Agency were among the victims of these attacks. Organizers enjoy the fact that DDoS attacks are simple and efficient. You press a button and within seconds the targeted website stops responding. Minutes later news portals report about the incident. Collateral customers are then affected, including banks, media, telecommunications companies, and Polish Railways.[67] Governments like Poland turn to their ISPs to assist in the defence of their infrastructure, and work proactively to establish countermeasures and incident response plans to mitigate and minimize the potentially devastating impact of a determined and well-resourced opponent.

As more of industry moves its services to an Internet-based infrastructure, one could envision a digital crisis similar to the ash clouds over Iceland that halted air traffic around the world for days in the spring of 2010. While US laws focus on shielding from liability those who choose to help in a situation they did not cause, European laws criminalize failure to help in such a situation.[68] What if, for example, the e-ticketing of several major airlines and train systems was taken off-line? The duty-to-assist obligation could be demanded to help restore that service so that passengers could be ticketed and tracked, and packages moved. This is not an impossible hypothetical situation because a reservation systems breakdown for United Airlines in fact stranded thousands of passengers and disrupted flights around the United States in January 2006.

Given the integrated and global nature of the Internet and the central role played by the large ISPs, it is incumbent on them to honour the duty to assist other ISPs both to warn of imminent danger, such as an emerging attack, and to

65 Safer Internet Surfing—Remove Threats, https://www.botfrei.de/en/ueber.html.

66 Finnish Communications Regulatory Authority, http://www.ficora.fi/en/index/saadokset/ohjeet.html.

67 "DDoS Against Polish Government Websites," http://www.cert.pl/news/4856/langswitch_lang/en.

68 Nancy Benac,. "Good Samaritan Laws Common in Europe but Rare in America," *Wisconsin State Journal* (1997-09-05): 7A, ISSN 0749405X, Retrieved 2010-01-07. (Registration Required)

help when an attack or outage occurs that seriously injures or disables a neighbour ISP. ISPs could deploy a hotline phone system, like the Inter-Network Operations Center Dial-By-ASN (INOC-DBA), that connects "Network Operations Centers (NOCs) and Security Incident Response Teams (IRTs) of Internet infrastructure providers, operators of Internet exchanges, critical individuals within the Internet security, policy, emergency-response, and governance community, and equipment vendors' support personnel." [69]

8. Duty to avoid aiding and abetting criminal activity

The recent settlement by Google with the United States Department of Justice underscores a new responsibility for ISPs—that they have a duty to avoid aiding and abetting criminal activity. From 2003 to 2009, Google permitted online Canadian pharmacies to place advertisements through Google's largest advertising program called AdWords. This service facilitated the unlawful importation of controlled pharmaceuticals into the United States. In the settlement agreement, Google admitted to its knowledge of, and participation in, unlawful advertising.[70] It is unlawful[71] for pharmacies outside the United States to ship prescription drugs to customers in the United States.[72] "The Department of Justice will continue to hold accountable companies who in their bid for profits violate federal law and put at risk the health and safety of American consumers," said Deputy Attorney General Cole. "This investigation is about the patently unsafe, unlawful, importation of prescription drugs by Canadian on-line pharmacies, with Google's knowledge and assistance, into the United States, directly to US consumers," said US Attorney Neronha. "It is about holding Google responsible for its conduct by imposing a $500 million forfeiture, the kind of forfeiture that will not only get Google's attention, but the attention of all those who contribute to America's pill problem."[73]

The Google case study suggests that as soon as the ISP or host becomes aware that a content or activity is unlawful, it could be found guilty of aiding and abetting the offence if it does not take immediate action to prevent the activity.[74] In 1999, the District Court, The Hague found that an access provider was liable for having maintained a link which connected to a site containing counterfeit material and

> declares it to be the law that by having a link on their computer systems which when activated brings about a reproduction of the works that CST (the plaintiff) has the copyright to on the screen of the user, without the consent of the plaintiffs, the Service Providers are acting unlawfully if and insofar that they have been notified of this, and moreover the correctness of the notification of this

69 INCO-DBA Hotline Phone Q&A, Packet Clearing House, https://www.pch.net/inoc-dba/docs/qanda.html (last visited 6 March 2011).

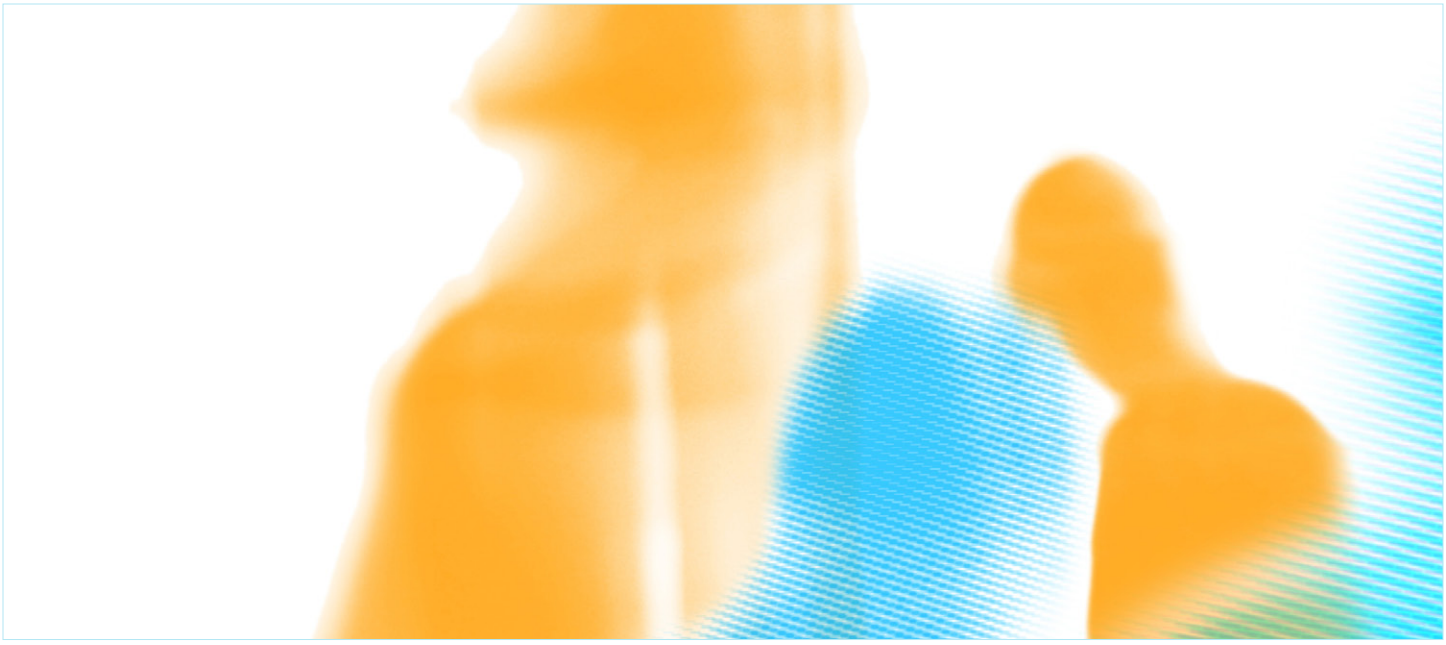
70 Non-prosecution Agreement, http://googlemonitor.com/wp-content/uploads/2011/05/Google%20Agreement.pdf.

71 These activities violate the Federal Food, Drug, and Cosmetic Act; Title 21 United States Code, section 331(a) and (d) (Introduction into Interstate Commerce of Misbranded or Unapproved Drugs). Where these prescription drugs are controlled substances, such conduct also violates the Controlled Substances Act, Title 21 United States Code, section 952 (Importation of Controlled Substances).

72 Google Non-prosecution Agreement, http://googlemonitor.com/wp-content/uploads/2011/05/Google%20Agreement.pdf and "DOJ Pharmacy Investigation Undermines Google Credibility," http://betanews.com/2011/08/28/doj-pharmacy-investigation-undermines-google-credibility/

73 "Google Forfeits $500 Million Generated by Online Ads and Prescription Drug Sales by Canadian Online Pharmacies," http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html.

74 On aiding and abetting, see the article by Sébastien Canevet, "Fourniture d'accès à l'Internet et responsabilité pénale" (Provision of access to the Internet and criminal liability).



> fact cannot be reasonably doubted, and the Service Providers have then not proceeded to remove this link from their computer system at the earliest opportunity.[75]

These cases can be extended to other forms of illicit or illegal behaviour conducted by customers or subscribers of those service providers. Other areas of the law substantiate this. For example, landlords can be held liable if they take inadequate precautions against criminal activity that harms tenants.[76] Entrepreneurs may be held liable if criminals use their premises to sell counterfeit or grey market goods.[77] Still others see it as a risk to their reputation. In March 2011, Microsoft decided that the Rustock botnet, the largest generator of spam in the world, was causing an Internet nuisance because it was damaging Microsoft products as well as its reputation. Accordingly, Microsoft turned to the courts to address the issue. On 16 March 2011, US Marshals accompanied employees of Microsoft's digital crimes unit into Internet hosting facilities in five US cities.[78] Using a federal court order, they seized the command-and-control servers that were responsible for manipulating an estimated one million computers worldwide.

Microsoft was not alone in its efforts to take down the Rustock infrastructure. The effort required collaboration between "industry, academic researchers, law enforcement agencies and governments worldwide."[79] Microsoft worked with pharmaceutical company Pfizer, the network security provider FireEye, Malware Intelligence Labs, and security experts at the University of Washington, each of whom attested in court to the dangers posed by Rustock and the impact

75 "Legal Instruments to Combat Racism on the Internet," European Commission against Racism and Intolerance, http://www.coe.int/t/dghl/monitoring/ecri/legal_research/combat_racism_on_internet/Internet_Chapter3_en.asp; see details of the case on http://www.juriscom/net/elaw/e-law11.htm.

76 See, for example, Sharp v. W.H. Moore, Inc., 796 p. 2d 506 (Idaho 1990); and Doug Lichtman and Eric Posner, "Holding Internet Service Providers Accountable," *John M. Olin Law & Economics Working Paper* no. 217 (July 2004): 9.

77 See, for example, Fonovisa v. Cherry Auction, 76 F.3d 259 (9th Cir. 1996) and Doug Lichtman and Eric Posner, "Holding Internet Service Providers Accountable," 9.

78 Bruce Sterling, "Microsoft Versus Rustock Botnet," *Wired*, 28 March 2011. http://www.wired.com/beyond_the_beyond/2011/03/microsoft-versus-rustock-botnet/.

79 Ibid.

on the Internet community. Additionally, Microsoft also worked with the Dutch High Tech Crime Unit within the Netherlands Police Agency to help dismantle part of the command structure for the botnet operating outside of the United States. Moreover, Microsoft worked with China's Computer Emergency Response Team (CN-CERT) to block registrations of domains in China, a pro-active approach aimed at preventing the stand-up of future command and control servers. Finally, Microsoft's digital crimes unit worked with global ISPs and CERTs around the world to remediate the infections.

Microsoft demonstrated that a multinational corporation can and should be responsible for discriminating against the illegal activity operating on service provider infrastructures. The global cooperation that it enjoyed during the takedown of the Rustock botnet suggests that others may follow suit with a duty to avoid aiding and abetting criminal activity.

Because ISPs are a platform for global access they can also become an instrument for illicit or illegal activity. Individually, law enforcement agencies will never be able to defeat the clever tactics and agile criminal infrastructures. Therefore, ISPs must have a duty to avoid aiding and abetting criminal activity and must play an important role in addressing and deterring illegal activity, fraud, and misleading and unfair practices conducted over their networks and services. Internet-based activities should comply with the law and all parties have responsibility to improve the safety and stability of the Internet of the future, including individuals, providers, ISPs, and judicial authorities.

CONCLUSION

The Internet is both a critical infrastructure in itself and a key component of other forms of critical infrastructure, underpinning economic and social activity at a global level. This paper exposes the gap between ISPs' written responsibilities and the unwritten, yet expected ones. As our examples illustrate, precedents are emerging around the world for ISPs to shoulder more responsibility for the stewardship of the Internet. The first three duties contain the basic functions, the expected services that an ISP should undertake as part of their participation in the global internet: (1) duty to provide a reliable and accessible conduit for traffic and services; (2) duty to provide authentic and authoritative routing information; and (3) duty to provide authentic and authoritative naming information. Networks and the platforms on which Internet users rely should not be susceptible to operator error or cyber attack. We can no longer be one click away from an infection or worse yet, no service. As such, many countries are turning to their regulatory authorities to apply pressure on their ISPs to facilitate the adoption of these core functions.

The next four duties usually fall outside of a regulatory regime, yet in many ways fall within our unwritten expectations or ISPs' social responsibility to maintain the security and integrity of the Internet as a global platform for communication and commerce. These duties are echoed in a recent OECD communiqué entitled, "Principles for Internet Policy Making."[80] The four duties of (4) duty to report anonymized statistics on security incidents to the public; (5) duty to educate customers about the threats; (6) duty to inform customers of apparent infections in their infrastructures; and (7) duty to warn other ISPs of imminent danger and help in emergencies, complement each other and help the Internet community to work together to stem the tide of the proliferating malicious activity that poisons our Internet experience and infects our Internet infrastructure. Today, some ISPs limit spam, notify customers of botnet infections, and partner with law enforcement to deny the distribution of child pornography. Some ISPs might participate strictly for business purposes— to reduce fraud, infections, and unnecessary bandwidth use. Others may engage for more altruistic purposes, like brand enhancement or a differentiated "secure" service by assuming responsibility for the safety of the Internet and their users, perhaps at their own expense.

Finally, while the Internet knows no specific geography, it facilitates activities between law-abiding nations. ISPs have a duty to avoid aiding and abetting criminal activity. Internet-based activities should comply with the law and all parties have the responsibility to improve the safety and stability of the Internet of the future, including individuals, providers, ISPs, and judicial authorities.

ISPs have an unparalleled access into and view of global networks, which gives them the proper tools to detect cyber intrusions and attacks as they are forming and transiting towards their targets. There are a limited number of ISPs that provide the world's Internet service (basic communication and enhanced services). If the leading fifteen or twenty companies were to become early adopters and market leaders for the eight duties of stewardship, they could make a significant difference in the overall security and resilience of the Internet. (The top twenty-five companies in 2009 by brand value are listed in Table 1.)

80 "Communiqué on Principles for Internet Policy Making,"
Delivered at an OECD High-Level Meeting, The Internet Economy: Generating Innovation and Growth, 28-29 June 2011, Paris, France.

TABLE 1: TOP 25 TELECOM COMPANIES IN THE WORLD, 2009

Rank	Brand	Parent Company	Brand Value ($bn)
1	Vodafone	Vodafone Group	26.59
2	AT&T	AT&T	24.6
3	Verizon	Verizon Comm	24.38
4	Orange	France Telecom	18.35
5	China Mobile	China Mobile	13.87
6	Telecom Italia	Telecom Italia	9.43
7	T-Mobile	Deutsche Telekom	8.96
8	Movistar	Telefonica	7.95
9	NTT DoCoMo	NTTC	7.54
10	BT	BT Group	7.29
11	Sprint	Sprint Nextel Corp.	7.07
12	Telefonica	Telefonica	6.33
13	Alcatel-Lucent	Alcatel-Lucent	5.16
14	America Movil	America Movil	5.08
15	Telstra	Telstra Corp.	4.64
16	O2	Telefonica	4.62
17	China Unicom	China Unicom	3.45
18	Qwest	Qwest Comm Intl	3.06
19	SoftBank	Softbank Corp.	3.02
20	KDDI	KDDI Corp.	3.01
21	Telenor	Telenor	2.97
22	Swisscom	Swisscom	2.96
23	MTS	Mobil TeleSystems	2.79
24	CNC	China Netcom Group	2.55
25	Airtel	Bharti Airtel Ltd	2.48

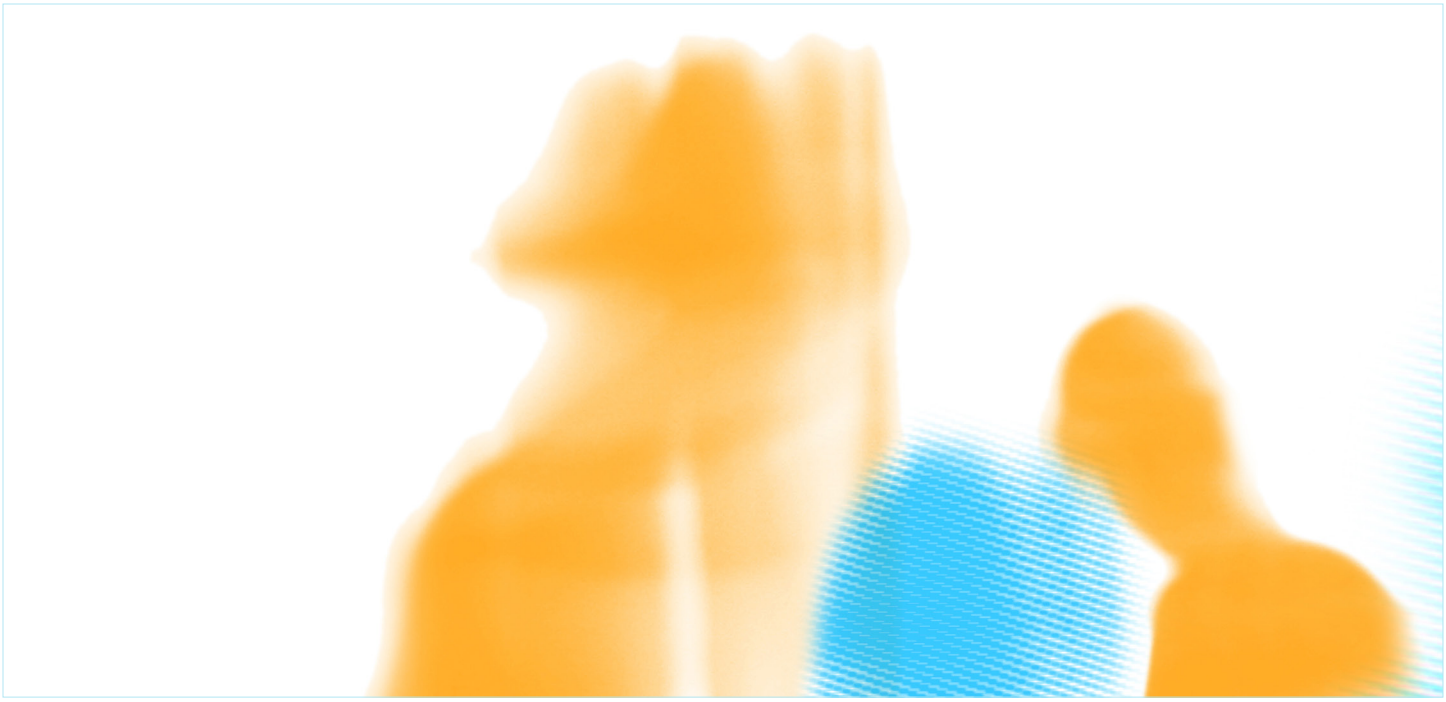
Alternatively, the top twenty Autonomous Systems (ASes) by customer cone size[81] could also assume broader responsibility for the health and hygiene of the Internet. These twenty ASes described in Table 2 (next page), which approximately map to ISPs, represent the broadest coverage of direct and indirect customer reach.[82]

81 Customer Cone refers to the set of ASes, IPv4 prefixes, or IPv4 addresses that can be reached from a given AS following only customer links.

82 The Cooperative Association for Internet Data Analysis (CAIDA) shows that the top twenty Autonomous Systems account for the majority of the IPv4 prefixes and addresses (http://as-rank.caida.org/).

TABLE 2: TOP TWENTY ASES BY CUSTOMER CONE

		Customer Cone		
AS Rank	AS Name	Number of ASes	Percentage of all ASes	Percentage of IPv4 addresses
1	Level 3 Communications	35,753	96%	97%
2	Hurricane Electric	33,621	91%	91%
3	Global Crossing Ltd.	33,427	90%	91%
4	Metromedia Fiber Net	30,524	82%	85%
5	Tinet SpA	29,989	81%	83%
6	Sprint	28,636	77%	82%
7	NTT America Inc.	28,501	77%	81%
8	Cogent/PSI	27,722	75%	73%
9	TeliaNet Global Network	27,573	74%	74%
10	AT&T Services, Inc.	27,375	74%	81%
11	Deutsche Telekom AG	27,114	73%	76%
12	Tata Communications	26,018	70%	73%
13	MCI Communications	25,632	69%	70%
14	ReTN.Net Autonomous	25,567	69%	68%
15	Savvis	25,077	67%	71%
16	Beyond The Network A	24,854	67%	70%
17	UPC Communications	24,538	66%	69%
18	XO Communications	24,364	66%	68%
19	Swisscom	23,944	64%	66%
20	Cable and Wireless	22,897	62%	68%



Regardless of the methodology chosen (i.e., market penetration or by topographic connectivity) a small number of ISPs could lead the way in ensuring the reliability, integrity, and security of the Internet as a critical infrastructure and thereby put pressure on the rest to follow. ISPs do come in many forms and sizes and go by many names: the phone company, the cable company, the wireless company, etc. They have become the stewards of the Internet: planning and managing resources, providing reliable connectivity, and ensuring delivery for traffic and services. In 2012 we should ask the ISPs to assume the explicit and implicit duties outlined in this paper to ensure the reliable delivery of an essential service—the Internet. Upon implementing these eight duties they will likely recognize one more unstated duty that is in the best interest of their business: to use their purchasing power to design and deploy the next generation of technology that protects users and accounts for security at the onset. After all, meeting tomorrow's demands for network capacity, new applications, and an expanding base of users requires extending and investing in the infrastructure. Anticipating the next-generation security requirements up front makes perfect business sense.

Melissa Hathaway is President of Hathaway Global Strategies LLC and a Senior Advisor at Harvard Kennedy School's Belfer Center. Ms. Hathaway served in the Obama Administration as Acting Senior Director for Cyberspace at the National Security Council and led the Cyberspace Policy Review. During the last two years of the administration of George W. Bush, Melissa served as Cyber Coordination Executive and Director of the Joint Interagency Cyber Task Force in the Office of the Director of National Intelligence where she led the development of the Comprehensive National Cybersecurity Initiative (CNCI).

Dr. John E. Savage is the An Wang Professor of Computer Science at Brown University. He earned his PhD in Electrical Engineering at MIT in coding and communication theory and joined Bell Laboratories in 1965 and Brown University in 1967. In 1979 he co-founded the Department of Computer Science at Brown and served as its second chair from 1985 to 1991. His research has centered on theoretical computer science and currently includes cybersecurity, computational nanotechnology, the performance of multicore chips, and reliable computing with unreliable elements.

ADMINISTRATION STRATEGY ON MITIGATING THE THEFT OF U.S. TRADE SECRETS

FEBRUARY 2013

Acknowledgement

This strategy is the product of a collaborative effort and reflects the recommendations and input from various entities of the U.S. government, including the Departments of Commerce, Defense, Homeland Security, Justice, State, Treasury, the Office of the Director of National Intelligence and the Office of the United States Trade Representative. This strategy reflects the research and reporting by the Departments of Commerce and Defense as well as the Office of the National Counterintelligence Executive respectively.

Table of Contents

Administration Strategy on Mitigating the Theft of U.S. Trade Secrets

"We are going to aggressively protect our intellectual property. Our single greatest asset is the innovation and the ingenuity and creativity of the American people. It is essential to our prosperity and it will only become more so in this century."

—President Barack Obama

Introduction

"We cannot look back years from now and wonder why we did nothing in the face of real threats to our security and our economy."

—President Barack Obama

The Administration is focused on protecting the innovation that drives the American economy and supports jobs in the United States. As a Nation, we create products and services that improve the world's ability to communicate, to learn, to understand diverse cultures and beliefs, to be mobile, to live better and longer lives, to produce and consume energy efficiently and to secure food, nourishment and safety. Most of the value of this work is intangible—it lies in America's entrepreneurial spirit, our creativity, ingenuity and insistence on progress and in creating a better life for our communities and for communities around the world. These intangible assets are often captured as intellectual property—copyrights, patents, trademarks and trade secrets, and reflect America's advantage in the global economy.

Emerging trends indicate that the pace of economic espionage and trade secret theft against U.S. corporations is accelerating.[1] There appears to be multiple vectors of attack for persons and governments seeking to steal trade secrets. Foreign competitors of U.S. corporations, some with ties to foreign governments, have increased their efforts to steal trade secret information through the recruitment of current or former employees.[2] Additionally, there are indications that U.S. companies, law firms, academia, and financial institutions are experiencing cyber intrusion activity against electronic repositories containing trade secret information.[3] Trade secret theft threatens American businesses, undermines national security, and places the security of the U.S. economy in jeopardy. These acts also diminish U.S. export prospects around the globe and put American jobs at risk.

As an Administration, we are committed to continuing to be vigilant in addressing threats—including corporate and state sponsored trade secret misappropriation—that jeopardize our status as the world's leader for innovation and creativity. We will continue to act vigorously to combat the theft of U.S. trade

1. The Office of the National Counterintelligence Executive (ONCIX), "Foreign Spies Stealing US Economic Secrets In Cyberspace", November 2011, at 1, available at http://www.ncix.gov/publications/reports/fecie_all/Foreign_Economic_Collection_2011.pdf

2. See ONCIX Report, supra note 1, at 8. When trade secrets are misappropriated by current or former employees, this method is referred to as an insider or "mole" operation.

3. See ONCIX Report, supra note 1, at 5.

secrets that could be used by foreign companies or foreign governments to gain an unfair economic edge. Departments across the U.S. government have roles in protecting trade secrets and preserving our nation's economic and national security. This strategy recognizes the crucial role of trade secrets in the U.S. economy and sets out a means for improved coordination within the U.S. government to protect them.

Strategy Action Items

1. Focus Diplomatic Efforts to Protect Trade Secrets Overseas

"Where every nation plays by the rules…and intellectual property and new technologies that fuel innovation are protected."

—President Barack Obama

In order to protect American innovation globally, trading partners must treat trade secret theft as a serious issue. The Administration, through the appropriate agencies, will take several steps to ensure this is the case.

Sustained and Coordinated International Engagement with Trading Partners

The Administration will continue to apply sustained and coordinated diplomatic pressure on other countries to discourage trade secret theft. This will be achieved by utilizing a whole of government approach directed at a sustained, consistent and coordinated message from all appropriate agencies to foreign governments where there are regular incidents of trade secret theft. Other governments must recognize that trade secret protection is vital to the success of our economic relationships and that they must take steps to strengthen their enforcement against trade secret theft.

The theft of U.S. trade secrets by foreign competitors or foreign governments has been and will continue to be raised by the most senior levels of the Administration with countries of concern. The relevant Federal agencies, including the Departments of Commerce, Defense, Justice, Homeland Security, State, Treasury and the U.S. Trade Representative, as appropriate, will continue to make it clear to the governments of those nations the importance the U.S. places on the protection of trade secrets and to press those governments to take action to reduce and resolve incidents of trade secret theft.

To assist in this effort, the Department of State will track scheduled diplomatic engagements and meetings by senior Administration officials with governments of countries where there are regular incidents of trade secret theft or that may be complicit in trade secret theft. During these meetings, senior Administration officials will deliver appropriate messages to their foreign counterparts to express the Administration's focus on reducing the incidents of trade secret theft, including improved legal frameworks, stronger enforcement of existing laws and strong and efficient remedies for trade secret owners.

Additionally, the Departments of Commerce and State and the U.S. Trade Representative will seek to build coalitions with other countries to deliver similar messages to countries of concern and to press jointly, or in coordination, for improved protection of trade secrets.

The Department of State and the U.S. Patent and Trademark Office (USPTO), through the USPTO's intellectual property Attachés, will also ensure that U.S. embassies located in countries that are known to present high-risk conditions for trade secret theft will incorporate trade secret protection into their established Intellectual Property Rights (IPR) Working Group plans, with input from appropriate agencies. The annual work plans will include concrete steps to work with the host government to address

trade secret theft. The identified embassies will also include discussions of trade secret issues as part of the IPR Working Groups' regular internal meetings in order to improve communication and coordination inside the embassies. The Embassy-led Working Groups will also enhance engagement with U.S. industry representatives in their host countries on trade secret theft issues.

Theft of Ford Motor Company Trade Secrets

In April 2011, Yu Xiang Dong was sentenced to 70 months in federal prison for theft of trade secrets and economic espionage. Yu was a former Ford Motor Company employee who resigned to work at Beijing Automotive Company. He copied 4,000 Ford documents onto an external hard drive, which he took to China. Ford valued the loss of the trade secrets at $50 million dollars.

Trade Policy Tools

The Administration will utilize trade policy tools to increase international enforcement against trade secret theft to minimize unfair competition against U.S. companies. The U.S. Trade Representative (USTR) will make additional efforts to promote adequate and effective protection and enforcement of trade secrets. These Administration efforts will include:

- Deeper cooperation with trading partners that share U.S. interests with the objective of promoting enhanced trade secret and other intellectual property protection in ways that are consistent with U.S. approaches and helpful in curbing trade in goods and services containing stolen trade secrets;
- Targeting weaknesses in trade secret protection through enhanced use of the annual Special 301 process[4], including the Special 301 Report, action plans and related tools to gather and, where appropriate, act upon information about the adequacy and effectiveness of trade secret protection by U.S. trading partners;
- Seeking, through USTR-led trade negotiations such as the Trans Pacific Partnership, new provisions on trade secret protections requiring parties to make available remedies similar to those provided for in U.S. law; and
- Continuing to raise trade secret protections as a priority issue in all appropriate bilateral, regional, and multilateral trade discussions and appropriate trade and IP-related forums, including the Trade-Related Aspects of Intellectual Property Rights Council and the Asia-Pacific Economic Cooperation, informed by interagency and stakeholder input regarding partners and issues of concern.

4. Through an extensive Special 301 interagency process, USTR publishes a report annually, known as the Special 301 Report, which designates countries of concern on different watch lists, referred to as "priority watch list" (PWL), "watch list" and "priority foreign country." Countries placed on the PWL are the focus of increased bilateral attention concerning the problem areas which will include trade secret protection. USTR also develops action plans and similar documents to establish benchmarks, such as legislative, policy or regulatory action, and as a tool to encourage improvements by countries in order to be removed from the Special 301 list.

International Law Enforcement Cooperation

International law enforcement cooperation is a critical part of combating the global nature of trade secret theft. To assist in domestic investigations of trade secret theft with an international element, Federal law enforcement agencies will also use, as appropriate, formal cooperative agreements or arrangements with foreign governments as a tool to strengthen relationships and investigative efforts. Federal law enforcement agencies will encourage cooperation with their foreign counterparts to:

- Enhance efforts to pursue domestic investigations of trade secret theft by foreign entities; and
- Encourage foreign law enforcement to pursue those targets themselves.

International Training and Capacity Building

The Department of Commerce will use existing programs[5] to educate foreign government officials and increase foreign capacity to protect trade secrets from theft and unlawful commercialization.

The Department of Justice and the Federal Bureau of Investigation, in collaboration with the Departments of Homeland Security and State, will include trade secret theft awareness and enforcement instruction in applicable international law enforcement training forums, such as the International Law Enforcement Academies and in country specific training missions.

International Organizations

The Administration will work with global organizations to strengthen international enforcement efforts and increase cross-border diplomatic and law enforcement cooperation. These efforts will include:

- The Departments of Commerce, Homeland Security, State, and Treasury and USTR will work with international organizations to ensure that there is robust trade secret protection abroad.
- The Department of Justice will continue to work with the European Police Organization and the International Criminal Police Organization on collaborative efforts to address trade secret misappropriation from the U.S. to recipients located abroad.

Theft of DuPont Trade Secrets

Hong Meng was a research chemist for DuPont. He was involved in researching Organic Light Emitting Diodes (OLED). DuPont's OLED research efforts resulted in the development of a breakthrough and proprietary chemical process for OLED displays. Mr. Meng stole trade secret compounds and passed them to a Chinese university. He was caught by the FBI and prosecuted by the U.S. Attorney's Office for the District of Delaware and was sentenced to 14 months in federal prison. DuPont valued the loss of the trade secrets at $400 million dollars.

5. The Department of Commerce has established the Intellectual Property Attaché Program and the USPTO Global Intellectual Property Academy to facilitate capacity building with foreign governments.

2. Promote Voluntary Best Practices by Private Industry to Protect Trade Secrets

"In America, innovation doesn't just change our lives. It's how we make a living."

—President Barack Obama

Advancements in technology, increased mobility, rapid globalization, and the anonymous or pseudonymous nature of the Internet create growing challenges in protecting trade secrets.[6] Companies need to consider whether their approaches to protecting trade secrets keeps pace with technology and the evolving techniques to acquire trade secrets enabled by technology. The Administration encourages companies to consider and share with each other practices that can mitigate the risk of trade secret theft. These best practices should encompass a holistic approach to protect trade secrets from theft via a wide array of vulnerabilities.

Support and Promote Voluntary Best Practices

The U.S. Intellectual Property Enforcement Coordinator (IPEC), working with appropriate U.S. government agencies, including the Departments of Justice and State, will help facilitate efforts by organizations and companies to develop industry led best practices to protect trade secrets. The Administration will encourage companies and industry associations to develop and adopt voluntary best practices, consistent with anti-trust laws, and help highlight those practices. Many private sector companies have recently begun to focus on examining their procedures in order to understand the threat and potential impact of trade secret misappropriation. These organizations are already working to develop best practices that companies can voluntarily implement to protect themselves against trade secret theft. The Administration will work to support groups crafting industry-driven initiatives that meet these objectives.

Identified best practices may not be suitable for every company or organization. Whether or not specific information is regarded as a trade secret is a matter determined by an individual company, not by industry at large. Additionally, for information to be legally protected as a trade secret, businesses need only take reasonable measures to protect the secrecy of such information which may vary by company and by industry. In practice, however, businesses may choose to take additional measures to protect trade secret information where appropriate. In identifying and promoting the adoption of best practices, it should be emphasized that such guidelines are intended solely to offer suggestions to assist businesses in safeguarding information they wish to keep secret and are not designed to be a minimum standard of protection.

The Administration encourages organizations and companies to examine internal operations and policies to determine if current approaches are mitigating the risks and factors associated with trade secret misappropriation committed by corporate and state sponsors. Some areas that private industries could consider for voluntary best practices include:

6. See ONCIX Report, supra note 1, at i-ii

- Research and development compartmentalization;
- Information security policies;
- Physical security policies;
- Human Resources policies; and

Theft of General Motors Trade Secrets

On November 30, 2012, a Federal jury in Detroit found Shanshan Du, a former General Motors (GM) engineer, and her husband, Yu Qin, both found guilty of stealing GM trade secrets related to hybrid vehicle technology worth $40 million. Du and Qin tried to pass the trade secrets to Chinese automaker Chery Automobile Company.

3. Enhance Domestic Law Enforcement Operations

"Our workers are the most productive on Earth, and if the playing field is level, I promise you—America will always win."

—President Barack Obama

As a result of the Attorney General's Task Force on Intellectual Property, established in 2010, the Federal Bureau of Investigation (FBI), which has primary responsibility for investigating domestic offenses under the Economic Espionage Act, increased the number of trade secret theft investigations by 29 percent from 2010.

Investigations and Prosecutions of Trade Secret Theft

The Department of Justice has made the investigation and prosecution of corporate and state sponsored trade secret theft a top priority. The Department of Justice and the FBI will continue to prioritize these investigations and prosecutions and focus law enforcement efforts on combating trade secret theft. The FBI is also expanding its efforts to fight computer intrusions that involve the theft of trade secrets by individual, corporate, and nation-state cyber hackers. The Department of Homeland Security component law enforcement agencies will continue to work cooperatively with the Department of Justice when its investigations uncover evidence of trade secret theft.

Theft of Cargill and Dow Chemical Trade Secrets

In October 2011, Kexue Huang, a former employee of both Cargill and Dow Chemical passed trade secret information to a Chinese university that was developing organic pesticides on behalf of China's government. Financial losses to both companies from his criminal acts exceed $7 million. In December 2011, after many months of hard work by FBI agents, CCIPS prosecutors and the U.S. Attorneys' Offices in Indiana and Minnesota, Huang was sentenced to 87 months in prison—the strongest sentence possible.

Law Enforcement and Intelligence Information Sharing

The Office of the Director of National Intelligence (ODNI)

ODNI will coordinate within the intelligence community to inform the private sector about ways to identify and prevent the theft of trade secrets that benefit a state sponsor or an entity with ties to a foreign government. ODNI will coordinate expanded discussions between the intelligence community and the private sector, focusing on four main aspects of the threat posed by trade secret theft:

- The number and identity of foreign governments involved in trade secret misappropriation;
- The industrial sectors and types of information and technology targeted by such espionage;
- The methods used to conduct such espionage; and
- The dissemination, use, and associated impact of information lost in trade secret misappropriation.

ODNI, though the Office of the National Counterintelligence Executive (ONCIX) will also counter the threat of trade secret misappropriation by sharing threat warning and awareness information with the private sector, as well as imparting counterintelligence tradecraft procedures tailored to the private sector.[7] In order to support this strategy, ONCIX will brief trade association groups and conferences on industry specific threats.

Report to Congress on Foreign Economic Collection & Industrial Espionage

In its November 2011 report to Congress, ONCIX determined that foreign collectors may have the greatest interest in the following areas:

- Information and communications technology;
- Business information that pertains to supplies of scarce natural resources or that provides foreign actors an edge in negotiations with U.S. businesses or the U.S. government;
- Military technologies, particularly marine systems, unmanned aerial vehicles, and other aerospace/aeronautic technologies; and
- Civilian and dual-use technologies in sectors likely to experience fast growth, such as clean energy and health care or pharmaceuticals.

The ONCIX also explored characteristics that make U.S. businesses more vulnerable to trade secret misappropriation including the use of portable devices; storage of information; globalization of economic activities; digitization of business records, research results, and other sensitive economic or technology-related information. A company within one of the four categories identified above is even more susceptible, when these high-risk factors are also present. The report also identified other risk factors. For example:

- The increase in data access points created by conducting business on smartphones and other mobile devices and storing information in the "cloud" increases the opportunities for malicious actors to steal or manipulate information.
- Companies with employees who work remotely are also likely to be at an increased risk of theft.

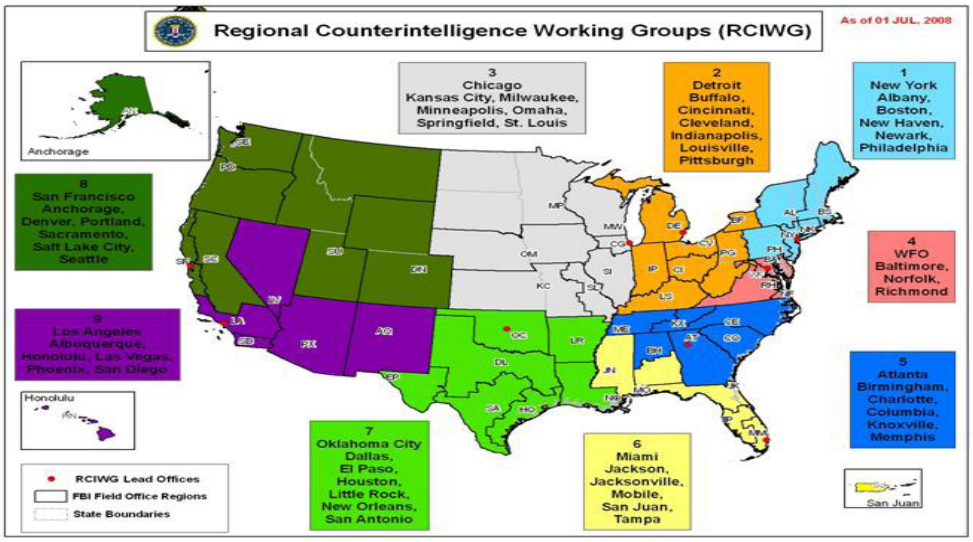
7. The Counterintelligence Enhancement Act of 2002, Public Law 107-306, authorizes ONCIX to carry out and coordinate outreach programs and activities on counterintelligence to other elements of the U.S. government and the private sector. These activities include vulnerability surveys of the private sector.

The Department of Justice

The Department of Justice and the FBI will continue to report on trade secret investigations and prosecutions.[8] Additionally, the FBI will continue its outreach and education efforts with the private sector through various local, regional and national initiatives. At the local level, each of the FBI's 56 field offices will continue to work with academic institutions, manufacturers, laboratories and other entities that are located within the field office's area of responsibility and are perceived as being potentially at risk for trade secret theft. At the regional level, the FBI will continue to meet regularly with other government agencies, industry, and academia to share information about insider threats, economic espionage and trade secret theft.

Theft of Valspar Trade Secrets

David Yen Lee worked for Valspar, an Indiana paint company. He stole trade secrets from Valspar and tried to pass them to Nippon Paint in China. Mr. Lee purchased a plane ticket to China, but was caught by the FBI before he could leave the U.S. On December 8, 2010, Mr. Lee was sentenced to 18 months in prison. Valspar valued the trade secrets between $7 and $20 million.



The FBI's headquarters will review the effectiveness of its local and regional efforts with a focus on the extent of outreach to companies and entities such as cleared defense contractors[9], universities, hospitals, high science companies, and emerging technology firms. The FBI will continue to engage with

8. The Department of Justice and the FBI are required to submit an annual report to the United States Congress pursuant to section 404 of the Prioritizing Resources and Organization for Intellectual Property Act of 2008, Public .Law 110-403.

9. The term "cleared defense contractor" means a private entity granted clearance by the Department of Defense to access, receive, or store classified information for the purpose of bidding for a contract or conducting activities in support of any program of the Department of Defense.

trade secrets owners through several national outreach organizations, including the Domestic Security Alliance Council, the National Security Business Alliance Council, and InfraGard, and will continue to work closely with various Information Sharing and Analysis Centers. These local, regional and national efforts will continue to reach a broad swath of companies in multiple sectors such as information technology, communications, aeronautics, engineering, energy, financial services, and consumer retail. The FBI's engagement with the private sector promotes reasonable safeguards based on recent intelligence, case studies, and emerging trends.

The Department of Justice and the FBI will continue to train prosecutors and investigators on trade secret theft with the goal of increasing the number of successful investigations and prosecutions for violations of the Economic Espionage Act. These training events will target domestic law enforcement officers, prosecutors, and international partners. These events will include both a trade secret specific curriculum as well as broader intellectual property rights enforcement themes in which trade secret theft is a component

The National Intellectual Property Rights Coordination Center

The National Intellectual Property Rights Coordination Center will obtain leads regarding trade secret misappropriation through its "Report IP Theft" Initiative.

Theft of Motorola Trade Secrets

In November 2011, Customs and Border Protection officers at Chicago's O'Hare Airport stopped Hanjuan Jin, a former Motorola software engineer, while she was allegedly carrying 1,000 sensitive Motorola documents, $30,000 in cash, and a one-way ticket to China. Jin was in the process of traveling to China to turn over stolen trade secret information relating to mobile telecommunications to Kai Sun News Technology Co., also known as SunKaisens, and to the Chinese military.

The Department of Defense

The Department of Defense, through the Defense Security Service, will collect, analyze and report on threat information to cleared industries that support Department of Defense programs and the missions of other U.S. government departments and agencies. The Defense Security Service, in coordination with its partner agencies, will continue to provide advice to those cleared industry partners and deliver security training and education on counterintelligence. Through its annual report on trend analysis of threats targeting to U.S. defense technologies, the Defense Security Service will continue to communicate its analysis to industrial partners of the U.S. government.

The Defense Intelligence Agency will co-chair the National Critical Systems and Joint Technology Task Force with the FBI. This effort will continue to provide a collaborative forum to provide input into the counterintelligence efforts to protect critical and emerging technologies by Federal agencies

4. Improve Domestic Legislation

"Congress should make sure that no foreign company has an advantage over American manufacturing."

—President Barack Obama

In March 2011, the Administration directed federal agencies to review relevant existing Federal intellectual property laws. The goal of this review was to assess if current laws were effective in combatting infringement and protected intellectual property rights. Based on that review, the IPEC sent to Congress the Administration's 2011 White Paper on Intellectual Property Enforcement Legislative Recommendations (White Paper). This document recommended legislation to increase the statutory maximum for economic espionage (18 USC §1831) from 15 years in prison to at least 20 years. Additionally, the Administration also recommended legislation to direct the U.S. Sentencing Commission to consider increasing the U.S. Sentencing Guideline range for the theft of trade secrets and economic espionage, including trade secrets transferred or attempted to be transferred outside the U.S.

The White Paper supported the efforts of Members of Congress who worked in a bicameral and bipartisan manner to introduce legislation to improve the protection of trade secrets in the 112th Congress. President Obama signed two important pieces of legislation into law that will have an immediate and positive impact on prospective trade secret prosecutions:

- **Public Law112-236—*The Theft of Trade Secrets Clarification Act of 2012* (S. 3642)**, closed a loophole in the Economic Espionage Act that had allowed the theft of valuable trade secret source code.[10] This legislation was introduced by Senate Judiciary Chairman Senator Patrick Leahy in response to the Second Circuit decision in United States v. Aleynikov, 676 F.3d 71 (2d Cir. 2012), which overturned a verdict that found that the defendant violated 18 U.S.C. §1832(a) by stealing proprietary computer code, a trade secret, from his employer. This legislation was in line with the overall IPEC objective of protecting trade secrets from misappropriation.
- **Public Law 112-269—*The Foreign and Economic Espionage Penalty Enhancement Act of 2012* (H.R. 6029/S. 678)**, bolstered criminal penalties for economic espionage and directed the Sentencing commission to consider increasing offense levels for trade secret crimes.[11] Its passage is an important step in ensuring that penalties are commensurate with the economic harm inflicted on trade secret owners. The passage of this legislation could not have been achieved without the efforts of former House of Representatives Judiciary Chairman Representative Lamar Smith and retired Senator Herb Kohl.

The Administration will continue to ensure that U.S. laws are as effective as possible and that they reflect the seriousness of these crimes and the economic harm inflicted on victims. To supplement the proposals contained in the 2011 White Paper, the IPEC will initiate and coordinate a process, working

10. P.L. 112-236, The Theft of Trade Secrets Clarification Act, available at http://www.gpo.gov/fdsys/pkg/BILLS-112s3642enr/pdf/BILLS-112s3642enr.pdf

11. H.R. 6029EH, Foreign and Economic Espionage Penalty Enhancement Act, available at http://www.gpo.gov/fdsys/pkg/BILLS-112hr6029eh/pdf/BILLS-112hr6029eh.pdf

with appropriate Executive Branch agencies, to review existing Federal laws to determine if legislative changes are needed to enhance enforcement against trade secret theft. The initial review process will conclude within 120 days from the date of the release of this Strategy. The Administration, coordinated through the IPEC, will recommend to Congress any proposed legislative changes resulting from this review process.

Theft of Goldman Sachs Trade Secret

Goldman Sachs spent $500 million dollars developing computer source code to support its high frequency trading program. Sergey Aleynikov, a Goldman Sachs computer programmer, resigned from his job to work for a competitor, and on his final day of employment transferred this extremely valuable proprietary computer code to an external computer server. Mr. Aleynikov had also transferred thousands of proprietary computer code files to his home computers. Mr. Aleynikov was investigated by the FBI and prosecuted by the U.S. Attorney's Office of the Southern District of New York. He was sentenced to 97 months in Federal prison. In February 2012, his conviction was overturned by the Second Circuit based on the court's interpretation of the Economic Espionage Act. This loophole was fixed when President Obama signed Public Law 112-236 *The Theft of Trade Secrets Clarification Act of 2012* (S. 3642) on December 28, 2012

5. Public Awareness and Stakeholder Outreach

"What we can do—what America does better than anyone—is spark the creativity and imagination of our people."

—President Barack Obama

Highlighting can help mitigate the theft of trade secrets by encouraging all stakeholders, including the general public, to be aware of the detrimental effects of misappropriation on trade secret owners and the U.S. economy in general. The Administration will continue to conduct education and outreach efforts through the following actions:

- The Department of Commerce will leverage existing resources like www.stopfakes.gov to provide useful information for the private sector such as general information on the threat of trade secret theft, expanded country specific toolkits with information on how to protect trade secrets in priority markets, developments in the laws and enforcement practices of significant trading partners and webinars on trade secret theft awareness.
- U.S. Patent and Trademark Office and International Trade Administration will utilize current "road show" trainings to provide forums to educate the private sector, particularly small and medium sized businesses, regarding the economic implications of corporate and state sponsored trade secret theft.
- The FBI will continue its current public awareness campaign on bringing public attention to the threat posed to the U.S. from trade secret theft.[12]

12. Federal Bureau of Investigation, "Economic Espionage—How To Spot An Insider Threat", May 11, 2012, http://www.fbi.gov/news/stories/2012/may/insider_051112/insider_051112

Appendix

For more information trade secret theft please visit these websites:

- Department of Commerce STOPfakes.gov IPR training module includes an introduction to trade secrets (available at http://www.stopfakes.gov/business-tools/sme-module).
- Special 301 Report released by the U.S. Trade Representative summarizes troubling trends involving trade secrets and forced technology transfer. Pages 17-19 (available at http://www.ustr.gov).
- The Department of State (available at http://www.state.gov/e/eb/tpp/ipe/).
- DOJ National Security Division (available at http://www.justice.gov/nsd/).
- DOJ Criminal Division—Computer Crimes and Intellectual Property Section (available at http://www.justice.gov/criminal/cybercrime/).
- FBI Counterintelligence Division (available at http://www.fbi.gov/about-us/investigate/counterintelligence/economic-espionage).
- National Intellectual Property Rights Coordination Center (available at http://www.iprcenter.gov/).
- The Office of the National Counterintelligence Executive (available at http://www.ncix.gov/issues/economic/index.php).
- The Department of Defense – Defense Security Service (available at http://www.dss.mil/documents/ci/Insider-Threats.pdf).
- Create.org study that includes recommendations for companies operating in foreign countries to mitigate the risk of trade secret theft (available at http://www.create.org/views-blog/trade-secret-theft-managing-growing-threat-supply-chains).
- The World Intellectual Property Organization (WIPO) has more trade secret information specifically designed for small and medium-sized enterprises (available at http://www.wipo.int/sme/en/ip_business/trade_secrets/trade_secrets.htm).

Annex

- **ANNEX A:** U.S. Patent and Trademark Office Overview of U.S. Trade Secret Laws and Changed Landscape
- **ANNEX B:** Summary of Department of Justice Trade Secret Theft Cases
- **ANNEX C:** 2011 Office of the National Counterintelligence Executive Report
- **ANNEX D:** 2012 Department of Defense – Defense Security Service Report

Administration Strategy on
Mitigating the Theft of U.S. Trade Secrets

Overview of U.S. Law and Changed Landscape

Overview of U.S. Law

Under U.S. law, trade secrets comprise commercially valuable information not generally known or readily ascertainable to the public that are subject to reasonable measures to maintain its confidentiality. Typical examples include confidential formulas, manufacturing techniques, and customer lists. Trade secret law offers protection from trade secret "misappropriation": the unauthorized acquisition, use, or disclosure of such secrets obtained by some improper means. But discovery of a trade secret by fair, lawful methods, such as reverse engineering or independent development, is permitted.

In the United States, civil private enforcement of trade secret protection is primarily a state law matter. However, the federal Economic Espionage Act of 1996 criminalizes some forms of trade secret theft and also empowers the U.S. Attorney General to initiate civil public enforcement proceedings. State law protection of trade secrets has its origin in the common law. These common law principles were first gathered and summarized in the 1939 Restatement (First) of Torts, and later in the 1995 Restatement (Third) of Unfair Competition. Beginning in the 1980s, states began to adopt provisions set forth in the Uniform Trade Secrets Act (UTSA) as a statutory basis for trade secret law. The UTSA, and various state measures provide for injunctive relief, damages, and in some instances attorney's fees as remedies to trade secret misappropriation. Under the UTSA, injunctive relief may be granted for "[a]ctual or threatened misappropriation." An injunction will be terminated when the trade secret ceases to be a trade secret. However, the injunction may be "continued for an additional reasonable period of time in order to eliminate commercial advantage that otherwise would be derived from the misappropriation," or "head start," that the misappropriator gained over one who set out to discover the trade secret through legitimate means such as reverse engineering. The UTSA also provides for recovery of damages, calculated by the actual loss caused by the misappropriation and any separate unjust enrichment. Exemplary damages up to twice that amount may be awarded in the case of willful and malicious misappropriation. Under the UTSA, a court may award attorney's fees to the prevailing party in instances of bad faith or willful and malicious misappropriation.

A controversial and regularly recurring issue in U.S. civil trade secret law is the doctrine of inevitable disclosure. Courts accepting the doctrine reason that an employee who learns a trade secret on the job and then leaves to work for a competitor may "inevitably" disclose the trade secret. To address this perceived problem, these courts frequently enter injunctions prohibiting such employees from working for competitors because the inevitable disclosure of the trade secrets would constitute misappropriation. The practical effect of adopting this principle is that, even absent a formal non-compete agreement, employers may be able to enjoin former employees from working for competitors, because the employee is bound by an implied covenant. Not all courts have adopted this principle.

The federal government currently protects trade secrets through both the criminal and the public civil enforcement sections of the Economic Espionage Act of 1996 ("EEA"), which is codified in 18 U.S.C. §§ 1831-39. Under section 1831, which addresses the more severe crime of economic espionage, it is a felony to knowingly steal or misappropriate a trade secret to "benefit any foreign government, foreign instrumentality, or foreign agent." Section 1832 addresses the theft of trade secrets "related to or included in a product that is produced for or placed in interstate or foreign commerce." It makes it a crime to knowingly steal or misappropriate a trade secret "to the economic benefit of anyone other

than the owner thereof" if the accused party "intend[s] or know[s] that the offense will . . . injure any owner of that trade secret."

The EEA applies to trade secret violations committed both domestically and outside the United States. However, it is only applicable to conduct occurring outside of the United States if the offender is a U.S. citizen or permanent resident alien or an organization organized under U.S. law, or if an act in furtherance of the offense was committed in the United States. The Attorney General may, in a public civil enforcement action, obtain injunctive relief to prevent further violations of the EEA, but the EEA does not provide a private civil right of action.

CHANGED LANDSCAPE

Current literature on trade secret protection points to significant shifts in the nature of trade secret theft and the resulting challenges presented. The nature, protection, and enforcement of a trade secret are distinct from other forms of intellectual property. Unlike other forms of intellectual property, once disclosed publicly, the property right itself ceases to exist. Protection is provided to trade secrets only when steps are taken by the owner to maintain the secrecy of the information. Liability is not imposed for mere theft absent a showing of reasonable efforts to maintain secrecy; continual vigilance is required. What constitutes reasonable efforts is often a pivotal issue in trade secret litigation and particularly important in the digital environment.

The technologies that have made the digital revolution possible also present significant threats to the protection of intellectual property, and trade secrets in particular. Advancements in technology, increased mobility, globalization, and the anonymous/pseudonymous nature of the internet are all working together to create growing challenges in protecting trade secrets. This technology has resulted in companies needing to re-evaluate what constitutes adequate protection of trade secrets in digital format and has impacted the manner in which the trade secrets are stolen. The same technologies that have been a catalyst to the economic growth of both businesses and economies have created a new and threatening environment for the protection of vital assets. These new technologies make it easier to store, access, disseminate, and publish confidential information, thereby enhancing the likelihood that a trade secret may be lost.

The internet in particular has become an innovation that can significantly affect trade secrets. Once a trade secret has been posted on the internet, it has the potential to become "generally known" within a short time period, thereby losing its status as a trade secret. It is in the best interest of the owner of the proprietary information to have the trade secret removed as quickly as possible. Many courts have taken the position that the publication of a trade secret on the internet results in the loss of the secret status of the information, making the claim unenforceable. Given the incomplete remedial nature of removing information from the internet, prevention from disclosure is the strongest weapon and immediate removal should be sought if prevention failed.

In the ONCIX 2011 *Report to Congress on Foreign Economic Collection and Industrial Espionage* there is a shift in focus from previous reports and the threat from cyberspace is highlighted. The report notes that:

> Nearly all business records, research results, and other sensitive economic or technology-related information now exist primarily in digital form. Cyberspace makes it possible for foreign collectors to gather enormous quantities of information quickly and with little risk, whether via remote exploitation of victim's computer networks,

downloads of data to external media devices, or e-mail messages transmitting sensitive information.

The pace of change in information and communications technology is projected to increase, bringing additional pressures on maintaining both the secrecy and ownership of trade secrets. The sharing of resources through cloud computing will facilitate a workforce even more mobile than today. Technologies providing greater access to information anytime and anywhere will increasingly rely on the internet, and present new challenges to companies seeking to protect information transmitted by, or contained on, mobile devices. This mobility will contribute to a future in which the defense provided by national borders to trade secret theft is diminished. Technology, however, can also provide tools to prevent and combat theft of electronic information. Through new technology, companies can better determine when and where confidential information has been accessed, copied, distributed, destroyed, etc. Companies can also better monitor the source of information that was misappropriated; for example, digital watermarking can assist in identifying the source of information. The threat to U.S. business of economic espionage coordinated by foreign governments, as opposed to industrial espionage, is of particular concern. Such acts would not only deprive U.S. companies of their valuable information, often to the benefit of foreign competitors who may receive that information from the foreign government, but countering the vast intelligence resources that a foreign government can utilize for such purpose may be a particular challenge for individual companies.

Summary of Department of Justice Economic Espionage and Trade Secret Criminal Cases January 2009 – Present (Updated January 2013)[1]

***Trade Secrets to China* –** On Nov. 30, 2012, a former General Motors engineer and her husband were convicted by a federal jury today in Detroit for conspiring to steal hybrid technology trade secrets from GM with the intent to use them in a joint venture with an automotive competitor in China. Shanshan Du and her husband, Yu Qin were convicted of unlawful possession of trade secrets. The evidence at trial showed that from December 2003 through May 2006, the defendants conspired to steal GM's trade secret information. Du, while employed with GM's hybrid vehicle technology group, provided GM trade secret information relating to hybrid vehicles to her husband, Qin, for the benefit of their private company, Millennium Technology International Inc. (MTI), which the defendants jointly owned and operated. Approximately five days after Du was offered a severance agreement by GM in January 2005, she copied more than 16,000 GM files, including trade secret documents, to an external computer hard drive used for MTI business. A few months later, Qin moved forward on a business venture to provide hybrid vehicle technology to Chery Automobile, an automotive manufacturer based in China and a competitor of GM. This investigation was conducted by the FBI.

Trade Secrets to South Korea – On Oct. 18, 2012, South Korea-based Kolon Industries Inc. and several of its executives and employees were indicted in the Eastern District of Virginia for allegedly engaging in a multi-year campaign to steal trade secrets related to DuPont's Kevlar para-aramid fiber and Teijin Limited's Twaron para-aramid fiber. The indictment seeks forfeiture of at least $225 million in proceeds from the alleged theft of trade secrets from Kolon's competitors and charges Kolon with one count of conspiring to convert trade secrets, four counts of theft of trade secrets and one count of obstruction of justice. Kolon makes a product called Heracron, which is a recent entrant into the para-aramid fiber market as a competitor to products called Kevlar and Twaron. Para-aramid fibers are used to make, for example, body armor, fiber optic cables and automotive and industrial products. Kevlar is produced by E. I. du Pont de Nemours and Company (DuPont), one of the largest chemical companies in the United States. For decades, Kevlar has competed against Twaron, a para-aramid fiber product produced by Teijin Limited, one of the largest chemical companies in Japan. According to the indictment, from July 2002 through February 2009, Kolon allegedly sought to improve its Heracron product by targeting current and former employees at DuPont and Teijin and hiring them to serve as consultants, then asking these consultants to reveal information that was confidential and proprietary. The indictment alleges that in July 2002, Kolon obtained confidential information related to an aspect of DuPont's manufacturing process for Kevlar, and within three years Kolon had replicated it. This successful misappropriation of DuPont's confidential information, the indictment alleges, spurred Kolon leadership to develop a multi-

[1] Available at http://www.justice.gov/nsd/docs/export-case-fact-sheet.pdf

phase plan in November 2005 to secure additional trade secret information from its competitors, by targeting people with knowledge of both pre-1990 para-aramid technology and post-1990 technologies. Kolon is alleged to have retained at least five former DuPont employees as consultants. Kolon allegedly met with these people individually on multiple occasions from 2006 through 2008 to solicit and obtain sensitive, proprietary information that included details about DuPont's manufacturing processes for Kevlar, experiment results, blueprints and designs, prices paid to suppliers and new fiber technology. This investigation was conducted by the FBI.

Military Technical Data and Trade Secrets to China – On Sept. 26, 2012, Sixing Liu, aka "Steve Liu," a native of China with a PhD in electrical engineering who worked as a senior staff engineer for Space & Navigation, a New Jersey-based division of L-3 Communications, was convicted in the District of New Jersey of exporting sensitive U.S. military technology to China, stealing trade secrets and lying to federal agents. The jury convicted Liu of nine of 11 counts of an April 5, 2012 second superseding indictment, specifically six counts of violating the Arms Export Control Act, one count of possessing stolen trade secrets in violation of the Economic Espionage Act, one count of transporting stolen property, and one count of lying to federal agents. The jury acquitted Liu on two counts of lying to federal agents. According to documents filed in the case and evidence presented at trial, in 2010, Liu stole thousands of electronic files from his employer, L-3 Communications, Space and Navigation Division. The stolen files detailed the performance and design of guidance systems for missiles, rockets, target locators, and unmanned aerial vehicles. Liu stole the files to position and prepare himself for future employment in China. As part of that plan, Liu delivered presentations about the technology at several Chinese universities, the Chinese Academy of Sciences, and conferences organized by Chinese government entities. However, Liu was not charged with any crimes related to those presentations. On Nov. 12, 2010, Liu boarded a flight from Newark to China. Upon his return to the United States on Nov. 29, 2010, agents found Liu in possession of a non-work-issued computer found to contain the stolen material. The following day, Liu lied to ICE agents about the extent of his work on U.S. defense technology. The State Department later verified that several of the stolen files on Liu's computer contained technical data that relates to defense items listed on the United States Munitions List. The jury also heard testimony that Liu's company trained him about the United States' export control laws and told him that most of the company's products were covered by those laws. Liu was first arrested on March 8, 2011, in Chicago on a complaint in the District of New Jersey charging him with one count of exporting defense-related technical data without a license. The investigation was conducted by the FBI, ICE and CBP.

Theft of Trade Secrets for Potential Use in China – On Sept. 19, 2012, Chunlai Yang, a former senior software engineer for Chicago-based CME Group, Inc., pleaded guilty in the Northern District of Illinois to two counts of theft of trade secrets for stealing source code and other proprietary information while at the same time pursuing plans to improve an electronic trading exchange in China. Yang admitted that he downloaded more than 10,000 files containing CME computer source code that made up a substantial part of the operating systems for the Globex electronic trading platform. The government maintains that the potential loss was between $50 million and $100 million. Yang began working for CME Group in 2000 and was a senior software engineer at the time of his arrest. Between late 2010, and June 30, 2011, Yang downloaded more than 10,000 computer files containing CME computer source code from CME's secure internal computer system to his CME-issued work computer. He then transferred many of these files from his work computer to his personal USB flash drives, and then transferred many of these files from his flash drives to his personal computers and hard drives at his home. Yang also admitted that he downloaded thousands of others CME files. Yang admitted that he and two unnamed business partners

developed plans to form a business referred to as the Tongmei (Gateway to America) Futures Exchange Software Technology Company (Gateway), whose purpose was to increase the trading volume at the Zhangjiagang, China, chemical electronic trading exchange (the Zhangjiagang Exchange.) The Zhangjiagang Exchange was to become a transfer station to China for advanced technologies companies around the world. Yang expected that Gateway would provide the exchange with technology through written source code to allow for high trading volume, high trading speeds, and multiple trading functions. Yang was indicted on Sept. 28, 2011. This investigation was conducted by the FBI.

Trade Secrets to China – On Sept. 4, 2012, Chinese citizens Ji Li Huang and Xiao Guang Qi were charged in a criminal complaint in the Western District of Missouri with attempting to purchase stolen trade secrets stolen from Pittsburgh Corning for the purpose of opening a plant in China to compete with Pittsburgh Corning. Pittsburgh Corning, headquartered in Pittsburgh, manufactures various grades or densities of cellular glass insulation sold under the trade name FOAMGLAS and had recently made technological advances in the formulation and manufacturing process of FOAMGLAS insulation. According to the complaint, the defendants attempted to pay $100,000 to an FBI cooperating source for confidential and proprietary information stolen from Pittsburgh Corning. The defendants were arrested on Sept. 2, 2012 after meeting with the confidential source who provided them documents that were purportedly stolen trade secrets from the company. The investigation was conducted by the FBI.

Motorola Trade Secrets to China – On Aug. 29, 2012, Hanjuan Jin, a former software engineer for Motorola, was sentenced in the Northern District of Illinois to four years in prison for stealing trade secrets from Motorola, specifically Motorola's proprietary iDEN telecommunications technology, for herself and for Sun Kaisens, a company that developed products for the Chinese military. According to court documents filed in the case, Motorola spent more than $400 million researching and developing iDEN technology in just a matter of years. On Feb. 8, 2012, Jin was found guilty of three counts of stealing trade secrets. Jin, a naturalized U.S. citizen born in China, possessed more than 1,000 electronic and paper Motorola proprietary documents when she was stopped by U.S. authorities at Chicago's O'Hare International Airport as she attempted to travel to China on Feb. 28, 2007. The judge presiding over the case found her not guilty of three counts of economic espionage for the benefit of the government of China and its military. According to the evidence at trial, Jin began working for Motorola in 1998, and took medical leave in February 2006. Between June and November 2006, while still on sick leave, Jin pursued employment in China with Sun Kaisens, a Chinese telecommunications firm that developed products for the Chinese military. Between November 2006 and February 2007, Jin returned to China and did work for Sun Kaisens on projects for the Chinese military. On Feb. 15, 2007, Jin returned to the United States from China and reserved a flight to China scheduled to depart on Feb. 28, 2007. Jin advised Motorola that she was ready to return to work at Motorola, without informing Motorola that she planned to return to China to work for Sun Kaisens. On Feb. 26, 2007, she returned to Motorola, and accessed hundreds of technical documents belonging to Motorola on its secure internal computer network. As she attempted to depart from Chicago to China, authorities seized numerous materials, some of which provided a description of communication feature that Motorola incorporates into its telecommunications products. Authorities also recovered classified Chinese documents describing telecommunication projects for the Chinese military. Jin was charged with theft of trade secrets in an April 1, 2008 indictment. A superseding indictment returned on Dec. 9, 2008 charged her with economic espionage. The investigation was conducted by the FBI, with assistance from U.S Customs and Border Protection.

Trade Secrets to Competitors in China – On May 7, 2012, an indictment returned in the District of Utah in April 2012 was unsealed charging two people and two companies with theft of trade secrets, wire fraud, and conspiracy to commit wire fraud in connection with the alleged theft of trade secrets from Orbit Irrigation Products, an irrigation company headquartered in Utah. The defendants are Janice Kuang Capener and Luo Jun, both citizens of China, as well as Sunhills International LLC, a California company established by Capener; and Zhejiang Hongchen Irrigation Equipment Co., LTD, a Chinese company under contract with Orbit. According to court documents, Capener worked at Orbit from June 2003 through Nov. 1, 2009, including serving chief of operations at Orbit's manufacturing plant in Ningbo, China. Capener allegedly stole Orbit trade secrets relating to sales and pricing and used that information for herself and others to the detriment of Orbit. Capener also allegedly worked with Jun, Sunhills International and Zhejiang Hongchen Irrigation Equipment to devise a scheme to undermine Orbit's position in the marketplace using illegally obtained proprietary pricing information. Capener and Jun were arrested on May 4, 2012. This case was investigated by the FBI.

Military Technical Data and Trade Secrets to China – On April 5, 2012, a second superseding indictment was returned in the District of New Jersey against Sixing Liu, aka "Steve Liu," a native of China with a PhD in electrical engineering who worked as a senior staff engineer for Space & Navigation, a New Jersey-based division of L-3 Communications, from March 2009 through Nov. 2010. The superseding indictment charged Liu with six counts of illegally exporting defense articles / technical data to China, one count of possessing stolen trade secrets, one count of interstate transportation of stolen property, and three counts of false statements to federal agents. Liu, of Deerfield, Ill., was first arrested on March 8, 2011 in Chicago on a criminal complaint filed in the District of New Jersey charging him with one count of exporting defense-related technical data without a license. At Space & Navigation, Liu allegedly worked on precision navigation devices for rocket launchers, missile launch systems, field artillery, smart munitions, and other components being used by and prepared for the U.S. Department of Defense. Liu was never approved to present information related to Space & Navigation's programs or the technology underlying its programs to any outside person or audience. In 2009 and again in 2010, the indictment alleges that Liu traveled to China where he attended and delivered presentations on export-restricted technical data at technology conferences sponsored by Chinese government entities, including the 863 Program. Before leaving for the 2010 conference in China, Liu allegedly downloaded some 36,000 computer files from Space & Navigation to his personal laptop. Upon his return to the United States in November 2010,
U.S. Customs inspectors found him to be in possession of a laptop computer that contained hundreds of documents related to the company's projects, as well as images of Liu making a presentation at a technology conference sponsored by the PRC government. Many of the documents on his computer were marked as containing sensitive proprietary company information and/or export-controlled technical data. The State Department verified that information on the Liu's computer was export-controlled technical data that relates to defense items on the U.S. Munitions List. The investigation was conducted by the FBI and ICE.

DuPont Trade Secrets to China – On March 2, 2012, former DuPont scientist Tze Chao pleaded guilty in the Northern District of California to conspiracy to commit economic espionage, admitting that he provided trade secrets concerning DuPont's proprietary titanium dioxide manufacturing process to companies he knew were controlled by the government of the People's Republic of China (PRC). On Feb. 7, 2012, a grand jury in San Francisco returned a superseding indictment charging Chao and four other individuals, as well as five companies, with economic espionage and theft of trade secrets for their roles in a long-running effort to obtain U.S. trade secrets from DuPont for the benefit of companies

controlled by the PRC. The five individuals named in the indictment were Walter Liew, his wife Christina Liew, Hou Shengdong, Robert Maegerle, and Tze Chao. The five companies named as defendants are Pangang Group Company Ltd; Pangang Group Steel Vanadium Industry Company Ltd; Pangang Group Titanium Industry Company Ltd., Pangang Group International Economic & Trading Co; and USA Performance Technology, Inc. According to the superseding indictment, the PRC government identified as a priority the development of chloride-route titanium dioxide (TiO2) production capabilities. TiO2 is a commercially valuable white pigment with numerous uses, including coloring paint, plastics and paper. To achieve that goal, companies controlled by the PRC government, specifically the Pangang Group companies named in the indictment, and employees of those companies conspired and attempted to illegally obtain TiO2 technology that had been developed over many years of research and development by DuPont. The Pangang Group companies were aided in their efforts by individuals in the United States who had obtained TiO2 trade secrets and were willing to sell those secrets for significant sums of money. Defendants Walter Liew, Christina Liew, Robert Maegerle and Tze Chao allegedly obtained and possessed TiO2 trade secrets belonging to DuPont. Each of these individuals allegedly sold information containing DuPont TiO2 trade secrets to the Pangang Group companies for the purpose of helping those companies develop large-scale chloride route TiO2 production capability in the PRC, including a planned 100,000 ton TiO2 factory at Chongqing, PRC. The Liews, USAPTI, and one of its predecessor companies, Performance Group, entered into contracts worth in excess of $20 million to convey TiO2 trade secret technology to Pangang Group companies. The Liews allegedly received millions of dollars of proceeds from these contracts. The proceeds were wired through the United States, Singapore and ultimately back into several bank accounts in the PRC in the names of relatives of Christina Liew. The object of the defendants' conspiracy was to convey DuPont's secret chloride-route technology to the PRC companies for the purpose of building modern TiO2 production facilities in the PRC without investing in time-consuming and expensive research and development. DuPont invented the chloride-route process for manufacturing TiO2 in the late-1940s and since then has invested heavily in research and development to improve that production process. The global titanium dioxide market has been valued at roughly $12 billion, and DuPont has the largest share of that market. This investigation was conducted by the FBI.

Trade Secrets to U.S. Subsidiary of Chinese Company – On Jan. 17, 2012, Yuan Li, a former research chemist with the global pharmaceutical company Sanofi-Aventis, pleaded guilty in the District of New Jersey to stealing Sanofi's trade secrets and making them available for sale through Abby Pharmatech, Inc., the U.S. subsidiary of a Chinese chemicals company. According to court documents, Li worked at Sanofi headquarters in Bridgewater, N.J., from August 2006 through June 2011, where she assisted in the development of several compounds (trade secrets) that Sanofi viewed as potential building blocks for future drugs. While employed at Sanofi, Li was a 50 percent partner in Abby, which sells and distributes pharmaceuticals. Li admitted that between Oct. 2008 and June 2011, she accessed internal Sanofi databases and downloaded information on Sanofi compounds and transferred this information to her personal home computer. She also admitted that she made the stolen compounds available for sale on Abby's website. This investigation was conducted by the FBI.

Dow Trade Secrets to China – On Jan. 12, 2012, Wen Chyu Liu, aka David W. Liou, a former research scientist at Dow Chemical Company in Louisiana, was sentenced in the Middle District of Louisiana to 60 months in prison, two years supervised release, a $25,000 fine and was ordered to forfeit $600,000. Liu was convicted on Feb. 7, 2011 of one count of conspiracy to commit trade secret theft for stealing trade secrets from Dow and selling them to companies in China, and he was also convicted of one count of perjury. According to the evidence presented in court, Liou came to the United States from China for

graduate work. He began working for Dow in 1965 and retired in 1992. Dow is a leading producer of the elastomeric polymer, chlorinated polyethylene (CPE). Dow's Tyrin CPE is used in a number of applications worldwide, such as automotive and industrial hoses, electrical cable jackets and vinyl siding. While employed at Dow, Liou worked as a research scientist on various aspects of the development and manufacture of Dow elastomers, including Tyrin CPE. The evidence at trial established that Liou conspired with at least four current and former employees of Dow's facilities in Plaquemine, Louisiana, and in Stade, Germany, who had worked in Tyrin CPE production, to misappropriate those trade secrets in an effort to develop and market CPE process design packages to Chinese companies. Liou traveled throughout China to market the stolen information, and he paid current and former Dow employees for Dow's CPE-related material and information. In one instance, Liou bribed a then-employee at the Plaquemine facility with $50,000 in cash to provide Dow's process manual and other CPE-related information. The investigation was conducted by the FBI.

Dow and Cargill Trade Secrets to China – On Dec. 21, 2011, Kexue Huang, a Chinese national and former resident of Indiana, was sentenced to 87 months in and three years supervised release on charges of economic espionage to benefit a foreign university tied to the People's Republic of China (PRC) and theft of trade secrets. On Oct. 18, 2011, Huang pleaded guilty in the Southern District of Indiana to these charges. In July 2010, Huang was charged in the Southern District of Indiana with misappropriating and transporting trade secrets to the PRC while working as a research scientist at Dow AgroSciences LLC. On Oct. 18, 2011, a separate indictment in the District of Minnesota charging Huang with stealing a trade secret from a second company, Cargill Inc., was unsealed. From January 2003 until February 2008, Huang was employed as a research scientist at Dow. In 2005, he became a research leader for Dow in strain development related to unique, proprietary organic insecticides marketed worldwide. Huang admitted that during his employment at Dow, he misappropriated several Dow trade secrets. According to plea documents, from 2007 to 2010, Huang transferred and delivered the stolen Dow trade secrets to individuals in Germany and the PRC. With the assistance of these individuals, Huang used the stolen materials to conduct unauthorized research to benefit foreign universities tied to the PRC. Huang also admitted that he pursued steps to develop and produce the misappropriated Dow trade secrets in the PRC. After Huang left Dow, he was hired in March 2008 by Cargill, an international producer and marketer of food, agricultural, financial and industrial products and services. Huang worked as a biotechnologist for Cargill until July 2009. Huang admitted that during his employment with Cargill, he stole one of the company's trade secrets – a key component in the manufacture of a new food product, which he later disseminated to another person, specifically a student at Hunan Normal University in the PRC. According to the plea agreement, the aggregated loss from Huang's conduct exceeds $7 million but is less than $20 million. This investigation was conducted by the FBI.

Trade Secrets to India – On Nov. 14, 2011, Prabhu Mohapatra was arrested on a criminal complaint in the District of Utah (filed on Nov. 10, 2011) charging him with stealing proprietary information from his employer, a Utah scientific company, and providing it to a relative in India who was starting up a competing company. According to the charges, Mohapatra worked as a senior scientist for Frontier Scientific, Inc., a company that makes large pure quantities of an organic chemical, 2,2'-dipyrromethane, that has several applications, including as an ingredient in new drugs, as well as in solar cells and batteries. The complaint alleges that Mohapatra emailed proprietary information from Frontier Scientific about the chemical to his brother-in-law in India, who was setting up an unregistered, competing company called Medchemblox. The complaint further alleges that Mohapatra had a financial interest in Medchemblox. This investigation was conducted by FBI.

Trade Secrets to Foreign Government – On Aug. 30, 2011, Elliot Doxer, of Brookline, Mass., pleaded guilty in the District of Massachusetts to one count of foreign economic espionage for providing trade secrets over an 18-month period to an undercover FBI agent posing as an Israeli intelligence officer. Neither the government of Israel nor anyone acting on its behalf committed any offense under U.S. laws in this case. Doxer was a former employee of Akamai Technologies, Inc., who in June 2006 sent an e-mail to the Israeli consulate in Boston stating that he worked in Akamai's finance department and was willing to provide information that might help Israel. In Sept. 2007, an undercover FBI agent posing as an Israeli intelligence officer spoke to Doxer and established a "dead drop" where the agent and Doxer could exchange information. From Sept. 2007 through March 2009, Doxer visited the dead drop at least 62 times to leave information, retrieve communications or check for new communications. Doxer provided the undercover agent with Akamai customer lists, employee lists, contract information and other trade secrets. He was arrested on Oct. 6, 2010 on a complaint charging him with wire fraud. That charge was dismissed as part of the plea agreement. Doxer was ultimately sentenced on Dec. 19, 2011 to six months in prison and two years supervised release. The case was investigated by the FBI.

Wire Fraud in Trade Secrets Case Involving China – On April 6, 2011, Yan Zhu, a Chinese citizen in the U.S. on a work visa, was convicted in the District of New Jersey on seven counts of wire fraud in connection with his scheme to steal confidential and proprietary business information relating to computer systems and software with environmental applications from his New Jersey employer. He was acquitted on the charge of conspiracy to steal trade secrets and two counts of unauthorized transmission of trade secrets in interstate or foreign commerce. April 10, 2009, Zhu was arrested on charges of theft of trade secrets, conspiracy, wire fraud, and theft of honest services fraud in connection with a plot to steal software from his former U.S. employer and sell a modified version to the Chinese government after he was fired. Zhu was employed as a senior environmental engineer from May of 2006 until his termination in July of 2008. Zhu worked for a comprehensive multi-media environmental information management portal that developed a proprietary software program for the Chinese market which allows users to manage air emissions, ambient water quality, and ground water quality. Zhu was sentenced on Jan. 5, 2012 to three years of probation and a $700 special assessment. This investigation was conducted by the FBI.

Valspar Trade Secrets to China – On Dec. 8, 2010, David Yen Lee, a former chemist for Valspar Corporation, a Chicago paint manufacturing company, was sentenced in the Northern District of Illinois to 15 months in prison for stealing trade secrets involving numerous formulas and other proprietary information valued up to $20 million as he prepared to go to work for a competitor in China. Lee, formerly a technical director in Valspar Corp's architectural coatings group since 2006, pleaded guilty in Sept. 2010 to using his access to Valspar's secure internal computer network to download approximately 160 original batch tickets, or secret formulas for paints and coatings. Lee also obtained raw materials information, chemical formulas and calculations, sales and cost data, and other internal memoranda, product research, marketing data, and other materials from Valspar. Lee admitted that between September 2008 and February 2009, he had negotiated employment with Nippon Paint, in Shanghai, China and accepted employment with Nippon as vice president of technology and administrator of research and development. Lee was scheduled to fly from Chicago to Shanghai on March 27, 2009. He did not inform Valspar that he had accepted a job at Nippon until he resigned on March 16, 2009. Between November 2008 and March 2009, Lee downloaded technical documents and materials belonging to Valspar, including the paint formula batch tickets. He further copied certain downloaded files to external thumb drives to store the data, knowing that he intended to use the confidential

information belong to Valspar for his own benefit. There was no evidence that he actually disclosed any of the stolen trade secrets. This investigation was conducted by the FBI.

Ford Motor Company Trade Secrets to China – On Nov. 17, 2010, Yu Xiang Dong, aka Mike Yu, a product engineer with Ford Motor Company pleaded guilty in the Eastern District of Michigan to two counts of theft of trade secrets. According to the plea agreement, Yu was a Product Engineer for Ford from 1997 to 2007 and had access to Ford trade secrets, including Ford design documents. In December 2006, Yu accepted a job at the China branch of a U.S. company. On the eve of his departure from Ford and before he told Ford of his new job, Yu copied some 4,000 Ford documents onto an external hard drive, including sensitive Ford design documents. Ford spent millions of dollars and decades on research, development, and testing to develop and improve the design specifications set forth in these documents. On Dec. 20, 2006, Yu traveled to the location of his new employer in Shenzhen, China, taking the Ford trade secrets with him. On Jan. 2, 2007, Yu emailed his Ford supervisor from China and informed him that he was leaving Ford's employ. In Nov. 2008, Yu began working for Beijing Automotive Company, a direct competitor of Ford. On Oct. 19, 2009, Yu returned to the U.S. Upon his arrival, he was arrested. At that time, Yu had in his possession his Beijing Automotive Company laptop computer. Upon examination of that computer, the FBI discovered that 41 Ford system design specifications documents had been copied to the defendant's Beijing Automotive Company work computer. The FBI also discovered that each of those design documents had been accessed by Yu during the time of his employment with Beijing Automotive Company. Yu was ultimately sentenced to 70 months in prison in April 2011. This case was investigated by the FBI.

DuPont Trade Secrets to China – On Oct. 26, 2010, Hong Meng, a former research chemist for DuPont, was sentenced in the District of Delaware to 14 months in prison and $58,621 in restitution for theft of trade secrets. Meng pleaded guilty on June 8, 2010. Meng was involved in researching Organic Light Emitting Diodes (OLED) during his tenure at DuPont. In early 2009, DuPont's OLED research efforts resulted in the development of a breakthrough chemical process (trade secret) that increased the performance and longevity of OLED displays. In the Spring of 2009, while still employed at DuPont and without DuPont's permission or knowledge, Meng accepted employment as a faculty member at Peking University (PKU) College of Engineering, Department of Nanotechnology in Beijing, China, and thereafter began soliciting funding to commercialize his OLED research at PKU. In June 2009, he emailed to his PKU account the protected chemical process from DuPont. He also downloaded the chemical process from his DuPont work computer to a thumb drive which he uploaded to his personal computer. In August 2009, he mailed a package containing 109 samples of DuPont intermediate chemical compounds to a colleague at Northwestern University and instructed his colleague at Northwestern to forward the materials to Meng's office at PKU. Eight of the 109 samples were trade secret chemical compounds. Meng also made false statements to the FBI when questioned about these samples. This investigation was conducted by the FBI.

GM Trade Secrets to China – On July 22, 2010, an indictment returned in the Eastern District of Michigan charging Yu Qin and his wife Shanshan Du, both of Troy, Michigan, was unsealed. The indictment charged the defendants with conspiracy to possess trade secrets without authorization, unauthorized possession of trade secrets and wire fraud. According to the indictment, from December 2003 through May 2006, the defendants conspired to possess trade secret information of General Motors Company relating to hybrid vehicles, knowing that the information had been stolen, converted, or obtained without authorization. The indictment alleges that Du, while employed with GM, provided GM trade secret information relating to hybrid vehicles to her husband, Qin, for his benefit and for the benefit of a company, Millennium Technology International Inc. (MTI), which the defendants owned and

operated. Five days after Du was offered a severance agreement by GM in January 2005, she copied thousands of GM documents, including trade secret documents, to a computer hard drive used for MTI business. A few months later, Qin moved forward on a new business venture to provide hybrid vehicle technology to Chery Automobile, a Chinese automotive manufacturer based in China and a competitor of GM. The indictment further alleges that in May 2006, the defendants possessed GM trade secret information without authorization on several computer and electronic devices located in their residence. Based on preliminary calculations, GM estimates that the value of the stolen GM documents is over $40 million. This investigation was conducted by the FBI.

Economic Espionage / Theft of Space Shuttle and Rocket Secrets for China – On Feb. 11, 2010 former Rockwell and Boeing engineer Dongfan "Greg" Chung was sentenced to 188 months imprisonment and three years supervised release after his July 16, 2009 conviction in the Central District of California. Chung was convicted of charges of economic espionage and acting as an illegal agent of the People's Republic of China (PRC), for whom he stole restricted technology and Boeing trade secrets, including information related to the Space Shuttle program and the Delta IV rocket. According to the judge's ruling, Chung served as an illegal agent of China for more than 30 years and kept more than 300,000 pages of documents reflecting Boeing trade secrets stashed in his home as part of his mission of steal aerospace and military trade secrets from Boeing to assist the Chinese government. Chung sent Boeing trade secrets to the PRC via the mail, via sea freight, via the Chinese consulate in San Francisco, and via a Chinese agent named Chi Mak. On several occasions, Chung also used the trade secrets that he misappropriated from Boeing to prepare detailed briefings that he later presented to Chinese officials in the PRC. Chung was originally arrested on Feb. 11, 2008, in Southern California after being indicted on eight counts of economic espionage, one count of conspiracy to commit economic espionage, one count of acting as an unregistered foreign agent, one count of obstruction of justice, and three counts of making false statements to the FBI. The investigation was conducted by the FBI and NASA.

COUNTERINTELLIGENCE SECURITY

OFFICE OF THE NATIONAL COUNTERINTELLIGENCE EXECUTIVE

FOREIGN SPIES STEALING US ECONOMIC SECRETS IN CYBERSPACE

Report to Congress on Foreign Economic Collection and Industrial Espionage, 2009-2011

October 2011

Table of Contents

Annex A

Annex B

List of Text Boxes

List of Charts

Executive Summary

Foreign economic collection and industrial espionage against the United States represent significant and growing threats to the nation's prosperity and security. Cyberspace—where most business activity and development of new ideas now takes place—amplifies these threats by making it possible for malicious actors, whether they are corrupted insiders or foreign intelligence services (FIS), to quickly steal and transfer massive quantities of data while remaining anonymous and hard to detect.

US Technologies and Trade Secrets at Risk in Cyberspace

Foreign collectors of sensitive economic information are able to operate in cyberspace with relatively little risk of detection by their private sector targets. The proliferation of malicious software, prevalence of cyber tool sharing, use of hackers as proxies, and routing of operations through third countries make it difficult to attribute responsibility for computer network intrusions. Cyber tools have enhanced the economic espionage threat, and the Intelligence Community (IC) judges the use of such tools is already a larger threat than more traditional espionage methods.

Economic espionage inflicts costs on companies that range from loss of unique intellectual property to outlays for remediation, but no reliable estimates of the monetary value of these costs exist. Many companies are unaware when their sensitive data is pilfered, and those that find out are often reluctant to report the loss, fearing potential damage to their reputation with investors, customers, and employees. Moreover, victims of trade secret theft use different methods to estimate their losses; some base estimates on the actual costs of developing the stolen information, while others project the loss of future revenues and profits.

Pervasive Threat from Adversaries and Partners

Sensitive US economic information and technology are targeted by the intelligence services, private sector companies, academic and research institutions, and citizens of dozens of countries.

- Chinese actors are the world's most active and persistent perpetrators of economic espionage. US private sector firms and cybersecurity specialists have reported an onslaught of computer network intrusions that have originated in China, but the IC cannot confirm who was responsible.
- Russia's intelligence services are conducting a range of activities to collect economic information and technology from US targets.
- Some US allies and partners use their broad access to US institutions to acquire sensitive US economic and technology information, primarily through aggressive elicitation and other human intelligence (HUMINT) tactics. Some of these states have advanced cyber capabilities.

Outlook

Because the United States is a leader in the development of new technologies and a central player in global financial and trade networks, foreign attempts to collect US technological and economic information will continue at a high level and will represent a growing and persistent threat to US economic security. The nature of the cyber threat will evolve with continuing technological advances in the global information environment.

- Over the next several years, the proliferation of portable devices that connect to the Internet and other networks will continue to create new opportunities for malicious actors to conduct espionage. The trend in both commercial and government organizations toward the pooling of information processing and storage will present even greater challenges to preserving the security and integrity of sensitive information.

- The US workforce will experience a cultural shift that places greater value on access to information and less emphasis on privacy or data protection. At the same time, deepening globalization of economic activities will make national boundaries less of a deterrent to economic espionage than ever.

We judge that the governments of China and Russia will remain aggressive and capable collectors of sensitive US economic information and technologies, particularly in cyberspace.

The relative threat to sensitive US economic information and technologies from a number of countries may change in response to international economic and political developments. One or more fast-growing regional powers may judge that changes in its economic and political interests merit the risk of aggressive cyber and other espionage against US technologies and economic information.

Although foreign collectors will remain interested in all aspects of US economic activity and technology, we judge that the greatest interest may be in the following areas:

- Information and communications technology (ICT), which forms the backbone of nearly every other technology.
- Business information that pertains to supplies of scarce natural resources or that provides foreign actors an edge in negotiations with US businesses or the US Government.
- Military technologies, particularly marine systems, unmanned aerial vehicles (UAVs), and other aerospace/aeronautic technologies.
- Civilian and dual-use technologies in sectors likely to experience fast growth, such as clean energy and health care/pharmaceuticals.

Cyberspace provides relatively small-scale actors an opportunity to become players in economic espionage. Under-resourced governments or corporations could build relationships with hackers to develop customized malware or remote-access exploits to steal sensitive US economic or technology information, just as certain FIS have already done.

- Similarly, political or social activists may use the tools of economic espionage against US companies, agencies, or other entities, with disgruntled insiders leaking information about corporate trade secrets or critical US technology to “hacktivist” groups like WikiLeaks.

Scope Note

This assessment is submitted in compliance with the Intelligence Authorization Act for Fiscal Year 1995, Section 809(b), Public Law 103-359, as amended, which requires that the President biennially submit to Congress updated information on the threat to US industry from foreign economic collection and industrial espionage. This report updates the *14th Annual Report to Congress on Foreign Economic Collection and Industrial Espionage, 2008* and draws primarily on data from 2009-2011.

New Focus and Additional Resources Used for This Year's Report

This report differs from previous editions in three important ways. The first and most significant is the focus. This report gives special attention to foreign collectors' exploitation of cyberspace, while not excluding other established tactics and methods used in foreign economic collection and industrial espionage. This reflects the fact that nearly all business records, research results, and other sensitive economic or technology-related information now exist primarily in digital form. Cyberspace makes it possible for foreign collectors to gather enormous quantities of information quickly and with little risk, whether via remote exploitation of victims' computer networks, downloads of data to external media devices, or e-mail messages transmitting sensitive information.

The second difference from prior reports is that, in addition to researching the large body of intelligence reporting and analysis on economic espionage produced by the Intelligence Community, the Department of Defense (DoD), and other US Government agencies, the drafters of this report consulted new sources of government information.

Third, the Office of the National Counterintelligence Executive (ONCIX) mobilized significant resources from outside the IC during the course of this study. This included outreach to the private sector and, in particular, sponsorship of a conference in November 2010 on cyber-enabled economic espionage at which 26 US Government agencies and 21 private-sector organizations were represented. ONCIX also contracted with outside experts to conduct studies of the academic literature on the cost of economic espionage and the role of the cyber "underground economy."

Definitions of Key Terms

For the purposes of this report, key terms were defined according to both legal and analytic criteria.

The **legal criteria** derive from the language in the Economic Espionage Act (EEA) of 1996 (18 USC §§ 1831-1839). The EAA is concerned in particular with economic espionage and foreign activities to acquire US **trade secrets**. In this context, trade secrets are all forms and types of financial, business, scientific, technical, economic, or engineering information, including patterns, plans, compilations, program devices, formulas, designs, prototypes, methods, techniques, processes, procedures, programs, or codes, whether tangible or intangible, and whether stored or unstored, compiled, or memorialized physically, electronically, graphically, photographically, or in writing, if the owner (the person or entity in whom or in which rightful legal or equitable title to, or license in, is reposed) has taken reasonable measures to keep such information secret and the information derives independent economic value, actual, or potential from not being generally known to, and not being readily ascertainable through, proper means by the public. Activities to acquire these secrets include the following criminal offenses:

- **Economic espionage** occurs when an actor, knowing or intending that his or her actions will benefit any foreign government, instrumentality or agent, knowingly: (1) steals, or without authorization appropriates, carries away, conceals, or obtains by deception or fraud a trade secret; (2) copies, duplicates, reproduces, destroys, uploads, downloads, or transmits that trade secret without authorization; or (3) receives a trade secret knowing that the trade secret had been stolen, appropriated, obtained or converted without authorization (Section 101 of the EEA, 18 USC § 1831).

- **Industrial espionage**, or theft of trade secrets, occurs when an actor, intending or knowing that his or her offense will injure the owner of a trade secret of a product produced for or placed in interstate or foreign commerce, acts with the intent to convert that trade secret to the economic benefit of anyone other than the owner by: (1) stealing, or without authorization appropriating, carrying away, concealing, or obtaining by deception or fraud information related to that secret; (2) copying, duplicating, reproducing, destroying, uploading, downloading, or otherwise transmiting that information without authorization; or (3) receiving that information knowing that that information had been stolen, appropriated, obtained or converted without authorization (Section 101 of the EEA, 18 USC § 1832).

The following definitions reflect the experience of IC cyber, counterintelligence, and economic analysts:

- **Cyberspace** is the interdependent network of information technology (IT) infrastructures, and includes the Internet, telecommunications networks, computer systems, and embedded processors and controllers in critical industries.
- **Sensitive** is defined as information or technology (a) that has been classified or controlled by a US Government organization or restricted in a proprietary manner by a US corporation or other institution, or (b) that has or may reasonably be expected to have military, intelligence, or other uses with implications for US national security, or (c) that may enhance the economic competitiveness of US firms in global markets.

Contributors

ONCIX compiled this report using inputs and reporting from many US Government agencies and departments, including the Air Force Office of Special Investigations (AFOSI), Army Counterintelligence Center (ACIC), Central Intelligence Agency (CIA), Defense Intelligence Agency (DIA), Defense Security Service (DSS), Department of Energy (DoE), Department of Health and Human Services (HHS), Department of State (DoS), Federal Bureau of Investigation (FBI), National Geospatial-Intelligence Agency (NGA), National Reconnaissance Office (NRO), National Security Agency (NSA), and Naval Criminal Investigative Service (NCIS).

Foreign Spies Stealing US Economic Secrets in Cyberspace

US Technologies and Trade Secrets at Risk in Cyberspace

The pace of foreign economic collection and industrial espionage activities against major US corporations and US Government agencies is accelerating. FIS, corporations, and private individuals increased their efforts in 2009-2011 to steal proprietary technologies, which cost millions of dollars to develop and represented tens or hundreds of millions of dollars in potential profits. The computer networks of a broad array of US Government agencies, private companies, universities, and other institutions—all holding large volumes of sensitive economic information—were targeted by cyber espionage; much of this activity appears to have originated in China.

Increasingly, economic collection and industrial espionage occur in cyberspace, reflecting dramatic technological, economic, and social changes that have taken place in recent years in the ways that economic, scientific, and other sensitive information is created, used, and stored. Today, nearly all business records, research results, and other sensitive economic data are digitized and accessible on networks worldwide. Cyber collection can take many forms, including: simple visits to a US company's website for the collection of openly available information; a corporate insider's downloading of proprietary information onto a thumb drive at the behest of a foreign rival; or intrusions launched by FIS or other actors against the computer networks of a private company, federal agency, or an individual.

The Appeal of Collecting in Cyberspace

Cyberspace is a unique complement to the espionage environment because it provides foreign collectors with relative anonymity, facilitates the transfer of a vast amount of information, and makes it more difficult for victims and governments to assign blame by masking geographic locations.

Security and attribution. Collectors operating in a cyber environment can collect economic information with less risk of detection. This is particularly true for remote computer network exploitation (CNE). Foreign collectors take advantage of the fact that it is difficult to detect and to attribute responsibility for these operations.

There is increasing similarity between the tools, tactics, and techniques used by various actors, which reduces the reliability of using these factors to identify those responsible for computer network intrusions.

- The proliferation of malicious software (malware) presents opportunities for intelligence services and other actors to launch operations with limited resources and without developing unique tools that can be associated with them.
- Hacker websites are prevalent across the Internet, and tool sharing is common, causing intrusions by unrelated actors to exhibit similar technical characteristics.
- FIS and other foreign entities have used independent hackers at times to augment their capabilities and act as proxies for intrusions, thereby providing plausible deniability.
- Many actors route operations through computers in third countries or physically operate from third countries to obscure the origin of their activity.

Another factor adding to the challenge of attribution is the diverging perspectives of the actual targets of economic espionage in cyberspace.

- At a conference sponsored by ONCIX in November 2010, US private industry representatives said they saw little difference between cybercrime—for example, identity theft or the misappropriation of intellectual property such as the counterfeiting of commercial video or audio recordings—and the collection of economic or technology information by intelligence services or other foreign entities. Private sector organizations are often less concerned with attribution and focus instead on damage control and prevention; moreover, few companies have the ability to identify cyber intruders.

- US Government law enforcement and intelligence agencies, on the other hand, seek to establish attribution as part of their mission to counter FIS and other clandestine information collectors. They, unlike companies, also have the intelligence collection authorities and capabilities needed to break multiple layers of cover and to establish attribution where possible.

Cyberspace also offers greater security to the perpetrator in cases involving insiders. Although audits or similar cyber security measures may flag illicit information downloads from a corporate network, a malicious actor can quickly and safely transfer a data set once it is copied. A physical meeting is unnecessary between the corrupted insider and the persons or organizations the information is being collected for, reducing the risk of detection.

Faster and cheaper. Cyberspace makes possible the near instantaneous transfer of enormous quantities of economic and other information. Until fairly recently, economic espionage often required that insiders pass large volumes of documents to their handlers in physical form—a lengthy process of collection, collation, transportation, and exploitation.

- Dongfan Chung was an engineer with Rockwell and Boeing who worked on the B-1 bomber, space shuttle, and other projects and was sentenced in early 2010 to 15 years in prison for economic espionage on behalf of the Chinese aviation industry. At the time of his arrest, 250,000 pages of sensitive documents were found in his house. This is suggestive of the volume of information Chung could have passed to his handlers between 1979 and 2006.[a] The logistics of handling the physical volume of these documents—which would fill nearly four 4-drawer filing cabinets—would have required considerable attention from Chung and his handlers. With current technology, all the data in the documents hidden in Chung's house would fit onto one inexpensive CD.[b]

[a]Chung was prosecuted only for possession of these documents with the intent to benefit the People's Republic of China (PRC) and acting as an unregistered foreign agent for China. He was not charged with communication of this information to the PRC or any other foreign entity.

[b]On average, one page of typed text holds 2 kilobytes (KB) of data; thus, 250,000 pages x 2 KB/page = 500,000 KB, or 488 megabytes (MB). A data CD with a capacity of 700 MB retails for $0.75, and a flashdrive with a capacity of 4 gigabytes costs about $13.00.

Extra-territoriality. In addition to the problem of attribution, it often is difficult to establish the geographic location of an act of economic espionage that takes place in cyberspace. Uncertainty about the physical location of the act provides cover for the perpetrators and complicates efforts by US Government law enforcement or intelligence agencies to respond.

Non-Cyber Methods of Economic Espionage

Although this assessment focuses on the use of cyber tools and the cyber environment in foreign efforts to collect sensitive US economic information and technologies, a variety of other methods also remain in use.

Requests for Information (RFI). *Foreign collectors make unsolicited direct and indirect requests for information via personal contacts, telephone, e-mail, fax, and other forms of communication and often seek classified, sensitive, or export-controlled information.*

Solicitation or Marketing of Services. *Foreign companies seek entrée into US firms and other targeted institutions by pursuing business relationships that provide access to sensitive or classified information, technologies, or projects.*

Conferences, Conventions, and Trade Shows. *These public venues offer opportunities for foreign adversaries to gain access to US information and experts in dual-use and sensitive technologies.*

Official Foreign Visitors and Exploitation of Joint Research. *Foreign government organizations, including intelligence services, use official visits to US Government and cleared defense contractor facilities, as well as joint research projects between foreign and US entities, to target and collect information.*

Foreign Targeting of US Visitors Overseas. *Whether traveling for business or personal reasons, US travelers overseas—businesspeople, US Government employees, and contractors—are routinely targeted by foreign collectors, especially if they are assessed*

as having access to some sensitive information. Some US allies engage in this practice, as do less friendly powers such as Russia and China. Targeting takes many forms: exploitation of electronic media and devices, surreptitious entry into hotel rooms, aggressive surveillance, and attempts to set up sexual or romantic entanglements.

Open Source Information. *Foreign collectors are aware that much US economic and technological information is available in professional journals, social networking and other public websites, and the media.*

Large but Uncertain Costs

Losses of sensitive economic information and technologies to foreign entities represent significant costs to US national security. The illicit transfer of technology with military applications to a hostile state such as Iran or North Korea could endanger the lives of US and allied military personnel. The collection of confidential US Government economic information—whether by a potential adversary or a current ally—could undercut US ability to develop and enact policies in areas ranging from climate change negotiations to reform of financial market regulations. The theft of trade secrets from US companies by foreign economic rivals undermines the corporate sector's ability to create jobs, generate revenues, foster innovation, and lay the economic foundation for prosperity and national security.

Data on the effects of the theft of trade secrets and other sensitive information are incomplete, however, according to an ONCIX-sponsored survey of academic literature on the costs of economic espionage.

- Many victims of economic espionage are unaware of the crime until years after loss of the information.
- Even when a company knows its sensitive information has been stolen by an insider or that its computer networks have been penetrated, it may choose not to report the event to the FBI or other law enforcement agencies. No legal requirement to report a loss of sensitive information or a remote computer intrusion exists, and announcing a security breach of this kind could tarnish a company's reputation and endanger its relationships with investors, bankers, suppliers, customers, and other stakeholders.
- A company also may not want to publicly accuse a corporate rival or foreign government of stealing its secrets from fear of offending potential customers or business partners.
- Finally, it is inherently difficult to assign an economic value to some types of information that are subject to theft. It would, for example, be nearly impossible to estimate the monetary value of talking points for a meeting between officials from a US company and foreign counterparts.

The Cost of Economic Espionage to One Company

Data exist in some specific cases on the damage that economic espionage or theft of trade secrets has inflicted on individual companies. For example, an employee of Valspar Corporation unlawfully downloaded proprietary paint formulas valued at $20 million, which he intended to take to a new job in China, according to press reports. This theft represented about one-eighth of Valspar's reported profits in 2009, the year the employee was arrested.

Even in those cases where a company recognizes it has been victimized by economic espionage and reports the incident, calculation of losses is challenging and can produce ambiguous results. Different methods can be used that yield divergent estimates, which adds to the difficulty of meaningfully comparing cases or aggregating estimated losses.

- An executive from a major industrial company told ONCIX representatives in late 2010 that his company has used historical costs—tallying salaries, supplies, utilities, and similar direct expenses—to estimate losses from cases of attempted theft of its trade secrets. This method has the advantage of using known and objective

data, but it underestimates the extent of losses in many cases because it does not capture the effect of lost intellectual property on future sales and profits.

- Harm is calculated in US civil court cases involving the theft of trade secrets by measuring the "lost profits" or "reasonable royalty" that a company is unable to earn because of the theft. Although this method requires subjective assumptions about market share, profitability, and similar factors, it does offer a more complete calculation of the cost than relying strictly on historical accounting data.
- Estimates from academic literature on the losses from economic espionage range so widely as to be meaningless—from $2 billion to $400 billion or more a year—reflecting the scarcity of data and the variety of methods used to calculate losses.

A Possible Proxy Measure of the Costs of Economic Espionage to the United States

New ideas are often a company's or an agency's most valuable information and are usually of greatest interest to foreign collectors. Corporate and government spending on research and development (R&D) is one measure of the cost of developing new ideas, and hence is an indicator of the value of the information that is most vulnerable to economic espionage. R&D spending has been tracked by the National Science Foundation (NSF) since 1953. For 2008, the most recent year available, the NSF calculated that US industry, the Federal Government, universities, and other nonprofit organizations expended $398 billion on R&D, or 2.8 percent of the US Gross Domestic Product.

Pervasive Threat from Intelligence Adversaries and Partners

Many states view economic espionage as an essential tool in achieving national security and economic prosperity. Their economic espionage programs combine collection of open source information, HUMINT, signals intelligence (SIGINT), and cyber operations—to include computer network intrusions and exploitation of insider access to corporate and proprietary networks—to develop information that could give these states a competitive edge over the United States and other rivals.

- China and Russia view themselves as strategic competitors of the United States and are the most aggressive collectors of US economic information and technology.
- Other countries with closer ties to the United States have conducted CNE and other forms of intelligence collection to obtain US economic and technology data, often taking advantage of the access they enjoy as allies or partners to collect sensitive military data and information on other programs.

Recent Insider Thefts of Corporate Trade Secrets with a Link to China

David Yen Lee…chemist with Valspar Corporation…between late 2008 and early 2009 used access to internal computer network to download about 160 secret formulas for paints and coatings to his own storage media…intended to take this proprietary information to a new job with Nippon Paint in Shanghai, China…arrested March 2009…pleaded guilty to one count of theft of trade secrets; sentenced in December 2010 to 15 months in prison.

Meng Hong…DuPont Corporation research chemist...in mid-2009 downloaded proprietary information on organic light-emitting diodes (OLED) to personal e-mail account and thumb drive…intended to transfer this information to Peking University, where he had accepted a faculty position; sought Chinese Government funding to commercialize OLED research…arrested October 2009…pleaded guilty to one count of theft of trade secrets; sentenced in October 2010 to 14 months in prison.

Yu Xiang Dong (aka Mike Yu)…product engineer wjth Ford Motor Company who in December 2006 accepted a job at Ford's China branch…copied approximately 4,000 Ford documents onto an external hard drive to help obtain a job with a Chinese automotive company…arrested in October 2009…pleaded guilty to two counts of theft of trade secrets; sentenced in April 2011 to 70 months in prison.

China: Persistent Collector

Chinese leaders consider the first two decades of the 21st century to be a window of strategic opportunity for their country to focus on economic growth, independent innovation, scientific and technical advancement, and growth of the renewable energy sector.

China's intelligence services, as well as private companies and other entities, frequently seek to exploit Chinese citizens or persons with family ties to China who can use their insider access to corporate networks to steal trade secrets using removable media devices or e-mail. Of the seven cases that were adjudicated under the Economic Espionage Act—both Title 18 USC § 1831 and § 1832—in Fiscal Year 2010, six involved a link to China.

US corporations and cyber security specialists also have reported an onslaught of computer network intrusions originating from Internet Protocol (IP) addresses in China, which private sector specialists call "advanced persistent threats." Some of these reports have alleged a Chinese corporate or government sponsor of the activity, but the IC has not been able to attribute many of these private sector data breaches to a state sponsor. Attribution is especially difficult when the event occurs weeks or months before the victims request IC or law enforcement help.

- In a February 2011 study, McAfee attributed an intrusion set they labeled "Night Dragon" to an IP address located in China and indicated the intruders had exfiltrated data from the computer systems of global oil, energy, and petrochemical companies. Starting in November 2009, employees of targeted companies were subjected to social engineering, spear-phishing e-mails, and network exploitation. The goal of the intrusions was to obtain information on sensitive competitive proprietary operations and on financing of oil and gas field bids and operations.
- In January 2010, VeriSign iDefense identified the Chinese Government as the sponsor of intrusions into Google's networks. Google subsequently made accusations that its source code had been taken—a charge that Beijing continues to deny.
- Mandiant reported in 2010 that information was pilfered from the corporate networks of a US Fortune 500 manufacturing company during business negotiations in which that company was looking to acquire a Chinese firm. Mandiant's report indicated that the US manufacturing company lost sensitive data on a weekly basis and that this may have helped the Chinese firm attain a better negotiating and pricing position.
- Participants at an ONCIX conference in November 2010 from a range of US private sector industries reported that client lists, merger and acquisition data, company information on pricing, and financial data were being extracted from company networks—especially those doing business with China.

Russia: Extensive, Sophisticated Operations

Motivated by Russia's high dependence on natural resources, the need to diversify its economy, and the belief that the global economic system is tilted toward US and other Western interests at the expense of Russia, Moscow's highly capable intelligence services are using HUMINT, cyber, and other operations to collect economic information and technology to support Russia's economic development and security.

- For example, the 10 Russian Foreign Intelligence Service (SVR) "illegals" arrested in June 2010 were tasked to collect economic and technology information, highlighting the importance of these issues to Moscow.[c]

[c] An illegal is an officer or employee of an intelligence organization who is dispatched abroad and who has no overt connection with the intelligence organization with which he or she is connected or with the government operating that intelligence organization.

Russian Leaders Link Intelligence Operations and Economic Interests

The SVR "must be able to swiftly and adequately evaluate changes in the international economic situation, understand the consequences for the domestic economy and…more actively protect the economic interests of our companies abroad."
—Vladimir Putin, President, Russian Federation, October 2007

"Intelligence…aims at supporting the process of modernization of our country and creating the optimal conditions for the development of its science and technology."
—Mikhail Fradkov, Director, SVR, December 2010

Source: Russian press reports.

US Partners: Leveraging Access

Certain allies and other countries that enjoy broad access to US Government agencies and the private sector conduct economic espionage to acquire sensitive US information and technologies. Some of these states have advanced cyber capabilities.

Outlook

Because the United States is a leader in the development of new technologies and a central player in global financial and trade networks, foreign attempts to collect US technological and economic information will remain at high levels and continue to threaten US economic security. The nature of these attempts will be shaped by the accelerating evolution of cyberspace, policy choices made by the economic and political rivals of the United States, and broad economic and technological developments.

Near Certainties

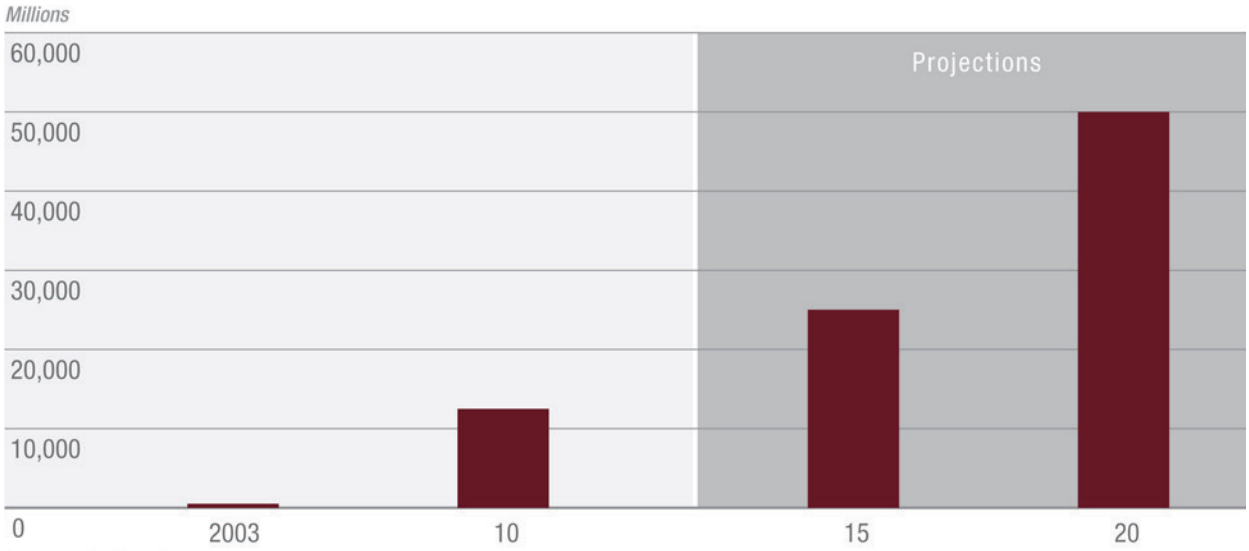
Evolving cyber environment. Over the next three to five years, we expect that four broad factors will accelerate the rate of change in information technology and communications technology in ways that are likely to disrupt security procedures and provide new openings for collection of sensitive US economic and technology information. These were identified in studies conducted by Cisco Systems and discussed at the ONCIX conference in November 2010. At the same time, the growing complexity and density of cyberspace will provide more cover for remote cyber intruders and make it even harder than today to establish attribution for these incidents.

The first factor is a *technological shift*. According to a Cisco Systems study, the number of devices such as smartphones and laptops in operation worldwide that can connect to the Internet and other networks is expected to increase from about 12.5 billion in 2010 to 25 billion in 2015. This will cause a proliferation in the number of operating systems and endpoints that malicious actors such as foreign intelligence services or corrupt insiders can exploit to obtain sensitive information. Meanwhile, the underlying hardware and software of information systems will become more complex.

- Marketing and revenue imperatives will continue to lead IT product vendors to release products with less than exhaustive testing, which will also create opportunities for remote exploitation.

An *economic shift* will change the way that corporations, government agencies, and other organizations share storage, computing, network, and application resources. The move to a "cloud computing" paradigm—which is much cheaper for companies than hosting computer services in-

Projected Growth in Number of IT Devices Connected to Networks and the Internet, 2003-2020



Source: CISCO Systems

house—will mean that employees will be able to work and access data anywhere and at any time, and not just while they are at the office, laboratory, or factory. Although cloud computing offers some security advantages, such as robust backup in the event of a systems disruption, the movement of data among multiple locations will increase the opportunities for theft or manipulation by malicious actors.

The *cultural shift* involves the rise in the US workforce of different expectations regarding work, privacy, and collaboration. Workers will tend to draw few distinctions between their home and work lives, and they will expect free access to any information they want—whether personal or professional—from any location.

- Current technology already enables many US workers to conduct business from remote locations and on-the-go at any time of day. This alteration relies on the ability of workers to connect to one another and their companies through the Internet—increasing their flexibility and corporate productivity but potentially increasing the risk of theft.

Finally, a *geopolitical shift* will continue the globalization of economic activities and knowledge creation. National boundaries will deter economic espionage less than ever as more business is conducted from wherever workers can access the Internet. The globalization of the supply chain for new—and increasingly interconnected—IT products will offer more opportunities for malicious actors to compromise the integrity and security of these devices.

Little change in principal threats. The IC anticipates that China and Russia will remain aggressive and capable collectors of sensitive US economic information and technologies, particularly in cyberspace. Both will almost certainly continue to deploy significant resources and a wide array of tactics to acquire this information from US sources, motivated by the desire to achieve economic, strategic, and military parity with the United States.

China will continue to be driven by its longstanding policy of "catching up fast and surpassing" Western powers. An emblematic program in this drive is Project 863, which provides funding and guidance for efforts to clandestinely acquire US technology and sensitive economic information. The project

was launched in 1986 to enhance China's economic competitiveness and narrow the science and technology gap between China and the West in areas such as nanotechnology, computers, and biotechnology.

- The growing interrelationships between Chinese and US companies—such as the employment of Chinese-national technical experts at US facilities and the off-shoring of US production and R&D to facilities in China—will offer Chinese Government agencies and businesses increasing opportunities to collect sensitive US economic information.
- Chinese actors will continue conducting CNE against US targets.

Two trends may increase the threat from Russian collection against US economic information and technology over the next several years.

- The many Russian immigrants with advanced technical skills who work for leading US companies may be increasingly targeted for recruitment by the Russian intelligence services.
- Russia's increasing economic integration with the West is likely to lead to a greater number of Russian companies affiliated with the intelligence services—often through their employment of ostensibly retired intelligence officers—doing business in the United States.

Technologies likely to be of greatest interest. Although all aspects of US economic activity and technology are of potential interest to foreign intelligence collectors, we judge that the highest interest may be in the following areas.

Information and communications technology (ICT). ICT is a sector likely to remain one of the highest priorities of foreign collectors. The computerization of manufacturing and the push for connectedness mean that ICT forms the backbone of nearly every other technology used in both civilian and military applications.

- Beijing's Project 863, for example, lists the development of "key technologies for the construction of China's information infrastructure" as the first of four priorities.

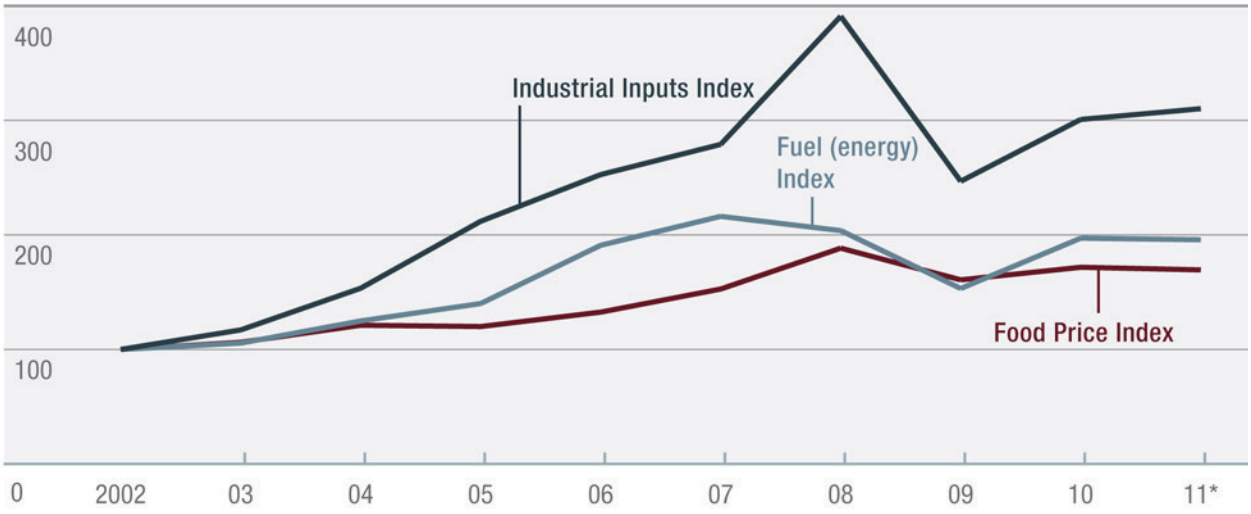
Military technologies. We expect foreign entities will continue efforts to collect information on the full array of US military technologies in use or under development. Two areas are likely to be of particular interest:

- *Marine systems.* China's desire to jump-start development of a blue-water navy—to project power in the Taiwan Strait and defend maritime trade routes—will drive efforts to obtain sensitive US marine systems technologies.
- *Aerospace/aeronautics.* The air supremacy demonstrated by US military operations in recent decades will remain a driver of foreign efforts to collect US aerospace and aeronautics technologies. The greatest interest may be in UAVs because of their recent successful use for both intelligence gathering and kinetic operations in Afghanistan, Iraq, and elsewhere.

Civilian and dual-use technologies. We expect that foreign collection on US civilian and dual-use technologies will follow overall patterns of investment and trade. The following sectors—which are expected to experience surges in investment and are priorities for China—may be targeted more aggressively.

- *Clean technologies.* Energy-generating technologies that produce reduced carbon dioxide and other emissions will be the fastest growing investment sectors in nine of 11 countries recently surveyed by a US consulting company—a survey that included China, France, and India.
- *Advanced materials and manufacturing techniques.* One focus of China's 863 program is achieving mastery of key new materials and advanced manufacturing technologies to boost industrial competitiveness, particularly in the aviation and high-speed rail sectors. Russia and Iran have aggressive programs for developing and collecting on one specific area of advanced materials development: nanotechnology.

Rising Prices Increase Value of Commodity Information to Foreign Collectors (Index, 2002=100)



*2011 values as of April.

Source: International Monetary Fund, World Economic Outlook Database.

- *Healthcare, pharmaceuticals, and related technologies.* Healthcare services and medical devices/equipment will be two of the five fastest growing international investment sectors, according to a US consulting firm. The massive R&D costs for new products in these sectors—up to $1 billion for a single drug—the possibility of earning monopoly profits from a popular new pharmaceutical, and the growing need for medical care by aging populations in China, Russia, and elsewhere are likely to drive interest in collecting valuable US healthcare, pharmaceutical, and related information.
- *Agricultural technology.* Surging prices for food—which have increased by 70 percent since 2002, according to the food price index published by the International Monetary Fund (IMF)—and for other agricultural products may increase the value of and interest in collecting on US technologies related to crop production, such as genetic engineering, improved seeds, and fertilizer.[d]

Business information. As with technologies, we assess that nearly all categories of sensitive US economic information will be targeted by foreign entities, but the following sectors may be of greatest interest:

Energy and other natural resources. Surging prices for energy and industrial commodities—which have increased by 210 percent and 96 percent, respectively, since 2002 according to IMF indices—may make US company information on these resources priority targets for intelligence services and other collectors.[e]

- As noted earlier, cyber intrusions originating in China, but not necessarily attributed to the Chinese Government, since at least 2009 have targeted sensitive operational and project-financing information of US and other international oil, energy, and petrochemical companies, according to reports published by McAfee.

[d]The IMF's Food Price Index is a weighted index that includes the spot prices of cereal grains, vegetable oils and protein meals, meat, seafood, sugar, bananas, and oranges.

[e]The Fuel (energy) index published by the IMF is a weighted index that includes the spot prices of crude oil, natural gas, and coal. The Industrial Inputs Index is a weighted index that includes the spot price of agricultural raw materials (timber, fibers, rubber and hides) and non-precious metals (such as copper, aluminum, and iron ore).

Business deals. Some foreign companies—at times helped by their home countries' intelligence services—will collect sensitive information from US economic actors that are negotiating contracts with or competing against them.

Macroeconomic information. In the wake of the global financial crisis of 2008-2009 and related volatility in the values of currencies and commodities, sensitive macroeconomic information held by the US private sector and government agencies is likely to remain a prime collection target for both intelligence services and foreign corporations. Chinese and Russian intelligence collectors may pursue, for example, non-public data on topics such as interest rate policy to support their policymakers' efforts to advance the role of their currencies and displace the dollar in international trade and finance. Such information also could help boost the performance of sovereign wealth funds controlled by governments like China's, whose China Investment Corporation managed more than $300 billion in investments as of late 2010.[f]

Possible Game Changers

Any of a range of less-likely developments over the next several years could increase the threat from economic espionage against US interests.

Emergence of new state threats. The relative threat to sensitive US economic information and technologies from different countries is likely to evolve as a function of international economic and political developments.

One or more fast-growing regional powers may judge that changes in its economic and political interests merit the risk of an aggressive program of espionage against US technologies and sensitive economic information.

Growing role of non-state and non-corporate actors. The migration of most business and technology development activities to cyberspace is making it easier for actors without the resources of a nation-state or a large corporation to become players in economic espionage. Such new actors may act as surrogates or contractors for intelligence services or major companies, or they could conduct espionage against sensitive US economic information and technology in pursuit of their own objectives.

Hackers for hire. Some intelligence services with less-developed cyber programs already use relationships with nominally independent hackers to augment their capabilities to target political and military information or to carry out operations against regime enemies. For example, the Iranian Cyber Army, a hacker group with links to the Iranian Government, has used social engineering techniques to obtain control over Internet domains and disrupt the political opposition, according to research conducted under an ONCIX contract.

No evidence of involvement by independent hackers in economic espionage has been found in intelligence or academic reporting to date, in large part due to the absence of a profitable market for the resale of stolen information. This "cyber underground" could, however, become a fruitful recruiting ground for the tools and talents needed to support economic espionage. Following the model used by some intelligence services in exploiting the cyber environment for political or military espionage, a foreign government or corporation could build relationships with hackers for the development of customized malware or remote access exploits for the exfiltration of sensitive US economic or technology information.

Hacktivists. Political or social activists also may use the tools of economic espionage against US companies, agencies, or other entities. The self-styled whistleblowing group WikiLeaks has already published computer files provided by corporate insiders indicating allegedly illegal or unethical behavior at a Swiss bank, a Netherlands-based commodities company, and an international pharmaceutical trade association. LulzSec—another hacktivist group—has exfiltrated data from several businesses that it posted for public viewing on its website.

[f] A sovereign wealth fund is a government investment fund, funded by foreign currency reserves but managed separately from official currency reserves. In other words, it is a pool of money that a government invests for profit.

Corporate trade secrets or information about critical US technology may be at similar risk of disclosure to activist groups by disgruntled insiders.

- Antipoverty activists, for instance, could seek to publish the details of a new medicine under development by a US pharmaceutical company, with the goal of ending the firm's "monopoly" profits and making the product more widely available.
- Antiwar groups could disclose information about a new weapons system in the hope of dissuading the United States from deploying it.

Annex A

Intelligence Community and Private Sector Measures to Counter Economic Espionage and Manage Collection in Cyberspace

The IC is working closely with all segments of the public and private sectors to try to counter espionage activities that target our sensitive economic data and technology. We cannot expect to stop entirely or prevent hostile activity to collect US public and private sector information, but we can work to minimize the activity and mitigate its effects.

Intelligence Community Responses

The IC and especially counterintelligence (CI) officers have already taken a number of steps to improve collaboration, collection, and analysis across the CI, economics, and cyber disciplines.

Improved collaboration. Over the past few years, the IC has established multiple organizations and working groups to better understand the cyber espionage threat. These have contributed to a better understanding of the use of cyber in economic espionage.

- The National Cyber Counterintelligence Working Group established in 2011 is composed of 16 IC and other federal agencies and is creating a coordinated response to the cyber intelligence threat.
- The FBI is leading the National Cyber Investigative Joint Task Force, which brings together multiple agencies to collaborate on intrusions into US systems.

CI officers are considering an expansion of collaboration to include enhanced information sharing with Department of Justice attorneys. CI officers could introduce questions for attorneys to pose to offenders during the investigation process. They might also look at ways to tie plea bargains and sentencing decisions to suspects' willingness to cooperate with the CI Community during damage assessments.

Improved analysis and collection. The IC has made great strides over the past few years in understanding the cyber espionage threat to US Government systems, but our knowledge of cyber-enabled economic espionage threats to the US private sector remains limited.

Defense Model Shows Limits to Mandatory Reporting Requirements

DoD's partnership with cleared defense contractors (CDCs) highlights difficulties in establishing an effective framework to improve the IC's understanding of foreign cyber threats and promote threat awareness in industry. The defense industrial base conducts $400 billion in business with the Pentagon each year and maintains a growing repository of government information and intellectual property on unclassified networks. CDCs are required to file reports of suspicious contacts indicative of foreign threats—including cyber—to their personnel, information, and technologies.

- *Despite stringent reporting requirements for CDCs, DSS reports that only 10 percent of CDCs actually provide any sort of reporting in a given year.*
- *Another shortcoming of the defense model is that contractors do not always report theft of intellectual property unless it relates specifically to Pentagon contracts, according to outreach discussions with corporate officers.*
- *Corporate security officers also have noted that US Government reporting procedures are often cumbersome and redundant, with military services and agencies such as DSS and the FBI often seeking the same information but in different formats.*

Operations. CI professionals are adapting how they detect, deter, and disrupt collection activity in cyberspace because of the challenges in detecting the traditional indicators of collection activity—spotting, assessing, and recruiting.

It is imperative that we improve our ability to attribute technical and human activity in the cyber environment so that we can improve our understanding of the threat and our ability to generate a greater number of offensive CI responses.

Training and awareness. Expanding our national education and awareness campaign aimed at individuals and corporations is an essential defensive strategy for countering threats from cyber-enabled economic collection and espionage. We are building on current outreach initiatives that the FBI and ONCIX have already initiated.

- IC outreach to all US Government agencies, state and local governments, academia, nongovernmental organizations, industry associations, and companies is critical for promoting threat awareness, as well as for a better understanding of nongovernmental perspectives. Partners outside the IC are becoming aware of the wide range of potentially sensitive information in their possession and the extent of foreign efforts to acquire it.
- Outreach efforts include awareness and mitigation strategies for insider threat issues. The unique access of insiders to information technology systems and organizational processes makes this the most dangerous approach to cyber economic collection and espionage, as insiders can act alone to guide CNE or to download sensitive data to portable media.

ONCIX already engages in dialogue with ASIS International—an industry association for security professionals—and the Department of State's Overseas Security Advisory Council on the challenges facing both the public and private sectors with regard to cyber-enabled economic collection and espionage.

Finally, IC outreach efforts to the private sector on economic espionage need to fully engage corporate and other partners in order to be credible. We can facilitate partnerships to share best practices, threat updates and analysis, and data on intrusions. One company security officer has suggested that the IC must speak to industry in language geared to the private sector's needs and experience and emphasize, for example, that the protection of trade secrets is critical to corporate profitability and growth.

As a follow-up to the public/private sector Workshop on Cyber-Enabled Economic Espionage held in 2010, ONCIX should consider sponsoring another conference with Department of Justice and private sector stakeholders on lessons learned regarding successful convictions under Section 1831 of the Economic Espionage Act.

Corporate Responses

The private sector already has a fiduciary duty to account for corporate risk and the bottom-line effects of data breaches, economic espionage, and loss or degradation of services. A key responsibility of chief executive officers and boards of directors is to ensure that the protection of trade secrets and computer networks is an integral part of all corporate decisions and processes and that all managers—not just security and information systems officials—have a stake in the outcome.[a] Viewing network security and data protection as a business matter that has a significant impact on profitability will lead to more effective risk management and ensure that adequate resources are allocated to address cyber threats to companies.

- Only 5 percent of corporate chief financial officers are involved in network security matters, and just 13 percent of companies have a cross-functional cyber risk team that bridges the technical, financial, and other elements of a company, according to a 2010 study.

Judicial Mandate for Boards of Directors To Secure Corporate Information

Delaware's Court of Chancery ruled in the 1996 Caremark *case that a director's good faith duty includes a duty to attempt to ensure that a corporate*

[a]Legal and human resources officers are two sets of key stakeholders given the role that corporate insiders have historically played in contributing to economic espionage and the theft of trade secrets.

information and reporting system exists and that failure to do so may render a director liable for losses caused by the illegal conduct of employees. The Delaware Supreme Court clarified this language in the 2006 Stone v. Ritter *case—deciding that directors may be liable for the damages resulting from legal violations committed by the employees of a corporation, if directors fail to implement a reporting system or controls or fail to monitor such systems.*

Companies that successfully manage the economic espionage threat realize and convey to their employees that threats to corporate data extend beyond company firewalls to include other locations where company data is moved or stored. These include cloud sites, home computers, laptops, portable electronic devices, portable data assistants, and social networking sites.

- A survey of 200 information technology and security professionals in February 2011 revealed that 65 percent do not know what files and data leave their enterprise.
- According to a March 2011 press report, 57 percent of employees save work files to external devices on a weekly basis.
- E-mail systems are often less protected than databases yet contain vast quantities of stored data. E-mail remains one of the quickest and easiest ways for individuals to collaborate—and for intruders to enter a company's network and steal data.

Cyber threats to company information are compounded when employees access data through portable devices or network connections while traveling overseas. Many FIS co-opt hotel staffs to allow access to portable devices left unattended in rooms. It is also much easier for FIS to monitor and exploit network connections within their own borders.

- Foreign collectors engage in virtual methods to collect sensitive corporate data and take advantage of victims' reluctance to report digital penetrations and low awareness of foreign targeting, according to legal academic research.

Corporate security officers have told ONCIX that US Government reporting procedures on economic espionage and cyber intrusions are often cumbersome and redundant. Agencies such as DSS and the FBI often seek the same information but in different formats.

Best Practices in Data Protection Strategies and Due Diligence for Corporations

Information Strategy
• Develop a "transparency strategy" that determines how closed or open the company needs to be based on the services provided.
Insider Threat Programs and Awareness
• Institute security training and awareness campaigns; convey threats to company information accessed through portable devices and when traveling abroad.
• Establish an insider threat program that consists of information technology-enabled threat detection, foreign travel and contact notifications, personnel security and evaluation, insider threat awareness and training, and reporting and analysis.
• Conduct background checks that vet users before providing them company information.
• Implement non-disclosure agreements with employees and business partners.
• Establish employee exit procedures; most employees who steal intellectual property commit the theft within one month of resignation.
Effective Data Management
• Get a handle on company data—not just in databases but also in e-mail messages, on individual computers, and as data objects in web portals; categorize and classify the data, and choose the most appropriate set of controls and markings for each class of data; identify which data should be kept and for how long. Understand that it is impossible to protect everything.
• Establish compartmentalized access programs to protect unique trade secrets and proprietary information; centralize intellectual property data—which will make for better security and facilitate information sharing.
• Restrict distribution of sensitive data; establish a shared data infrastructure to reduce the quantity of data held by the organization and discourage unnecessary printing and reproduction.
Network Security, Auditing, and Monitoring
• Conduct real-time monitoring/auditing of the networks; maintain thorough records of who is accessing servers, and modifying, copying, deleting, or downloading files.
• Install software tools—content management, data loss prevention, network forensics—on individual computer workstations to protect files.

- Encrypt data on servers and password-protect company information.
- Incorporate multi-factor authentication measures—biometrics, PINs, and passwords combined with knowledge-based questions—to help verify users of information and computer systems.
- Create a formal corporate policy for mobility—develop measures for centrally controlling and monitoring which devices can be attached to corporate networks and systems and what data can be downloaded, uploaded, and stored on them.
- Formalize a social media policy for the company and implement strategies for minimizing data loss from on-line social networking.

Contingency Planning

- Establish a continuity of operations plan—back up data and systems; create disaster recovery plans; and plan for data breach contingencies.
- Conduct regular penetration testing of company infrastructure as well as of third-party shared service provider systems.
- Establish document creation, retention, and destruction policies.

Resources for Help

- Contact ONCIX or the FBI for assistance in developing effective data protection strategies. If a data breach is suspected, contact the FBI or other law enforcement/organizations for help in identifying and neutralizing the threat.

Annex B

West and East Accuse China and Russia of Economic Espionage

Other advanced industrial countries principally blame China and Russia for economic espionage that results in large but uncertain monetary costs and job losses. They perceive that China and Russia continue to use traditional human and technical collection methods—particularly against small- and medium-sized businesses—to gather economic information and technologies that save them research and development (R&D) resources and provide entrepreneurial and marketing advantage for their corporate sectors.

- Germany's Federal Office for the Protection of the Constitution (BfV) estimates that German companies lose $28 billion-$71 billion and 30,000-70,000 jobs per year from foreign economic espionage. Approximately 70 percent of all cases involve insiders.
- South Korea says that the costs from foreign economic espionage in 2008 were $82 billion, up from $26 billion in 2004. The South Koreans report that 60 percent of victims are small- and medium-sized businesses and that half of all economic espionage comes from China.[a]
- Japan's Ministry of Economy, Trade, and Industry conducted a survey of 625 manufacturing firms in late 2007 and found that more than 35 percent of those responding reported some form of technology loss. More than 60 percent of those leaks involved China.

France's Renault Affair Highlights Tendency to Blame China

Broad French concerns with Chinese economic espionage formed the background of the hasty—and subsequently retracted—accusations by corporate and political leaders in January 2011 that three top executives with the Renault automobile company had taken bribes from China in exchange for divulging technology.

- *An investigation by the French internal security service revealed that the accusations against China lacked substance and may have stemmed from a corrupt corporate security officer's attempts to generate investigative work for a friend's consulting business.*

Past Chinese economic espionage against the French automotive industry—including the parts manufacturer Valeo—probably made the French willing to give credence to any accusation of similar malfeasance against China.

Countries acknowledge the growing use of cyber tools for foreign economic collection and espionage and often note difficulties in understanding losses associated with these cyber collection methods. A 2010 survey of 200 industry executives from the power, oil, gas, and water sectors in 12 Western countries, China, and Russia indicates that 85 percent of respondents experienced network intrusions and that government-sponsored sabotage and espionage was the most often cited cyber threat.

- A 2010 Canadian Government report claimed that 86 percent of large Canadian corporations had been hit and that cyber espionage against the private sector had doubled in two years, according to a press report.
- The German BfV offers no reliable figures on the number of cases and amount of damage caused by cyber-enabled economic espionage, adding that their intelligence services are "groping in the dark." The German Government has noted the use of CNE tools and removable media devices, claiming that $99 million are spent annually for IT security.
- UK officials note that the cost of an information security incident averages between $16,000 and $32,000 for a small company and between

[a]We have no information on the methodologies that the Germans and South Koreans used to calculate their losses.

$1.6 million and $3.2 million for firms with more than 500 employees. The United Kingdom estimates that attacks on computer systems, including industrial espionage and theft of company trade secrets, cost the private sector $34 billion annually, of which more than 40 percent represents theft of intellectual property such as designs, formulas, and company secrets.

- Germany and South Korea judge that China, in particular, increasingly uses cyber tools to steal trade secrets and achieve plausible deniability, according to press reporting.[b]
- Unidentified CNE operators have accessed more than 150 computers at France's Finance Ministry since late 2010, exfiltrating and redirecting documents relating to the French G-20 presidency to Chinese sites, according to a press report.
- The British Security Service's Center for the Protection of National Infrastructure warned hundreds of UK business leaders in 2010 of Chinese economic espionage practices, including giving gifts of cameras and memory sticks equipped with cyber implants at trade fairs and exhibitions. This followed similar notification sent to 300 UK business leaders in 2007 warning them of a coordinated cyber espionage campaign against the British economy.
- German officials also noted that business travelers' laptops are often stolen during trips to China. The Germans in 2009 highlighted an insider case in which a Chinese citizen downloaded highly sensitive product data from the unidentified German company where he worked to 170 CDs.

China's Response to Allegations of Economic Espionage

China usually responds to public allegations of economic espionage with outright denial and counteraccusations. In 2009 China claimed the Australian mining giant Rio Tinto engaged in six years of espionage activities—bribery and information gathering—that resulted in a loss of iron ore imports for the Chinese steel industry as large as $107 billion. This loss was more than twice the total profits generated by the Chinese steel industry over that same six-year period, according to the Chinese Government.

Russia also is seen as an important actor in cyber-enabled economic collection and espionage against other countries, albeit a distant second to China. Germany's BfV notes that Russia uses CNE and e-mail interception to save billions of dollars on R&D in the energy, information technology, telecommunications, aerospace, and security sectors.

- The Director-General of the British Security Service publicly stated that Russia, as well as China, is targeting the UK's financial system.
- A Russian automotive company bribed executives at South Korea's GM-Daewoo Auto and Technology to pass thousands of computer files on car engine and component designs in 2009, according to a press report.
- A German insider was convicted of economic espionage in 2008 for passing helicopter technology to the Russian SVR in exchange for $10,000. The insider communicated with his Russian handler through anonymous e-mail addresses.

Countries Suspect Each Other of Committing Economic Espionage

Allies often suspect each other of economic espionage—underlining how countries can be partners in traditional security matters yet competitors in business and trade. Foreign corporate leaders may make accusations that are not publicly endorsed by their governments.

- *According to a 2010 press report, the Germans view France and the United States as the primary perpetrators of economic espionage "among friends."*
- *France's Central Directorate for Domestic Intelligence has called China and the United States the leading "hackers" of French businesses, according to a 2011 press report.*

[b] We lack insight on the processes that the Germans and South Koreans used to attribute cyber activities to China.

Some countries exercise various legislative, intelligence, and diplomatic options to respond to the threat of cyber-enabled economic collection and espionage.

- France and South Korea have proposed new legislation or changes to existing laws to help mitigate the effects of economic espionage. France also is considering a public economic intelligence policy and a classification system for business information.
- France, the United Kingdom, and Australia have issued strategies and revamped bureaucracies to better align resources against cyber and economic espionage threats. France created a 12-person Economic Intelligence Office in 2009 to coordinate French corporate intelligence efforts. The United Kingdom established an Office of Cyber Security to coordinate Whitehall policy under a senior official and a Cyber Security Operations Centre within the Government Communications Headquarters (GCHQ) SIGINT unit. Australia created a cyber espionage branch within its Security Intelligence Organization in 2010.
- The United Kingdom is mobilizing its intelligence services to gather intelligence on potential threats and for operations against economic collection and espionage in cyberspace, according to press reports.

German Espionage Legislation Has Limited Results

Germany's Federal Prosecutor General initiated 31 preliminary proceedings on espionage in 2007, resulting in just one arrest and one conviction. German authorities note that espionage cases are often hindered by diplomatic immunity protections and by attribution issues from operating abroad through cyberspace.

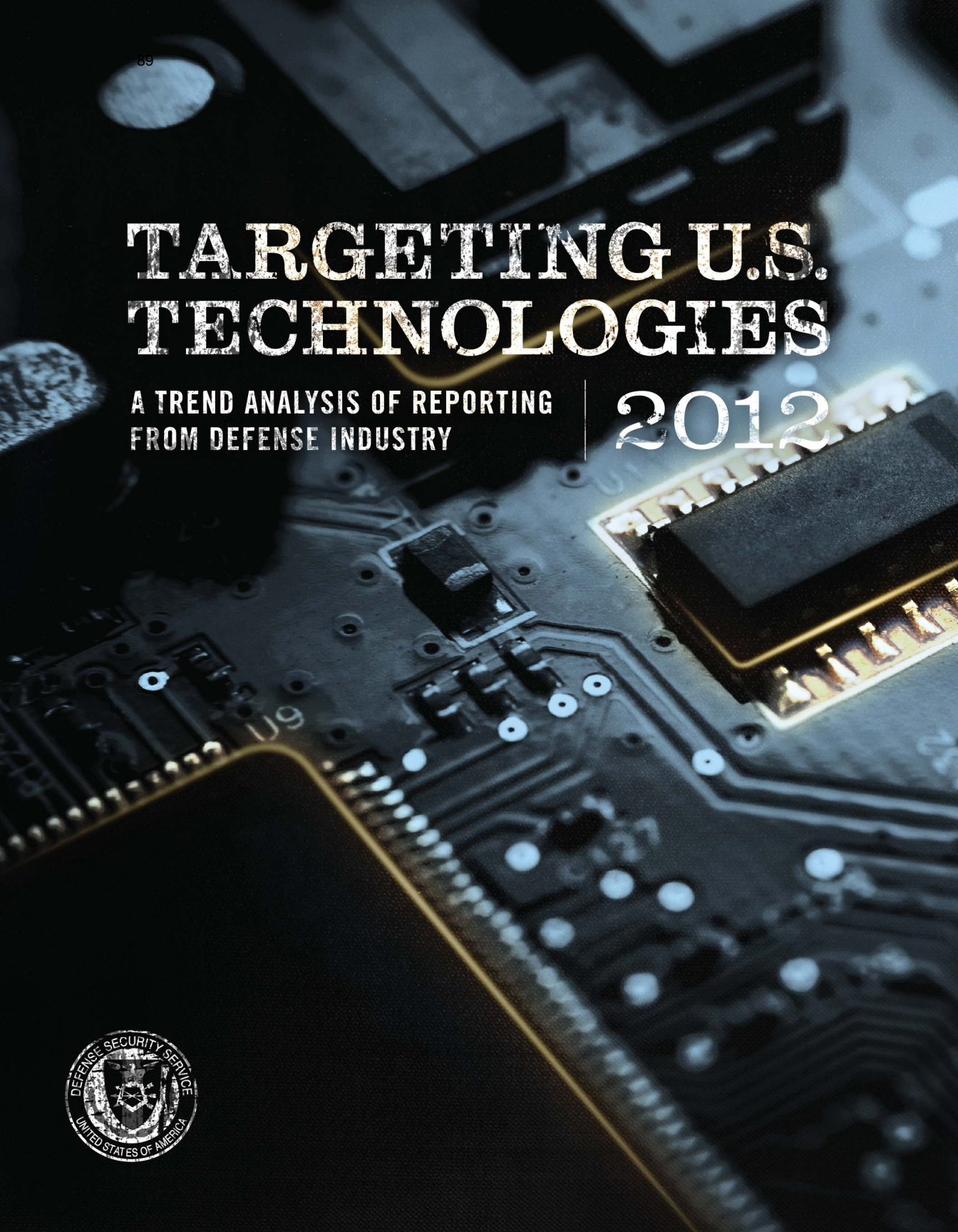
Nearly all countries realize that public and private partnerships are crucial to managing the effects of cyber-enabled economic collection and espionage. The United Kingdom notes that 80 percent of its critical national infrastructure is owned and operated by the private sector. German authorities would like more corporate feedback and say that most enterprises either do not know when they are victims of cyber espionage or do not want to publicly admit their weaknesses. Most countries engage in some form of corporate outreach.

- The French intelligence services offer regular threat briefings to private companies, according to press reports.
- German authorities regularly exchange information with corporate security officers through a private/public working group that includes Daimler AG, Volkswagen, Porsche, Bayer, the German post office, and the railroad industry.

Corporate Leaders Speak Out on Chinese Espionage

Some foreign corporate executives have singled out Chinese espionage as a threat to their companies.

- *British entrepreneur James Dyson—inventor of the bagless vacuum cleaner—warned in 2011 that Chinese students were stealing technological and scientific secrets from UK universities, according to a press report. He noted that Chinese students were also planting software bugs that would relay information to China even after their departure from the universities.*
- *The CEO of an Australian mining firm said that worries over Chinese and other corporate espionage drove him to adopt a more transparent quarterly pricing mechanism for commodities such as iron ore. He claimed that selling products at market-clearing prices visible to all would minimize the impact of differential information that one party may hold, according to a press article.*



TARGETING U.S. TECHNOLOGIES

A TREND ANALYSIS OF REPORTING FROM DEFENSE INDUSTRY

2012

DEFENSE SECURITY SERVICE
UNITED STATES OF AMERICA

DSS MISSION

DSS supports national security and the warfighter, secures the nation's technological base, and oversees the protection of sensitive and classified information and technology in the hands of industry.

We accomplish this mission by: clearing industrial facilities, personnel, and associated information systems; collecting, analyzing, and providing threat information to industry and government partners; managing foreign ownership control and influence in cleared industry; providing advice and oversight to industry; delivering security education and training; and, providing information technology services that support the industrial security mission of the Department of Defense and its partner agencies.

THIS PRODUCT WAS COORDINATED WITH: ACIC, AFOSI, DIA, & NGA

Produced by the Defense Security Service
Counterintelligence Directorate
www.DSS.mil

TARGETING U.S. TECHNOLOGIES

A TREND ANALYSIS OF REPORTING FROM DEFENSE INDUSTRY | 2012

TABLE OF CONTENTS

FIGURES

IN THE INTERESTS OF READABILITY AND COMPREHENSION, THE EDITORS HAVE DEFERRED THE CONVENTIONAL STYLISTIC USE OF REPEATED ACRONYMS IN FAVOR OF A FULL EXPOSITION OF TERMS AS THEY ARE FIRST USED WITHIN EACH SECTION.

PREFACE

The stakes are high in the battle against foreign collection efforts and espionage that target U.S. technology, intellectual property, trade secrets, and proprietary information. Our national security relies on our collective success at thwarting these persistent attacks. Every time our adversaries gain access to sensitive or classified information and technology, it jeopardizes the lives of our warfighters, since these adversaries can exploit the information and technology to develop more lethal weapons or countermeasures to our systems. Our national security is also at risk in the potential loss of our technological edge, which is closely tied to the economic success of the cleared contractor community and the well-being of our economy.

Preventing such losses takes a team effort. The Defense Security Service (DSS) builds on the information contained in reports from industry to develop analytical assessments that articulate the threat to U.S. information and technology resident in cleared industry. This annual publication, *Targeting U.S. Technologies: A Trend Analysis of Reporting from Defense Industry*, presents DSS' analysis of those industry reports. Like any analysis, this one is only as good as the information that goes into it. Timely and accurate initial reports of illicit collection attempts are the foundation upon which this process rests, and it is cleared contractor employees who originate those suspicious contact reports.

When this process works well, our national security, warfighters, cleared industry partners, and local communities all benefit. The information contained in this report helps employees, companies, and intelligence and law enforcement professionals better understand the continuing yet changing nature of the threats we face. Increased awareness of the U.S. technologies being targeted by foreign entities and the methods of operation they use in their efforts to acquire those technologies can only make us better at identifying and thwarting illicit collection attempts. In fiscal year 2011, our combined efforts produced 485 operations or investigations based on information that industry provided. Over three-quarters of these are still undergoing significant action, with many foreign collectors already identified, isolated, diverted, or otherwise thwarted.

But these combined efforts face a threat that is growing, persistent, pervasive, and insidious. Cleared industry, DSS, and the intelligence and law enforcement communities continue their efforts to further expand, develop, and refine their methods of defending our national security. Yet the response by foreign collectors who seek to illicitly acquire U.S. information and technology despite those efforts also continues to undergo expansion, development, and refinement.

During fiscal year 2011, the persistent, pervasive, and insidious nature of that threat became particularly noteworthy, and the pattern became even more firmly established. Foreign collectors seek to elude the protective efforts of industry, DSS, the Intelligence Community, and law enforcement by concealing their activities behind various covers, such as third countries, front companies, and cyber identities. This report will present various examples of such activities.

Increasingly, the result of all this foreign collection activity is like malignant plants with multiple interlocking roots and branches. These noxious weeds root in unexpected places, then send out shoots and tendrils that encroach through any crack or gap into the nurseries and gardens of our industrial base. We may pull out some parts of a plant by the roots and lop off the leaves of others, but the pervasive, penetrating weeds remain.

It is only by the continued vigilance and focused and unstinting effort of those of you in cleared industry—by "tending your garden" assiduously and reporting incursions of "weeds" promptly and fully—that the rest of the nation's defenders can help protect its security.

Stanley L. Sims

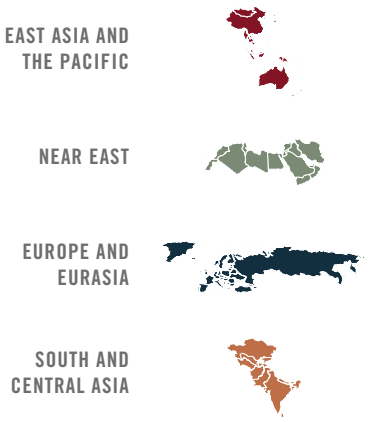
DIRECTOR
DEFENSE SECURITY SERVICE

EXECUTIVE SUMMARY

In one way, the data concerning industry reports of foreign attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base remained very consistent between fiscal year 2010 (FY10) and FY11. The East Asia and the Pacific region accounted for 43 percent of the total in both years; the Near East accounted for 18 percent in both years; Europe and Eurasia dropped only slightly, from 15 percent to 13 percent; and South and Central Asia was reasonably stable, rising from nine percent to twelve percent.

REGIONAL TRENDS

FIGURE 1



But this seeming stability in the data does not reflect the overall phenomena in the past year. The total number of reports received from industry increased over 75 percent from FY10. In the past year, reports from the East Asia and the Pacific and Near East regions increased by around 75 percent, from Europe and Eurasia by over 60 percent, and from South and Central Asia by a steep 129 percent. All other regions increased in number of reports as well. Thus, the only stability in the data is the relentless upward trend.

Considerable diversity exists within each region. Countries vary in size, resources, economic development, political system, degree of militarization, and foreign policy orientation and goals. And the situation is not static; change continues in these variables as well. Some countries are on the way "up," others "down," however defined. Some are satisfied with their place and role in the world; others aspire to change them, and work aggressively to do so. Any of these factors can lead to attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base.

Despite the diversity between regions and countries discussed above, collectors continue to expand the degree of interaction between them in their attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. Whether working with each other, working through each other, buying from each other, or attempting to throw suspicion on each other, these convoluted pathways make it more difficult to ascribe collection attempts to a particular country, region, or collector affiliation.

KEY FINDINGS

The order of the regions linked to the most prolific collectors of U.S. information and technology remained unchanged from fiscal year 2010 (FY10); commercial remained the most common collector affiliation; and the top four most targeted technology categories remained the same.

Constancy of the order of the regions represents the most enduring trend. Over the past five years, East Asia and the Pacific and the Near East have remained the first and second most prolific collector regions, responsible for at least 56 percent of all reported collection attempts each year, including 61 percent in FY11. However, industry reports of collection attempts originating from South and Central Asia increased by 129 percent, reflecting aggressive collection efforts.

Commercial entities constituted the most common affiliation in FY11 industry reporting, residing at the top of the ranking in five of the six regions.

Collectors' most frequently applied methods of operation (MO) sought information or technology directly, whether by attempted acquisition of technology or request for information (RFI). Combined, these MOs accounted for 43 percent of reported collection attempts in FY11. A DSS redefinition of attempted acquisition led to different apportionment of cases in FY11 than in previous years, but taken together these two MOs represent direct overt contact with cleared industry in an attempt to receive information or acquire technology—by simply asking for it.

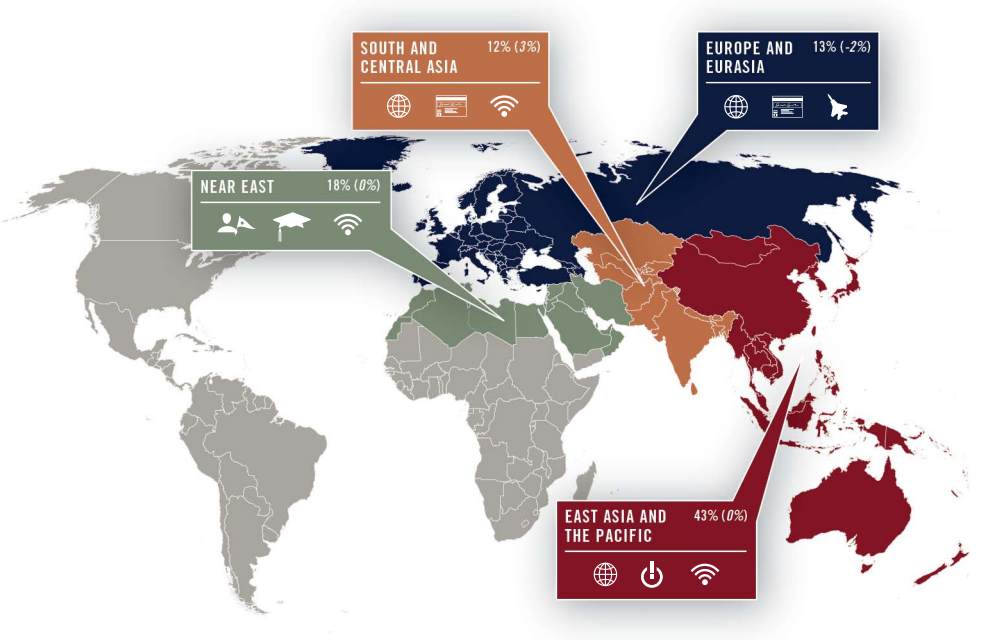
In FY11, suspicious network activity (SNA) was the most prevalent collection MO for entities originating from East Asia and the Pacific; SNA figured no more prominently than fifth in any other region. Due to the nature of SNA, it remains difficult to attribute such collection attempts to an entity or even to a region of origin.

The top four most targeted technology categories in FY11—information systems (IS); lasers, optics, and sensors (LO&S); aeronautics systems; and electronics—remained unchanged. Armaments and energetic materials replaced marine systems as the fifth most targeted category of the Militarily Critical Technologies List (MCTL). But there was a broadening of reported interest in technology to space systems, processing and manufacturing, and directed energy systems in FY11.

Foreign governments are beginning to move into space for commercial telecommunications, increased command and control, and intelligence, surveillance, and reconnaissance (ISR), and the demand for radiation-hardened (rad-hard) microelectronics is likely to dramatically rise over the coming years. Foreign entities' interest in these technologies rose over the past year, and collectors will likely increase their targeting of cleared contractors' design, manufacturing, and packaging of rad-hard microelectronics.

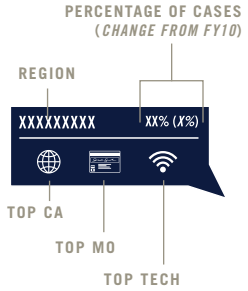
FISCAL YEAR 2011 COLLECTION TRENDS

FIGURE 2



COLLECTOR AFFILIATIONS*

- COMMERCIAL
- INDIVIDUAL
- GOVERNMENT AFFILIATED
- GOVERNMENT
- UNKNOWN



METHODS OF OPERATION*

- ATTEMPTED ACQUISITION OF TECHNOLOGY
- REQUESTS FOR INFORMATION
- SUSPICIOUS NETWORK ACTIVITY
- ACADEMIC SOLICITATION
- SOLICITATION OR MARKETING
- OFFICIAL FOREIGN VISITS AND TARGETING
- CONFERENCES, CONVENTIONS, AND TRADE SHOWS
- EXPLOITATION OF RELATIONSHIPS
- SEEKING EMPLOYMENT
- CRIMINAL ACTIVITIES
- TARGETING U.S. TRAVELERS OVERSEAS

TOP TARGETED TECHNOLOGIES*

- INFORMATION SYSTEMS
- LASERS, OPTICS, AND SENSORS
- AERONAUTICS SYSTEMS
- ELECTRONICS
- ARMAMENTS AND ENERGETIC MATERIALS
- SPACE SYSTEMS
- MARINE SYSTEMS
- POSITIONING, NAVIGATION, AND TIME
- MATERIALS AND PROCESSES
- GROUND SYSTEMS
- INFORMATION SECURITY
- PROCESSING AND MANUFACTURING

*Categories of affiliations, methods, and technologies listed above appear in order of prevalence in overall FY11 reporting statistics.

BACKGROUND

THE ROLE OF THE DEFENSE SECURITY SERVICE

DSS supports national security and the warfighter, secures the nation's technological base, and oversees the protection of U.S. and foreign classified information and technology in the hands of industry. The DSS Counterintelligence (CI) Directorate seeks to identify unlawful penetrators of cleared U.S. industry and stop foreign collection attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. DSS CI articulates the threat for industry and U.S. Government leaders.

THE ROLE OF INDUSTRY

In carrying out its mission, DSS relies on the support of cleared contractor employees and the U.S. intelligence and law enforcement communities. Chapter 1, Section 3 of Department of Defense (DoD) Instruction 5220.22-M, *National Industrial Security Program Operating Manual (NISPOM)*, dated February 28, 2006, requires cleared contractors to remain vigilant and report suspicious contacts. The process that begins with initial industry reporting and continues with ongoing and collective analysis reaches its ultimate stage in successful investigations or operations by federal investigative or intelligence agencies.

In accordance with the reporting requirements laid out in the *NISPOM*, DSS receives and analyzes reports from cleared contractors and categorizes them as suspicious, unsubstantiated, or of no value. For each reported collection attempt, DSS data aggregation and analysis methodologies seek to gather as much information as possible. The analysis of this information forms the basis for this report.

Such cleared contractor reporting provides information concerning actual, probable, or possible espionage, sabotage, terrorism, or subversion activities to DSS and the Federal Bureau of Investigation. When indicated, DSS refers cases of CI concern to its partners in the law enforcement and intelligence communities for potential exploitation or neutralization. DSS follows up with remedial actions for industry to decrease the threat in the future. This builds awareness and understanding of the individual and collective threats and actions and informs our defenses.

THE REPORT

DoD Instruction 5200.39, *Critical Program Information (CPI) Protection within the Department of Defense*, dated July 16, 2008, requires DSS to publish a report that details suspicious contacts occurring within the cleared contractor community. The focus of the report is on efforts to compromise or exploit cleared personnel or to obtain illegal or unauthorized access to classified information and technology resident in the U.S. cleared industrial base.

Each year DSS publishes *Targeting U.S. Technologies: A Trend Analysis of Reporting*

from Defense Industry. In this report, the 14th annual *Targeting U.S. Technologies*, DSS provides a snapshot of its findings on foreign collection attempts. It provides a statistical and trend analysis that covers the most prolific foreign collectors targeting the cleared contractor community during fiscal year 2011 (FY11), compares that information to the previous year's report, and places that comparison into a larger context.

DoD Instruction 5200.39 requires DSS to provide its reports to the DoD CI community, national entities, and the cleared contractor community. This unclassified version of the report constitutes part of DSS' ongoing effort to assist in better protecting the U.S. cleared industrial base by raising general threat awareness, encouraging the reporting of incidents as they occur, identifying specific technologies at risk, and applying appropriate countermeasures. DSS intends the report to be a ready reference tool for security professionals in their efforts to detect, deter, mitigate, or neutralize the effects of foreign targeting. DSS released a classified version of this report earlier this year.

SCOPE/METHODOLOGY

DSS bases this report primarily on SCRs collected from the cleared contractor community. It also includes references to all-source Intelligence Community (IC) reporting.

DSS considers all SCRs received from cleared industry. It then applies analytical processes to them, including the DSS foreign intelligence threat assessment methodology. This publication is organized first by targeting region, then by collector affiliation, methodologies employed, and technologies, including the specific technology sectors targeted. It incorporates statistical and trend analyses on each of these areas. Each section also contains a forecast of potential future collection attempts against the cleared contractor community, based on analytical assessments.

COLLECTOR AFFILIATION DEFINITIONS

FIGURE 3

COMMERCIAL
Entities whose span of business includes the defense sector

GOVERNMENT AFFILIATED
Research institutes, laboratories, universities, or contractors funded by, representing, or otherwise operating in cooperation with a foreign government agency, whose shared purposes may include acquiring access to U.S. sensitive, classified, or export-controlled information

GOVERNMENT
Ministries of Defense and branches of the military, as well as foreign military attachés, foreign liaison officers, and the like

INDIVIDUAL
Persons who, for financial gain or ostensibly for academic or research purposes, seek to acquire access to U.S. sensitive, classified, or export-controlled information or technology, or the means of transferring it out of the country

UNKNOWN
Instances in which no attribution of a contact to a specific end user could be directly made

METHOD OF OPERATION DEFINITIONS

FIGURE 4

ACADEMIC SOLICITATION

Via requests for or arrangement of peer or scientific board reviews of academic papers or presentations, or requests to study or consult with faculty members, or applications for admission into academic institutions, departments, majors, or programs, as faculty members, students, fellows, or employees

ATTEMPTED ACQUISITION OF TECHNOLOGY

Via direct purchase of firms or the agency of front companies or third countries, these are attempts to acquire protected information in the form of controlled technologies, whether the equipment itself or diagrams, schematics, plans, spec sheets, or the like

CONFERENCES, CONVENTIONS, AND TRADE SHOWS

This refers to suspicious activity at such events—especially those involving dual-use or sensitive technologies that involve protected information—such as taking of photographs, making sketches, or asking of detailed technical questions

CRIMINAL ACTIVITIES

Via theft, these are attempts to acquire protected information with no pretense or plausibility of legitimate acquisition

EXPLOITATION OF RELATIONSHIPS

Via establishing connections such as joint ventures, official agreements, foreign military sales, business arrangements, or cultural commonality, these are attempts to play upon existing legitimate or ostensibly innocuous relationships to gain unauthorized access

OFFICIAL FOREIGN VISITS AND TARGETING

Via visits to cleared contractor facilities that are either pre-arranged by foreign contingents or unannounced, these are attempts to gain access to and collect protected information that goes beyond that permitted and intended for sharing

REQUESTS FOR INFORMATION

Via phone, email, or webcard approaches, these are attempts to collect protected information under the guise of price quote, marketing surveys, or other direct and indirect efforts

SEEKING EMPLOYMENT

Via résumé submissions, applications, and references, these are attempts to introduce persons who, wittingly or unwittingly, will thereby gain access to protected information which could prove useful to agencies of a foreign government

SOLICITATION OR MARKETING

Via sales, representation, or agency offers, or response to tenders for technical or business services, these are attempts by foreign entities to establish a connection with a cleared contractor vulnerable to the extraction of protected information

SUSPICIOUS NETWORK ACTIVITY

Via cyber intrusion, viruses, malware, backdoor attacks, acquisition of user names and passwords, and similar targeting, these are attempts to carry out intrusions into cleared contractor networks and exfiltrate protected information

TARGETING U.S. TRAVELERS OVERSEAS

Via airport searches, hotel room incursions, computer/device accessing, telephone monitoring, personal interchange, and the like, these are attempts to gain access to protected information through the presence of cleared contractor employees traveling abroad as a result of invitations and/or payment to attend seminars, provide training, deliver speeches, and the like

Pending a transition in technology categorization schemes, DSS continues to analyze foreign interest in U.S. defense technology in terms of the 20 sections in the Militarily Critical Technologies List (MCTL). The MCTL is a compendium of the science and technology capabilities under development worldwide that have the potential to significantly enhance or degrade U.S. military capabilities in the future. It provides categories and subcategories for DSS to use in identifying and defining targeted technologies.

This publication also makes reference to the Department of Commerce's Entity List. This list provides public notice that certain exports, re-exports, and transfers (in-country) to entities included on the Entity List require a license from the Bureau of Industry and Security. An End-User Review Committee (ERC) annually examines and makes changes to the list, as required. The ERC includes representatives from the Departments of Commerce, Defense, Energy, State, and, when appropriate, Treasury.

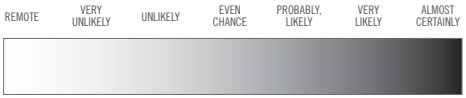
For FY11, the categories DSS used to identify methods of operation remained unchanged from the previous year. However, improved industry reporting and a refinement in DSS methodology resulted in more cases falling into the attempted acquisition of technology category that might previously have been labeled requests for information.

ESTIMATIVE LANGUAGE AND ANALYTIC CONFIDENCE

DSS uses the IC estimative language standard. The phrases used, such as *we judge*, *we assess*, or *we estimate*, and terms such as *likely* or *indicate* represent the agency's effort to convey a particular analytical assessment or judgment.

Because DSS bases these assessments on incomplete and at times fragmentary information, they do not constitute facts nor provide proof, nor do they represent empirically based certainty or knowledge. Some analytical judgments are based directly on collected information, others rest on previous judgments, and both types serve as building blocks. In either variety of judgment, the agency may not have evidence showing something to be a fact or that definitively links two items or issues.

Intelligence judgments pertaining to likelihood are intended to reflect the approximate level of probability of a development, event, or trend. Assigning precise numerical ratings to such judgments would imply more rigor than the agency intends. The chart below provides a depiction of the relationship of terms to each other.



The report uses *probably* and *likely* to indicate that there is a greater than even chance of an event happening. However, even when the authors use terms such as *remote* and *unlikely*, they do not intend to imply that an event will not happen. The report uses phrases such as *we cannot dismiss*, *we cannot rule out*, and *we cannot discount* to reflect that, while some events are unlikely or even remote, their consequences would be such that they warrant mentioning.

DSS uses words such as *may* and *suggest* to reflect situations in which DSS is unable to assess the likelihood of an event, generally because relevant information is sketchy, fragmented, or nonexistent.

In addition to using words within a judgment to convey degrees of likelihood, DSS also assigns analytic confidence levels based on the scope and quality of information supporting DSS judgments:

HIGH CONFIDENCE

- Well-corroborated information from proven sources, minimal assumptions, and/or strong logical inferences
- Generally indicates that DSS based judgments on high-quality information, and/or that the nature of the issue made it possible to render a solid judgment

MODERATE CONFIDENCE

- Partially corroborated information from good sources, several assumptions, and/or a mix of strong and weak inferences
- Generally means that the information is credibly sourced and plausible but not of sufficient quality or corroborated sufficiently to warrant a higher level of confidence

LOW CONFIDENCE

- Uncorroborated information from good or marginal sources, many assumptions, and/or mostly weak inferences
- Generally means that the information's credibility or plausibility is questionable, or that the information is too fragmented or poorly corroborated to make solid analytic inferences, or that we have significant concerns or problems with the sources

SPECIAL FOCUS AREA:

RADIATION-HARDENED MICROELECTRONICS

OVERVIEW

Ionizing radiation affects microelectronics and electronic systems during high-altitude flights and space operations, in particle accelerators, and in the proximity of fission or fusion reactions. In environments of high ionizing radiation, non radiation-hardened (rad-hard) microelectronics or insufficiently rad-hard microelectronics operationally degrade or fail due to single-event effects (SEEs).

Radiation hardening, by process or design, protects microelectronics and electronic systems from the effects of ionizing radiation. The Defense Security Service (DSS) produced this Special Focus Area assessment to alert cleared industry to the increasing foreign threat to rad-hard microelectronics and facilitate the implementation of mitigation strategies to counter that threat.

RADIATION HARDENING BY PROCESS – This method requires a foundry dedicated to hardened microelectronics. Recipe steps are the proprietary information of the manufacturer or classified by the U.S. Government. Radiation hardening by process can consist of proprietary steps added to a standard process of manufacturing a wafer so as to make it rad-hard. In such a case, there is no distinction between standard wafers and rad-hard wafers during much of the process.

RADIATION HARDENING BY DESIGN – This method relies solely on integrated circuit design and layout techniques to mitigate damage caused by ionizing radiation. Manufacturers design custom circuits for optimal performance in a targeted radiation environment, then fabricate them separately in a high-volume commercial approach. Radiation hardening by design presumes no access or visibility into the manufacturing process to enhance radiation tolerance reliability.

Foreign entities' interest in rad-hard microelectronics has risen over the past year, a trend reflected in industry reporting from fiscal year 2011 (FY11), which saw a 17 percent rise in reported targeting of rad-hard microelectronics from FY10. When analyzed collectively, these reports show a particularly strong interest in these technologies from regions with active or maturing space programs. Acquisition of a relatively small number of rad-hard microelectronics would likely assist foreign governments in developing their own radiation hardening processes or increase the reliability and effectiveness of their indigenous technologies already in use. Foreign entities focused collection activities on cleared contractors producing rad-hard memory whose resistance to the effects of ionizing radiation make them suitable for supporting manned and unmanned space activities.

Foreign governments are beginning to move to space for commercial telecommunications, increased command and control, and intelligence, surveillance and reconnaissance (ISR). Failure of microelectronics in space is costly. Whether SEEs are non-destructive or destructive, they can result in the total abandonment of a space system versus spending the time and money to fix the problem.

DSS analysis of industry documentation reveals that reported foreign collection attempts directed at cleared contractors that design, manufacture, and package rad-hard microelectronics increased 17 percent from FY10 to FY11. Near East and Europe and Eurasia collectors targeting rad-hard microelectronics, who were frequently noted in reporting in previous years, emerged as the most active collectors, with each region accounting for 26 percent of FY11 reports. Entities connected to East Asia and the Pacific, however, remained the top collectors, as represented by their 40 percent of total industry reporting.

Foreign entities appear to rely on three methods of operation (MOs) when targeting rad-hard designers, manufacturers, and packers: requests for information (RFIs); attempted acquisitions of technology; and academic solicitations. These MOs account for 97 percent of FY11 collection attempts reported by industry.

SINGLE-EVENT EFFECTS

SOFT ERRORS (non-destructive)

- SINGLE-EVENT TRANSIENT – Discharge of collected charges from an ionizing event
- SINGLE-EVENT UPSET – Changes of memory or register bits caused by a single ion interaction on the chip
- SINGLE-EVENT FUNCTIONAL INTERRUPTION – Ionizing events cause temporary loss of device functionality

HARD ERRORS (destructive)

- SINGLE-EVENT LATCHUP – Ionizing events cause circuit lockup and/or catastrophic device failure
- SINGLE-EVENT BURNOUT – Destructive burnout due to high current conditions
- SINGLE-EVENT GATE RUPTURE – Rupture of gate dielectric due to high electrical field conditions
- STUCK BITS – Unalterable change of state in a memory element

The packaging of microelectronics is as important as the design and manufacturing of integrated circuitry. Timothy May of Intel Corporation noted the first packaging-induced soft errors in 1979. In an article entitled "Alpha-Particle-Induced Soft Errors in Dynamic Memories," first published in *IEEE Transactions on Electron Devices*, May analyzed single-event upsets occurring due to uranium and thorium decay in microelectronics packaging.

REGIONS OF ORIGIN

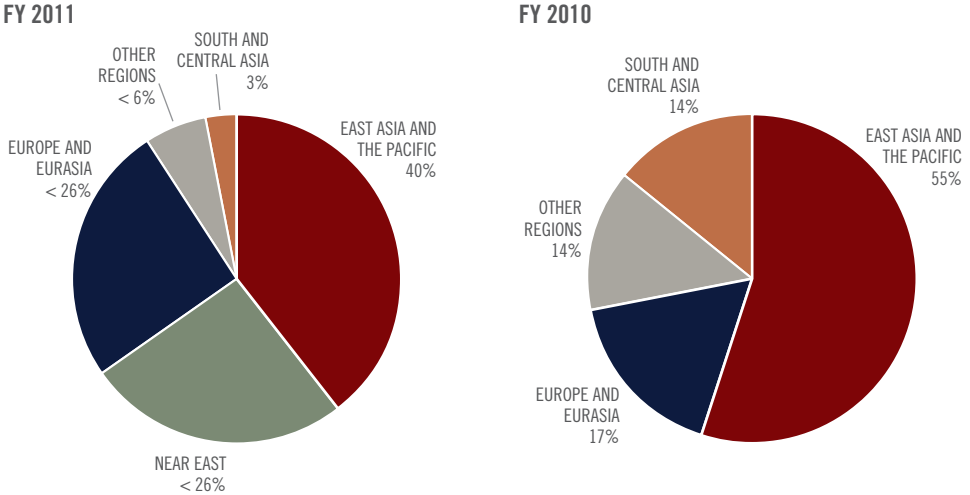
- East Asia and the Pacific
  - Requested specific quantities of rad-hard static random-access memory (SRAM), optical transceivers, and databus controllers
  - Primarily used commercial entities with RFI as MO
- Near East
  - Primarily used student requests to attempt to elicit information from leading experts
- Europe and Eurasia
  - Attempted to acquire specific quantities of rad-hard SRAM and optical transceivers
  - Primarily attempted acquisition by commercial entities

EAST ASIA AND THE PACIFIC

East Asia and the Pacific entities accounted for 40 percent of FY11 industry reporting on the targeting of rad-hard microelectronics. In many requests for rad-hard microelectronics from entities in this region, the requestor solicited the U.S. cleared manufacturer for a specific quantity of the product, implying that there was an immediate need from a customer for the microelectronics.

Twelve East Asia and the Pacific countries have active or planned space programs. Three with the most active space programs are spending $4 billion annually for launching space platforms, controlling satellites, and observing space. Expanding East Asia and the Pacific economies are using space-based technologies to communicate, command, and control across growing land and sea lines of communications.

REGIONS OF ORIGIN
FIGURE 5



However, many of these countries do not possess the technical proficiency to design, manufacture, and rad-hard microelectronics capable of withstanding sustained cosmic radiation. These countries seek Western Hemisphere and Europe and Eurasia rad-hard microelectronic suppliers to enable them to assemble space-worthy systems that will withstand high radiation for a sustained period.

Analyst Comment: Based on reporting from cleared industry, it is likely that East Asia and the Pacific collectors have immediate needs for rad-hard microelectronics for various commercial and military programs. The lack of East Asia and the Pacific technical proficiency to design and manufacture space-worthy rad-hard microelectronics coupled with East Asia and the Pacific commercial entities' specific requests for the technology from cleared industry likely signify that the microelectronics markets in East Asia and the Pacific are unable to meet the strategic goals of included countries. (Confidence Level: Moderate)

NEAR EAST

Near East entities were the second most active FY11 collectors of rad-hard microelectronics information, as reflected in attempts reported by industry. Near East entities are consistently among the most active collectors of U.S. technology overall, but this is the first year in which industry reporting portrayed a particular and deliberate effort to obtain restricted rad-hard information from U.S. universities researching radiation hardening. To do so, Near East entities relied on academic solicitation, in the form of student requests seeking restricted rad-hard information

CASE STUDY

On November 11, 2010, a Colorado-based cleared contractor received a request from an individual representing an East Asia and the Pacific commercial entity for rad-hard SRAM. The individual did not specify the end use or end user of the rad-hard SRAM; however, according to the commercial entity's website, an East Asia and the Pacific military is a customer of the company.

On November 17, 2010, the same Colorado-based cleared contractor reported receiving an almost identical request from another individual representing a separate East Asia and the Pacific commercial entity. In an email, the individual requested to purchase a large number of the company's rad-hard microelectronics for an East Asia and the Pacific customer. The individual did not further identify the intended end use or end user of the requested products. The quantities and specifications of the requested rad-hard microelectronics follow:

- 2000 pieces of 512K rad-hard SRAMs with a standard microcircuit drawing
- 2000 pieces of 256K rad-hard SRAMs with a standard microcircuit drawing

These commercial entities' collection activities demonstrate the aggressive nature of the attempts to acquire U.S. rad-hard microelectronics from cleared contractors.

Analyst Comment: Although a connection between these commercial entities cannot be confirmed, given the similarity of the requests over a relatively short period, it is likely that the end user of the rad-hard SRAM would have been customers within the same East Asia and the Pacific country. East Asia and the Pacific commercial entities and their proxies will likely continue to employ these MOs in attempts to circumvent U.S. export laws covering this restricted technology. (Confidence Level: High)

from cleared contractors and research and employment opportunities at facilities specializing in radiation hardening.

Analyst Comment: Near East governments' association with universities likely provides an avenue for procurement of restricted rad-hard microelectronics research and development under the guise of academic cooperation for the advancement of sciences and technologies. Rad-hard microelectronic information garnered through academic cooperation with U.S. universities would almost certainly advance current Near East space capabilities and provide a foundation for long-term space and military advancements in hardening of microelectronics. (Confidence Level: High)

EUROPE AND EURASIA

Europe and Eurasia entities' targeting of rad-hard microelectronics increased from the previous year, now representing 26 percent of the FY11 reported total. Although collectors connected to Europe and Eurasia are consistently among the top foreign entities attempting to collect U.S. technology, this is the first year that reporting suggested a concerted effort by Europe and Eurasia collectors to acquire rad-hard microelectronics from cleared contractors. In almost every reported incident, Europe and Eurasia commercial entities attempted to acquire specific numbers of rad-hard microelectronics.

Europe and Eurasia leaders have stated their beliefs that national defensive capabilities are directly related to strong microelectronics design and manufacturing processes. For over ten years, Europe and Eurasia leaders have discussed the need to end reliance on foreign microelectronics. In some countries, over 90 percent of the microelectronics used in defense systems are imported.

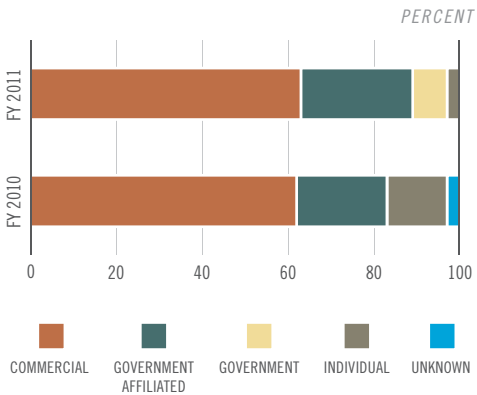
Analyst Comment: Although indigenous microelectronics design and manufacturing and radiation hardening research appear to be a priority among Europe and Eurasia strategic technology pursuits, regional producers almost certainly cannot provide U.S.-quality and -quantity rad-hard microelectronics. The attempted acquisition of specific numbers of rad-hard microelectronics probably means there is a specific Europe or Eurasia program requiring certain capabilities to be found only in U.S. cleared contractor-manufactured rad-hard microelectronics. (Confidence Level: Moderate)

AFFILIATIONS AND METHODS OF OPERATION

Once DSS established the collecting entity's country of origin, it identified its affiliation and the MO used. The following paragraphs detail the top affiliations and MOs identified in FY11 reporting from cleared industry.

COLLECTOR AFFILIATIONS

FIGURE 6



DSS analysis of industry reporting shows that collectors affiliated with East Asia and the Pacific primarily relied on commercial entities to obtain sensitive or classified U.S. information and technology in FY11. They did so using two MOs. The RFI was used most often, employing email to seek price quotes and technical information regarding rad-hard technology. At 45 percent, attempted acquisition of technology via email was the other MO East Asia and the Pacific commercial entities used to attempt

to circumvent U.S. laws restricting the export of rad-hard microelectronics. In emails, when individuals representing commercial entities were notified that the U.S. cleared contractor would need an export determination prior to a transaction, the U.S. manufacturer either did not receive a response or the suspicious entity provided a U.S. address and reiterated the same request.

In contrast, Near East entities' efforts, as reflected in industry reporting, relied solely on government-affiliated university students who made academic solicitations to rad-hard research facilities. Radiation reliability experts at a cleared U.S. university received numerous emails and curricula vitae (CVs) from Near East university students expressing interest in obtaining research positions under their supervision. Often the résumé or CV demonstrated a history of research in microelectronics and radiation effects on microelectronics. In one email, the collector cited experience working in a laboratory studying space radiation effects on satellite systems.

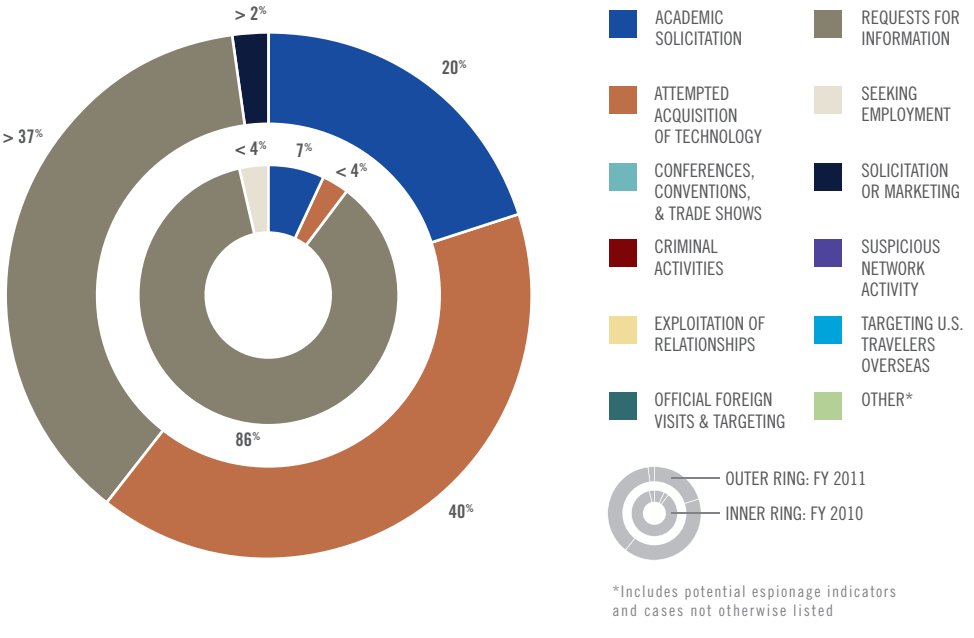
Europe and Eurasia entities, like East Asia and the Pacific entities, relied on RFIs and the attempted acquisition of technology through commercial collectors in attempting to acquire sensitive rad-hard technology in FY11. RFIs ranged from requesting data sheets for a U.S. contractor's rad-hard technology to requesting a list of a U.S. company's distributors in a particular foreign country.

TARGETING RAD-HARD SRAM

Reporting from cleared industry pointed to SRAM being the most sought after rad-hard microelectronics technology. SRAM is a type of memory that is faster and more reliable than the more common dynamic random-access memory (DRAM). While DRAM supports access times of about 60 nanoseconds, SRAM can support

METHODS OF OPERATION

FIGURE 7



*Includes potential espionage indicators and cases not otherwise listed

access times as low as 10 nanoseconds. In addition, its cycle time is much shorter than that of DRAM because it does not need to pause between accesses. It is also much more expensive to produce, so SRAM is usually employed only as a memory cache.

The following table shows the number and type of rad-hard SRAMs that entities from East Asia and the Pacific and Europe and Eurasia requested, according to reporting from cleared industry in FY11.

TARGETED STATIC RANDOM ACCESS MEMORY

COUNTRY	TYPE	QUANTITY
EAST ASIA AND THE PACIFIC	128K, 256K, 512K	> 4338
EUROPE AND EURASIA	128K, 512K	> 6640

Analyst Comment: Although previous assessments found that these requests for rad-hard microelectronics were likely intended to fill immediate requirements in commercial and military programs, there is an even chance that the requestor could divert rad-hard microelectronics to commercial or government organizations specializing in reverse-engineering. (Confidence Level: Moderate)

According to IC reporting, multiple foreign companies and government labs conduct failure and vulnerability analysis and reverse-engineering (FAVA-RE) to validate microelectronics design. Although the FAVA-RE process is legal in the United States to discover and analyze circuit designs, it can reveal sensitive information contained in microelectronics and proprietary fabrication processes.[1,2]

Analyst Comment: Success by East Asia and the Pacific and Europe and Eurasia companies in the illegal acquisition of U.S. rad-hard SRAM would probably result in the revelation of sensitive information and proprietary fabrication processes. The likely diversion of these items to university or government labs capable of conducting FAVA-RE analysis would probably spur indigenous development of rad-hard microelectronics. This would likely decrease the funding that entities in these regions would have to dedicate to researching radiation hardening techniques and increase world-wide competition to supply rad-hard microelectronics, potentially impacting U.S. companies' sales. (Confidence Level: Moderate)

OUTLOOK

Reporting from industry confirms that U.S. rad-hard microelectronics are of significant interest to collecting entities in several regions. They are likely to use a variety of MOs by commercial, government-affiliated, government, and individual entities to attempt to collect rad-hard microelectronics information or technology. **(Confidence Level: High)**

DSS assesses that agents from East Asia and the Pacific, the Near East, and Europe and Eurasia in particular will likely continue their efforts to collect U.S. rad-hard microelectronics in the immediate future, largely reliant on the RFI and attempted acquisition of technology MOs. **(Confidence Level: High)**

With more countries moving toward conducting space activities and operations, DSS assesses that it is likely the demand for rad-hard microelectronics will dramatically rise over the coming years, especially as once-torpid economies grow and outdated militaries modernize and move terrestrial communication and ISR activities into space. As U.S. companies continue to increase rad-hard microelectronics' speed and decrease their susceptibility to ionizing radiation, foreign entities will likely increase their targeting of cleared contractors' design, manufacturing, and packaging of rad-hard microelectronics. **(Confidence Level: Moderate)**

CASE STUDY: A DATE FOR THE PROM

On September 30, 2011, two Chinese nationals were sentenced to 24 months in prison for participating in a conspiracy to violate the Arms Export Control Act. Hong Wei Xian, also known as Harry Zan, and co-conspirator Li Li, also known as Lea Li, attempted to acquire and smuggle rad-hard microchips out of the United States for an agency controlled by the Chinese government.

Xian and Li, representing Beijing Starcreates Space Science and Technology Development Company Limited, engaged in the importing and selling of programmable read-only memory (PROM) to China Aerospace Science and Technology Corporation. Between April 2009 and September 2010, they contacted a company in the Eastern District of Virginia requesting to purchase thousands of rad-hard PROMs. China Aerospace is controlled by the government of China and researches, designs, develops, and produces strategic and tactical missiles and exo-atmospheric launch vehicles.

Xian and Li sought PROMs specifically designed to withstand sustained radiation bombardment in space. The conspirators knew the PROMs were export-controlled, but they did not seek licenses because doing so would have revealed the ultimate end user of the rad-hard microelectronics—China Aerospace. Xian and Li conspired to break up orders into multiple shipments in an attempt to circumvent U.S. export-control restrictions on the sale of U.S. Munitions List technology to China.[3]

Analyst Comment: This collection attempt and thwarted scheme demonstrate an approach used by collectors to illegally acquire rad-hard microelectronics. Based on investigations, it is almost certain that China Aerospace is driving its commercial suppliers to collect U.S.-manufactured rad-hard microelectronics. (Confidence Level: High)

EAST ASIA AND THE PACIFIC

OVERVIEW

Foreign collectors connected to this region remain dominant among those attempting to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. The East Asia and the Pacific region maintained the same 43 percent share of the total in fiscal year 2011 (FY11) as in FY10. This steady share represented an increase in the number of reported cases of more than 75 percent from FY10 to FY11.

Despite this continuity in East Asia and the Pacific's portion of the year's total reports from industry, some interesting shifts occurred from FY10 to FY11 within the data applicable to the region. The most significant overall trend within industry reporting was the increased clustering in the data among collector affiliations as well as methods of operation (MOs).

There was also increased quality of industry reporting, especially of the new top MO, suspicious network activity (SNA), which resulted in improved attribution by the Defense Security Service (DSS) Counterintelligence (CI) Directorate.

Commercial entities, in their 31 percent of total reported attempts in FY11, were probably attempting to gain opportunistic access to sensitive technologies for subsequent resale to other East Asia and the Pacific actors.

Additionally, industry reporting reflects a large number of cases (a combined 27 percent of the year's total) in which East Asia and the Pacific-connected entities reportedly attempted to establish a relationship with a cleared contractor, either through academic exchange, commercial deals, or individual employment. While these incidents did not suggest specific targeting of technology, they remain of interest due to the possibility that such relationships could lead to future opportunities for exploitation.

Multiple countries within East Asia and the Pacific perceive themselves as being surrounded by threats, including from each other. This leads them to believe that they must significantly upgrade their military capabilities, building their capacity for deterrence. Many of those countries also desire to make their militaries more self-reliant, although at present they remain significantly dependent on the acquisition of military technology from abroad.

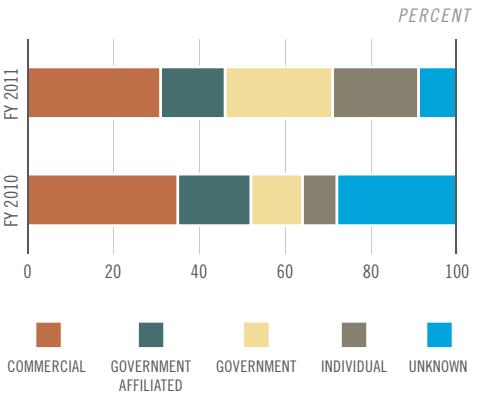
Reflecting the significant scope of these military modernization efforts ongoing in the region, requests originating in East Asia and the Pacific sought technologies found in nearly every section of the Militarily Critical Technologies List (MCTL). As in FY10, information systems (IS) was the single most targeted technology category, although reduced from FY10's 25 percent to 13 percent. However, the majority of those incidents were attributed to cyber actors and were non-specific in nature. In addition to IS technology, lasers, optics, and sensors (LO&S) technology remained a top identifiable targeting priority.

Despite the aforementioned frictions that exist between some countries in the East Asia and the Pacific region, unique relationships continue to exist between some of its geopolitical entities. Close economic ties between some of these entities continue to make third-party technology transfers a possibility. Some locations within the region are not governed by enforcement regimes that are sufficiently robust to adequately implement transit/transshipment license controls, creating popular diversion points for third-party transfers. Some East Asia and the Pacific collectors attempt to obtain U.S. technology to sell to third countries hostile to U.S. interests.

COLLECTOR AFFILIATIONS

East Asia and the Pacific entities targeting cleared industry were characterized by significantly variegated affiliations. Of the five categories of collector affiliation, four increased in number of reports from industry concerning East Asia and the Pacific, while the unknown category decreased in number of reports, and its share of all FY11 reports went from 28 to nine percent. Two categories—the top category, commercial, and government-affiliated—increased in number of reports, but fell slightly in percentage of the total, commercial from 35 to 31 percent and government-affiliated

COLLECTOR AFFILIATIONS
FIGURE 8



from 17 to 15 percent. The final two categories increased both in number of reports and share of the total, government from 12 to 25 percent and individual from eight to 20 percent.

Analyst Comment: The pattern revealed in industry reporting is that some East Asia and the Pacific collection entities use a diversified and persistent approach, often employing multiple collector types and MOs at the same time. When one entity fails, a second entity, often with a different affiliation, reengages the cleared contractor in pursuit of the same technology. DSS assesses that some East Asia and the Pacific collection campaigns probably represent coordinated national strategies. (Confidence Level: Moderate)

In particular, increased attribution of cyber incidents highlights the multifaceted nature of the threat to U.S. information and technology originating in East Asia and the Pacific. Overt collection efforts by commercial entities run in parallel with aggressive cyber collection activities, which target cleared contractor networks in attempts to exfiltrate data relating to sensitive U.S. information and technologies and the companies that produce them.

In some cases within East Asia and the Pacific, commercial entities are known to be tightly interwoven with other East Asia and the Pacific actors, relationships that cleared industry reporting and subsequent DSS analysis confirmed in FY11. This is especially so with regard to attempted technology collection and defense sales, as other collectors often use commercial entities to collect information on U.S. technology and programs. Commercial entities involved ran the gamut from large international corporations to small, privately owned companies with ten or fewer employees.

Analyst Comment: In many cases involving commercial entities, requestors failed to identify intended end users or uses. However, cleared industry reported frequent demonstrations of interest in a very specific system or capability from multiple separate entities, making it likely that they were acting on behalf of a common end user. (Confidence Level: Moderate)

While some requests for information (RFIs) from or attempts to purchase components by commercial entities resolved to innocuous entities, industry reporting cited a significant number of instances in which the companies and individuals behind these requests had appeared in previous industry or Intelligence Community (IC) reporting. Many of these entities were based in third countries, including the United States, Canada, and European nations, but could be attributed to East Asia and the Pacific end users.

Analyst Comment: Some collectors were likely attempting to circumvent U.S. export laws that apply different regulations to different locations within East Asia and the Pacific. It is likely that many of these collectors were acting as illicit technology brokers for other East Asia and the Pacific actors. DSS assessed that most requests made by entities identified in IC reporting as illicit technology brokers very likely reflected tasking by end users to acquire specific components, systems, or technologies. Additionally, inquiries from technology brokers associated with particular East Asia and the Pacific entities which mirrored otherwise innocuous requests can identify otherwise unidentified or intentionally misidentified end users. (Confidence Level: High)

A substantial minority of the commercial cases consisted of interest from companies in establishing business relationships with cleared contractors, either as distributors in the East Asia and the Pacific market or as suppliers of components for integration into systems under development by the contractors.

Analyst Comment: Integration of foreign-manufactured components into U.S. defense systems is a growing concern within the

IC and U.S. cleared industrial base. While the majority of cases in which companies attempted to establish supply chain relationships with cleared contractors appear unlikely to be directed efforts to infiltrate the contractors, DSS CI deemed these cases likely to be of intelligence value due to the identity of the companies interested in establishing connections with cleared contractors. (Confidence Level: Moderate)

In many other cases the acquisition mechanisms employed by East Asia and the Pacific militaries are manifestations of complex and very opaque systems of competing interests sharing common goals and end users. There are many unknowns concerning commercial entities, other collectors, and the varying nature of the relationships between them. This frequently makes specific attribution of commercially originated requests to the ultimate requestors and end users uncertain at best, and concrete findings of any kind difficult to establish.

Overt requests usually come from non-traditional collectors, such as commercial and academic entities. In the majority of cases associated with commercial entities, East Asia and the Pacific companies contacted cleared contractors and attempted to acquire sensitive, export-controlled, or dual-use components and systems by overt means.

Analyst Comment: Most separate incidents appeared to be innocuous, involving entities which did not appear to be acting in a duplicitous manner and which had not been cited in classified reporting for previous suspicious activities. Most of the commercial collectors involved maintain no apparent ties to intelligence services, and in many instances are likely motivated by financial gain. (Confidence Level: Moderate)

Additionally, the sharp rise from FY10 to FY11 in the number of reported cases attributed to government entities and the doubling of their share of the total, while noteworthy, should not be viewed as reflecting new entry into attempted technology collection by governments from East Asia and the Pacific, but rather as the result of refined attribution by DSS and increased quality of reporting from industry. Through security education and other means of generating increased awareness, cleared contractors increasingly recognized the threat posed by seemingly innocuous contacts and reported these incidents with greater frequency and attention to technical indicators. As a result of this increased fidelity, DSS attributed a large number of cases to government entities which would likely have been designated with the unknown affiliation in FY10.

Industry-reported cases attributed to individuals provided 20 percent of the FY11 total. Students attempting to obtain postdoctoral positions or other employment opportunities with cleared contractors dominated reported attempts, and the majority of these reports came from cleared contractors associated with U.S. universities. While available information can seldom establish a direct connection between foreign intelligence services and most, if any, of the students and academics who contacted cleared contractors, IC and law enforcement reporting provides numerous instances in which East Asia and the Pacific students have exploited access to sensitive or classified technologies to support parallel research and development (R&D) efforts in their home countries.

Analyst Comment: While most or all of these individuals are likely legitimately interested in obtaining positions with cleared contractors, placement within those facilities would likely offer academics the opportunity to exploit their access to personnel, information, and technologies resident in those facilities. Moreover, some individuals have used the bona fides of U.S. universities to acquire otherwise inaccessible components, materials, and systems for end users in their home countries. Review of industry and IC reporting leads DSS to assess that many

academics and their sponsoring institutions very likely view placement in U.S. facilities as supporting national technology collection goals. (Confidence Level: High)

METHODS OF OPERATION

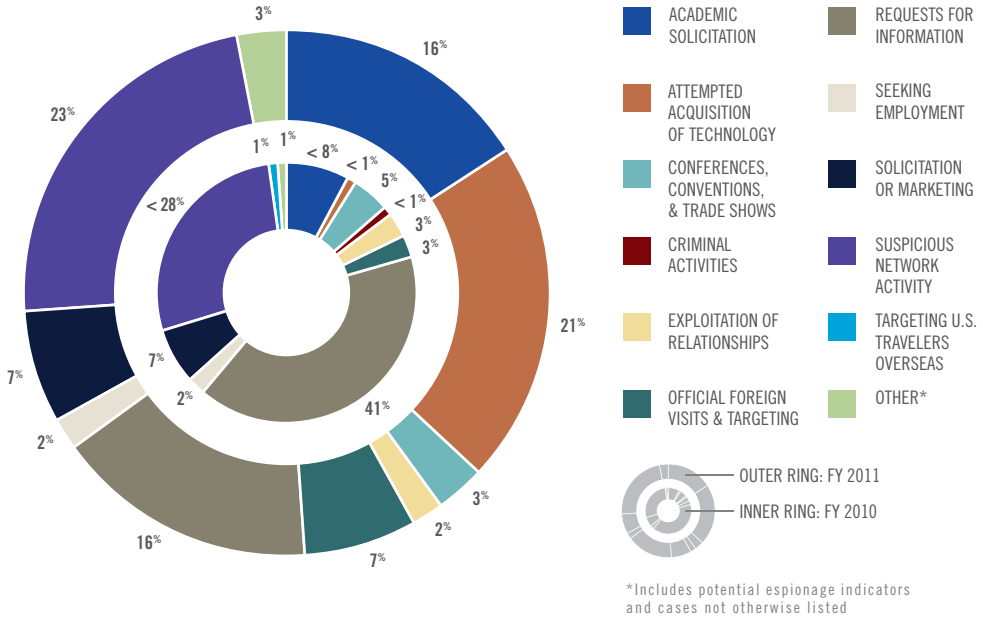
The data on frequency of use of different MOs by collectors from East Asia and the Pacific fell into two tiers. SNA, attempted acquisition of technology, RFI, and academic solicitation each accounted for 16 percent of the total or more, whereas the portion that all other MOs accounted for individually remained in the single digits.

A major change in the DSS categorization method led to many reports that in previous years would have been labeled RFI being listed as attempted acquisition of technology, moving the latter category from low in the second tier in FY10 to the second highest category in FY11, at 21 percent. Within the upper tier, this dropped RFI from the top position in FY10 at 41 percent down to third in FY11 at 16 percent. It was joined at 16 percent by academic solicitation, up from eight percent in FY10. Partly as a result of the accounting change but also due to a continued increase in illicit cyber activity, SNA surged to the top of the region's MO list in FY11 at 23 percent of the total.

Together these four MOs accounted for over three-quarters of the East Asia and Pacific total. The next most common MOs, as measured by reports from industry, were official foreign visits and targeting and solicitation or marketing services, at only seven percent apiece.

The new top MO practiced by East Asia and the Pacific collectors, SNA, experienced increased quality of reporting from industry, which led to significant refinement in attribution. Increased clarity allowed DSS analysts to discard most reports of unsophisticated attempts to access cleared

METHODS OF OPERATION
FIGURE 9



*Includes potential espionage indicators and cases not otherwise listed

contractor networks through tactics such as brute-force attacks, attributing these actions to criminal rather than intelligence actors. Notably, almost all of the SNA reporting deemed to be of intelligence value resulted from spear phishing emails with malicious attachments received by cleared contractors.

Analyst Comment: While FY11 industry reporting of spear phishing emails significantly increased over FY10, this probably does not represent greater use of that vector, or delivery mechanism, but rather increased cleared contractor awareness, recognition, and acknowledgement of such collection attempts. In those instances in which a compromise occurred but no vector was identified, DSS CI assessed that the initial intrusion was likely achieved through an unidentified spear phishing email. (Confidence Level: Moderate)

Although attempted acquisition of technology and RFIs (accounting for 21 and 16 percent, respectively, of the total collection attempts reported by industry in FY11) are separated into different reporting categories, these MOs are employed very similarly, and both are associated very closely with commercial entities. Typically, reports of either type resulted from commercial entities requesting sensitive components or specifications through the cleared contractor's sales department, with many initial contacts failing to disclose the intended end user and use.

In most instances of attempted acquisition of technology, the entity sent an email with a purchase order for the cleared contractor's products. The second most common MO reported was RFI, again most commonly executed via email, web-card submission, or telephone call. RFIs often begin with general questions whose answers, if supplied, could be used to confirm or deny information on the technology or system, opening the way to more pointed and sensitive questions.

Both attempted acquisition and RFI represent a low-risk, high-reward approach to collection. If the request is questioned or deemed inappropriate, the entity can claim it was made in good faith with no knowledge of restrictions. If the request goes unchallenged, it provides immediate reward as well as building a potential relationship that can be exploited in the future. If the acquisition attempt is successful, it provides opportunity for reverse-engineering and significant savings in R&D costs.

Some requests initially appear innocuous, but gradually reveal themselves as apparent attempts to acquire sensitive or controlled technology for East Asia and the Pacific end users. In a handful of reports, entities openly or implicitly stated their intention to circumvent export controls by transshipping purchased components through third countries.

Analyst Comment: While U.S. export controls prevent many collection entities from purchasing sensitive, dual-use components and systems, it is likely that unauthorized East Asia and the Pacific end users have acquired components through entities located in countries without such restrictions and the falsification of end-use documents. (Confidence Level: High)

Academic solicitations jumped significantly as a percentage of industry reports, from eight percent in FY10 to 16 percent in FY11, and more than tripled in the number of reported approaches. This largely resulted from increased industry reporting of attempts by students and postdoctoral researchers to obtain positions with cleared contractors. U.S. universities reported receiving by far the greatest number of academic solicitations noted in DSS reporting. Reporting also reflected a significant number of solicitations in which individuals affiliated with East Asia and the Pacific universities and institutes requested research and other academic information produced by cleared contractor employees.

Analyst Comment: While much of the requested research material was both publicly available and basic in nature, attempts to acquire information directly from the author present the opportunity to expand conversations into areas outside the scope of the initial paper and into more sensitive areas of the cleared contractor employee's current research. Taking advantage of the academic predilection to share information in this way almost certainly presents an excellent avenue to support military research.
(Confidence Level: Moderate)

TARGETED TECHNOLOGIES

In FY11, the four most common targeted technologies by collectors connected to East Asia and the Pacific were IS; LO&S; electronics; and aeronautics systems, just as they were in FY10. However, the top technology, IS, actually fell slightly in number of reports and by almost half in share, from 25 to 13 percent. Technologies in the next three sections of the MCTL all increased in number of reports, but LO&S decreased in percentage from 13 to ten percent, electronics increased slightly from seven to eight percent, and aeronautics maintained its share unchanged at eight percent.

Even more interesting variation occurred in the second tier of technologies. The next four most commonly targeted technologies all increased in number of reports from industry. But while positioning, navigation, and time merely maintained its five percent share of the total and marine systems declined to five percent, two categories, armaments and energetic materials and space systems, doubled in the number of reports year over year and increased in proportional share; the former actually doubled its share to six percent.

East Asia and the Pacific's increased practice of the SNA MO meant that there were more incidents in which the specific data targeted could not be determined; in such cases, DSS analysts frequently

TARGETED TECHNOLOGY
FIGURE 10

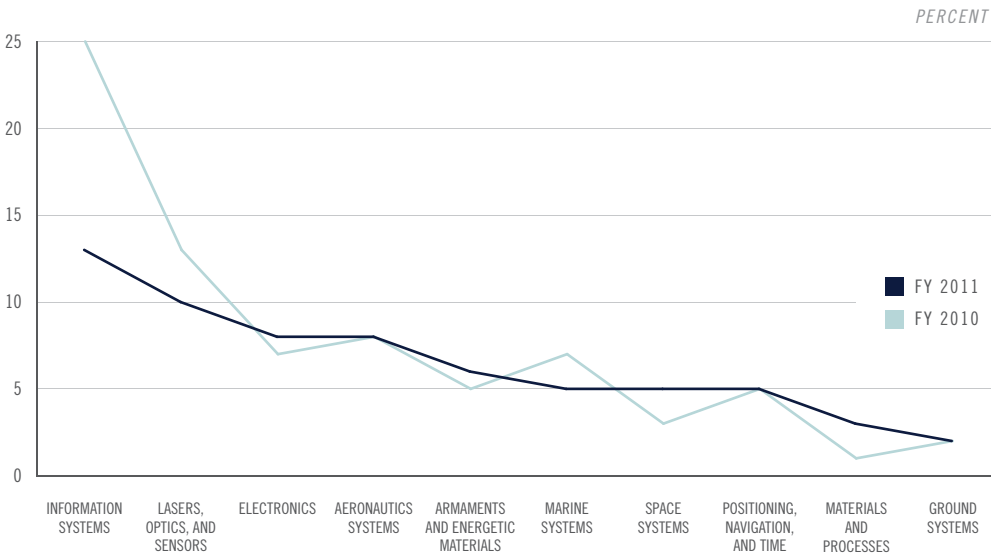


Figure illustrates the top ten most targeted technology categories in FY11 compared with the reporting statistics for the same categories from FY10.

identified attempts with the primary technologies affiliated with the subject facility. However, this determination was made on a case-by-case basis, and in many instances it was not possible to associate a cyber incident with a specific MCTL section. In instances in which entities contacted a facility more directly but still did not mention a specific product or technology, such as the case with many student requests and business solicitations, analysts regarded the request as undefined with regard to MCTL category.

Despite the fall in the number of cases attributed to IS technology in both numbers and share of the total, it remained the most commonly targeted section. Most cases involving IS as well as aeronautics technology originated from cyber actors and were nonspecific in nature. In the non-cyber cases in which entities expressed interest in specific IS technologies, software programs were the most common targets, particularly those supporting satellites, radar, and signals.

Analyst Comment: Whereas in FY10 DSS analysts frequently assessed the targeted technologies based on incomplete information regarding the incident or targeted facility, the improvement in the quality of industry reporting allowed for better identification of targeted technologies. DSS analysts did not designate a targeted MCTL category in those incidents which suggested multiple targeted technologies or targeting of non-MCTL information. The sharp decrease in reporting regarding IS compared to the previous year is almost certainly due to improved attribution techniques and results. (Confidence Level: Moderate)

Requests from East Asia and the Pacific that fell within the category of LO&S encompassed a wide range of technologies. Some of the most commonly targeted were advanced radar and sonar systems suitable for upgrading and modernizing the region's sometimes antiquated missile, air, and maritime defense networks and improving command and control capabilities. Similarly, interest in unmanned aerial systems within the aeronautics category is consistent with a need to survey and monitor against neighbors' infiltration and attack. The geographical, topographical, and geopolitical landscape of the East Asia and the Pacific region makes such technologies a matter of high priority to regimes and militaries within the region.

Among the most targeted LO&S technologies were shortwave infrared optical systems, which are useful for measurements imaging for a variety of civilian and military purposes, ranging from agricultural to battlefield targeting applications. While some of the requests received were ostensibly civilian in nature, others made specific reference to military hardening and specifications exceeding those required for civilian use. Collection entities also sought a variety of laser technologies.

Requests for electronics technology accounted for eight percent of the total, slightly more than the previous year. This section also encompassed a wide range of sub-technologies. Many requests targeted a variety of antenna systems as well as space-qualified equipment. Based on the specifications requested, the items were appropriate for use in satellite communications, but could also be applied to a number of other end uses. Of additional note, industry reporting concerning attempts by East Asia and the Pacific students to obtain placement with cleared contractors showed that a large number of requests were sent to professors and employees working in areas of study that have both commercial and military uses, including sensors, positioning, and detection technologies.

Analyst Comment: Many East Asia and the Pacific universities and research institutes have associations with their nations' militaries. Reported interest by such institutions in the study of the

technologies and applications discussed above is noteworthy, as many of the requesting academics and students are likely to contribute to military R&D following completion of their studies. (Confidence Level: Moderate)

The noted increase in industry reports concerning the attempted collection of armaments and energetic materials and space systems technology involved integrated circuits, switches, amplifiers, and other electronic components with applications to a variety of systems, which could include missile systems or weapons countermeasure systems.

Analyst Comment: It is not apparent what, if any, specific requirements have driven these increases. They are likely a result, at least in part, of general modernization and upgrade efforts and/or attempts to reverse-engineer any technology obtained to produce indigenous variants for domestic use and foreign sales. (Confidence Level: Moderate)

OUTLOOK

As anticipated in previous years' versions of this publication, industry awareness of the threat posed by entities from East Asia and the Pacific has consistently increased year over year, and will likely lead to greater numbers of reports from cleared contractors and further identification of entities of concern. However, even as this awareness has grown, DSS has not observed any discontinuities from the reported MOs that entities from the region have used over preceding years, providing evidence that those methods continue to be useful in acquiring U.S. technologies. Therefore, DSS CI assesses that East Asia and the Pacific entities will continue to aggressively target cleared contractors through both computer network exploitation activities and the overt means used predominantly by non-cyber actors. **(Confidence Level: High)**

East Asia and the Pacific commercial entities continued to lead all other collector affiliations. This points to some degree of success by those entities, so their collection efforts are likely to continue. The opaque but arguably close relationship between governments and industry within East Asia and the Pacific often manifests itself in collection patterns characteristic of coordinated collection strategies. DSS assesses it is very likely that commercial entities will lead the accounting of the East Asia and the Pacific collection effort in FY12, but may receive significant support from government, government-affiliated, and individual entities. **(Confidence Level: High)**

This year's industry reporting does not suggest any single, common, driving goal behind technology collection efforts beyond the continuing frictions in relations between countries within the region and between countries within the region and those from outside. It is likely that these general concerns will continue to drive the great scope of efforts to modernize and upgrade the somewhat backward and antiquated existing militaries of countries within East Asia and the Pacific, and thus collection attempts related to them. **(Confidence Level: High)**

The breadth of systems, components, and capabilities that East Asia and the Pacific collection entities target underscores these frictions and the dangers to which various regimes consider themselves subject. The immediacy of the perceived threats calls for a high priority on border surveillance and air and maritime defenses. Consequently, LO&S, particularly sensor technologies, will almost certainly remain a high priority. **(Confidence Level: High)**

Multiple regimes within East Asia and the Pacific seek advanced technology to transform their militaries from quantitative to qualitative forces. Technology can be a force multiplier crucial to success in that transformation. In pursuit of this, collection

entities will almost certainly continue to place a high priority on IS technologies and aeronautics systems technologies. However, the broadness of the goals pursued will likely drive collection entities, whether tasked or not, to target a very wide array of technology categories across nearly the entire MCTL. **(Confidence Level: High)**

Requests for sensitive or classified information and technology resident in the U.S. cleared industrial base, if successful, would likely directly support development of new military systems or upgrades to existing capabilities. Such requests also emphasize the degree to which indigenous research capabilities in the region, while improving, continue to rely on acquisition of foreign technology to further ongoing development efforts and will likely continue to do so in the foreseeable future. **(Confidence Level: High)**

Similarly, based on industry reporting, East Asia and the Pacific collection entities practice a diverse suite of collection methodologies, with significant effort exerted in SNA, attempted acquisition of technology, RFIs, and academic solicitation. These MOs are either "stand-off" methods practiced from a distance or arguably innocuous, and the use of this combination of methods is very likely to continue. **(Confidence Level: High)**

CASE STUDY: "IF AT FIRST YOU DON'T SUCCEED..."

In September 2011, a Massachusetts-based cleared contractor received a request for an export-controlled amplifier from a company based in East Asia and the Pacific. The company did not state the intended end user or end use in the initial contact.

Reporting from the same cleared contractor indicated that the model of amplifier requested had been the subject of numerous previous requests, including several from companies located in the U.S. and third countries. Several of these requests listed other East Asia and the Pacific actors as the intended end users.

Reporting from other cleared contractors cited several of the entities requesting the equipment as having contacted separate facilities seeking other particular items of sensitive, dual-use technologies. IC reporting identified several of those entities as suspected technology brokers for East Asia and the Pacific actors and enterprises associated with multiple military development projects.

Analyst Comment: Requests such as this one were typical of overt attempts by East Asia and the Pacific entities to acquire sensitive or classified information and technology resident in the U.S. cleared industrial base. While the contacting entities were likely unaware of each others' requests, viewing the requests together allowed DSS CI to establish a likely connection between the soliciting entities and end users associated with the national military in question. (Confidence Level: Low)

NEAR EAST

OVERVIEW

The Near East accounted for 18 percent of the worldwide total of industry reports to the Defense Security Service (DSS) for fiscal year 2011 (FY11), just as it did in FY10. The aggressive efforts of collectors associated with this region to obtain illegal or unauthorized access to sensitive or classified information and technology resulted in almost 75 percent more reports in FY11 than FY10.

Near Eastern collectors' steadily increasing volume of suspicious contacts over the last several years signifies a continued high value placed on the acquisition of U.S. defense technology and technological know-how. This is despite national goals, in several cases, of achieving greater self-sufficiency in the production of defense equipment. While the region produces some of its own defense equipment, the technology remains foreign-influenced, and rapid advances in defense technology mean the Near East continues to rely on accessing foreign sources.

At present, increased perceived threats from regional neighbors and/or the United States may have temporarily taken precedence over longer-term goals of self-sufficiency. Near East short-term collection efforts may be driven by the perceived need to quickly improve national defense infrastructures, particularly air defense-related technologies.

DSS continues to receive reports of Near Eastern entities' attempts to acquire U.S. technology by subterfuge. Near Eastern collectors have become exceptionally adept at using complex networks of front companies, shell companies, brokers, and procurement agents in their efforts to acquire U.S. technology. These collectors continue their attempts to acquire U.S.-origin technology through third countries, leveraging relaxed export-control laws.

Sometimes the subterfuge is somewhat more direct. Some Near East collectors attempt to exploit established trade assistance agreements (TAAs) with U.S. cleared contractors. Official visits and targeting was also prevalent in reporting as collectors sought to leverage official facility visits to gain unauthorized access to U.S. technology information.

Other frequently attempted MOs manifested themselves in FY11 when Near Eastern commercial entities sought to acquire U.S. technology, requested sensitive information, or solicited marketing relationships. Although not as prevalent as in FY10, targeting by Near East government agents of U.S. travelers on official business overseas, usually as cleared contractor personnel were departing the region, remained a threat.

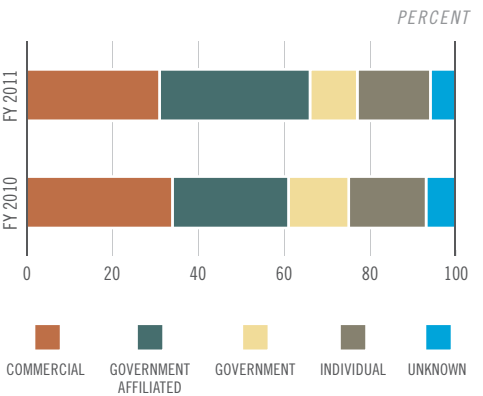
In FY11, Near Eastern collectors targeted a wide array of defense technologies, ranging from antiquated U.S. military hardware to new, state-of–the-art military technologies. Consistent with previous years' reporting, Near East collection targets spanned nearly all the sections of the Militarily Critical Technologies List (MCTL).

Students from the Near East continue to show interest in conducting postgraduate-level research in emerging technologies. Reporting received from industry shows evidence of an increase in academic solicitations from students seeking to conduct postgraduate research in cleared university-affiliated research centers. Near East student enrollment in these programs poses a technology transfer risk when students return home equipped with the knowledge and technological know-how to design and develop new defense technologies necessary to help their countries achieve self-sufficiency.

COLLECTOR AFFILIATIONS

All five collector affiliations increased year over year in number of reports from industry. The top two affiliations swapped places in the ranking from FY10 to FY11, while the remaining three maintained their positions. The interesting changes were in the percentages of the total each category accounted for. The new most common identification, government-affiliated, increased in percentage from 27 to 35 percent, while all four other categories, including the former top category, commercial, declined in share, by one to three percentage points each.

COLLECTOR AFFILIATIONS
FIGURE 11



Within the government-affiliated category for the Near East, the two main models involved affiliations between the government and either academics or commercial firms. Government-affiliated academics, purporting to be students and professors, tended to be associated with major universities; government-affiliated firms tended to

be major commercial companies. The academics typically requested access to cleared contractors' postgraduate research, placements for sabbaticals, assistance with or collaboration on research and scientific publications, and/or employment; government-affiliated firms tended to attempt to exploit established relationships with cleared contractors and leverage official cleared facility visits.

Analyst Comment: Near Eastern countries desiring to maintain or enhance their status as regional powers likely seek to establish technological autonomy and gain recognition as scientific and technological achievers, which requires the ability to independently develop advanced and innovative technologies. Currently, their education systems, scientific establishments, industrial bases, and/or forces of skilled workers probably lack the resources, equipment, and technical expertise to achieve such goals. Therefore they likely continue to rely on collections against western countries' industrial bases to cultivate the necessary knowledge and technical abilities and keep current on technology advances. (Confidence Level: Moderate)

Although industry reports identified with the commercial affiliation declined in their share of the Near East total and the affiliation fell from the top spot, it still accounted for nearly a third of all reports. During FY11, commercial entities maintained a consistent targeting of cleared contractors by seeking dual-use technologies. Sometimes the same individual attempted to acquire the same technology while purportedly representing multiple companies; sometimes multiple companies from the same country attempted to acquire the same technology.

Using commercial firms for collection attempts can constitute an effort to obscure government involvement in attempted collection against U.S. information and technology. Near East companies sometimes contacted cleared contractors in an attempt to either procure an export-controlled technology or solicit an opportunity to market the cleared contractor's technology within the country or region.

Various industry reports recounted incidents in which Near East commercial distributors requested U.S. technology in what would nominally be a legal and permitted acquisition. However, the purchases sometimes were on behalf of end users from other regions, after multiple attempts by entities in those regions to procure the same technology themselves had failed. In such cases, any subsequent transfers of defense technology violated signed agreements requiring U.S. approval. In other cases, regimes' acquisition of U.S. technology itself was illicit, and was then followed by a sharing of U.S. technology with the third parties that sought it indirectly.

Analyst Comment: DSS assesses that aggressive collectors from other regions likely exploit Near East relations with the United States to acquire U.S. defense technology for misrepresented end uses, as well as employ other successful MOs. When Near Eastern states obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base, it is likely to undergo further illicit transfer. (Confidence Level: Moderate)

FY11 industry reporting attributed to individual collectors represented 17 percent of suspicious contacts to cleared industry, remaining proportionally consistent with last year's results. These collectors continue to provide little to no information to indicate ties to commercial or government entities. Individual collectors typically employed the academic solicitation, the request for information (RFI), and seeking employment MOs.

Analyst Comment: Individual collectors likely attempt to increase their chances of successfully obtaining U.S. sensitive or classified information and technology

by obscuring ties to governments and commercial firms known to the United States. (Confidence Level: High)

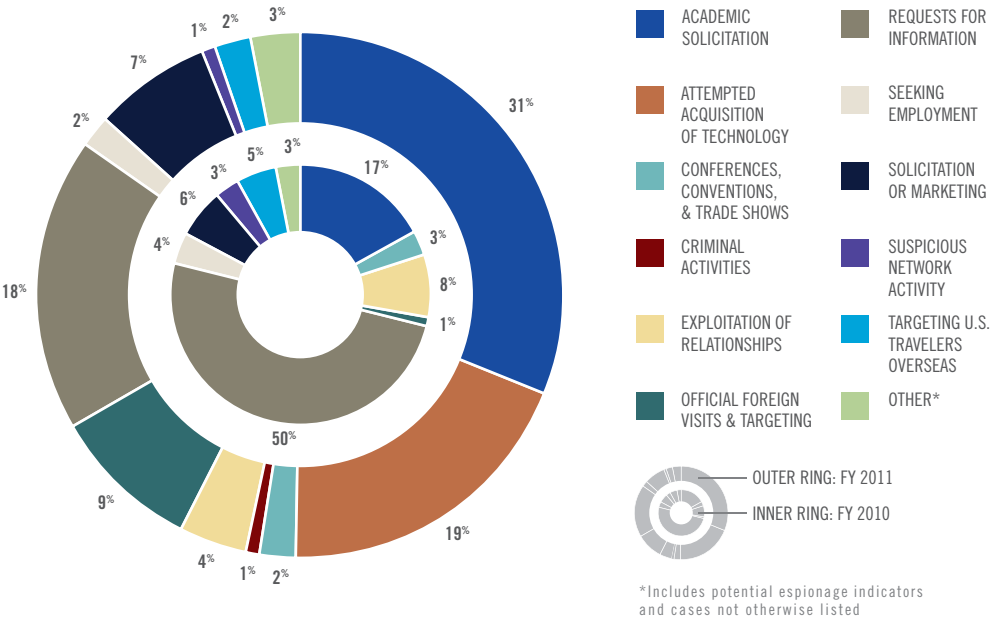
Reported collection attempts associated with Near Eastern government entities declined slightly in proportion of the whole, from 14 percent in FY10 to 11 percent in FY11, even as the number of industry reports in that category increased by over 40 percent. In FY11, governments with access to cleared contractor facilities via established TAAs continued to attempt to leverage them to collect against U.S. information and technology. Known or suspected intelligence officers (IOs) supplemented official delegations in visits to cleared facilities, typically under the guise of official representation. Also, in conformance with a FY10 trend, in some countries airport security continued targeting cleared contractor personnel while on official business in-country. Industry reporting documented multiple incidents of cleared contractor personnel receiving intense scrutiny from airport security elements when attempting to depart the country.

METHODS OF OPERATION

An adjustment in the DSS accounting system for MOs resulted in many FY11 cases that would previously have been categorized as RFIs being labeled instead as attempted acquisitions of technology. This had a major effect on the top of the listing of MOs used by Near East collectors, as represented by industry reporting. The RFI, which had been at the top of the FY10 listing by accounting for half of that year's reports, fell to third at 18 percent of the total in FY11. The corresponding rise in attempted acquisition of technology reports was from no reports in FY10 to 19 percent of the total in FY11.

The most notable change, however, occurred with regard to reports on academic solicitation. The number of such reports

METHODS OF OPERATION
FIGURE 12



*Includes potential espionage indicators and cases not otherwise listed

more than tripled, an increase of 220 percent. Academic solicitation accounted for 31 percent of this year's larger total.

Recent changes in U.S. visa requirements loosened restrictions on foreign students, allowing more to remain in the United States after graduation, and available statistics verify that the number of such students staying has increased in recent years.

Analyst Comment: Some of the statistical shift can be attributed to the recategorization of many reports from RFIs to the attempted acquisition of technology. However, it is likely that a greater part of the explanation for the diversion of Near Eastern collection efforts into solicitations aimed at exploiting U.S. academia lies in loosened visa requirements. (Confidence Level: Moderate)

Of the lower tier of MOs, as measured by the number of industry reports, none individually accounted for more than nine percent of reports. Some increased in number of reports, some declined, and there were resultant adjustments in proportional share. Two of these lesser MOs deserve specific comment, however. There were 11 times as many reports of targeting during official foreign visits to cleared contractors in FY11 than in FY10. In contrast, the suspicious network activity that represents such a noted threat from other regions actually declined in number of reports related to the Near East, measured year over year, and in FY11 accounted for only one percent of reports.

As noted in the previous section, academics, both students and professors, constituted a major bloc of Near Eastern government-affiliated collectors. Students sent emails to cleared U.S. professors requesting to join research programs in technology areas related to energy, materials, electronics, and mechanical and aerospace engineering.

Analyst Comment: The levels at which students from the Near East are contacting U.S. professors engaged in classified research are alarming. Almost exclusively, such programs are classified because the research they conduct is defense-related. It is noteworthy that the U.S. universities targeted are not commensurate with the top universities attended by Near Eastern students in the United States or located in areas with large home-country expatriate communities where foreign students typically seek to live. It is likely that many of the approaches to particular U.S. professors by Near Eastern students are intended to gain illicit access to sensitive or classified information and technology in targeted technology areas. (Confidence Level: Moderate)

Although no available evidence corroborates that Near Eastern government agencies are presently tasking student placement at cleared contractor facilities, some intelligence reporting suggests that the practice has occurred in the recent past. Students may be recruited, trained, and tasked as sources, and receive financial aid and support.

Analyst Comment: Some Near Eastern students seeking placement at cleared contractor facilities receive financial support from their governments. Government-sponsored students would likely attempt to collect technical information on behalf of their government in return for its sponsorship. (Confidence Level: Moderate)

Although attempted acquisition of technology was the second most prevalent MO practiced by Near Eastern collectors as represented by industry reporting, at 19 percent of the total it fell far behind academic solicitation. Intelligence Community (IC) reporting showed that some countries' collectors attempted to purchase sensitive or classified U.S. technology directly, usually via email or telephone, whereas others made their approaches indirectly, using front companies or third-country entities to make contact with U.S. companies. Industry reporting during FY11 corroborates IC reporting, with requests for

export-controlled technology linked to the Near East originating from at least a dozen foreign countries.

Analyst Comment: Because of the nature of clandestine attempts to acquire sensitive or classified U.S. technology, DSS assesses that FY11 industry reporting almost certainly does not provide a complete representation of this aspect of Near Eastern collection activities. Some companies in other regions have a documented history of providing Near Eastern collectors with U.S. technology, and during FY11 DSS analysis found a substantial increase in such links. DSS assesses it as very likely that some portion of the reported attempts to acquire U.S. technology that DSS attributed to collectors from other regions had intended Near Eastern end users. (Confidence Level: Moderate)

Direct attempts to acquire sensitive or classified U.S. technology via purchase, usually requested via email or telephone, were most often made by commercial firms overtly requesting to purchase export-controlled technology. When commercial entities target U.S. technology, it is often for competitive advantage, with the export of defense production in mind. Most often, such Near Eastern collection entities attempted to procure U.S. technology in a seemingly innocuous and legitimate manner. Similarly, commercial firms constitute the affiliation of Near Eastern collectors predominantly employing the relatively similar and seemingly straightforward MOs of the RFI and solicitation and marketing.

As noted earlier, however, such seemingly innocuous, legitimate, and straightforward requests can be the result of deliberate efforts to minimize the signature of government involvement. DSS evaluation of information concerning certain Near Eastern firms reveals the likelihood that the government in question had a hand in certain requests, such as requests to market a cleared contractor's global positioning system or act as an intermediary for brokering aerospace and defense deals with the United States.

As represented by FY11 industry reporting, the RFI was Near Eastern collectors' third most frequently used MO, representing 18 percent of reported cases. These contacts consisted of web-card submissions that requested the cleared contractor to provide more information regarding its products and emails to cleared contractor employees to obtain additional information. For example, in May 2011, a California-based cleared contractor facility received an unsolicited email request for information regarding ship technology. The sender stated that he was studying naval architecture and drafting an article about such technology for a home-country newspaper.

Analyst Comment: The email to the cleared employee was likely an attempt to obtain specific information about such ships under the guise of drafting an article. Any information provided to the sender probably would have been used to determine specifications and aid in reverse-engineering a ship for home-country use. (Confidence Level: Moderate)

Although FY11 industry reporting registered official visits and targeting at only nine percent of the year's total, that represented a noteworthy increase in both number of cases and percentage share from FY10. This MO is typically employed by governments or defense firms that maintain defense relationships with cleared contractors. In FY11, under the auspices of official delegation visits, Near Eastern entities made numerous attempts, in multiple variants, to leverage their admission to cleared contractor facilities to gain illegal or unauthorized access to sensitive or classified U.S. information and technology. For example, some visitors, typically through casual conversation, persistently queried cleared contractor personnel for sensitive information that fell outside the agreed-upon topic or scope of discussion. Additionally,

delegations attempted to make last-minute revisions to the approved list of individuals visiting the facility so as to insert known or suspected IOs into their delegations.

In FY11, some Near Eastern entities employed additional methods to exploit established trade agreements. Typically, employees of privileged firms would contact cleared contractor personnel via email and attempt to leverage an established relationship by inquiring about sensitive information beyond the scope of the TAA. The pattern in previous incidents using this approach has been for foreign personnel to deliberately solicit multiple cleared contractor personnel through casual conversation in pursuit of the same information.

Analyst Comment: DSS assesses that some Near Eastern entities likely prefer using the official foreign visit MO over email contact to target cleared industry because in-person requests appear less premeditated. (Confidence Level: Moderate)

Although constituting only two percent of industry reports, the targeting of cleared contractor personnel traveling overseas on official business did still occur in the Near East in FY11. Industry reported multiple instances of airport security personnel selecting cleared contractor employees for enhanced scrutiny as they attempted to depart for home. Actions included invasive questioning regarding classified and proprietary information and occasional seizure and exploitation of contractor-issued laptops and electronic devices.

Of collection activity ascribed to Near Eastern entities in FY11, suspicious network activity remained at a low level. Reported Near East-originating cyber activity directed against cleared contractors included brute-force password attacks against internet-accessible servers and spear phishing emails that sent back information on recipients but contained no malware in attachments or links. Some Near Eastern actors conducted an intelligence campaign that consisted of relatively innocuous but extensive collection efforts, including on social network sites. Directed against the Department of Defense and its personnel as well as some cleared contractors, they sought to gather email contact lists and similar information. Such tracking and reconnaissance-type activities posed a low threat and did not result in any confirmed intrusions into cleared contractor networks in FY11.

Analyst Comment: While limited in number, the recent Near East-originating spear phishing campaigns likely served to collect information about the recipients so as to check the accuracy of target lists and the effectiveness of the messages in getting recipients to open them. Collection agents almost certainly sought this data in order to more effectively target particular employees when conducting future spear phishing operations against cleared contractors. (Confidence Level: Moderate)

TARGETED TECHNOLOGIES

The top six technology categories targeted by collectors from the Near East, as measured by FY11 reports from cleared industry, were the same as in FY10. The numbers of reports relating to all six sectors of the MCTL increased in FY11, by percentages ranging from 45 to 210 percent. However, while four of these technologies (lasers, optics, and sensors [LO&S]; space systems; armaments and energetic materials; and electronics) also increased their share of the total, two sectors (including the top category, information systems [IS], as well as aeronautics systems) declined in share. The result was that these top six targeted technologies became more tightly bunched, ranging from eight to 14 percent apiece in FY11 in contrast to five to 16 percent in FY10.

Thus, reporting showed that Near East entities' technology interests became more evenly spread across the field, with collectors seeking more U.S. information

TARGETED TECHNOLOGY
FIGURE 13

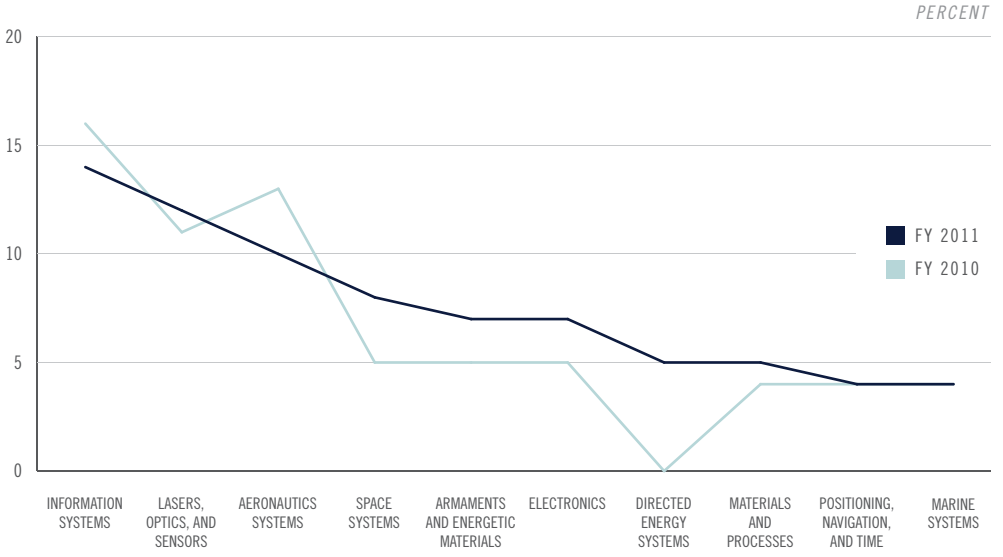


Figure illustrates the top ten most targeted technology categories in FY11 compared with the reporting statistics for the same categories from FY10.

and technology almost impartially. Within this wide range of technology sectors, some particular technologies came to the fore, including unmanned aerial vehicles, underwater autonomous vehicles, night vision devices, modeling and simulation (M&S) software, radiation-hardened (rad-hard) electronics, commercial aircraft, missile technology, and radar components. On the other hand, while a particular technology, an inertial navigation system, was absent from this year's industry reporting on the Near East, it should be noted that some requests for the technology resolved to companies from other regions which have a history of conducting business with Near Eastern entities and which failed to identify an end user.

Analyst Comment: It is likely that some of the third-country requests for the system were intended to supply Near Eastern end users. (Confidence Level: Moderate)

The IS technology sector received the most attention from Near Eastern collectors, as reflected in FY11 industry reporting. The number of reported collection attempts from this region rose 75 percent from FY10, representing 14 percent of the total in FY11.

Potential Near Eastern collectors practicing the longer-term MO of academic solicitation showed a high level of interest in academic programs addressing radar, communications, antenna, and radio technologies. Other Near Eastern collectors attempted to acquire IS technologies more directly. They specifically targeted U.S.-developed command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) systems technology, and requested to purchase dual-use telecommunications equipment.

Analyst Comment: The courses of academic study in question are likely targeted for their defense applications that support advances in wireless networking and communication. Those who sought C4ISR technology probably did so to enhance battle space awareness, airborne electronic warfare systems, and naval electronic support measures systems. (Confidence Level: Moderate)

LO&S technology, at 12 percent of the total, was the second highest Near Eastern collection priority in FY11, based on industry reports that doubled year over year. The attention was attributable largely to interest in U.S.-developed radar technologies associated with naval and ground vehicle applications.

Examples of attempts against LO&S include a Near Eastern research company contacting multiple cleared contractors in FY11, requesting detailed technical information regarding U.S. naval radar platforms. In other cases, Near Eastern delegations visiting cleared contractors conducted entire facility visits in a mode of aggressive attempts to gain unauthorized access to particular technology assembly processes.

Analyst Comment: Near Eastern delegation personnel were likely attempting to acquire the schematics and learn about the assembly processes to provide insight into the functioning of the specified technology. Collecting entities likely sought to obtain such technical know-how to strengthen their country's indigenous development and production capability and decrease the vulnerabilities inherent in relying on foreign sources for military equipment. (Confidence Level: Moderate)

In FY11, aeronautics systems technology remained a noteworthy Near Eastern collection target, accounting for 10 percent of industry-reported incidents, even though the category experienced a relatively "low" 45 percent increase in number of reports. Unmanned aerial systems were a primary target, including associated technologies and short-range unmanned aerial vehicles (UAVs) normally used for surveillance purposes.

Analyst Comment: Border security and terrorist threat concerns have likely heightened Near Eastern nations' interest in enhanced surveillance capability, leading to attempts to acquire U.S.-developed mini-UAVs to strengthen their security presence along their borders. (Confidence Level: Moderate)

Based on industry reporting, Near Eastern entities targeted space systems technology in FY11 at eight percent of the total, but with a 210 percent increase in number of cases. Powers within the region that are investing heavily in space programs have plans to launch several indigenous satellites for military and civilian use over the next several years.

In a space systems-related trend, Near Eastern students also demonstrated significant interest in conducting research related to rad-hard electronics, which are necessary to withstand the high levels of radiation encountered during space flight (see the Special Focus Area section). In one case, a national of a Near Eastern country attempted to acquire a free trial of a cleared contractor's version of M&S software (which satellite programs require) by creating a fictitious web-based email account using the name of a U.S. cleared employee. The attempt failed only because an employee of the cleared contractor recognized the name and asked the cleared employee whether he had sent the email.

Analyst Comment: For any country within the Near East, having a successful space program would be a substantial source not only of military benefit but also of national pride. Interested collectors likely target U.S. space technology through a variety of means. Collector attempts to acquire satellite and M&S software are likely linked to aspiring national satellite programs. (Confidence Level: Moderate)

Armaments and energetic materials technology increased from being the subject of five percent of total industry reports related to the Near East in FY10 to seven percent in FY11, representing an increase of over 152 percent in number of reports. Near Eastern government and government-affiliated entities attempted to leverage official facility visits to gain access to U.S. classified or export-restricted technology information and data, including U.S. missile defense technology and technical data and source codes of missile defense systems.

Analyst Comment: Near Eastern governments are likely concerned with countering missile attacks. They probably sought to enhance their missile defense platforms' capability to withstand rocket and missile threats by correcting deficiencies in missile defense capabilities, leading to their active attempts to address these deficiencies through system upgrades. (Confidence Level: Moderate)

Electronics technology also received substantial interest from Near Eastern collectors, representing seven percent of the year's reports, a rise in number of over 150 percent. Near Eastern collectors focused their efforts on various microwave, radar, antenna, and other specialized electronics systems and components.

OUTLOOK

The Near East contains several countries that harbor hostility toward each other, and perceive threats against their safety and security to be immediate. Some are on good terms with the United States, others are not. These countries strive not only to counter any regional attack by one another and in some cases from the United States as well, but also to achieve regional dominance in the Middle East. All Near Eastern collectors will likely remain reliant on acquiring U.S. information and technology to enhance their defensive and offensive capabilities and support their own military industrial bases, so will almost certainly continue to target U.S. technology. **(Confidence Level: Moderate)**

The Near East includes countries that are or strive to be technologically competitive with U.S. defense industries, and even to establish and maintain a global economic advantage in the field of defense exports. Stable economic success can come to rely heavily on indigenous manufacturing entities successfully collecting against equivalent, rival U.S. technologies. Technologies targeted by Near Eastern interests in FY12 will likely include a wide variety of U.S. systems and equipment in pursuit of modernization and enhancement of their own forces, as well as their likely goal, moving forward, of dominating specific defense markets for economic gain. **(Confidence Level: Moderate)**

Near Eastern collector affiliations will likely settle into the new pattern established in FY11. Government-affiliated entities will likely remain the top category, largely due to the number of Near Eastern students and professors requesting some sort of association with cleared contractors, which requires them to provide some identifying information. In contrast, individual and unknown collectors will likely remain noteworthy as some near Eastern entities strive to provide minimal or no information linking them to their home countries. **(Confidence Level: High)**

Commercial firms will very likely contribute a noteworthy share of overall reported Near Eastern collection attempts again in FY12. Governments will likely continue to attempt to exploit official facility visits so as to gain unauthorized access to U.S. information and technology. Perceptions of success in employing this tactic will likely result in the continuation of its use from FY11 into FY12. Other Near Eastern commercial entities

will very likely continue to use companies located outside of the region to request U.S. technology. **(Confidence Level: High)**

Near Eastern MOs will also likely remain stable in the near future. Numbers of reported academic solicitations will almost certainly remain at high levels as students continue to seek entry into cleared research programs and request technology under the guise of academic cooperation. Recent changes in U.S. visa requirements will very likely continue to make U.S. research programs a prime target for Near Eastern collection activity. **(Confidence Level: High)**

In their attempted acquisition of U.S. technology and information from cleared contractors, some Near Eastern collectors will probably take very direct approaches, combining this MO with RFIs; official foreign visits; solicitation or marketing; exploitation of relationships; conferences, conventions, and trade shows; and targeting U.S. travelers overseas. Other collectors will likely continue to use a variety of circuitous methods to procure technologies, relying heavily on front companies, procurement agents, and brokers located abroad. As more of these procurement networks are exposed, Near Eastern acquisition methods will likely evolve even further in the direction of advanced techniques to attempt to delude U.S. companies, such as the use of western-style aliases and company names from non-threatening countries.
(Confidence Level: High)

Current events and the need to defend their countries against the aforementioned perceived threat of military strikes within the region or by the United States will almost certainly continue to focus Near Eastern technology collection efforts in FY12. These would likely be aimed, first, toward addressing any previously identified limitations in indigenously produced missile defense systems, then on further enhancing missile defense platforms' capability against rocket and missile threats. Collection attempts against cleared contractors will likely target missile technologies and radar components. **(Confidence Level: High)**

Given various Near Eastern governments' desires to strengthen their border security, a revived focus on aerial and underwater autonomous vehicles for surveillance purposes will likely reemerge in FY12, leading to continued targeting of U.S.-manufactured unmanned systems.
(Confidence Level: Moderate)

Any country within the Near East desiring to launch indigenously produced satellites will likely continue to target U.S.-derived rad-hard electronics.
(Confidence Level: Moderate)

CASE STUDY: PERMUTATIONS

The following case demonstrates the convoluted mechanisms by which some Near Eastern entities seek to acquire U.S. export-controlled technology. During FY11, a suspected procurement agent for a Near Eastern regime was seeking various radar, microwave, and electronic components. He contacted several cleared contractor facilities and U.S. businesses, using various company names and email addresses in his requests.

In June 2011, the agent, purportedly representing a Near Eastern company, contacted a New York-based cleared contractor facility seeking the price and availability of two items of an export-controlled technology. On the same day, another individual, representing a commercial entity in another region, contacted the cleared contractor facility regarding the acquisition of two items of the export-controlled technology as well as other electronic components. The items requested by both procurement agents had the same specifications. According to the cleared contractor, specifications for the items were uncommon, as none with those specifications had been sold before.

Analyst Comment: Considering the unusual specifications of the requested items, combined with the similarity of the two requests, DSS assesses that the two suspicious contacts were likely related. The out-of-region firm was probably seeking to procure export-controlled items on behalf of Near Eastern entities.
(Confidence Level: Moderate)

In December 2010, the same Near Eastern procurement agent contacted the same cleared contractor facility, this time claiming to represent a company located in a different Near Eastern country. He requested a quote for six amplifiers of an advanced type. IC reporting revealed that he had made multiple previous solicitations as well. In December 2010, the agent—purportedly representing both the same company and yet another company in yet another Near Eastern country—contacted U.S. businesses seeking a variety of export-controlled advanced amplifiers.

An available business directory classifies the procurement agent's company as trading in textiles, clothing, and footwear. However, DSS records reveal it is linked to multiple requests for U.S. electronic components with warfare applications.

Analyst Comment: Reporting from cleared industry continues to illustrate Near Eastern collectors' use of complicated networks consisting of third-party intermediaries, front companies, brokers, and procurement agents to attempt to illicitly acquire U.S. technology. DSS assesses that the individual in question is almost certainly a procurement agent for his government, specializing in radar and microwave components that could be used for electronic warfare operations. He probably uses various company names, email addresses, and locations to facilitate attempts to illegally acquire U.S. export-controlled technology. It is likely that Near Eastern entities also use brokers or intermediaries based in other regions to further their acquisition of U.S. technology.
(Confidence Level: High)

EUROPE AND EURASIA

OVERVIEW

Europe and Eurasia was the third most active region in fiscal year 2011 (FY11) in terms of reports from industry concerning collectors attempting to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. Yet as significant as that is, it might seem less consequential when compared to the approximately 75 percent increases by the two most active regions, East Asia and the Pacific and the Near East, and a 129 percent increase by the fourth-ranking South and Central Asia. In this context, industry reports on collection attempts originating in Europe and Eurasia increased by "only" 60 percent in FY11.

Yet some factors serve to heighten concerns about Europe and Eurasia. The region contains some of the most advanced technological and economic competitors to the United States, as well as some of the most skillful and clever human and cyber intelligence collectors. It is likely that even industry reporting and other counterintelligence contributions combined underestimate the totality of the ongoing Europe and Eurasia efforts to gain illicit access to U.S. industrial secrets.

In industry reporting, commercial entities and government-affiliated institutions (often involved in research and education) were the top two Europe and Eurasia collector affiliations, at 45 and 19 percent of the total, with individuals and government following. From FY10, the number of reported contacts by entities with unknown affiliation decreased and the proportion of the total accounted for by that category went from second position at 28 percent to fifth position at ten percent. This could reflect collectors' greater willingness to disclose association with government-affiliated research organizations due to deepening economic ties between the United States and Europe and Eurasia.

Attempted acquisition of technology was the method of operation (MO) Europe and Eurasia entities used most, as reflected in industry reporting, accounting for over a third of the FY11 total, followed by the request for information (RFI) at 29 percent. The relative prominence of these categories is consistent with the previous year's data. New Defense Security Service (DSS) categorization guidelines require that a contact formerly considered an RFI now be identified as an attempted acquisition of technology if it solely sought to purchase the technology.

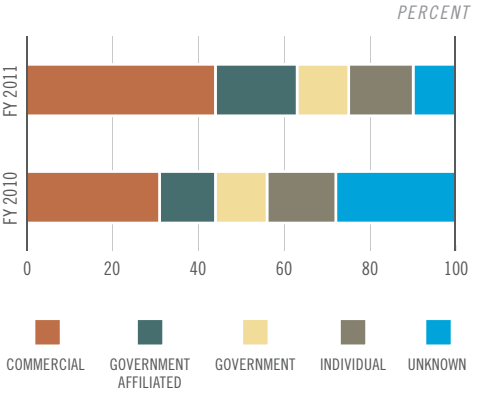
Based on industry reporting, Europe and Eurasia collectors targeted aeronautics systems and lasers, optics, and sensors (LO&S) almost equally, followed closely by information systems (IS) technology and electronics technology. They were all clustered within a narrow range, each accounting for 10 to 16 percent of the FY11 total.

The implied continuity in Europe and Eurasia collection emphases is attributable to the ongoing efforts to upgrade military technology. Europe and Eurasia countries seek to accomplish a variety of goals, whether reducing dependence on natural resource exports, decreasing dependence on foreign supply sources and thus foreign influence, boosting domestic production of military goods for both domestic use and export, and/or creating indigenous high-technology sectors.

COLLECTOR AFFILIATIONS

Collector affiliations reflected in industry reporting linked to Europe and Eurasia became ever more concentrated in FY11 on commercial collectors. Whereas in FY10 reports were fairly well distributed between the five categories, from 31 percent for commercial down into the teens, in FY11 the commercial category accounted for 45 percent of the total, with no other category exceeding 19 percent.

COLLECTOR AFFILIATIONS
FIGURE 14



Beyond this basic observation, there was some interesting movement within the statistics. In numbers of reports during the most recent year, the unknown category decreased by 43 percent, while the other four all increased, two of them by around 50 percent and two by over 100 percent. In terms of change from FY10 in percentage of the total, in addition to a decrease in the share accounted for by unknown collectors from 28 to ten percent, the individual collectors' share also decreased, while the government collectors' share was unchanged at 12 percent. This left the government-affiliated collector category to increase from 13 to 19 percent

and the commercial category from 31 to the aforementioned 45 percent.

Consistent with the previous year's reporting, commercial entities remained the primary affiliation of collectors associated with Europe and Eurasia in FY11, with the number of reported cases more than doubling year over year. Many Europe and Eurasia commercial entities identify neither end users nor specific technologies in their requests.

Analyst Comment: Some ostensibly commercial and individual Europe and Eurasia collectors demonstrated a level of knowledge about technologies that was consistent with that of intelligence officers (IOs). DSS assesses that the continued increase in reported activity by Europe and Eurasia commercial collectors likely reflects an effort to upgrade military technology. Certain aspects of the effort to modernize civilian economies likely dovetail with military requirements for improved technology. (Confidence Level: High)

Interest by Europe and Eurasia commercial entities in developing business ties to the United States is increasing, and contacts by collectors affiliated with them are as well. Simultaneous with the 14 percentage point increase in the share of contacts made by commercial collectors, government-affiliated collectors became the second most common category, with the number of such cases in FY11 more than doubling over FY10. The share of contacts from unknown collectors decreased by almost half, and those from individuals slightly.

Analyst Comment: Economic ties between the United States and most countries in Europe and Eurasia are close, and in some cases are growing closer. The significant increase in reports linked to government-affiliated entities likely reflects a greater willingness by collectors to disclose association with government-affiliated research centers in light of these closer economic ties. Simultaneously, multiple countries within Europe and Eurasia almost certainly intend to remain competitive in the world arms market with the United States; Intelligence Community (IC) reporting indicates that such countries view the United States as a market competitor for the sale of military equipment. (Confidence Level: High)

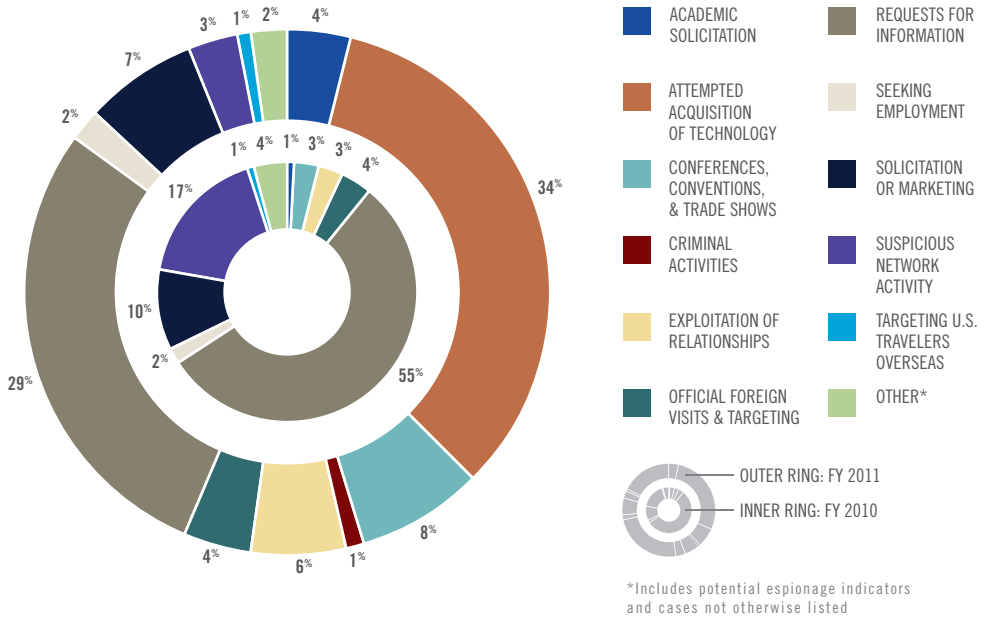
METHODS OF OPERATION

Regarding the MOs that collectors linked to Europe and Eurasia were reported by industry as using, attempted acquisition of technology at 34 percent and RFI at 29 percent combined to account for even more of the total in FY11 (63 percent) than they did in FY10 (55 percent). In the interim, DSS changed its accounting methodology such that many collection attempts that would previously have been labeled as RFIs are now categorized as attempted acquisition of technology.

The increase in these two categories might seem to make all the other MO categories relatively unimportant, with each of them accounting for only one to eight percent of the total. But if the two most common MOs represent the simplest, most straightforward method of attempting to obtain illegal or unauthorized access to sensitive or classified information and technology, it should still be noted that the wide range of other MOs recorded in industry reports represent all the "Plan B" methods. In other words, one may not succeed in gaining the desired information or technology by buying it or asking about it outright. In that case, the next-most-likely-to-succeed method is to somehow get someone close to a cleared contractor, then seek opportunities to gain illegal or unauthorized access to the desired materials. Whether at a conference, convention, or trade show, via a delegation visiting a cleared contractor in the United States, by targeting a U.S. traveler overseas, or by obtaining a job or academic placement or setting up a marketing arrangement, collectors seek to insinuate themselves into a position or relationship they can exploit

METHODS OF OPERATION

FIGURE 15



*Includes potential espionage indicators and cases not otherwise listed

to their benefit. Success for them almost certainly results in harm to the interests of cleared contractors as well as the larger military, technological, and economic well-being of the United States. All of these "lesser" MOs together accounted for a not inconsiderable one-third of the year's reported collection attempts originating in Europe and Eurasia.

The exception to this discussion is the suspicious network activity (SNA) MO. By definition, it involves attempts to work through computers and networks, not human beings directly, and at a distance, not in person. In FY11, industry reports of collection activities categorized as SNA decreased markedly in number from FY10, amounting to a drop of over 70 percent. As a category, SNA went from being the second most common in FY10 at 17 percent to only three percent of the total in FY11. The worrisome possibility is that this change did not occur because industry, DSS, and others in the IC got better at detecting and defeating SNA from Europe and Eurasia, but that the region's cyber collectors, already highly skillful, got even better at concealing their illicit activities.

The MO that Europe and Eurasia collectors were most commonly reported as using in FY11 was the attempted acquisition of technology, accounting for 34 percent of total contacts. Attempted acquisition of technology is defined as expressing interest in purchasing, or actually placing an order for, export-controlled technology.

RFIs comprised 29 percent of FY11 Europe and Eurasia contacts reported by industry. RFIs often target technical specifications of sensitive military systems, but stop short of attempting to purchase an export-controlled item. Receipt of such an RFI could mean that an intelligence service has already obtained a restricted piece of technology and is seeking information on its use.

A possible example occurred in January 2011 when a Europe and Eurasia national contacted a cleared contractor, claimed to possess one of its export-controlled transceivers, and requested the data transfer protocol for use with the module. The transceiver is a component in several military applications, including bombots and other unmanned vehicles. The individual did not reveal how he obtained the transceiver, but it may have been lost or stolen in a combat zone.

Conferences, conventions, and trade shows rose to be the third most used MO in the reporting data for the year. Such venues continued to be used to solicit information and technology in FY11. IC reporting noted that government representatives and civilian journalists from Europe and Eurasia who questioned unmanned aerial system (UAS) industry experts at military expositions and conferences frequently asked questions beyond the scope of their responsibilities and showed an unusual breadth of knowledge.

Together the solicitation or marketing services and the exploitation of relationships MOs accounted for 13 percent of reported Europe and Eurasia collection attempts in FY11. Reporting from cleared contractors suggests that collectors attempt to exploit government cooperation agreements and legitimate business exchanges to collect intelligence. Throughout 2010 and 2011, DSS received several reports that Europe and Eurasia commercial delegations visiting cleared contractors included government IOs.

In their simplest manifestation and deployment, collectors find electronic methods of contact such as unsolicited emails and phone calls to be attractive, as they can be conducted inexpensively, with a low risk of adverse consequences combined with the potential for high gain. Yet the more advanced types of Europe and Eurasia cyber espionage against U.S. cleared contractors essentially represent a current intelligence gap for DSS. In FY11, several cyber attacks against cleared contractor networks, such as those using the Zeus Trojan banking malware, were linked to criminal hacking. Thus, even when contacts are categorized as SNA, incidents cannot necessarily be attributed to foreign intelligence entities.

Analyst Comment: Such cyber espionage may cause malicious activity targeting cleared contractors that is conducted by Europe and Eurasia collectors to be incorrectly attributed to actors in a different country or region. (Confidence Level: Low)

Some Europe and Eurasia countries may attempt insider-enabled network attacks, which prevent the observation of suspicious indicators normally associated with network attacks. Additionally, such attacks may enable the compromise of computer networks that are sufficiently hardened to withstand attacks originating over the Internet, but remain vulnerable to subversion by a malicious employee or contractor, constituting a significant insider threat. (Confidence Level: Low)

TARGETED TECHNOLOGIES

The top four targeted technologies in FY11 industry reports were the same as in FY10: aeronautics systems; LO&S; IS; and electronics. However, they became much more bunched at the top, with the share accounted for by the former top category, IS, cut in half from 26 to 13 percent, leaving aeronautics systems unchanged at 16 percent and tied with LO&S at the top of the list; electronics accounted for ten percent. No other individual technology section accounted for more than five percent of the total.

Aeronautics rose to the top of the list of Europe and Eurasia-targeted technologies at 16 percent of all reports. Some Europe and Eurasia countries that do not have the resources to produce all the weapon systems and technologies they consider vital to their national interest seek out U.S.-developed UASs to support their armed forces

deployed in various spots around the globe. In FY11, there was some focus on long-endurance unmanned aerial vehicles.

LO&S accounted for 16 percent of the reported total. On one occasion, Europe and Eurasia government collectors questioned a cleared contractor employee working at an exhibit booth at the Euronavale Trade Show in Paris about operating frequencies used in tactical missile defense systems.

Last year's top technology category, IS, was FY11's third most targeted sector, accounting for 13 percent of industry reporting. Most of the contacts involved invitations to conferences or foreign visits to cleared contractors specializing in IS; thus, targeting of specific items was difficult to verify. Optical communications technology with civilian and military applications was a specific focus identified in several reports, with one collecting entity withdrawing its request after the cleared contractor insisted on end user information.

Analyst Comment: The consistent collection emphasis on the IS sector probably reflects the priority of Europe and Eurasia militaries to upgrade their communication technologies. (Confidence Level: Moderate)

Radiation-hardened (rad-hard) circuits (see the Special Focus Area section of this publication) for space-based applications have been a consistent target of some Europe and Eurasia collectors for several years. Within the last decade, a company from the region proposed to a cleared contractor a joint venture to create a facility in its country to produce rad-hard circuits, but this did not transpire. Subsequently, Europe and Eurasia entities sought rad-hard circuits from cleared contractors at least 11 times from FY08 to FY11, as reported by industry to DSS. Four of those requests, made to three separate cleared contractors, occurred in FY11. Most of these requests for rad-hard circuits requested between 20 and 42 pieces, although one sought 3,200.

TARGETED TECHNOLOGY

FIGURE 16

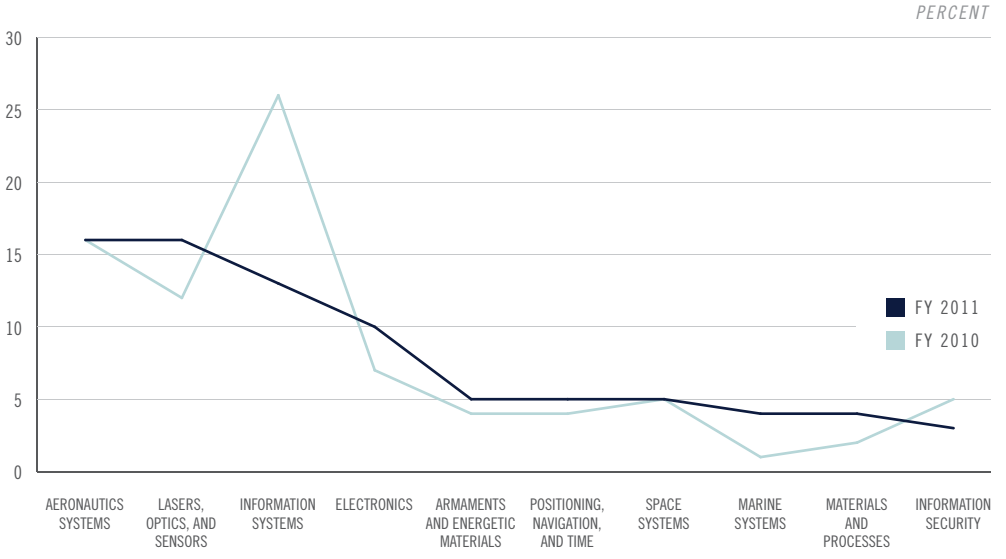


Figure illustrates the top ten most targeted technology categories in FY11 compared with the reporting statistics for the same categories from FY10.

Historical reporting shows that elements within Europe and Eurasia have pursued alternative means to develop or acquire desired technology, including operating facilities to reverse-engineer any Western technology acquired.

Analyst Comment: Those seeking the rad-hard circuits were likely unable to establish an indigenous capability to produce technology that met a desired standard. DSS cannot rule out the possibility that Europe and Eurasia entities are still seeking rad-hard components for reverse-engineering. (Confidence Level: Moderate)

OUTLOOK

DSS assesses that Europe and Eurasia collectors will likely continue to emphasize legitimate commercial exchanges to upgrade their military technology, and those requirements will in turn likely draw upon commercial ties to foreign businesses. Cleared contractors conducting business in Europe and Eurasia will likely be subject to unabated, aggressive collection efforts via all means available. **(Confidence Level: Moderate)**

Several Europe and Eurasia countries view the United States as their foremost economic competitor, and will likely continue to seek information to help them compete politically, economically, and militarily in world affairs. One way in which Europe and Eurasia entities are likely to continue to be a significant threat to U.S. information and technology resident in cleared industry in the coming years is by some companies from the region attempting to purchase U.S. companies. Their likely intent in doing so would be to appropriate U.S. technologies that can then be legally used in Europe and Eurasia exports. **(Confidence Level: Moderate)**

Europe and Eurasia entities' targeting of U.S. information and technology will likely continue to focus on aeronautics systems and IS, with emphasis on UASs and the Joint Tactical Radio System. DSS assesses that additional attention to LO&S will probably continue. Collectors will likely continue to emphasize microelectronics, including the rad-hard variety, due to their importance in bringing militaries in the region into the 21st century. **(Confidence Level: Moderate)**

The U.S. IC will likely face continuing challenges in attempting to attribute cyber attacks against cleared contractors to identifiable Europe and Eurasia entities. Most such SNA will likely appear to support criminal activity, but may occasionally address information falling within the scope of technology requirements set by governments in the region. **(Confidence Level: Moderate)**

CASE STUDY: "WON'T YOU COME INTO MY PARLOR…?"

Between November 2010 and February 2011, a U.S. cleared contractor employee received three email invitations to an international science conference, to be held in Europe and Eurasia. The invitations were sent to the employee's work email address.

IC reporting shows that in 2010, employees from two separate cleared contractors received invitations to the previous conference, held the year before, also in Europe and Eurasia.

Such conferences hosted in Europe and Eurasia may have indirect connections with Europe and Eurasia intelligence services, although the full extent of the relationship is unknown.

Analyst Comment: Scientific conferences present opportunities for foreign intelligence services to spot and assess persons with access to technology intelligence. The successive iterations of this Europe and Eurasia conference may be used to elicit technology information that is responsive to government collection requirements. (Confidence Level: Low)

SOUTH AND CENTRAL ASIA

OVERVIEW

South and Central Asia made the most noteworthy change from fiscal year 2010 (FY10) to FY11—in an unfortunate direction, as far as U.S. cleared contractors are concerned. This region more than doubled year over year in number of reports ascribed to it for attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. In so doing, it increased its share of the world's larger total for FY11 from nine percent to 12 percent.

Increased regional instability and conflicts, counterterrorism efforts, and defense modernization initiatives continue to impact South and Central Asia defense industries, driving efforts to obtain U.S. information and technology. These attempts to maintain and upgrade military capabilities can be accomplished through the purchase of new technologies as well as the upgrading or replacing of older systems. Any sensitive or classified U.S. information and technology acquired could assist greatly with such modernization efforts.

South and Central Asia government entities that experience difficulty in obtaining the licenses and paperwork necessary to purchase dual-use technology are able and willing to exploit their relationships with the U.S. government and commercial entities to circumvent export-restriction laws. South and Central Asia entities still on U.S. export-restriction lists remain a threat to attempt to illicitly acquire U.S. technology.

Commercial companies remained the top South and Central Asia collector category in reported attempts in FY11. The private sector often contacted U.S. cleared contractors in an attempt to win contracts with government agencies in their countries. Intelligence Community (IC) reporting indicates that South and Central Asia intelligence and security services likely work closely with these government agencies on certain matters; however, no evidence suggests that commercial companies have contacted cleared contractors on behalf of or at the urging of intelligence services.

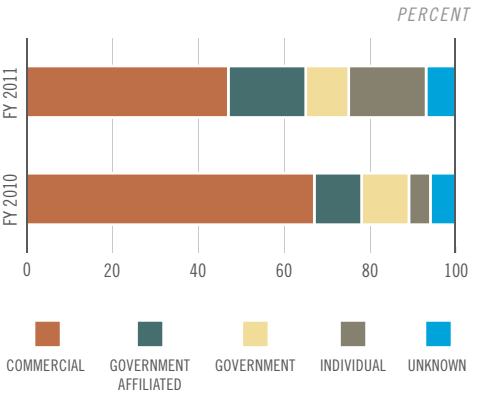
As reflected in FY11 industry reporting on South and Central Asia, the combination of commercial entities using the attempted acquisition of technology and request for information (RFI) methods of operation (MO) accounted for the majority of suspicious contacts. These commercial entities were largely procurement agents who identified military and other government agency end users for the materials sought. In FY11, South and Central Asia commercial companies commonly used direct contact methods, primarily email, to attempt to acquire technology from cleared contractors.

FY11 industry reporting showed that South and Central Asia entities targeted technologies across the Militarily Critical Technologies List (MCTL), most notably in the sections encompassing information systems (IS); lasers, optics, and sensors (LO&S); aeronautics systems; and electronics systems.

COLLECTOR AFFILIATIONS

Given the large overall increase in number of industry reports ascribed to collectors from South and Central Asia, it is not surprising that the number of reports went up in all five affiliation categories from FY10 to FY11. However, the top category in both years, commercial, decreased in share of the total from nearly two-thirds to under one-half. The shares for the government and unknown categories varied by only a percentage point year over year. The significant changes were in the government-affiliated and individual categories, which rose from 11 to 18 percent and from five to 18 percent, respectively, of the more recent year's total, now tying them for a distant second place behind the commercial category's 47 percent.

COLLECTOR AFFILIATIONS
FIGURE 17



Although the percentage share of the year's totals accounted for by commercial entities decreased, the number of reports nonetheless increased by over 60 percent. The majority of the commercial entities making requests for U.S. technology in FY11 were procurement agents acting on behalf of, or in response to requirements from, elements of South and Central Asia governments, including military, security, and intelligence services.

The dominant pattern practiced by governments in South and Central Asia for procuring defense technology consists of state-run organizations issuing tenders to secure military equipment, both systems and subcomponents. Such tenders are often accessible to the public on official government websites and frequently include specifications for the requested technologies. Procurement agents respond to the tenders, attempting to fill the requirements by contacting companies discovered through open-source research that market products matching the tender specifications.

Analyst Comment: The Defense Security Service (DSS) assesses it is very likely that the majority of the suspicious contacts reported by cleared contractors represented efforts to respond to South and Central Asia government tenders and meet government requirements. (Confidence Level: High)

Typically, once a South and Central Asia commercial entity identifies a U.S. company producing technology responsive to the tender requirements, it seeks to contact the company. The commercial agent either attempts to acquire the system outright or requests information on the technology.

Analyst Comment: Queries regarding a technology in question likely constituted attempts to determine whether it could ultimately meet the needs of the South and Central Asia end user; however, DSS cannot rule out that such RFIs represented attempts to obtain sensitive or classified information from the cleared contractor. (Confidence Level: Moderate)

Most South and Central Asia requests for information or technology received by U.S. cleared contractors identified a military service or other government entity as the end user. Several of the commercial collectors that did not identify an end user have ties to the military or are procurement agents with a history of making requests on behalf of the government. Open-source searches provide evidence that in many such cases South and Central Asia companies were referencing tenders put out by specific government agencies, as their requests to U.S. cleared contractors cited specific technologies sought by those tenders.

Analyst Comment: Considering the similarities between the commercial requests and the government tenders, it is likely that South and Central Asia government agencies were the intended end users for the technologies requested in a majority of the cases in the commercial category. (Confidence Level: Moderate)

Some governments within the region are promoting policies to encourage involvement by a wider array of private, commercial companies in defense procurement, including the bidding on government tenders issued by defense agencies.

Analyst Comment: This policy probably contributed to the rise in the number of reported acquisition attempts by South and Central Asia commercial entities. Furthermore, it was also likely responsible for a rise in the overall number of firms and procurement agents that contacted U.S. cleared contractors, as more firms and procurement agents became active in the market. (Confidence Level: Moderate)

Government-affiliated entities followed commercial entities in reported suspicious requests to cleared contractors, constituting 18 percent of the FY11 South and Central Asia total. The number of reports from government-affiliated entities rose by 280 percent over FY10 figures. South and Central Asia collecting entities in this category in FY11 included government-owned companies and government-affiliated technological institutes, other universities, and research organizations.

From such entities, students, researchers, engineers, and others initiated numerous unsolicited contacts to cleared contractors.

They requested jobs, internships, research positions, and other assistance with research; such inquiries often sought information on the pricing or availability of sensitive or classified U.S. technology as well. According to IC reporting, some of the government-affiliated entities in question encourage South and Central Asia students studying in the United States to transfer information and/or technology back to their homelands.

Analyst Comment: Many South and Central Asia students who initiate contacts to cleared contractors likely have a working relationship with defense agencies in their countries, which sometimes fund research and development (R&D) programs at the government-affiliated institutions, then use students and resources from them. (Confidence Level: High)

While the individual category of collector, like government-affiliated, at 18 percent accounted for a considerably smaller share of the total than commercial, the number of reports on individuals soared by over 600 percent since FY10. Entries in the individual category include student requests that DSS counterintelligence analysis connected to independent South and Central Asia universities rather than government-affiliated ones, or cases in which no affiliation with a specific university could be determined. The largest part of these individual requests consisted of résumé submissions to cleared contractors soliciting employment or to U.S. university-affiliated research centers seeking research-related positions.

The small FY11 amount of cyber activity that could be traced to South and Central Asia but no farther is represented in the individual category as well. The remaining individual contacts consisted of RFIs or attempted acquisitions, including requests during which individuals provided no affiliation with a specific company or organization, but their email addresses, mailing addresses, and/or telephone numbers traced back to South and Central Asia. While these requests sought disparate technologies, they tended to mirror requests made by commercial entities.

Analyst Comment: For South and Central Asia collectors in the individual category, DSS could not connect the person to any company. However, there is an even chance that these individuals were independent or new procurement agents responding to government tenders. (Confidence Level: Moderate)

When South and Central Asia government entities themselves contacted cleared contractors, the requests were largely in pursuit of technology systems that are of interest to researchers for space and satellite systems, or consisted of military officers making inquiries about military platforms.

METHODS OF OPERATION

There was a real contrast between FY10 and FY11 in the reported MOs collectors linked to South and Central Asia used in their attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base. Partly this was due to a change in the statistical accounting method used by DSS, which resulted in many contacts that had previously been categorized as RFIs now being labeled attempted acquisitions of technology; the latter category went from no reports in FY10 to ranking first in FY11, with nearly one-third of the total. Mostly this was at the expense of the RFI category, which went from 78 percent of the total in FY10 to 29 percent in FY11.

But the South and Central Asia statistics concerning other MOs experienced change as well. Academic solicitations, which had registered a negligible one percent of the total in FY10, rose to nine percent of the FY11 total, and seeking employment went from three percent to ten percent of the year's total. In comparison, solicitation or

marketing services remained stable within the listing, with only a percentage point increase in share from eight to nine percent.

Attempted acquisition of technology was the most common MO South and Central Asia entities used in FY11, comprising 32 percent of reported collection attempts associated with the region. Generally, South and Central Asia entities sent unsolicited emails to cleared contractors requesting to purchase specific technology, usually in a specific quantity as well. While not all of the unsolicited emails referenced a particular government tender, some cited the exact specifications and quantities listed in such tenders.

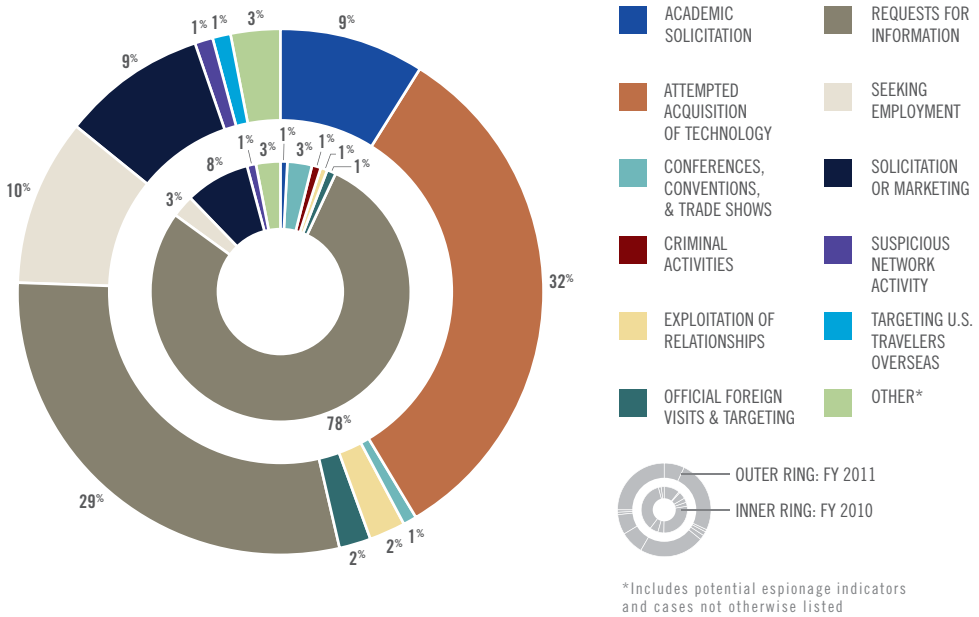
Closely following attempted acquisitions of technology were RFIs, at 29 percent of reported South and Central Asia-originating collection attempts. Commercial entities used unsolicited emails as the primary mechanism to submit purchase requests, ask technical questions, and/or gather information about specific technologies.

Analyst Comment: More South and Central Asia entities are now attempting to develop business relationships with cleared contractors. It is likely that the attempted acquisition of technology MO surpassed reported RFIs in part because of the more amicable relationships between the United States and some South and Central Asia countries, which encourage technology transfer. (Confidence Level: Moderate)

Although the attempted acquisition of technology and RFI categories are separate for the purposes of increasingly discriminating reporting, the means by which these two MOs are employed are very similar. In both methods, an entity contacts a cleared contractor requesting certain sensitive components or platforms, or asking for information such as pricing or technical specifications. The entities

METHODS OF OPERATION

FIGURE 18



*Includes potential espionage indicators and cases not otherwise listed

making these requests mostly appear legitimate; inquiries only occasionally reveal a nefarious or suspicious end user. The difference between the MOs is that in the case of attempted acquisition, the suspicious entity is more likely to be aware that it is not an authorized recipient.

Analyst Comment: Most South and Central Asia procurement agents very likely view RFIs and attempted acquisitions of technology as legitimate and potentially successful means of obtaining sensitive or classified U.S. information and technology. (Confidence Level: Moderate)

FY11 saw the emergence of academic solicitations by South and Central Asia actors, totaling nine percent of reported collection attempts linked to that region in comparison to one percent the year before. Governments within the region are engaged in expanding institutions of higher learning in number and enrollment, to more closely parallel enrollment of students in Western countries. University requirements of an internship for students—a majority of whom seek to fulfill their internship requirement at a U.S. company—contributed to the number of academic solicitations made to cleared contractors.

Analyst Comment: In addition to the creation of additional South and Central Asia universities, better awareness among cleared contractors concerning foreign students likely contributed to the increase in the number of student résumés, job applications, and inquiries reported by cleared industry in FY11. (Confidence Level: Moderate)

While solicitation or marketing was only the fifth most common MO South and Central Asia collectors used in FY11 as reflected in industry reporting, it remains noteworthy. Although it represented eight percent of the reporting last year and nine percent in FY11, due to the overall increase in reporting related to South and Central Asia the number of cases in this category more than doubled year over year. In most suspicious contact incidents reported by cleared industry involving this MO, a South and Central Asia company offered to act as the cleared contractor's agent or distributor in a particular country or the region.

Analyst Comment: While South and Central Asia entities' attempts to form business partnerships may be legitimate, it is likely that they are intended more to promote an additional avenue to access sensitive or classified U.S. information and technology. Were cleared contractors to enter into such agreements, the South and Central Asia entity would likely request an exchange of personnel or even access to controlled U.S. information and technology as a condition of the deal; either situation could result in unauthorized access to sensitive or classified U.S. information and technology. (Confidence Level: Moderate)

TARGETED TECHNOLOGIES

The top of the list of technologies most frequently reported by industry as having been targeted by collectors from South and Central Asia was fairly stable from FY10 to FY11. Just as in FY10, in FY11 the IS and LO&S sections were tied at the top, at 19 percent. Aeronautics, in third place with ten percent, increased only one percentage point from the year before. Last year's fourth-place technology, positioning, navigation, and time, slid to seventh place in the new listing, now at five percent, allowing electronics to move up one spot from last year, with nine percent of the total. Industry reporting shows that South and Central Asia entities continue to seek a wide and diverse range of dual-use technologies from cleared contractors.

The IS technologies South and Central Asia collectors targeted in FY11 included modeling and simulation (M&S) software, used for range-testing of aircraft and missiles. Existing South and Central Asia missile systems may lack radar and testing equipment adequate to track, review, and improve test results

TARGETED TECHNOLOGY
FIGURE 19

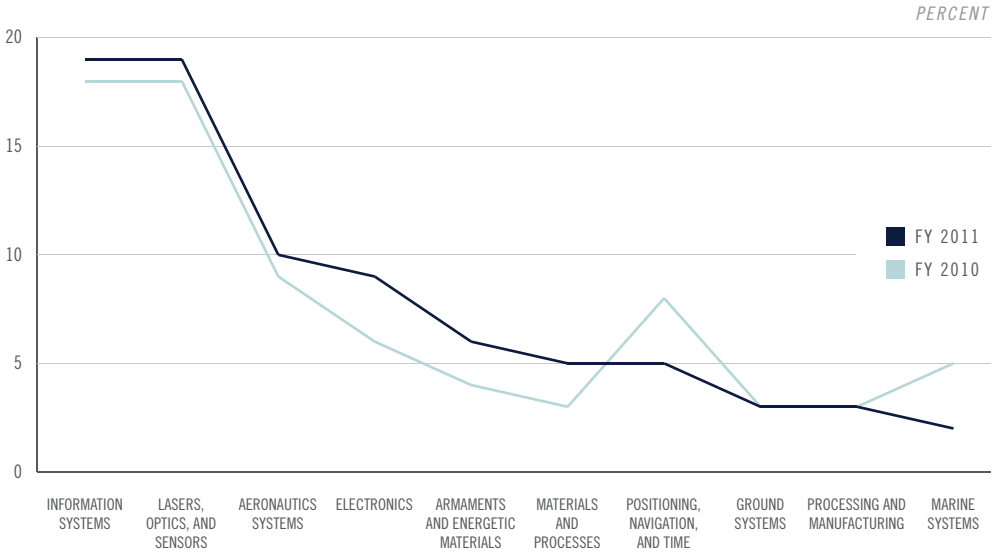


Figure illustrates the top ten most targeted technology categories in FY11 compared with the reporting statistics for the same categories from FY10.

accurately. To achieve a competitive military and economic edge in the region with regard to radar capabilities and products, collectors seek enhanced tracking capability.

Analyst Comment: It is likely that acquisition of more advanced M&S software would improve South and Central Asia entities' radar capabilities, which would likely assist in correcting deficiencies in a multitude of areas, including missiles, surveillance systems, and training programs. (Confidence Level: Moderate)

Additionally, in a large number of cases, South and Central Asia students sent résumés requesting positions in the information technology (IT) field, including programming, software development, and network systems engineering, any of which could facilitate access to cleared contractors' IS-related materials, software, and technologies.

Analyst Comment: While some of the requested positions do not directly involve classified material, they may allow access to proprietary and/or export-controlled information. When students in such positions complete their internships or employment, they possess the potential to either knowingly or unknowingly transfer sensitive information back to their home countries. There is an even chance that relationships opened by such student contacts with U.S. cleared contractors are exploited by the students' home countries. (Confidence Level: Moderate)

Technology areas within the LO&S and electronics systems sections of the MCTL that South and Central Asia entities specifically targeted in FY11 included thermal imaging cameras. South and Central Asia is characterized by security concerns from perceived threats both within and outside the region. Persistent and continuous requests for thermal imaging

systems, considered in the context of IC and open-source reporting, indicate that South and Central Asia actors are probably continuing to build their surveillance technology base for application to border security, and in response to a growing threat of missile deployment along those borders.

Other systems targeted within the LO&S section of the MCTL included fire control radar, airborne warning radar, medium wave infrared cameras, and battlefield surveillance radar (BSR). The volume of requests from South and Central Asia for BSR technology reported by industry, combined with 2010 IC reporting, indicates that some South and Central Asia militaries have a limited BSR capability but are seeking to upgrade it, including to achieve systems networking capability.

Analyst Comment: South and Central Asia nations likely view BSR systems as crucial to protecting their borders. There is an even chance that many South and Central Asia-connected attempts to acquire U.S. BSR systems are a response to similar efforts by their neighbors to improve their own BSR systems. (Confidence Level: Moderate)

To further support goals regarding border security, as well as intelligence, surveillance, and reconnaissance (ISR), weapons mobility/deployment, and the conduct of terrain studies, South and Central Asia companies and agencies targeted unmanned aerial vehicles (UAVs). Entities continued to request not only whole UAV systems but also increased their requests for UAV components, as defense industries and laboratories in the region worked toward self-production of complete UAVs. Some in the region have developed indigenous unmanned aerial systems (UASs), but have experienced difficulties in developing advanced systems.

Analyst Comment: South and Central Asia entities have made multiple attempts to acquire U.S. long-range, ISR-capable UAVs, including those that can be launched from either ship or coastal installations. Their targeting of U.S. UAVs almost certainly reflects an effort to support force modernization plans and upgrades. (Confidence Level: High)

OUTLOOK

DSS assesses that South and Central Asia entities almost certainly perceive an enduring need for foreign, in particular U.S., technology. Ongoing and intensifying conflicts in the region, border issues with neighbors within and outside the region, frictions with the United States, and internal security concerns are likely to motivate South and Central Asia countries. As neighbors and rivals continue efforts to collect and advance upon multiple technology platforms, countries desire to counter with capable technologies of their own. To counter perceived threats, South and Central Asia collectors will almost certainly continue to attempt acquisition of and collection against U.S. information and technology. **(Confidence Level: Moderate)**

Given the perceived imperative to improve military capabilities, there is an even chance that South and Central Asia entities that encounter what they perceive as delays in acquiring desired technology, including dual-use systems, through legitimate avenues will turn to illicit methods. There is an even chance that South and Central Asia agencies' and companies' motivations to protect their own interests will outweigh their inclination to follow U.S. export laws, especially if they risk compromising security within the region, hampering defense industry development, and reducing their own revenue. In order to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base, some South and Central Asia entities will likely attempt to exploit relationships with the United States. **(Confidence Level: Moderate)**

If successful at illicitly acquiring U.S. information and technology from cleared

contractors, some South and Central Asia entities are likely to share such materials with intra- and interregional allies. Such alliance relationships are likely to continue to develop, and the out-of-region parties involved would thereby gain illicit access to U.S. military technology, even that which was legally acquired originally. **(Confidence Level: Moderate)**

There is also an even chance of increased exploitation attempts from South and Central Asia cyber actors. The region's active and growing IT sector produces and employs individuals capable of hacking computer systems. According to industry reporting from FY11, such individuals contacted cleared contractors to establish business relationships with their companies. While no reporting indicates these South and Central Asia IT companies are acting as intelligence collection sources at this time, their capabilities are likely advanced enough for them to be exploited as a collection tool. **(Confidence Level: Moderate)**

The existing and expanding technical institutes that graduate those with such capabilities are likely to produce an increase in student requests to U.S. cleared contractors. Government initiatives will probably enhance R&D partnerships between South and Central Asia training institutes and government agencies, which are then likely to increase their combined outreach to U.S. cleared contractors. **(Confidence Level: Moderate)**

DSS assesses that South and Central Asia collection efforts will probably continue to rely heavily on commercial entities acting as government procurement agents to acquire U.S. technology. These entities will almost certainly continue to use RFIs and attempted acquisitions of technology to obtain sensitive or classified U.S. information and technology for their defense industries. By and large, such requests will very likely continue via email and web card, with occasional in-person contact. While most such approaches will almost certainly be made by legitimate entities, it cannot be discounted that disreputable actors will attempt to obscure the illicit nature of their acquisition attempts amid the increasing volume of reports from commercial entities. **(Confidence Level: High)**

DSS assesses that South and Central Asia entities will very likely continue their collections against U.S. cleared contractors' LO&S systems, software testing materials, infrared and surveillance technologies, and UAS components. Much of this effort will very likely be the result of force modernization requirements and upgrades, but will also likely reflect the perceived need to maintain parity with or even outpace neighbors' capabilities in these areas. **(Confidence Level: Moderate)**

CASE STUDY: GROUND (RADAR) ATTACK

The following is an example of South and Central Asia use of a procurement agent to obtain information regarding a sensitive U.S. technology. This collector has a history of making inquiries on behalf of the military.

In December 2010, a representative of a South and Central Asia company visited the booth of a cleared contractor at the Defence Security and Equipment International conference in London and followed up with an email requesting to market the cleared contractor's ground surveillance radar (GSR) and other technologies to his country.

The individual in question had previously used the same MO at a 2010 Washington, D.C., conference, visiting the booth of the same cleared contractor and following up with an email to inquire about marketing GSR to his country's army. IC reporting indicates that he is a procurement agent for his country's intelligence service and the country's military. Over the last few years, he has attended various defense shows attempting to procure equipment for his country's military.

Analyst Comment: Based on the agent's ties to his government, DSS assesses that he probably conducts his attempts to acquire sensitive or classified information and technology at the behest of his country's military and intelligence establishments. DSS assesses that there is an even chance that his government uses him and his company to leverage the process of legitimate inquiry to obtain information and technologies from U.S. businesses. (Confidence Level: Moderate)

OTHER REGIONS

Entities originating from the Western Hemisphere and Africa accounted for just seven percent of the collection attempts targeting U.S. information and technology reported by cleared industry in fiscal year 2011 (FY11). This was a marginal decrease from previous years in the share of overall reported collection attempts, down from representing eight percent of all attempts in FY10 and ten percent of all attempts in FY09.

DSS attributed a larger number of suspicious reports to entities from both of these regions in FY11 than previously. However, the increase in reports linked to these regions was far lower than the overall increase in reporting from FY10 to FY11, which increased by 75 percent, while reporting attributed to the Western Hemisphere increased by just 49 percent and that from Africa by just six percent.

Commercial entities from both of these regions were the most active at attempting to collect U.S. technologies, as reported by industry. Commercial entities from Africa conducted over half of the collection attempts attributed to this region, while commercial entities accounted for 35 percent of the attempts originating from the Western Hemisphere. Government entities were the second most common affiliation for entities from Africa, accounting for over a fifth of all reported attempts from this region. In contrast, individual was the second most common affiliation for entities from the Western Hemisphere, tallying one-third of all reported collection attempts linked to it.

Entities from both of these regions relied heavily on the request for information followed by attempted acquisition of technology as their primary methods of operation.

Based on industry reporting in FY11, entities from the Western Hemisphere most actively targeted information systems (IS), aeronautics systems, electronics technology, and lasers, optics, and sensors (LO&S), in that order. These four categories accounted for 40 percent of the collection attempts attributed to the region. Similarly, entities from Africa favored information pertaining to aeronautics systems, IS, LO&S, and armaments and energetic materials technology, in that order. Collection attempts targeting those four categories accounted for over two-thirds of those attributed to Africa.

Analyst Comment: The number of attempts to target U.S. technologies originating from these two regions will likely continue to increase, albeit at a slower pace than those from the other four regions. Countries in the Western Hemisphere and Africa largely possess smaller armed forces and less developed defense industrial bases than those in East Asia and the Pacific, the Near East, and Europe and Eurasia. (Confidence Level: Moderate)

CONCLUSION

Technologies resident in U.S. cleared industry remain highly sought after. Foreign intelligence entities (FIEs) continue to expand their collection networks and activities. These networks are growing like a malignant vine. This ongoing theft—FIEs' pilfering of U.S. technologies from cleared industry—could reduce or even end advantages in military capabilities the United States possesses over potential adversaries, thereby adversely affecting U.S. battlefield dominance. It also could strangle U.S economic growth, vitiating the nation's economic health.

The overall number of reports submitted by cleared industry to the Defense Security Service (DSS) in FY11 increased by nearly 65 percent over FY10, and the number that actually became suspicious contact reports (SCRs) increased by 75 percent, likely due in large part to increased awareness and reporting by industry.

Many of the attributes of the entities targeting U.S. technologies remained constant from FY10 through FY11. The order of the regions linked to the most prolific collectors of U.S. information and technology remained unchanged from FY10; commercial remained the most common collector affiliation; and the top four most targeted technology categories remained the same. A modest change in the favored method of operation (MO) occurred, with attempted acquisition of technology becoming the most common MO. This largely reflected a change in terminology, in that DSS would have classified many incidents of attempted acquisition of technology as a request for information (RFI) in previous years. In FY11, RFI became the second most common MO.

Constancy of the order of the regions represents the most enduring trend. Over the past five years, the only change in the order occurred in FY07 and FY09, when South and Central Asia was the third most prolific and Europe and Eurasia the fourth; the other three years, Europe and Eurasia has been the third most prolific. East Asia and the Pacific and the Near East have remained the first and second most prolific collector regions throughout the five years, responsible for at least 56 percent of all reported collection attempts each year.

As previously noted, entities linked to East Asia and the Pacific remained the preeminent attempted collectors of U.S. technology. Over the past five years, entities from this region accounted for 42 percent of all collection attempts reported to DSS. Entities from the Near East consistently represented the second most active collectors, but accounted for just slightly over 18 percent of all reporting.

Analyst comment: East Asia and the Pacific features many areas with a permissive environment in which collectors can operate. In some areas, collection efforts, even those by commercial and individual entities, have government sanction, or at least tacit approval; in some cases, collection is conducted at government direction. In other areas, lax export controls provide collectors a permissive environment from which to acquire technology and subsequently forward

it to entities in other areas of East Asia and the Pacific or beyond to other regions. (Confidence Level: Moderate)

In FY11, foreign entities identified as commercial made that affiliation the most common one in industry reporting for collectors targeting U.S. information and technology. Commercial entities have constituted the most common affiliation in each of the past five years, accounting for over 36 percent of all the reported collection attempts during that period. In FY11, commercial entities were the most common affiliation in five of the six regions, the only exception being government-affiliated entities in the Near East region.

In FY11, the individual affiliation accounted for the second most reported attempts to collect U.S. technology, as reported by industry. This was a significant shift from previous years. Over the five-year period FY07 through FY11, the individual affiliation accounted for just over 13 percent of all collection attempts, the fifth most common. From FY07 through FY09, entities identified as individuals accounted for no more than nine percent of the attempts to collect U.S technologies, and was consistently the fifth most common affiliation. In FY10, the individual affiliation was the fourth most common for attempted collectors and accounted for 12 percent of the collection attempts. In FY11, the number of collection attempts attributed to individuals increased by more than 160 percent over the total from FY10, and accounted for 18 percent of the total collection attempts. This may be related to the increase in academic solicitation.

The third and fourth most common affiliations, government-affiliated and government, both significantly increased in number of reported attempts to target cleared industry. Government-affiliated experienced a 100 percent increase and reported attempts conducted by government entities increased by 165 percent in FY11 over FY10. Much of the increase in attempts attributed to government reflects better reporting and attribution, which reduced the number of attempts credited to unknown entities. Over the past five years, attempts by unknown entities accounted for over 17 percent of all reported collection attempts, and was the second most common affiliation over that period. However, in FY11, unknown was the fifth most common affiliation, accounting for 14 percent of the collection attempts.

Consistently throughout the past five years, the most frequently applied MOs for collectors have been to directly request information or attempt to acquire technology. Attempted acquisition of technology and request for information (RFI) were the two most common MOs. Together in FY11 they accounted for 43 percent of reported collection attempts. A redefinition of attempted acquisitions led to DSS attributing many cases in FY11 to that category that would have been considered RFIs in previous years. Thus, reported efforts via attempted acquisition of technology jumped from less than one percent in FY10 to 23 percent in FY11. Consequently, RFIs plummeted over the same period from representing 48 percent of reported attempted collections to 20

percent. Collectively, these MOs represent direct overt contact with cleared industry in an attempt to receive information or acquire technology by asking for it.

Suspicious network activity (SNA) continued to be a growing phenomenon in FY11. The number of reported SNA collection attempts increased by 36 percent in FY11 over FY10. Better detection and reporting by industry has contributed greatly to improved identification of SNA and the ability to attribute it to particular regions. In FY11, SNA was the most prevalent collection MO for entities originating from East Asia and the Pacific. This is the only region identified as leveraging SNA so heavily; SNA figured no more prominently than fifth in any other region. However, in reports for which the region of origin is unknown, SNA was again the most prominent MO. Due to the nature of SNA, it is difficult to attribute some collection attempts to an entity or even to a region of origin.

The most sought after technologies in FY11 remained largely the same. The top four most targeted technology categories—information systems (IS); lasers, optics, and sensors (LO&S); aeronautics systems; and electronics—remained unchanged. Armaments and energetic materials replaced marine systems as the fifth most targeted category of the Militarily Critical Technologies List (MCTL). The top five in FY11 were the most commonly targeted technologies for the last five years.

A trend for the past three years is an apparent broadening of the targeting of technology: the focus of collectors seems to be diffusing. In FY07, the top five targeted technologies accounted for 67 percent of all reported collection attempts. In FY09, these technologies continued to represent over 66 percent of reported collection attempts. However, in FY10, the top five targeted technology categories accounted for 57 percent, and this dropped further in FY11, with the top five categories accounting for just over 51 percent of reported collection attempts.

This apparent broadening of interest in technology has made space systems, processing and manufacturing, and directed energy systems more common targets for collectors. In FY09, collectors targeted space systems in fewer than two percent of reported collection attempts, whereas in FY11, collectors targeted space systems in almost five percent of reported attempts. In the same period, collection attempts aimed at directed energy systems went from one-sixth of one percent to over two percent of all reported collection attempts.

Analyst Comment: If this diffusion of interest continues in FY12 and beyond, it may signify that some competitor countries now consider themselves peers to the United States in those technologies that formerly were the most highly sought after, such as IS technology. Such countries may further turn the focus of their collection efforts to other technology categories, such as space systems, in which the United States retains an advantage. (Confidence Level: Low)

OUTLOOK

Access to and application of the latest technologies is a vital component of being victorious on the battlefield and competitive economically. The technologies resident in U.S. cleared industry represent the latest and greatest advances. But this technological advantage is under perpetual attack from foreign intelligence entities (FIEs) representing political adversaries and economic competitors. This onslaught of espionage targeting U.S. technologies is constant and unwavering. In fact, this year's reporting suggests this persistent attack on U.S. technologies continues to grow.

A technological advantage can be devastating on the battlefield, providing one side with a decisive victory while it suffers limited losses. In 1991, Coalition forces, led by the United States and armed with the most advanced weapons systems, crushed an Iraqi army that had established itself in defensive positions in Kuwait and southern Iraq. The Iraqi army deployed aging equipment, most of which was a generation older than that wielded by the United States and its NATO allies in the coalition. Coalition soldiers, sailors, airmen, Marines, and Coast Guardsmen used stealth technology, precision weapons systems, and superior battlefield surveillance technology to their advantage, helping to lead to a decisive victory.

Conversely, conflict between opponents sharing technologic parity can lead to bloody, costly, and enervating conflagrations. On July 1, 1916, to relieve the pressure on the French army fighting near Verdun, the British army initiated an offensive against German lines near the Somme River. During the week prior to the offensive, the British fired over 1.7 million artillery rounds against the German lines. On the first day of the battle, the British advanced with over 100,000 men—and suffered an estimated 60,000 casualties, including 20,000 deaths. The Battle of the Somme would last until November of 1916 and cost the British 420,000, the French 200,000, and the Germans 500,000 casualties.[4] The Battle of the Somme featured opposing forces largely armed with the same generation of weaponry. It also demonstrated that the offensive tactics of the day could not match the modern firepower wielded by the defense.

Advances in technology are equally important to the economic health of a country. The fortunes of a country can hinge upon an advantage in industry. In 1789, Samuel Slater (1768-1835) emigrated from England to a young and newly independent United States. Prior to leaving England, while working in the textile industry, he had memorized the design and workings of the water mill designed by Richard Arkwright. At that time, England strictly restricted the export of textile machinery or technology. Slater claimed to be a farmer when leaving England, fearing he would not be allowed to leave if authorities knew his true profession. After arriving in the United States, Slater was instrumental in establishing the first water-powered cotton-spinning mill in the country.[5] This violation of export controls, along with Slater's ability to replicate the mill machinery, greatly accelerated the industrial revolution in America. Furthermore, this story demonstrates that it can often be as

important to obtain information and design details of a given technology as the actual piece of equipment.

The battlefield and economic advantage enjoyed by the United States is precarious, and the loss of the advantage on the battlefield would likely have disastrous results for U.S. forces. Concurrently, the continuing invasive collection of U.S. technologies would likely further erode the U.S. technological advantage and cause severe repercussions to the U.S. economy. **(Confidence Level: Moderate)**

Those who attempt to collect U.S. technologies will almost certainly continue to target a wide variety of them, spanning the entire spectrum delineated in the Militarily Critical Technology List (MCTL). Collectors will very likely target, to some extent, technologies in all 20 MCTL sections, in addition to sensitive and classified information held in cleared industry. **(Confidence Level: High)**

Collectors will likely continue to focus greater attention on particular technology sections of the MCTL. Overall, information systems (IS); lasers, optics, and sensors (LO&S); aeronautics systems; and electronics technology will very likely experience the most targeting attempts from foreign entities. **(Confidence Level: High)**

IS technology will almost certainly remain the most sought after category of technology by foreign collectors. The category encompasses a wide range of enabling technologies that can provide military and commercial advantage. Collectors will likely continue to target command, control, communications, computers, intelligence, surveillance, and reconnaissance technologies; modeling and simulation software; and advanced radio technologies. **(Confidence Level: High)**

LO&S technology has held its position as the second most sought after category for the last two years, and will very likely remain a highly targeted MCTL sector. In fiscal year 2009 (FY09), the Defense Security Service treated LO&S as two separate categories, which, if combined, would have been the most targeted technology category. **(Confidence Level: Moderate)**

While IS, LO&S, and aeronautics systems technology will likely remain the most targeted, FIEs will probably increase their targeting of information and technology relating to space systems technology as well as technologies in other MCTL categories with application to the space industry, including radiation-hardened integrated circuits. **(Confidence Level: Moderate)**

Although the methods of operation (MOs) used by collectors will very likely continue to evolve, it is almost certain that attempted acquisition of technology and request for information will continue to be the most prominent MOs. **(Confidence Level: High)**

Cyber-based collection, characterized as suspicious network activity (SNA), will almost certainly continue to increase as adversaries apply new malicious programs to target the vulnerabilities inherent in systems connected to the Internet. **(Confidence Level: High)**

Academic solicitation will likely remain a common MO for entities originating in East Asia and the Pacific and the Near East. **(Confidence Level: Moderate)**

In FY11 reporting, commercial entities were the most common attempted collectors of U.S. technologies in all but one of the six regions. It is very likely that commercial will continue to be the most common collector affiliation overall in reporting data. Some companies seek U.S. sensitive and classified information and technology to develop and sell their own products for profit. But commercial entities can also provide a layer of separation between the collector and the foreign government. This affords the foreign government the ability to deny involvement in the targeting of U.S. information and technology. In addition, collectors likely employ commercial entities in third countries to target U.S. technology in order to hide the identity of the intended end user and circumvent export controls. **(Confidence Level: Moderate)**

Outside the continued predominance of commercial entities as collectors, the number of government entities identified as collecting will likely increase with improved reporting of SNA by industry. Government entities identified as targeting U.S. technology, especially via SNA, will likely continue to most frequently originate in East Asia and the Pacific. **(Confidence Level: Moderate)**

In the other regions, government-affiliated entities such as academic and research institutions or individuals will probably be the next most common type of entities targeting U.S. technologies, after commercial. **(Confidence Level: Moderate)**

Entities from East Asia and the Pacific will almost certainly remain the most prolific in collection attempts reported by cleared industry. This region features contentious boundaries and encompasses economic rivals of the United States. The perceived need within this region for modern militaries combined with growing economies will very likely fuel the continued targeting of U.S. technologies as an efficient and effective method of abbreviating research and development of new and emerging technologies. **(Confidence Level: High)**

The Near East will probably continue to account for the second most reported collection attempts targeting cleared industry. Adversarial forces in the region seek the latest in technology to enhance their security, to re-package and re-sell for commercial gain, and to circumvent international sanctions. **(Confidence Level: Moderate)**

Persistent and pervasive foreign collection attempts to obtain illegal or unauthorized access to sensitive or classified information and technology resident in the U.S. cleared industrial base will almost certainly continue unabated in the future. FIE MOs will likely evolve and the specific technologies targeted will probably change, but the constancy and aggressiveness of the campaign of collection attempts will almost certainly not subside. **(Confidence Level: High)**

EXPLANATION OF ABBREVIATIONS AND ACRONYMS

ALL ARE U.S. UNLESS OTHERWISE INDICATED

OMITTED: FOREIGN ACRONYMS THAT APPEAR IN ONLY ONE PLACE

BSR	battlefield surveillance radar
C4ISR	command, control, communications, computers, intelligence, surveillance, and reconnaissance
CI	counterintelligence
CPI	critical program information
CV	curriculum vitae
DoD	Department of Defense
DRAM	dynamic random-access memory
DSS	Defense Security Service
ERC	End-User Review Committee
FAVA-RE	failure and vulnerability analysis and reverse-engineering
FIE	foreign intelligence entity
FY	fiscal year
GSR	ground surveillance radar
IC	Intelligence Community
IO	intelligence officer
IS	information systems
ISR	intelligence, surveillance, and reconnaissance
IT	information technology
LO&S	lasers, optics, and sensors
MCTL	Militarily Critical Technologies List
MO	method of operation
M&S	modeling and simulation
NISPOM	National Industrial Security Program Operating Manual
PROM	programmable read-only memory
RAD-HARD	radiation-hardened
R&D	research and development
RFI	request for information
SCR	suspicious contact report
SEE	single-event effect
SNA	suspicious network activity
SRAM	static random-access memory
TAA	trade assistance agreement
UAS	unmanned aerial system
UAV	unmanned aerial vehicle

AFRICA	EAST ASIA AND THE PACIFIC	EUROPE AND EURASIA	NEAR EAST	SOUTH AND CENTRAL ASIA	WESTERN HEMISPHERE
Angola	Australia	Albania	Algeria	Afghanistan	Antigua and Barbuda
Benin	Brunei	Andorra	Bahrain	Bangladesh	Argentina
Botswana	Burma	Armenia	Egypt	Bhutan	Aruba
Burkina Faso	Cambodia	Austria	Iran	India	Bahamas, The
Burundi	China	Azerbaijan	Iraq	Kazakhstan	Barbados
Cameroon	Fiji	Belarus	Israel	Kyrgyz Republic	Belize
Cape Verde	Indonesia	Belgium	Jordan	Maldives	Bermuda
Central African Republic	Japan	Bosnia and Herzegovina	Kuwait	Nepal	Bolivia
Chad	Kiribati	Bulgaria	Lebanon	Pakistan	Brazil
Comoros	Korea, North	Croatia	Libya	Sri Lanka	Canada
Congo, Democratic Republic of the	Korea, South	Cyprus	Morocco	Tajikistan	Cayman Islands
Congo, Republic of the	Laos	Czech Republic	Oman	Turkmenistan	Chile
Cote d'Ivoire	Malaysia	Denmark	Palestinian Territories	Uzbekistan	Colombia
Djibouti	Marshall Islands	Estonia	Qatar		Costa Rica
Equatorial Guinea	Micronesia	European Union	Saudi Arabia		Cuba
Eritrea	Mongolia	Finland	Syria		Dominica
Ethiopia	Nauru	France	Tunisia		Dominican Republic
Gabon	New Zealand	Georgia	United Arab Emirates		Ecuador
Gambia, The	Palau	Germany	Yemen		El Salvador
Ghana	Papua New Guinea	Greece			Grenada
Guinea	Philippines	Greenland			Guatemala
Guinea-Bissau	Samoa	Holy See			Guyana
Kenya	Singapore	Hungary			Haiti
Lesotho	Solomon Islands	Iceland			Honduras
Liberia	Taiwan	Ireland			Jamaica
Madagascar	Thailand	Italy			Mexico
Malawi	Timor-Leste	Kosovo			Netherlands Antilles
Mali	Tonga	Latvia			Nicaragua
Mauritania	Tuvalu	Liechtenstein			Panama
Mauritius	Vanuatu	Lithuania			Paraguay
Mozambique	Vietnam	Luxembourg			Peru
Namibia		Macedonia			St. Kitts and Nevis
Niger		Malta			St. Lucia
Nigeria		Moldova			St. Vincent and the Grenadines
Rwanda		Monaco			Suriname
Sao Tome and Principe		Montenegro			Trinidad and Tobago
Senegal		Netherlands			United States
Seychelles		Norway			Uruguay
Sierra Leone		Poland			Venezuela
Somalia		Portugal			
South Africa		Romania			
Sudan		Russia			
Swaziland		San Marino			
Tanzania		Serbia			
Togo		Slovakia			
Uganda		Slovenia			
Zambia		Spain			
Zimbabwe		Sweden			
		Switzerland			
		Turkey			
		Ukraine			
		United Kingdom			

REFERENCES

[1] Source redacted; Available upon request from DSS

[2] BBN Technologies; Internet Security Glossary, May 2000; Accessed on June 6, 2012; tools.ietf.org/html/rfc2828

[3] U.S. Attorney's Office, Eastern District of Virginia; September 20, 2011; press release; Chinese Nationals Sentenced to 24 Months for Illegally Attempting to Export Radiation-Hardened Microchips to PRC; http://www.justice.gov/usao/vae/news/2011/09/20110930chinese.nr.html; News; Unclassified

[4] Open source website; History Learning Site; Battle of Somme; http://historylearningsite.co.uk/somme.htm; Background; UNCLASSIFIED

[5] Open source website; Public Broadcasting Service; Who Made America? – Samuel Slater; http://www.pbs.org/wgbh/theymadeamerica/whomade/slater_hi.html; Background; UNCLASSIFIED

DSS MISSION

DSS supports national security and the warfighter, secures the nation's technological base, and oversees the protection of sensitive and classified information and technology in the hands of industry.

We accomplish this mission by: clearing industrial facilities, personnel, and associated information systems; collecting, analyzing, and providing threat information to industry and government partners; managing foreign ownership control and influence in cleared industry; providing advice and oversight to industry; delivering security education and training; and, providing information technology services that support the industrial security mission of the Department of Defense and its partner agencies.

THIS PRODUCT WAS COORDINATED WITH: ACIC, AFOSI, DIA, & NGA

Produced by the Defense Security Service
Counterintelligence Directorate
www.DSS.mil


DEFENSE SECURITY SERVICE
UNITED STATES OF AMERICA

Administration Strategy on Mitigating the Theft of U.S. Trade Secrets

Arizona State Law Journal
Winter 2009
41 Ariz. St. L.J. 933

TRADE SECRECY AS AN INSTRUMENT OF NATIONAL SECURITY? RETHINKING THE FOUNDATIONS OF ECONOMIC ESPIONAGE

Aaron J. Burstein

Excerpt:

A. The Incompatible Goals and Structure of Regulations Protecting Trade Secrets and National Security Information

The goals and design of trade secret law are fundamentally different from national security information regulations. Though the purpose(s) of trade secret as a legal doctrine are still widely debated, most commentators agree that it helps to order commercial relationships among private parties. Private parties decide whether to treat information under their control as a trade secret. They decide whether and under what circumstances to disclose the secret to others. And finally, private parties decide whether to take action to enforce their trade secret rights. None of this changed when the EEA made it a federal crime to misappropriate a trade secret.

This subpart lays out the major theories of trade secrecy and relates the text, history, and enforcement of the economic espionage statute to those theories. This Article argues that trade secret law differs fundamentally from the laws that protect national security information. Trade secret law does not provide any of the mechanisms that facilitate the protection of national security information. This subpart also argues that none of the principal theories of trade secret protection are consonant with what Congress hoped to achieve by passing the economic espionage statute, or what prosecutors hope to achieve by enforcing it. The structure and rationale of trade secret protection simply make it unsuitable to do the work of national security.

1. Mechanisms for Trade Secret Protection

Trade secret law contains nothing remotely comparable to the system of centralized assessment of sensitivity and control over dissemination that characterizes classification. Indeed, the absence of any centralized recording of information about trade secrets is part of what distinguishes them from other forms of IP. Unlike patents and trademarks, trade secrets need not be disclosed to, reviewed by, or registered with any agency in order to be enforceable. Also, in contrast to copyright protection, which protects published works with exclusive rights of reproduction and distribution, trade secret protection is lost if the owner fails to take reasonable measures to keep it secret or discloses it to another person without creating at least an implicit duty to keep it confidential.

Any comparable structure--some sort of confidential trade secret registry, for example--would be inimical to the broader scheme of the law. Individual firms decide whether and how to protect trade secrets. Many companies do not maintain lists of their trade secrets, preferring

instead to define broad confidentiality obligations with employees and business partners, and define a trade secret more specifically after they suspect misappropriation.

It would probably be counterproductive to require firms to take these steps. An important policy consideration in courts' efforts to relate trade secrets to patent rights in particular has been that they offer a low-cost way of protecting information that is valuable but not patentable. Imposing greater costs on trade secret holders would likely result in some firms deciding not to protect their trade secrets at all.

An alternative to knowledge management is to induce trade secret holders to be more aggressive in reporting suspected misappropriation to law enforcement agencies, which could then investigate likely instances of economic espionage. Such a voluntary mechanism would at least be consistent with the structure of trade secret law, but it would have to overcome the significant downside that private firms see in reporting trade secret theft to law enforcement agencies. A persistent complaint among private sector representatives is that criminal investigations draw unwanted attention to a company and, even worse, sometimes end up exposing the information that the company sought to protect.

2. Rationales for Trade Secrecy

In addition to providing enforcement mechanisms that are incompatible with protecting national security information, the rationales for trade secret law also are mostly irreconcilable with national security. As a result, trade secret law is not up to the task of channeling the energy of trade secret owners toward the ends of national security; nor is trade secret law successful in changing the incentives of would-be economic spies. Which rationale, if any, best fits trade secret law is a topic of continuing scholarly and jurisprudential debate. The discussion here does not attempt to settle these debates. Instead, the discussion shows that none of the rationales advanced for trade secret protection is commensurate with the goal of preventing the disclosure of information when it would harm national interests.

Efficiency: Providing Incentives to Create and Disclose Information. One commonly cited rationale for trade secret law is that it enhances economic efficiency. A world with trade secret law, the argument goes, provides greater incentives to invent and lowers the costs of exchanging information, relative to a world without it. Thus, the efficiency rationale holds that trade secrecy encourages both information creation and information disclosure.

The argument from efficiency is that trade secret protection provides some marginal incentive to invent or disclose, compared to a world without trade secrecy. For example, trade secrecy provides some exclusivity for inventions that are not or might not be patentable. In these cases, some inventors might refrain from disclosing their inventions to others in order to wait for the Patent Office to decide whether to issue a patent. Other inventors simply might choose not to disclose information at all, because the risk of unauthorized use or disclosure is too great, and the deterrent provided by a purely contractual substitute for trade secrecy is too small. Still others might resort to costly security measures or a highly restricted employee pool. These restrictions are likely to raise the cost of developing new inventions and thus reduce the level of inventive activity.

Both aspects of the efficiency rationale--promoting invention and promoting disclosure--are in obvious conflict with the national security paradigm. The objective of protecting information under the national security paradigm is to prevent disclosures that would impair national interests. This objective is both negative and binary, and a single decisionmaker--the government--decides which disclosures are permissible. In contrast, the efficiency rationale for trade secrecy holds that private parties are best situated to decide whether disclosing information is to their advantage. To first order, trade secret owners do not care whether a prospective trade secret licensee is a national friend or foe, foreign or domestic. What matters is whether a trade secret licensee will respect the terms of the agreement under which it obtains access to the secret. Conversely, trade secret owners may care little about the identity of a trade secret misappropriator. Indeed, the owner might be harmed more by a competitor's misappropriation than by misappropriation from a country in which the firm does not compete.

Additionally, the efficiency rationale views legal protection as a partial substitute for the trade secret owner's investment in security. All the owner has to do is take “reasonable” measures to maintain secrecy. If trade secret owners were encouraged to add the potential harm to national security to the potential costs of a breach, one would expect them to invest more in security and, correspondingly, less in invention. In practice, of course, trade secret owners are unlikely to have good information about how the information security threats they face intersect with national security, which might further distort their investments in secrecy.

Upholding Commercial Morality. While the mismatch between the efficiency rationale and national security centers on the incentives of inventors, the mismatch between the commercial morality rationale centers on the assumptions about the incentives to which adversaries are likely to respond. The commercial morality justification for trade secrecy holds that the law should protect a party that discloses information to another party as part of a confidential relationship. As the Supreme Court has written, “[t]he necessity of good faith and honest, fair dealing, is the very life and spirit of the commercial world.”

Even within the confines of U.S. trade secret law, defining generally accepted industry norms, and determining whether a firm complies with them, is difficult. When globalized industries, multinational firms, and foreign actors enter the picture, the question of where to look for industry norms becomes even more daunting.

In practice, the EEA might obviate some of this inquiry. The offenses are defined under the Act in terms of taking a trade secret “without authorization.” This standard broadens the civil conduct standard of “improper means,” but it does not necessarily clarify what constitutes “authorization.”

The few decisions issued in trade secret theft cases do not address this element of the statute. Indictments simply allege that the defendants “stole[], appropriated, and obtained without authorization” the information that was at issue in each case. In most cases, confidentiality agreements appear to provide the basis for determining that the defendants acted “without authorization.” Thus, the course of EEA enforcement has not clarified which norms the Act seeks to protect.

It is doubtful that devoting more prosecutorial resources to using the EEA to define norms for international economic information collection would succeed. The basic problem of economic espionage is that agents act on behalf of principals who stand outside a shared system of formal and informal constraints on their conduct. Though EEA supporters emphasized that the law was justified to prevent other countries from benefiting from U.S. firms' research and development, they ignored the fact that foreign economic information collection is a game played by different rules. A law passed in the United States does not address these rules.

Punishing Unfair Competition. Finally, some scholars and courts argue that trade secret law should punish unfair methods of competition. Allowing business firms to breach duties of "good faith and honest, fair dealing," the Supreme Court has noted, would exact an "inevitable cost to the basic decency of society when one firm steals from another."

This rationale fails to reconcile trade secret protection with the ends of national security for reasons similar to those discussed in connection with industry norms. Foreign governments seeking know-how from U.S. targets are unlikely to be moved by a domestic policy judgment that this behavior is unfair. These governments simply are not playing the same game as private businesses.

The course of the economic espionage statute's enforcement has built on and enhanced this rhetoric, even when the substance of some economic espionage cases appear to have a tenuous connection to national security interests. Finally, in its oversight of economic espionage threats and enforcement of the statute, Congress has relied mainly upon sources whose perspectives are oriented toward national security.

REGULATING CYBERSECURITY

Nathan Alexander Sales,
George Mason University School of Law

***Northwestern University Law Review*,**
Forthcoming

George Mason University Law and Economics Research Paper Series

12-35

REGULATING CYBERSECURITY

Forthcoming, 107 Northwestern University Law Review (2013)

Nathan Alexander Sales
George Mason University School of Law

ABSTRACT

The conventional wisdom is that this country's privately owned critical infrastructure – banks, telecommunications networks, the power grid, and so on – is vulnerable to catastrophic cyberattacks. The existing academic literature does not adequately grapple with this problem, however, because it conceives of cybersecurity in unduly narrow terms: Most scholars understand cyberattacks as a problem of either the criminal law or the law of armed conflict. Cybersecurity scholarship need not run in such established channels. This article argues that, rather than thinking of private companies merely as potential victims of cyber crimes or as possible targets in cyber conflicts, we should think of them in administrative law terms. Firms that operate critical infrastructure tend to underinvest in cyberdefense because of problems associated with negative externalities, positive externalities, free riding, and public goods – the same sorts of challenges the modern administrative state faces in fields like environmental law, antitrust law, products liability law, and public health law. These disciplines do not just yield a richer analytical framework for thinking about cybersecurity, they also expand the range of possible responses. Understanding the problem in regulatory terms allows us to adapt various regulatory solutions for the cybersecurity context, such as monitoring and surveillance to detect malicious code, hardening vulnerable targets, and building resilient and recoverable systems. In short, an entirely new conceptual approach to cybersecurity is needed.

REGULATING CYBERSECURITY

Nathan Alexander Sales[†]

TABLE OF CONTENTS

INTRODUCTION

The Red Army had been gone for years, but it still had the power to inspire controversy – and destruction.[1] In April 2007, the government of Estonia announced plans to relocate a contentious Soviet-era memorial in its capital city of Tallinn. Known as the Bronze Soldier, the Soviets erected the statue in 1947 to commemorate their sacrifices in the Great Patriotic War and their "liberation" of their Baltic neighbors. The local population, which suffered under the Bolshevik boot for decades, understandably saw the monument in a rather different light. Not long after the announcement, the tiny nation was hit with a massive cyberattack. Estonia, sometimes nicknamed "E-stonia," is one of the most networked countries in the world – its citizens bank, vote, and pay taxes online[2] – and it ground to a halt for weeks. The country's largest bank was paralyzed. Credit card companies took their systems down to keep them from

[†] Assistant Professor of Law, George Mason University School of Law. Thanks to Jonathan Adler, Stewart Baker, Eric Claeys, Bruce Johnsen, Orin Kerr, Bruce Kobayashi, Michael Krauss, Adam Mossoff, Steve Prior, Jeremy Rabkin, Paul Rosenzweig, J.W. Verret, Ben Wittes, and Todd Zywicki for their helpful comments. I'm also grateful to participants in a workshop at the Republic of Georgia's Ministry of Justice. Special thanks to the Center for Infrastructure Protection and Homeland Security for generous financial support.

[1] The events in this paragraph are described in JOEL BRENNER, AMERICA THE VULNERABLE 127-30 (2011); RICHARD A. CLARKE & ROBERT K. KNAKE, CYBERWAR 11-16 (2010); and Ian Traynor, *Russia Accused of Unleashing Cyberwar to Disable Estonia*, GUARDIAN (U.K.), May 16, 2007.

[2] Kelly A. Gable, *Cyber-Apocalypse Now*, 43 VAND. J. TRANSNAT'L L. 57, 61 & n.14 (2010).

being attacked. The telephone network went dark. Newspapers and television stations were knocked offline. Who was responsible for launching what has come to be known as Web War I?[3] The smart money is on Russia, though no one can say for sure.

It could happen here. Government officials like Richard Clarke, the former White House cybersecurity czar, have been warning of an "electronic Pearl Harbor" for years.[4] Others lament the "gaping vulnerabilit[ies]"[5] in America's cyberdefenses and speculate that the economic effect of a major assault could be "an order of magnitude" greater than the September 11, 2001 terrorist attacks.[6] Academic commentators generally agree. Some see the danger as "monumental"[7] and the country's "most pervasive and pernicious threat."[8] Others predict that America's failure to secure its cyber assets could "take down the nation's entire security and economic infrastructure"[9] and "bring this country to its knees."[10] It has even been suggested that "[t]he very future of the Republic" depends on "protect[ing] ourselves from enemies armed with cyber weapons."[11] There are some naysayers,[12] but the consensus that we stand on the brink of a cyber calamity is both broad and deep.

A large scale cyberattack on this country, as in Estonia, likely would target privately held critical infrastructure – banks, telecommunications carriers, power companies, and other firms whose compromise would cause widespread harm.[13] Indeed, America's critical infrastructure,

---

[3] *War in the Fifth Domain*, THE ECONOMIST, Jul. 1, 2010, at 4; *see also* CLARKE & KNAKE, *supra* note 1, at 30; David W. Opderbeck, *Cybersecurity and Executive Power*, 89 WASH. L. REV. __, 3 (forthcoming 2012).

[4] Richard Clarke, *Threats to U.S. National Security*, 12 DEPAUL BUS. L.J. 33, 38 (2000).

[5] Joby Warrick & Walter Pincus, *Senate Legislation Would Federalize Cybersecurity*, WASH. POST, Apr. 1, 2009.

[6] Max Fisher, *Fmr. Intelligence Director: New Cyberattack May Be Worse than 9/11*, THE ATLANTIC (__) (quoting former Director of National Intelligence Mike McConnell); *see also Cyberspace Policy Review* 1 (2009) ("Threats to cyberspace pose one of the most serious economic and national security challenges of the 21st Century for the United States and our allies.").

[7] William C. Banks & Elizabeth Rindskopf Parker, *Introduction*, 4 J. NAT'L SEC. L & POL'Y 7, 11 (2010).

[8] Walter Gary Sharp, Sr., *The Past, Present, and Future of Cybersecurity*, 4 J. NAT'L SEC. L & POL'Y 13, 13 (2010); *see also* CSIS, *Securing Cyberspace for the 44th Presidency* 11 (Dec. 2008); Greg Rattray et al., *American Security in the Cyber Commons*, *in* CONTESTED COMMONS 139, 145 (Abraham M. Denmark & James Mulvenon eds. 2010).

[9] Opderbeck, *supra* note 3, at 2.

[10] Neal Kumar Katyal, *Criminal Law in Cyberspace*, 149 U. PA. L. REV. 1003, 1020 n.45 (2001).

[11] Stephen Dycus, *Congress's Role in Cyber Warfare*, 4 J. NAT'L SEC. L & POL'Y 155, 156 (2010).

[12] *See, e.g.*, Derek E. Bambauer, *Conundrum*, 96 MINN. L. REV. 584, 590 (2011); Jerry Brito & Tate Watkins, *Loving the Cyber Bomb?* (Apr. 26, 2011); Charles J. Dunlap, Jr., *Meeting the Challenge of Cyberterrorism*, 76 INT'L L. STUD. 353, 361 (2002); Seymour H. Hersh, *The Online Threat*, NEW YORKER, Nov. 1, 2010; Martin Libicki, *Rethinking War*, FOREIGN POL'Y 30, 38 (winter 1999/2000).

[13] Davis Brown, *A Proposal for an International Convention to Regulate the Use of Information Systems in Armed Conflict*, 47 HARV. INT'L L.J. 179, 182 (2006); CLARKE & KNAKE, *supra* note 1, at xiii. Federal law defines "critical infrastructure" as "systems and assets, whether physical or virtual, so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters." 42 U.S.C. § 5195c. Some types of critical infrastructure are more important, and less likely to be adequately defended, than others. *See infra* Part I.

approximately 85 percent of which is owned by private firms,[14] already faces constant intrusions.[15] Yet the private sector's defenses are widely regarded as inadequate. Companies are essentially on their own when it comes to protecting their computer systems, with the government neither imposing security requirements nor bearing a share of the resulting costs.[16] According to Bruce Smith, the United States follows a "bifurcated approach to network security" that "relie[s] predominantly on private investment in prevention and public investment in prosecution."[17] Christopher Coyne and Peter Leeson likewise stress that our defensive strategy "is simply the sum of dispersed decisions of individual users and businesses."[18] Regular firms that operate in competitive markets (such as online retailers) may be more likely to effectively protect their systems against ordinary intruders (such as recreational hackers). But strategically significant firms in uncompetitive markets (such as power companies and other public utilities) seem especially unlikely to maintain defenses capable of protecting their systems against skilled and determined adversaries (such as foreign intelligence services).

The poor state of America's cyberdefenses is partly due to the fact that the analytical framework used to understand the problem is incomplete. The law and policy of cybersecurity are undertheorized. Virtually all legal scholarship approaches cybersecurity from the standpoint of the criminal law or the law of armed conflict.[19] Given these analytical commitments, it is

---

[14] Todd A. Brown, *Legal Propriety of Protecting Defense Industrial Base Information Infrastructure*, 64 AIR FORCE L. REV. 214, 220 (2009); Gus P. Coldebella & Brian M. White, *Foundational Questions Regarding the Federal Role in Cybersecurity*, 4 J. NAT'L SEC. L & POL'Y 233, 240 (2010); Coyne & Leeson, *supra* note 18, at 476; Gregory T. Nojeim, *Cybersecurity and Freedom on the Internet*, 4 J. NAT'L SEC. L & POL'Y 119, 135 (2010); Benjamin Powell, *Is Cybersecurity a Public Good?*, 1 J.L. ECON. & POL'Y 497, 497 (2005); Paul Rosenzweig, *Cybersecurity and Public Goods* at 2 (2011), *reprinted in* PAUL ROSENZWEIG, CYBERWARFARE (forthcoming 2012).

[15] Eric Talbot Jensen, *Cyber Warfare and Precautions Against the Effects of Attacks*, 88 TEX. L. REV. 1533, 1537 (2010); Neal Kumar Katyal, *Digital Architecture as Crime Control*, 112 YALE L.J. 2261, 2263 (2003); McAfee, *In the Dark* at 6 (2011); Debra Wong Yang & Brian M. Hoffstadt, *Countering the Cyber-Crime Threat*, 43 AM. CRIM. L. REV. 201, 201, 205 (2006).

[16] Yasuhide Yamada et al., *A Comparative Study of the Information Security Policies of Japan and the United States*, 4 J. NAT'L SEC. L & POL'Y 217, 219 (2010).

[17] Bruce Smith, *Hacking, Poaching, and Counterattacking*, 1 J.L. ECON. & POL'Y 171, 173 (2005).

[18] Christopher J. Coyne & Peter T. Leeson, *Who's to Protect Cyberspace?*, 1 J.L. ECON. & POL'Y 473, 475-76 (2005); *see also* Am. Bar Ass'n, *National Security Threats in Cyberspace* 8 (Sept. 2009); Banks & Parker, *supra* note 7, at 11; Nojeim, *supra* note 14, at 121.

[19] Bambauer, *supra* note 12, at 588-89. For examples of the criminal law approach, see Banks & Parker, *supra* note 7, at 9; Mary M. Calkins, *They Shoot Trojan Horses, Don't They?*, 89 GEO. L.J. 171, 190-97 (2000); Sean M. Condron, *Getting It Right*, 20 HARV. J.L. & TECH. 403, 407 (2007); Katyal, *Criminal Law*, *supra* note 10; Katyal, *Digital Architecture*, *supra* note 15; Michael Edmund O'Neill, *Old Crimes in New Bottles: Sanctioning Cybercrime*, 9 GEO. MASON. L. REV. 237 (2000); Opderbeck, *supra* note 3, at 2l; Yang & Hoffstadt, *supra* note 15. For examples of the armed conflict, see Davis Brown, *supra* note 13; Condron, *supra*, at 408; David E. Graham, *Cyber Threats and the Law of War*, 4 J. NAT'L SEC. L & POL'Y 87 (2010); Eric Talbot Jensen, *Computer Attacks on Critical National Infrastructure*, 38 STAN. J. INT'L L. 207 (2002); Herbert S. Lin, *Offensive Cyber Operations and the Use of Force*, 4 J. NAT'L SEC. L & POL'Y 63 (2010); William J. Lynn, *Defending a New Domain*, 89 FOREIGN AFF. 97 (2010); Matthew Waxman, *Cyber-Attacks and the Use of Force*, 36 YALE J. INT'L L. 421 (2011); Michael N. Schmitt, *Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework*, 37 COLUM. J. TRANSNAT'L L. 885 (1999); Matthew J. Sklerov, *Solving the Dilemma of State Responses to Cyberattacks*, 201 MIL. L. REV. 1 (2009). There are exceptions. Some scholars understand cybersecurity in public health terms. Jeffrey Hunker, *U.S. International Policy for Cybersecurity*, 4 J. NAT'L SEC. L & POL'Y 197,

inevitable that academics and lawmakers will tend to favor law enforcement and military solutions to cybersecurity problems. These are important perspectives, but cybersecurity scholarship need not run in such narrow channels. An entirely new approach is needed. Rather than conceiving of private firms merely as possible victims of cyber crimes, or as potential targets in cyber conflicts, we should think of them in regulatory terms.[20] Many companies that operate critical infrastructure tend to underinvest in cyberdefense because of problems associated with negative externalities, positive externalities, free riding, and public goods – the same sorts of challenges the modern administrative state encounters in a variety of other contexts, such as environmental law, antitrust law, products liability law, and public health law.

For instance, cybersecurity resembles environmental law in that both fields are primarily concerned with negative externalities. Just as firms tend to underinvest in pollution controls because some costs of their emissions are borne by those who are downwind, they also tend to underinvest in cyberdefenses because some costs of intrusions are externalized onto others. An attack on a power company will not just harm the intended target; it will also harm the target's customers and those with whom the power company has no relationship. Because firms do not bear the full costs of their vulnerabilities, they have weaker incentives to secure their systems. Cybersecurity also resembles an antitrust problem. Antitrust law seeks to prevent anticompetitive behavior, and it traditionally has been skeptical of coordination among competitors. Some inter-firm cooperation could improve cybersecurity – sharing information about vulnerabilities and threats, for example, or developing industry wide security standards. Yet firms are reluctant to do so because they fear liability under the antitrust laws. Next, cybersecurity raises tort problems. Products liability law uses the threat of money damages to incentivize firms to take reasonable precautions when designing their products. This threat is almost entirely absent in the cybersecurity context, as companies face little risk of liability to those who are harmed by attacks on their systems or products. The incentive to patch vulnerabilities thus is weaker than it would be under a meaningful liability regime. Finally, cybersecurity resembles public health. A key goal of public health law is prevention – keeping those who have contracted a disease from spreading it to the healthy, a form of negative externality. Public health law uses vaccinations to promote immunity, biosurveillance to detect outbreaks, and quarantines to contain infectious diseases. Cybersecurity has similar goals – ensuring that critical systems are immune to malware, quickly detecting outbreaks of malicious code, and preventing contaminated computers from infecting clean systems.

Approaching cybersecurity from a regulatory vantage point does not just yield a richer analytical framework. It also expands the range of possible responses. The available solutions are determined by the threshold choice of analytical models; the more frameworks, the longer the menu of policy choices. If cyber insecurity resembles problems that arise in other regulatory contexts, then perhaps some of their solutions can be adapted here. Taken together, these

---

202-04 (2010); Rattray et al., *supra* note 8; IBM, *Meeting the Cybersecurity Challenge* (Feb. 2010). Others approach cybersecurity from an economic perspective. Coyne & Leeson, *supra* note 18; THE LAW AND ECONOMICS OF CYBERSECURITY (Mark F. Grady & Francesco Parisi eds., 2006); Powell, *supra* note 14; Rosenzweig, *supra* note 14; Supriya Sarnikar & D. Bruce Johnsen, *Cyber Security in the National Market System*, 6 RUTGERS BUS. L.J. 1 (2009).

[20] *Cf.* Samuel J. Rascoff, *Domesticating Intelligence*, 83 S. CAL. L. REV. 575 (2010) (proposing an administrative law framework for understanding domestic intelligence).

disciplines suggest four groups of responses: (1) monitoring and surveillance to detect malicious code; (2) hardening vulnerable targets and enabling them to defeat intrusions; (3) building resilient systems that can function during attacks and recover quickly; and (4) responding in the aftermath of attacks.

In particular, public health law's distributed biosurveillance network might be used as a model for detecting cyber intrusions. Rather than empowering a single regulator to monitor internet traffic for outbreaks of malicious code, private firms could be tasked with reporting information about the vulnerabilities and threats they experience in much the same way hospitals report to public health authorities. To incentivize participation in this distributed surveillance network, firms might be offered various subsidies (on the theory that cybersecurity data is a public good that the market will tend to underproduce) and liability protections (such as an exemption from the antitrust laws). As for hardening targets, we might adopt industrywide security standards for companies that operate critical infrastructure. These protocols need not be issued in the form of traditional regulatory commands. Instead, as is sometimes the case in environmental law and other fields, the private sector should actively participate in formulating the standards. Tort law has a role to play as well: Threats of liability and offers of immunity might be used to incentivize firms to implement the protocols. Next, because it is inevitable that some cyberattacks will succeed, it is important that critical systems are able to survive and recover. Public health law offers several strategies for improving resilience. Systems that are infected with malware might be temporarily isolated to prevent them from spreading the contagion. Or firms might build excess capacity into their systems that can be deployed in emergencies – the equivalent of stockpiling vaccines and medicines. Finally, retaliation is thoroughly addressed in the existing criminal law and armed conflict literatures, but there is one response that deserves a brief mention here: "hackbacks," in which a victim counterattacks the attacker. Because the counterattack might fall on a third party whose system unwittingly is being used by the assailant, hackbacks can incentivize firms to prevent their systems from being so commandeered. Hackbacks also might weaken attackers' incentives. If assailants know that counterattacks can render their intrusions ineffective, they are less likely to commit them in the first place.

This article proceeds in three parts. Part I considers whether private companies are investing socially optimal amounts in cyberdefenses. Part II describes four regulatory frameworks – environmental law, antitrust law, products liability law, and public health law – and explains their relevance to cybersecurity. Part III surveys solutions used by these regulatory disciplines and considers how to adapt them for the cybersecurity context.

Several preliminary observations are needed. First, I use the terms "cyberattack" and "cyber intrusion" interchangeably to denote any effort by an unauthorized user to affect the data on, or to take control of, a computer system. As used here, the terms include all of the following: "viruses" (a piece of code that "infects a software program and then ensures that the infected program reproduces the virus"[21]), "worms" ("a standalone program that replicates itself"[22]),

[21] O'Neill, *supra* note 19, at 246; *see also* Katyal, *Criminal Law*, *supra* note 10, at 1023; Sklerov, *supra* note 19, at 14-15.

[22] Katyal, *Criminal Law*, *supra* note 10, at 1024; *see also* Sklerov, *supra* note 19, at 15. Viruses and worms are similar. A principal difference is that viruses require human action to propagate – such as clicking on a link or

"logic bombs" (malware that "tells a computer to execute a set of instructions at a certain time or under certain specified conditions"[23]), and "distributed denial of service" ("DDOS") attacks (in which a "master" computer conscripts "zombies" and orders them to disable a victim by flooding it with traffic[24]). Second, this article emphatically is not a paean to traditional, command and control regulation. The conventional wisdom is to avoid cybersecurity regulation,[25] in part because of doubts about the government's ability to manage such a dynamic field. But as I hope to show in the following pages, cybersecurity need not, and in many cases should not, be pursued with heavy handed regulatory tools. It is possible to promote better cyberdefenses with private law, such as by modifying traditional tort law doctrines. As for public law, regulation need not take the form of rigid legal commands backed by threat of sanction; regulatory objectives often can be attained by appealing to private firms' self interest – by offering positive incentives to improve their defenses, not just by punishing them when they fall short. The private sector's poor defenses may represent a market failure, as some have argued,[26] but "[t]here's not much point in replacing a predictable market failure with an equally predictable government failure."[27]

I. AN EFFICIENT LEVEL OF CYBERSECURITY

Our national security "depends heavily on privately owned critical infrastructure."[28] A cyberattack on these private assets could be devastating: With a few keystrokes, adversaries could hack into banks and corrupt customer data, take control of power plants and bring down the electricity grid, open the floodgates of dams, and take telecommunications networks offline.[29] Or worse. Despite the magnitude of the threat, the conventional wisdom is that the private sector is not adequately protecting itself. This section surveys the available evidence on the extent of private cybersecurity expenditures. It then predicts that ordinary firms in competitive markets (like online retailers) are more likely to be investing socially optimal amounts in cyberdefense, while strategically significant firms in uncompetitive markets (like public utilities) are more likely to be underinvesting.

The optimal level of cyber intrusions is not zero, and the optimal level of cybersecurity expenditures is not infinity. From an economic perspective, the goal is to achieve an efficient

---

opening an attachment – but worms replicate on their own. CLARKE & KNAKE, *supra* note 1, at 81; Katyal, *Criminal Law*, *supra* note 10, at 1024; O'Neill, *supra* note 19, at 247.

[23] Katyal, *Criminal Law*, *supra* note 10, at 1025; *see also* O'Neill, *supra* note 19, at 248.

[24] STEWART A. BAKER, SKATING ON STILTS 202-03 (2010); BRENNER, *supra* note 1, at 38-39; CLARKE & KNAKE, *supra* note 1, at 13-14; Lin, *supra* note 19, at 70; Yamada et al., *supra* note 16, at 226.

[25] CLARKE & KNAKE, *supra* note 1, at 108-09.

[26] ABA, *supra* note 18, at 8; BAKER, *supra* note 24, at 237; CSIS, *supra* note 8, at 50, Katyal, *Digital Architecture*, *supra* note 15, at 2285.

[27] BAKER, *supra* note 24, at 237; *see also* Coyne & Leeson, *supra* note 18, at 490; Powell, *supra* note 14, at 507.

[28] BRENNER, *supra* note 1, at 223; *cf.* ABA, *supra* note 18, at 8 ("[P]rivate sector security is often governmental security.").

[29] Stewart Baker, *Denial of Service*, FOREIGN POL'Y 2 (Sept. 30, 2011); BRENNER, *supra* note 1, at 137-54; CLARKE & KNAKE, *supra* note 1, at 64-68;

level of attacks, not to prevent all attacks.[30] Suppose that the expected cost to society of a given cyberattack – its cost discounted by the probability that it will occur – is $5 billion. It would be efficient for society to invest up to $5 billion in countermeasures to prevent the attack. If the necessary countermeasures cost more than $5 billion, the cost of preventing the attack would exceed the resulting security gains. In short, it is worthwhile to invest in cyberdefenses whose marginal costs are less than the marginal benefits of preventing the attacks.[31] Relatedly, some intrusions are more problematic than others. Cybersecurity is a form of risk management, where risk is a function of three variables: vulnerabilities, threats, and consequences. A company with easily hacked systems, that faces a high probability of attacks from sophisticated foreign intelligence services, and whose compromise would cause severe social harm, raises very different problems than a company with relatively robust defenses, that is unlikely to face skilled intruders, and whose compromise would have few consequences for society.

Are individual firms, and society as a whole, investing the right amount in cyberdefense? Most observers believe that firms are underinvesting – and are missing the mark by a wide margin. Richard Clarke proclaims the private sector response an "unmitigated failure,"[32] and scholars generally agree.[33] Very little empirical data is available, but the consensus view has at least some anecdotal support. Studies conducted by McAfee (a computer security firm) in 2010 and 2011 revealed low levels of investment in cyberdefense. The studies found that many firms regard cybersecurity as little more than "a last box they have to check,"[34] and that they neglect network security because they find it too expensive.[35] In particular, McAfee found that companies often have weak authentication requirements[36] – tools that can verify that the person who is accessing a system is who he says he is, and is authorized to access the system. Even fewer have systems that can monitor network activity and identify anomalies.[37] Other studies

[30] Coyne & Leeson, *supra* note 18, at 477-78.

[31] Coyne & Leeson, *supra* note 18, at 478.

[32] CLARKE & KNAKE, *supra* note 1, at 104.

[33] ABA, *supra* note 18, at 8; Banks & Parker, *supra* note 7, at 9; Katyal, *Criminal Law*, *supra* note 10, at 1019; Bruce K. Kobayashi, *Private Versus Social Incentives in Cybersecurity*, *in* Grady & Parisi, *supra* note 19, at 14; Sarnikar & Johnsen, *supra* note 19, at 3, 16; Bruce Schnieier, *Computer Security: It's the Economics, Stupid* 1 (May 16, 2002). *But see* Coldebella & White, *supra* note 14, at 240; Smith, *supra* note 17, at 173 n.12. Some scholars argue that companies are providing a suboptimally *high* level of cybersecurity. Benjamin Powell reports that a 2000 study found that firms would invest in cyberdefenses if they were expected to produce a 20 percent return on investment, which was considerably lower than the 30 percent ROI typically required for information technology investments. Powell, *supra* note 14, at 504. What mechanism could account for a tendency to overinvest? A firm's IT department has incentives to overstate the vulnerabilities the company faces, as cybersecurity fears translate into a larger share of the company's budget; for outside security vendors, such fears mean brisker business. Ross Anderson, *Unsettling Parallels Between Security and the Environment* 2 (May 16, 2002); Bambauer, *supra* note 12, at 604-06; Calkins, *supra* note 19, at 198-99.

[34] McAfee 2011, *supra* note 15, at 1.

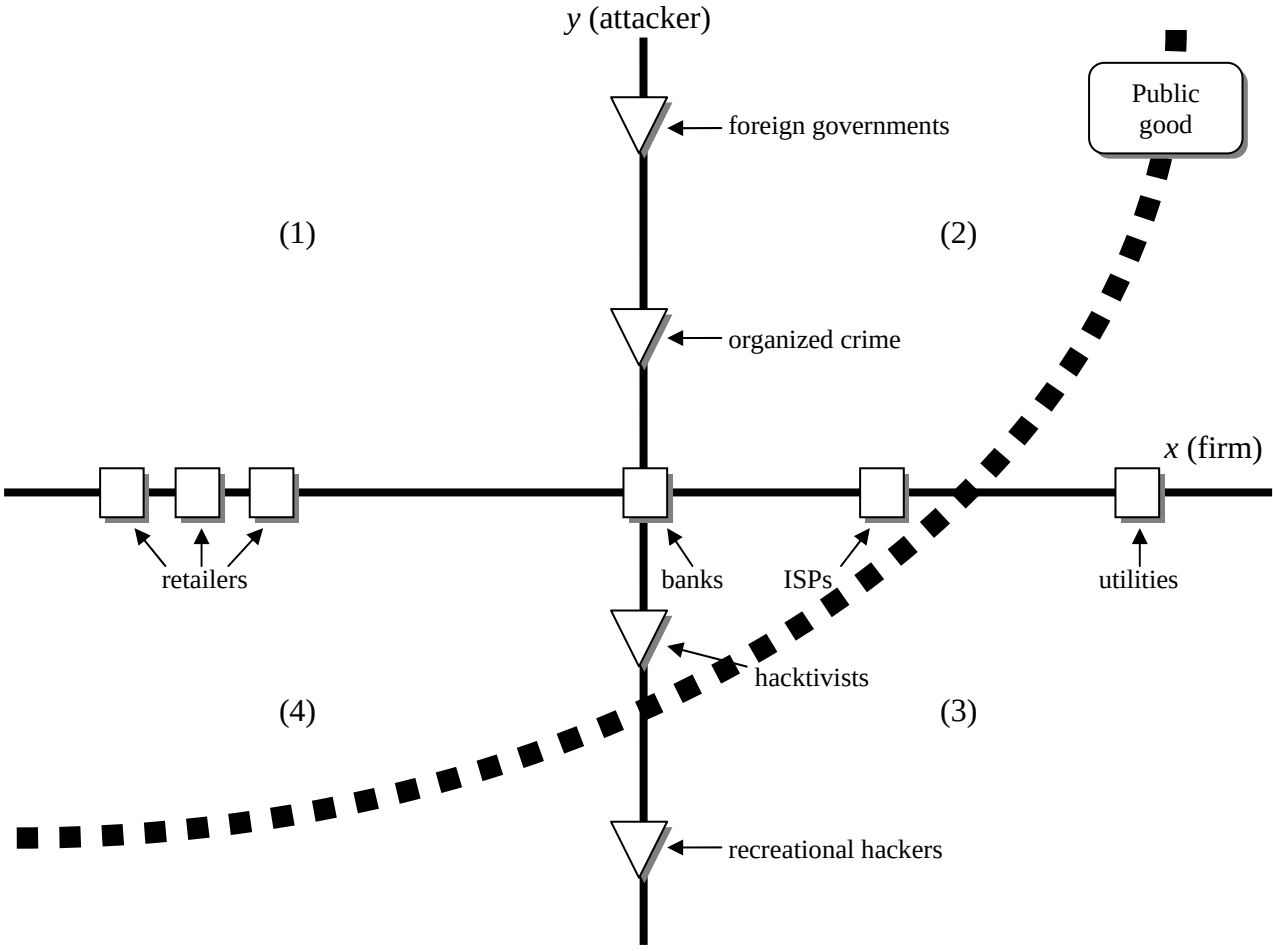
[35] McAfee, *In the Crossfire* at 14 (2010).

[36] McAfee 2011, *supra* note 15, at 14.

[37] McAfee 2011, *supra* note 15, at 15. It would be a mistake to read too much into these findings. The study's methodology was to survey business executives in about a dozen countries, McAfee 2010, *supra* note 35, at 1, 41; McAfee 2011, *supra* note 15, at 3, and it "was not designed to be a statistically valid opinion poll with sampling and

reveal that some companies' defenses are so poor they don't even know when they have suffered an attack. Verizon reported that "fully 75 percent of the intrusions they investigated were discovered by people other than the victims and 66 percent of victims did not even know an intrusion occurred on the system."[38] Finally, a 2011 study by the Ponemon Institute found that "73 percent of companies surveyed had been hacked, but 88 percent of them spent more money on coffee than on securing their Web applications."[39]

Are these levels of investment efficient? Whether a particular firm is making socially optimal investments in cybersecurity – and the related issue of who should pay for that company's cyberdefenses – is a function of two intersecting questions. First, what is the defending firm? Is it a regular company in a competitive market, an operator of critical infrastructure in an uncompetitive market, or something in between? Second, who is the anticipated attacker? Is it a recreational hacker, a foreign intelligence service, or someone in between? The range of possibilities can be depicted in a simple graph:



error margins," McAfee 2010, *supra* note 35, at 1. Moreover, a computer security company obviously stands to benefit from public perceptions that security is lacking.

[38] Rattray et al., *supra* note 8, at 155; *see also* Jensen, *Cyber Warfare*, *supra* note 15, at 1536.

[39] BRENNER, *supra* note 1, at 239.

The *x* axis depicts the firms that might be subject to a cyberattack. They are arranged from left to right in order of increasing strategic significance. (A strategically significant company is one whose compromise would result in substantial social harms.) On the far left are relatively insignificant firms in competitive markets – i.e., markets in which many companies offer the same good or service, and where disappointed consumers therefore may defect from one to another. An example would be online retailers, such as Amazon.com. To the right are financial institutions. These firms rate high on the strategic significance scale; a former CIA director predicted that an attack on a single bank "would have an order-of-magnitude greater impact on the global economy" than 9/11.[40] Banks operate in fairly competitive markets, as consumers can easily move their accounts from one to another. Another step to the right are ISPs and telecommunications carriers. They, too, are strategically significant. When Russia crippled Georgia's communications systems during their 2008 war, citizens "could not connect to any outside news or information sources and could not send e-mail out of the country."[41] These markets are less competitive; consumers typically have only a handful of internet providers or telephone companies to choose from. At the far right are power companies and other public utilities. These firms rate high on the strategic significance scale. A cyberattack on the power grid would be truly catastrophic. The industrial control, or SCADA,[42] systems used by power plants and other utilities are increasingly connected to the internet.[43] Hackers could exploit this connectivity to disrupt power generation and leave tens of millions of people in the dark for months[44]; they could even destroy key system components like turbines.[45] (In 2009, the Stuxnet worm – "the most sophisticated cyberweapon ever deployed"[46] – caused similar physical damage to Iran's nuclear program.[47]) Utility markets are uncompetitive. Municipalities typically have only one power company or natural gas supplier, and there is no meaningful prospect that disappointed consumers will switch to a competitor.

---

[40] *Quoted in* David E. Sanger et al., *U.S. Steps Up Effort on Digital Defenses*, N.Y. TIMES, Apr. 27, 2009, at A1; *see also* Sarnikar & Johnsen, *supra* note 19, at 1; Sklerov, *supra* note 19, at 19-20.

[41] CLARKE & KNAKE, *supra* note 1, at 19; *see also* BRENNER, *supra* note 1, at 39-40; Jensen, *Cyber Warfare*, *supra* note 15, at 1540.

[42] The acronym stands for "supervisory control and data acquisition." CLARKE & KNAKE, *supra* note 1, at 98; CSIS, *supra* note 8, at 54; Randal C. Picker, *Cybersecurity: Of Heterogeneity and Autarky*, *in* Grady & Parisi, *supra* note 19, at 126.

[43] BRENNER, *supra* note 1, at 97; Steven R. Chabinsky, *Cybersecurity Strategy*, 4 J. NAT'L SEC. L & POL'Y 27, 28 n.1 (2010); Condron, *supra* note 19, at 407; Coyne & Leeson, *supra* note 18, at 474; CSIS, *supra* note 8, at 54; Sklerov, *supra* note 19, at 18.

[44] BRENNER, *supra* note 1, at 105; CLARKE & KNAKE, *supra* note 1, at 99; Ellen Nakashima & Steven Mufson, *Hackers Have Attacked Foreign Utilities, CIA Analyst Says*, WASH. POST, Jan. 19, 2008; Sean Watts, *Combatant Status and Computer Network Attack*, 50 VA. J. INT'L L. 391, 404-05 (2010).

[45] BRENNER, *supra* note 1, at 110; CLARKE & KNAKE, *supra* note 1, at 100, 107; ECONOMIST, *supra* note 3, at 4; Gable, *supra* note 2, at 59-60.

[46] William J. Broad et al., *Israeli Test on Worm Called Crucial in Iran Nuclear Delay*, N.Y. TIMES, Jan. 15, 2011; *see also* BRENNER, *supra* note 1, at 102; Ellen Nakashima, *Homeland Security Tries to Shore up Nation's Cyber Defenses*, WASH POST., Oct. 1, 2011; Kim Zetter, *How Digital Detectives Deciphered Stuxnet, the Most Menacing Malware in History*, WIRED, Jul. 11, 2011.

[47] Bambauer, *supra* note 12, at 585-86; BRENNER, *supra* note 1, at 103; John Markoff, *A Silent Attack, but not a Subtle One*, N.Y. TIMES, Sept. 27, 2010.

The *y* axis depicts the assailants that might commit a cyberattack. They are arranged from bottom to top in order of increasing sophistication. (A sophisticated attacker is one who is capable of compromising the most secure systems; unsophisticated attackers are only capable of compromising relatively unsecured systems.) At the bottom are recreational hackers – the stereotypical teenagers out for "a digital joy ride."[48] One step above are "hacktivists." Hacktivists are relatively skilled hackers who use cyber intrusions to advance a political agenda; they typically do not group themselves into formal organizations.[49] An example is "Anonymous," a group that launched DDOS attacks on financial institutions that stopped allowing customers to send money to WikiLeaks, an anti-secrecy group that had published a number of classified documents.[50] Next are organized crime syndicates, such as those operating out of Russia.[51] They, too, are fairly sophisticated; they engage in cyber intrusions primarily for financial gain; and by definition they are structured organizations.[52] (International terrorists might be placed here as well, though they have shown little enthusiasm or aptitude for cyberattacks thus far.[53] On the other hand, al Qaeda reportedly established an "academy of cyberterrorism" in Afghanistan,[54] and computers taken from members contained information about SCADA systems in the United States.[55]) At the top are foreign governments' militaries and intelligence services. These are the most sophisticated adversaries of all and they are capable of breaking into even highly secure systems. Internet giant Google recently saw its Gmail service penetrated by Chinese spies who wanted to eavesdrop on the Dalai Lama.[56] Similarly, RSA – a software firm that issues online security credentials for the Pentagon, defense contractors, and other sensitive enterprises – was compromised so badly (probably by China) that it had to offer new credentials to all its customers.[57]

The curve roughly predicts the combinations of victims and attackers that are likely to occur. Quadrant (4) involves high frequency, low severity attacks. Retailers and other relatively insignificant firms can expect to be targeted fairly often by relatively unsophisticated recreational hackers and by more sophisticated hacktivists. Quadrant (2) involves attacks that are low frequency and high severity. More strategically significant firms like ISPs and public utilities will face attacks from sophisticated militaries and intelligence services, and perhaps from organized crime syndicates. These attacks will only occur rarely, but they are likely to be

[48] Dunlap, *supra* note 12, at 358.

[49] Byron Acohido, *Cyberattacks Likely to Escalate this Year,* USA TODAY, Jan. 10, 2012.

[50] Somini Sengupta, *16 People Arrested in Wave of Attacks on Web Sites,* N.Y. TIMES, Jul. 20, 2011.

[51] Brian Krebs, *Shadowy Russian Firm Seen as Conduit for Cybercrime,* WASH. POST, Oct. 13, 2007.

[52] BRENNER, *supra* note 1, at 7, 25.

[53] Condron, *supra* note 19, at 405; Dunlap, *supra* note 12, at 359-60.

[54] Joel P. Trachtman, *Global Cyberterrorism, Jurisdiction, and International Organization*, *in* Grady & Parisi, *supra* note 19, at 260-61.

[55] BRENNER, *supra* note 1, at 106.

[56] BAKER, *supra* note 24, at 208-13; BRENNER, *supra* note 1, at 46-47; Ellen Nakashima, *Google to Enlist NSA to Help It Ward off Cyberattacks*, WASH POST., Feb. 4, 2010; Rosenzweig, *supra* note 14, at 6.

[57] Baker, *supra* note 29, at 2-3.

devastating. In quadrant (3), recreational hackers and hacktivists might launch attacks against utilities and similarly significant enterprises, but these targets are probably less attractive to them than they are to foreign militaries or intelligence services.[58] In quadrant (1), foreign governments are unlikely to target insignificant firms like retailers, because they gain little by compromising them, though organized crime may do so.

We are now in a position to make predictions about various companies' cybersecurity expenditures. The closer we are on the curve to the lower left corner, the higher the probability that the firm is investing a socially optimal amount in cyberdefense. This is so in part because the expected social cost of an attack on an ordinary company is fairly low. Society will not grind to a halt if Amazon.com is knocked offline; bookworms might experience minor annoyance but they will still be able to buy a copy of *Macbeth* from Barnes & Noble. In addition, these companies are unlikely to face attacks by skilled and determined foreign governments, so it is not necessary for them to spend huge sums of money on the very best and most impregnable defenses. The efficient level of cybersecurity investment for them thus is fairly low. Importantly, market forces may provide these firms with meaningful incentives to protect their systems against cyberattacks. Retailers, banks, and similar companies operate in competitive markets. The risk of customer exit provides them with strong incentives to cater to customer demand. If consumers want the companies with which they do business to provide better security against cyberattacks – the jury is out on that question, incidentally[59] – they will have good reason do so. (Note that current liability rules both diminish and augment these incentives. The federal wiretap act makes it a crime to intercept electronic communications, and some ISPs fear that this prohibition prevents them from filtering botnet traffic or other malware; the threat of liability undermines their incentives to improve the security of their systems.[60] By contrast, the Gramm-Leach-Bliley Act requires banks, on pain of significant money damages, to protect customer data against unauthorized access; the threat of liability amplifies their incentives to improve the security of their systems.[61])

The closer we are on the curve to the upper right corner, the lower the probability that the firm is adequately investing in cybersecurity.[62] Quadrant (2) – low frequency, high severity – is the opposite of quadrant (4). First, the expected social cost of a cyberattack is monumental. The consequences of an attack on, say, the power grid, would reverberate throughout the economy, causing harm to the utility and its customers and also to third parties with which the company has no contractual relationships. Because the expected cost of an attack on these firms is so high, it is efficient to invest greater sums in securing them against intruders. In addition, the modest (and low cost) defenses that are usually capable of thwarting recreational hackers will do nothing to prevent intrusions by foreign governments; more expensive countermeasures are

[58] Zetter, *supra* note 46 ("[C]ontrol systems aren't a traditional hacker target, because there's no obvious financial gain in hacking them . . . .").

[59] *Compare* BRENNER, *supra* note 1, at 225, 226; *and* Paul M. Schwartz & Edward J. Janger, *Notification of Data Security Breaches*, 105 MICH. L. REV. 913, 946-47 (2007); *with* Dunlap, *supra* note 12, at 361; *and* Doug Lichtman & Eric P. Posner, *Holding Internet Service Providers Accountable*, *in* Grady & Parisi, *supra* note 19, at 256.

[60] *See infra* notes 194 to 201 and accompanying text.

[61] *See infra* notes 202 to 210 and accompanying text.

[62] Sarnikar & Johnsen, *supra* note 19, at 17, 22-23.

needed to protect against these exceptionally sophisticated adversaries. The socially optimal level of cybersecurity investment for these firms thus is fairly high. Second, power companies and other utilities are not subject to market forces that might incentivize them to improve their cyberdefenses. Utilities face little if any competition; a given customer typically will be served by only one power company. Customer exit is essentially impossible, and the utility therefore has weaker incentives to supply what its customers are demanding. (This absence of favorable market forces may help explain why public utilities often fail to implement even relatively costless security measures.[63] Many electric companies use vendor default passwords to protect their SCADA systems,[64] and a recent study found that they take an average of 331 days to implement security patches for these systems.[65] Perhaps not coincidentally, hackers – most likely Chinese and Russian spies – have been able to insert logic bombs into the power grid.[66])

If this analysis is correct, then strategically significant firms in uncompetitive markets are less likely to adequately invest in cybersecurity than ordinary firms in competitive markets. The question then becomes who should be responsible for securing these most sensitive companies against the most dangerous adversaries. Economists often argue that risk should be allocated to the low cost avoider.[67] If the government can reduce a vulnerability more efficiently than a firm, it should pay; if the firm can reduce the vulnerability more efficiently, it should pay. There is no single low cost avoider in this context. Defending critical infrastructure against sophisticated cyberattackers is a task that features dueling comparative advantages. Private firms typically know more than outsiders, including the government, about the architecture of their systems, so they often are in a better position to know about weaknesses that intruders might exploit.[68] The private sector thus has a comparative advantage at identifying cyber vulnerabilities. On the other hand, the government's highly skilled intelligence agencies typically know more than the private sector about malware used by foreign governments and how to defeat it.[69] The government thus has a comparative advantage at detecting sophisticated attacks and developing countermeasures. This suggests that responsibility for defending the most sensitive systems against the most sophisticated adversaries should be shared.

What might such a partnership look like? All private firms might be asked to provide a baseline level of cybersecurity – modestly effective (and modestly expensive) defenses that are capable of thwarting intrusions by adversaries of low to medium sophistication. The government then would assume responsibility for defending public utilities and other sensitive enterprises

[63] Availability bias is another reason why firms might tend to underinvest in cyberdefense. The United States has not experienced a major cyber incident that has captured the public's imagination, so firms might irrationally discount the probability that they will suffer a catastrophic attack. John Grant, *Will There Be Cybersecurity Legislation?*, 4 J. NAT'L SEC. L & POL'Y 103, 111 (2010); McAfee 2010, *supra* note 35, at 14.

[64] McAfee 2011, *supra* note 15, at 8.

[65] BRENNER, *supra* note 1, at 98.

[66] Siobhan Gorman, *Electricity Grid in U.S. Penetrated by Spies*, WALL ST. J., Apr. 8, 2009.

[67] Katyal, *Criminal Law*, *supra* note 10, at 1095-96; LAWRENCE LESSIG, CODE 2.0 at 169-17 (2006).

[68] *See infra* notes 263 to 265 and accompanying text.

[69] *See infra* notes 266 to 268 and accompanying text.

against catastrophic attacks by foreign militaries and other highly sophisticated adversaries.[70] This arrangement – basic security provided by firms, supplemental security provided by the government – is in a sense the opposite of what we see in realspace criminal law. In realspace, the government offers all citizens a baseline level of protection against criminals (in the form of police officers, prosecutors, and courts). Individuals may supplement these protections at their own expense, such as by installing alarm systems in their homes or hiring private security guards.[71] This arrangement also is consistent with our intuitions about the respective roles of government and the private sector. Consider another realspace analogy. In World War II, factories weren't expected to install antiaircraft batteries to defend themselves against Luftwaffe bombers.[72] Nor would we expect power plants to defend themselves against foreign governments' cyberattacks. Protecting vital national assets from destruction by foreign militaries is a quintessential, perhaps *the* quintessential, government function.[73]

The division of labor I am suggesting also seems sound from an economic standpoint. If a firm invested in extraordinarily expensive cyberdefenses capable of thwarting doomsday attacks by China's intelligence service and Russia's military, it would effectively be subsidizing the rest of the population. The company would capture some benefits of increased security, but a large portion of the benefits would be in the form of a positive externality conferred on others.[74] In other words, the firm would be providing a public good (a good that is both non-rivalrous and non-excludable).[75] Economic theory predicts that public goods will be underprovided on the market; a standard response is to subsidize them. So the government might provide a sensitive enterprise with a subsidy equal in value to its costs of defending against the most sophisticated cyberattackers.[76] This subsidy could take many forms. The government could either pay for the firm's defenses directly or reimburse it for its cybersecurity expenditures. Or the company could be offered various tax credits, deductions, and other benefits. Or it could be granted immunity from certain forms of legal liability. (In that case, the subsidy would not run from society as a whole, but from those who were injured by the firm's otherwise unlawful conduct and whose entitlement to redress has been extinguished. This sort of subsidy is potentially regressive.) Or the government might provide the company with intelligence about the types of attacks it is likely to face. (This sort of subsidy appears to be occurring already. The NSA reportedly is providing malware signature files to Google and certain banks to help them detect sophisticated intrusions into their systems.[77])

---

[70] Rabkin & Rabkin, at 4; Trachtman, *supra* note 54, at 272.

[71] Rosenzweig, *supra* note 14, at 20.

[72] CLARKE & KNAKE, *supra* note 1, at 144; Rosenzweig, *supra* note 14, at 5-26.

[73] BRENNER, *supra* note 1, at 223; CSIS, *supra* note 8, at 15; Katyal, *Digital Architecture*, *supra* note 15, at 2282.

[74] Sarnikar & Johnsen, *supra* note 19, at 17.

[75] *See infra* notes 129 to 135 and accompanying text.

[76] Amitai Aviram, *Network Responses to Network Threats*, *in* Grady & Parisi, *supra* note 19, at 149, 156; Bambauer, *supra* note 12, at 658; CLARKE & KNAKE, *supra* note 1, at 113-14; Rosenzweig, *supra* note 14, at 10; Sarnikar & Johnsen, *supra* note 19, at 22-23.

[77] *See infra* notes 267 to 268 and accompanying text.

II. CYBERSECURITY FRAMEWORKS, CONVENTIONAL AND UNCONVENTIONAL

The vast majority of academic commentary regards cybersecurity as a problem of the criminal law or the law of armed conflict.[78] The problem is not that these conventional approaches are mistaken. The problem is that they are incomplete. Treating cybersecurity as a matter for law enforcement or the military brings certain challenges into sharper focus. But it tends to obscure others.

Cybersecurity is beset by externalities; it's "externalities galore."[79] An externality is "an effect on the market the source of which is external to the market"[80]; it occurs when an actor's conduct results in the imposition of a cost or benefit on a nonconsenting third party. Externalities can be either positive or negative. "Positive externalities occur whenever an activity generates benefits that the actor is unable to internalize," such as through prices; "[n]egative externalities occur when one's activity imposes costs on others" that likewise are not transmitted through prices.[81] Economic theory predicts that the market will oversupply negative externalities, relative to socially optimal levels, "because the producer will internalize all the benefits of the activity but not all of the costs."[82] It also predicts that the market will undersupply positive externalities because third parties will free ride. Externalities thus represent a form of market failure.[83] The standard government response to a negative externality is to discourage the responsible conduct (as with taxation or regulation); the standard response to a positive externality is to encourage the responsible conduct (as with a subsidy).[84]

Cybersecurity can be understood in these terms. If a company suffers an intrusion, much of the harm will fall on remote third parties; the attack results in a negative externality.[85] It can be extraordinarily difficult to internalize these costs. The class of persons affected by the intrusion is likely to be so large that it would be prohibitively expensive to use market exchanges to internalize the resulting externalities; the transaction costs are simply too great. Nor can tort law internalize the costs, as firms generally do not face liability for harms that result from cyberattacks on their systems or products. Because many companies do not bear these externalized costs, they ignore them when deciding how much to spend on cyberdefense. They therefore tend to underinvest relative to socially optimal levels. (This is true both of companies that produce computer products, such as software manufacturers, and companies that use them,

---

[78] *See* sources cited *supra* note 19.

[79] Picker, *supra* note 42, at 115.

[80] Niva Elkin-Koren & Eli M. Salzberger, *Law and Economics in Cyberspace*, 19 INT'L REV. L. & ECON. 553, 563 (1999).

[81] Elkin-Koren & Salzberger, *supra* note 80, at 563.

[82] Coyne & Leeson, *supra* note 18, at 479.

[83] Coyne & Leeson, *supra* note 18, at 479; Timothy F. Malloy, *Regulating by Incentives*, 80 TEX. L. REV. 531 534-35 n.13 (2002).

[84] Coyne & Leeson, *supra* note 18, at 479; Rosenzweig, *supra* note 14, at 10.

[85] *See infra* notes 118 to 124 and accompanying text.

such as ISPs and electric companies.) Cyberattacks also involve positive externalities.[86] A company that secures itself against intruders makes it harder for assailants to use its systems to attack others. Investments in cyberdefense thus effectively subsidize other firms. Because the investing company doesn't capture the full benefit of its expenditures, it has weaker incentives to secure its systems. And because other companies are able to free ride on the investing firm's expenditures, they have weaker incentives to adopt defenses of their own.

These externality and free rider problems are largely overlooked by the conventional approaches to cybersecurity, but they can be illuminated if we consult alternative regulatory frameworks – frameworks like environmental law, antitrust law, products liability law, and public health law. In short, a wider selection of analytical lenses is needed to fully comprehend cybersecurity challenges in all their complexity.

A. The Conventional Approaches: Law Enforcement and Armed Conflict

Scholars typically use two analytical frameworks to understand cyberattacks: criminal law and the law of armed conflict.[87] Consider the former first. Broadly speaking, the criminal law seeks to protect members of society from unjustified acts of violence against their persons or property. The criminal law pursues this objective by imposing sanctions, such as incarceration, on those adjudged to have violated the law. These penalties, it is alternatively said, will either punish those who have transgressed society's moral code (retribution), or dissuade the perpetrator or others from committing similar offenses in the future (specific or general deterrence), or isolate the dangerous perpetrator from society (incapacitation), or teach the misguided perpetrator the error of his ways (rehabilitation). Cyberattacks fit into this conceptual framework fairly comfortably. A person who hacks into another's computer may have thereby violated any number of laws, such as the federal Computer Fraud and Abuse Act.[88] Society regards this sort of conduct as sufficiently blameworthy that it proscribes it and subjects those who engage in it to criminal penalties of varying severity.

Scholars who approach cybersecurity from a law enforcement perspective focus on the "whodunit" questions. Who was responsible for launching this particular attack? Was it an individual hacker or a larger criminal enterprise? The law enforcement framework also emphasizes jurisdictional questions.[89] Which court (or courts) properly may exercise subject matter jurisdiction over a given cyberattack?[90] State courts, federal courts, or perhaps an international tribunal? Should jurisdiction be determined by the location of the target? By the location of the attacker? By the location in which the effects of the attack are felt? Should cyberattacks be subject to universal jurisdiction – the notion that a court may try certain crimes

[86] *See infra* notes 126 to 135 and accompanying text.

[87] *See* sources cited *supra* note 19.

[88] 18 U.S.C. § 1030.

[89] Gable, *supra* note 2, at 99-117

[90] Jack L. Goldsmith, *Against Cyberanarchy*, 65 U. CHI. L. REV. 1199, 1200-01 (1998); DAVID G. POST, IN SEARCH OF JEFFERSON'S MOOSE 163-71 (2009).

regardless of where in the world they occurred?[91] How might courts gain personal jurisdiction over those suspected of committing the attack, especially if they are overseas? Do existing extradition treaties cover the range of offenses that cybercriminals might commit? Should the United States negotiate new bilateral agreements with key international partners (such as our European allies), or with countries in which cyberattacks are likely to originate (such as China and Russia)? Or should there be a multilateral global convention on cybercrime, one that will facilitate extradition of suspects from their home countries to the states in which they will stand trial for their alleged crimes?

The law enforcement framework also emphasizes punishment and deterrence.[92] Certain economic theories of criminal law posit that a person's willingness to commit crimes is a function of the expected penalty for that activity – i.e., the sanction for particular offense discounted by the probability he will get caught.[93] The greater the sanction, and the greater the likelihood of detection and punishment, the less likely a person will choose to commit that crime. The question then becomes what should be done to increase the deterrent effect of laws that proscribe various cyber intrusions? Should the penalties for violating these statutes be increased? To what level? Should society invest more resources in detecting cyber crime, thereby increasing the probability that perpetrators will be caught and punished? Or should lawmakers pursue "cost deterrence,"[94] the objective of which is to increase the costs one must incur to perpetrate cybercrime?

The second conventional approach is to regard cyberattacks from the standpoint of the law of armed conflict (LOAC). The LOAC, also known as international humanitarian law (IHL), is a body of international law that regulates a state's ability to use force in several ways. First, it sets forth the circumstances in which a state lawfully may engage in armed conflict – the *jus ad bellum* regulations. For instance, the United Nations Charter forbids signatories "from the threat or use of force against the territorial integrity or political independence of any state"[95] but also recognizes an inherent right to use force in self defense: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs . . . ."[96] Second, the LOAC regulates what kinds of force may be used during an authorized armed conflict – the *jus in bello* regulations. For instance, a state may not deliberately kill civilians or destroy civilian infrastructure (the *distinction* or *discrimination* principle), may not inadvertently inflict harm on civilian populations and structures that is disproportionate to the importance of

[91] *See generally* Eugene Kontorovich, *The Piracy Analogy: Modern Universal Jurisdiction's Hollow Foundation*, 45 HARV. INT'L L.J. 183, 190-92 (2004).

[92] ABA, *supra* note 18, at 13; Gable, *supra* note 2, at 65; Katyal, *Criminal Law*, *supra* note 10, at 1006, 1011, 1040; O'Neill, *supra* note 19, at 265-68; K.A. Taipale, *Cyber-Deterrence* 18 (Apr. 2010).

[93] *See generally* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968); George J. Stigler, *The Optimum Enforcement of Laws*, 78 J. POL. ECON. 526 (1970).

[94] Katyal, *Criminal Law*, *supra* note 10, at 1012; *see also* O'Neill, *supra* note 19, at 265-88.

[95] U.N. Charter Art. 2.

[96] U.N. Charter Art. 51.

the military objective (*proportionality*), and may not cause more harm to legitimate targets than is needed to achieve the military objective (*necessity*).[97]

Cybersecurity is often described in LOAC terms. Scholars who see cybersecurity as an armed conflict problem focus on determining who was responsible for a particular attack.[98] Was this attack launched by a state or an international terrorist organization, in which case the LOAC would probably permit some form of military retaliation? Or was it carried out by criminals, in which case the distinction principle likely would rule out a military response? If the attacker was in fact a state or terrorist group, which one? Was it China, or maybe Russia, or perhaps North Korea? Or was it al Qaeda, or al Qaeda in the Arabian Peninsula, or Hezbollah? Until the identity of the assailant it known, it will be unclear against whom to retaliate – or even whether retaliation is lawful at all.[99]

Another set of important questions concerns how to characterize a cyber incident. Is a given intrusion an act of espionage or an attack? It can be quite difficult to answer that question, because the steps an intruder would take to steal information often are identical to the steps it would take to bring down a system. If the intrusion is properly understood as an attack, does it rise to the level of an "armed attack" that triggers the right of self defense?[100] Should these questions be resolved with an "instrument based" test (a cyber intrusion counts as an armed attack when it causes harms that previously could have been caused only by a kinetic attack[101])? Or a less demanding "effects" or "consequence based" test (a cyber intrusion counts as an armed attack when it has a sufficiently harmful effect on the targeted state[102])? Or an even less demanding "intent" test (a cyber intrusion counts as an armed attack whenever it evinces a hostile intent, regardless of whether it causes actual damage[103]). The LOAC approach also emphasizes possible responses. When a nation suffers a cyberattack, is it limited to responding with a cyber intrusion of its own?[104] Or may a victim retaliate by launching a kinetic attack?[105] How severe must the cyberattack be before a kinetic response would be justified?

---

[97] *See generally* Eric A. Posner, *A Theory of the Laws of War*, 70 U. CHI. L. REV. 297, 298-99 (2003); ERIC A. POSNER & ADRIAN VERMEULE, TERROR IN THE BALANCE 261-66 (2007).

[98] Graham, *supra* note 19, at 92; Lin, *supra* note 19, at 77.

[99] Condron, *supra* note 19, at 414.

[100] Condron, *supra* note 19, at 412-13; Graham, *supra* note 19, at 90-92; Jensen, *Computer Attacks*, *supra* note 19, at 221; Lin, *supra* note 19, at 74; Sklerov, *supra* note 19, at 50-59.

[101] Graham, *supra* note 19, at 91; Sklerov, *supra* note 19, at 54.

[102] Michael N. Schmitt, *Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework*, 37 COLUM. J. TRANSNAT'L L. 885, 913-15 (1999); *see also* Graham, *supra* note 19, at 91; Sklerov, *supra* note 19, at 54-55.

[103] WALTER GARY SHARP, SR., CYBERSPACE AND THE USE OF FORCE 129-31 (1999). Some scholars describe the intent test as a form of "strict liability." *See, e.g.*, Graham, *supra* note 19, at 91; Sklerov, *supra* note 19, at 55. This seems incorrect. A strict liability regime imposes liability solely on the basis of the social harm produced by the actor's conduct, without reference to his mens rea. WAYNE R. LAFAVE, CRIMINAL LAW § 5.5, at 288-89 (5th ed. 2010). It would be more accurate to say that the intent test imposes liability solely on the basis of mens rea, without any requirement that the actor's conduct result in social harm.

[104] Condron, *supra* note 19, at 415-16; Graham, *supra* note 19, at 89-90.

[105] Jensen, *Computer Attacks*, *supra* note 19, at 229-30.

Other problems arise from the fact that much of the world's critical infrastructure is "dual use" – it serves a state's civilian population but also is relied upon by the state's political leadership and armed forces.[106] (In the United States, civilian networks carry up to 98 percent of the federal government's communications traffic, including 95 percent of defense-related traffic.[107]) When, if ever, may a combatant direct a cyberattack at an adversary's dual use infrastructure?[108] Finally, the LOAC approach to cybersecurity focuses on deterrence. Given the differences between cyber conflicts and kinetic ones, how can a state dissuade its adversaries from committing cyberattacks? Key differences include the fact that it is difficult to determine who was responsible for a given intrusion, the possibility that a retaliatory cyber strike might end up harming innocent third parties more than the actual assailant, and the fact that different nations are more (or less) dependent on cyber infrastructure and therefore have more (or less) to lose from an exchange of cyber weapons.[109]

A central problem for both the law enforcement and armed conflict approaches to cybersecurity is determining the identity of the assailant. Yet attribution is extraordinarily difficult; the challenges are "staggering"[110] and "[n]o one has come close to solving" them.[111] This is so because of the basic architecture of the internet. The internet's TCP/IP protocol was designed to move packets of data as efficiently as possible; it is utterly unconcerned with who sent them.[112] As such, it is fairly easy for attackers to obscure their true identities by routing their intrusions through a series of dispersed intermediary computers.[113] These attribution difficulties can severely frustrate the law enforcement and armed conflict approaches to cybersecurity.

B. Cybersecurity as an Environmental Law Problem

A principal goal of environmental law is to regulate externalities. Various forms of environmental degradation can be described as negative externalities – i.e., spillover costs that are imposed on third parties and that are not transmitted through prices.[114] A coal fired power plant imposes negative externalities on those who live downwind when, as a byproduct of

---

[106] Davis Brown, *supra* note 13, at 193-94; CLARKE & KNAKE, *supra* note 1, at 242.

[107] Condron, *supra* note 19, at 407; Jensen, *Computer Attacks*, *supra* note 19, at 211; Jensen, *Cyber Warfare*, *supra* note 15, at 1534.

[108] Davis Brown, *supra* note 13, at 194; CLARKE & KNAKE, *supra* note 1, at 243; Jensen, *Cyber Warfare*, *supra* note 15, at 1543-46.

[109] CSIS, *supra* note 8, at 25-27; Lynn, *supra* note 19, at 99-100; James P. Terry, *Responding to Attacks on critical Computer Infrastructure*, 76 INT'L L. STUD. 421, 432-33 (2002).

[110] Jensen, *Computer Attacks*, *supra* note 19, at 234.

[111] Lin, *supra* note 19, at 77; *see also* Dycus, *supra* note 11, at 163; Katyal, *Criminal Law*, *supra* note 10, at 1047-48; O'Neill, *supra* note 19, at 275.

[112] Bambauer, *supra* note 12, at 595-96; LESSIG, *supra* note 67, at 44.

[113] BRENNER, *supra* note 1, at 32; Gable, *supra* note 2, at 101; Graham, *supra* note 19, at 92; Ruth G. Wedgwood, *Proportionality, Cyberwar, and the Law of War*, 76 INT'L L. STUD. 219, 227 (2002).

[114] *See supra* notes 79 to 84 and accompanying text.

productive activity, it emits pollutants that increase the incidence of asthma. Sometimes these externalities are geographic; pollutants emitted by a factory in Ohio might affect residents of New York.[115] Sometimes they are temporal; carbon emissions today might affect the planet's climate for future generations.[116] The critical point is that these costs are borne by people other than those who are responsible for the pollution, and it is usually impossible to use market transactions to internalize the costs onto the polluter. Many scholars therefore believe that regulatory controls are necessary. For instance, Richard Lazarus cites a "need for government regulation because of the spatial and temporal spillovers caused by unrestricted resource exploitation."[117] These controls often take the form of strict limits on the regulated activity, backed by the threat of civil damages or criminal sanctions, though less coercive forms of regulation exist.

Cybersecurity can be understood in these terms. First, consider negative externalities.[118] A given firm – whether it is a company that produces computer products (such as a software manufacturer) or a company that uses them (such as an ISP or electric company) – will not bear the full costs of its cyber insecurities. (By "cyber insecurity," I refer to a firm that suffers a cyberattack after failing to implement defenses capable of defeating the attack.) Instead, some of these costs are borne by third parties; they are partially externalized.[119] Imagine a cyberattack that disables a power plant. The intrusion would harm the utility as well as consumers who buy electricity from it[120] – hospitals, manufacturers, and others. The attack also would harm a number of third parties who have no relationship with the power company – hospital patients, downstream manufacturers in the supply chain, and so on. These "indirect effects of a cyberattack are almost always more important to the attacker than the direct effects."[121] And it would be prohibitively expensive to internalize them through market exchanges; the transaction costs would be staggering, as it is extraordinarily difficult to identify the universe of third parties affected by the intrusion.

The fact that many costs of cyberattack are externalized onto third parties is enormously significant. Some commentators have argued that firms have strong "financial incentives to

[115] *See, e.g.*, Massachusetts v. EPA.

[116] Richard J. Lazarus, *A Different Kind of "Republican Moment" in Environmental Law*, 87 MINN. L. REV. 999, 1000, 1005 (2003).

[117] Lazarus, *supra* note 116, at 1005-06.

[118] Anderson, *supra* note 33, at 1. One potential difference between pollution and cyber insecurity is that pollution is a harmful byproduct of socially beneficial activity (such as manufacturing) whereas cyberattacks involve intentionally malicious conduct. Rattray et al., *supra* note 8, at 171. Yet cyber intrusions likewise may be seen as a harmful byproduct of beneficial activity. A cyberattack on a computer is a byproduct of the computer being connected to the internet. And connecting a computer to the internet is socially beneficial because it produces network effects; by joining the network, the user increases its value to all users. POST, *supra* note 90, at 47-49.

[119] ABA, *supra* note 18, at 8; Anderson, *supra* note 33, at 1; Jim Harper, *Government-Run Cyber Security? No Thanks* 1 (Mar. 13, 2009); Rosenzweig, *supra* note 14, at 9-10; Schwartz & Janger, *supra* note 59, at 928.

[120] Aviram, *supra* note 76, at 155; Lin, *supra* note 19, at 68.

[121] Lin, *supra* note 19, at 68.

protect [their systems] from cyberattacks."[122] Those incentives are weaker than might be supposed. A firm that is deciding how much to invest in securing its systems will not account for the costs that an attack will impose on third parties.[123] Firms tend to oversupply pollution, since they capture all the benefits of the associated productive activity but not all of the resulting costs. In a similar way, firms tend to oversupply cyber insecurity – or, to say the same thing, they tend to undersupply cyberdefense – because they internalize all of the benefits but only some of the costs.[124] Many firms thus tend to invest less in cyberdefense than would be optimal from a societal standpoint.

The point can be illustrated with a simple hypothetical. Imagine a cyberattack that will result in $1 million in expected costs for the target firm and $10 million in expected costs for third parties. From a societal standpoint, it would be worthwhile to invest up to $11 million to prevent the attack – the sum of the expected harms to the firm and third party victims. But from the company's standpoint, it would only be worthwhile to invest up to $1 million to prevent the attack. If the firm spent more than that, the cost to it of the precautions would exceed the benefit to it, and the firm would be conferring uncompensated benefits on third parties. The firm effectively would be providing a security subsidy. Thus, there is a gap between the welfare of the company and the welfare of society as a whole. Levels of cybersecurity investment that are efficient for particular firms may turn out to be inefficient for society at large.[125]

Cybersecurity also can be understood as a positive externality. When a firm expends resources to defend itself against intruders, that investment makes other users' systems marginally more secure as well. This is so because the defenses not only help prevent harm to the company's system, they also help prevent the firm's system from being used to inflict harm on others' systems.[126] If Pepsi's network is well defended, it is less likely to be infected by a worm and thus less likely to transmit the malware to Coke. The effect is to decrease the overall incidence of infection on the internet, but the investing firm does not capture the full benefit. A classic positive externality. Cyberdefenses can differ from realspace defenses in this respect. If I install an alarm in my home, that might prevent burglars from breaking into my house, but it won't necessarily decrease the overall incidence of burglary. The alarm might simply displace the burglar who would have targeted me onto my neighbor[127] – a form of negative externality. By contrast, cyberdefenses can make my system more secure at the same time they increase the overall security of the internet.[128]

---

[122] Nojeim, *supra* note 14, at 134; *see also* Coldebella & White, *supra* note 14, at 236, 241; Dunlap, *supra* note 12, at 361; Yang & Hoffstadt, *supra* note 15, at 203.

[123] ABA, *supra* note 18, at 8; Coyne & Leeson, *supra* note 18, at 479; Rosenzweig, *supra* note 14, at 9-10.

[124] Coyne & Leeson, *supra* note 18, at 480.

[125] Sarnikar & Johnsen, *supra* note 19, at __.

[126] Katyal, *Criminal Law*, *supra* note 10, at 1081-82; O'Neill, *supra* note 19, at 278; Rosenzweig, *supra* note 14, at 9; Sarnikar & Johnsen, *supra* note 19, at 15-16.

[127] Neal Kumar Katyal, *Community Self-Help*, 1 J.L. ECON. & POL'Y 33, 46 (2005); Katyal, *Criminal Law*, *supra* note 10, at 1081; O'Neill, *supra* note 19, at 278.

[128] *But see* Kobayashi, *supra* note 33, at 16; Rosenzweig, *supra* note 14, at 9.

Relatedly, some aspects of cybersecurity resemble public goods.[129] A public good is both non-rivalrous (one person's use of the good doesn't reduce its availability for use by others) and non-excludable (the owner of the good can't prevent particular persons from using it).[130] A classic example of a public good is a large municipal park – open to all comers, and one person enjoying a crisp fall afternoon on a park bench (generally) doesn't prevent anyone else from doing the same. Some scholars argue that cybersecurity information is a public good – e.g., information about the vulnerability of particular system, or the most effective way to counter a particular cyberthreat – that the market will tend to underproduce.[131] There is also a sense in which defensive measures themselves are public goods. Like a municipal park, cyberdefenses can be non-rivalrous.[132] When Pepsi expends resources to secure its computer network, that doesn't decrease the amount of security available for Coke; it actually can *increase* security for third parties. Cyberdefenses also can be non-excludable.[133] When Pepsi secures its system against conscription into a botnet, it isn't possible to specify which third parties will enjoy the benefit of Pepsi's immunity – Coke, but not Snapple. *All* such users are thereby protected from a DDOS attack launched from Pepsi's system.

Understanding cybersecurity in terms of positive externalities and public goods can help explain why many firms underinvest in defense. It's a free rider problem.[134] A company that decides to better secure its computer systems thereby produces benefits that accrue to a number of third parties, and it is impossible to exclude them from receiving those benefits. The investing firm therefore has a weaker incentive to expend resources on cyberdefenses because such expenditures effectively subsidize other firms, including its competitors. Third parties likewise have weaker incentives to secure their own systems; they would prefer to free ride on a rival firm's investment in, say, anti-spyware software than to purchase the product on their own. The overall effect is to weaken the incentive of all firms to invest in protecting their networks. "The individual undertaking the security precautions does not internalize all the benefits, and will seek to free-ride off the efforts taken by others"; as a result, "theory predicts that security will be undersupplied on the market."[135]

---

[129] CSIS, *supra* note 8, at 50; Kobayashi, *supra* note 33, at 15; Powell, *supra* note 14, at 498.

[130] Elkin-Koren & Salzberger, *supra* note 80, at 559; James Grimmelman, *The Internet Is a Semicommons*, 78 FORD. L. REV. 2799, 2806 (2010); Rosenzweig, *supra* note 14, at 8-9; *see also* Harold Demsetz, *The Private Production of Public Goods*, 13 J.L. & ECON. 293, 295 (1970) (distinguishing between non-rivalrous goods, which are properly characterized as public goods, and non-exclusive goods, which are properly characterized as "collective goods").

[131] Kobayashi, *supra* note 33, at 16; Rosenzweig, *supra* note 14, at 9. *But see* Amitai Aviram & Avishalom Tor, *Overcoming Impediments to Information Sharing*, 55 ALA. L. REV. 231, 234-35 (2004) (arguing that information can be a rivalrous good, insofar as sharing it can cause a firm to "los[e] a competitive edge over rivals that benefit from the information").

[132] Kobayashi, *supra* note 33, at 20-21; Trachtman, *supra* note 54, at 270.

[133] Trachtman, *supra* note 54, at 270.

[134] Aviram & Tor, *supra* note 131, at 238; CSIS, *supra* note 8, at 50; Elkin-Koren & Salzberger, *supra* note 80, at 559; Sarnikar & Johnsen, *supra* note 19, at 16; Trachtman, *supra* note 54, at 281. *But see* Powell, *supra* note 14, at 504-05.

[135] Coyne & Leeson, *supra* note 18, at 480; *see also* Sarnikar & Johnsen, *supra* note 19, at 16.

Environmental law and the underlying economic principles it reflects thus provide an important framework through which we might better understand the problem of cybersecurity. Firms tend to underinvest in cyberdefenses for the same reason they tend to underinvest in pollution control technologies – because insecurities that result in successful cyberattacks produce negative externalities that are borne by third parties. Firms also tend to underinvest in cyberdefenses because such expenditures create positive externalities and provide opportunities for free riding. Understood in these terms, the challenge for a cybersecurity regime is to internalize the externalities – to ensure that firms that impose negative externalities by failing to secure their systems are made to bear the resulting costs.

C. . . . as an Antitrust Problem

The ultimate goal of antitrust law, promoting consumer welfare, is achieved by restraining businesses from engaging in anticompetitive conduct. Antitrust law is especially concerned about the possibility that firms will take coordinated action that undermines competition – an agreement by firms to divide a market among themselves, for instance. Antitrust also is apprehensive about information sharing among competitor firms; such exchanges, it is feared, can "facilitate anti-competitive collusion or unilateral oligopolistic behavior."[136] Hence section 1 of the Sherman Act sweepingly prohibits "[e]very contract, combination in the form of trust or otherwise, or conspiracy, in restraint of trade or commerce among the several States."[137]

Antitrust law often subjects coordinated conduct by multiple firms to stricter scrutiny than isolated conduct by a single firm; "the legal presumption against" joint arrangements "is sometimes thought to be strong."[138] Many such arrangements – namely, coordinated action that can be characterized as a "naked" restraint, or a restraint that "is formed with the objectively intended purpose or likely effect of increasing price or decreasing output in the short run"[139] – are condemned under a "per se rule."[140] With a per se rule, there is no need to inquire whether a particular arrangement actually has anticompetitive effects. Instead, antitrust law takes a shortcut and presumes that the conduct is harmful.[141] This approach may lead to the occasional false positive – coordinated action that is actually beneficial to consumers but that nevertheless is condemned as unlawful. But the conventional wisdom is that the costs of these false positives would be dwarfed by the decision costs of distinguishing the small number of naked restraints that are procompetitive from the much larger number that are anticompetitive.

Yet some inter-firm cooperation is beneficial to consumers,[142] and antitrust law can struggle to determine whether a given instance of joint action is pro- or anticompetitive.[143] In the

---

[136] Aviram & Tor, *supra* note 131, at 236.

[137] 15 U.S.C. § 1.

[138] HOVENKAMP § 5.1, at 211; *see also id.* § 5.1b, at 214-16.

[139] HOVENKAMP § 5.1a, at 212.

[140] HOVENKAMP § 5.1, at 211.

[141] HOVENKAMP § 5.1, at 211-12.

[142] Aviram & Tor, *supra* note 131, at 231; HOVENKAMP § 5.1, at 211.

cybersecurity context, various forms of coordination and information sharing can help firms better defend themselves against intrusions and thus prevent consumers from incurring losses. Firms in a particular industry might agree to exchange threat information.[144] An ISP that discovers it has been victimized by a particular form of malware could alert others to be on the lookout for the same threat. Or firms could share vulnerability information.[145] A power plant that learns that its SCADA system can be compromised by a particular type of intrusion could tell other companies about the vulnerability. Firms also might share countermeasure information.[146] A company might discover that a particular security solution is an especially effective way to defend against, say, a DDOS attack, and the company might notify other firms to use the same technique. Finally, firms might agree to establish a uniform set of cybersecurity standards for their industry, along with a monitoring and enforcement mechanism to ensure that all members are implementing the agreed upon measures. They might, in other words, form something like a cartel.

Which brings us to the problem. Coordinating on cyberdefense could give rise to antitrust liability, and firms therefore are reluctant to share information with their competitors or to adopt common security standards.[147] These liability fears appear to be fairly widespread. A 2002 analysis found that, among the private sector's "major concerns about fully communicating cybervulnerabilities," one of the most important is "the potential for antitrust action against cooperating companies."[148] In a 2009 report, the American Bar Association likewise recounted the concerns of several firms that "antitrust laws created a barrier to some forms of sharing" cybersecurity information.[149] Government officials have reported the same fears. The White House's 2009 Cyberspace Policy Review acknowledged that some inter-firm coordination takes place, but went on to report that "some in industry are concerned that the information sharing and collective planning that occurs among members of the same sector under existing partnership models might be viewed as 'collusive' or contrary to laws forbidding restraints on trade."[150]

---

[143] Aviram & Tor, *supra* note 131, at 236; HOVENKAMP § 5.1, at 211.

[144] Emily Frye, *The Tragedy of the Cybercommons*, 58 BUS. LAW. 349, 369 (2002); Lichtman & Posner, *supra* note 59, at 236.

[145] Aviram & Tor, *supra* note 131, at 263.

[146] Kobayashi, *supra* note 33, at 23.

[147] *Cf.* Jonathan H. Adler, *Conservation through Collusion: Antitrust as an Obstacle to Marine Resource Conservation*, 61 WASH. & LEE L. REV. 3 (2004) (arguing that antitrust regulation discourages cooperative inter-firm efforts to control pollution).

[148] Frye, *supra* note 144, at 374. The other two concerns are "an increased risk of liability" and the "loss of proprietary information." *Id*.

[149] ABA, *supra* note 18, at 10.

[150] *Cyberspace Policy Review* 18-19 (2009). *But see* BRENNER, *supra* note 1, at 228 (dismissing the possibility that cybersecurity coordination might give rise to antitrust liability); Rosenzweig, *supra* note 14, at 16 (same). Cybersecurity experts sometimes exchange information about threats and vulnerabilities notwithstanding the antitrust laws. For instance, an informal collaboration between researchers at Symantec, the computer security company, and several freelance computer experts in Europe revealed that Stuxnet, originally thought to be a "routine and unambitious" piece of malware, was in fact a sophisticated cyberweapon aimed at Iran's nuclear program. Zetter, *supra* note 46. This episode is important for two reasons. First, it confirms that information sharing can produce significant cybersecurity gains. Second, it suggests that information sharing is more likely to take place

These concerns seem well founded. There are a number of scenarios in which cybersecurity coordination conceivably could trigger liability under federal antitrust statutes. For instance, suppose that firms in a particular industry agree to implement a uniform set of cybersecurity practices. It is improbable that these new standards will be costless. Whether the companies have agreed to purchase and install new firewall software, or to transition from vulnerable commercial off the shelf ("COTS") operating systems to more expensive proprietary operating systems, the measures are likely to affect their bottom lines. Industry members might decide to absorb these increased costs, depending on the elasticity of consumer demand for the goods or services they offer. But they might further decide to pass on these costs to consumers, either in the form of a general price hike or as a free standing surcharge.

Would the arrangement be lawful? This sort of venture may well amount to price fixing in violation of the Sherman Act. Even if the participating firms are not setting a specific price for their products (everyone will now charge $50 for widgets instead of $45), they are still establishing a premium that will be assessed for their products (everyone will increase the price they charge for their widgets by $5). The economic effect is the same. Indeed, the arrangement may even amount to a "naked" restraint that triggers review – and therefore probably reflexive condemnation – under the per se rule.[151] The venture also might stand condemned as an unlawful tying arrangement. Tying occurs when a firm requires a consumer to purchase one product as a condition of purchasing another[152]; Canon refuses to sell you a camera unless you also buy a flash. Like naked restraints, tying arrangements typically are reviewed under a per se rule.[153] Transferring the increased costs of cybersecurity to consumers might be seen as an effort to force them to buy a new security product in addition to the firm's basic product. Imagine a bank that previously would have offered financial services, such as the ability to use a credit card, for $45 a year. After the agreement, it now sells financial services *plus* enhanced security for $50 a year. That additional $5 represents the price for a separate product, cybersecurity, which consumers may or may not independently wish to purchase.[154]

As a second example of how shared cybersecurity standards might violate the antitrust laws, consider an arrangement that imposes no new costs on consumers – at least not directly. Suppose firms in a particular industry agree to install intrusion-detection or -prevention capabilities to scan for malware on their networks.[155] These systems rely on a technique known as "deep packet inspection," in which all data traversing the network is scanned and checked against signature files of known malware.[156] The effect is often to slow down the network's

---

where there is little risk of antitrust liability. Symantec and European researchers could freely exchange information because they did not offer competing goods or services, so the arrangement was unlikely to be condemned as a contract, combination, or conspiracy in restraint of trade. 15 U.S.C. § 1.

[151] HOVENKAMP § 5.1a, at 212.

[152] HOVENKAMP § 10.1, at 435.

[153] *But see* Jefferson Parish Hosp. Dist. No. 2 v. Hyde, 466 U.S. 2, 40 (1984) (O'Connor, J., concurring) (arguing that tying arrangements should be reviewed under a rule of reason).

[154] *See* sources cited *supra* note 59.

[155] POST, *supra* note 90, at 85.

[156] CLARKE & KNAKE, *supra* note 1, at161-62; LESSIG, *supra* note 67, at 55-56; Lynn, *supra* note 19, at 103.

performance, sometimes dramatically.[157] Suppose further that the firms decide to absorb the costs of the monitoring or detection system rather than pass them on to their consumers. Would that forbearance save the arrangement from antitrust liability? Not necessarily. The shared security standards still plausibly could be described as an unlawful price fixing agreement. While the participating companies have not agreed to raise prices directly, they have indirectly accomplished something similar; instead of requiring consumers to pay a *higher* price for the *same* product, the firms have agreed to require consumers to pay the *same* price for a *lesser* product (where speed is an important component of the product's value).

Notice that clear and unambiguous prohibitions on inter-firm coordination may not be necessary to deter businesses from cooperating with one another. Mere uncertainty about the applicability of the antitrust laws – and the corresponding risk of liability – may be enough to dissuade firms from participating in joint cybersecurity ventures. The deterrent effect of legal ambiguity is likely to be especially strong in this context because of the sanctions that may be imposed on antitrust defendants. Firms that are alleged to have violated federal antitrust laws face criminal prosecutions[158] as well as federal civil actions,[159] state civil actions,[160] and lawsuits by aggrieved private parties,[161] and each type of civil litigation carries the prospect of treble damages payouts to the successful plaintiffs.[162] In light of these potential sanctions, private firms will have good reasons to avoid coordinating their efforts to improve cybersecurity.

One final observation about cybersecurity and antitrust. Beyond liability concerns, there are other serious impediments to coordination and information sharing. The difficulties of forming and maintaining cartels are well known. Among other problems, individual cartel members have strong incentives to cheat, such as by offering a greater quantity of product or by charging a higher price than allotted by the cartel.[163] In the cybersecurity context, businesses will have comparable incentives to shirk their responsibilities to implement any agreed upon (and likely costly) security standards. In addition, firms may be especially reluctant to share information with their competitors.[164] If a firm discovers an effective way to defend its systems against a particular form of cyber intrusion, that information gives it a comparative advantage over rivals that may not be as adept at protecting their own networks. Sharing the information with competitors enables them to free ride and thereby eliminates the firm's comparative advantage. As such, even if fears of liability under the antitrust laws were eliminated completely, it is doubtful that firms would fully cooperate with one another. Nevertheless, liability concerns appear to be a significant impediment to cybersecurity coordination and

---

[157] CLARKE & KNAKE, *supra* note 1, at 81; Smith, *supra* note 17, at 180.

[158] CITE

[159] 15 U.S.C. § 15a.

[160] 15 U.S.C. § 15c.

[161] 15 U.S.C. § 15.

[162] *Compare* 15 U.S.C. § 15(a) (treble damages in private lawsuits), *with id*. § 15a (treble damages in lawsuits by United States), *with id*. § 15c(a)(2) (treble damages in lawsuits by state attorneys general).

[163] HOVENKAMP § 4.1a, at 161-68.

[164] Aviram & Tor, *supra* note 131, at 234; *see also* Nathan Alexander Sales, *Share and Share Alike*, 78 GEO. WASH. L. REV. 279, 319-20 (2010).

information sharing. Reducing these fears would not by itself ensure cooperation, but doing so would make it more likely at the margin.

D. . . . as a Products Liability Problem

Private investment in cybersecurity also resembles a tort problem – more precisely, a products liability problem. Broadly speaking, the law of products liability has two complementary goals. First, from an ex post perspective, the law seeks to compensate consumers injured by products that did not perform as expected. Second, from an ex ante perspective, products liability law uses the risk of money damages to incentivize firms to take reasonable precautions when designing and manufacturing products.[165]

The branch of products liability law that is most relevant to cybersecurity concerns design defects. In a design defect case, the theory is that "the intended design of the product line itself is inadequate and needlessly dangerous.[166] (By contrast, a manufacturing defect occurs when a product suffers from "a random failing or imperfection,"[167] such as a crack in a Coke bottle that causes it to explode,[168] and a marketing defect occurs when an otherwise safe product "become[s] unreasonably dangerous and defective if no information explains [its] use or warns of [its] dangers."[169]) In its infancy, products liability law typically assigned blame on a theory of strict liability.[170] A plaintiff could recover damages by establishing that a given product had a defective design and that he was injured by that defect; it wasn't necessary to show that the manufacturer was negligent, or otherwise blameworthy, in producing the defect.[171] The modern approach abandons strict liability in favor of a negligence standard.[172] How do courts determine whether a manufacturer was at fault when it produced a product with a design defect? One common approach is the risk-utility test.[173] The test, which has its roots in Learned Hand's negligence formula,[174] compares "the risks of the product as designed against the costs of making the product safer."[175] If the risks associated with a product can be reduced by a significant amount at a relatively low cost, a manufacturer that declines to do so is negligent. If

---

[165] DOBBS § 353, at 975-76; WILLIAM M. LANDES & RICHARD A. POSNER, THE ECONOMIC STRUCTURE OF TORT LAW 4-5 (1987).

[166] DOBBS § 355, at 980; MICHAEL I. KRAUSS, PRINCIPLES OF PRODUCTS LIABILITY 81 (2011).

[167] DOBBS § 355, at 979.

[168] Lee v. Crookson Coca-Cola Bottling Co., 188 N.W.2d 426 (Minn. 1971).

[169] DOBBS § 355, at 981.

[170] *See, e.g.*, Greenman v. Yuba Power Prods., 377 P.2d 897 (Cal. 1963); RESTATEMENT (SECOND) OF TORTS § 402A (1965).

[171] DOBBS § 353, at 974-75.

[172] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § __ (1998); *see also* DOBBS § 353, at 977; KRAUSS, *supra* note 166, at 40; LANDES & POSNER, *supra* note 165, at 292.

[173] RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2, cmts. a & f (1998); *see also* DOBBS § 357, at 985-87; LANDES & POSNER, *supra* note 165, at 291-92.

[174] United States v. Carroll Towing Co. 159 F.2d 169 (2d. Cir. 1947).

[175] DOBBS § 357, at 985.

the risks can be reduced only by a small amount at a relatively high cost, a manufacturer that declines to do so is not negligent.

This system of tort liability creates important incentives for manufacturers to prevent or eliminate design defects.[176] Imagine a company that makes residential furnaces; it is trying to decide whether to remedy a design defect that increases the probability the furnaces will explode. The company will do so if the expected benefits of reducing the risk of explosion exceed the expected costs of making the fix. Without tort liability, the benefit of making defect-free furnaces is lower than it otherwise would be. Furnaces that occasionally explode would damage the firm's reputation, and some consumers likely would buy competitors' products instead. The manufacturer benefits to the extent it reduces these harms. But it doesn't face the prospect of paying money damages to homeowners whose houses burned down. The cost-benefit calculus looks very different once a products liability regime is in place. Now, a decision to eliminate the design defect will be more beneficial. The company thereby reduces its exposure to potentially ruinous money damages awards, at least where the defect can be cured in a sufficiently low-cost way that the failure to do so would be deemed negligent. In short, a products liability regime increases a firm's expected benefit of remedying design defects – namely, the benefit of foregone money damages, discounted by the probability that they would be awarded. It thus increases the number of circumstances in which firms will find it welfare maximizing (benefit > cost) to improve the safety of their products. The result is that, at the margin, products will be safer than they otherwise would be.

Internet-related goods and services sometimes suffer from design defects that increase their vulnerability to cyberattacks.[177] Perhaps the best known example is Microsoft Windows. The operating system software, which accounts for more than 90 percent of the PC market,[178] is notoriously riddled with vulnerabilities.[179] These vulnerabilities stem in part from the software's size. Windows Vista, released in 2007, featured some 50 million lines of code, compared to 35 million for Windows XP (released in 2001) and just 15 million for Windows 95 (released in 1995).[180] It is more or less inevitable that the programmers who write these millions of lines will make mistakes, and it can be quite difficult to detect and repair them.[181] (Given that it probably would cost a great deal to eliminate all of these vulnerabilities, the failure to do so may not be negligent under the risk-utility test.[182]) Other examples abound. Indeed, many of the vulnerabilities described in Part I can be understood as the results of design defects. Consider the decision by power companies to connect generators and other elements of the electrical grid to the internet. This can be seen as a form of defective system design, in that internet

[176] LANDES & POSNER, *supra* note 165, at 10-11.

[177] Lichtman & Posner, *supra* note 59, at 255.

[178] Steve Lohr & John Markoff, *Windows Is so Slow, but Why?*, N.Y. TIMES, Mar. 27, 2006.

[179] *See, e.g.*, __.

[180] Lohr & Markoff, *supra* note 178.

[181] DOROTHY DENNING, INFORMATION WARFARE AND SECURITY 12 (1999); Katyal, *Digital Architecture*, *supra* note 15, at 2264-65; Sarnikar & Johnsen, *supra* note 19, at 18.

[182] *But see* Lichtman & Posner, *supra* note 59, at 255 (arguing that improving the security of Windows is "simply a matter of investing more resources in product design as well as testing").

connectivity exposes the nation's power grid to potentially catastrophic cyberattacks in exchange for relatively minor benefits.[183] The same can be said of companies that continue to protect their SCADA systems with vendor supplied default passwords[184] – a defect, incidentally, that could be remedied at a negligible cost.

The incentives to cure these design defects are fairly weak, because poor cyber security generally does not trigger civil liability. "[L]iability has played virtually no role in achieving greater Internet security."[185] One reason for this is a venerable chestnut of tort law known as the economic loss doctrine. The economic loss doctrine provides that, while a defendant who causes physical injuries is also liable for any resulting economic harms, he generally is not liable for freestanding economic harms. "When commercial or economic harm stands alone, divorced from injury to person or property, courts have not imposed a general duty of reasonable care."[186] Many of the harms that would result from a cyberattack on, say, the power grid or the financial sector would be purely economic in nature. An automobile manufacturer might be unable to run its assembly line because the power is out, or a consumer might default on a loan because he can't make a payment online. Few of these harms would derive from a physical injury, and they therefore would not be actionable under the economic loss doctrine. For instance, in 2009, the Massachusetts Supreme Judicial Court dismissed a lawsuit brought by credit unions against a retailer after hackers accessed the retailer's computer systems and stole customer credit card data. The court emphasized that, because "the plaintiffs suffered only economic harm due to the theft of the credit card account information," the "economic loss doctrine barred recovery on their negligence claims."[187] (Cyberattacks that cause injuries to person or property presumably would remain actionable, as would any resulting economic harms. So, for instance, if an attacker exploited a design defect in a dam's control system and opened the floodgates,[188] the dam operator might be held liable for the deaths of the downstream landowners and any corresponding economic losses.)

The problem can be understood in Coasean terms.[189] Consider the famous example of a train that emits sparks that burn the wheat in neighboring fields. Regardless of whether the legal entitlement is initially assigned to the railroad (a right to emit sparks) or the farmers (a right to be free from incinerated crops), the parties will bargain to reallocate the entitlement to its socially most efficient use (assuming that the transaction costs are sufficiently small). In the

[183] *See supra* notes 42 to 47 and accompanying text.

[184] *See supra* notes 63 to 65 and accompanying text.

[185] BRENNER, *supra* note 1, at 224; *see also* Schnieier, *supra* note 33, at 2.

[186] DOBBS § 452, at 1282; *see also id*. § 452, at 1285-87 (discussing exceptions to the economic loss doctrine); LANDES & POSNER, *supra* note 165, at 251. The rule's familiar rationales are, first, the fact that "financial harm tends to generate other financial harm endlessly and often in many directions" and the corresponding recognition that liability "would be onerous for defendants and burdensome for courts"; and, second, the notion that "contract law is adequate to deal with the problem and also usually more appropriate." DOBBS § 452, at 1283.

[187] Cumis Ins. Soc'y, Inc. v. BJ's Wholesale Club, Inc., 918 N.E.2d 36, 46 (Mass. 2009); *accord* Pennsylvania State Employees Credit Union v. Fifth Third Bank, 398 F. Supp. 2d 317, 330 (M.D. Pa. 2005), *aff'd in part*, Sovereign Bank v. BJ's Wholesale Club, Inc., 533 F.3d 162, 176-78 (3d Cir.2008).

[188] Frye, *supra* note 144, at 350; *see also* Sklerov, *supra* note 19, at 20.

[189] *See generally* Ronald H. Coase, *The Problem of Social Cost*, 3 J. L. & ECON. 1 (1960).

cybersecurity context, the absence of tort liability essentially grants firms a legal right to refrain from taking precautions that would protect third parties from attacks on their systems or products. This may be an efficient allocation of the legal entitlement in some contexts, but not always. In these latter circumstances, companies and third parties theoretically should negotiate and establish a new legal right to be free from harm due to cyber intrusions. But Coasean bargaining over cybersecurity seems unlikely to occur because of the staggering transaction costs: It would be prohibitively expensive, if not impossible, for companies to bargain with everyone who conceivably could be injured by cyberattacks on their systems or products.

Beyond tort, it is doubtful that other sources of law credibly will threaten cybersecurity shirkers with liability. Contract law does not seem well suited to the task. Software manufacturers typically do not offer warranties that their products are secure.[190] Indeed, some do not "sell" software at all. They merely grant a license, and users cannot install the software unless they click a button to accept the terms and conditions of the license – which very often include a limit on the manufacturer's liability.[191] Likewise, federal law extends broad immunity to internet service providers. Section 230 of the Communications Decency Act provides that an ISP will not be "treated as the publisher or speaker of any information" that transits its network.[192] This statute has been interpreted by at least one federal appellate court to foreclose a lawsuit alleging that an ISP negligently failed to prevent malware from being sent over its network.[193] From the standpoint of a profit-maximizing firm, then, the expected benefits of remedying a cyber vulnerability often will be lower than the expected costs. Without the prospect of civil liability, firms have weaker incentives to invest in measures to secure their systems and products against cyberattacks.

Not only are liability fears failing to incentivize firms to take better precautions against cyberattacks, they are actually discouraging them from doing so. Companies sometimes are reluctant to better secure their systems because of concerns that these steps could expose them to civil liability. For instance, ISPs typically do not offer assistance if they discover that their customers' PCs have been infected by malware. ISPs often are able to tell, through routine traffic analysis, that a particular machine on the network is part of a botnet or has been infected by a worm.[194] "[B]ut they don't dare inform the customer (much less cut off access) out of fear that customers would . . . try to sue them for violating their privacy."[195] Doing so might even be a crime. The federal wiretap act makes it unlawful to "intentionally intercept[] . . . any wire, oral, or electronic communication,"[196] and some companies fear that filtering botnet traffic or

[190] Frye, *supra* note 144, at 367.

[191] BRENNER, *supra* note 1, at 224.

[192] 47 U.S.C. § 203(c)(1).

[193] Green v. America Online, 318 F.3d 465, 471 (3d Cir. 2003); *see generally* Lichtman & Posner, *supra* note 59, at 247-52.

[194] BRENNER, *supra* note 1, at 229; CLARKE & KNAKE, *supra* note 1, at 164.

[195] CLARKE & KNAKE, *supra* note 1, at 164-65; *see also* BRENNER, *supra* note 1, at 229; Coldebella & White, *supra* note 14, at 236-37.

[196] 18 U.S.C. § 2511(1)(a)

other malware might fall within this prohibition.[197] And while federal law makes an exception for ISPs that intercept communications to protect their own property,[198] there is no parallel exception for intercepts intended to protect the property of subscribers. Likewise, some ISPs are using deep packet inspection to examine the data streams on their networks for malicious code (this is probably lawful under the exception mentioned above, or a separate exception for "mechanical or service quality control checks"[199]). But even when they uncover malware, ISPs "have been reluctant to 'black hole' (or kill) malicious traffic because of the risk that they might be sued by customers whose service is interrupted."[200] Again, as in the antitrust context,[201] even if the applicable service contracts or state and federal laws do not clearly forbid these measures, the mere risk of liability may be enough to dissuade firms from undertaking them.

While firms with poor cyberdefenses generally do not face the prospect of civil lawsuits, there is one context in which a credible liability threat exists: data breaches in the financial services sector. The Gramm-Leach-Bliley Act of 1999 directs a group of federal agencies, such as the Federal Trade Commission and the Federal Deposit Insurance Corporation, to issue data security regulations for financial institutions.[202] In particular, the act mandates the adoption of "administrative, technical, and physical safeguards" that will, among other things, "insure the security and confidentiality of customer records and information" and "protect against unauthorized access to or use of such records."[203] The sanctions for violating these data security requirements can be severe. The GLB Act does not enumerate specific penalties, but rather directs the enforcing agencies to apply the act's requirements according to their respective enabling statutes.[204] Thus, for example, a bank subject to FTC jurisdiction would face a civil penalty of up to $16,000 for each violation.[205] If the FTC treated every customer affected by a cyber intrusion as a separate violation, the penalties very quickly would become staggering. If 100 customers have their data compromised the bank would face up to a $1.6 million penalty, 10,000 customers would mean up to a $160 million penalty, and so on.

Not coincidentally, financial institutions are widely believed to do at better job of protecting customer data than members of other industries.[206] Unlike other firms, which typically spend only modest sums on cybersecurity, most banks make large investments,

---

[197] BRENNER, *supra* note 1, at 229-30.

[198] 18 U.S.C. § 2511(2)(a)(i).

[199] 18 U.S.C. § 2511(2)(a)(i).

[200] CLARKE & KNAKE, *supra* note 1, at 163; *see also* McAfee 2010, *supra* note 35, at 5.

[201] *See supra* notes 158 to 162 and accompanying text.

[202] 15 U.S.C. § 6805. *See generally, e.g.*, 16 C.F.R. pt. 314 (Federal Trade Commission rule).

[203] 15 U.S.C. § 6801(b); *see* Kenneth A. Bamberger, *Regulation as Delegation*, 56 DUKE L.J. 377, 391 (2006); Schwartz & Janger, *supra* note 59, at 920.

[204] 15 U.S.C. § 6805(b).

[205] 16 C.F.R. § 1.98.

[206] ABA, *supra* note 18, at 21; Frye, *supra* note 144, at 367-68; Powell, *supra* note 14, at 501-05. *But see* Gable, *supra* note 2, at 84 (emphasizing that banks remain vulnerable to cyberattack).

"between 6 and 7 percent of their entire information technology budgets."[207] Financial institutions also are more likely to adopt specific security measures like intrusion detection and prevention systems, antivirus software, smart cards, and biometrics.[208] The unique risk of liability that banks face may be responsible, at least in part, for that record. The GLB Act has the effect of increasing the expected benefit of cybersecurity – namely, avoiding potentially crippling civil penalties – and thus creates strong incentives for banks to invest in defenses. (Another explanation is the risk of customer exit. It is relatively easy for a customer who fears cyber intrusions to switch banks, so the bank has an incentive to maintain data integrity.[209]) Of course, the GLB Act's emphasis on protecting consumer data might distort firms' cybersecurity investments. Rather than expending resources on defenses against the attacks they regard as the most dangerous, or the most likely to occur, financial institutions will tend to prioritize defenses against the one form of intrusion singled out by their regulators – the compromise of customer data.[210] The effect may be to ensure that firms are well defended against one threat at the expense of increased exposure to many other threats.[211] Even so, Gramm-Leach-Bliley remains an example of how the risk of civil liability might be used to incentivize firms to improve (at least some of) their cyberdefenses.

E. . . . as a Public Health Problem

As several scholars have noted, in more or less detail, cybersecurity can be thought of in terms of public health.[212] A critically important goal for any cybersecurity regime is to keep attacks from happening and to contain their ill effects.[213] The same is true of public health, the ultimate goal of which is prevention.[214] Unlike medical practice, which typically has an ex post orientation (treating illnesses that have already occurred), public health is primarily oriented toward ex ante solutions – preventing people from contracting infectious diseases, preventing pathogens from spreading, and so on. Broadly summarized, public health law – including the subset known as public health emergency law – involves government efforts "to persuade, create incentives, or even compel individuals and businesses to conform to health and safety standards for the collective good."[215] These interventions are sometimes defended, controversially, on paternalistic grounds. The notion is that the state may curtail individuals' freedoms to promote their own interests, which in this context means their physical health and safety.[216] By far the

---

[207] Powell, *supra* note 14, at 502

[208] Powell, *supra* note 14, at 503.

[209] *See supra* notes 40 to 47 and accompanying text.

[210] Similar distortions may arise at the state level, as a number of states have enacted laws requiring designated companies to disclose breaches of customer data. Vincent R. Johnson, *Cybersecurity, Identity Theft, and the Limits of Tort Liability*, 57 S.C. L. REV. 255, __ (2005); Schwartz & Janger, *supra* note 59, at 917.

[211] *Cf.* BAKER, *supra* note 24, at 238-39; McAfee 2010, *supra* note 35, at 29.

[212] *See generally* Rattray et al., *supra* note 8; IBM, *supra* note 19; *see also* Coyne & Leeson, *supra* note 18, at 480; Hunker, *supra* note 19, at 202-03; Katyal, *Criminal Law*, *supra* note 10, at 1081; Rosenzweig, *supra* note 14, at 19.

[213] Katyal, *Community*, *supra* note 127, at 34; Katyal, *Criminal Law*, *supra* note 10, at 1078-79.

[214] LAWRENCE O. GOSTIN, PUBLIC HEALTH LAW 19 (2d ed. 2008).

[215] GOSTIN, *supra* note 214, at xxii.

[216] GOSTIN, *supra* note 214, at 50-54.

more common, and widely accepted, justification for public health law is the risk of harm to others. The state may coerce persons who have contracted an infectious disease or are at risk of doing so, the theory goes, to prevent them from transmitting the disease to, and thereby harming, others.[217] Seen in this light, a principal objective of public health law is to internalize negative externalities – in particular, the costs associated with spreading infections to others.

Public health law contemplates three specific measures that are relevant here: mandatory inoculations to reduce susceptibility to infectious diseases, biosurveillance to monitor for epidemics and other outbreaks, and isolation and quarantine to treat those who have been infected and prevent them from spreading the pathogen.[218] We will consider each in turn along with their potential relevance to cybersecurity.

Inoculation, in which a healthy subject is exposed to a pathogen, helps prevent disease both directly (a person who is inoculated against a disease is thereby rendered immune) and indirectly (the person's immunity reduces the risk that he will transmit the disease to others). Inoculation mandates can take several forms. In the nineteenth and early twentieth centuries, state and local governments sometimes opted for direct regulation – a firm legal requirement that citizens must receive a particular vaccine, backed by the threat of sanctions.[219] (In the 1905 case of *Jacobson v. Massachusetts*,[220] the Supreme Court upheld such a requirement against a lawsuit invoking Fourteenth Amendment's privileges or immunities, due process, and equal protection clauses. According to the Court, mandatory inoculation is a permissible exercise of the states' police powers.[221]) The modern approach usually involves a lighter touch. Now, state and local governments typically create incentives for citizens to undergo inoculation by making it a condition of eligibility for certain valuable benefits. The best known example is to deny children access to public schools unless they have been vaccinated.[222] (The Supreme Court upheld such a scheme in 1922 in *Zucht v. King*.[223])

It isn't necessary to inoculate all members of a population to frustrate the transmission of a given disease. This is due to "herd immunity." Herd immunity theory proposes that, for a contagious disease that is transmitted from person to person, chains of infection are likely to be disrupted when large numbers of a population are immune or less susceptible to the disease.[224] The critical number is typically around 85 percent of the population, but it can be as low as 75 percent for some diseases (such as mumps) and as high as 95 percent for others (such as pertussis – i.e., whooping cough).[225] Herd immunity is a form of positive externality – those who undergo

[217] GOSTIN, *supra* note 214, at 49.

[218] GOSTIN, *supra* note 214, at 11, 39.

[219] GOSTIN, *supra* note 214, at 379.

[220] 197 U.S. 11 (1905).

[221] 197 U.S. at 34.

[222] GOSTIN, *supra* note 214, at 379-80, 382; Hunker, *supra* note 19, at 203; Lichtman & Posner, *supra* note 59, at 255.

[223] 260 U.S. 174 (1922).

[224] Katyal, *Criminal Law*, *supra* note 10, at 1081.

[225] http://www.bt.cdc.gov/agent/smallpox/training/overview/pdf/eradicationhistory.pdf.

vaccination provide an uncompensated benefit to those who do not – and, as a result, there is a free rider problem.[226] Many people would prefer to enjoy the benefits of herd immunity without themselves undergoing vaccination, which is costly (money, discomfort, risk of reaction, etc.); they would rather be part of the 15 percent than the 85 percent. The effect is to weaken each person's incentive to undergo vaccination, and overall vaccinations may drop below the levels needed to support herd immunity. The need to overcome this free rider problem helps explain why state and local governments sometimes use their coercive powers to require inoculation. (Another approach would be to provide subsidies to those who have been inoculated. Public school vaccination requirements can be understood in these terms; the government is subsidizing the education of children who are inoculated.)

Ensuring widespread immunity – not to disease, but to malicious code – is also an important goal of cybersecurity. The average internet-connected computer may be even more susceptible to infection by malware than the average person is to infection by a pathogen, because malicious code can propagate more efficiently than disease. Many pathogens are transmitted by person-to-person contact; you are unlikely to contract polio unless you come into close proximity with someone who is already infected. But one can contract malware from virtually any (networked) computer in the world. In effect, the internet brings dispersed systems into direct contact with one another; alternatively, the internet is a disease vector that, like mosquitoes and malaria, can transmit a contagion between dispersed systems. It is therefore essential for the elements at the edge of the network (such as the SCADA system that runs the local power plant) to maintain effective defenses against cyber intrusions (such as disconnecting the power plant's controls from the public internet). And there's the rub. As with herd immunity, cybersecurity raises free rider problems.[227] A user who takes steps to prevent his computer from being infected by a worm or impressed into a botnet thereby makes other systems more secure; if the user's machine is not infected, it cannot transmit the malware to others. But the user receives no compensation from those who receive this benefit; he does not internalize the positive externality. He therefore has weaker incentives to secure his system, as he – like everyone else – would prefer to free ride on others' investments. A critical challenge for any cybersecurity regime, then, is to reverse these incentives.

The second key element of public health law is biosurveillance. "Biosurveillance is the systematic monitoring of a wide range of health data of potential value in detecting emerging health threats."[228] Public health officials collect and analyze data to determine a given disease's incidence, or "the rate at which new cases occur in a population during a specified period," as well as its prevalence, or "the proportion of a population that are cases at a point in time."[229] Effective biosurveillance is a vital first step in managing an epidemic or other outbreak.[230] Biosurveillance takes place through a partnership among the U.S. Centers for Disease Control

[226] Coyne & Leeson, *supra* note 18, at 480; GOSTIN, *supra* note 214, at 378-79. *See generally supra* notes 134 to 135 and accompanying text.

[227] *See supra* notes 126 to 135 and accompanying text.

[228] GOSTIN, *supra* note 214, at 291.

[229] Rattray et al., *supra* note 8, at 152.

[230] IBM, *supra* note 19, at 11.

and Prevention, the CDC's state level counterparts, and front line health care providers (such as hospitals, clinics, and individual medical practitioners). Many, if not all, states have enacted legislation requiring specified health care professionals to notify state authorities if their patients have contracted any number of infectious diseases,[231] such as smallpox or polio.[232] These reports typically include the patient's name, the type of disease, his medical history, and other personal information.[233] State authorities then share the data with the CDC. These reports are not required by law, but most states appear to be fairly conscientious about them.[234] Public health law thus relies on a system of distributed surveillance. No central regulator is responsible for collecting all the data needed to detect and respond to infectious disease outbreaks. Instead, the system relies on individual nodes within a far flung network – from state agencies to hospitals to individual doctors – to gather the necessary information and route it to the CDC's central storehouse. CDC then analyzes the data and issues alerts advising state agencies and medical practitioners about disease trends and offering recommendations about how to respond.[235]

The third public health intervention involves containing infectious diseases once an outbreak has occurred, and preventing them from spreading further.[236] Two key measures are isolation and quarantine. Isolation and quarantine differ in subtle ways,[237] though in colloquial usage the terms are essentially synonymous. The goal of each is to segregate from the population those who have contracted or been exposed to an infectious disease and thus prevent them from transmitting it to those who are well. Isolation and quarantine are often coupled with mandatory treatment, which helps reduce the risk of further contagion; a person who has been cured of an infectious disease cannot transmit it to others.[238] The rationale for these interventions is the familiar harm principle – i.e., the risk that a person who has contracted or been exposed to a pathogen will infect others.[239] Isolation and quarantine thus seek to reduce negative externalities.

At the federal level, isolation and quarantine are accomplished under the Public Health Service Act of 1944. The Secretary of Health and Human Services has authority under the act to

---

[231] GOSTIN, *supra* note 214, at 295-96.

[232] http://www.cdc.gov/mmwr/pdf/wk/mm5853.pdf

[233] GOSTIN, *supra* note 214, at 297.

[234] GOSTIN, *supra* note 214, at 296; Hunker, *supra* note 19, at 202-03.

[235] This reporting scheme is permissible under the Health Insurance Portability and Accountability Act privacy rule, which generally limits the use and disclosure of protected health information, 45 C.F.R. § 164.502(a), but which contains an exception for disclosures to public health authorities, 45 C.F.R. § 164.512(b). The reporting is probably constitutional as well. The Supreme Court in *Whalen v. Roe*, 429 U.S. 589 (1977), upheld, against a Fourth Amendment challenge, a similar New York law requiring physicians to report information about drug prescriptions.

[236] Rattray et al., *supra* note 8, at 154.

[237] Isolation involves separating persons who are known to be infected with a disease, for as long as the disease remains communicable. GOSTIN, *supra* note 214, at 429. Quarantine involves separating persons who, though asymptomatic, may have been exposed to a disease, for the period of communicability. GOSTIN, *supra* note 214, at 429.

[238] GOSTIN, *supra* note 214, at 411-12.

[239] GOSTIN, *supra* note 214, at 414-15.

"make and enforce such regulations as in his judgment are necessary to prevent the introduction, transmission, or spread of communicable diseases" into or within the United States.[240] The law further provides for the "apprehension, detention, or conditional release" of persons who may have been exposed to any one of several communicable diseases that the president has specified by executive order.[241] The list, which was updated most recently in 2005,[242] includes cholera, tuberculosis, plague, smallpox, SARS, and several other diseases.[243] Large scale isolation and quarantine are rarely used; the most recent example is from the 1918 Spanish flu pandemic (and it was carried out under different legal authorities). However, isolation and quarantine are sometimes used for particular individuals. In May 2007, HHS issued an isolation order for an American with multi-drug resistant tuberculosis who flew from the Czech Republic to Canada and then crossed the land border into the United States.[244] Violations of the quarantine regulations carry criminal penalties – a fine of up to $1000 and incarceration for up to a year.[245]

Both biosurveillance and isolation/quarantine have important lessons for cybersecurity. Like the public health system, effective cyberdefenses depend on information about the incidence and prevalence of various kinds of malware. Users – both individuals on their PCs and firms that operate extensive networks – need to know what new forms of malicious code are circulating on the internet in order to secure their systems against them. As for isolation and quarantine, a critical cybersecurity challenge is to ensure that systems infected with malicious code do not spread the contagion to other, healthy computers. In both cases, cybersecurity faces the same problems that arise in realspace disease control, and public health solutions therefore might be adapted for the cyber context.

There is, of course, a significant difference between infectious diseases and malicious computer code: Diseases (typically) develop and spread on their own, whereas malware is created by human beings and (sometimes) requires human intervention to propagate. This is true as far as it goes, but the differences between cyberspace and realspace pathogens can be overstated. Infectious diseases can be engineered (e.g., biological weapons), and sometimes malware is able to spread on its own (e.g., a worm that is programmed to search for other computers on which to replicate itself[246]). Another potential obstacle is the antiquity of public health statutes.[247] Many of these laws have been on the books for decades, even a century, and they do not necessarily reflect contemporary scientific understandings of disease.[248] Plus they often restrict civil liberties and privacy to a degree rarely seen today.[249] The judicial precedents

[240] 42 U.S.C. § 264(a).

[241] 42 U.S.C. § 264(b).

[242] Exec. Order No. 13375 (Apr. 1, 2005).

[243] Exec. Order No. 13295 (Apr. 4, 2003).

[244] http://www.hhs.gov/asl/testify/2007/06/t20070606b.html.

[245] 42 U.S.C. § 271(a).

[246] *See supra* note 22 and accompanying text.

[247] INSTITUTE OF MEDICINE, THE FUTURE OF THE PUBLIC'S HEALTH IN THE 21ST CENTURY 4 (2002).

[248] GOSTIN, *supra* note 214, at 24.

[249] GOSTIN, *supra* note 214, at 24.

upholding these statutes against various constitutional challenges typically antedate the Supreme Court's modern civil rights and liberties jurisprudence. It is not clear that today's Court would uphold, say, mandatory vaccination of adults as readily as it did in 1905.[250] Yet even if public health law fits uneasily into contemporary constitutional law, it can still be a useful framework for cybersecurity. This is so because, as explained below, the cyber versions of public health interventions can be friendlier to civil liberties and privacy than their realspace counterparts.[251]

III. REGULATORY PROBLEMS, REGULATORY SOLUTIONS

This concluding section examines the responses of environmental, antitrust, products liability, and public health law to the various challenges that arise in those fields, and it considers how those solutions might be adapted to the field of cybersecurity. The range of possible responses to cyber insecurity depends on our understanding of that problem. The security measures we choose are determined by our antecedent choice of how to describe the problem in the first place. If we regard cybersecurity from the standpoint of law enforcement and armed conflict, we will tend to favor the responses of law enforcement and armed conflict – stronger penalties for cyber intrusions, say, or retaliating with kinetic attacks, and so on. Those are plausible frameworks and equally plausible solutions. But they are not the only ones. A wider angle lens is needed. Going beyond the conventional approaches, and conceiving of cybersecurity in terms of the regulatory disciplines surveyed above, brings into focus certain aspects of the problem that otherwise might have gone unnoticed. A more comprehensive set of analytical frameworks also enlarges the menu of legal and policy responses available to decisionmakers.

Taken together, the frameworks described in Part II suggest that an effective cybersecurity regime should include four components: (1) monitoring and surveillance to detect malicious code; (2) hardening vulnerable targets and enabling them to defeat intrusions; (3) building resilient systems that can function during an attack and recover quickly; and (4) responding in the aftermath of an attack.[252] There are two complementary objectives here: preventing intrusions from happening at all, and enabling firms to withstand the intrusions that do take place.[253] Stronger defenses would provide an obvious, first order level of protection: Better defense means less damage. They also would provide an important second order level of protection: Stronger defenses can help achieve deterrence. By enabling victims to defeat, survive, and recover from cyberattacks, these measures increase the expected costs of an intrusion to an attacker (i.e., the costs one must bear to overcome the defenses) and also decrease its expected benefits.[254] And that means weaker incentives to attack in the first place; why expend your scarce resources to try to take down the power grid if the effort is likely to fail?

---

[250] Jacobson v. Massachusetts, 197 U.S. 11 (1905). *But see* GOSTIN, *supra* note 214, at 130 (proposing that the Court "indisputably" would reach the same result if it decided *Jacobson* today).

[251] *See infra* notes 272 to 273 and accompanying text.

[252] *Cf.* Nojeim, *supra* note 14, at 131; Trachtman, *supra* note 54, at 265.

[253] Bambauer, *supra* note 12, at 673; Yochai Benkler, *Peer Production of Survivable Critical Infrastructures*, *in* Grady & Parisi, *supra* note 19, at 76-77; BRENNER, *supra* note 1, at 214; CLARKE & KNAKE, *supra* note 1, at 159.

[254] CSIS, *supra* note 8, at 26; Lynn, *supra* note 19, at 99-100; Taipale, *supra* note 92, at 36.

Of course, it is inevitable that some attacks will succeed. Some intrusions can be prevented or mitigated but others cannot, and any defensive scheme necessarily will be imperfect.[255] (This is so because, in cyberspace, offense is much less costly than defense. "Defending a modern information system" is like "defending a large, thinly populated territory like the nineteenth century Wild West: the men in black hats can strike anywhere, while the men in white hats have to defend everywhere."[256]) The goal therefore is not to develop impregnable defenses. Doing so may be impossible from a technological standpoint, and even if they were feasible they may be inefficiently costly.[257] Instead, the goal is to adopt efficient levels of investment in defenses that are better at protecting society's critical systems than current defenses are.[258] Another important point is that cyberdefense is not a one-size-fits-all proposition. Security measures should be tailored to the unique risks faced by specific firms or industries – their combinations of vulnerabilities, threats, and consequences.[259] The strongest (and presumably most costly) defenses should be reserved for the firms that are most vulnerable to cyberattacks, that face the most severe threats (e.g., from foreign intelligence services as opposed to recreational hackers), and whose compromise would have the most devastating consequences for society. Strategically unimportant firms might get by with modest defenses, whereas robust defenses may be needed for critical industries.[260] Finally, what follows is by no means an exhaustive list of possible responses to cyber insecurity. It is merely a list of responses that are implied if we conceive of cybersecurity in environmental, antitrust, products liability, and public health terms. Other solutions, suggested by other analytical frameworks, may be just as promising.

A. Monitoring and Surveillance

Effective cybersecurity depends on the generation and exchange of information about cyber threats.[261] An ideal system would create and distribute vulnerability data (the holes intruders might exploit to gain access to computer systems), threat data (the types of malware circulating on the internet and the types of attacks firms have suffered), and countermeasure data (steps that can be taken to prevent or combat infection by a particular piece of malicious code).[262] Perhaps the best way to collect this information is through a distributed surveillance network akin to the biosurveillance system at the heart of public health law. Companies are unlikely to participate in this sort of arrangement due to fears of liability under antitrust and other laws. A suite of measures is therefore needed to help foster favorable incentives, including

---

[255] Bambauer, *supra* note 12, at 673; CSIS, *supra* note 8, at 51; Gable, *supra* note 2, at 65; IBM, *supra* note 19, at 12; Lynn, *supra* note 19, at 99; Sklerov, *supra* note 19, at 8; Taipale, *supra* note 92, at 9.

[256] Ross Anderson, *Why Information Security Is Hard – An Economic Perspective* 5 (__); *see also* BAKER, *supra* note 24, at 213; Jensen, *Cyber Warfare*, *supra* note 15, at 1536. *But see* Libicki, *supra* note 12, at 38.

[257] *See supra* notes 30 to 31 and accompanying text.

[258] DOROTHY DENNING, INFORMATION WARFARE AND SECURITY 12 (1999).

[259] ABA, *supra* note 18, at 21; Katyal, *Criminal Law*, *supra* note 10, at 1080; Nojeim, *supra* note 14, at 119.

[260] *See supra* notes 59 to 66 and accompanying text.

[261] *But see* CSIS, *supra* note 8, at 45 (information sharing should not be "a primary goal").

[262] *See supra* notes 144 to 146 and accompanying text.

subsidies, threats of liability, and offers of immunity. These steps won't guarantee that firms will collect and share cybersecurity data, but they will make such arrangements more viable than they are at present.

Public health law's system of distributed biosurveillance seems well suited to the challenge of gathering and disseminating data about a vast range of cyber threats. Like health care providers who diagnose and then report their patients' infectious diseases, firms could be tasked with monitoring their systems for vulnerabilities and intrusions, then reporting their findings (as well as the countermeasures they have implemented) to designated recipients. Such a system would take advantage of important information asymmetries. Individual companies often know more than outsiders about the vulnerabilities in their systems and the types of intrusions they have faced; they have a comparative advantage in compiling this data.[263] The principal alternative – surveillance by a single, central regulator – is unlikely to be as effective. As Hayek emphasized, "the knowledge of the [economic] circumstances of which we must make use never exists in concentrated or integrated form but solely as the dispersed bits of incomplete and frequently contradictory knowledge which all the separate individuals possess."[264] The same is true of cybersecurity data. A central regulator lacks the capacity to examine each device that is connected to the internet to determine its vulnerabilities, nor can it inspect every data packet transiting the internet to determine whether it contains malicious code. And even if the scope of the project wasn't prohibitively vast, the privacy costs associated with a central monitor – especially a government monitor – likely would be intolerable. Instead, the more efficient course would be to rely on individual firms to gather the relevant information.[265]

While firms would be responsible for the lion's share of monitoring, there is still an important role for the government: providing especially sensitive companies (such as power companies and ISPs) with information about especially sophisticated forms of malware. Here, the comparative advantage is reversed; the government's highly resourceful intelligence agencies are simply better than the private sector at detecting intrusions by sophisticated adversaries like foreign militaries and developing countermeasures.[266] The government can provide these firms with the signatures of malware used in previous attacks, and firms can use the signature files to detect future intrusions. In 2010, for instance, the National Security Agency began assisting Google in detecting intrusions into its systems. The partnership was announced in the wake of reports that sophisticated hackers, most likely affiliated with China's intelligence service, had broken into Google's systems and collected data about users, including a number of human rights activists.[267] The NSA reportedly has entered a similar partnership with a number of large banks.[268]

---

[263] Bamberger, *supra* note 203, at 391-92; CSIS, *supra* note 8, at 53; Katyal, *Criminal Law*, *supra* note 10, at 1091. *See generally* Bamberger, *supra* note 203, at 399 (emphasizing "the information asymmetries between regulated firms and administrative agencies").

[264] F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, __ (1945).

[265] CLARKE & KNAKE, *supra* note 1, at 162.

[266] Condron, *supra* note 19, at 407; Coldebella & White, *supra* note 14, at 240. *But see* O'Neill, *supra* note 19, at 265, 27; Taipale, *supra* note 92, at 9.

[267] Nakashima, *Google*, *supra* note 56.

[268] Andrew Shalal-Esa & Jim Finkle, *NSA Helps Banks Battle Hackers*, REUTERS, Oct. 26, 2011.

What should be the architecture of the system used to disseminate the vulnerability, threat, and countermeasure information compiled by private firms? At least two possibilities exist. Some commentators have called for the creation of a central repository of cybersecurity data – a "cyber-CDC,"[269] as it were. Under such a system, an individual firm would notify the clearinghouse if it discovers a new vulnerability in its systems, or a new type of malicious code, or a particular countermeasure that is effective against a particular kind of threat. The repository would analyze the information, looking for broader trends in vulnerabilities and threats, then issue alerts and recommendations to other firms. This clearinghouse might be a government entity, as in public health law, but it need not be. An alternative architecture would be for firms to exchange cybersecurity information with one another directly, on a peer-to-peer basis, rather than first routing it through a central CDC-type storehouse. One advantage of the peer-to-peer approach is that it may be more resilient. A central storehouse would be an attractive target for cyber adversaries, and the entire system would fail if it were compromised.

Distributed surveillance may be an even better fit for cybersecurity than for public health, for several reasons. First, malicious computer code often can be detected more quickly than biological pathogens,[270] which means that countermeasures can be put in place rapidly. Biosurveillance can be slow because the incubation period for certain diseases – i.e., the amount of time between when a disease is contracted and when its symptoms first manifest – can be days or weeks. By contrast, it is possible to detect known malware in real time, as the code is passing through a company's system (assuming, of course, that the firm's intrusion detection systems know what to look for, which is not always the case[271]). Second, cyberthreat monitoring has the potential to raise fewer privacy concerns than biosurveillance.[272] Health care providers often give authorities intensely sensitive information about individual patients, such as their names, Social Security numbers, and other personally identifiable information, as well the diseases they have contracted.[273] A properly designed cyber monitoring system need not compile and disseminate information of the same sensitivity. Collection and sharing could be limited to information about the incidence and prevalence of known malware. The fact that a particular system has been infected by the "ILoveYou" worm exposes a great deal less personal information, and thus raises weaker privacy concerns, than the fact that a particular patient suffers from HIV or breast cancer.

This framework has important limitations. Malware detection is an inexact science.[274] One challenge is that deep packet inspection and other forms of network monitoring typically work by comparing streams of data against signature files of known malicious code.[275] These systems are only as good as their underlying signatures. If there is no signature for a particular

[269] IBM, *supra* note 19, at 13-14; *see also* Sharp, *supra* note 8, at 25.

[270] Rattray et al., *supra* note 8, at 152.

[271] *See infra* notes 274 to 277 and accompanying text.

[272] *But see* Nojeim, *supra* note 14, at 126.

[273] GOSTIN, *supra* note 214, at 297.

[274] CLARKE & KNAKE, *supra* note 1, at 126; Sklerov, *supra* note 19, at 74.

[275] *See supra* note 156 and accompanying text.

type of malware, chances are it will not be detected. As a result, sophisticated "zero day" attacks – so called because they occur before the first day on which security personnel become aware of them and begin to develop countermeasures – may well go unnoticed.[276] Former CIA director Jim Woolsey emphasizes that "[i]f you can't deal with a zero-day attack coming from a thumb drive, you have nothing."[277] Of course, these are the very sorts of attacks likely to be launched by sophisticated adversaries like foreign intelligence services. Public health law's biosurveillance framework thus is probably better at detecting intrusions of low to modest complexity than those undertaken by foreign governments.

The challenge, then, is to provide firms with incentives to collect and disseminate information about cyber vulnerabilities, threats, and countermeasures.[278] At present companies have strong disincentives to do so, partly due to fears of legal liability,[279] but also because of concerns about compromising trade secrets, losing customer goodwill, reputational harms, and so on.[280] Public health law facilitates collection and sharing through direct regulation, such as state statutes requiring health care providers to notify authorities about patients who have contracted various infectious diseases.[281] A similar arrangement might be adopted for cyberspace. The government could require firms to gather information about the vulnerabilities in their systems, the types of attacks they have suffered, and the countermeasures they have used to combat malware, and then to disseminate the data to designated recipients.[282] (Imposing such an obligation would not eliminate companies' incentives to withhold cybersecurity data. It would simply make it more costly for them to do so, where cost is equal to the sanctions for hoarding discounted by the probability of punishment. Increased costs mean that firms are more likely to collect and share cybersecurity data, but some will still find it advantageous to hoard.) There is also a less coercive, and probably more effective, alternative. Cybersecurity data is a sort of public good, and economic theory therefore predicts that it will be underproduced.[283] One way to encourage the provision of public goods is to subsidize them, so firms might be offered bounties to compile and exchange the needed information.[284] These bounties could be direct payments from the government or, more probably, tax credits or deductions. The subsidies also could take the form of enhanced intellectual property protections for the cybersecurity information firms generate. If the subsidies are large enough, firms will have an incentive not

[276] Rosenzweig, *supra* note 14, at 28 n.23; Zetter, *supra* note 46.

[277] Quoted in McAfee 2011, *supra* note 15, at 1.

[278] Nojeim, *supra* note 14, at 128.

[279] *See supra* notes 148 to 150, 194 to 201 and accompanying text.

[280] ABA, *supra* note 18, at 10; Aviram, *supra* note 76, at 154; Aviram & Tor, *supra* note 131, at 240; Bambauer, *supra* note 12, at 611; Todd Brown, *supra* note 14, at 232; Coldebella & White, *supra* note 14, at 236; Frye, *supra* note 144, at 369; Katyal, *Digital Architecture*, *supra* note 15, at 2278; Nojeim, *supra* note 14, at 135; Powell, *supra* note 14, at 501; Rosenzweig, *supra* note 14, at 9; Sarnikar & Johnsen, *supra* note 19, at 19; Schwartz & Janger, *supra* note 59, at 931; Smith, *supra* note 17, at 172-73. *But see* O'Neill, *supra* note 19, at 281.

[281] *See supra* notes 228 to 235 and accompanying text.

[282] Frye, *supra* note 144, at 370-71.

[283] *See supra* notes 129 to 131 and accompanying text. *But see* Aviram & Tor, *supra* note 131, at 234-35 (arguing that information can be a rivalrous good).

[284] Nojeim, *supra* note 14, at 128.

just to report the data they have already compiled, but to invest in discovering previously unknown vulnerabilities, threats, and countermeasures.[285]

Antitrust law also can help recalibrate firms' incentives.[286] Antitrust is often skeptical of information sharing and other forms of cooperation among companies.[287] But in the cybersecurity context, such arrangements can enhance consumer welfare: Agreements to exchange vulnerability, threat, and countermeasure information can help prevent cyberattacks from taking place (or at least mitigate their effects) and thereby minimize the resulting consumer losses.[288] One way to incentivize companies to cooperate is to alleviate their apparently widespread fears of antitrust liability. This could be accomplished through judicial, administrative, or legislative action. Federal courts could expressly discard the per se approach and substitute a rule of reason when reviewing private sector agreements to share cybersecurity data or to adopt common security protocols. This doctrinal shift would reduce the risk of false positives – i.e., the danger that the coarse grained per se rule might invalidate a cybersecurity initiative that is actually welfare enhancing. Instead, arrangements would be judged on a case by case basis, and would stand or fall based on the degree to which they actually advance or hinder consumer welfare. While this approach shows promise, it also has some significant drawbacks. The judicial route may not do enough to remove legal uncertainty. At the time companies are deciding whether to enter a cybersecurity venture, it will not always be possible to predict whether reviewing courts would sustain or invalidate it. A wrong guess could have severe consequences: Firms that are found to have violated federal antitrust law must pay treble damages to the successful plaintiffs.[289] This continued risk of legal liability will dissuade companies from entering arrangements that ultimately may withstand judicial scrutiny but, then again, possibly might not. In short, the uncertain prospects of ex post judicial approval may not provide firms with enough assurance ex ante.

A more promising approach would be for administrative agencies to sponsor cybersecurity exchanges, as some in Congress have proposed.[290] Agencies with special expertise in cybersecurity (such as the NSA and DHS) could partner with the agencies that are responsible for enforcing federal antitrust laws (the Federal Trade Commission and the Justice Department's antitrust division) to establish fora in which companies could establish common security standards and exchange information. The government's participation in these fora would offer assurances that they are being used for legitimate purposes and not as vehicles for anticompetitive conduct. From the standpoint of participating firms, this approach is advantageous because it offers them de facto antitrust immunity.[291] It is unlikely that an FTC or DOJ that sponsored a cooperative cybersecurity arrangement later would go to court to have it

[285] *But see* Malloy, *supra* note 83, at 572-73 (predicting that firms will tend to neglect "regulatory investments" – i.e., expending scarce resources to obtain benefits offered to those who comply with government regulations).

[286] *Cf.* Adler, *supra* note 147, at __.

[287] *See supra* notes 136 to 141 and accompanying text.

[288] *See supra* notes 144 to 146 and accompanying text.

[289] 15 U.S.C. § 15(a); *id*. § 15a; *id*. § 15c(a)(2).

[290] Cybersecurity Act of 2012, S. 2102, 112th Cong. § 301 (2012).

[291] BRENNER, *supra* note 1, at 228.

invalidated. And while the blessing of these agencies does not formally bind other potential plaintiffs, such as state attorneys general or private parties, their determination that a proposed venture is permissible under federal antitrust laws probably would receive a healthy dose of judicial deference. Government sponsorship has another advantage: It can help solve the coordination and free rider problems associated with collective action.[292] Left to their own devices, companies would prefer for their competitors to bear the expense of implementing new security standards and then free ride on the resulting security gains; they also would prefer to gain a competitive advantage by using other firms' information and then withholding their own data from their competitors. A regulator can mitigate these tendencies. It can coerce firms into participating in the forum and complying with its requirements; it also can withhold the forum's benefits from firms that shirk.

A third measure would be for Congress to enact a cybersecurity exception to the antitrust laws.[293] The upside of a legislative carve out is that it would eliminate virtually all risk of liability and thus remove one powerful disincentive for companies to cooperate on cybersecurity initiatives. Ideally, such a measure would be narrowly tailored to the precise sort of inter-firm cooperation that is desired – the exchange of vulnerability, threat, and countermeasure information and the development of common security protocols. In other words, the exemption would be pegged to specific conduct, and would not extend antitrust immunity to entire industries (as used to be the case with major league baseball[294]). A broader exception would offer few additional cybersecurity gains and could open the door to anticompetitive conduct.

We also might consult products liability law for ideas on how to incentivize companies to exchange cybersecurity data. Firms have weaker incentives to search for vulnerabilities in the systems they operate or the products they offer, because they typically cannot be sued by those injured by any resulting breaches.[295] At the same time, other companies, such as ISPs, are reluctant to monitor network traffic for malicious code because of fears that doing so could expose them to legal liability.[296] Adjusting the liability rules could help recalibrate these incentives.[297] Lawmakers might use a combination of carrots and sticks. Offers of immunity would increase companies' expected benefits of compiling and sharing cybersecurity data; threats of liability would increase their expected costs of failing to do so.[298]

Consider the carrots first. Firms could be offered immunity from various laws that presently inhibit them from collecting and exchanging certain information about cyber vulnerabilities and threats. One way to accomplish this would be for Congress to expand the

---

[292] Kobayashi, *supra* note 33, at 23; Sarnikar & Johnsen, *supra* note 19, at 22-23.

[293] Katyal, *Community*, *supra* note 127, at 52.

[294] Federal Baseball Club v. National League, 259 U.S. 200 (1922), *reaffirmed by* Flood v. Kuhn, 407 U.S. 258 (1972), *abrogated by* 15 U.S.C. § 26B.

[295] *See supra* notes 185 to 193 and accompanying text.

[296] *See supra* notes 194 to 201 and accompanying text.

[297] Sarnikar & Johnsen, *supra* note 19, at 22.

[298] Malloy, *supra* note 83, at 531-32. *But see id.* at 573 (predicting that firms will tend to neglect "regulatory investments" – i.e., complying with regulations to receive the benefits they offer).

service provider exception to the federal wiretap act's general ban on intercepting electronic communications.[299] The exception could be broadened to authorize ISPs to monitor network traffic for malicious code that threatens their subscribers' systems, not just their own systems. Congress also could authorize (or perhaps even require) ISPs to notify customers whose systems are found to be infected by malware.[300] It further could expressly preempt any state laws to the contrary. This would foreclose any claims that monitoring for malware violates a given state's privacy law or breaches the terms of service between an ISP and its subscribers. In all cases, eligibility for these forms of immunity could be conditioned on information sharing: A company would not be able to take advantage of the safe harbor unless it shared the information it discovered with other firms. The result would be to create strong incentives to exchange data about threats and vulnerabilities

As for the sticks, below I propose modifying tort law's traditional economic loss doctrine (under which a defendant generally is not liable for freestanding economic harms, only economic harms that result from physical injuries) in the cybersecurity context.[301] Firms that implement approved industry wide security standards would enjoy immunity from lawsuits seeking redress for injuries sustained from an intrusion; companies that disregard the protocols would be subject to lawsuits for any resulting damages. Under such a scheme, a company that has implemented the standards might have its immunity stripped if it fails to share information about known weaknesses in its systems or products. As for firms that fail to adopt the security standards, the lack of information sharing could be treated as an aggravating factor; extra damages could be imposed on firms that are aware of vulnerabilities or threats but fail to share that information with other companies. This series of tiered penalties would result in marginal deterrence; firms would have good reason not only to implement the approved security standards, but to exchange the threat and vulnerability information on which those protocols depend.

B. Hardening Targets

A second objective for a cybersecurity regime is to harden critical systems against attack by developing effective security protocols.[302] The goal of such measures is to prevent cyber intruders from infecting these systems at all, as opposed to limiting the amount of damage intrusions can do; the objective is to increase the impregnability of critical systems, as opposed to their survivability.[303] (Of course, some cyberattacks inevitably will succeed, so enhancing survivability – discussed below[304] – is an essential goal as well.) The regulatory disciplines surveyed above suggest various techniques for encouraging companies to adequately secure their networks. Environmental law suggests the need for industry wide security standards; these rules should be developed through collaborative partnerships between regulatory agencies and private firms, rather than imposed via direct regulation. Products liability law suggests that pairing

[299] 18 U.S.C. § 2511(2)(a)(i).

[300] BRENNER, *supra* note 1, at 229-31; CLARKE & KNAKE, *supra* note 1, at 164-65.

[301] *See infra* notes 332 to 338 and accompanying text.

[302] CLARKE & KNAKE, *supra* note 1, at 159.

[303] *See supra* note 253 and accompanying text.

[304] *See infra* Part III.C.

threats of liability with offers of immunity can incentivize firms to implement the security standards. And public health law's use of mandatory vaccinations might be adapted by incentivizing firms to take certain minimum steps to secure their systems. Again, different firms and industries face different vulnerabilities, threats, and consequences, so the resulting security standards should be calibrated to the particular conditions in individual industries.

Regulators could improve critical systems' defenses by establishing and enforcing new cybersecurity protocols, akin to the environmental regulations that restrict, say, the amount of sulfur dioxide a given source may emit into the atmosphere.[305] A cyberattack on critical infrastructure will not just harm the targeted company, it also will impose negative externalities on a number of remote third parties. It is usually impossible to internalize the resulting costs through market exchanges, because the transaction costs would be prohibitive. Regulatory standards can help manage these spillovers. It should be emphasized at the outset that the specific content of any cybersecurity standards is well beyond the scope of this article.[306] My focus here is not on the technical feasibility or policy advantages of any particular defensive measure. Instead, the focus of this article is establishing regulatory mechanisms by which new cybersecurity standards – whatever their content – may be adopted.

Turning to that question, one obvious option would be for administrative agencies to use traditional "command and control" regulation – i.e., issue a set of mandatory standards and incentivize firms to comply with them by threatening civil or criminal penalties.[307] This is a fairly common approach in environmental law, and some scholars have urged the government to adopt it here. Neal Katyal argues that "direct government regulation" of cybersecurity "is the best solution," and calls for regulatory agencies to issue "the equivalent of building codes to

---

[305] It is also possible to develop new cybersecurity standards through litigation. Harper, *supra* note 119, at 2; Johnson, *supra* note 210, at 275-76; Rosenzweig, *supra* note 14, at 23. A court might hold, for instance, that a given firm's failure to adopt a particular security measure breaches a general duty of care. This option seems less promising than the regulatory approach for several reasons. First, courts may not have the technical expertise to fashion detailed security protocols for complicated systems and products. Second, there is the problem of legal uncertainty. A regulation is likely to be more comprehensive than a series of incremental judicial opinions, especially in the context of a highly complex subject matter like cybersecurity, and relying on litigation thus runs the risk that firms will not know what is expected of them. There is, of course, an important role for litigation – the prospect of civil liability creates incentives for firms to comply with the regulatory standards. *See infra* notes 332 to 344 and accompanying text. But litigation should be limited to enforcing the standards, not formulating them in the first place.

[306] Just within the law review literature – to say nothing of computer science, economics, and other fields – authors have debated relatively modest regulations such as mandating that firms use encryption, firewalls, and intrusion detection systems, Condron, *supra* note 19, at 410; Gable, *supra* note 2, at 94-95, requiring companies that operate certain sensitive systems to authenticate users before granting them access, Nojeim, *supra* note 14, at 131-33; Sklerov, *supra* note 19, at 22-24, and disconnecting vulnerable SCADA systems from the internet, CLARKE & KNAKE, *supra* note 1, at __; McAfee 2010, *supra* note 35, at 34. Others have debated even more dramatic proposals, such as requiring ISPs to monitor the traffic that flows over their networks for malicious code, Lichtman & Posner, *supra* note 59, at 222; Katyal, *Criminal Law*, *supra* note 10, at 1007, 1095-1101; Taipale, *supra* note 92, at 34, or moving to an entirely new internet architecture (such as IPv6) in which anonymity is reduced and user activity is capable of being traced, Bambauer, *supra* note 12, at 590, 601; BAKER, *supra* note 24, at 231-32; Frye, *supra* note 144, at 354; Katyal, *Digital Architecture*, *supra* note 15, at 2269-70; LESSIG, *supra* note 67, at 45, 54; POST, *supra* note 90, at 84; Taipale, *supra* note 92, at 31.

[307] Malloy, *supra* note 83, at 531.

require proper design and performance standards for software."[308] Likewise, a prominent think tank argues that "the federal government bears primary responsibility" for cybersecurity and that "it is completely inadequate" to leave the matter "to the private sector and the market."[309] Some have even called for the federal government to take over certain sectors of the economy in the name of cybersecurity. According to an ABA task force, "government may also need to 'semi-nationalize' some sectors (like the electricity grid) where isolation is not an option and the adverse consequences of certain low probability events are likely to be very high."[310] It isn't steel mills, but Harry Truman would have admired the proposal.[311]

Traditional command and control regulation seems ill suited to the task of securing the nation's critical cyber infrastructure. The better course would be to involve the firms that operate these assets in establishing and implementing new security protocols. Private sector participation – an approach sometimes seen in environmental law – is desirable for several familiar reasons. First, information asymmetries: Companies often (though not always) know more than regulators about the vulnerabilities in their systems, the types of intrusions they have faced, and the most effective countermeasures for dealing with those threats.[312] A related concern is that regulators probably lack the knowledge necessary to determine the socially optimal level of cyber breaches and set the security standards accordingly.[313] The market, through the price system, is capable of aggregating and processing this information in a way that central planners cannot. Third, rapid technological change makes it difficult for regulators to formulate durable security rules.[314] Vulnerabilities, threats, and countermeasures are in a constant state of flux, and regulatory standards cannot keep pace with these developments. Notice and comment rulemaking rarely takes less than 18 months, sometimes much longer,[315] and the rules likely would be obsolete before the ink in the *Federal Register* was dry. Fourth, there is a risk that government protocols will stifle innovation.[316] If regulatory agencies promulgate a set of mandatory standards, regulated firms will have less reason to search for newer and more efficient countermeasures; they will simply implement the government's directives.

What specific role should private firms have in developing and implementing cybersecurity standards? At least two possibilities come to mind.[317] First, regulators could

---

[308] Katyal, *Digital Architecture*, *supra* note 15, at 2284, 2286. *But see* Katyal, *Criminal Law*, *supra* note 10, at 1091

[309] CSIS, *supra* note 8, at 15; *see also* Frye, *supra* note 144, at __.

[310] ABA, *supra* note 18, at 27.

[311] Youngstown Sheet & Tube Co. v. Sawyer, 343 U.S. 579 (1952).

[312] *See supra* notes 263 to 265 and accompanying text.

[313] Coyne & Leeson, *supra* note 18, at 488-89; Kobayashi, *supra* note 33, at 26-27; Powell, *supra* note 14, at 502, 505.

[314] BAKER, *supra* note 24, at 235, 237; CSIS, *supra* note 8, at 51; Rosenzweig, *supra* note 14, at 10.

[315]

[316] CSIS, *supra* note 8, at 51; Kobayashi, *supra* note 33, at 26.

[317] *See also* Kobayashi, *supra* note 33, at 27 (calling for the creation of private "security cooperatives" as an "alternative to government standards"). *But see* Aviram, *supra* note 76, at 150, 182 (arguing that private security standard setting is unlikely to be effective due to high enforcement costs).

practice a form of "delegated regulation"[318] in which they mandate broad security goals and establish the penalties for falling short, then leave it up to companies to achieve those goals in whatever manner they deem most effective.[319] Regulation by delegation is said to be appropriate where administrative agencies have the capacity to "identify specific outcomes but cannot easily codify in generally-applicable rules the means for achieving them."[320] It is also desirable where, as is often the case, private firms have a comparative advantage at developing innovative and efficient ways to achieve regulatory goals. Environmental law sometimes follows this approach (as do other fields such as food safety[321] and securities regulation[322]). For instance, the EPA's "bubble" approach to the Clean Air Act allowed polluters to offset increased emissions from one source with decreased emissions from other sources, providing them with an incentive to experiment with new technologies that could reduce emissions at lower cost.[323] Delegated regulation seems a good fit for cybersecurity, though not perfect one. Giving companies discretion to implement the government's security standards achieves three of the four benefits of private action mentioned above: It avoids (some) problems with information asymmetries, allows for flexibility in reacting to fast changing technologies, and promotes rather than stifles private sector innovation. However, difficulties would remain with formulating the standards. As is true of much command and control regulation, agencies lack the knowledge needed to determine the socially optimal level of cyber breaches and set the security standards accordingly.

An alternative would be a form of "enforced self-regulation"[324] in which private companies develop the new cybersecurity protocols in tandem with the government.[325] These requirements would not be handed down by administrative agencies, but rather would be developed through a collaborative partnership in which both regulators and regulated would play a role. In particular, firms might prepare sets of industrywide security standards. (The National Industrial Recovery Act, famously invalidated by the Supreme Court in 1935, contained such a mechanism,[326] and today the energy sector develops reliability standards in the same way.[327]) Or agencies could sponsor something like a negotiated rulemaking in which regulators, firms, and

---

[318] Schwartz & Janger, *supra* note 59, at 919; *see also* Bamberger, *supra* note 203, at 385, Jody Freeman, *The Private Role in Public Governance*, 75 N.Y.U. L. REV. 543 (2000).

[319] Bamberger, *supra* note 203, at 380; *see also* ABA, *supra* note 18, at 9; CLARKE & KNAKE, *supra* note 1, at 134; Jensen, *Cyber Warfare*, *supra* note 15, at 1565.

[320] Bamberger, *supra* note 203, at 389.

[321] Cary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management to Achieve Public Goals*, 37 LAW & SOC'Y REV. 691, 696-700 (2003).

[322] Bamberger, *supra* note 203, at 390-91.

[323] Chevron U.S.A. Inc. v. Natural Resources Defense Council, Inc., 467 U.S. 837 (1984); Malloy, *supra* note 83, at 536, 541; *see also* Malloy, *supra* note 83, at 547-49 (discussing conflicting accounts of whether bubble approach actually promoted innovation).

[324] Bamberger, *supra* note 203, at 461 (citing IAN AYRES & JOHN BRAITHWAITE, RESPONSIVE REGULATION: TRANSCENDING THE DEREGULATION DEBATE 101-32 (1995)).

[325] ABA, *supra* note 18, at 9; Coldebella & White, *supra* note 14, at 241-42; *Katyal, Criminal Law*, *supra* note 10, at 1099.

[326] A.L.A. Schechter Poultry Corp. v. United States, 295 U.S. 495 (1935).

[327] CSIS, *supra* note 8, at 52-53.

other stakeholders forge a consensus on new security protocols.[328] In either case, agencies then would ensure compliance through standard administrative techniques like audits, investigations, and enforcement actions.[329] This approach would achieve all four of the benefits of private action mentioned above: It avoids (some) problems with information asymmetries, takes advantage of distributed private sector knowledge about vulnerabilities and threats, accommodates rapid technological change, and promotes innovation. On the other hand, allowing firms to help set the standards that will be enforced against them may increase the risk of regulatory capture – the danger that agencies will come to promote the interests of the companies they regulate instead of the public's interests.[330] The risk of capture is always present in regulatory action, but it is probably even more acute when regulated entities are expressly invited to the decisionmaking table.[331]

Products liability law likewise offers several strategies for hardening critical infrastructure against cyberattacks. The prospect that a company might be required to pay money damages to those who have been injured by an attack on the systems they operate or the products they offer would internalize costs that are now externalized onto others. Liability thus would incentivize firms to offer goods (such as computer software) and services (such as online banking) that are more secure.[332] At present, companies face little risk of liability for the injuries that result from their failure to prevent cyber intrusions, because the economic loss doctrine generally limits them to paying damages for physical injuries and associated economic harms, not for freestanding economic injuries.[333] Modifying this default rule of de facto immunity could help foster incentives for firms to improve their cyberdefenses.

What could a recalibrated liability regime for cybersecurity look like? Again, a combination of carrots and sticks could be used. Congress might abolish the economic loss doctrine for injuries that result from a given company's (wrongful) failure to prevent a cyberattack. In its place, lawmakers could substitute a regime that imposes liability or offers immunity based on what steps a company has taken to secure its products or systems. As for the

---

[328] 5 U.S.C. §§ 561-70.

[329] CSIS, *supra* note 8, at 52.

[330] George Stigler, *The Theory of Economic Regulation*, 2 BELL J. ECON. & MANAGEMENT SCI. 3 (1971). A related problem is that, because of information asymmetries, agencies often depend on the companies they regulate to provide the data they need to formulate rules. Yet firms will have an incentive to underestimate vulnerabilities and threats to persuade regulators to approve lenient (and less costly) security protocols. Coyne & Leeson, *supra* note 18, at 489. (Of course, that concern is also present in traditional regulation.) There are also doctrinal difficulties. Depending on how the public-private partnership is structured, it conceivably could violate what remains of the nondelegation doctrine. *See, e.g.*, Carter v. Carter Coal Co., 298 U.S. 238 (1936) (striking down a statute that authorized coal producers to establish minimum prices in certain geographic regions on the ground that it was an unconstitutional delegation of legislative power to private companies).

[331] USA Group Loan Servs., Inc. v. Riley, 82 F.3d 708, __ (7th Cir. 1996) (Posner, C.J.) (describing negotiated rulemaking as "an abdication of regulatory authority to the regulated, the full burgeoning of the interest-group state, and the final confirmation of the 'capture' theory of administrative regulation").

[332] Coyne & Leeson, *supra* note 18, at 492; Lichtman & Posner, *supra* note 59, at 232-39; Hunker, *supra* note 19, at 211; Johnson, *supra* note 210, at 260; Rosenzweig, *supra* note 14, at 23; Schnieier, *supra* note 33, at 2; Yang & Hoffstadt, *supra* note 15, at 207-10.

[333] *See supra* notes 185 to 187 and accompanying text.

carrots, firms that implement the security standards that are developed in tandem with regulators, but nevertheless suffer cyberattacks, could be offered immunity from lawsuits seeking redress for the resulting damages.[334] This cyber safe harbor could extend not just to purely economic injuries (for which firms currently enjoy de facto immunity) but also to physical injuries and the associated economic harms (for which firms presently may be held liable). The scope of immunity thus would be broader than under current law, but it would only be available to companies that take the desired steps to improve their cyberdefenses. (Lawmakers might use the SAFETY Act as a model.[335] The Support Anti-terrorism by Fostering Effective Technologies Act of 2002[336] grants immunity to firms that sell certain anti-terrorism goods and services, so long as they comply with various standards, including a requirement that they carry liability insurance.)

As for the sticks, firms that fail to implement the agreed security measures and then suffer cyberattacks could be exposed to liability for the full range of injuries that result from the intrusions. The severity of the damages could be pegged to the severity of their misconduct, thereby achieving marginal deterrence. So, for instance, a company that fails to adopt the approved security standards might be made to pay compensatory damages or even a smaller fixed sum set by statute, but a company whose conduct is more egregious – one that fails to share information about known vulnerabilities or threats, for instance[337] – might be eligible for exemplary damages. (For inspiration, lawmakers might look to the Gramm-Leach-Bliley Act, which imposes liability on banks that fail to protect consumer data,[338] resulting in relatively robust defenses in the financial services sector against cyber intrusions.) Such a liability regime would increase both a firm's expected benefits of implementing the security protocols that are developed in tandem with regulators, as well as the expected costs of defying them.

Civil liability also would help promote a more robust market for cybersecurity insurance. Insurers can have a profound effect on the steps firms take to secure their systems and products against cyber intrusions, because they can insist that companies implement various security measures as a condition of coverage or charge higher premiums of those that do not.[339] In effect, insurance companies provide a sort of second order regulation, enforcing cybersecurity standards by refusing to bear the losses of firms with poor records or engaging in price discrimination against them. The result is to provide insureds with financial incentives to implement the defenses their insurers are calling for. These incentives have already borne fruit. According to Bruce Schnieier, "[f]irewalls are ubiquitous because auditors started demanding firewalls. This changed the cost equation for businesses. The cost of adding a firewall was expense and user annoyance, but the cost of not having a firewall was failing an audit."[340] Enforcement by

[334] Coldebella & White, *supra* note 14, at 235.

[335] BAKER, *supra* note 24, at 234-35.

[336] 6 U.S.C. §§ 441-44.

[337] *See supra* note 301 and accompanying text.

[338] 15 U.S.C. § 6801(b); *see supra* notes 202 to 210 and accompanying text.

[339] Bamberger, *supra* note 203, at 456; BRENNER, *supra* note 1, at 225; Coyne & Leeson, *supra* note 18, at 491-92; Rosenzweig, *supra* note 14, at 23-24.

[340] Schnieier, *supra* note 33, at 1.

insurers also can decrease the government's enforcement costs; there is less need for regulators to verify that firms are complying with the agreed security standards if insurers, pursuing their own financial interests, are already doing so.

At present, the market for cybersecurity insurance is fairly underdeveloped (though some insurance companies have begun to offer coverage[341]). This is so in large part because firms currently face very little risk of liability for injuries resulting from cyberattacks on their systems or products; why insure when one is effectively immune?[342] The prospect of civil liability is a critical first step in creating a viable market for cybersecurity insurance.[343] Lawmakers might further stimulate the market by offering various kinds of subsidies. For instance, the government might provide insurers with more information (including, perhaps, classified information) about the incidence, prevalence, and consequences of various sorts of malicious code. Insurers could use this data to assess more accurately the probability of cyber intrusions and their potential costs, which would help in setting insurance premiums.[344] Or the government might offer tax benefits to insurers that offer cybersecurity policies. Or it might require certain companies (such as strategically important firms like public utilities or companies that supply goods or services to the government) to carry cybersecurity insurance.

Public health law suggests a final approach to hardening critical infrastructure against cyberattacks. Most states have enacted laws requiring schoolchildren to be vaccinated against various diseases, and lawmakers might adopt similar measures for cyberspace. In both contexts, compulsory inoculation helps reduce negative externalities. In the same way that an unvaccinated child might infect his classmates with a pathogen, a computer system that lacks effective cyberdefenses might be commandeered into a botnet. Compulsory inoculation also helps create positive externalities. A child who has been vaccinated contributes to herd immunity and thereby decreases the probability that other, unvaccinated students will contract the disease. In the same way, companies that adopt effective cyberdefenses make it less likely that their systems will be used to transmit malware to other users.

What would mandatory vaccination look like in cyberspace? Several variants exist. The most coercive approaches involve direct regulation, akin to a requirement that all citizens receive a particular vaccine. One option would be for lawmakers to mandate that every computer user (or, less dramatically, firms in particularly sensitive industries such as the telecommunications sector) install certain security products on their systems (such as antivirus software or firewalls). Think of it as a digital equivalent of the Patient Protection and Affordable Care Act's "individual mandate" to purchase health insurance.[345] An alternative would be for the government to require ISPs to provide their customers with a specified security software package.[346] ISPs presumably

[341] Coyne & Leeson, *supra* note 18, at 491; Yang & Hoffstadt, *supra* note 15, at 208-09.

[342] ABA, *supra* note 18, at 8; BRENNER, *supra* note 1, at 225. Another challenge is that it is difficult for insurers to write policies when – as is often the case with cyberattacks – the probability and consequences of an incident are uncertain.

[343] Rosenzweig, *supra* note 14, at 23.

[344] Coyne & Leeson, *supra* note 18, at 491-92; Frye, *supra* note 144, at 366-67.

[345] 26 U.S.C. § 5000A(a).

[346] CLARKE & KNAKE, *supra* note 1, at 165; Sharp, *supra* note 8, at 25.

would pass on the costs of the software to their subscribers, so the effect would be the same as the individual mandate approach – users would be made to pay a premium for a security product they previously declined to purchase. Or, the government could compensate the ISPs for the costs of making the security package available to their subscribers. In that event, the scheme would represent a (likely regressive) wealth transfer from taxpayers who do not use computers to those who do.

Another, less coercive, set of options would withhold or offer certain benefits to incentivize security improvements; they are the equivalent of making vaccination a condition of eligibility to attend public schools. The ability to access the internet (as opposed to local or proprietary networks) is a valuable benefit of the service one receives from an ISP – for many subscribers it is the most valuable benefit ISPs offer – and it might be conditioned on a subscriber taking steps to improve cyber security. In particular, regulators could direct ISPs to refuse to route users' traffic to the public internet unless they are able to verify that the users have installed specified security software on their systems.[347] Alternatively, government web sites could refuse any traffic sent from a system that has not adopted specified security measures. Users thus would be unable to, for example, post comments in an online rulemaking docket or check the status of a tax refund unless they adopted the security measures. (This sort of measure depends on the ability to authenticate the identity of the sender, as well as the presence of various cyberdefenses on its system. That capability does not presently exist, because the TCP/IP routing protocol is unconcerned with the sender's identity,[348] though some scholars believe an authenticated internet is inevitable.[349]) Finally, the government could offer tax credits or deductions to firms (or individual users) that install the specified security software on their systems – another (likely regressive) wealth transfer.

C. Survivability and Recovery

The third thing an ideal cybersecurity regime would do is promote resilience, thus limiting the amount of damage attackers can do to critical infrastructure. Here, the goals are survivability and recovery, not impregnability.[350] The need to build resilience into the nation's cyberdefenses is based on the hard reality that, no matter how good one's defenses are, some attackers will be able to breach them. As a result, it is not enough to try to prevent attacks altogether. It is also necessary to minimize the amount of harm that the inevitably successful intrusions can do, and to restore victims to the status quo ante as quickly as possible. Public health law offers several strategies for improving resilience. Quarantine and isolation laws, which help limit the spread of infectious diseases during outbreaks, might be adapted for cyberspace. In the event of a cyberattack, systems that become infected with malware might be temporarily disconnected from the internet. Or certain especially sensitive systems (such as the power grid or financial networks) that have not been infected might nevertheless be isolated as a preventive measure. Finally, firms might undertake the cyberspace equivalent of stockpiling

[347] Rattray et al., *supra* note 8, at 160.

[348] *See supra* notes 112 to 113 and accompanying text.

[349] BAKER, *supra* note 24, at 231-32; LESSIG, *supra* note 67, at 45.

[350] *See supra* note 253 and accompanying text.

vaccines and medicines – they might build excess capacity into their systems that can be called into service in emergencies.

In realspace, quarantine and isolation aim at minimizing the harm a pathogen can do; once an outbreak is underway, we want to contain the disease and limit the number of people to whom it can spread. In other words, the objective is to regulate negative externalities. Quarantine and isolation might be adapted for cyberspace – where the goal is to prevent malicious code from infecting more machines – in any number of ways. The most straightforward approach would be for authorities, in the event of a cyberattack, to order systems that are known or suspected to be infected with malware to temporarily disconnect from the internet. While in quarantine, the systems could be inspected to see if they are in fact carrying malicious code. If not, they could be reconnected; if so they could be repaired. The analogy to public health law is fairly exact: Separation of the infected, whether physical or virtual, prevents them spreading the contagion to others and presents an opportunity for treatment. While potentially effective, this approach has a significant drawback – legitimate users will be unable to access the infected system while it is offline. Putting a bank into cyber quarantine doesn't just keep hackers from stealing money, it also keeps a customer from logging on to pay his credit card bill. A less drastic way of preventing the spread of malware would be to isolate *traffic* rather than *systems*. Infected systems would remain connected to the internet, but authorities could use (or require firms to use) deep packet inspection to determine if the data the systems are sending and receiving contains malware. If a given packet is found to be carrying malicious code, it could be interdicted; if not, it would be allowed to continue on its way. The public health analogy is allowing a man infected with SARS to leave an isolation facility and go about his business, while wearing a surgical mask that intercepts the respiratory droplets through which the virus is spread. The virtue of this finer grained variant is that it allows legitimate users to continue to access an infected system even as attackers are prevented from using it for their malign purposes; the hackers are thwarted, but customers can still access their accounts (although perhaps a bit more slowly than usual). On the other hand, traffic quarantines will only be as effective as the packet sniffers and malware signature files on which they rely, and sophisticated adversaries might be able to defeat both.

Another, more controversial set of options involves preventive quarantine – separating systems that have not been infected but that are vulnerable. This approach would turn public health law on its head; rather than isolating the sick, authorities would isolate the healthy. The most aggressive variant would be for regulators to require a select group of strategically significant firms, such as the power grid, financial institutions, telecommunications carriers, to temporarily disconnect from the internet if a cyberattack takes place.[351] (Senator Nelson Rockefeller introduced legislation along these lines in 2009[352]; critics denounced it as an "internet kill switch."[353]) Preventive quarantine would be a fairly effective way of preventing malware from spreading to critical infrastructure – a system that isn't on the internet cannot contract a virus that spreads via the internet. But it wouldn't be infallible. Even "air gapped"

[351] BRENNER, *supra* note 1, at 234; CLARKE & KNAKE, *supra* note 1, at 167; Picker, *supra* note 42, at 126.

[352] Cybersecurity Act of 2010, S. 773, 111th Cong. (2009).

[353] Mark Gibbs, *The Internet Kill Switch*, NETWORK WORLD, Apr. 13, 2009.

systems – those that are physically separated from the internet[354] – are vulnerable to infection via USB devices and other removable media.[355] A disconnection requirement also could prove quite costly: The affected systems would be unavailable to legitimate users for as long as the order remained in effect. There is also a risk that regulators might pull the disconnection trigger too readily. As an alternative to a strict disconnection requirement, regulators might direct firms to implement security countermeasures of their own devising. (Senator Joseph Lieberman introduced legislation along these lines in 2010[356]; it likewise was denounced as an internet kill switch.[357]) Whatever the content of these security protocols – encrypting data to prevent its theft, for instance, or requiring users to authenticate themselves before gaining access to the system – they might be established through the collaborative regulatory partnership described above.[358] An even more modest version of preventive quarantine would be, as above, to segregate traffic rather than entire systems. In the event of a cyberattack, packet sniffers might be used to inspect all traffic that is sent to and from designated systems. This would allow the systems to continue to operate more or less as usual, though perhaps at a cost of less security.

Another important goal is to ensure that critical systems are able to continue functioning during a cyberattack and recover quickly thereafter. One way to achieve this is to build systems with excess capacity – i.e., to include more capabilities than a firm needs for its day to day operations, but that can be held in reserve and called into service if an attack takes place.[359] In particular, regulators might require certain companies to build their systems with excess bandwidth. A "strategic reserve of bandwidth" is an especially useful countermeasure for defending against denial of service attacks[360]; if a company's servers are being overwhelmed, the reserve bandwidth can be brought into service to process the requests. Regulators also might require certain companies to maintain redundant data storage capabilities. These firms might routinely back up their data to servers that are dispersed, both geographically and in network terms. If a cyberattack corrupted their systems, it would be relatively easy to wipe them clean and restore the data from an uncorrupted backup.[361] An attacker thus might succeed in taking down one site "only to find that the same content continues to appear through other servers. This is like playing electronic Whac-A-Mole on a global scale . . . ."[362] These sorts of measures can be thought of as akin to the public health practice of stockpiling medicines and vaccines for use in a crisis. The CDC may not need 300 million doses of smallpox vaccine in its everyday operations, but they would prove critical in the event of an outbreak.

---

[354] BRENNER, *supra* note 1, at 84; Ellen Nakashima, *Cyber-Intruder Sparks Massive Federal Response*, WASH. POST, Dec. 8, 2011.

[355] BAKER, *supra* note 24, at 216; Baker, *supra* note 29, at 3; BRENNER, *supra* note 1, at 61; CLARKE & KNAKE, *supra* note 1, at 127; Nakashima, *Cyber-Intruder*, *supra* note 354.

[356] Protecting Cyberspace as a National Asset Act of 2010, S. 3480, 111th Cong. (2010).

[357] Adam Cohen, *What's Missing in the Internet Kill-Switch Debate*, TIME, Aug. 11, 2010.

[358] *See supra* Part III.B.

[359] Benkler, *supra* note 253, at 75.

[360] Taipale, *supra* note 92, at 37.

[361] Bambauer, *supra* note 12, at 637; Sarnikar & Johnsen, *supra* note 19, at 2, 21; Taipale, *supra* note 92, at 38.

[362] BRENNER, *supra* note 1, at 179.

Excess capacity can be expensive; requiring firms to keep reserves of (largely unused) bandwidth costs money, and "[h]aving information located in multiple places makes it more costly to maintain."[363] How to pay for these requirements? One obvious option would be for companies to pass their costs of complying with resilience mandates to their customers in the form of price increases, service decreases, or both. A difficulty with this approach is that improving a given company's ability to withstand an attack does not just confer benefits on its customers. It also confers benefits on those with whom the company has no relationship; if Citibank can continue to operate notwithstanding a DDOS, its customers will still be able to pay their bills, and third party vendors will still be able to receive payments. Excess capacity thus creates positive externalities, and the customers who pay higher prices for excess capacity are effectively subsidizing the third parties. Another option would be for the government to offer various subsidies to firms that are subject to survivability mandates. This approach is based on a recognition that excess capacity is, in a sense, a public good that the market will tend to undersupply.[364] In part because excess capacity requirements can be costly, it wouldn't be advisable for regulators to apply them to all firms in the marketplace.

D. Responding to Cyberattacks

The fourth and final component of an effective cybersecurity regime is responding to individuals, groups, and states that have committed cyberattacks. This topic is exhaustively covered in the existing literature, and it naturally lends itself to analysis under the law enforcement and armed conflict approaches to cybersecurity. For instance, scholars have proposed better international cooperation on cybercrime investigations, increasing the penalties for certain computer related offenses, increasing the costs that perpetrators must bear to commit cybercrimes, treating intrusions as "armed attacks" that trigger the right to self defense under the United Nations Charter, treating cyberattacks as acts of aggression that justify retaliating with conventional military force, and so on.[365] This article does not seek to add to this already voluminous literature. There is, however, one type of response that deserves a brief mention: active self defense, or "hackbacks." Like the regulatory solutions discussed above, hackbacks can incentivize firms to better secure their systems against cyber intruders.

A hackback is an in-kind response to a cyberattack. The victim essentially mounts a counterattack against the assailant, "shutting down the attack before it can do further harm and/or damaging the perpetrator's system to stop it from launching future attacks."[366] This might be accomplished in several ways. If a victim detects that it is experiencing a cyberattack, it might direct a flood of traffic to the servers through which the attack is being routed, temporarily overwhelming them and preventing them from continuing the intrusion.[367] Or it might hack into

[363] Bambauer, *supra* note 12, at 637.

[364] *See supra* notes 129 to 135 and accompanying text.

[365] *See supra* Part II.A.

[366] Sklerov, *supra* note 19, at 25; *see also* Condron, *supra* note 19, at 410-11; O'Neill, *supra* note 19, at 280.

[367] Condron, *supra* note 19, at 410-11.

the responsible servers, taking control of them or damaging them.[368] Some scholars believe that hackbacks are the most effective defense against cyberattacks.[369] This is so in part because active self defense can avoid the attribution problem; a victim firm that is experiencing an intrusion could retaliate against any computer that is attacking it without knowing who is behind the incident or his purposes.[370] (Needless to say, active self defense is only possible if the victim is aware that it is under attack. Hackback will not be an option if, as is sometimes the case, the intrusion goes undetected.)

Active self defense fits into the law enforcement framework fairly comfortably. Although hackbacks are probably illegal under the Computer Fraud and Abuse Act[371] – the victims are, after all, perpetrating cyber intrusions of their own – fundamental principles of criminal law can explain why they might be acceptable. The basic idea is justification. Conduct that ordinarily is condemned can become permissible, or even desirable, in certain circumstances.[372] Homicide is typically illegal, as society deems it blameworthy to take the life of another human being, but we are allowed to use deadly force against those who pose a threat to our lives or the lives of others. The same might be said of hackbacks. Society ordinarily condemns those who break into others' computers, but one might be justified in hacking an intruder's machine to protect one's own system.[373] Hackbacks also might be described in armed conflict terms, though perhaps not as neatly. Scholars have extensively debated whether the law of armed conflict allows a government to engage in active self defense against those who have launched cyberattacks against its computer systems.[374] A related question is whether the LOAC allows *private individuals* to respond in kind when their systems are attacked. Initially one might think the answer is no; hackbacks by civilians against foreign adversaries can be seen as a use of force by noncombatants in violation of the LOAC.[375] However, the LOAC allows noncombatants to participate in hostilities in certain circumstances. *Levee en masse* refers to the right of a civilian population to resist invaders by force, provided that they organize themselves into units, carry arms openly, and otherwise obey the laws of war.[376] Active self defense against

[368] Smith, *supra* note 17, at 177-78.

[369] O'Neill, *supra* note 19, at 240, 280; Sklerov, *supra* note 19, at 25 & n.160; *cf.* Richard A. Epstein, *The Theory and Practice of Self-Help*, 1 J.L. ECON. & POL'Y 1, 30 (2005) (emphasizing the need for "self-help remedies").

[370] Condron, *supra* note 19, at 415-16; Jensen, *Computer Attacks*, *supra* note 19, at 232. *See generally supra* notes 110 to 113 and accompanying text.

[371] ABA, *supra* note 18, at 18; BAKER, *supra* note 24, at 212; CLARKE & KNAKE, *supra* note 1, at 214; Smith, *supra* note 17, at 180, 182.

[372] *See generally* Dressler, 33 WAYNE L. REV. 1155, __ (1987).

[373] Katyal, *Community*, *supra* note 127, at 61; O'Neill, *supra* note 19, at 280; Smith, *supra* note 17, at 190-91. *But see* Susan W. Brenner, *"At Light Speed"*, 97 J. CRIM. L. & CRIMINOLOGY 379, 448 (2007) (condemning active self defense as "vigilantism"); Orin S. Kerr, *Virtual Crime, Virtual Deterrence*, 1 J.L. ECON. & POL'Y 197, 204 (2005) (same).

[374] Condron, *supra* note 19, at 416; Sklerov, *supra* note 19, at 11-13; Taipale, *supra* note 92, at 35; Terry, *supra* note 109, at 424.

[375] Brenner, *supra* note 373, at 441; Watts, *supra* note 44, at 395-96, 424-25.

[376] Third Geneva Convention art. 4A(6); *see also* Watts, *supra* note 44, at 435. *Cf.* Rabkin & Rabkin, at 11-12 (analogizing private citizens who conduct cyber intrusions with a state's blessing to privateers who operate under letters of marquee).

cyberattacks might be thought of as a digital (though far less organized) version of the *levee en masse*. Just as civilians are entitled to use pistols and rifles to repel invading armies, they also might be entitled to use hackers' tools to stop attacks by foreign governments or groups – especially when those attacks are directed at their systems.[377]

Active self defense is controversial, but it offers one potential benefit that has been largely overlooked in the literature. Like the other regulatory solutions discussed in this article, hackbacks can incentivize firms to improve the security of their systems. Cyber perpetrators typically do not launch attacks directly; to obscure their responsibility, they usually route an attack through a chain of unsecured intermediary systems before reaching the ultimate target.[378] If a victim firm responds to an intrusion with active self defense, it is likely that these third party systems will suffer harm.[379] (The realspace equivalent would be a driver who leaves his car unlocked; the car is then stolen by bank robbers and destroyed when the thieves open fire and the bank's security guards shoot back.) Many scholars regard this third party problem as a sufficient reason to forbid hackbacks.[380] Yet the prospect of damage to third parties may actually be desirable. The threat of harm provides firms with incentives to better secure their systems and prevent them from being used as conduits for attacks on others. Suppose Citibank knows that, if attackers gain control of its computers and use them to conduct DDOS attacks, the victims will be allowed to retaliate against Citibank's machines. Citibank will have a fairly strong incentive to ensure that its computers are not commandeered into botnets. Damage from hackbacks thus would internalize some of the costs that third parties impose on others by maintaining insecure systems.[381] (Likewise in realspace. If drivers know that security guards are allowed to damage getaway cars even if they are stolen, they will lock their doors.) Active self defense also might weaken attackers' incentives to commit cyberattacks. If assailants know that victims will be able to use hackbacks to render their attacks ineffective, or less effective, they will have less reason to undertake them in the first place. By increasing the futility of intrusions, hackbacks can help achieve deterrence.[382] Active self defense thus can simultaneously foster favorable incentives to improve security and weaken unfavorable incentives to commit attacks.[383]

---

[377] *But see* Davis Brown, *supra* note 13, at 191-92.

[378] *See supra* note 113 and accompanying text.

[379] Epstein, *supra* note 369, at 31; Katyal at 62-63; Kerr, *supra* note 373, at 205; Smith, *supra* note 17, at 180.

[380] Katyal, *Community*, *supra* note 127, at 60-66; Kerr, *supra* note 373, at 205-06; Bruce Schnieier, *Counterattack*, CRYPTO-GRAM NEWSLETTER (Dec. 15, 2002).

[381] *Cf.* Picker, *supra* note 42, at 116, 136.

[382] O'Neill, *supra* note 19, at 280; Sklerov, *supra* note 19, at 10. *See generally supra* note 254 and accompanying text.

[383] To be sure, active self defense might foster negative incentives. As Orin Kerr points out, allowing hackbacks "would create an obvious incentive for attackers to be extra careful to disguise their location or use someone else's computer to launch the attack." Kerr, *supra* note 373, at 205. Allowing hackbacks also would "encourage foul play designed to harness the new privileges"; one example is the "bankshot attack," in which an assailant who wants a computer to be attacked "can route attacks through that one computer toward a series of victims, and then wait for the victims to attack back at that computer." Kerr, *supra* note 373, at 205-06; *see also* Katyal, *Community*, *supra* note 127, at 63. It cannot be predicted *a priori* whether the harmful conduct produced by these negative incentives would be greater or lesser than the beneficial conduct produced by the positive incentives.

CONCLUSION

Cyberthreats aren't going away. As society increasingly comes to rely on networked critical infrastructure such as banks and the power grid, assailants will find that they have ever more to gain by attacking these digital assets. And we will find that we have ever more to lose.

It therefore becomes essential to think about cybersecurity using an analytical framework that is rich enough to account for the problem in all its complexity. Cybersecurity is too important, and too intricate, to leave to the criminal law and the law of armed conflict. Instead, as this article has proposed, an entirely new conceptual approach is needed – an approach that can account for the systematic tendency of many private firms to underinvest in cyberdefense. Companies sometimes fail to secure their systems against attackers because they do not bear the full costs of the resulting intrusions; the harms are partially externalized onto third parties. Firms also tend to neglect cybersecurity because by improving their own defenses they contribute to the security of others' systems; the benefits are partially externalized, which creates opportunities for free riding. If these problems sound familiar, that's because they are. These challenges of negative externalities, positive externalities, and free riding are similar to challenges that the modern administrative state encounters in a number of other settings. Cybersecurity thus resembles the problems that arise in environmental law, antitrust law, products liability law, and public health law. Scholars and lawmakers might look to these other fields for suggestions on how to incentivize private firms to improve their defenses; conceiving of cybersecurity in regulatory terms opens the door to regulatory solutions.

Of course, "regulatory solutions" need not mean "command and control solutions." Often it will be possible to promote better cybersecurity by appealing to firms' self interest – encouraging them to improve their defenses by immunizing them from liability or offering other subsidies, not just sanctioning them when they fail to do so. For instance, rather than empowering a central regulator to monitor the internet for outbreaks of malicious code, companies should use something like public health law's distributed biosurveillance network to collect and share information about cyberthreats with one another. Similarly, the private sector should play an active role in establishing industrywide cybersecurity standards, as it frequently does in environmental law and other regulatory contexts. Offers of immunity and threats of liability then would be used to encourage companies to adopt the agreed upon standards. As for improving the ability of critical systems to survive intrusions, infected computers could be temporarily disconnected from the internet to keep them from spreading the malware, and companies should be encouraged to build their systems with excess capacity (such as reserve bandwidth and remote backups) that can be called into service during cyberattacks. Finally, lawmakers might loosen the restrictions on "hackbacks," to incentivize firms to protect their systems from being commandeered into attacks on third parties.

Virtually no one is happy with the state of America's cyberdefenses, and scholars have felled entire forests exploring how to prosecute cyber criminals more effectively or retaliate against countries that launch cyberattacks. Maybe we've been asking the wrong questions. Maybe what we need to secure cyberspace isn't cops, spies, or soldiers. Maybe what we need is administrative law.

Syllabus

NOTE: Where it is feasible, a syllabus (headnote) will be released, as is being done in connection with this case, at the time the opinion is issued. The syllabus constitutes no part of the opinion of the Court but has been prepared by the Reporter of Decisions for the convenience of the reader. See *United States* v. *Detroit Timber & Lumber Co.,* 200 U. S. 321, 337.

SUPREME COURT OF THE UNITED STATES

Syllabus

CITY OF ONTARIO, CALIFORNIA, ET AL. *v.* QUON ET AL.

CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

No. 08–1332. Argued April 19, 2010—Decided June 17, 2010

Petitioner Ontario (hereinafter City) acquired alphanumeric pagers able to send and receive text messages. Its contract with its service provider, Arch Wireless, provided for a monthly limit on the number of characters each pager could send or receive, and specified that usage exceeding that number would result in an additional fee. The City issued the pagers to respondent Quon and other officers in its police department (OPD), also a petitioner here. When Quon and others exceeded their monthly character limits for several months running, petitioner Scharf, OPD's chief, sought to determine whether the existing limit was too low, *i.e.,* whether the officers had to pay fees for sending work-related messages or, conversely, whether the overages were for personal messages. After Arch Wireless provided transcripts of Quon's and another employee's August and September 2002 text messages, it was discovered that many of Quon's messages were not work related, and some were sexually explicit. Scharf referred the matter to OPD's internal affairs division. The investigating officer used Quon's work schedule to redact from his transcript any messages he sent while off duty, but the transcript showed that few of his on-duty messages related to police business. Quon was disciplined for violating OPD rules.

He and the other respondents—each of whom had exchanged text messages with Quon during August and September—filed this suit, alleging, *inter alia*, that petitioners violated their Fourth Amendment rights and the federal Stored Communications Act (SCA) by obtaining and reviewing the transcript of Quon's pager messages, and that Arch Wireless violated the SCA by giving the City the transcript. The District Court denied respondents summary judgment on the

constitutional claims, relying on the plurality opinion in *O'Connor* v. *Ortega*, 480 U. S. 709, to determine that Quon had a reasonable expectation of privacy in the content of his messages. Whether the audit was nonetheless reasonable, the court concluded, turned on whether Scharf used it for the improper purpose of determining if Quon was using his pager to waste time, or for the legitimate purpose of determining the efficacy of existing character limits to ensure that officers were not paying hidden work-related costs. After the jury concluded that Scharf's intent was legitimate, the court granted petitioners summary judgment on the ground they did not violate the Fourth Amendment. The Ninth Circuit reversed. Although it agreed that Quon had a reasonable expectation of privacy in his text messages, the appeals court concluded that the search was not reasonable even though it was conducted on a legitimate, work-related rationale. The opinion pointed to a host of means less intrusive than the audit that Scharf could have used. The court further concluded that Arch Wireless had violated the SCA by giving the City the transcript.

Held: Because the search of Quon's text messages was reasonable, petitioners did not violate respondents' Fourth Amendment rights, and the Ninth Circuit erred by concluding otherwise. Pp. 7–17.

(a) The Amendment guarantees a person's privacy, dignity, and security against arbitrary and invasive governmental acts, without regard to whether the government actor is investigating crime or performing another function. *Skinner* v. *Railway Labor Executives' Assn.*, 489 U. S. 602, 613–614. It applies as well when the government acts in its capacity as an employer. *Treasury Employees* v. *Von Raab*, 489 U. S. 656, 665. The Members of the *O'Connor* Court disagreed on the proper analytical framework for Fourth Amendment claims against government employers. A four-Justice plurality concluded that the correct analysis has two steps. First, because "some [government] offices may be so open . . . that no expectation of privacy is reasonable," a court must consider "[t]he operational realities of the workplace" to determine if an employee's constitutional rights are implicated. 480 U. S., at 718. Second, where an employee has a legitimate privacy expectation, an employer's intrusion on that expectation "for noninvestigatory, work-related purposes, as well as for investigations of work-related misconduct, should be judged by the standard of reasonableness under all the circumstances." *Id.,* at 725–726. JUSTICE SCALIA, concurring in the judgment, would have dispensed with the "operational realities" inquiry and concluded "that the offices of government employees . . . are [generally] covered by Fourth Amendment protections," *id.,* at 731, but he would also have held "that government searches to retrieve work-related materials or

to investigate violations of workplace rules—searches of the sort that are regarded as reasonable and normal in the private-employer context—do not violate the . . . Amendment," *id.,* at 732. Pp. 7–9.

(b) Even assuming that Quon had a reasonable expectation of privacy in his text messages, the search was reasonable under both *O'Connor* approaches, the plurality's and JUSTICE SCALIA's. Pp. 9–17.

(1) The Court does not resolve the parties' disagreement over Quon's privacy expectation. Prudence counsels caution before the facts in this case are used to establish far-reaching premises that define the existence, and extent, of privacy expectations of employees using employer-provided communication devices. Rapid changes in the dynamics of communication and information transmission are evident not just in the technology itself but in what society accepts as proper behavior. At present, it is uncertain how workplace norms, and the law's treatment of them, will evolve. Because it is therefore preferable to dispose of this case on narrower grounds, the Court assumes, *arguendo,* that: (1) Quon had a reasonable privacy expectation; (2) petitioners' review of the transcript constituted a Fourth Amendment search; and (3) the principles applicable to a government employer's search of an employee's physical office apply as well in the electronic sphere. Pp. 9–12.

(2) Petitioners' warrantless review of Quon's pager transcript was reasonable under the *O'Connor* plurality's approach because it was motivated by a legitimate work-related purpose, and because it was not excessive in scope. See 480 U. S., at 726. There were "reasonable grounds for [finding it] necessary for a noninvestigatory work-related purpose," *ibid.,* in that Chief Scharf had ordered the audit to determine whether the City's contractual character limit was sufficient to meet the City's needs. It was also "reasonably related to the objectives of the search," *ibid.,* because both the City and OPD had a legitimate interest in ensuring that employees were not being forced to pay out of their own pockets for work-related expenses, or, on the other hand, that the City was not paying for extensive personal communications. Reviewing the transcripts was an efficient and expedient way to determine whether either of these factors caused Quon's overages. And the review was also not "excessively intrusive." *Ibid.* Although Quon had exceeded his monthly allotment a number of times, OPD requested transcripts for only August and September 2002 in order to obtain a large enough sample to decide the character limits' efficaciousness, and all the messages that Quon sent while off duty were redacted. And from OPD's perspective, the fact that Quon likely had only a limited privacy expectation lessened the risk that the review would intrude on highly private details of Quon's life. Similarly, because the City had a legitimate reason for the search

Syllabus

and it was not excessively intrusive in light of that justification, the search would be "regarded as reasonable and normal in the private-employer context" and thereby satisfy the approach of JUSTICE SCALIA's concurrence, *id.,* at 732. Conversely, the Ninth Circuit's "least intrusive" means approach was inconsistent with controlling precedents. See*, e.g., Vernonia School Dist. 47J* v. *Acton*, 515 U. S. 646, 663. Pp. 12–16.

(c) Whether the other respondents can have a reasonable expectation of privacy in their text messages to Quon need not be resolved. They argue that because the search was unreasonable as to Quon, it was also unreasonable as to them, but they make no corollary argument that the search, if reasonable as to Quon, could nonetheless be unreasonable as to them. Given this litigating position and the Court's conclusion that the search was reasonable as to Quon, these other respondents cannot prevail. Pp. 16–17.

529 F. 3d 892, reversed and remanded.

KENNEDY, J., delivered the opinion of the Court, in which ROBERTS, C. J., and STEVENS, THOMAS, GINSBURG, BREYER, ALITO, and SOTOMAYOR, JJ., joined, and in which SCALIA, J., joined except for Part III–A. STEVENS, J., filed a concurring opinion. SCALIA, J., filed an opinion concurring in part and concurring in the judgment.

SUPREME COURT OF THE UNITED STATES

No. 08–1332

CITY OF ONTARIO, CALIFORNIA, ET AL.,
PETITIONERS *v.* JEFF QUON ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

[June 17, 2010]

JUSTICE KENNEDY delivered the opinion of the Court.

This case involves the assertion by a government employer of the right, in circumstances to be described, to read text messages sent and received on a pager the employer owned and issued to an employee. The employee contends that the privacy of the messages is protected by the ban on "unreasonable searches and seizures" found in the Fourth Amendment to the United States Constitution, made applicable to the States by the Due Process Clause of the Fourteenth Amendment. *Mapp* v. *Ohio*, 367 U. S. 643 (1961). Though the case touches issues of far-reaching significance, the Court concludes it can be resolved by settled principles determining when a search is reasonable.

I

A

The City of Ontario (City) is a political subdivision of the State of California. The case arose out of incidents in 2001 and 2002 when respondent Jeff Quon was employed by the Ontario Police Department (OPD). He was a police ser-

geant and member of OPD's Special Weapons and Tactics (SWAT) Team. The City, OPD, and OPD's Chief, Lloyd Scharf, are petitioners here. As will be discussed, two respondents share the last name Quon. In this opinion "Quon" refers to Jeff Quon, for the relevant events mostly revolve around him.

In October 2001, the City acquired 20 alphanumeric pagers capable of sending and receiving text messages. Arch Wireless Operating Company provided wireless service for the pagers. Under the City's service contract with Arch Wireless, each pager was allotted a limited number of characters sent or received each month. Usage in excess of that amount would result in an additional fee. The City issued pagers to Quon and other SWAT Team members in order to help the SWAT Team mobilize and respond to emergency situations.

Before acquiring the pagers, the City announced a "Computer Usage, Internet and E-Mail Policy" (Computer Policy) that applied to all employees. Among other provisions, it specified that the City "reserves the right to monitor and log all network activity including e-mail and Internet use, with or without notice. Users should have no expectation of privacy or confidentiality when using these resources." App. to Pet. for Cert. 152a. In March 2000, Quon signed a statement acknowledging that he had read and understood the Computer Policy.

The Computer Policy did not apply, on its face, to text messaging. Text messages share similarities with e-mails, but the two differ in an important way. In this case, for instance, an e-mail sent on a City computer was transmitted through the City's own data servers, but a text message sent on one of the City's pagers was transmitted using wireless radio frequencies from an individual pager to a receiving station owned by Arch Wireless. It was routed through Arch Wireless' computer network, where it remained until the recipient's pager or cellular telephone

was ready to receive the message, at which point Arch Wireless transmitted the message from the transmitting station nearest to the recipient. After delivery, Arch Wireless retained a copy on its computer servers. The message did not pass through computers owned by the City.

Although the Computer Policy did not cover text messages by its explicit terms, the City made clear to employees, including Quon, that the City would treat text messages the same way as it treated e-mails. At an April 18, 2002, staff meeting at which Quon was present, Lieutenant Steven Duke, the OPD officer responsible for the City's contract with Arch Wireless, told officers that messages sent on the pagers "are considered e-mail messages. This means that [text] messages would fall under the City's policy as public information and [would be] eligible for auditing." App. 30. Duke's comments were put in writing in a memorandum sent on April 29, 2002, by Chief Scharf to Quon and other City personnel.

Within the first or second billing cycle after the pagers were distributed, Quon exceeded his monthly text message character allotment. Duke told Quon about the overage, and reminded him that messages sent on the pagers were "considered e-mail and could be audited." *Id.,* at 40. Duke said, however, that "it was not his intent to audit [an] employee's text messages to see if the overage [was] due to work related transmissions." *Ibid.* Duke suggested that Quon could reimburse the City for the overage fee rather than have Duke audit the messages. Quon wrote a check to the City for the overage. Duke offered the same arrangement to other employees who incurred overage fees.

Over the next few months, Quon exceeded his character limit three or four times. Each time he reimbursed the City. Quon and another officer again incurred overage fees for their pager usage in August 2002. At a meeting in October, Duke told Scharf that he had become "'tired of

being a bill collector.'" *Id.,* at 91. Scharf decided to determine whether the existing character limit was too low—that is, whether officers such as Quon were having to pay fees for sending work-related messages—or if the overages were for personal messages. Scharf told Duke to request transcripts of text messages sent in August and September by Quon and the other employee who had exceeded the character allowance.

At Duke's request, an administrative assistant employed by OPD contacted Arch Wireless. After verifying that the City was the subscriber on the accounts, Arch Wireless provided the desired transcripts. Duke reviewed the transcripts and discovered that many of the messages sent and received on Quon's pager were not work related, and some were sexually explicit. Duke reported his findings to Scharf, who, along with Quon's immediate supervisor, reviewed the transcripts himself. After his review, Scharf referred the matter to OPD's internal affairs division for an investigation into whether Quon was violating OPD rules by pursuing personal matters while on duty.

The officer in charge of the internal affairs review was Sergeant Patrick McMahon. Before conducting a review, McMahon used Quon's work schedule to redact the transcripts in order to eliminate any messages Quon sent while off duty. He then reviewed the content of the messages Quon sent during work hours. McMahon's report noted that Quon sent or received 456 messages during work hours in the month of August 2002, of which no more than 57 were work related; he sent as many as 80 messages during a single day at work; and on an average workday, Quon sent or received 28 messages, of which only 3 were related to police business. The report concluded that Quon had violated OPD rules. Quon was allegedly disciplined.

B

Raising claims under Rev. Stat. §1979, 42 U. S. C. §1983; 18 U. S. C. §2701 *et seq.*, popularly known as the Stored Communications Act (SCA); and California law, Quon filed suit against petitioners in the United States District Court for the Central District of California. Arch Wireless and an individual not relevant here were also named as defendants. Quon was joined in his suit by another plaintiff who is not a party before this Court and by the other respondents, each of whom exchanged text messages with Quon during August and September 2002: Jerilyn Quon, Jeff Quon's then-wife, from whom he was separated; April Florio, an OPD employee with whom Jeff Quon was romantically involved; and Steve Trujillo, another member of the OPD SWAT Team. Among the allegations in the complaint was that petitioners violated respondents' Fourth Amendment rights and the SCA by obtaining and reviewing the transcript of Jeff Quon's pager messages and that Arch Wireless had violated the SCA by turning over the transcript to the City.

The parties filed cross-motions for summary judgment. The District Court granted Arch Wireless' motion for summary judgment on the SCA claim but denied petitioners' motion for summary judgment on the Fourth Amendment claims. *Quon* v. *Arch Wireless Operating Co.,* 445 F. Supp. 2d 1116 (CD Cal. 2006). Relying on the plurality opinion in *O'Connor* v. *Ortega*, 480 U. S. 709, 711 (1987), the District Court determined that Quon had a reasonable expectation of privacy in the content of his text messages. Whether the audit of the text messages was nonetheless reasonable, the District Court concluded, turned on Chief Scharf's intent: "[I]f the purpose for the audit was to determine if Quon was using his pager to 'play games' and 'waste time,' then the audit was not constitutionally reasonable"; but if the audit's purpose "was to determine the efficacy of the existing character

limits to ensure that officers were not paying hidden work-related costs, . . . no constitutional violation occurred." 445 F. Supp. 2d, at 1146.

The District Court held a jury trial to determine the purpose of the audit. The jury concluded that Scharf ordered the audit to determine the efficacy of the character limits. The District Court accordingly held that petitioners did not violate the Fourth Amendment. It entered judgment in their favor.

The United States Court of Appeals for the Ninth Circuit reversed in part. 529 F. 3d 892 (2008). The panel agreed with the District Court that Jeff Quon had a reasonable expectation of privacy in his text messages but disagreed with the District Court about whether the search was reasonable. Even though the search was conducted for "a legitimate work-related rationale," the Court of Appeals concluded, it "was not reasonable in scope." *Id.,* at 908. The panel disagreed with the District Court's observation that "there were no less-intrusive means" that Chief Scharf could have used "to verify the efficacy of the 25,000 character limit . . . without intruding on [respondents'] Fourth Amendment rights." *Id.,* at 908–909. The opinion pointed to a "host of simple ways" that the chief could have used instead of the audit, such as warning Quon at the beginning of the month that his future messages would be audited, or asking Quon himself to redact the transcript of his messages. *Id.,* at 909. The Court of Appeals further concluded that Arch Wireless had violated the SCA by turning over the transcript to the City.

The Ninth Circuit denied a petition for rehearing en banc. *Quon* v. *Arch Wireless Operating Co.,* 554 F. 3d 769 (2009). Judge Ikuta, joined by six other Circuit Judges, dissented. *Id.,* at 774–779. Judge Wardlaw concurred in the denial of rehearing, defending the panel's opinion against the dissent. *Id.,* at 769–774.

This Court granted the petition for certiorari filed by the City, OPD, and Chief Scharf challenging the Court of Appeals' holding that they violated the Fourth Amendment. 558 U. S. ___ (2009). The petition for certiorari filed by Arch Wireless challenging the Ninth Circuit's ruling that Arch Wireless violated the SCA was denied. *USA Mobility Wireless, Inc.* v. *Quon,* 558 U. S. ___ (2009).

II

The Fourth Amendment states: "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated . . . ." It is well settled that the Fourth Amendment's protection extends beyond the sphere of criminal investigations. *Camara* v. *Municipal Court of City and County of San Francisco*, 387 U. S. 523, 530 (1967). "The Amendment guarantees the privacy, dignity, and security of persons against certain arbitrary and invasive acts by officers of the Government," without regard to whether the government actor is investigating crime or performing another function. *Skinner* v. *Railway Labor Executives' Assn.*, 489 U. S. 602, 613–614 (1989). The Fourth Amendment applies as well when the Government acts in its capacity as an employer. *Treasury Employees* v. *Von Raab*, 489 U. S. 656, 665 (1989).

The Court discussed this principle in *O'Connor*. There a physician employed by a state hospital alleged that hospital officials investigating workplace misconduct had violated his Fourth Amendment rights by searching his office and seizing personal items from his desk and filing cabinet. All Members of the Court agreed with the general principle that "[i]ndividuals do not lose Fourth Amendment rights merely because they work for the government instead of a private employer." 480 U. S., at 717 (plurality opinion); see also *id.,* at 731 (SCALIA, J., concurring in judgment); *id.,* at 737 (Blackmun, J., dissenting). A major-

ity of the Court further agreed that "'special needs, beyond the normal need for law enforcement,'" make the warrant and probable-cause requirement impracticable for government employers. *Id.,* at 725 (plurality opinion) (quoting *New Jersey* v. *T. L. O.*, 469 U. S. 325, 351 (1985) (Blackmun, J., concurring in judgment); 480 U. S., at 732 (opinion of SCALIA, J.) (quoting same).

The *O'Connor* Court did disagree on the proper analytical framework for Fourth Amendment claims against government employers. A four-Justice plurality concluded that the correct analysis has two steps. First, because "some government offices may be so open to fellow employees or the public that no expectation of privacy is reasonable," *id.,* at 718, a court must consider "[t]he operational realities of the workplace" in order to determine whether an employee's Fourth Amendment rights are implicated, *id.,* at 717. On this view, "the question whether an employee has a reasonable expectation of privacy must be addressed on a case-by-case basis." *Id.,* at 718. Next, where an employee has a legitimate privacy expectation, an employer's intrusion on that expectation "for noninvestigatory, work-related purposes, as well as for investigations of work-related misconduct, should be judged by the standard of reasonableness under all the circumstances." *Id.,* at 725–726.

JUSTICE SCALIA, concurring in the judgment, outlined a different approach. His opinion would have dispensed with an inquiry into "operational realities" and would conclude "that the offices of government employees . . . are covered by Fourth Amendment protections as a general matter." *Id.,* at 731. But he would also have held "that government searches to retrieve work-related materials or to investigate violations of workplace rules—searches of the sort that are regarded as reasonable and normal in the private-employer context—do not violate the Fourth Amendment." *Id.,* at 732.

Later, in the *Von Raab* decision, the Court explained that "operational realities" could diminish an employee's privacy expectations, and that this diminution could be taken into consideration when assessing the reasonableness of a workplace search. 489 U. S., at 671. In the two decades since *O'Connor*, however, the threshold test for determining the scope of an employee's Fourth Amendment rights has not been clarified further. Here, though they disagree on whether Quon had a reasonable expectation of privacy, both petitioners and respondents start from the premise that the *O'Connor* plurality controls. See Brief for Petitioners 22–28; Brief for Respondents 25–32. It is not necessary to resolve whether that premise is correct. The case can be decided by determining that the search was reasonable even assuming Quon had a reasonable expectation of privacy. The two *O'Connor* approaches—the plurality's and JUSTICE SCALIA's—therefore lead to the same result here.

III

A

Before turning to the reasonableness of the search, it is instructive to note the parties' disagreement over whether Quon had a reasonable expectation of privacy. The record does establish that OPD, at the outset, made it clear that pager messages were not considered private. The City's Computer Policy stated that "[u]sers should have no expectation of privacy or confidentiality when using" City computers. App. to Pet. for Cert. 152a. Chief Scharf's memo and Duke's statements made clear that this official policy extended to text messaging. The disagreement, at least as respondents see the case, is over whether Duke's later statements overrode the official policy. Respondents contend that because Duke told Quon that an audit would be unnecessary if Quon paid for the overage, Quon reasonably could expect that the contents of his messages

would remain private.

At this point, were we to assume that inquiry into "operational realities" were called for, compare *O'Connor*, 480 U. S., at 717 (plurality opinion), with *id.,* at 730–731 (opinion of SCALIA, J.); see also *id.,* at 737–738 (Blackmun, J., dissenting), it would be necessary to ask whether Duke's statements could be taken as announcing a change in OPD policy, and if so, whether he had, in fact or appearance, the authority to make such a change and to guarantee the privacy of text messaging. It would also be necessary to consider whether a review of messages sent on police pagers, particularly those sent while officers are on duty, might be justified for other reasons, including performance evaluations, litigation concerning the lawfulness of police actions, and perhaps compliance with state open records laws. See Brief for Petitioners 35–40 (citing Cal. Public Records Act, Cal. Govt. Code Ann. §6250 *et seq.* (West 2008)). These matters would all bear on the legitimacy of an employee's privacy expectation.

The Court must proceed with care when considering the whole concept of privacy expectations in communications made on electronic equipment owned by a government employer. The judiciary risks error by elaborating too fully on the Fourth Amendment implications of emerging technology before its role in society has become clear. See, *e.g., Olmstead* v. *United States*, 277 U. S. 438 (1928), overruled by *Katz* v. *United States*, 389 U. S. 347, 353 (1967). In *Katz*, the Court relied on its own knowledge and experience to conclude that there is a reasonable expectation of privacy in a telephone booth. See *id.,* at 360–361 (Harlan, J., concurring). It is not so clear that courts at present are on so sure a ground. Prudence counsels caution before the facts in the instant case are used to establish far-reaching premises that define the existence, and extent, of privacy expectations enjoyed by employees when using employer-provided communication devices.

Rapid changes in the dynamics of communication and information transmission are evident not just in the technology itself but in what society accepts as proper behavior. As one *amici* brief notes, many employers expect or at least tolerate personal use of such equipment by employees because it often increases worker efficiency. See Brief for Electronic Frontier Foundation et al. 16–20. Another *amicus* points out that the law is beginning to respond to these developments, as some States have recently passed statutes requiring employers to notify employees when monitoring their electronic communications. See Brief for New York Intellectual Property Law Association 22 (citing Del. Code Ann., Tit. 19, §705 (2005); Conn. Gen. Stat. Ann. §31–48d (West 2003)). At present, it is uncertain how workplace norms, and the law's treatment of them, will evolve.

Even if the Court were certain that the *O'Connor* plurality's approach were the right one, the Court would have difficulty predicting how employees' privacy expectations will be shaped by those changes or the degree to which society will be prepared to recognize those expectations as reasonable. See 480 U. S., at 715. Cell phone and text message communications are so pervasive that some persons may consider them to be essential means or necessary instruments for self-expression, even self-identification. That might strengthen the case for an expectation of privacy. On the other hand, the ubiquity of those devices has made them generally affordable, so one could counter that employees who need cell phones or similar devices for personal matters can purchase and pay for their own. And employer policies concerning communications will of course shape the reasonable expectations of their employees, especially to the extent that such policies are clearly communicated.

A broad holding concerning employees' privacy expectations vis-à-vis employer-provided technological equipment

might have implications for future cases that cannot be predicted. It is preferable to dispose of this case on narrower grounds. For present purposes we assume several propositions *arguendo:* First, Quon had a reasonable expectation of privacy in the text messages sent on the pager provided to him by the City; second, petitioners' review of the transcript constituted a search within the meaning of the Fourth Amendment; and third, the principles applicable to a government employer's search of an employee's physical office apply with at least the same force when the employer intrudes on the employee's privacy in the electronic sphere.

B

Even if Quon had a reasonable expectation of privacy in his text messages, petitioners did not necessarily violate the Fourth Amendment by obtaining and reviewing the transcripts. Although as a general matter, warrantless searches "are *per se* unreasonable under the Fourth Amendment," there are "a few specifically established and well-delineated exceptions" to that general rule. *Katz, supra,* at 357. The Court has held that the "'special needs'" of the workplace justify one such exception. *O'Connor*, 480 U. S., at 725 (plurality opinion); *id.,* at 732 (SCALIA, J., concurring in judgment); *Von Raab,* 489 U. S., at 666–667.

Under the approach of the *O'Connor* plurality, when conducted for a "noninvestigatory, work-related purpos[e]" or for the "investigatio[n] of work-related misconduct," a government employer's warrantless search is reasonable if it is "'justified at its inception'" and if "'the measures adopted are reasonably related to the objectives of the search and not excessively intrusive in light of'" the circumstances giving rise to the search. 480 U. S., at 725–726. The search here satisfied the standard of the *O'Connor* plurality and was reasonable under that ap-

proach.

The search was justified at its inception because there were "reasonable grounds for suspecting that the search [was] necessary for a noninvestigatory work-related purpose." *Id.,* at 726. As a jury found, Chief Scharf ordered the search in order to determine whether the character limit on the City's contract with Arch Wireless was sufficient to meet the City's needs. This was, as the Ninth Circuit noted, a "legitimate work-related rationale." 529 F. 3d, at 908. The City and OPD had a legitimate interest in ensuring that employees were not being forced to pay out of their own pockets for work-related expenses, or on the other hand that the City was not paying for extensive personal communications.

As for the scope of the search, reviewing the transcripts was reasonable because it was an efficient and expedient way to determine whether Quon's overages were the result of work-related messaging or personal use. The review was also not "'excessively intrusive.'" *O'Connor, supra,* at 726 (plurality opinion). Although Quon had gone over his monthly allotment a number of times, OPD requested transcripts for only the months of August and September 2002. While it may have been reasonable as well for OPD to review transcripts of all the months in which Quon exceeded his allowance, it was certainly reasonable for OPD to review messages for just two months in order to obtain a large enough sample to decide whether the character limits were efficacious. And it is worth noting that during his internal affairs investigation, McMahon redacted all messages Quon sent while off duty, a measure which reduced the intrusiveness of any further review of the transcripts.

Furthermore, and again on the assumption that Quon had a reasonable expectation of privacy in the contents of his messages, the extent of an expectation is relevant to assessing whether the search was too intrusive. See *Von*

Raab, *supra,* at 671; cf. *Vernonia School Dist. 47J* v. *Acton*, 515 U. S. 646, 654–657 (1995). Even if he could assume some level of privacy would inhere in his messages, it would not have been reasonable for Quon to conclude that his messages were in all circumstances immune from scrutiny. Quon was told that his messages were subject to auditing. As a law enforcement officer, he would or should have known that his actions were likely to come under legal scrutiny, and that this might entail an analysis of his on-the-job communications. Under the circumstances, a reasonable employee would be aware that sound management principles might require the audit of messages to determine whether the pager was being appropriately used. Given that the City issued the pagers to Quon and other SWAT Team members in order to help them more quickly respond to crises—and given that Quon had received no assurances of privacy—Quon could have anticipated that it might be necessary for the City to audit pager messages to assess the SWAT Team's performance in particular emergency situations.

From OPD's perspective, the fact that Quon likely had only a limited privacy expectation, with boundaries that we need not here explore, lessened the risk that the review would intrude on highly private details of Quon's life. OPD's audit of messages on Quon's employer-provided pager was not nearly as intrusive as a search of his personal e-mail account or pager, or a wiretap on his home phone line, would have been. That the search did reveal intimate details of Quon's life does not make it unreasonable, for under the circumstances a reasonable employer would not expect that such a review would intrude on such matters. The search was permissible in its scope.

The Court of Appeals erred in finding the search unreasonable. It pointed to a "host of simple ways to verify the efficacy of the 25,000 character limit . . . without intruding on [respondents'] Fourth Amendment rights." 529 F. 3d,

at 909. The panel suggested that Scharf "could have warned Quon that for the month of September he was forbidden from using his pager for personal communications, and that the contents of all his messages would be reviewed to ensure the pager was used only for work-related purposes during that time frame. Alternatively, if [OPD] wanted to review past usage, it could have asked Quon to count the characters himself, or asked him to redact personal messages and grant permission to [OPD] to review the redacted transcript." *Ibid.*

This approach was inconsistent with controlling precedents. This Court has "repeatedly refused to declare that only the 'least intrusive' search practicable can be reasonable under the Fourth Amendment." *Vernonia*, *supra,* at 663; see also, *e.g.*, *Board of Ed. of Independent School Dist. No. 92 of Pottawatomie Cty.* v. *Earls*, 536 U. S. 822, 837 (2002); *Illinois* v. *Lafayette*, 462 U. S. 640, 647 (1983). That rationale "could raise insuperable barriers to the exercise of virtually all search-and-seizure powers," *United States* v. *Martinez-Fuerte*, 428 U. S. 543, 557, n. 12 (1976), because "judges engaged in *post hoc* evaluations of government conduct can almost always imagine some alternative means by which the objectives of the government might have been accomplished," *Skinner*, 489 U. S., at 629, n. 9 (internal quotation marks and brackets omitted). The analytic errors of the Court of Appeals in this case illustrate the necessity of this principle. Even assuming there were ways that OPD could have performed the search that would have been less intrusive, it does not follow that the search as conducted was unreasonable.

Respondents argue that the search was *per se* unreasonable in light of the Court of Appeals' conclusion that Arch Wireless violated the SCA by giving the City the transcripts of Quon's text messages. The merits of the SCA claim are not before us. But even if the Court of Appeals was correct to conclude that the SCA forbade

Arch Wireless from turning over the transcripts, it does not follow that petitioners' actions were unreasonable. Respondents point to no authority for the proposition that the existence of statutory protection renders a search *per se* unreasonable under the Fourth Amendment. And the precedents counsel otherwise. See *Virginia* v. *Moore*, 553 U. S. 164, 168 (2008) (search incident to an arrest that was illegal under state law was reasonable); *California* v. *Greenwood*, 486 U. S. 35, 43 (1988) (rejecting argument that if state law forbade police search of individual's garbage the search would violate the Fourth Amendment). Furthermore, respondents do not maintain that any OPD employee either violated the law him- or herself or knew or should have known that Arch Wireless, by turning over the transcript, would have violated the law. The otherwise reasonable search by OPD is not rendered unreasonable by the assumption that Arch Wireless violated the SCA by turning over the transcripts.

Because the search was motivated by a legitimate work-related purpose, and because it was not excessive in scope, the search was reasonable under the approach of the *O'Connor* plurality. 480 U. S., at 726. For these same reasons—that the employer had a legitimate reason for the search, and that the search was not excessively intrusive in light of that justification—the Court also concludes that the search would be "regarded as reasonable and normal in the private-employer context" and would satisfy the approach of JUSTICE SCALIA's concurrence. *Id.,* at 732. The search was reasonable, and the Court of Appeals erred by holding to the contrary. Petitioners did not violate Quon's Fourth Amendment rights.

C

Finally, the Court must consider whether the search violated the Fourth Amendment rights of Jerilyn Quon, Florio, and Trujillo, the respondents who sent text mes-

sages to Jeff Quon. Petitioners and respondents disagree whether a sender of a text message can have a reasonable expectation of privacy in a message he knowingly sends to someone's employer-provided pager. It is not necessary to resolve this question in order to dispose of the case, however. Respondents argue that because "the search was unreasonable as to Sergeant Quon, it was also unreasonable as to his correspondents." Brief for Respondents 60 (some capitalization omitted; boldface deleted). They make no corollary argument that the search, if reasonable as to Quon, could nonetheless be unreasonable as to Quon's correspondents. See *id.,* at 65–66. In light of this litigating position and the Court's conclusion that the search was reasonable as to Jeff Quon, it necessarily follows that these other respondents cannot prevail.

* * *

Because the search was reasonable, petitioners did not violate respondents' Fourth Amendment rights, and the court below erred by concluding otherwise. The judgment of the Court of Appeals for the Ninth Circuit is reversed, and the case is remanded for further proceedings consistent with this opinion.

It is so ordered.

SUPREME COURT OF THE UNITED STATES

No. 08–1332

CITY OF ONTARIO, CALIFORNIA, ET AL.,
PETITIONERS *v.* JEFF QUON ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

[June 17, 2010]

JUSTICE STEVENS, concurring.

Although I join the Court's opinion in full, I write separately to highlight that the Court has sensibly declined to resolve whether the plurality opinion in *O'Connor* v. *Ortega*, 480 U. S. 709 (1987), provides the correct approach to determining an employee's reasonable expectation of privacy. See *ante*, at 9. Justice Blackmun, writing for the four dissenting Justices in *O'Connor*, agreed with JUSTICE SCALIA that an employee enjoys a reasonable expectation of privacy in his office. 480 U. S., at 737. But he advocated a third approach to the reasonable expectation of privacy inquiry, separate from those proposed by the *O'Connor* plurality and by JUSTICE SCALIA, see *ante*, at 8. Recognizing that it is particularly important to safeguard "a public employee's expectation of privacy in the workplace" in light of the "reality of work in modern time," 480 U. S., at 739, which lacks "tidy distinctions" between workplace and private activities, *ibid.*, Justice Blackmun argued that "the precise extent of an employee's expectation of privacy often turns on the nature of the search," *id.,* at 738. And he emphasized that courts should determine this expectation in light of the specific facts of each particular search, rather than by announcing a categorical standard. See *id.*, at 741.

For the reasons stated at page 13 of the Court's opinion,

it is clear that respondent Jeff Quon, as a law enforcement officer who served on a SWAT Team, should have understood that all of his work-related actions—including all of his communications on his official pager—were likely to be subject to public and legal scrutiny. He therefore had only a limited expectation of privacy in relation to this particular audit of his pager messages. Whether one applies the reasoning from Justice O'Connor's opinion, JUSTICE SCALIA's concurrence, or Justice Blackmun's dissent* in *O'Connor*, the result is the same: The judgment of the Court of Appeals in this case must be reversed.

---

*I do not contend that Justice Blackmun's opinion is controlling under *Marks* v. *United States*, 430 U. S. 188, 193 (1977), but neither is his approach to evaluating a reasonable expectation of privacy foreclosed by *O'Connor*. Indeed, his approach to that inquiry led to the conclusion, shared by JUSTICE SCALIA but not adopted by the *O'Connor* plurality, that an employee had a reasonable expectation of privacy in his office. See *O'Connor* v. *Ortega*, 480 U. S. 709, 718 (1987) (plurality opinion). But Justice Blackmun would have applied the Fourth Amendment's warrant and probable-cause requirements to workplace investigatory searches, *id.*, at 732 (dissenting opinion), whereas a majority of the Court rejected that view, see *id.*, at 722, 725 (plurality opinion); *id.*, at 732 (SCALIA, J., concurring in judgment). It was that analysis—regarding the proper standard for evaluating a search when an employee has a reasonable expectation of privacy—that produced the opposite result in the case. This case does not implicate that debate because it does not involve an investigatory search. The jury concluded that the purpose of the audit was to determine whether the character limits were sufficient for work-related messages. See *ante*, at 6.

SUPREME COURT OF THE UNITED STATES

No. 08–1332

CITY OF ONTARIO, CALIFORNIA, ET AL.,
PETITIONERS *v.* JEFF QUON ET AL.

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE NINTH CIRCUIT

[June 17, 2010]

JUSTICE SCALIA, concurring in part and concurring in the judgment.

I join the Court's opinion except for Part III–A. I continue to believe that the "operational realities" rubric for determining the Fourth Amendment's application to public employees invented by the plurality in *O'Connor* v. *Ortega*, 480 U. S. 709, 717 (1987), is standardless and unsupported. *Id.*, at 729–732 (SCALIA, J., concurring in judgment). In this case, the proper threshold inquiry should be not whether the Fourth Amendment applies to messages on *public* employees' employer-issued pagers, but whether it applies *in general* to such messages on employer-issued pagers. See *id.*, at 731.

Here, however, there is no need to answer that threshold question. Even accepting at face value Quon's and his co-plaintiffs' claims that the Fourth Amendment applies to their messages, the city's search was reasonable, and thus did not violate the Amendment. See *id.*, at 726 (plurality opinion); *id.*, at 732 (SCALIA, J., concurring in judgment). Since it is unnecessary to decide whether the Fourth Amendment applies, it is unnecessary to resolve which approach in *O'Connor* controls: the plurality's or mine.*

*Despite his disclaimer, *ante*, at 2, n. (concurring opinion), JUSTICE STEVENS' concurrence implies, *ante*, at 1–2, that it is also an open

That should end the matter.

The Court concedes as much, *ante*, at 9, 12–17, yet it inexplicably interrupts its analysis with a recitation of the parties' arguments concerning, and an excursus on the complexity and consequences of answering, that admittedly irrelevant threshold question, *ante*, at 9–12. That discussion is unnecessary. (To whom do we owe an *additional* explanation for declining to decide an issue, once we have explained that it makes no difference?) It also seems to me exaggerated. Applying the Fourth Amendment to new technologies may sometimes be difficult, but when it is necessary to decide a case we have no choice. The Court's implication, *ante*, at 10, that where electronic privacy is concerned we should decide less than we otherwise would (that is, less than the principle of law necessary to resolve the case and guide private action)—or that we should hedge our bets by concocting case-specific standards or issuing opaque opinions—is in my view indefensible. The-times-they-are-a-changin' is a feeble excuse for disregard of duty.

Worse still, the digression is self-defeating. Despite the Court's insistence that it is agnostic about the proper test, lower courts will likely read the Court's self-described "instructive" expatiation on how the *O'Connor* plurality's approach would apply here (if it applied), *ante*, at 9–11, as a heavy-handed hint about how *they* should proceed. Litigants will do likewise, using the threshold question whether the Fourth Amendment is even implicated as a

---

question whether the approach advocated by Justice Blackmun in his *dissent* in *O'Connor* is the proper standard. There is room for reasonable debate as to which of the two approaches advocated by Justices whose votes supported the judgment in *O'Connor*—the plurality's and mine—is controlling under *Marks* v. *United States*, 430 U. S. 188, 193 (1977). But unless *O'Connor* is overruled, it is assuredly false that a test that would have produced the *opposite* result in that case is still in the running.

basis for bombarding lower courts with arguments about employer policies, how they were communicated, and whether they were authorized, as well as the latest trends in employees' use of electronic media. In short, in saying why it is not saying more, the Court says much more than it should.

The Court's inadvertent boosting of the *O'Connor* plurality's standard is all the more ironic because, in fleshing out its fears that applying that test to new technologies will be too hard, the Court underscores the unworkability of that standard. Any rule that requires evaluating whether a given gadget is a "necessary instrumen[t] for self-expression, even self-identification," on top of assessing the degree to which "the law's treatment of [workplace norms has] evolve[d]," *ante*, at 11, is (to put it mildly) unlikely to yield objective answers.

I concur in the Court's judgment.

Syllabus

NOTE: Where it is feasible, a syllabus (headnote) will be released, as is being done in connection with this case, at the time the opinion is issued. The syllabus constitutes no part of the opinion of the Court but has been prepared by the Reporter of Decisions for the convenience of the reader. See *United States* v. *Detroit Timber & Lumber Co.,* 200 U. S. 321, 337.

SUPREME COURT OF THE UNITED STATES

Syllabus

UNITED STATES *v.* JONES

CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

No. 10–1259. Argued November 8, 2011—Decided January 23, 2012

The Government obtained a search warrant permitting it to install a Global-Positioning-System (GPS) tracking device on a vehicle registered to respondent Jones's wife. The warrant authorized installation in the District of Columbia and within 10 days, but agents installed the device on the 11th day and in Maryland. The Government then tracked the vehicle's movements for 28 days. It subsequently secured an indictment of Jones and others on drug trafficking conspiracy charges. The District Court suppressed the GPS data obtained while the vehicle was parked at Jones's residence, but held the remaining data admissible because Jones had no reasonable expectation of privacy when the vehicle was on public streets. Jones was convicted. The D. C. Circuit reversed, concluding that admission of the evidence obtained by warrantless use of the GPS device violated the Fourth Amendment.

Held: The Government's attachment of the GPS device to the vehicle, and its use of that device to monitor the vehicle's movements, constitutes a search under the Fourth Amendment. Pp. 3–12.

(a) The Fourth Amendment protects the "right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures." Here, the Government's physical intrusion on an "effect" for the purpose of obtaining information constitutes a "search." This type of encroachment on an area enumerated in the Amendment would have been considered a search within the meaning of the Amendment at the time it was adopted. Pp. 3–4.

(b) This conclusion is consistent with this Court's Fourth Amendment jurisprudence, which until the latter half of the 20th century was tied to common-law trespass. Later cases, which have deviated from that exclusively property-based approach, have applied the

Syllabus

analysis of Justice Harlan's concurrence in *Katz* v. *United States*, 389 U. S. 347, which said that the Fourth Amendment protects a person's "reasonable expectation of privacy," *id.*, at 360. Here, the Court need not address the Government's contention that Jones had no "reasonable expectation of privacy," because Jones's Fourth Amendment rights do not rise or fall with the *Katz* formulation. At bottom, the Court must "assur[e] preservation of that degree of privacy against government that existed when the Fourth Amendment was adopted." *Kyllo* v. *United States*, 533 U. S. 27, 34. *Katz* did not repudiate the understanding that the Fourth Amendment embodies a particular concern for government trespass upon the areas it enumerates. The *Katz* reasonable-expectation-of-privacy test has been added to, but not substituted for, the common-law trespassory test. See *Alderman* v. *United States*, 394 U. S. 165, 176; *Soldal* v. *Cook County*, 506 U. S. 56, 64. *United States* v. *Knotts*, 460 U. S. 276, and *United States* v. *Karo*, 468 U. S. 705—post-*Katz* cases rejecting Fourth Amendment challenges to "beepers," electronic tracking devices representing another form of electronic monitoring—do not foreclose the conclusion that a search occurred here. *New York* v. *Class*, 475 U. S. 106, and *Oliver* v. *United States*, 466 U. S. 170, also do not support the Government's position. Pp. 4–12.

(c) The Government's alternative argument—that if the attachment and use of the device was a search, it was a reasonable one—is forfeited because it was not raised below. P. 12.

615 F. 3d 544, affirmed.

SCALIA, J., delivered the opinion of the Court, in which ROBERTS, C. J., and KENNEDY, THOMAS, and SOTOMAYOR, JJ., joined. SOTOMAYOR, J., filed a concurring opinion. ALITO, J., filed an opinion concurring in the judgment, in which GINSBURG, BREYER, and KAGAN, JJ., joined.

SUPREME COURT OF THE UNITED STATES

No. 10–1259

UNITED STATES, PETITIONER *v.* ANTOINE JONES

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

[January 23, 2012]

JUSTICE SCALIA delivered the opinion of the Court.

We decide whether the attachment of a Global-Positioning-System (GPS) tracking device to an individual's vehicle, and subsequent use of that device to monitor the vehicle's movements on public streets, constitutes a search or seizure within the meaning of the Fourth Amendment.

I

In 2004 respondent Antoine Jones, owner and operator of a nightclub in the District of Columbia, came under suspicion of trafficking in narcotics and was made the target of an investigation by a joint FBI and Metropolitan Police Department task force. Officers employed various investigative techniques, including visual surveillance of the nightclub, installation of a camera focused on the front door of the club, and a pen register and wiretap covering Jones's cellular phone.

Based in part on information gathered from these sources, in 2005 the Government applied to the United States District Court for the District of Columbia for a warrant authorizing the use of an electronic tracking device on the Jeep Grand Cherokee registered to Jones's

wife. A warrant issued, authorizing installation of the device in the District of Columbia and within 10 days.

On the 11th day, and not in the District of Columbia but in Maryland,[1] agents installed a GPS tracking device on the undercarriage of the Jeep while it was parked in a public parking lot. Over the next 28 days, the Government used the device to track the vehicle's movements, and once had to replace the device's battery when the vehicle was parked in a different public lot in Maryland. By means of signals from multiple satellites, the device established the vehicle's location within 50 to 100 feet, and communicated that location by cellular phone to a Government computer. It relayed more than 2,000 pages of data over the 4-week period.

The Government ultimately obtained a multiple-count indictment charging Jones and several alleged coconspirators with, as relevant here, conspiracy to distribute and possess with intent to distribute five kilograms or more of cocaine and 50 grams or more of cocaine base, in violation of 21 U. S. C. §§841 and 846. Before trial, Jones filed a motion to suppress evidence obtained through the GPS device. The District Court granted the motion only in part, suppressing the data obtained while the vehicle was parked in the garage adjoining Jones's residence. 451 F. Supp. 2d 71, 88 (2006). It held the remaining data admissible, because "'[a] person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another.'" *Ibid.* (quoting *United States* v. *Knotts*, 460 U. S. 276, 281 (1983)). Jones's trial in October 2006 produced a hung jury on the conspiracy count.

In March 2007, a grand jury returned another indict-

---

[1] In this litigation, the Government has conceded noncompliance with the warrant and has argued only that a warrant was not required. *United States* v. *Maynard*, 615 F. 3d 544, 566, n. (CADC 2010).

ment, charging Jones and others with the same conspiracy. The Government introduced at trial the same GPS-derived locational data admitted in the first trial, which connected Jones to the alleged conspirators' stash house that contained $850,000 in cash, 97 kilograms of cocaine, and 1 kilogram of cocaine base. The jury returned a guilty verdict, and the District Court sentenced Jones to life imprisonment.

The United States Court of Appeals for the District of Columbia Circuit reversed the conviction because of admission of the evidence obtained by warrantless use of the GPS device which, it said, violated the Fourth Amendment. *United States* v. *Maynard*, 615 F. 3d 544 (2010). The D. C. Circuit denied the Government's petition for rehearing en banc, with four judges dissenting. 625 F. 3d 766 (2010). We granted certiorari, 564 U. S. ___ (2011).

II

A

The Fourth Amendment provides in relevant part that "[t]he right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated." It is beyond dispute that a vehicle is an "effect" as that term is used in the Amendment. *United States* v. *Chadwick*, 433 U. S. 1, 12 (1977). We hold that the Government's installation of a GPS device on a target's vehicle,[2] and its use of that device to monitor the vehicle's movements, constitutes a "search."

---

[2] As we have noted, the Jeep was registered to Jones's wife. The Government acknowledged, however, that Jones was "the exclusive driver." *Id.*, at 555, n. (internal quotation marks omitted). If Jones was not the owner he had at least the property rights of a bailee. The Court of Appeals concluded that the vehicle's registration did not affect his ability to make a Fourth Amendment objection, *ibid.*, and the Government has not challenged that determination here. We therefore do not consider the Fourth Amendment significance of Jones's status.

It is important to be clear about what occurred in this case: The Government physically occupied private property for the purpose of obtaining information. We have no doubt that such a physical intrusion would have been considered a "search" within the meaning of the Fourth Amendment when it was adopted. *Entick* v. *Carrington*, 95 Eng. Rep. 807 (C. P. 1765), is a "case we have described as a 'monument of English freedom' 'undoubtedly familiar' to 'every American statesman' at the time the Constitution was adopted, and considered to be 'the true and ultimate expression of constitutional law'" with regard to search and seizure. *Brower* v. *County of Inyo*, 489 U. S. 593, 596 (1989) (quoting *Boyd* v. *United States*, 116 U. S. 616, 626 (1886)). In that case, Lord Camden expressed in plain terms the significance of property rights in search-and-seizure analysis:

> "[O]ur law holds the property of every man so sacred, that no man can set his foot upon his neighbour's close without his leave; if he does he is a trespasser, though he does no damage at all; if he will tread upon his neighbour's ground, he must justify it by law." *Entick, supra,* at 817.

The text of the Fourth Amendment reflects its close connection to property, since otherwise it would have referred simply to "the right of the people to be secure against unreasonable searches and seizures"; the phrase "in their persons, houses, papers, and effects" would have been superfluous.

Consistent with this understanding, our Fourth Amendment jurisprudence was tied to common-law trespass, at least until the latter half of the 20th century. *Kyllo* v. *United States*, 533 U. S. 27, 31 (2001); Kerr, The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution, 102 Mich. L. Rev. 801, 816 (2004). Thus, in *Olmstead* v. *United States*, 277 U. S.

438 (1928), we held that wiretaps attached to telephone wires on the public streets did not constitute a Fourth Amendment search because "[t]here was no entry of the houses or offices of the defendants," *id.*, at 464.

Our later cases, of course, have deviated from that exclusively property-based approach. In *Katz* v. *United States*, 389 U. S. 347, 351 (1967), we said that "the Fourth Amendment protects people, not places," and found a violation in attachment of an eavesdropping device to a public telephone booth. Our later cases have applied the analysis of Justice Harlan's concurrence in that case, which said that a violation occurs when government officers violate a person's "reasonable expectation of privacy," *id.*, at 360. See, *e.g.*, *Bond* v. *United States*, 529 U. S. 334 (2000); *California* v. *Ciraolo*, 476 U. S. 207 (1986); *Smith* v. *Maryland*, 442 U. S. 735 (1979).

The Government contends that the Harlan standard shows that no search occurred here, since Jones had no "reasonable expectation of privacy" in the area of the Jeep accessed by Government agents (its underbody) and in the locations of the Jeep on the public roads, which were visible to all. But we need not address the Government's contentions, because Jones's Fourth Amendment rights do not rise or fall with the *Katz* formulation. At bottom, we must "assur[e] preservation of that degree of privacy against government that existed when the Fourth Amendment was adopted." *Kyllo*, *supra*, at 34. As explained, for most of our history the Fourth Amendment was understood to embody a particular concern for government trespass upon the areas ("persons, houses, papers, and effects") it enumerates.[3] *Katz* did not repudiate

---

[3]JUSTICE ALITO's concurrence (hereinafter concurrence) doubts the wisdom of our approach because "it is almost impossible to think of late-18th-century situations that are analogous to what took place in this case." *Post*, at 3 (opinion concurring in judgment). But in fact it posits a situation that is not far afield—a constable's concealing himself

that understanding. Less than two years later the Court upheld defendants' contention that the Government could not introduce against them conversations between *other* people obtained by warrantless placement of electronic surveillance devices in their homes. The opinion rejected the dissent's contention that there was no Fourth Amendment violation "unless the conversational privacy of the homeowner himself is invaded."[4] *Alderman* v. *United States*, 394 U. S. 165, 176 (1969). "[W]e [do not] believe that *Katz*, by holding that the Fourth Amendment protects persons and their private conversations, was intended to withdraw any of the protection which the Amendment extends to the home . . . ." *Id.*, at 180.

More recently, in *Soldal* v. *Cook County*, 506 U. S. 56 (1992), the Court unanimously rejected the argument that although a "seizure" had occurred "in a 'technical' sense" when a trailer home was forcibly removed, *id.*, at 62, no Fourth Amendment violation occurred because law enforcement had not "invade[d] the [individuals'] privacy," *id.*, at 60. *Katz*, the Court explained, established that "property rights are not the sole measure of Fourth

---

in the target's coach in order to track its movements. *Ibid.* There is no doubt that the information gained by that trespassory activity would be the product of an unlawful search—whether that information consisted of the conversations occurring in the coach, or of the destinations to which the coach traveled.

In any case, it is quite irrelevant whether there was an 18th-century analog. Whatever new methods of investigation may be devised, our task, *at a minimum*, is to decide whether the action in question would have constituted a "search" within the original meaning of the Fourth Amendment. Where, as here, the Government obtains information by physically intruding on a constitutionally protected area, such a search has undoubtedly occurred.

[4] Thus, the concurrence's attempt to recast *Alderman* as meaning that individuals have a "legitimate expectation of privacy in all conversations that [take] place under their roof," *post*, at 6–7, is foreclosed by the Court's opinion. The Court took as a given that the homeowner's "conversational privacy" had not been violated.

Amendment violations," but did not "snuf[f] out the previously recognized protection for property." 506 U. S., at 64. As Justice Brennan explained in his concurrence in *Knotts*, *Katz* did not erode the principle "that, when the Government *does* engage in physical intrusion of a constitutionally protected area in order to obtain information, that intrusion may constitute a violation of the Fourth Amendment." 460 U. S., at 286 (opinion concurring in judgment). We have embodied that preservation of past rights in our very definition of "reasonable expectation of privacy" which we have said to be an expectation "that has a source outside of the Fourth Amendment, either by reference to concepts of real or personal property law or to understandings that are recognized and permitted by society." *Minnesota* v. *Carter*, 525 U. S. 83, 88 (1998) (internal quotation marks omitted). *Katz* did not narrow the Fourth Amendment's scope.[5]

The Government contends that several of our post-*Katz* cases foreclose the conclusion that what occurred here constituted a search. It relies principally on two cases in

---

[5]The concurrence notes that post-*Katz* we have explained that "'an actual trespass is neither necessary *nor sufficient* to establish a constitutional violation.'" *Post*, at 6 (quoting *United States* v. *Karo*, 468 U. S. 705, 713 (1984)). That is undoubtedly true, and undoubtedly irrelevant. *Karo* was considering whether a seizure occurred, and as the concurrence explains, a seizure of property occurs, not when there is a trespass, but "when there is some meaningful interference with an individual's possessory interests in that property." *Post*, at 2 (internal quotation marks omitted). Likewise with a search. Trespass alone does not qualify, but there must be conjoined with that what was present here: an attempt to find something or to obtain information.

Related to this, and similarly irrelevant, is the concurrence's point that, if analyzed separately, neither the installation of the device nor its use would constitute a Fourth Amendment search. See *ibid.* Of course not. A trespass on "houses" or "effects," or a *Katz* invasion of privacy, is not alone a search unless it is done to obtain information; and the obtaining of information is not alone a search unless it is achieved by such a trespass or invasion of privacy.

which we rejected Fourth Amendment challenges to "beepers," electronic tracking devices that represent another form of electronic monitoring. The first case, *Knotts*, upheld against Fourth Amendment challenge the use of a "beeper" that had been placed in a container of chloroform, allowing law enforcement to monitor the location of the container. 460 U. S., at 278. We said that there had been no infringement of Knotts' reasonable expectation of privacy since the information obtained—the location of the automobile carrying the container on public roads, and the location of the off-loaded container in open fields near Knotts' cabin—had been voluntarily conveyed to the public.[6] *Id.*, at 281–282. But as we have discussed, the *Katz* reasonable-expectation-of-privacy test has been *added to*, not *substituted for*, the common-law trespassory test. The holding in *Knotts* addressed only the former, since the latter was not at issue. The beeper had been placed in the container before it came into Knotts' possession, with the consent of the then-owner. 460 U. S., at 278. Knotts did not challenge that installation, and we specifically declined to consider its effect on the Fourth Amendment analysis. *Id.*, at 279, n. *Knotts* would be relevant, perhaps, if the Government were making the argument that what would otherwise be an unconstitutional search is not such where it produces only public information. The Government does not make that argument, and we know of no case that would support it.

The second "beeper" case, *United States* v. *Karo*, 468 U. S. 705 (1984), does not suggest a different conclusion. There we addressed the question left open by *Knotts*, whether the installation of a beeper in a container

---

[6] *Knotts* noted the "limited use which the government made of the signals from this particular beeper," 460 U. S., at 284; and reserved the question whether "different constitutional principles may be applicable" to "dragnet-type law enforcement practices" of the type that GPS tracking made possible here, *ibid.*

amounted to a search or seizure. 468 U. S., at 713. As in *Knotts*, at the time the beeper was installed the container belonged to a third party, and it did not come into possession of the defendant until later. 468 U. S., at 708. Thus, the specific question we considered was whether the installation "*with the consent of the original owner* constitute[d] a search or seizure . . . when the container is delivered to a buyer having no knowledge of the presence of the beeper." *Id.*, at 707 (emphasis added). We held not. The Government, we said, came into physical contact with the container only before it belonged to the defendant Karo; and the transfer of the container with the unmonitored beeper inside did not convey any information and thus did not invade Karo's privacy. See *id.*, at 712. That conclusion is perfectly consistent with the one we reach here. Karo accepted the container as it came to him, beeper and all, and was therefore not entitled to object to the beeper's presence, even though it was used to monitor the container's location. Cf. *On Lee* v. *United States*, 343 U. S. 747, 751–752 (1952) (no search or seizure where an informant, who was wearing a concealed microphone, was invited into the defendant's business). Jones, who possessed the Jeep at the time the Government trespassorily inserted the information-gathering device, is on much different footing.

The Government also points to our exposition in *New York* v. *Class*, 475 U. S. 106 (1986), that "[t]he exterior of a car . . . is thrust into the public eye, and thus to examine it does not constitute a 'search.'" *Id.*, at 114. That statement is of marginal relevance here since, as the Government acknowledges, "the officers in this case did *more* than conduct a visual inspection of respondent's vehicle," Brief for United States 41 (emphasis added). By attaching the device to the Jeep, officers encroached on a protected area. In *Class* itself we suggested that this would make a difference, for we concluded that an officer's momentary reaching into the interior of a vehicle did constitute a

search.[7] 475 U. S., at 114–115.

Finally, the Government's position gains little support from our conclusion in *Oliver* v. *United States*, 466 U. S. 170 (1984), that officers' information-gathering intrusion on an "open field" did not constitute a Fourth Amendment search even though it was a trespass at common law, *id.*, at 183. Quite simply, an open field, unlike the curtilage of a home, see *United States* v. *Dunn*, 480 U. S. 294, 300 (1987), is not one of those protected areas enumerated in the Fourth Amendment. *Oliver, supra*, at 176–177. See also *Hester* v. *United States*, 265 U. S. 57, 59 (1924). The Government's physical intrusion on such an area—unlike its intrusion on the "effect" at issue here—is of no Fourth Amendment significance.[8]

B

The concurrence begins by accusing us of applying "18th-century tort law." *Post*, at 1. That is a distortion. What we apply is an 18th-century guarantee against unreasonable searches, which we believe must provide *at*

---

[7]The Government also points to *Cardwell* v. *Lewis*, 417 U. S. 583 (1974), in which the Court rejected the claim that the inspection of an impounded vehicle's tire tread and the collection of paint scrapings from its exterior violated the Fourth Amendment. Whether the plurality said so because no search occurred or because the search was reasonable is unclear. Compare *id.*, at 591 (opinion of Blackmun, J.) ("[W]e fail to comprehend what expectation of privacy was infringed"), with *id.*, at 592 ("Under circumstances such as these, where probable cause exists, a warrantless examination of the exterior of a car is not unreasonable . . . ").

[8]Thus, our theory is *not* that the Fourth Amendment is concerned with "*any* technical trespass that led to the gathering of evidence." *Post*, at 3 (ALITO, J., concurring in judgment) (emphasis added). The Fourth Amendment protects against trespassory searches only with regard to those items ("persons, houses, papers, and effects") that it enumerates. The trespass that occurred in *Oliver* may properly be understood as a "search," but not one "in the constitutional sense." 466 U. S., at 170, 183.

a minimum the degree of protection it afforded when it was adopted. The concurrence does not share that belief. It would apply *exclusively Katz*'s reasonable-expectation-of-privacy test, even when that eliminates rights that previously existed.

The concurrence faults our approach for "present[ing] particularly vexing problems" in cases that do not involve physical contact, such as those that involve the transmission of electronic signals. *Post*, at 9. We entirely fail to understand that point. For unlike the concurrence, which would make *Katz* the *exclusive* test, we do not make trespass the exclusive test. Situations involving merely the transmission of electronic signals without trespass would *remain* subject to *Katz* analysis.

In fact, it is the concurrence's insistence on the exclusivity of the *Katz* test that needlessly leads us into "particularly vexing problems" in the present case. This Court has to date not deviated from the understanding that mere visual observation does not constitute a search. See *Kyllo*, 533 U. S., at 31–32. We accordingly held in *Knotts* that "[a] person traveling in an automobile on public thoroughfares has no reasonable expectation of privacy in his movements from one place to another." 460 U. S., at 281. Thus, even assuming that the concurrence is correct to say that "[t]raditional surveillance" of Jones for a 4-week period "would have required a large team of agents, multiple vehicles, and perhaps aerial assistance," *post*, at 12, our cases suggest that such visual observation is constitutionally permissible. It may be that achieving the same result through electronic means, without an accompanying trespass, is an unconstitutional invasion of privacy, but the present case does not require us to answer that question.

And answering it affirmatively leads us needlessly into additional thorny problems. The concurrence posits that "relatively short-term monitoring of a person's movements

on public streets" is okay, but that "the use of longer term GPS monitoring in investigations *of most offenses*" is no good. *Post*, at 13 (emphasis added). That introduces yet another novelty into our jurisprudence. There is no precedent for the proposition that whether a search has occurred depends on the nature of the crime being investigated. And even accepting that novelty, it remains unexplained why a 4-week investigation is "surely" too long and why a drug-trafficking conspiracy involving substantial amounts of cash and narcotics is not an "extraordinary offens[e]" which may permit longer observation. See *post*, at 13–14. What of a 2-day monitoring of a suspected purveyor of stolen electronics? Or of a 6-month monitoring of a suspected terrorist? We may have to grapple with these "vexing problems" in some future case where a classic trespassory search is not involved and resort must be had to *Katz* analysis; but there is no reason for rushing forward to resolve them here.

III

The Government argues in the alternative that even if the attachment and use of the device was a search, it was reasonable—and thus lawful—under the Fourth Amendment because "officers had reasonable suspicion, and indeed probable cause, to believe that [Jones] was a leader in a large-scale cocaine distribution conspiracy." Brief for United States 50–51. We have no occasion to consider this argument. The Government did not raise it below, and the D. C. Circuit therefore did not address it. See 625 F. 3d, at 767 (Ginsburg, Tatel, and Griffith, JJ., concurring in denial of rehearing en banc). We consider the argument forfeited. See *Sprietsma* v. *Mercury Marine*, 537 U. S. 51, 56, n. 4 (2002).

* * *

The judgment of the Court of Appeals for the D. C. Circuit is affirmed.

It is so ordered.

SUPREME COURT OF THE UNITED STATES

No. 10–1259

UNITED STATES, PETITIONER *v.* ANTOINE JONES

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

[January 23, 2012]

JUSTICE SOTOMAYOR, concurring.

I join the Court's opinion because I agree that a search within the meaning of the Fourth Amendment occurs, at a minimum, "[w]here, as here, the Government obtains information by physically intruding on a constitutionally protected area." *Ante,* at 6, n. 3. In this case, the Government installed a Global Positioning System (GPS) tracking device on respondent Antoine Jones' Jeep without a valid warrant and without Jones' consent, then used that device to monitor the Jeep's movements over the course of four weeks. The Government usurped Jones' property for the purpose of conducting surveillance on him, thereby invading privacy interests long afforded, and undoubtedly entitled to, Fourth Amendment protection. See, *e.g., Silverman* v. *United States*, 365 U. S. 505, 511–512 (1961).

Of course, the Fourth Amendment is not concerned only with trespassory intrusions on property. See, *e.g., Kyllo* v. *United States*, 533 U. S. 27, 31–33 (2001). Rather, even in the absence of a trespass, "a Fourth Amendment search occurs when the government violates a subjective expectation of privacy that society recognizes as reasonable." *Id.,* at 33; see also *Smith* v. *Maryland*, 442 U. S. 735, 740–741 (1979); *Katz* v. *United States*, 389 U. S. 347, 361 (1967) (Harlan, J., concurring). In *Katz*, this Court enlarged its then-prevailing focus on property rights by announcing

that the reach of the Fourth Amendment does not "turn upon the presence or absence of a physical intrusion." *Id.,* at 353. As the majority's opinion makes clear, however, *Katz*'s reasonable-expectation-of-privacy test augmented, but did not displace or diminish, the common-law trespassory test that preceded it. *Ante,* at 8. Thus, "when the Government *does* engage in physical intrusion of a constitutionally protected area in order to obtain information, that intrusion may constitute a violation of the Fourth Amendment." *United States* v. *Knotts*, 460 U. S. 276, 286 (1983) (Brennan, J., concurring in judgment); see also, *e.g., Rakas* v. *Illinois*, 439 U. S. 128, 144, n. 12 (1978). JUSTICE ALITO's approach, which discounts altogether the constitutional relevance of the Government's physical intrusion on Jones' Jeep, erodes that longstanding protection for privacy expectations inherent in items of property that people possess or control. See *post,* at 5–7 (opinion concurring in judgment). By contrast, the trespassory test applied in the majority's opinion reflects an irreducible constitutional minimum: When the Government physically invades personal property to gather information, a search occurs. The reaffirmation of that principle suffices to decide this case.

Nonetheless, as JUSTICE ALITO notes, physical intrusion is now unnecessary to many forms of surveillance. *Post,* at 9–12. With increasing regularity, the Government will be capable of duplicating the monitoring undertaken in this case by enlisting factory- or owner-installed vehicle tracking devices or GPS-enabled smartphones. See *United States* v. *Pineda-Moreno*, 617 F. 3d 1120, 1125 (CA9 2010) (Kozinski, C. J., dissenting from denial of rehearing en banc). In cases of electronic or other novel modes of surveillance that do not depend upon a physical invasion on property, the majority opinion's trespassory test may provide little guidance. But "[s]ituations involving merely the transmission of electronic signals without trespass

would *remain* subject to *Katz* analysis." *Ante,* at 11. As JUSTICE ALITO incisively observes, the same technological advances that have made possible nontrespassory surveillance techniques will also affect the *Katz* test by shaping the evolution of societal privacy expectations. *Post,* at 10–11. Under that rubric, I agree with JUSTICE ALITO that, at the very least, "longer term GPS monitoring in investigations of most offenses impinges on expectations of privacy." *Post,* at 13.

In cases involving even short-term monitoring, some unique attributes of GPS surveillance relevant to the *Katz* analysis will require particular attention. GPS monitoring generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about her familial, political, professional, religious, and sexual associations. See, *e.g., People* v. *Weaver*, 12 N. Y. 3d 433, 441–442, 909 N. E. 2d 1195, 1199 (2009) ("Disclosed in [GPS] data . . . will be trips the indisputably private nature of which takes little imagination to conjure: trips to the psychiatrist, the plastic surgeon, the abortion clinic, the AIDS treatment center, the strip club, the criminal defense attorney, the by-the-hour motel, the union meeting, the mosque, synagogue or church, the gay bar and on and on"). The Government can store such records and efficiently mine them for information years into the future. *Pineda-Moreno*, 617 F. 3d, at 1124 (opinion of Kozinski, C. J.). And because GPS monitoring is cheap in comparison to conventional surveillance techniques and, by design, proceeds surreptitiously, it evades the ordinary checks that constrain abusive law enforcement practices: "limited police resources and community hostility." *Illinois* v. *Lidster*, 540 U. S. 419, 426 (2004).

Awareness that the Government may be watching chills associational and expressive freedoms. And the Government's unrestrained power to assemble data that reveal private aspects of identity is susceptible to abuse. The net

result is that GPS monitoring—by making available at a relatively low cost such a substantial quantum of intimate information about any person whom the Government, in its unfettered discretion, chooses to track—may "alter the relationship between citizen and government in a way that is inimical to democratic society." *United States* v. *Cuevas-Perez*, 640 F. 3d 272, 285 (CA7 2011) (Flaum, J., concurring).

I would take these attributes of GPS monitoring into account when considering the existence of a reasonable societal expectation of privacy in the sum of one's public movements. I would ask whether people reasonably expect that their movements will be recorded and aggregated in a manner that enables the Government to ascertain, more or less at will, their political and religious beliefs, sexual habits, and so on. I do not regard as dispositive the fact that the Government might obtain the fruits of GPS monitoring through lawful conventional surveillance techniques. See *Kyllo*, 533 U. S., at 35, n. 2; *ante,* at 11 (leaving open the possibility that duplicating traditional surveillance "through electronic means, without an accompanying trespass, is an unconstitutional invasion of privacy"). I would also consider the appropriateness of entrusting to the Executive, in the absence of any oversight from a coordinate branch, a tool so amenable to misuse, especially in light of the Fourth Amendment's goal to curb arbitrary exercises of police power to and prevent "a too permeating police surveillance," *United States* v. *Di Re*, 332 U. S. 581, 595 (1948).*

---

* *United States* v. *Knotts*, 460 U. S. 276 (1983), does not foreclose the conclusion that GPS monitoring, in the absence of a physical intrusion, is a Fourth Amendment search. As the majority's opinion notes, *Knotts* reserved the question whether "'different constitutional principles may be applicable'" to invasive law enforcement practices such as GPS tracking. See *ante,* at 8, n. 6 (quoting 460 U. S., at 284).

United States v. *Karo*, 468 U. S. 705 (1984), addressed the Fourth

More fundamentally, it may be necessary to reconsider the premise that an individual has no reasonable expectation of privacy in information voluntarily disclosed to third parties. *E.g., Smith*, 442 U. S., at 742; *United States* v. *Miller*, 425 U. S. 435, 443 (1976). This approach is ill suited to the digital age, in which people reveal a great deal of information about themselves to third parties in the course of carrying out mundane tasks. People disclose the phone numbers that they dial or text to their cellular providers; the URLs that they visit and the e-mail addresses with which they correspond to their Internet service providers; and the books, groceries, and medications they purchase to online retailers. Perhaps, as JUSTICE ALITO notes, some people may find the "tradeoff" of privacy for convenience "worthwhile," or come to accept this "diminution of privacy" as "inevitable," *post,* at 10, and perhaps not. I for one doubt that people would accept without complaint the warrantless disclosure to the Government of a list of every Web site they had visited in the last week, or month, or year. But whatever the societal expectations, they can attain constitutionally protected status only if our Fourth Amendment jurisprudence ceases

---

Amendment implications of the installation of a beeper in a container with the consent of the container's original owner, who was aware that the beeper would be used for surveillance purposes. *Id.,* at 707. Owners of GPS-equipped cars and smartphones do not contemplate that these devices will be used to enable covert surveillance of their movements. To the contrary, subscribers of one such service greeted a similar suggestion with anger. Quain, Changes to OnStar's Privacy Terms Rile Some Users, N. Y. Times (Sept. 22, 2011), online at http://wheels.blogs.nytimes.com/2011/09/22/changes-to-onstars-privacy-terms-rile-some-users (as visited Jan. 19, 2012, and available in Clerk of Court's case file). In addition, the bugged container in *Karo* lacked the close relationship with the target that a car shares with its owner. The bugged container in *Karo* was stationary for much of the Government's surveillance. See 468 U. S., at 708–710. A car's movements, by contrast, are its owner's movements.

to treat secrecy as a prerequisite for privacy. I would not assume that all information voluntarily disclosed to some member of the public for a limited purpose is, for that reason alone, disentitled to Fourth Amendment protection. See *Smith*, 442 U. S., at 749 (Marshall, J., dissenting) ("Privacy is not a discrete commodity, possessed absolutely or not at all. Those who disclose certain facts to a bank or phone company for a limited business purpose need not assume that this information will be released to other persons for other purposes"); see also *Katz*, 389 U. S., at 351–352 ("[W]hat [a person] seeks to preserve as private, even in an area accessible to the public, may be constitutionally protected").

Resolution of these difficult questions in this case is unnecessary, however, because the Government's physical intrusion on Jones' Jeep supplies a narrower basis for decision. I therefore join the majority's opinion.

SUPREME COURT OF THE UNITED STATES

No. 10–1259

UNITED STATES, PETITIONER *v.* ANTOINE JONES

ON WRIT OF CERTIORARI TO THE UNITED STATES COURT OF APPEALS FOR THE DISTRICT OF COLUMBIA CIRCUIT

[January 23, 2012]

JUSTICE ALITO, with whom JUSTICE GINSBURG, JUSTICE BREYER, and JUSTICE KAGAN join, concurring in the judgment.

This case requires us to apply the Fourth Amendment's prohibition of unreasonable searches and seizures to a 21st-century surveillance technique, the use of a Global Positioning System (GPS) device to monitor a vehicle's movements for an extended period of time. Ironically, the Court has chosen to decide this case based on 18th-century tort law. By attaching a small GPS device[1] to the underside of the vehicle that respondent drove, the law enforcement officers in this case engaged in conduct that might have provided grounds in 1791 for a suit for trespass to chattels.[2] And for this reason, the Court concludes, the installation and use of the GPS device constituted a search. *Ante*, at 3–4.

---

[1] Although the record does not reveal the size or weight of the device used in this case, there is now a device in use that weighs two ounces and is the size of a credit card. Tr. of Oral Arg. 27.

[2] At common law, a suit for trespass to chattels could be maintained if there was a violation of "the dignitary interest in the inviolability of chattels," but today there must be "some actual damage to the chattel before the action can be maintained." W. Keeton, D. Dobbs, R. Keeton, & D. Owen, Prosser & Keeton on Law of Torts 87 (5th ed. 1984) (hereinafter Prosser & Keeton). Here, there was no actual damage to the vehicle to which the GPS device was attached.

This holding, in my judgment, is unwise. It strains the language of the Fourth Amendment; it has little if any support in current Fourth Amendment case law; and it is highly artificial.

I would analyze the question presented in this case by asking whether respondent's reasonable expectations of privacy were violated by the long-term monitoring of the movements of the vehicle he drove.

I

A

The Fourth Amendment prohibits "unreasonable searches and seizures," and the Court makes very little effort to explain how the attachment or use of the GPS device fits within these terms. The Court does not contend that there was a seizure. A seizure of property occurs when there is "some meaningful interference with an individual's possessory interests in that property," *United States* v. *Jacobsen,* 466 U. S. 109, 113 (1984), and here there was none. Indeed, the success of the surveillance technique that the officers employed was dependent on the fact that the GPS did not interfere in any way with the operation of the vehicle, for if any such interference had been detected, the device might have been discovered.

The Court does claim that the installation and use of the GPS constituted a search, see *ante,* at 3–4, but this conclusion is dependent on the questionable proposition that these two procedures cannot be separated for purposes of Fourth Amendment analysis. If these two procedures are analyzed separately, it is not at all clear from the Court's opinion why either should be regarded as a search. It is clear that the attachment of the GPS device was not itself a search; if the device had not functioned or if the officers had not used it, no information would have been obtained. And the Court does not contend that the use of the device constituted a search either. On the contrary, the Court

accepts the holding in *United States* v. *Knotts*, 460 U. S. 276 (1983), that the use of a surreptitiously planted electronic device to monitor a vehicle's movements on public roads did not amount to a search. See *ante*, at 7.

The Court argues—and I agree—that "we must 'assur[e] preservation of that degree of privacy against government that existed when the Fourth Amendment was adopted.'" *Ante*, at 5 (quoting *Kyllo* v. *United States*, 533 U. S. 27, 34 (2001)). But it is almost impossible to think of late-18th-century situations that are analogous to what took place in this case. (Is it possible to imagine a case in which a constable secreted himself somewhere in a coach and remained there for a period of time in order to monitor the movements of the coach's owner?[3]) The Court's theory seems to be that the concept of a search, as originally understood, comprehended any technical trespass that led to the gathering of evidence, but we know that this is incorrect. At common law, any unauthorized intrusion on private property was actionable, see Prosser & Keeton 75, but a trespass on open fields, as opposed to the "curtilage" of a home, does not fall within the scope of the Fourth Amendment because private property outside the curtilage is not part of a "hous[e]" within the meaning of the Fourth Amendment. See *Oliver* v. *United State*s, 466 U. S. 170 (1984); *Hester* v. *United States*, 265 U. S. 57 (1924).

B

The Court's reasoning in this case is very similar to that in the Court's early decisions involving wiretapping and electronic eavesdropping, namely, that a technical trespass followed by the gathering of evidence constitutes a

---

[3] The Court suggests that something like this might have occurred in 1791, but this would have required either a gigantic coach, a very tiny constable, or both—not to mention a constable with incredible fortitude and patience.

search. In the early electronic surveillance cases, the Court concluded that a Fourth Amendment search occurred when private conversations were monitored as a result of an "unauthorized physical penetration into the premises occupied" by the defendant. *Silverman* v. *United States*, 365 U. S. 505, 509 (1961). In *Silverman*, police officers listened to conversations in an attached home by inserting a "spike mike" through the wall that this house shared with the vacant house next door. *Id.,* at 506. This procedure was held to be a search because the mike made contact with a heating duct on the other side of the wall and thus "usurp[ed] . . . an integral part of the premises." *Id.,* at 511.

By contrast, in cases in which there was no trespass, it was held that there was no search. Thus, in *Olmstead* v. *United States*, 277 U. S. 438 (1928), the Court found that the Fourth Amendment did not apply because "[t]he taps from house lines were made in the streets near the houses." *Id.*, at 457. Similarly, the Court concluded that no search occurred in *Goldman* v. *United States*, 316 U. S. 129, 135 (1942), where a "detectaphone" was placed on the outer wall of defendant's office for the purpose of overhearing conversations held within the room.

This trespass-based rule was repeatedly criticized. In *Olmstead*, Justice Brandeis wrote that it was "immaterial where the physical connection with the telephone wires was made." 277 U. S., at 479 (dissenting opinion). Although a private conversation transmitted by wire did not fall within the literal words of the Fourth Amendment, he argued, the Amendment should be understood as prohibiting "every unjustifiable intrusion by the government upon the privacy of the individual." *Id.,* at 478. See also, *e.g.*, *Silverman*, *supra*, at 513 (Douglas, J., concurring) ("The concept of 'an unauthorized physical penetration into the premises,' on which the present decision rests seems to me beside the point. Was not the wrong . . . done when the

intimacies of the home were tapped, recorded, or revealed? The depth of the penetration of the electronic device—even the degree of its remoteness from the inside of the house—is not the measure of the injury"); *Goldman*, *supra*, at 139 (Murphy, J., dissenting) ("[T]he search of one's home or office no longer requires physical entry, for science has brought forth far more effective devices for the invasion of a person's privacy than the direct and obvious methods of oppression which were detested by our forebears and which inspired the Fourth Amendment").

Katz v. *United States*, 389 U. S. 347 (1967), finally did away with the old approach, holding that a trespass was not required for a Fourth Amendment violation. *Katz* involved the use of a listening device that was attached to the outside of a public telephone booth and that allowed police officers to eavesdrop on one end of the target's phone conversation. This procedure did not physically intrude on the area occupied by the target, but the *Katz* Court "repudiate[ed]" the old doctrine, *Rakas* v. *Illinois*, 439 U. S. 128, 143 (1978), and held that "[t]he fact that the electronic device employed . . . did not happen to penetrate the wall of the booth can have no constitutional significance," 389 U. S., at 353 ("[T]he reach of th[e] [Fourth] Amendment cannot turn upon the presence or absence of a physical intrusion into any given enclosure"); see *Rakas*, *supra*, at 143 (describing *Katz* as holding that the "capacity to claim the protection for the Fourth Amendment depends not upon a property right in the invaded place but upon whether the person who claims the protection of the Amendment has a legitimate expectation of privacy in the invaded place"); *Kyllo, supra,* at 32 ("We have since decoupled violation of a person's Fourth Amendment rights from trespassory violation of his property"). What mattered, the Court now held, was whether the conduct at issue "violated the privacy upon which [the defendant] justifiably relied while using the telephone booth." *Katz, supra*,

at 353.

Under this approach, as the Court later put it when addressing the relevance of a technical trespass, "an actual trespass is neither necessary *nor sufficient* to establish a constitutional violation." *United States* v. *Karo,* 468 U. S. 705, 713 (1984) (emphasis added). *Ibid.* ("Compar[ing] *Katz* v. *United States*, 389 U. S. 347 (1967) (no trespass, but Fourth Amendment violation), with *Oliver* v. *United States*, 466 U. S. 170 (1984) (trespass, but no Fourth Amendment violation)"). In *Oliver,* the Court wrote:

> "The existence of a property right is but one element in determining whether expectations of privacy are legitimate. 'The premise that property interests control the right of the Government to search and seize has been discredited.' *Katz*, 389 U. S., at 353, (quoting *Warden* v. *Hayden*, 387 U. S. 294, 304 (1967); some internal quotation marks omitted)." 466 U. S., at 183.

II

The majority suggests that two post-*Katz* decisions—*Soldal* v. *Cook County*, 506 U. S. 56 (1992), and *Alderman* v. *United States*, 394 U. S. 165 (1969)—show that a technical trespass is sufficient to establish the existence of a search, but they provide little support.

In *Soldal*, the Court held that towing away a trailer home without the owner's consent constituted a seizure even if this did not invade the occupants' personal privacy. But in the present case, the Court does not find that there was a seizure, and it is clear that none occurred.

In *Alderman,* the Court held that the Fourth Amendment rights of homeowners were implicated by the use of a surreptitiously planted listening device to monitor third-party conversations that occurred within their home. See 394 U. S., at 176–180. *Alderman* is best understood to

mean that the homeowners had a legitimate expectation of privacy in all conversations that took place under their roof. See *Rakas*, 439 U. S., at 144, n. 12 (citing *Alderman* for the proposition that "the Court has not altogether abandoned use of property concepts in determining the presence or absence of the privacy interests protected by that Amendment"); 439 U. S., at 153 (Powell, J., concurring) (citing *Alderman* for the proposition that "property rights reflect society's explicit recognition of a person's authority to act as he wishes in certain areas, and therefore should be considered in determining whether an individual's expectations of privacy are reasonable); *Karo*, *supra*, at 732 (Stevens, J., concurring in part and dissenting in part) (citing *Alderman* in support of the proposition that "a homeowner has a reasonable expectation of privacy in the contents of his home, including items owned by others").

In sum, the majority is hard pressed to find support in post-*Katz* cases for its trespass-based theory.

III

Disharmony with a substantial body of existing case law is only one of the problems with the Court's approach in this case.

I will briefly note four others. First, the Court's reasoning largely disregards what is really important (the *use* of a GPS for the purpose of long-term tracking) and instead attaches great significance to something that most would view as relatively minor (attaching to the bottom of a car a small, light object that does not interfere in any way with the car's operation). Attaching such an object is generally regarded as so trivial that it does not provide a basis for recovery under modern tort law. See Prosser & Keeton §14, at 87 (harmless or trivial contact with personal property not actionable); D. Dobbs, Law of Torts 124 (2000) (same). But under the Court's reasoning, this conduct

may violate the Fourth Amendment. By contrast, if long-term monitoring can be accomplished without committing a technical trespass—suppose, for example, that the Federal Government required or persuaded auto manufacturers to include a GPS tracking device in every car—the Court's theory would provide no protection.

Second, the Court's approach leads to incongruous results. If the police attach a GPS device to a car and use the device to follow the car for even a brief time, under the Court's theory, the Fourth Amendment applies. But if the police follow the same car for a much longer period using unmarked cars and aerial assistance, this tracking is not subject to any Fourth Amendment constraints.

In the present case, the Fourth Amendment applies, the Court concludes, because the officers installed the GPS device after respondent's wife, to whom the car was registered, turned it over to respondent for his exclusive use. See *ante*, at 8. But if the GPS had been attached prior to that time, the Court's theory would lead to a different result. The Court proceeds on the assumption that respondent "had at least the property rights of a bailee," *ante*, at 3, n. 2, but a bailee may sue for a trespass to chattel only if the injury occurs during the term of the bailment. See 8A Am. Jur. 2d, Bailment §166, pp. 685–686 (2009). So if the GPS device had been installed before respondent's wife gave him the keys, respondent would have no claim for trespass—and, presumably, no Fourth Amendment claim either.

Third, under the Court's theory, the coverage of the Fourth Amendment may vary from State to State. If the events at issue here had occurred in a community property State[4] or a State that has adopted the Uniform Marital

---

[4] See, *e.g.*, Cal. Family Code Ann. §760 (West 2004).

Property Act,[5] respondent would likely be an owner of the vehicle, and it would not matter whether the GPS was installed before or after his wife turned over the keys. In non-community-property States, on the other hand, the registration of the vehicle in the name of respondent's wife would generally be regarded as presumptive evidence that she was the sole owner. See 60 C. J. S., Motor Vehicles §231, pp. 398–399 (2002); 8 Am. Jur. 2d, Automobiles §1208, pp. 859–860 (2007).

Fourth, the Court's reliance on the law of trespass will present particularly vexing problems in cases involving surveillance that is carried out by making electronic, as opposed to physical, contact with the item to be tracked. For example, suppose that the officers in the present case had followed respondent by surreptitiously activating a stolen vehicle detection system that came with the car when it was purchased. Would the sending of a radio signal to activate this system constitute a trespass to chattels? Trespass to chattels has traditionally required a physical touching of the property. See Restatement (Second) of Torts §217 and Comment *e* (1963 and 1964); Dobbs, *supra*, at 123. In recent years, courts have wrestled with the application of this old tort in cases involving unwanted electronic contact with computer systems, and some have held that even the transmission of electrons that occurs when a communication is sent from one computer to another is enough. See, *e.g., CompuServe, Inc.* v. *Cyber Promotions*, *Inc.* 962 F. Supp. 1015, 1021 (SD Ohio 1997); *Thrifty-Tel, Inc.* v. *Bezenek*, 46 Cal. App. 4th 1559, 1566, n. 6 (1996). But may such decisions be followed in applying the Court's trespass theory? Assuming that what matters under the Court's theory is the law of trespass as it existed at the time of the adoption of the Fourth

---

[5] See Uniform Marital Property Act §4, 9A U. L. A. 116 (1998).

Amendment, do these recent decisions represent a change in the law or simply the application of the old tort to new situations?

IV

A

The *Katz* expectation-of-privacy test avoids the problems and complications noted above, but it is not without its own difficulties. It involves a degree of circularity, see *Kyllo*, 533 U. S., at 34, and judges are apt to confuse their own expectations of privacy with those of the hypothetical reasonable person to which the *Katz* test looks. See *Minnesota* v. *Carter*, 525 U. S. 83, 97 (1998) (SCALIA, J., concurring). In addition, the *Katz* test rests on the assumption that this hypothetical reasonable person has a well-developed and stable set of privacy expectations. But technology can change those expectations. Dramatic technological change may lead to periods in which popular expectations are in flux and may ultimately produce significant changes in popular attitudes. New technology may provide increased convenience or security at the expense of privacy, and many people may find the tradeoff worthwhile. And even if the public does not welcome the diminution of privacy that new technology entails, they may eventually reconcile themselves to this development as inevitable.[6]

On the other hand, concern about new intrusions on privacy may spur the enactment of legislation to protect against these intrusions. This is what ultimately happened with respect to wiretapping. After *Katz*, Congress

---

[6] See, *e.g.*, NPR, The End of Privacy http://www.npr.org/series/114250076/the-end-of-privacy (all Internet materials as visited Jan. 20, 2012, and available in Clerk of Court's case file); Time Magazine, Everything About You Is Being Tracked—Get Over It, Joel Stein, Mar. 21, 2011, Vol. 177, No. 11.

did not leave it to the courts to develop a body of Fourth Amendment case law governing that complex subject. Instead, Congress promptly enacted a comprehensive statute, see 18 U. S. C. §§2510–2522 (2006 ed. and Supp. IV), and since that time, the regulation of wiretapping has been governed primarily by statute and not by case law.[7] In an ironic sense, although *Katz* overruled *Olmstead*, Chief Justice Taft's suggestion in the latter case that the regulation of wiretapping was a matter better left for Congress, see 277 U. S., at 465–466, has been borne out.

B

Recent years have seen the emergence of many new devices that permit the monitoring of a person's movements. In some locales, closed-circuit television video monitoring is becoming ubiquitous. On toll roads, automatic toll collection systems create a precise record of the movements of motorists who choose to make use of that convenience. Many motorists purchase cars that are equipped with devices that permit a central station to ascertain the car's location at any time so that roadside assistance may be provided if needed and the car may be found if it is stolen.

Perhaps most significant, cell phones and other wireless devices now permit wireless carriers to track and record the location of users—and as of June 2011, it has been reported, there were more than 322 million wireless devices in use in the United States.[8] For older phones, the accuracy of the location information depends on the density of the tower network, but new "smart phones," which

---

[7] See Kerr, The Fourth Amendment and New Technologies: Constitutional Myths and the Case for Caution, 102 Mich. L. Rev. 801, 850–851 (2004) (hereinafter Kerr).

[8] See CTIA Consumer Info, 50 Wireless Quick Facts, http://www.ctia.org/consumer_info/index.cfm/AID/10323.

are equipped with a GPS device, permit more precise tracking. For example, when a user activates the GPS on such a phone, a provider is able to monitor the phone's location and speed of movement and can then report back real-time traffic conditions after combining ("crowdsourcing") the speed of all such phones on any particular road.[9] Similarly, phone-location-tracking services are offered as "social" tools, allowing consumers to find (or to avoid) others who enroll in these services. The availability and use of these and other new devices will continue to shape the average person's expectations about the privacy of his or her daily movements.

V

In the pre-computer age, the greatest protections of privacy were neither constitutional nor statutory, but practical. Traditional surveillance for any extended period of time was difficult and costly and therefore rarely undertaken. The surveillance at issue in this case—constant monitoring of the location of a vehicle for four weeks—would have required a large team of agents, multiple vehicles, and perhaps aerial assistance.[10] Only an investigation of unusual importance could have justified such an

---

[9] See, *e.g.*, The bright side of sitting in traffic: Crowdsourcing road congestion data, Google Blog, http://googleblog.blogspot.com/2009/08/bright-side-of-sitting-in-traffic.html.

[10] Even with a radio transmitter like those used in *United States* v. *Knotts*, 460 U. S. 276 (1983), or *United States* v. *Karo,* 468 U. S. 705 (1984), such long-term surveillance would have been exceptionally demanding. The beepers used in those cases merely "emit[ted] periodic signals that [could] be picked up by a radio receiver." *Knotts*, 460 U.S., at 277. The signal had a limited range and could be lost if the police did not stay close enough. Indeed, in *Knotts* itself, officers lost the signal from the beeper, and only "with the assistance of a monitoring device located in a helicopter [was] the approximate location of the signal . . . picked up again about one hour later." *Id.*, at 278.

expenditure of law enforcement resources. Devices like the one used in the present case, however, make long-term monitoring relatively easy and cheap. In circumstances involving dramatic technological change, the best solution to privacy concerns may be legislative. See, *e.g.*, Kerr, 102 Mich. L. Rev., at 805–806. A legislative body is well situated to gauge changing public attitudes, to draw detailed lines, and to balance privacy and public safety in a comprehensive way.

To date, however, Congress and most States have not enacted statutes regulating the use of GPS tracking technology for law enforcement purposes. The best that we can do in this case is to apply existing Fourth Amendment doctrine and to ask whether the use of GPS tracking in a particular case involved a degree of intrusion that a reasonable person would not have anticipated.

Under this approach, relatively short-term monitoring of a person's movements on public streets accords with expectations of privacy that our society has recognized as reasonable. See *Knotts,* 460 U. S., at 281–282. But the use of longer term GPS monitoring in investigations of most offenses impinges on expectations of privacy. For such offenses, society's expectation has been that law enforcement agents and others would not—and indeed, in the main, simply could not—secretly monitor and catalogue every single movement of an individual's car for a very long period. In this case, for four weeks, law enforcement agents tracked every movement that respondent made in the vehicle he was driving. We need not identify with precision the point at which the tracking of this vehicle became a search, for the line was surely crossed before the 4-week mark. Other cases may present more difficult questions. But where uncertainty exists with respect to whether a certain period of GPS surveil

lance is long enough to constitute a Fourth Amendment search, the police may always seek a warrant.[11] We also need not consider whether prolonged GPS monitoring in the context of investigations involving extraordinary offenses would similarly intrude on a constitutionally protected sphere of privacy. In such cases, long-term tracking might have been mounted using previously available techniques.

* * *

For these reasons, I conclude that the lengthy monitoring that occurred in this case constituted a search under the Fourth Amendment. I therefore agree with the majority that the decision of the Court of Appeals must be affirmed.

---

[11] In this case, the agents obtained a warrant, but they did not comply with two of the warrant's restrictions: They did not install the GPS device within the 10-day period required by the terms of the warrant and by Fed. Rule Crim. Proc. 41(e)(2)(B)(i), and they did not install the GPS device within the District of Columbia, as required by the terms of the warrant and by 18 U. S. C. §3117(a) and Rule 41(b)(4). In the courts below the Government did not argue, and has not argued here, that the Fourth Amendment does not impose these precise restrictions and that the violation of these restrictions does not demand the suppression of evidence obtained using the tracking device. See, *e.g.*, *United States* v. *Gerber*, 994 F. 2d 1556, 1559–1560 (CA11 1993); *United States* v. *Burke*, 517 F. 2d 377, 386–387 (CA2 1975). Because it was not raised, that question is not before us.

FOR PUBLICATION

UNITED STATES COURT OF APPEALS
FOR THE NINTH CIRCUIT

UNITED STATES OF AMERICA, *Plaintiff-Appellant*, v. HOWARD WESLEY COTTERMAN, *Defendant-Appellee*.	No. 09-10139 D.C. No. 4:07-cr-01207- RCC-CRP-1 OPINION

Appeal from the United States District Court
for the District of Arizona
Raner C. Collins, District Judge, Presiding

Argued and Submitted En Banc
June 19, 2012—Pasadena, California

Filed March 8, 2013

Before: Alex Kozinski, Chief Judge, Sidney R. Thomas, M. Margaret McKeown, Kim McLane Wardlaw, Raymond C. Fisher, Ronald M. Gould, Richard R. Clifton, Consuelo M. Callahan, Milan D. Smith, Jr., Mary H. Murguia, and Morgan Christen, Circuit Judges.[1]

Opinion by Judge McKeown;
Partial Concurrence and Partial Dissent by Judge Callahan;
Dissent by Judge Milan D. Smith, Jr.

---

[1] Judge Betty B. Fletcher was a member of the en banc panel but passed away after argument of the case. Judge Wardlaw was drawn as her replacement.

SUMMARY*

Criminal Law

The en banc court reversed the district court's order suppressing evidence of child pornography obtained from a forensic examination of the defendant's laptop, which was seized by agents at the U.S.-Mexico border in response to an alert based in part on a prior conviction for child molestation.

The en banc court explained that a border search of a computer is not transformed into an "extended border search" requiring particularized suspicion simply because the device is transported and examined beyond the border. The en banc court wrote that the fact that the forensic examination occurred 170 miles away from the border did not heighten the interference with the defendant's privacy, and the extended border search doctrine does not apply, in this case in which the defendant's computer never cleared customs and the defendant never regained possession.

The en banc court held that the forensic examination of the defendant's computer required a showing of reasonable suspicion, a modest requirement in light of the Fourth Amendment. The en banc court wrote that it is the comprehensive and intrusive nature of forensic examination – not the location of the examination – that is the key factor triggering the requirement of reasonable suspicion here. The en banc court wrote that the uniquely sensitive nature of data on electronic devices, which often retain information far

* This summary constitutes no part of the opinion of the court. It has been prepared by court staff for the convenience of the reader.

beyond the perceived point of erasure, carries with it a significant expectation of privacy and thus renders an exhaustive exploratory search more intrusive than with other forms of property.

The en banc court held that the border agents had reasonable suspicion to conduct an initial search at the border (which turned up no incriminating material) and the forensic examination. The en banc court wrote that the defendant's Treasury Enforcement Communication System alert, prior child-related conviction, frequent travels, crossing from a country known for sex tourism, and collection of electronic equipment, plus the parameters of the Operation Angel Watch program aimed at combating child sex tourism, taken collectively, gave rise to reasonable suspicion of criminal activity.

The en banc court wrote that password protection of files, which is ubiquitous among many law-abiding citizens, will not in isolation give rise to reasonable suspicion, but that password protection may be considered in the totality of the circumstances where, as here, there are other indicia of criminal activity. The en banc court wrote that the existence of password-protected files is also relevant to assessing the reasonableness of the scope and duration of the search of the defendant's computer.

The en banc court concluded that the examination of the defendant's electronic devices was supported by reasonable suspicion and that the scope and manner of the search were reasonable under the Fourth Amendment.

Concurring in part, dissenting in part, and concurring in the judgment, Judge Callahan (with whom Judge Clifton

joined and with whom Judge M. Smith joined as to all but Part II.A) wrote that the majority's new rule requiring reasonable suspicion for any thorough search of electronic devices entering the United States flouts more than a century of Supreme Court precedent, is unworkable and unnecessary, and will severely hamstring the government's ability to protect our borders.

Judge M. Smith (with whom Judges Clifton and Callahan joined with respect to Part I) dissented. Judge Smith wrote that the majority's decision to create a reasonable suspicion requirement for some property searches at the border so muddies current border search doctrine that border agents will be left to divine on an ad hoc basis whether a property search is sufficiently "comprehensive and intrusive" to require suspicion, or sufficiently "unintrusive" to come within the traditional border search exception. Judge Smith also wrote that the majority's determination that reasonable suspicion exists under the exceedingly weak facts of this case undermines the liberties of U.S. citizens generally – not just at the border, and not just with regard to our digital data – but on every street corner, in every vehicle, and wherever else we rely on the doctrine of reasonable suspicion to safeguard our legitimate privacy interests.

COUNSEL

Dennis K. Burke, Christina M. Cabanillas, Carmen F. Corbin, John S. Leonardo, John J. Tuchi, United States Attorney's Office for the District of Arizona, Tucson, Arizona, for Appellant.

William J. Kirchner, Law Office of Nash & Kirchner, P.C., Tucson, Arizona, for Appellee.

David M. Porter, Malia N. Brink, National Association of Criminal Defense Lawyers, Washington, D.C.; Michael Price, Brennan Center for Justice, New York, New York; Hanni M. Fakhoury, Electronic Frontier Foundation, San Francisco, California, for Amicus Curiae National Association of Criminal Defense Lawyers and Electronic Frontier Foundation.

Christopher T. Handman, Mary Helen Wimberly, Hogan Lovells US LLP, Washington, D.C.; Sharon Bradford Franklin, The Constitution Project, Washington, D.C., for Amicus Curiae The Constitution Project.

OPINION

McKEOWN, Circuit Judge:

Every day more than a million people cross American borders, from the physical borders with Mexico and Canada to functional borders at airports such as Los Angeles (LAX), Honolulu (HNL), New York (JFK, LGA), and Chicago (ORD, MDW). As denizens of a digital world, they carry with them laptop computers, iPhones, iPads, iPods, Kindles,

Nooks, Surfaces, tablets, Blackberries, cell phones, digital cameras, and more. These devices often contain private and sensitive information ranging from personal, financial, and medical data to corporate trade secrets. And, in the case of Howard Cotterman, child pornography.

Agents seized Cotterman's laptop at the U.S.-Mexico border in response to an alert based in part on a fifteen-year-old conviction for child molestation. The initial search at the border turned up no incriminating material. Only after Cotterman's laptop was shipped almost 170 miles away and subjected to a comprehensive forensic examination were images of child pornography discovered.

This watershed case implicates both the scope of the narrow border search exception to the Fourth Amendment's warrant requirement and privacy rights in commonly used electronic devices. The question we confront "is what limits there are upon this power of technology to shrink the realm of guaranteed privacy." *Kyllo v. United States*, 533 U.S. 27, 34 (2001). More specifically, we consider the reasonableness of a computer search that began as a cursory review at the border but transformed into a forensic examination of Cotterman's hard drive.

Computer forensic examination is a powerful tool capable of unlocking password-protected files, restoring deleted material, and retrieving images viewed on web sites. But while technology may have changed the expectation of privacy to some degree, it has not eviscerated it, and certainly not with respect to the gigabytes of data regularly maintained as private and confidential on digital devices. Our Founders were indeed prescient in specifically incorporating "papers" within the Fourth Amendment's guarantee of "[t]he right of

the people to be secure in their persons, houses, papers, and effects." U.S. Const. amend. IV. The papers we create and maintain not only in physical but also in digital form reflect our most private thoughts and activities.

Although courts have long recognized that border searches constitute a "historically recognized exception to the Fourth Amendment's general principle that a warrant be obtained," *United States v. Ramsey*, 431 U.S. 606, 621 (1977), reasonableness remains the touchstone for a warrantless search. Even at the border, we have rejected an "anything goes" approach. *See United States v. Seljan*, 547 F.3d 993, 1000 (9th Cir. 2008) (en banc).

Mindful of the heavy burden on law enforcement to protect our borders juxtaposed with individual privacy interests in data on portable digital devices, we conclude that, under the circumstances here, reasonable suspicion was required for the forensic examination of Cotterman's laptop. Because border agents had such a reasonable suspicion, we reverse the district court's order granting Cotterman's motion to suppress the evidence of child pornography obtained from his laptop.

I. FACTUAL BACKGROUND AND PROCEDURAL HISTORY[2]

Howard Cotterman and his wife were driving home to the United States from a vacation in Mexico on Friday morning, April 6, 2007, when they reached the Lukeville, Arizona, Port of Entry. During primary inspection by a border agent, the

[2] The facts related here are drawn from the record of the evidentiary hearing held before the magistrate judge.

Treasury Enforcement Communication System ("TECS")[3] returned a hit for Cotterman. The TECS hit indicated that Cotterman was a sex offender—he had a 1992 conviction for two counts of use of a minor in sexual conduct, two counts of lewd and lascivious conduct upon a child, and three counts of child molestation—and that he was potentially involved in child sex tourism. Because of the hit, Cotterman and his wife were referred to secondary inspection, where they were instructed to exit their vehicle and leave all their belongings in the car. The border agents called the contact person listed in the TECS entry and, following that conversation, believed the hit to reflect Cotterman's involvement "in some type of child pornography." The agents searched the vehicle and retrieved two laptop computers and three digital cameras. Officer Antonio Alvarado inspected the electronic devices and found what appeared to be family and other personal photos, along with several password-protected files.

Border agents contacted Group Supervisor Craig Brisbine at the Immigration and Customs Enforcement ("ICE") office in Sells, Arizona, and informed him about Cotterman's entry and the fact that he was a sex offender potentially involved in child sex tourism. The Sells Duty Agent, Mina Riley, also spoke with Officer Alvarado and then contacted the ICE Pacific Field Intelligence Unit, the office listed on the TECS hit, to get more information. That unit informed Riley that the alert was part of Operation Angel Watch, which was aimed at combating child sex tourism by identifying registered sex offenders in California, particularly those who travel frequently outside the United States. She was advised

[3] The TECS is an investigative tool of the Department of Homeland Security that keeps track of individuals entering and exiting the country and of individuals involved in or suspected to be involved in crimes.

to review any media equipment, such as computers, cameras, or other electronic devices, for potential evidence of child pornography. Riley then spoke again to Alvarado, who told her that he had been able to review some of the photographs on the Cottermans' computers but had encountered password-protected files that he was unable to access.

Agents Brisbine and Riley departed Sells for Lukeville at about 1:30 p.m. and decided en route to detain the Cottermans' laptops for forensic examination. Upon their arrival, they gave Cotterman and his wife *Miranda* warnings and interviewed them separately. The interviews revealed nothing incriminating. During the interview, Cotterman offered to help the agents access his computer. The agents declined the offer out of concern that Cotterman might be able to delete files surreptitiously or that the laptop might be "booby trapped."

The agents allowed the Cottermans to leave the border crossing around 6 p.m., but retained the Cottermans' laptops and a digital camera.[4] Agent Brisbine drove almost 170 miles from Lukeville to the ICE office in Tucson, Arizona, where he delivered both laptops and one of the three digital cameras to ICE Senior Special Agent & Computer Forensic Examiner John Owen. Agent Owen began his examination on Saturday, the following day. He used a forensic program to copy the hard drives of the electronic devices. He determined that the digital camera did not contain any contraband and released the camera that day to the Cottermans, who had traveled to Tucson from Lukeville and planned to stay there a few days. Agent Owen then used forensic software that often must run for several hours to examine copies of the laptop hard drives.

[4] The other two cameras were returned to the Cottermans.

He began his personal examination of the laptops on Sunday. That evening, Agent Owen found seventy-five images of child pornography within the unallocated space of Cotterman's laptop.[5]

Agent Owen contacted the Cottermans on Sunday evening and told them he would need Howard Cotterman's assistance to access password-protected files he found on Cotterman's laptop. Cotterman agreed to provide the assistance the following day, but never showed up. When Agent Brisbine called again to request Cotterman's help in accessing the password-protected files, Cotterman responded that the computer had multiple users and that he would need to check with individuals at the company from which he had retired in order to get the passwords. The agents had no further contact with Cotterman, who boarded a flight to Mexico from Tucson the next day, April 9, and then flew onward to Sydney, Australia. On April 11, Agent Owen finally managed to open twenty-three password-protected files on Cotterman's laptop. The files revealed approximately 378 images of child pornography. The vast majority of the images were of the same girl, approximately 7–10 years of age, taken over a two- to three-year period. In many of the images, Cotterman was sexually molesting the child. Over the next few months, Agent Owen discovered hundreds more pornographic images, stories, and videos depicting children.

[5] "Unallocated space is space on a hard drive that contains deleted data, usually emptied from the operating system's trash or recycle bin folder, that cannot be seen or accessed by the user without the use of forensic software. Such space is available to be written over to store new information." *United States v. Flyer*, 633 F.3d 911, 918 (9th Cir. 2011).

A grand jury indicted Cotterman for a host of offenses related to child pornography. Cotterman moved to suppress the evidence gathered from his laptop and the fruits of that evidence. The magistrate judge filed a Report and Recommendation finding that the forensic examination was an "extended border search" that required reasonable suspicion. He found that the TECS hit and the existence of password-protected files on Cotterman's laptop were suspicious, but concluded that those facts did not suffice to give rise to reasonable suspicion of criminal activity. The district judge adopted the Report and Recommendation and granted Cotterman's motion to suppress.

In its interlocutory appeal of that order, the government characterized the issue as follows: "Whether the authority to search a laptop computer *without reasonable suspicion* at a border point of entry permits law enforcement to take it to another location to be forensically examined, when it has remained in the continuous custody of the government." A divided panel of this court answered that question in the affirmative and reversed. *United States v. Cotterman*, 637 F.3d 1068 (9th Cir. 2011). The panel concluded that reasonable suspicion was not required for the search and that "[t]he district court erred in suppressing the evidence lawfully obtained under border search authority." *Id.* at 1084. In dissent, Judge Betty B. Fletcher wrote that "officers must have some level of particularized suspicion in order to conduct a seizure and search like the one at issue here." *Id.* (B. Fletcher, J., dissenting). By a vote of a majority of nonrecused active judges, rehearing en banc was ordered. 673 F.3d 1206 (9th Cir. 2012). Following en banc oral argument, we requested supplemental briefing on the issue of whether reasonable suspicion existed at the time of the search.

II. WAIVER

The government argued below that the forensic examination was part of a routine border search not requiring heightened suspicion and, alternatively, that reasonable suspicion justified the search. Before the district court, the government maintained "the facts of this case clearly establish that there was reasonable suspicion." However, having failed to obtain a favorable ruling on that ground, the government did not challenge on appeal the conclusion that there was no reasonable suspicion. Rather, it sought a broad ruling that no suspicion of any kind was required. Cotterman thus argued in his answering brief that the government had waived the issue—an assertion that the government did not address in its reply brief. Cotterman contends that the government has abandoned and conceded the issue of reasonable suspicion and that this court may not address that issue. We disagree.

We review de novo the ultimate question of whether a warrantless search was reasonable under the Fourth Amendment. *United States v. Johnson*, 256 F.3d 895, 905 (9th Cir. 2001) (en banc). Our review necessarily encompasses a determination as to the applicable standard: no suspicion, reasonable suspicion or probable cause. That the government may hope for the lowest standard does not alter our de novo review, particularly when the issue was fully briefed and argued below. Further, we may consider an issue that has not been adequately raised on appeal if such a failure will not prejudice the opposing party. *United States v. Ullah*, 976 F.2d 509, 514 (9th Cir. 1992). Where, as here, we "called for and received supplemental briefs by both parties," *Alcarez v. INS*, 384 F.3d 1150, 1161 (9th Cir. 2004), the government's failure to address the issue does not prejudice

Cotterman. *See also United States v. Resendiz-Ponce*, 549 U.S. 102, 103–04 (2007).

III. THE BORDER SEARCH

The broad contours of the scope of searches at our international borders are rooted in "the long-standing right of the sovereign to protect itself by stopping and examining persons and property crossing into this country." *Ramsey*, 431 U.S. at 616. Thus, border searches form "a narrow exception to the Fourth Amendment prohibition against warrantless searches without probable cause." *Seljan*, 547 F.3d at 999 (internal quotation marks and citation omitted). Because "[t]he Government's interest in preventing the entry of unwanted persons and effects is at its zenith at the international border," *United States v. Flores-Montano*, 541 U.S. 149, 152 (2004), border searches are generally deemed "reasonable simply by virtue of the fact that they occur at the border." *Ramsey*, 431 U.S. at 616.

This does not mean, however, that at the border "anything goes." *Seljan*, 547 F.3d at 1000. Even at the border, individual privacy rights are not abandoned but "[b]alanced against the sovereign's interests." *United States v. Montoya de Hernandez*, 473 U.S. 531, 539 (1985). That balance "is qualitatively different . . . than in the interior" and is "struck much more favorably to the Government." *Id.* at 538, 540. Nonetheless, the touchstone of the Fourth Amendment analysis remains reasonableness. *Id.* at 538. The reasonableness of a search or seizure depends on the totality of the circumstances, including the scope and duration of the deprivation. *See United States v. Jacobsen*, 466 U.S. 109, 124 (1984); *see also United States v. Duncan*, 693 F.2d 971, 977 (9th Cir. 1982).

In view of these principles, the legitimacy of the initial search of Cotterman's electronic devices at the border is not in doubt. Officer Alvarado turned on the devices and opened and viewed image files while the Cottermans waited to enter the country. It was, in principle, akin to the search in *Seljan*, where we concluded that a suspicionless cursory scan of a package in international transit was not unreasonable. 547 F.3d at 1004. Similarly, we have approved a quick look and unintrusive search of laptops. *United States v. Arnold*, 533 F.3d 1003, 1009 (9th Cir. 2008) (holding border search reasonable where "CBP officers simply 'had [traveler] boot [the laptop] up, and looked at what [he] had inside.'") (second alteration in original).[6] Had the search of Cotterman's laptop ended with Officer Alvarado, we would be inclined to conclude it was reasonable even without particularized suspicion. *See id.* But the search here transformed into something far different. The difficult question we confront is the reasonableness, without a warrant, of the forensic examination that comprehensively analyzed the hard drive of the computer.

A. The Forensic Examination Was Not An Extended Border Search

Cotterman urges us to treat the examination as an extended border search that requires particularized suspicion.

---

[6] Although the *Arnold* decision expressed its conclusion in broad terms, stating that, "reasonable suspicion is not needed for customs officials to search a laptop or other personal electronic storage devices at the border," *Arnold*, 533 F.3d at 1008, the facts do not support such an unbounded holding. As an en banc court, we narrow *Arnold* to approve only the relatively simple search at issue in that case, not to countenance suspicionless forensic examinations. The dissent's extensive reliance on *Arnold* is misplaced in the en banc environment.

Although the semantic moniker "extended border search" may at first blush seem applicable here, our jurisprudence does not support such a claim. We have "define[d] an extended border search as any search away from the border where entry is not apparent, but where the dual requirements of reasonable certainty of a recent border crossing and reasonable suspicion of criminal activity are satisfied." *United States v. Guzman-Padilla*, 573 F.3d 865, 878–79 (9th Cir. 2009) (internal quotation marks and citations omitted). The key feature of an extended border search is that an individual can be assumed to have cleared the border and thus regained an expectation of privacy in accompanying belongings. *See United States v. Abbouchi*, 502 F.3d 850, 855 (9th Cir. 2007) ("Because the *delayed* nature of an extended border search . . . necessarily entails a greater level of intrusion on legitimate expectations of privacy than an ordinary border search, the government must justify an extended border search with reasonable suspicion that the search may uncover contraband or evidence of criminal activity.") (internal quotation marks omitted) (emphasis added).

Cotterman's case is different. Cotterman was stopped and searched at the border. Although he was allowed to depart the border inspection station after the initial search, some of his belongings, including his laptop, were not. The follow-on forensic examination was not an "extended border search." A border search of a computer is not transformed into an extended border search simply because the device is transported and examined beyond the border.

To be sure, our case law has not always articulated the "extended border search" doctrine with optimal clarity. But the confusion has come in distinguishing between facts

describing a functional border search and those describing an extended border search, not in defining the standard for a search at the border. *See, e.g.*, *United States v. Cardona*, 769 F.2d 625, 628 (9th Cir. 1985) ("We have recently recognized the difficulty of making sharp distinctions between searches at the functional equivalent of the border and extended border searches."). The "functional equivalent" doctrine effectively extends the border search doctrine to all ports of entry, including airports. *See Almeida-Sanchez v. United States*, 413 U.S. 266, 273 (1973). A routine customs search at the "functional equivalent" of the border is "analyzed as a border search" and requires neither probable cause nor reasonable suspicion. *Seljan*, 547 F.3d at 999. This case involves a search initiated at the actual border and does not encounter any of the difficulties surrounding identification of a "functional" border. As to the extended border search doctrine, we believe it is best confined to cases in which, after an apparent border crossing or functional entry, an attenuation in the time or the location of conducting a search reflects that the subject has regained an expectation of privacy.[7]

In his dissent, Judge Smith advocates applying the extended border search doctrine because the forensic examination occurred 170 miles from the border and days after Cotterman's entry. Moving the laptop to a specialized

[7] This characterization is consistent with how our circuit and others have articulated the doctrine. *See, e.g.*, *United States v. Villasenor*, 608 F.3d 467, 471–72 (9th Cir. 2010); *United States v. Yang*, 286 F.3d 940, 945–46 (7th Cir. 2002); *United States v. Hyde*, 37 F.3d 116, 120 n.2 (3d Cir. 1994); *United States v. Santiago*, 837 F.2d 1545, 1548 (11th Cir. 1988); *United States v. Gaviria*, 805 F.2d 1108, 1112 (2d Cir. 1986); *United States v. Niver*, 689 F.2d 520, 526 (5th Cir. 1982); *United States v. Bilir*, 592 F.2d 735, 739–40 (4th Cir. 1979).

lab at a distant location might highlight that the search undertaken there was an extensive one, but it is not the dispositive factor here. Because Cotterman never regained possession of his laptop, the fact that the forensic examination occurred away from the border, in Tucson, did not heighten the interference with his privacy. Time and distance become relevant to determining whether there is an adequate nexus to a recent border crossing only after the subject or items searched have entered. *See Villasenor*, 608 F.3d at 471 (explaining that reasonableness of extended border search depends on "whether the totality of the surrounding circumstances, including the time and distance elapsed" establish that items to be searched have recently entered the country) (internal quotation marks omitted). Cotterman's computer never cleared customs so entry was never effected. In short, the extended border search doctrine does not fit the search here.

B. Forensic Examination At The Border Requires Reasonable Suspicion

It is the comprehensive and intrusive nature of a forensic examination—not the location of the examination—that is the key factor triggering the requirement of reasonable suspicion here.[8] *See Cotterman*, 637 F.3d at 1086–87 n.6 (B. Fletcher, J., dissenting) (recognizing that "[a] computer search in a forensic lab will *always* be equivalent to an *identical* search at the border. The duration of a computer search is not

[8] The concurrence goes to great lengths to "refute any such notion" that location and duration contributed to our holding reasonable suspicion required here. Concurrence at 40–43. We see no reason for such an exegesis; our opinion is clear on the point that these factors are not at issue.

controlled by *where* the search is conducted. The duration of a computer search is controlled by what one is looking for and how one goes about searching for it.") (emphasis in original). The search would have been every bit as intrusive had Agent Owen traveled to the border with his forensic equipment. Indeed, Agent Owen had a laptop with forensic software that he could have used to conduct an examination at the port of entry itself, although he testified it would have been a more time-consuming effort. To carry out the examination of Cotterman's laptop, Agent Owen used computer forensic software to copy the hard drive and then analyze it in its entirety, including data that ostensibly had been deleted. This painstaking analysis is akin to reading a diary line by line looking for mention of criminal activity—plus looking at everything the writer may have erased.[9]

Notwithstanding a traveler's diminished expectation of privacy at the border, the search is still measured against the Fourth Amendment's reasonableness requirement, which considers the nature and scope of the search. Significantly, the Supreme Court has recognized that the "dignity and privacy interests of the person being searched" at the border will on occasion demand "some level of suspicion in the case of highly intrusive searches of the person." *Flores-Montano*, 541 U.S. at 152. Likewise, the Court has explained that "some searches of property are so destructive," "particularly offensive," or overly intrusive in the manner in which they

[9] Agent Owen used a software program called EnCase that exhibited the distinctive features of computer forensic examination. The program copied, analyzed, and preserved the data stored on the hard drive and gave the examiner access to far more data, including password-protected, hidden or encrypted, and deleted files, than a manual user could access.

are carried out as to require particularized suspicion. *Id.* at 152, 154 n.2, 155–56; *Montoya de Hernandez*, 473 U.S. at 541. The Court has never defined the precise dimensions of a reasonable border search, instead pointing to the necessity of a case-by-case analysis. As we have emphasized, "[r]easonableness, when used in the context of a border search, is incapable of comprehensive definition or of mechanical application." *Duncan*, 693 F.2d at 977 (internal quotation marks and citation omitted).

Over the past 30-plus years, the Supreme Court has dealt with a handful of border cases in which it reaffirmed the border search exception while, at the same time, leaving open the question of when a "particularly offensive" search might fail the reasonableness test. The trail begins with *United States v. Ramsey*, where the Court reserved judgment on this question: "We do not decide whether, and under what circumstances, a border search might be deemed 'unreasonable' because of the particularly offensive manner in which it is carried out." 431 U.S. at 618 n.13. Of note, the Court cited two cases, albeit non-border cases, as examples: *Kremen v. United States*, 353 U.S. 346, 347–48 (1957) (holding unconstitutional an exhaustive warrantless search of a cabin and seizure of its entire contents that were moved 200 miles away for examination) and *Go-Bart Importing Co. v. United States*, 282 U.S. 344, 358 (1931) (condemning as "lawless invasion of the premises and a general exploratory search" a warrantless "unlimited search, ransacking the desk, safe, filing cases and other parts of [an] office").

Less than ten years later, in 1985, the Court observed that it had "not previously decided what level of suspicion would justify a seizure of an incoming traveler for purposes other than a routine border search" and then went on to hold in the

context of an alimentary canal search that reasonable suspicion was required for "the detention of a traveler at the border, beyond the scope of a routine customs search and inspection." *Montoya de Hernandez*, 473 U.S. at 540–41. The Court's reference to "routine border search" was parsed in a later case, *Flores-Montano*, where the Court explained that "the reasons that might support a requirement of some level of suspicion in the case of highly intrusive searches of the person—dignity and privacy interests of the person being searched—simply do not carry over to vehicles," and, more specifically, to the gas tank of a car. 541 U.S. at 152. Accordingly, the Court rejected a privacy claim vis-a-vis an automobile gas tank.

We are now presented with a case directly implicating substantial personal privacy interests. The private information individuals store on digital devices—their personal "papers" in the words of the Constitution—stands in stark contrast to the generic and impersonal contents of a gas tank. *See, e.g.*, *United States v. Jones*, 132 S. Ct. 945, 957 (2012) (Sotomayor, J., concurring) (expressing "doubt that people would accept without complaint the warrantless disclosure to the Government of a list of every Web site they had visited in the last week, or month, or year"). We rest our analysis on the reasonableness of this search, paying particular heed to the nature of the electronic devices and the attendant expectation of privacy.

The amount of private information carried by international travelers was traditionally circumscribed by the size of the traveler's luggage or automobile. That is no longer the case. Electronic devices are capable of storing warehouses full of information. The average 400-gigabyte laptop hard drive can store over 200 million pages—the

equivalent of five floors of a typical academic library. *See* Orin S. Kerr, *Searches and Seizures in a Digital World*, 119 Harv. L. Rev. 531, 542 (2005) (explaining that an 80 GB hard drive is equivalent to 40 million pages or one floor of an academic library); *see also* LexisNexis, *How Many Pages in a Gigabyte?*, http://www.lexisnexis.com/applieddiscovery/lawlibrary/whitePapers/ADI_FS_PagesInAGigabyte.pdf. Even a car full of packed suitcases with sensitive documents cannot hold a candle to the sheer, and ever-increasing, capacity of digital storage.[10]

The nature of the contents of electronic devices differs from that of luggage as well. Laptop computers, iPads and the like are simultaneously offices and personal diaries. They contain the most intimate details of our lives: financial records, confidential business documents, medical records and private emails. This type of material implicates the Fourth Amendment's specific guarantee of the people's right to be secure in their "papers." U.S. Const. amend. IV. The express listing of papers "reflects the Founders' deep concern with safeguarding the privacy of thoughts and ideas—what we might call freedom of conscience—from invasion by the government." *Seljan*, 547 F.3d at 1014 (Kozinski, C.J., dissenting); *see also New York v. P.J. Video, Inc.*, 475 U.S. 868, 873 (1986). These records are expected to be kept

[10] We are puzzled by the dissent's speculation about "how many gigabytes of storage [one must] buy to secure the guarantee that reasonable suspicion will be required before one's devices are searched." Dissent at 68. We discuss the typical storage capacity of electronic devices simply to highlight the features that generally distinguish them from traditional baggage. Indeed, we do not and need not determine whether Cotterman's laptop possessed unusually large or simply "average" capacity in order to resolve that the forensic examination of it required reasonable suspicion.

private and this expectation is "one that society is prepared to recognize as 'reasonable.'" *Katz v. United States*, 389 U.S. 347, 361 (1967) (Harlan, J., concurring).[11]

Electronic devices often retain sensitive and confidential information far beyond the perceived point of erasure, notably in the form of browsing histories and records of deleted files. This quality makes it impractical, if not impossible, for individuals to make meaningful decisions regarding what digital content to expose to the scrutiny that accompanies international travel. A person's digital life ought not be hijacked simply by crossing a border. When packing traditional luggage, one is accustomed to deciding what papers to take and what to leave behind. When carrying a laptop, tablet or other device, however, removing files unnecessary to an impending trip is an impractical solution given the volume and often intermingled nature of the files. It is also a time-consuming task that may not even effectively erase the files.

The present case illustrates this unique aspect of electronic data. Agents found incriminating files in the unallocated space of Cotterman's laptop, the space where the computer stores files that the user ostensibly deleted and maintains other "deleted" files retrieved from web sites the user has visited. Notwithstanding the attempted erasure of material or the transient nature of a visit to a web site,

[11] The dissent's discussion about Facebook and other platforms where the user voluntarily transmits personal data over the Internet, often oblivious to privacy issues, Dissent at 65–66, is a red herring. Of course, willful disclosure of electronic data, like disclosure of other material, undercuts an individual's expectation of privacy. But there was no such disclosure here. Nor does the border search implicate such an affirmative disclosure.

computer forensic examination was able to restore the files. It is as if a search of a person's suitcase could reveal not only what the bag contained on the current trip, but everything it had ever carried.

With the ubiquity of cloud computing, the government's reach into private data becomes even more problematic.[12] In the "cloud," a user's data, including the same kind of highly sensitive data one would have in "papers" at home, is held on remote servers rather than on the device itself. The digital device is a conduit to retrieving information from the cloud, akin to the key to a safe deposit box. Notably, although the virtual "safe deposit box" does not itself cross the border, it may appear as a seamless part of the digital device when presented at the border. With access to the cloud through forensic examination, a traveler's cache is just a click away from the government.

As Justice Scalia wrote, "It would be foolish to contend that the degree of privacy secured to citizens by the Fourth Amendment has been entirely unaffected by the advance of technology." *Kyllo*, 533 U.S. at 33–34. Technology has the dual and conflicting capability to decrease privacy and augment the expectation of privacy. While the thermal imaging device in *Kyllo* threatened to expose the hour at

[12] "The term 'cloud computing' is based on the industry usage of a cloud as a metaphor for the ethereal internet. . . . An external cloud platform is storage or software access that is essentially rented from (or outsourced to) a remote public cloud service provider, such as Amazon or Google. . . . By contrast, an internal or private cloud is a cluster of servers that is networked behind an individual or company's own firewall." David A. Couillard, *Defogging the Cloud: Applying Fourth Amendment Principles to Evolving Privacy Expectations in Cloud Computing*, 93 Minn. L. Rev. 2205, 2216 (2009) (internal citations omitted).

which "the lady of the house" took her daily "sauna and bath," *id.* at 38, digital devices allow us to carry the very papers we once stored at home.

The point is technology matters. The Department of Homeland Security has acknowledged as much in the context of international travelers:

> Where someone may not feel that the inspection of a briefcase would raise significant privacy concerns because the volume of information to be searched is not great, that same person may feel that a search of their laptop increases the possibility of privacy risks due to the vast amount of information potentially available on electronic devices.

DHS, Privacy Impact Assessment for the Border Searches of Electronic Devices 2 (Aug. 25, 2009), *available at* http://www.dhs.gov/xlibrary/assets/privacy/privacy_pia_cbp_laptop.pdf.

This is not to say that simply because electronic devices house sensitive, private information they are off limits at the border. The relevant inquiry, as always, is one of reasonableness. But that reasonableness determination must account for differences in property. *See Samson v. California*, 547 U.S. 843, 848 (2006) ("Under our general Fourth Amendment approach, *we examine the totality of the circumstances* to determine whether a search is reasonable . . . .") (internal quotation marks, citation, and alterations omitted) (emphasis added). Unlike searches involving a reassembled gas tank, *Flores-Montano*, 541 U.S. at 150, or

small hole in the bed of a pickup truck, *United States v. Chaudhry*, 424 F.3d 1051, 1054 (9th Cir. 2005), which have minimal or no impact beyond the search itself—and little implication for an individual's dignity and privacy interests—the exposure of confidential and personal information has permanence. It cannot be undone. Accordingly, the uniquely sensitive nature of data on electronic devices carries with it a significant expectation of privacy and thus renders an exhaustive exploratory search more intrusive than with other forms of property.

After their initial search at the border, customs agents made copies of the hard drives and performed forensic evaluations of the computers that took days to turn up contraband. It was essentially a computer strip search. An exhaustive forensic search of a copied laptop hard drive intrudes upon privacy and dignity interests to a far greater degree than a cursory search at the border. It is little comfort to assume that the government—for now—does not have the time or resources to seize and search the millions of devices that accompany the millions of travelers who cross our borders. It is the potential unfettered dragnet effect that is troublesome.

We recognize the important security concerns that prevail at the border. The government's authority to protect the nation from contraband is well established and may be "heightened" by "national cris[e]s," such as the smuggling of illicit narcotics, *Montoya de Hernandez*, 473 U.S. at 538, the current threat of international terrorism and future threats yet to take shape. But even in the face of heightened concerns, we must account for the Fourth Amendments rights of travelers. *Id.* at 539.

The effort to interdict child pornography is also a legitimate one. But legitimate concerns about child pornography do not justify unfettered crime-fighting searches or an unregulated assault on citizens' private information. Reasonable suspicion is a modest, workable standard that is already applied in the extended border search, *Terry* stop,[13] and other contexts. Its application to the forensic examination here will not impede law enforcement's ability to monitor and secure our borders or to conduct appropriate searches of electronic devices.

Nor does applying this standard impede the deterrent effect of suspicionless searches, which the dissent contends is critical to thwarting savvy terrorists and other criminals. Dissent at 63. The Supreme Court has never endorsed the proposition that the goal of deterring illegal contraband at the border suffices to justify any manner of intrusive search. Rather, reasonableness remains the touchstone and the Court has expressed support for the deterrence value of suspicionless searches of a routine nature, such as vehicle checkpoints near the border. *See United States v. Martinez-Fuerte*, 428 U.S. 543, 556 (1976) ("We note here *only the substantiality of the public interest in the practice of routine stops* for inquiry at permanent checkpoints, a practice which the Government identifies as the most important of the traffic-checking operations.") (emphasis added). In practical terms, suspicionless searches of the type approved in *Arnold* will continue; border officials will conduct further, forensic examinations where their suspicions are aroused by what they find or by other factors. Reasonable suspicion leaves ample room for agents to draw on their expertise and experience to pick up on subtle cues that criminal activity may be afoot.

[13] *Terry v. Ohio*, 392 U.S. 1, 30 (1983).

See United States v. Tiong, 224 F.3d 1136, 1140 (9th Cir. 2000).[14]

We have confidence in the ability of law enforcement to distinguish a review of computer files from a forensic examination. We do not share the alarm expressed by the concurrence and the dissent that the standard we announce will prove unmanageable or give border agents a "Sophie's choice" between thorough searches and *Bivens* actions. Concurrence at 48–49; Dissent at 65. Determining whether reasonable suspicion is required does not necessitate a "complex legal determination[]" to be made on a "moment-by-moment basis." Dissent at 61. Rather, it requires that officers make a commonsense differentiation between a manual review of files on an electronic device and application of computer software to analyze a hard drive, and utilize the latter only when they possess a "particularized and objective

[14] The greatest obstacle to ferreting out contraband at the border has always been the sheer number of international travelers. Any contention that national security will be critically hampered by stripping border agents of a critical law enforcement tool—suspicionless forensic examinations of electronics—is undermined by the fact that, as a matter of commonsense and resources, it is only when reasonable suspicion is aroused that such searches typically take place. *See, e.g.*, *Chaudhry*, 424 F.3d at 1054 (B. Fletcher, J., concurring) ("As a practical matter, border agents are too busy to do extensive searches (removing gas tanks and door panels, boring holes in truck beds) unless they have suspicion."). As Judge Callahan acknowledges in her separate opinion, the record suggests that "remote and/or intensive searches of electronic devices crossing the border do not occur all that often." Concurrence at 50 n.11. The reference that only a small fraction of travelers at the border have their devices searched simply reinforces our point—our ruling will not place an undue burden on border agents who already rely on a degree of suspicion in referring travelers to secondary inspection.

basis for suspecting the person stopped of criminal activity." *Tiong*, 224 F.3d at 1140 (internal quotation marks omitted).

International travelers certainly expect that their property will be searched at the border. What they do not expect is that, absent some particularized suspicion, agents will mine every last piece of data on their devices or deprive them of their most personal property for days (or perhaps weeks or even months, depending on how long the search takes). *United States v. Ramos-Saenz*, 36 F.3d 59, 61 n.3 (9th Cir. 1994) ("Intrusiveness includes both the extent of a search as well as the degree of indignity that may accompany a search."). Such a thorough and detailed search of the most intimate details of one's life is a substantial intrusion upon personal privacy and dignity. We therefore hold that the forensic examination of Cotterman's computer required a showing of reasonable suspicion, a modest requirement in light of the Fourth Amendment.

IV. REASONABLE SUSPICION

Reasonable suspicion is defined as "a particularized and objective basis for suspecting the particular person stopped of criminal activity." *United States v. Cortez*, 449 U.S. 411, 417–18 (1981). This assessment is to be made in light of "the totality of the circumstances." *Id.* at 417. "[E]ven when factors considered in isolation from each other are susceptible to an innocent explanation, they may collectively amount to a reasonable suspicion." *United States v. Berber-Tinoco*, 510 F.3d 1083, 1087 (9th Cir. 2007). We review reasonable suspicion determinations de novo, reviewing findings of historical fact for clear error and giving "due weight to inferences drawn from those facts by resident judges and

local law enforcement officers." *Ornelas v. United States*, 517 U.S. 690, 699 (1996).

In the district court and in supplemental briefing, the government argued that the border agents had reasonable suspicion to conduct the initial search and the forensic examination of Cotterman's computer. We agree.

The objective facts reflect that both the agents at the border and the agents who arrived later from Sells based their decision to search Cotterman's belongings on the TECS hit. Officer Alvarado was told by those in charge of administering the TECS database that he should search Cotterman's property because the TECS hit indicated "that [Cotterman] appeared to [have] been involved in some type of child pornography." Agent Riley also looked up Cotterman's criminal record and understood that he had a prior conviction for child pornography. As it turned out, Cotterman's previous conviction was not for pornography, but for child molestation. Nonetheless, the agents' *understanding* of the objective facts, albeit mistaken, is the baseline for determining reasonable suspicion. *See Liberal v. Estrada*, 632 F.3d 1064, 1077 (9th Cir. 2011) ("Even if an officer makes a mistake of fact, that mistake 'will not render a stop illegal, if the objective facts known to the officer gave rise to a reasonable suspicion that criminal activity was afoot.'" (quoting *United States v. Mariscal*, 285 F.3d 1127, 1131 (9th Cir. 2002))).

By itself, Cotterman's 1992 conviction for child molestation does not support reasonable suspicion to conduct an extensive forensic search of his electronic devices. "Although a prior criminal history cannot alone establish reasonable suspicion . . . it is permissible to consider such a

fact as part of the total calculus of information in th[at] determination[].” *Burrell v. McIlroy*, 464 F.3d 853, 858 n.3 (9th Cir. 2006). The TECS alert was not based merely on Cotterman’s conviction—the agents were aware that the alert targeted Cotterman because he was a sex offender “who travel[ed] frequently out of the country” and who was “possibly involved in child sex tourism.” Further, Agent Riley testified that an examination of Cotterman’s passport confirmed that he had traveled in and out of the country frequently since his conviction in 1992.

In further support of reasonable suspicion, the government asserts that Mexico, from which the Cottermans were returning, is “a country associated with sex tourism.”[15] The ICE field office specifically informed Agent Riley that the alert was part of Operation Angel Watch, which targeted individuals potentially involved in sex tourism and alerted officials to be on the lookout for laptops, cameras and other paraphernalia of child pornography. *See* 156 Cong. Rec. S9581-03 (daily ed. Dec. 14, 2010) (describing Operation Angel Watch as a program “help[ing] ICE [to] identify travel patterns of convicted sex offenders who may attempt to exploit children in foreign countries”). Cotterman’s TECS alert, prior child-related conviction, frequent travels, crossing from a country known for sex tourism, and collection of electronic equipment, plus the parameters of the Operation

[15] It is ironic that the dissent expresses concern that, by factoring in the incidence of crime in particular countries, “thousands of individuals . . . will now be forced to reconsider traveling to entire countries . . . or will need to leave all their electronic equipment behind, to avoid arousing a ‘reasonable’ suspicion,” Dissent at 78, when, if forensic examination of those travelers’ electronics occurs at the border, the dissent would require *no suspicion at all*.

Angel Watch program, taken collectively, gave rise to reasonable suspicion of criminal activity.

To these factors, the government adds another—the existence of password-protected files on Cotterman's computer.[16] We are reluctant to place much weight on this factor because it is commonplace for business travelers, casual computer users, students and others to password protect their files. Law enforcement "cannot rely solely on factors that would apply to many law-abiding citizens," *Berber-Tinoco*, 510 F.3d at 1087, and password protection is ubiquitous. National standards require that users of mobile electronic devices password protect their files. *See generally* United States Department of Commerce, Computer Security Division, National Institute of Standards and Technology, Computer Security (2007) (NIST Special Publication 800-111). Computer users are routinely advised—and in some cases, required by employers—to protect their files when traveling overseas. *See, e.g.*, Michael Price, *National Security Watch*, 34-MAR Champion 51, 52 (March 2010) ("[T]here is one relatively simple thing attorneys can do [when crossing the border] to protect their privacy and the rights of their clients: password-protect the computer login and any sensitive files or folders.").

Although password protection of files, in isolation, will not give rise to reasonable suspicion, where, as here, there are other indicia of criminal activity, password protection of files

[16] Agent Riley testified that Alvarado told her that he had "encounter[ed] some *files* that were password protected," while Agent Alvarado testified that he found one file.

may be considered in the totality of the circumstances.[17] To contribute to reasonable suspicion, encryption or password protection of files must have some relationship to the suspected criminal activity. Here, making illegal files difficult to access makes perfect sense for a suspected holder of child pornography. When combined with the other circumstances, the fact that Officer Alvarado encountered at least one password protected file on Cotterman's computer contributed to the basis for reasonable suspicion to conduct a forensic examination.

The existence of the password-protected files is also relevant to assessing the reasonableness of the scope and duration of the search of Cotterman's computer. The search was necessarily protracted because of the password protection that Cotterman employed. After Cotterman failed to provide agents with the passwords to the protected files and fled the country, it took Agent Owen days to override the computer security and open the image files of child pornography.

Although we must take into account factors weighing both in favor and against reasonable suspicion, Cotterman's innocent explanation does not tip the balance. *See Tiong*, 224 F.3d at 1140 (recognizing that "innocent possibilities . . . do not undermine reasonable suspicion"). The dissent suggests that Cotterman's offer at the border "to help the agents access his computer" counsels against a finding of reasonable suspicion. Dissent at 80. The agents were

[17] We do not suggest that password protecting an entire device—as opposed to files within a device—can be a factor supporting a reasonable suspicion determination. Using a password on a device is a basic means of ensuring that the device cannot be accessed by another in the event it is lost or stolen.

appropriately wary of such an offer due to concerns that Cotterman could tamper with the devices. Nor did the agents' discovery of vacation photos eliminate the suspicion that Cotterman had engaged in criminal activity while abroad or might be importing child pornography into the country. Because the first examination of Cotterman's laptop, by Officer Alvarado, turned up nothing incriminating, Cotterman urges that any suspicion prompted by the TECS alert was dispelled by this initial failure. But the nature of the alert on Cotterman, directing agents to review media and electronic equipment for child pornography, justified conducting the forensic examination despite the failure of the first search to yield any contraband.

Collectors of child pornography can hardly be expected to clearly label such files and leave them in readily visible and accessible sections of a computer's hard drive, particularly when they are traveling through border crossings, where individuals ordinarily anticipate confronting at least a cursory inspection. Officer Alvarado, who was responsible for conducting the initial search, was specifically looking for photographs as described in the TECS hit but testified that he had only a slightly above-average familiarity with laptops. He could do no more than open a file, look at it and see if he could access it. He testified that "[i]f [he] encountered something that [he] could not access, then [he] would reference it to somebody that may have that ability to look at [it]." That is precisely what occurred here. Officer Alvarado came across password-protected files but, unable to open them, moved on to other files. Alvarado told Agent Riley about the password protection, and she and Agent Brisbine decided to seize the computers for further examination. The border agents "certainly had more than an inchoate and unparticularized suspicion or hunch" of criminal activity to

support their decision to more carefully search for evidence of child pornography. *Montoya de Hernandez*, 473 U.S. at 542 (internal quotation marks and citation omitted). An alert regarding possession of this type of criminal contraband justified obtaining additional resources, here available in Tucson, to properly determine whether illegal files were present.

Unlike the dissent, we credit the agents' observations and experience in acting upon significant myriad factors that support reasonable suspicion. It is not our province to nitpick the factors in isolation but instead to view them in the totality of the circumstances. For the above reasons, we conclude that the examination of Cotterman's electronic devices was supported by reasonable suspicion and that the scope and manner of the search were reasonable under the Fourth Amendment. Cotterman's motion to suppress therefore was erroneously granted.

REVERSED.

CALLAHAN, Circuit Judge, concurring in part, dissenting in part, and concurring in the judgment, with whom CLIFTON, Circuit Judge, joins, and with whom M. SMITH, Circuit Judge, joins as to all but Part II.A:

Whether it is drugs, bombs, or child pornography, we charge our government with finding and excluding any and all illegal and unwanted articles and people before they cross our international borders. Accomplishing that Herculean task requires that the government be mostly free from the Fourth Amendment's usual restraints on searches of people and their

property. Today the majority ignores that reality by erecting a new rule requiring reasonable suspicion for any thorough search of electronic devices entering the United States. This rule flouts more than a century of Supreme Court precedent, is unworkable and unnecessary, and will severely hamstring the government's ability to protect our borders.

I therefore dissent from Part III of the majority's opinion. I concur in Parts I, II, and IV, and in particular the majority's conclusion in Part IV that the government had reasonable suspicion to conduct the forensic examination of Howard Cotterman's electronic devices. I therefore also concur in the judgment.

I.

Over the last 125 years, the Supreme Court has explained that the United States and its people have a "paramount interest" in national self-protection and an "inherent" right to exclude illegal and "unwanted persons and effects." *United States v. Flores-Montano*, 541 U.S. 149, 152–53 (2004); *see also United States v. Montoya de Hernandez*, 473 U.S. 531, 537–40 (1985); *United States v. Ramsey*, 431 U.S. 606, 616–18 (1977); *United States v. Thirty-Seven (37) Photographs*, 402 U.S. 363, 376 (1971); *Carroll v. United States*, 267 U.S. 132, 154 (1925); *Boyd v. United States*, 116 U.S. 616, 623 (1886). Accordingly, "[t]he Government's interest in preventing the entry of unwanted persons and effects is at its zenith at the international border." *Flores-Montano*, 541 U.S. at 152.

To effectuate this interest, the Supreme Court has recognized a broad exception to the Fourth Amendment's requirement of probable cause or a warrant for searches

conducted at the border. Under that exception, searches of people and their property at the United States borders and their functional equivalents are *per se* reasonable, meaning that they typically do not require a warrant, probable cause, or even reasonable suspicion. *Montoya de Hernandez*, 473 U.S. at 538; *see also Flores-Montano*, 541 U.S. at 152–53; *Ramsey*, 431 U.S. at 616–18; *United States v. Seljan*, 547 F.3d 993, 999–1000 (9th Cir. 2008) (en banc), *cert. denied*, 129 S. Ct. 1368 (2009).

In the long time that the Court has recognized the border search doctrine, the Court has found just *one* search at the border that required reasonable suspicion. *See Montoya de Hernandez*, 473 U.S. at 541 (upholding the 24-hour detention of a woman suspected of smuggling illegal drugs in her digestive system, followed by a pregnancy test and rectal examination, based on reasonable suspicion). In the remaining cases, the Court consistently has described the government's border search authority in very broad terms[1]

[1] *See, e.g.*, *Flores-Montano*, 541 U.S. at 152 ("The Government's interest in preventing the entry of unwanted persons and effects is at its zenith at the international border."); *id.* at 153 ("It is axiomatic that the United States, as sovereign, has the inherent authority to protect, and a paramount interest in protecting, its territorial integrity."); *Ramsey*, 431 U.S. at 617 ("This interpretation, that border searches were not subject to the warrant provisions of the Fourth Amendment and were 'reasonable' within the meaning of that Amendment, has been faithfully adhered to by this Court."); *id.* at 620 ("The border-search exception is grounded in the recognized right of the sovereign to control, subject to substantive limitations imposed by the Constitution, who and what may enter the country."); *Thirty-Seven (37) Photographs*, 402 U.S. at 376 ("[A traveler's] right to be let alone neither prevents the search of his luggage nor the seizure of unprotected, but illegal, materials when his possession of them is discovered during such a search. Customs officers characteristically inspect luggage and their power to do so is not

and overturned the lower courts' attempts to cabin that authority.[2] The Court also repeatedly has gone out of its way to explain that border searches generally are exempt from the limits it imposes on domestic searches. *See, e.g.*, *Flores-Montano*, 541 U.S. at 154 ("[O]n many occasions, we have noted that the expectation of privacy is less at the border than it is in the interior."); *Montoya de Hernandez*, 473 U.S. at 539–40 ("But not only is the expectation of privacy less at the border than in the interior, the Fourth Amendment balance between the interests of the Government and the privacy right of the individual is also struck much more favorably to the Government at the border." (internal and external citations omitted)); *United States v. 12 200-Foot Reels of Super 8mm. Film*, 413 U.S. 123, 125 (1973) ("Import restrictions and searches of persons or packages at the national borders rest on

questioned in this case; it is an old practice and is intimately associated with excluding illegal articles from the country."); *Carroll*, 267 U.S. at 154 ("Travelers may be so stopped in crossing an international boundary because of national self-protection reasonably requiring one entering the country to identify himself as entitled to come in, and his belongings as effects which may be lawfully brought in."). Even in *Montoya de Hernandez* the Court described the government's border search authority expansively. *See* 473 U.S. at 539–40, 542–44.

[2] *See, e.g.*, *Flores-Montano*, 541 U.S. at 152–55 (overturning the Ninth Circuit's conclusion that the border search of a gas tank required reasonable suspicion); *Ramsey*, 431 U.S. at 616–22 (overturning the D.C. Circuit's conclusion that the search of international mail required probable cause); *Thirty-Seven (37) Photographs*, 402 U.S. at 376 (relying in part on border search doctrine to overturn lower court's decision that statute barring the importation of obscene material was unconstitutional).

different considerations and different rules of constitutional law from domestic regulations.").[3]

II.

It is against this legal backdrop that we must assess the constitutionality of the government's search in this case. As with all searches subject to Fourth Amendment review, the constitutionality of a border search turns on whether it is reasonable. *See Brigham City, Utah v. Stuart*, 547 U.S. 398, 403 (2006) ("[T]he ultimate touchstone of the Fourth Amendment is 'reasonableness.'"). Under the border search doctrine, suspicionless border searches are *per se* reasonable. However, the Supreme Court has identified three situations in which they might not be *per se* reasonable, *i.e.*, at least reasonable suspicion is required: (1) "highly intrusive searches of the person;" (2) destructive searches of property;

[3] *See also City of Indianapolis v. Edmond*, 531 U.S. 32, 47–48 (2000) (explaining that decision barring domestic drug interdiction checkpoints "does not affect the validity of border searches or searches at places like airports"); *United States v. Ross*, 456 U.S. 798, 823 (1982) (explaining that while the Fourth Amendment gives protection to containers in domestic vehicles, "[t]he luggage carried by a traveler entering the country may be searched at random by a customs officer"); *Torres v. Puerto Rico*, 442 U.S. 465, 472–74 (1979) (distinguishing between United States–Puerto Rico border and international borders in holding unconstitutional the search of a traveler's luggage without "articulable suspicion"); *United States v. Brignoni-Ponce*, 422 U.S. 873, 884 (1975) ("Except at the border and its functional equivalents, officers on roving patrol may stop vehicles" only with reasonable suspicion they contain illegal aliens); *Almeida-Sanchez v. United States*, 413 U.S. 266, 272–76 (1973) (distinguishing searches of vehicles at the border from a search that occurred 25 miles away); *Carroll*, 267 U.S. at 151–54 (distinguishing between interior and border searches of vehicles and persons).

and (3) searches conducted in a "particularly offensive" manner. *Flores-Montano*, 541 U.S. at 152–56 & n.2.

Although its opinion is not entirely clear, the majority appears to rely on the first and third exceptions to hold that the search at issue in this case required reasonable suspicion. (There is no claim that the government damaged or destroyed Cotterman's property.) But the exception for "highly intrusive searches of the person," *Flores-Montano*, 541 U.S. at 152, cannot apply here; "papers," even private ones in electronic format, are not a "person." *See id.* ("The reasons that might support a requirement of some level of suspicion in the case of highly intrusive searches of the person—dignity and privacy interests of the person being searched—simply do not carry over to vehicles."). That leaves the exception for searches conducted in a "particularly offensive" manner. *Id.* at 154 n.2. The majority relies primarily on the notion that electronic devices are special to conclude that reasonable suspicion was required. Majority at 20–28. The majority is mistaken.

A.

The majority correctly concludes that the government's forensic search in Tucson was not an extended border search, as the border agents retained custody of Cotterman's laptop.[4]

[4] I agree with the majority that this case does not involve an extended border search. Unlike a border search, an extended border search takes place at a location "away from the border where entry is not apparent, but where the dual requirements of reasonable certainty of a recent border crossing and reasonable suspicion of criminal activity are satisfied." *United States v. Guzman-Padilla*, 573 F.3d 865, 878–79 (9th Cir. 2009) (internal quotation marks and citation omitted), *cert. denied*, 131 S. Ct. 67 (2010). Reasonable suspicion is required precisely because the individual

Id. at 9, 14–15. The majority also states that "[i]t is the comprehensive and intrusive nature of a forensic examination—not the location of the examination—that is the key factor triggering the requirement of reasonable suspicion here." Majority at 17. The inclusion of the word "key" might be read to imply that some other factor, such as the location and duration of the search, contributed to its purported unreasonableness. I write to refute any such notion.

First consider the facts. The border agents took Cotterman's electronic devices to the nearest computing center (to Tucson, where Cotterman and his wife were already traveling), before clearing them for entry into the United States. The computer specialist moved the search ahead of his other work and conducted it over the weekend. Although the forensic search lasted five days, it took only 48 hours to discover the initial 75 images of child pornography. The agents were reasonably reluctant to rely on Cotterman's offer to help, since he might have deleted or otherwise made unrecoverable any contraband that his devices contained. The agents returned the devices as soon as they cleared them.

---

has regained an expectation of privacy by moving away from the border. *See United States v. Villasenor*, 608 F.3d 467, 471–72 (9th Cir.), *cert. denied*, 131 S. Ct. 547 (2010); *United States v. Whiting*, 781 F.2d 692, 695 (9th Cir. 1986). Here, there was no attenuation between Cotterman's border crossing and the forensic search of his electronic property; the government conducted that search *before* clearing the property for entry and *before* Cotterman could regain an expectation of privacy in that property. *See* 19 U.S.C. § 1499 (providing that imported goods are permitted entry only after Customs clears them); *United States v. Alfonso*, 759 F.2d 728, 734 (9th Cir. 1985) ("Extended border searches occur after the actual entry has been effected and intrude more on an individual's normal expectation of privacy.").

Now consider the law. The Supreme Court has upheld the constitutionality of a police search of packages retrieved from an automobile, even though the police conducted their search three days after the police stopped the vehicle and at the police station. *United States v. Johns*, 469 U.S. 478, 485–88 (1985). The Court rejected the argument that "searches of containers discovered in the course of a vehicle search are subject to temporal restrictions not applicable to the vehicle search itself." *Id.* at 485. Although *Johns* involved a domestic automobile search based on probable cause, it still stands for the proposition, equally applicable to this case, that "the legality of the search was determined by reference to the [applicable] exception to the warrant requirement." *Id.*

In the border search context, the Supreme Court, in upholding the lengthy detention of a person reasonably suspected of smuggling drugs in her digestive system at an airport, addressed whether that detention was "reasonably related in scope to the circumstances which justified it initially." *Montoya de Hernandez*, 473 U.S. at 542. The Court explained that: (1) "courts should not indulge in unrealistic second-guessing" when answering this question, as "[a]uthorities must be allowed to graduate their response to the demands of any particular situation;" (2) the Court consistently has "refused to charge police with delays in investigatory detention attributable to the suspect's evasive actions;" and (3) "we have also consistently rejected hard-and-fast time limits." *Id*. at 542–43 (quotation marks and citations omitted). The Court emphasized that, at the international border, "the Fourth Amendment balance of interests leans heavily to the Government" because the government is charged not just with investigating crime but with "protecting this Nation from entrants who may bring anything harmful into this country." *Id.* at 544. Finally, any

"length" or "discomfort" associated with a border search does not offend the Fourth Amendment when it "result[s] solely from the method by which [a traveler] cho[oses] to smuggle [contraband] into this country." *Id.*

Any suggestion that the government's search here was "particularly offensive" due to the location and duration of the search runs counter to the Supreme Court's admonitions in *Johns* and *Montoya de Hernandez*. It also effectively requires the government to supply every port of entry with the equipment and staff needed to conduct forensic electronic searches, or at least to have such equipment and staff waiting at a nearby location. Such a requirement is unreasonable, particularly since the record in this case suggests that a forensic search of Cotterman's electronic devices at the border station would have taken *longer* than the search at the Tucson computing center.[5] *See United States v. Hill*, 459 F.3d 966, 974–75 (9th Cir. 2006), *cert. denied*, 127 S. Ct. 1863 (2007) (discussing problems inherent in requiring police to bring with them equipment to search electronic media); *cf. Johns*, 469 U.S. at 486–87 (explaining that requiring police

[5] The district court found that the government could have conducted the forensic search at the Lukeville border station. *United States v. Cotterman*, No. CR 07-1207-TUC-RCC, 2009 WL 465028, at *1 (D. Ariz. Feb. 24, 2009). The court presumably based this finding on testimony that the computer specialist who conducted the forensic examination had a specially-equipped laptop. However, the specialist testified that using his laptop at the border station, rather than transporting Cotterman's electronic devices to the Tucson computer center, would have taken "a lot longer" because the laptop was "not nearly as extensive as what I have in my lab," the "processor in my laptop is much slower" than the lab equipment, and "I could only do one computer at a time with the laptop." Technical difficulties also could have slowed down an examination conducted at the border station.

officers to immediately inspect all packages "would be of little benefit to the person whose property is searched").

B.

The majority's opinion turns primarily on the notion that electronic devices deserve special consideration because they are ubiquitous and can store vast quantities of personal information. That idea is fallacious and has no place in the border search context.

The Supreme Court has been willing to distinguish only between border searches of people and property, not between different types of property. In 2004, in *Flores-Montano*, the Court explained that

> the reasons that might support a requirement of some level of suspicion in the case of highly intrusive searches of the person—dignity and privacy interests of the person being searched—simply do not carry over to vehicles. Complex balancing tests to determine what is a "routine" search of a vehicle, as opposed to a more "intrusive" search of a person, have no place in border searches of vehicles.

541 U.S. at 152. We have since applied *Flores-Montano* to hold that any distinction between "routine" and "nonroutine" searches does not apply to searches of property, and that there can be no "least restrictive means" test for border searches. *United States v. Chaudhry*, 424 F.3d 1051, 1054 (9th Cir. 2005), *cert. denied*, 547 U.S. 1083 (2006); *United States v. Cortez-Rocha*, 394 F.3d 1115, 1122–23 (9th Cir. 2004), *cert.*

denied, 546 U.S. 849 (2005).[6] Put another way, the Supreme Court—and, reluctantly, this court—have refused to adopt a sliding "intrusiveness" scale for border searches of property. Thus, the Court has all but held that property that crosses the border, whatever it is, does not merit Fourth Amendment protection.

Of course, *Flores-Montano*, *Chaudhry*, and *Cortez-Rocha* involved vehicles or parts of vehicles, not electronic devices, and the other border search cases that have reached the Supreme Court all involved containers of some sort. *See, e.g.*, *Ramsey*, 431 U.S. at 616–22 (mail); *Thirty-Seven (37) Photographs*, 402 U.S. at 376 (luggage). And yes, the Court has left open the possibility that a border search might be "'"'unreasonable' because of the particularly offensive manner in which it is carried out.'" *Flores-Montano*, 541 U.S. at 154 n.2 (quoting *Ramsey*, 431 U.S. at 618 n.13). But is the mere fact that Cotterman chose to save his child pornography electronically, rather than print it out on paper, enough to invoke that exception?

The two courts of appeals—including this court—that have had occasion to address whether electronic devices

[6] In 1985, the Supreme Court wrote about the government's "plenary authority to conduct *routine* searches and seizures at the border." *Montoya de Hernandez*, 473 U.S. at 537 (emphasis added); *see also id.* at 541 n.4 ("Because the issues are not presented today we suggest no view on what level of suspicion, if any, is required for *nonroutine* border searches such as strip, body-cavity, or involuntary x-ray searches.") (emphasis added). We unfortunately seized on the word "routine" to establish a sliding scale of intrusiveness, with more intrusive (*i.e.*, less "routine") searches requiring reasonable suspicion. *See, e.g.*, *United States v. Molina-Tarazon*, 279 F.3d 709, 711–13 (9th Cir. 2002). *Flores-Montano* plainly repudiated that approach.

deserve special consideration have correctly concluded that they do not. In *United States v. Arnold*, 533 F.3d 1003, 1008–10 (9th Cir. 2008), *cert. denied*, 555 U.S. 1176 (2009), we held that laptops are like other property, relying on the reasoning and language in *Flores-Montano*, *Chaudhry*, and *Cortez-Rocha* discussed above (among other cases). Similarly, in *United States v. Ickes*, 393 F.3d 501, 503–07 (4th Cir. 2005), the Fourth Circuit upheld an extensive border search of the defendant's laptop that revealed child pornography. Notably, the court held that the border agents had reasonable suspicion to search the defendant's laptop, but explained why that did not matter:

> The agents did not inspect the contents of Ickes's computer until they had already discovered marijuana paraphernalia, photo albums of child pornography, a disturbing video focused on a young ball boy, and an outstanding warrant for Ickes's arrest. As a practical matter, computer searches are most likely to occur where—as here—the traveler's conduct or the presence of other items in his possession suggest the need to search further. However, to state the probability that reasonable suspicions will give rise to more intrusive searches is a far cry from enthroning this notion as a matter of constitutional law. The essence of border search doctrine is a reliance upon the trained observations and judgments of customs officials, rather than upon constitutional requirements applied to the inapposite context of this sort of search.

Id. at 507. Thus, the Fourth Circuit has recognized what the majority does not: electronic devices are like any other container that the Supreme Court has held may be searched at the border without reasonable suspicion.[7] Though we are not bound by *Arnold* nor *Ickes* in this en banc proceeding, we *are* bound by what the Supreme Court has said: in the unique context of border searches, property is property and we may not chip away at the government's authority to search it by adopting a sliding scale of intrusiveness. It's the border, not the technology, that "matters." Majority at 24; *cf. Ramsey*, 431 U.S. at 620 ("It is clear that there is nothing in the rationale behind the border-search exception which suggests that the mode of entry will be critical.").

Logic and commonsense, not just Supreme Court precedent, reveal the flaws in the majority's opinion. The fact that electronic devices are capable of storing a lot of personal information does not make an extensive search of them "particularly offensive." We have squarely rejected the idea that the "intrusiveness" of a search depends in whole or in part on the nature of the property being searched. In *United States v. Giberson*, 527 F.3d 882 (9th Cir. 2008), we specifically rebuffed the argument that computers are special for Fourth Amendment purposes by virtue of how much information they store; "neither the quantity of information, nor the form in which it is stored, is legally relevant in the Fourth Amendment context." *Id.* at 888; *see also California v. Carney*, 471 U.S. 386, 393–94 (1985) (rejecting applying

[7] I agree with Judge Smith that the majority's opinion appears to create an imprudent split with the Fourth Circuit. *See* Dissent at 58.

Fourth Amendment protection to property (a mobile home) that is "capable of functioning as a home" simply on account of the property's size or "worth[iness]" as a container); *United States v. Payton*, 573 F.3d 859, 864 (9th Cir. 2009) ("*Giberson* held that computers were not entitled to a special categorical protection of the Fourth Amendment."); *Kyllo v. United States*, 533 U.S. 27, 41 (2001) (Stevens, J., dissenting) (explaining that Fourth Amendment exceptions and distinctions based solely on a type of technology are "unwise[ ] and inconsistent with the Fourth Amendment").

While *Giberson* and *Carney* involved domestic searches, their reasoning applies equally in the border search context. If the government may search the contents of a briefcase, car, or mobile home that transits the border, there is no reason it should not also be able to search the contents of a camera, tablet, or laptop that enters the country. All of those things are capable of storing, and often do store, private information. *See Ross*, 456 U.S. at 823 ("The luggage carried by a traveler entering the country may be searched at random by a customs officer; *the luggage may be searched no matter how great the traveler's desire to conceal the contents may be.*" (emphasis added)). The majority points out that electronic devices can and usually do store much *more* private information than their non-electronic counterparts. Majority at 17–24. But "a port of entry is not a traveler's home," *Thirty-Seven (37) Photographs*, 402 U.S. at 376, even if a traveler chooses to carry a home's worth of personal information across it.[8]

[8] The element of choice is crucial. The fact that border searches occur at fixed times and checkpoints makes them inherently less intrusive; a person "with advance notice of the location of a permanent checkpoint has an opportunity to avoid the search entirely, or at least to prepare for, and limit, the intrusion on her privacy." *Mich. Dep't of State Police v. Sitz*,

Moreover, a bright-line rule distinguishing electronic from non-electronic devices—of the sort the Supreme Court has made clear has no place in Fourth Amendment jurisprudence, *Ohio v. Robinette*, 519 U.S. 33, 39 (1996)—is arbitrary; there is no reason someone carrying a laptop should receive greater privacy protection than someone who chooses (or can only afford) to convey his or her personal information on paper.

In short, today the court erects a new bright-line rule: "forensic examination" of electronic devices "at the border requires reasonable suspicion." Majority at 17; *see also id.* at 21 n.10. The majority never defines "forensic," leaving border agents to wonder exactly what types of searches are

---

496 U.S. 444, 463 (1990) (Stevens, J., dissenting); *see also Montoya de Hernandez*, 473 U.S. at 544 ("Respondent's detention was long, uncomfortable, indeed, humiliating; but both its length and its discomfort resulted solely from the method by which she chose to smuggle illicit drugs into this country.").

The element of choice goes to the more fundamental issue of whether someone can have any reasonable expectation of privacy when he or she voluntarily carries electronic equipment across the border. Border officers are permitted to examine a written diary, and someone who wants to keep the contents of a diary secret should know not to take it across the border. The same should be true for personal data stored on a laptop or other electronic device rather than a written diary.

Moreover, the fact that the Fourth Amendment does not apply in foreign countries further weakens any claim to a reasonable expectation of privacy in property that crosses the United States border. Carrying an electronic device outside the United States almost always entails carrying it into another country, making it subject to search under that country's laws. Travelers expect these intrusions, or at least their possibility.

off-limits.[9] Even if the majority means to require reasonable suspicion for *any* type of digital forensic border search, no court has ever erected so categorical a rule, based on so general a type of search or category of property, and the Supreme Court has rightly slapped down anything remotely similar. The majority invites—indeed, requires—the Court to do so again.[10]

III.

The majority's holding contravenes Supreme Court precedent, defies logic and commonsense, and is unworkable. It is also unnecessary and will impair the federal government's ability to protect our borders.

As Judge Smith points out in his dissent, "[b]order patrol agents process hundreds of thousands of travelers each day and conduct thousands of searches on electronic devices each year." Dissent at 61–62 (citation omitted). All the evidence in this case suggests that the government does not have the resources—time, personnel, facilities, or technology—to exhaustively search every (or even a majority) of the electronic devices that cross our borders. *Cf. Ickes*, 393 F.3d at 507. Unless we somehow manage to solve our fiscal problems, and unless the government somehow manages to

---

[9] *See* Darrin J. Behr, *Anti-Forensics: What it Does and Why You Need to Know*, 255 N.J. Law. 9, 10 (Dec. 2008) ("Due to the fact that there are hundreds of digital forensic investigation procedures developed all over the world, digital forensics has yet to be defined.").

[10] I note that a case currently pending in the Sixth Circuit appears to raise similar issues as this case. *See United States v. Stewart*, No. 12-1427 (6th Cir. filed Apr. 5, 2012); *see also United States v. Stewart*, 715 F. Supp. 2d 750 (E.D. Mich. 2010).

acquire better technology at a faster pace than the rest of us, these restraints will continue. That means border agents must prioritize who, what, and how they search. By and large, border agents will conduct forensic electronic searches of people who, like Howard Cotterman, the agents reasonably suspect may be trying to carry illegal articles into, or themselves illegally enter, the country.[11] That agents typically will have reasonable suspicion is, of course, "a far cry from enthroning this notion as a matter of constitutional law." *Ickes*, 393 F.3d at 507.

The majority finds this reality check to be of "little comfort[;] [i]t is the potential unfettered dragnet effect that is troublesome." Majority at 25. But that abstract risk, which exists with any exception to the Fourth Amendment, does not justify a bright-line rule requiring reasonable suspicion for any thorough search of electronic devices entering the United

[11] Testimony from the suppression hearing in this case suggests that remote and/or intensive searches of electronic devices crossing the border do not occur all that often. For example, the computer specialist who conducted the forensic search of Cotterman's laptop testified that the search was the first one he was asked to conduct in his 18 months on the job at the Tucson computer center. (He added that at his previous post at San Francisco International Airport, forensic searches were done right at the airport.) Similarly, one of the border agents testified that this was the first case he was aware of in which electronic devices were turned over to Immigrations and Customs Enforcement for forensic examination, and that even cursory reviews of laptops for information about illegal drug trading occurred "no more than five" times during agent's three-plus years at the Lukeville border station. *See* Michael Chertoff, Secretary of Homeland Security, *Searches Are Legal, Essential*, USA Today, July 16, 2008 ("Of the approximately 400 million travelers who entered the country last year, only a tiny percentage were referred to secondary baggage inspection for a more thorough examination. Of those, only a fraction had electronic devices that may have been checked.").

States. *See Robinette*, 519 U.S. at 39 ("[W]e have consistently eschewed bright-line rules, instead emphasizing the fact-specific nature of the reasonableness inquiry."); *see also Lyng v. Nw. Indian Cemetery Protective Ass'n*, 485 U.S. 439, 445 (1988) ("A fundamental and longstanding principle of judicial restraint requires that courts avoid reaching constitutional questions in advance of the necessity of deciding them.").

Moreover, border agents are not free to undertake "unfettered crime-fighting searches or an unregulated assault on citizens' private information." Majority at 26. As I explained in my concurrence in *Seljan*, Congress and the Executive Branch have (and have exercised) the authority to restrict when and how border agents conduct searches. *See Seljan*, 547 F.3d at 1012 (Callahan, J., concurring) (citing, e.g., 19 U.S.C. § 1583; 19 C.F.R. § 145.3(b)-(c)); *see also* Yule Kim, Cong. Research Serv. RL34404, *Border Searches of Laptop Computers and Other Electronic Storage Devices*, 13–14 (2009) (describing recent legislative proposals to limit border searches of electronic devices). In a similar vein, Justice Breyer has noted that "Customs keeps track of the border searches its agents conduct, including the reasons for the searches. This administrative process should help minimize concerns that [border] searches might be undertaken in an abusive manner." *Flores-Montano*, 541 U.S. at 156 (Breyer, J., concurring) (internal citation omitted).[12]

[12] *See also* U.S. Customs & Border Protection, Directive No. 3340-049, *Border Search of Electronic Devices Containing Information*, 3–9 (2009) (describing procedures for, and limits on, border searches of electronic devices).

Apart from being unnecessary, the majority's new limits on the government's border search authority will make it much harder for border agents to do their jobs, for at least two reasons. First, it is common knowledge that border agents at security checkpoints conduct more thorough searches not simply of those persons who arouse suspicion but also of a percentage of travelers on a random basis. Otherwise, a person who appears entirely innocent will have nothing to fear and will not be deterred from carrying something that should not be brought into the country. A checkpoint limited to searches that can be justified by articulable grounds for "reasonable suspicion" is bound to be less effective.

Second, courtesy of the majority's decision, criminals now know they can hide their child pornography or terrorist connections in the recesses of their electronic devices, while border agents, fearing Fourth Amendment or *Bivens* actions, will avoid conducting the searches that could find those illegal articles. The result will be that people and things we wish to keep out of our country will get in—a result hardly in keeping with our "inherent authority to protect, and a paramount interest in protecting," the "territorial integrity" of the United States. *Flores-Montano*, 541 U.S. at 153. The border search doctrine *must* account for the fact that border agents may need time and forensics to bypass "evasive actions" a criminal has taken to hide contraband or other illegal articles from plain view. *Montoya de Hernandez*, 473 U.S. at 542–43. I would rather leave those difficult decisions "to the discretion of the officers in the field who confront myriad circumstances we can only begin to imagine from the relative safety of our chambers." *United States v.*

Williams, 419 F.3d 1029, 1034 (9th Cir.), *cert. denied*, 546 U.S. 1081 (2005).[13]

IV.

The border search exception to the Fourth Amendment may be just that—an exception—but it is, and must be, a mighty one. The government's right and duty to protect our nation's territorial integrity demand that the government have clear authority to exclude—*and thus to find*—those people and things we have decided are offensive, threatening, or otherwise unwanted. Recognizing this, the Supreme Court has only once required reasonable suspicion for border searches in the 125 years it has been reviewing them. In the remaining cases, the Court has eschewed bright-line rules, balancing tests, and sliding intrusiveness scales, alluding to the possibility of, but never finding, a "particularly offensive"

[13] The majority insists that reasonable suspicion is a "modest, workable standard" that is applied in domestic stops of automobiles "and other contexts," and that still allows "agents to draw on their expertise and experience." Majority at 26, 27 n.14. The majority is wrong for at least three reasons. First, in making this argument, the majority reveals that it does not appreciate the crucial differences between domestic and border searches, despite those differences being spelled out in a century of case law. Those differences range from the legitimate expectation of privacy that people have in their property to the constraints government officials face in searching it. Second, a reasonable suspicion standard injects unnecessary judicial review where previously it was absent. Third, just because border agents *could* apply the reasonable suspicion standard does not mean they are, or should be, *constitutionally compelled* to do so. *See Ickes*, 393 F.3d at 507; *cf. Seljan*, 547 F.3d at 1011 (Callahan, J. concurring) (explaining that requiring border agents to apply a First Amendment exception to border searches "would require them to engage in the sort of decision-making process that the Supreme Court wished to avoid in sanctioning expansive border searches").

search. The fact that electronic devices can store large amounts of private information, or that the government can search them forensically, does not make a thorough search of such devices "particularly offensive." Rather, the Supreme Court and this court have wisely avoided making the reasonableness of a search turn on the nature of the property being searched, for the many reasons discussed above. The result has been a clear, well-understood, efficient, and effective rule that border searches are *per se* reasonable.

Regrettably the majority, dispensing with these well-settled, sensible, and *binding* principles, lifts our anchor and charts a course for muddy waters. Now border agents, instead of knowing that they may search any and all property that crosses the border for illegal articles, must ponder whether their searches are sufficiently "comprehensive and intrusive," Majority at 17, to require reasonable suspicion, and whether they have such suspicion. In most cases the answer is going to be as clear as, well, mud. We're due for another course correction.

M. SMITH, Circuit Judge, dissenting, with whom CLIFTON and CALLAHAN, Circuit Judges, join with respect to Part I:

I respectfully dissent. Until today, federal courts have consistently upheld suspicionless searches of electronic storage devices at the border. *See United States v. Arnold*, 533 F.3d 1003, 1008 (9th Cir. 2008), *cert. denied*, 555 U.S. 1176 (2009) ("[R]easonable suspicion is not needed for customs officials to search a laptop or other personal electronic storage devices at the border."); *see also United States v. Ickes*, 393 F.3d 501, 507 (4th Cir. 2005) (no finding

of reasonable suspicion required to search personal computers and disks at border); *United States v. Linarez-Delgado*, 259 Fed. Appx. 506, 508 (3d Cir. 2007); *United States v. McAuley*, 563 F. Supp. 2d 672, 677–78 (W.D. Tex. 2008); *United States v. Bunty*, 617 F. Supp. 2d 359, 365 (E.D. Pa. 2008). Yet the majority ignores these cases, rewrites long standing Fourth Amendment jurisprudence, and, in narrowing *Arnold*, creates a circuit split.

While I share some of the majority's concerns about the steady erosion of our personal privacy in this digital age, the majority's decision to create a reasonable suspicion requirement for some property searches at the border so muddies current border search doctrine that border agents will be left to divine on an ad hoc basis whether a property search is sufficiently "comprehensive and intrusive" to require reasonable suspicion, or sufficiently "unintrusive" to come within the traditional border search exception. Requiring border patrol agents to determine that reasonable suspicion exists prior to performing a basic forensic examination of a laptop or other electronic devices discourages such searches, leaving our borders open to electronically savvy terrorists and criminals who may hereafter carry their equipment and data across our borders with little fear of detection. In fact, the majority opinion makes such a legal bouillabaisse out of the previously unambiguous border search doctrine, that I sincerely hope the Supreme Court will grant certiorari, and reverse the holding in this case regarding the level of suspicion necessary to search electronic devices at the border, for the sake of our national security, and the consistency of our national border search law.

The Supreme Court rejected our last attempt to narrow the border search exception, cautioning us not to create "complex

balancing tests" for border searches of property except in the rarest of cases, where the search is "so destructive as to require" reasonable suspicion. *United States v. Flores-Montano*, 541 U.S. 149, 152, 156 (2004) (rejecting our proposed reasonable suspicion requirement in *United States v. Molina-Tarazon*, 279 F.3d 709, 713–17 (9th Cir. 2002)). "Time and again" the Court has concluded that border searches are "'reasonable simply by virtue of the fact that they occur at the border.'" *Id.* at 152–53 (quoting *United States v. Ramsey*, 431 U.S. 606, 616 (1977)).

Despite the Court's clear ruling on the issue, the majority again seeks to whittle away at the border search exception, this time by conjuring a reasonable suspicion requirement for border searches that employ computer software to search an electronic storage device. Why the use of computer software to analyze a hard drive triggers a reasonable suspicion requirement while a "manual review" of the same hard drive requires no suspicion, is left unexplained. Although technology may serve as a useful proxy for the intrusiveness of a search today, in the future even cursory searches might be more efficiently conducted by the use of such technology. Under the majority's reasonable suspicion standard, individuals' privacy rights are only as secure as the sophistication of the government's current search mechanism.

Moreover, the task of distinguishing these "comprehensive and intrusive" laptop searches from the "unintrusive search" of a laptop affirmed in *Arnold*, 533 F.3d at 1008, or the search of a private letter affirmed in *United States v. Seljan*, 547 F.3d 993, 1003 (9th Cir. 2008) (en banc), leaves border patrol officers with a difficult choice: either protect our nation from those who mean us harm, or risk their own jobs and livelihood in a *Bivens* action, or disciplinary

proceedings. Apart from being administratively impractical, the majority's reasonable suspicion requirement disregards well established border search jurisprudence, and undermines vital national security interests. Ironically, the majority did not even need to consider the border search doctrine in this case because *the search at issue in this case did not occur at the border*.

Separately, but importantly, the majority's application of the reasonable suspicion requirement to Cotterman is also troubling. The majority purports to be concerned with travelers' "personal privacy and dignity," but its determination that reasonable suspicion exists under the exceedingly weak facts of this case undermines the liberties of U.S. citizens generally—not just at the border, and not just with regard to our digital data—but on every street corner, in every vehicle, and wherever else we rely on the doctrine of reasonable suspicion to safeguard our legitimate privacy interests.

I. The Border Search Doctrine

The majority heralds this as a "watershed" case that requires a narrowing of the border search exception to accommodate the privacy interests allegedly created by new technologies. Yet despite the majority's attempts to avoid the fact, the border search exception is clear and inflexible. The Supreme Court has repeatedly affirmed the breadth of the border search doctrine, extending a reasonable suspicion requirement only to: (1) "highly intrusive searches *of the person*"; (2) "searches of property [that] are so destructive as to require" reasonable suspicion; and (3) searches carried out in a "particularly offensive manner"—of which the Court has yet to find an example. *Flores-Montano*, 541 U.S. at 152,

154 n.2, 156 (quotations and citations omitted) (emphasis added).

The majority misconstrues these narrowly-defined exceptions, reading *Flores-Montano* to require reasonable suspicion whenever a search of property is deemed "overly intrusive." Majority at 18–19. Yet, the exceptions articulated in *Flores-Montano* are far more circumscribed—applying not to "overly intrusive" searches of property, like the search of Cotterman's computer, but only to "*highly* intrusive searches *of the person*." *Flores-Montano*, 541 U.S. at 152 (emphasis added). The majority's adoption of a reasonable suspicion requirement to "comprehensive forensic examination[s]" of *property* is irreconcilable with *Flores-Montano*. Majority at 6.

We have consistently rejected a reasonable suspicion requirement for border searches of expressive materials, such as papers and their modern-day equivalent—the data contained on electronic storage devices. *See, e.g.*, *Seljan*, 547 F.3d at 1003 ("An envelope containing personal correspondence is not uniquely protected from search at the border."); *Arnold*, 533 F.3d at 1008 ("[R]easonable suspicion is not needed for customs officials to search a laptop or other personal electronic storage devices at the border."). The majority states that its en banc decision narrows *Arnold* to permit only "relatively simple" border searches of laptops, and "not to countenance suspicionless forensic examinations." Majority at 14 n.6. In narrowing *Arnold*, however, the court creates a circuit split regarding the application of reasonable suspicion to border searches of electronic devices. *See United States v. Ickes*, 393 F.3d 501 (4th Cir. 2005); *see also United States v. Linarez-Delgado*, 259 Fed. Appx. 506, 508 (3d Cir. 2007).

For instance, in *Ickes* (as in *Arnold*) the defendant-appellant argued that a reasonable suspicion requirement was necessary for laptop searches at the border because otherwise "any person carrying a laptop computer [] on an international flight would be subject to a search of the files on the computer hard drive." *Ickes*, 393 F.3d at 506–07. The Fourth Circuit rejected this argument, noting that

> "[a]s a practical matter, computer searches are most likely to occur where—as here—the traveler's conduct or the presence of other items in his possession suggest the need to search further. However, to state the probability that reasonable suspicions will give rise to more intrusive searches *is a far cry from enthroning this notion as a matter of constitutional law*. The essence of border search doctrine is a reliance upon the trained observations and judgments of customs officials, *rather than upon constitutional requirements applied to the inapposite context of this sort of search*."

Id. at 507 (emphasis added). The Third Circuit similarly rejected a reasonable suspicion requirement for border searches of electronic data, albeit in an unpublished opinion. *See United States v. Linarez-Delgado*, 259 Fed. Appx. 506, 508 (3d Cir. 2007) ("Data storage media and electronic equipment, such as films, computer devices, and videotapes, may be inspected and viewed during a reasonable border search.") (citing *Ickes*, 393 F.3d 501). Because the majority has narrowed our holding in *Arnold* that "reasonable suspicion is not needed for customs officials to search a laptop or other personal electronic storage devices at the

border," *Arnold*, 533 F.3d at 1008, the Ninth Circuit stands alone, as it so often does.

The majority likens the search of Cotterman's laptop to a "computer strip search," Majority at 25, and proceeds to conflate the law regarding property searches with that regarding "highly intrusive searches of the person." *Flores-Montano*, 541 U.S. at 152. However, the "reasons that might support a requirement of some level of suspicion in the case of highly intrusive searches *of the person*—dignity and privacy interests *of the person* being searched—simply do not carry over" to laptops, which know no dignity or shame, and thus have neither of those interests. *Flores-Montano*, 541 U.S. at 152 (emphasis added). Moreover, even genuine strip searches do not necessarily require reasonable suspicion at the border. *See United States v. Montoya de Hernandez*, 473 U.S. 531, 541 n.4 (1985) (expressly declining to decide "what level of suspicion, *if any*, is required for . . . strip, body cavity, or involuntary x-ray searches") (emphasis added).

The majority's decision to insulate electronic storage devices from the border search exception unsettles the border search doctrine, places inappropriate burdens on law enforcement, reduces deterrence, and raises serious national security concerns. It also ignores the realities of electronic data transmission and the reduced privacy expectations that accompany much of this data, particularly at the border where "[t]he government's interest in preventing the entry of unwanted persons and effects is at its zenith." *Flores-Montano*, 541 U.S. at 152.

A. Burdens on Law Enforcement

The majority's holding cripples law enforcement at the border by depriving border patrol agents of the clear administrative guidance they need to carry out core law enforcement activities. "Officers who interact with those suspected of violating the law have an essential interest in readily administrable rules." *Florence v. Bd. of Chosen Freeholders of Cnty. of Burlington*, 132 S. Ct. 1510, 1522 (2012). Yet the majority's holding requires border patrol agents to determine on a case-by-case and moment-by-moment basis whether a search of digital data remains "unintrusive," *a la Arnold*, or has become "comprehensive and intrusive," *a la Cotterman*. Majority at 14, 17. Requiring law enforcement to make such complex legal determinations on the spot, and in the face of potentially grave national security threats, strips agents of their necessary discretion and deprives them of an efficient and administrable rule.

The majority dismisses the burden its reasonable suspicion requirement places on law enforcement, asserting that agents can simply "draw on their expertise and experience" to make the necessary judgment calls. Majority at 26. Yet rather than actually deferring to this expertise and experience, the majority forces border patrol agents to justify their decisions under a heightened standard that has never before been applied to border searches of property.

Border patrol agents process hundreds of thousands of travelers each day and conduct thousands of searches on

electronic devices each year.[1] Identifying national security and criminal threats at the border requires a high level of experience and discretion in order to recognize and respond to the ever-changing tactics of those who seek to enter our country with nefarious intent. In recognition of these crucial interests, the border search exception provides law enforcement with broad discretion to conduct border searches of property without resorting to case-by-case determinations of reasonable suspicion—determinations border patrol agents are ill-equipped to handle. *See generally Florence*, 132 S. Ct. at 1522 (rejecting reasonable suspicion requirement for prison strip-searches under this rationale). Moreover, as a practical matter, suspicionless border searches of property make sense, in light of the sheer number of individuals crossing the border with electronic devices each day. *See United States v. Martinez-Fuerte*, 428 U.S. 543, 557 (1976) (requiring reasonable suspicion for vehicle checkpoints near the Mexican border "would be impractical because the flow of traffic tends to be too heavy to allow the particularized study of a given car"). Given these realities of law enforcement at the border, a reasonable suspicion requirement for all "overly intrusive" electronic searches is simply not practicable.

B. National Security Concerns

The majority's decision to insulate electronic devices from search at the border creates serious national security concerns. An "ever present threat exists from the potential for terrorists to employ the same smuggling and transportation networks, infrastructure, drop houses, and other support" as other illegal aliens. U.S. Customs and

[1] Department of Homeland Security Privacy Office, Annual Report to Congress 54 (2009).

Border Protection, National Border Patrol Strategy 5 (2005). The Department of Homeland Security has found that border searches of electronic storage devices are "essential" for "detect[ing] evidence relating to terrorism and other national security matters."[2] Terrorists rely on electronic storage devices, for example, to copy and alter passports and other travel documents.[3] By providing special privacy protections for electronic devices at the border, the majority eliminates the powerful deterrent of suspicionless searches and significantly aids technologically savvy terrorists and criminals who rely on encryption and other surreptitious forms of data storage in their efforts to do harm. *See Martinez-Fuerte*, 428 U.S. at 557 (rejecting reasonable suspicion requirement for vehicle checkpoints near the Mexican border because to hold otherwise "would largely eliminate any deterrent to the conduct of well-disguised smuggling operations").

The majority contends that the goal of deterrence does not justify "any manner of intrusive search" at the border. Majority at 26. Although I certainly agree with the majority that a policy objective like deterrence cannot justify an otherwise unconstitutional "highly intrusive search[] *of the person*" at the border, *Flores-Montano*, 541 U.S. at 152, the crucial role of deterrence cannot, and should not, be understated. In fact, the Supreme Court recently affirmed the importance of deterrence in upholding suspicionless *strip*

[2] U.S. Customs and Border Protection, Border Search of Electronic Devices Containing Information, CBP Directive No. 3340-049 § 1 (2009).

[3] Thomas R. Eldridge, *et al.*, 9/11 and Terrorist Travel: Staff Report of the National Commission on Terrorist Attacks Upon the United States 60 (2004).

searches—the apotheosis of an intrusive search. *Florence*, 132 S. Ct. at 1516 (rejecting reasonable suspicion requirement for prison strip searches and reasoning that "deterring the possession of contraband depends in part on the ability to conduct searches without predictable exceptions"). The suspicionless strip search upheld in *Florence*, which included a close visual inspection of "the buttocks or genital areas," was unquestionably more intrusive than the so-called "computer strip search" at issue here. *Id.* at 1515.

The majority contends that the deterrence function of suspicionless searches will not be hampered by the requirement of reasonable suspicion because, "as a matter of commonsense and resources, it is only when reasonable suspicion is aroused that such searches typically take place." Majority at 27 n.14. This is, of course, the very argument rejected by the Fourth Circuit in *Ickes*. *See Ickes*, 393 F.3d at 507 ("As a practical matter, computer searches are most likely to occur where—as here—the traveler's conduct or the presence of other items in his possession suggest the need to search further. However, to state the probability that reasonable suspicions will give rise to more intrusive searches is a far cry from enthroning this notion as a matter of constitutional law.").

In addition to undermining deterrence, a reasonable suspicion requirement will likely disincentivize agents to conduct laptop searches in close cases. *See Florence*, 132 S. Ct. at 1522 ("To avoid liability" if required to find reasonable suspicion, "officers might be inclined not to conduct a thorough search in any close case, thus creating unnecessary risk for the entire jail population."). Border patrol agents accused of conducting an "unreasonable" search

face very real consequences—as federal officials, for example, they may be sued in their individual capacities for civil damages, as part of a *Bivens*[4] action. *See* Ronald J. Sievert, *Meeting the Twenty-First Century Terrorist Threat Within the Scope of Twentieth Century Constitutional Law*, 37 Hous. L. Rev. 1421, 1424 (2000). The majority's reasonable suspicion requirement saddles border patrol agents with a "Sophie's choice" between securing our nation, and protecting their own livelihoods. These misaligned incentives create unnecessary risk, not just for a prison population, as in *Florence*, 132 S. Ct. at 1522, but for our entire nation.

C. Expectation of Privacy in Electronic Data at the Border

The majority suggests that travelers at the border have a heightened expectation of privacy in their electronic storage devices, due to the "uniquely sensitive nature of [this] data." Majority at 25. There is no question that searches of electronic data are protected by the Fourth Amendment, but we have never found this data to be immune from the border search exception. In fact, these electronic storage devices are hardly a bastion of privacy. When connected to the Internet, they transmit a massive amount of intimate data to the public on an almost constant basis, rendering it unremarkable that they can be searched at the border, where "[t]he government's interest in preventing the entry of unwanted persons and effects is at its zenith." *Flores-Montano*, 541 U.S. at 152.

Indeed, Facebook, for example, now has more than 500 million users, who share more than 25 billion pieces of data

[4] *Bivens v. Six Unknown Fed. Narcotics Agents*, 403 U.S. 388 (1971).

each month.[5] Those who opt out of social networking sites are no less susceptible to the ubiquitous Internet cookie, which collects data on users' Internet activities to share or sell with other organizations. Max Stul Oppenheimer, *Consent Revisited*, 13 No. 12 J. Internet L. 3, 4 (2010). Until recently, a federally funded data accumulation system allowed clients to "search tens of billions of data records on individuals and businesses in mere seconds."[6] Considering the steady erosion of our privacy on the Internet, searches of electronic storage devices may be increasingly akin to a well-placed Internet search. Ironically, the majority creates a zone of privacy in electronic devices at the border that is potentially greater than that afforded the Google searches we perform in our own homes, and elsewhere.

The majority muses that "[a] person's digital life ought not be hijacked simply by crossing the border," Majority at 22, but it fails to explain why electronic data deserves special protections when we have never extended such protections to the same data in written form. *See Seljan*, 547 F.3d at 1003 ("An envelope containing personal correspondence is not uniquely protected from search at the border."); *see also United States v. Tsai*, 282 F.3d 690, 696 (9th Cir. 2002) (no reasonable suspicion needed to search a traveler's briefcase); *United States v. Grayson*, 597 F.2d 1225, 1228–29 (9th Cir. 1979) (no reasonable suspicion needed to search papers found

[5] Jeffrey Rosen, The Deciders: Facebook, Google, and the Future of Privacy and Free Speech, in *Constitution 3.0: Freedom and Technological Change* (*Constitution 3.0*) 76 (Jeffrey Rosen & Benjamin Wittes eds., Brookings Institution Press 2011).

[6] Christopher Slobogin, Is the Fourth Amendment Relevant?, in *Constitution 3.0* 18 (citing Laura K. Donohue, *Anglo-American Privacy and Surveillance*, 96 J. Crim. L. & Criminology 1059, 1150–51 (2006)).

in a shirt pocket); *Henderson v. United States*, 390 F.2d 805, 808 (9th Cir. 1967) (no reasonable suspicion needed to search a traveler's "purse, wallet, or pockets"). The documents carried on today's smart phones and laptops are different only in form, but not in substance, from yesterday's papers, carried in briefcases and wallets. The majority contends that electronic devices hold data of a "uniquely sensitive nature" and that, inexplicably, these devices have the "capability to . . . augment the expectation of privacy." Majority at 23, 25. *Under the majority's reasoning, the mere process of digitalizing our diaries and work documents somehow increases the "sensitive nature" of the data therein, providing travelers with a greater expectation of privacy in a diary that happens to be produced on an iPad rather than a legal pad.* Such artificial and arbitrary distinctions cannot serve as a reasonable basis for determining privacy rights at the border.

The majority attempts to distinguish electronic devices from papers by the vast amount of data they can hold, noting that "[a] car full of packed suitcases . . . cannot hold a candle to the sheer, and ever-increasing, capacity of digital storage." Majority at 21. Yet, "case law does not support a finding that a search which occurs in an otherwise ordinary manner, is 'particularly offensive' simply due to the storage capacity of the object being searched." *Arnold*, 533 F.3d at 1010. The majority contends that it "discuss[es] the typical storage capacity of electronic devices simply to highlight the features that generally distinguish them from traditional baggage." Majority at 21 n.10. Yet why the majority would bother to distinguish between the storage capacities of electronic devices and traditional luggage is a mystery, unless to support its enhanced protections for electronic devices based on their greater storage capacity.

Mapping our privacy rights by the amount of information we carry with us leads to unreasonable and absurd results. Under the majority's reasoning, a Mini Cooper filled with documents is entitled to less privacy protection at the border than a stretch Rolls-Royce filled with documents; a pickup truck filled with documents is entitled to less protection than an 18 wheeler filled with documents. It appears that those who cannot afford a 64 gigabyte iPad, or the "average" 400 gigabyte hard drive discussed by the majority, Majority at 20, will alone be subject to suspicionless searches. The majority's reasoning also protects the rich (who can generally afford more sophisticated devices) to a greater extent than the poor (who are presumably less able to afford those more capable devices.) *See United States v. Ross*, 456 U.S. 798, 822 (1982) ("[A] traveler who carries a toothbrush and a few articles of clothing in a paper bag or knotted scarf claim[s] an equal right to conceal his possessions from official inspection as the sophisticated executive with the locked attache case.").

If our privacy interests are to be dictated by the quantity of data we possess, the question then becomes, how many gigabytes of storage must one buy to secure the guarantee that reasonable suspicion will be required before one's devices are searched? The majority gives us no firm basis for deciding how much storage space is necessary—32 gigabytes? 64 gigabytes? 400 gigabytes? Who knows? Moreover, the majority's test must constantly change to accommodate the ever-increasing capacity of electronic storage and new technologies. Before we know it, today's "average" 400 gigabyte hard drive will look like yesterday's diary next to tomorrow's "average" 2 terabyte hard drive.

The majority asserts that our "reasonableness determination must account for differences in property."

Majority at 24. This assertion has no basis in law, however, since *Flores-Montano* distinguished not between types of property, but between searches of property and "searches *of the person*." *Flores-Montano*, 541 U.S. at 152 (emphasis added). In any event, it appears that the majority's reasonableness requirement accounts not for "differences in property," as it suggests, but rather for differences in the *intrusiveness* of a particular property search. As discussed *supra*, however, these intrusiveness-based tests have no place in border searches of property and have been explicitly rejected by the Supreme Court as "[c]omplex balancing tests." *Flores-Montano*, 541 U.S. at 152.

The majority additionally speculates about the privacy implications of searching an external cloud platform, which may "includ[e] the same kind of highly sensitive data one would have in 'papers' at home." Majority at 23. I share the majority's keen interest in the Fourth Amendment implications of this burgeoning technology, but the reasonableness of cloud computing has no bearing on the case at hand, absent any facts that Cotterman utilized such a platform, or that such a platform was searched.

II. Waiver

There is another important issue in this case that is separate from the majority's new standard for border searches. Specifically, I refer to the majority's finding that there was reasonable suspicion to search Cotterman's computer and other electronic devices, miles from the border. In its zeal to cripple the application of the current border search doctrine, while still securing Cotterman's conviction, the majority turns on their heads all the parties' arguments about reasonable suspicion as to Cotterman, and the findings

made by the lower courts concerning that suspicion. First, the majority now stakes its holding on a finding of reasonable suspicion—despite the fact that the government knowingly and unequivocally *conceded* on appeal any argument that the computer search was supported by reasonable suspicion. Second, the majority's determination that reasonable suspicion was required under the border search exception is contrary to every argument raised by either party in its briefs *prior to* our request for supplemental briefing. Third, even the majority seems to concede that the search of Cotterman's own computer that *actually occurred at the border* did not involve a computer with sufficient storage capacity, and was not sufficiently intrusive, to require reasonable suspicion, under its "new" border search doctrine. Thus, it need not have treated, nor altered, the current border search exception. Fourth, the Magistrate Judge's Report and Recommendation, adopted by the District Judge, did not conclude that reasonable suspicion was required under the border search exception. Despite all the above, the majority upholds Cotterman's conviction on grounds that the government had reasonable suspicion to extensively search his computer 170 miles from the border. Being mindful that the government has the burden of proof in this case, not the majority of our panel, I would have heeded the government's strategic, good faith decision to abandon on appeal its argument that reasonable suspicion existed.[7]

The majority claims that Cotterman has not been prejudiced—despite the fact that the majority's finding of reasonable suspicion is the *raison d'être* for his

[7] When asked during oral argument why it failed to argue reasonable suspicion on appeal, the government acknowledged that the issue was a "close" one.

conviction—because Cotterman was allowed to file a supplemental brief on the matter after oral argument. Although I concede that what the majority did is technically permissible, *see U.S. Nat'l Bank of Oregon v Indep. Ins. Agents of Am., Inc.*, 508 U.S. 439, 446 (1993) ("When an issue or claim is properly before the court, the court is not limited to the particular legal theories advanced by the parties, but rather retains the independent power to identify and apply the proper construction of governing law") (citations and quotations omitted), it is clear to me that *Cotterman has been severely prejudiced*, because his conviction is based solely on an issue the government conceded, and that Appellant, and the lower courts, took for granted because it was not needed for a border search. It is the majority of our panel, not the government, that prosecuted the reasonable suspicion issue in this case.

III. Extended Border Search

The extended border search doctrine applies to "searches that do not occur at the time of entry or in the immediate vicinity of the border." *United States v. Alfonso*, 759 F.2d 728, 735 (9th Cir. 1985). Because these searches "intrude more on an individual's normal expectation of privacy," reasonable suspicion is required. *Id.* at 734.

The majority's mutation of the border search exception is especially unnecessary given that *this search did not occur at the border*, but rather 170 miles away from the border and five days after the border was crossed. Indeed, the majority concedes that the government *could have* performed the forensic computer search at the border, but instead chose to transport Cotterman's electronics more than 170 miles away. By labeling this a border search, the majority has conjured a

sort of "floating border," whereby any item initially seized at the border, but not cleared there, can be transported thousands of miles away and searched anywhere, and at any time, simply because the government did not find anything (or enough) during its original search at the border. Because the search at issue occurred neither "at the time of entry or in the immediate vicinity of the border," it is more appropriately analyzed as an extended border search. *See Alfonso*, 759 F.2d at 735.

The majority asserts that this case cannot be analyzed as an extended border search because Cotterman's computer was never "cleared" at the border prior to search. Majority at 15. The majority is mistaken. In *United States v. Cardona*, 769 F.2d 625, 628 (9th Cir. 1985), we applied the extended border search doctrine to a search of a Federal Express package that occurred twenty-four hours *before* the scheduled border crossing, and 3,000 miles from the border. *See* 769 F.2d at 628 ("Considering the distance and time factors of the present case, we conclude that the facts of this case should be analyzed under the extended border search doctrine.").

While this case presents issues we have not yet addressed in the context of an extended border search, *United States v. Alfonso* is squarely on point. In *Alfonso*, the government conducted an initial, cursory search of a ship upon its arrival at the Los Angeles harbor. *Alfonso*, 759 F.2d at 732. Thirty-six hours later, while still docked at the port, officials conducted a second, more intrusive search. *Id.* Tasked with determining whether the second search was an extended border search or a search at the functional equivalent of the border, we noted that "the instant case illustrates the difficulty of making sharp distinctions in this area." *Id.* at

735. We determined that "[a]lthough we have no difficulty in relating this site with the border, we shall, because of the time factor—the lapse of thirty-six hours in conducting the searches—examine the facts under the rules of extended border search." *Id.* at 734. The majority suggests that cases like *Alfonso* are distinguishable from the case at issue because those cases wrestled with distinguishing between a functional border search and an extended border search, whereas this case involves distinguishing between a traditional border search and an extended border search. This is a distinction without a difference since, as the majority acknowledges, there is no operative difference between border searches and searches that occur at the functional equivalent of the border, at least for purposes of determining whether a search is an extended border search.

I would hold that the search which took place 170 miles from the border, five days after crossing—a much greater lapse than the thirty-six hours in *Alfonso*—requires this case to be analyzed as an extended border search. Additionally, the reasonable suspicion requirement already applies to extended border searches, in recognition of the fact that such searches "intrude more on an individual's normal expectation of privacy." *Id.* As such, the extended border search doctrine is aptly suited to address the privacy expectations at issue in this case.

IV. Reasonable Suspicion

Irrespective of the government's concession of the issue, the evidence in this case falls woefully short of reasonable suspicion. "[R]easonable suspicion exists when an officer is aware of specific, articulable facts which, when considered with objective and reasonable inferences, form a basis for

particularized suspicion." *United States v. Montero-Camargo*, 208 F.3d 1122, 1129 (9th Cir. 2000) (en banc). We assess reasonable suspicion under the totality of the circumstances, "tak[ing] into account both factors weighing for and against reasonable suspicion." *United States v. Manzo-Jurado*, 457 F.3d 928, 938 (9th Cir. 2006). We "will defer to officers' inferences only when such inferences rationally explain how the objective circumstances 'aroused a reasonable suspicion that *the particular person being stopped* had committed or was about to commit a crime.'" *Manzo-Jurado*, 457 F.3d at 934–35 (quoting *Montero-Camargo*, 208 F.3d at 1129) (alterations omitted). "Reasonable suspicion may not be based on broad profiles which cast suspicion on entire categories of people without any individualized suspicion of the particular person to be stopped." *United States v. Sigmond-Ballesteros*, 285 F.3d 1117, 1121 (9th Cir. 2001) (internal quotations and citations omitted).

I agree with the majority that reasonable suspicion was not needed to conduct the initial search of Cotterman's computer at the border, and that we analyze reasonable suspicion only as to the second search (the majority would say a continuation of the initial search,) which took place 170 miles from the border and several days after the border crossing. The majority's reasonable suspicion finding appears to be based *solely* on the TECS alert: it states that "the nature of the alert on Cotterman, directing agents to review media and electronic equipment for child pornography, justified conducting the forensic examination despite the failure of the first search to yield any contraband." Majority at 33. Thus, the majority pins reasonable suspicion on the TECS alert, dismisses out of hand the numerous factors weighing against reasonable suspicion, *and paves the*

way for a government database to target "entire categories of people without any individualized suspicion of the particular person to be stopped." Sigmond-Ballesteros, 285 F.3d at 1121 (internal quotations and citations omitted) (emphasis added). The majority considers the TECS alert to be a sufficient basis for reasonable suspicion, but in reality, it is nothing more than a mechanism that automatically flags all individuals who are registered sex offenders in California—no matter the nature of the sex offense or how old the conviction—who travel frequently.[8] California is home to more than 106,000 sex offenders.[9] Some of these individuals are required to register as sex offenders for life. Depending on how many of them travel frequently, a TECS hit could affect tens of thousands of Californians—many with decades-old convictions. The TECS database clearly hits on "a very large category of presumably innocent travelers, who would be subject to virtually random seizures were the Court to conclude that as little foundation as there was in this case could justify a seizure." *Reid v. Georgia*, 448 U.S. 438, 441 (1980). By allowing reasonable suspicion to rest entirely on the TECS alert, the majority rules that a decades-old conviction can serve as a basis to deprive a person of his or

[8] The TECS alert is part of Operation Angel Watch, a program that targets California residents who are registered sex offenders based on the suspicion that those individuals who travel internationally are engaging in child sex tourism. The majority at one point improperly lists "the parameters of the Operation Angel Watch program" as an independent factor supporting reasonable suspicion. Majority at 30–31. We must look solely at the underlying *facts* supporting reasonable suspicion—*i.e.*, Cotterman's status as a sex offender and his frequent travel—rather than the database or mechanisms used to deliver that information.

[9] Press Release, National Center for Missing and Exploited Children, Number of Registered Sex Offenders in the US Nears Three-Quarters of a Million (Jan. 23, 2012).

her property for an indefinite period, so that a "border search" may be conducted hundreds of miles from the border.

The majority suggests that the TECS alert informed border patrol agents of the nature of Cotterman's conviction. In fact, the TECS hit did not state the nature of Cotterman's conviction, although one agent mistakenly recollected that "it stated that [Cotterman] appeared to [sic] been involved in some type of child pornography." Curiously, another agent stated that a criminal history check on Cotterman revealed that "that he had a prior conviction pertaining to child pornography." In fact, and despite the erroneous contentions of the referenced agents, Cotterman had no prior child pornography conviction; he had a 15-year-old conviction for sexual conduct with a minor. While we generally give "due weight to inferences drawn" by law enforcement, *Ornelas v. United States*, 517 U.S. 690, 699 (1996), the case for deference is questionable here in the absence of any rational explanation as to how the officers could have read the TECS alert and criminal history check, neither of which listed a conviction for child pornography, and come away thinking that Cotterman was guilty of that offense. *See Manzo-Jurado*, 457 F.3d at 934–35 ("[W]e will defer to officers' inferences only when such inferences rationally explain how the objective circumstances aroused a reasonable suspicion."); *see also Liberal v. Estrada*, 632 F.3d 1064, 1078 (9th Cir. 2011) (mistake of fact must be reasonable).

All things considered, the fact that an individual with a 15-year-old sex conviction was also a frequent traveler appears to be a rather weak lynchpin for reasonable suspicion. Yet, other than Cotterman's prior conviction and travels, the factors cited by the majority are far too generalized to provide even an indicia of suspicion that Cotterman was involved in

sex tourism. For instance, the majority considers Cotterman's "collection of electronic equipment" to be a factor supporting reasonable suspicion. In today's world, the fact that Cotterman and his wife each carried a laptop and digital camera when traveling internationally, as well as one video camera between them,[10] is no more evidence of "sex tourism" than of any other kind of tourism.

Similarly, the fact that Cotterman was returning from Mexico fails to support a finding of reasonable suspicion. Mexico is a popular travel destination for Californians, including those who travel to Mexico for its beaches, culture and weather, and not for its sex tourism. Travel to Mexico simply does not support reasonable suspicion without more specific evidence that Cotterman traveled to a particular establishment, city, or even region, associated with sex tourism. *See United States v. Irving*, 452 F.3d 110, 114, 124 (2d Cir. 2006) (finding reasonable suspicion based on knowledge that suspect, a convicted pedophile and the subject of criminal investigation, had visited an orphanage in Mexico and had luggage with children's books and drawings). According to the Department of Justice, American sex tourism is a problem not only in Mexico, but also in Southeast Asia, Central and South America, and, to a lesser extent, Eastern Europe.[11] Under the majority's application of reasonable suspicion, an individual who committed a sex offense 30 years ago cannot visit the Charles Bridge in Prague, the Cristo Redentor in Rio de Janeiro, or even the "lost city" of Machu Picchu, without arousing a "reasonable"

[10] The video camera was apparently Mrs. Cotterman's.

[11] U.S. Department of Justice, The National Strategy for Child Exploitation Prevention and Interdiction, A Report to Congress 36 (2010).

suspicion of sex tourism. Someone who was convicted of tax evasion 15 years ago, or any other kind of conviction listed on a federal database, and particularly one that involved the use of a computer, should also probably avoid visiting Switzerland or Luxemburg under the majority's new standard. The bottom line is that thousands of individuals—many with decades-old convictions—will now be forced to reconsider traveling to entire countries or even *continents*, or will need to leave all their electronic equipment behind, to avoid arousing a "reasonable" suspicion.

Perhaps the most concerning aspect of the majority's opinion, especially given its stated stance on privacy rights at the border, is its readiness to strip former sex offenders and others convicted of past crimes (and who are, theoretically, entitled to be presumption of innocence) of even the most basic of privacy rights, such as the right to password-protect their electronic devices. The majority acknowledges that "it is commonplace for business travelers, casual computer users, students and others to password protect their files" and that "password protection is ubiquitous." Majority at 31. It avers that "[n]ational standards *require* that users of mobile electronic devices password protect their files," and that "[c]omputer users are routinely advised—and in some cases, required by employers—to protect their files when traveling overseas." Majority at 31 (emphasis added). Yet because border patrol agents encountered a *single password-protected file* on Cotterman's computer, the majority considers password protection a factor contributing to reasonable suspicion.[12] Worse still, the majority contends that it is

[12] The unequivocal testimony of Agent Antonio Alvarado confirms that only a single password-protected file was discovered on Cotterman's computer at the border.

justified in considering the password-protected file because "making illegal files difficult to access makes perfect sense for a suspected holder of child pornography." Majority at 32. I fail to see how the agents had reasonable suspicion that Cotterman's computer contained "illegal files" based solely on his 15-year-old sex offense, travel to Mexico with his wife, and the "ubiquitous" act of password-protection. Indeed, as the majority acknowledges, making *legal* files difficult to access makes "perfect sense" for anyone, even former sex offenders.

I would find a password-protected file to be *not at all* suspicious, unless we want to start basing reasonable suspicion on locked diaries and briefcases. Registered sex offenders face numerous consequences as a result of their convictions, but the law has never before punished them for using "ubiquitous" and "commonplace" password-protection. Yet under the majority's analysis, an individual traveling to Southeast Asia for business, who happens to be subject to one of TECS's broad-based alerts, and who follows his company's security protocols, should expect to have his electronic equipment seized and transported hundreds of miles away.[13]

Moreover, the majority fails to consider reasonable suspicion in light of the totality of the circumstances because

[13] The majority finds ironic my concern about the expansiveness of its reasonable suspicion standard, since at the border, I would advocate for no suspicion at all. The majority is correct that at the border, my concern is simply with following *Flores-Montano* and maintaining national security. I view the majority's application of its reasonable suspicion requirement as a separate issue, and my concern there is that the majority has so diluted the reasonable suspicion requirement as to undermine the rights of U.S. citizens generally.

it dismisses without explanation numerous factors that weigh *against* a finding of reasonable suspicion in this case. *See Manzo-Jurado*, 457 F.3d at 938 (the reasonable suspicion determination must "take[] into account both factors weighing for *and against* reasonable suspicion.") (emphasis added). At the time the border patrol agents commenced the second search, 170 miles away from the border, any suspicions they may have initially harbored against Cotterman would have been largely addressed by their interrogations of Cotterman and his wife, which produced nothing suspicious. An initial search of Cotterman's computer and the digital cameras turned up nothing more than a single password protected file and photos of "whale hunting and various excursions," all of which corroborated Cotterman's story about vacationing in Mexico. Also during this initial search, one of the border patrol agents did a records check and discovered that Cotterman's conviction for the sex offense had occurred more than 15 years ago. Cotterman was fully cooperative and even offered to help the agents access his computer. The majority contends that Cotterman's offer to help does not weigh against a finding of reasonable suspicion because the agents declined Cotterman's offer based on the possibility—however slight—that Cotterman could "booby trap" the devices. That the agents were unable to accept Cotterman's offer, however, does not change the reasonable inference that his offer was a genuine one.

Accordingly, it is irrelevant whether there was reasonable suspicion for the initial search, because I agree with the majority that reasonable suspicion was not required. The relevant inquiry here is what suspicion existed after all of Cotterman's electronics were searched, and he and his wife were interrogated separately, and every piece of evidence

obtained corroborated the Cottermans' story about vacationing in Mexico. The only hint of suspicion remaining at that point—after the initial border search and interrogations—was the single password-protected file, which I agree with the majority is insufficient, by itself, to sustain a finding of reasonable suspicion. *See Manzo-Juardo*, 457 F.3d at 935 ("[T]o establish reasonable suspicion, an officer cannot rely solely on generalizations that, if accepted, would cast suspicion on large segments of the lawabiding population.").

V. Conclusion

Reasonable suspicion has no place in property searches at the border, as the Supreme Court has consistently held. *See Flores-Montano*, 541 U.S. at 152–53 ("Time and time again, we have stated that searches made at the border, pursuant to the longstanding right of the sovereign to protect itself by stopping and examining persons and property crossing into this country, are reasonable simply by virtue of the fact that they occur at the border."). Imposing a reasonable suspicion requirement here forces courts and border patrol agents to engage in just the "sort of decision-making process that the Supreme Court wished to avoid in sanctioning expansive border searches." *Seljan*, 547 F.3d at 1011 (citation omitted) (Callahan, J. concurring). Rather than rewrite the border search exception, as the majority does, I would affirm the district court's application of the extended border search doctrine to Cotterman's case, which appears most appropriate given the extensive lapse in distance and time between the first and the second search. Additionally, I would hold the government to its burden of proof in determining that reasonable suspicion was absent here. Under the doctrine of this case, the majority sweeps in thousands of innocent individuals whose electronic equipment can now be taken

away from the border and searched indefinitely, under the border search exception.

I respectfully dissent.

ARTICLE

EMERGING TECHNOLOGY AND CLIENT CONFIDENTIALITY: HOW CHANGING TECHNOLOGY BRINGS ETHICAL DILEMMAS

LOUISE L. HILL*

* Professor of Law, School of Law, Widener University. The author would like to thank John Nivala for his time and thoughtful reflections when reading drafts of this article. The author would also like to thank Maggie Stewart for her diligent and tireless research assistance on this project.

I. INTRODUCTION

As technology advances, new methods for transmitting communications are created. With this emerging technology, lawyers face ethical issues associated with the conveyance of communications, many of which relate to confidentiality and privilege. Because of concern that third parties may intercept or have access to transmitted material, lawyers tend to tread with caution, especially when sensitive information is at issue.

As new tools for communicating information become a part of everyday legal practice, they challenge the parameters of client confidentiality. A matter drawing significant attention today relates to the transmission of documents in electronic form and "metadata," which is hidden information contained in digital documents. Questions arise about lawyers' responsibilities relating to hidden data imbedded in documents. Should liability attach to a lawyer who transmits a document containing hidden sensitive material? Should a lawyer who receives a digital document search for information that might benefit his client? Questions also arise about the effect of the transmission of hidden material that is confidential. Of particular concern is whether transmission of this information, which may be available to a non-privileged viewer, can destroy the privileged nature of a document or communication.

This article will begin by addressing the law of confidentiality in the United States. It will then consider recent changes in technology, and the impact these changes have had on contemporary legal practice. It raises issues associated with the transmission of electronic documents and metadata, with a focus on matters relating to confidentiality and privilege, as well as lawyer responsibility. This article will also examine the divergent positions that different jurisdictions have taken regarding metadata, as well as the effects of those positions. The article will conclude by positing a position for the treatment of metadata, highlighting the duty attaching to lawyers in its treatment.

II. CLIENT CONFIDENTIALITY

In the United States, the law of confidentiality is composed of three key doctrines: attorney-client privilege; lawyer work-product immunity; and a lawyer's ethical duty to maintain client confidences.[1] These are concepts which are related, but distinct.[2] The attorney-client privilege applies "in judicial and other proceedings in which a lawyer may be called as a witness or otherwise required to produce evidence concerning a client."[3] It essentially protects against compelled disclosure of confidential communications exchanged between lawyer and client.[4] A lawyer's ethical duty to maintain client confidences is "not limited to judicial or other proceedings, but rather applies in all representational contexts,"[5] covering all information relating to the representation, not just client communications.[6] Work-product immunity also extends beyond client communications, protecting material from discovery that a lawyer generates in preparing a matter for litigation.[7]

A. Attorney-Client Privilege

The attorney-client privilege is one of the most recognized of the privileges,[8] referred to by Dean John Wigmore as "the oldest of the privileges

---

[1] *See* Charles W. Wolfram, *The U.S. Law of Client Confidentiality: Framework for an International Perspective*, 15 FORDHAM INT'L L.J. 529, 540-44 (1992). A lawyer's ethical duty to maintain client confidences has been referred to as "the agency law of confidentiality." *Id*. at 545.

[2] Arthur Garwin, *Confidentiality and Its Relationship to the Attorney-Client Privilege*, *in* ATTORNEY-CLIENT PRIVILEGE IN CIVIL LITIGATION 31 (Vincent S. Walkowiak ed., 2004). That which is privileged also is "protected by the confidentiality principle but the reverse is not true." *Id.* at 32; *see* Louise L. Hill, *Disparate Positions on Confidentiality and Privilege Across National Boundaries Create Danger and Uncertainty for In-House Counsel and Their Clients*, *in* LEGAL ETHICS FOR IN-HOUSE CORPORATE COUNSEL A-127, 128 (BNA, Corp. Practice Series No. 87, 2007).

[3] MODEL RULES OF PROF'L CONDUCT R. 1.6 cmt. 3 (2009).

[4] *See* GEOFFREY C. HAZARD, JR. & W. WILLIAM HODES, THE LAW OF LAWYERING, §. 9.2, 9-6 (3d ed. Supp. 2003). The privilege protects only the communication, not the underlying facts. Thus there is a distinction between information about a client and communication about that information. *See* Garwin, *supra* note 2, at 32.

[5] Garwin, *supra* note 2, at 31.

[6] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6 (2009).

[7] *See* Wolfram, *supra* note 1, at 543.

[8] *See* J. Triplett Mackintosh & Kristen M. Angus, *Conflict in Confidentiality: How E.U. Laws Leave In-House Counsel Outside the Privilege*, 38 INT'L LAW. 35, 38 (2004); Joseph Pratt, *The Parameters of the Attorney-Client Privilege for In-house Counsel at the International Level: Protecting the Company's Confidential Information*, 20 NW. J. INT'L L. & BUS. 145, 149 (1999).

for confidential communications."[9] Dean Wigmore traces the attorney-client privilege to sixteenth century England where a solicitor was exempted from offering evidence.[10] The privilege has also been traced to Roman times where attorneys were servants of those whose affairs they managed, and under Roman law, could not testify for or against their masters since the relationship created a duty of loyalty.[11] In the United States, the attorney-client privilege is the only communications privilege recognized in every state.[12] Each state has its own privilege rules which generally follow the common law doctrine, while Rule 501 of the Federal Rules of Evidence governs federal courts.[13]

Rather than establishing fixed rules for attorney-client privilege, the United States Supreme Court determined that Rule 501 allows privilege issues to be decided on a case-by-case basis.[14] The Court acknowledged that this approach could "undermine desirable certainty in the boundaries of the attorney-client privilege."[15] Some feel this has come to fruition, in that there is "inconsistency and confusion at the margins of the privilege,"[16] which creates "practical difficulties for attorneys and other legal advisors."[17]

---

[9] 8 JOHN HENRY WIGMORE, EVIDENCE § 2290 (McNaughton Rev. 1961).

[10] *Id.*

[11] *See* Max Radin, *The Privilege of Confidential Communication Between Lawyer and Client*, 16 CAL. L. REV. 487, 487-88 (1927).

[12] *See* Daiske Yoshida, *The Applicability of the Attorney Privilege to Communications with Foreign Legal Professionals*, 66 FORDHAM L. REV. 209, 212 (1997).

[13] *E.g.*, *id*. at 213; Pratt, *supra* note 8, at 151. The Federal Rules of Evidence provide:

> Except as otherwise required by the Constitution of the United States or provided by Act of Congress or in rules prescribed by the Supreme Court pursuant to statutory authority, the privilege of a witness, person, government, State, or political subdivision thereof shall be governed by principles of the common law as they may be interpreted by the courts of the United States in light of reason and experience. However, in civil actions and proceedings with respect to an element of a claim or defense as to which State law supplies the rule of decision, the privilege of witness, person, government, State, or political subdivision thereof shall be determined in accordance with State law.

FED. R. EVID. 501.

[14] Upjohn Co. v. United States, 449 U.S. 383, 396 (1981).

[15] *Id*. at 396-97.

[16] Yoshida, *supra* note 12, at 213. Whether communications of patent agents are entitled to privilege, and whether a general privilege is recognized for communications between in-house counsel and corporate employees, are examples of inconsistencies and confusion created by the case-by-case approach. *Id*. at 214. Third-party disclosure constituting waiver is also an issue on which the Circuits differ. *See* Mackintosh & Angus, *supra* note 8, at 43. The split on third-party disclosures is due in part to ambiguities in statutes that require disclosure of relevant documents and voluntary disclosure provisions of some government agencies. *Id.*

[17] Yoshida, *supra* note 12, at 214.

Attorney-client privilege is based on "a pragmatic judgment that confidentiality is necessary in order to encourage client communication."[18] It is recognized "to promote open and uninhibited consultations with lawyers," which is acknowledged "as providing a significant benefit to society."[19] While acknowledging these attributes, at the same time we are cautioned that the privilege, grounded on subjective considerations, is "an obstacle to the investigation of the truth," which "ought to be strictly confined within the narrowest possible limits consistent with the logic of its principle."[20] It is a rule of evidence, applicable in civil and criminal court proceedings, limiting "the extent to which a party in litigation can force from an unwitting witness a statement or document that is protected as confidential."[21]

As a general premise, the privilege attaches to confidential communications made between privileged persons, for the purpose of obtaining or providing legal assistance.[22] A standard rule of attorney-client privilege in the United States, formulated by Dean Wigmore,[23] is as follows:

(1) Where legal advice of any kind is sought
(2) from a professional legal advisor in his capacity as such,
(3) the communications relating to that purpose,
(4) made in confidence,
(5) by the client,
(6) are at his instance permanently protected
(7) from disclosure by himself or by his legal advisor,
(8) except the protection be waived.[24]

A version of the Wigmore rule, put forward by Judge Wyzanski of the United States District Court, District of Massachusetts,[25] finds there is

[18] Wolfram, *supra* note 1, at 544.

[19] Mackintosh & Angus, *supra* note 8, at 38. The public needs to know the law for society to function smoothly. This is furthered by consultation with attorneys, whose counsel should not result in greater liability. *Id.*

[20] WIGMORE, *supra* note 9, § 2291 ("The policy of the privilege has been plainly grounded since the latter part of the 1700s on subjective considerations." Prior to that, its theory was objective rather than subjective, "a consideration for *the oath and the honor* of the attorney rather than for the apprehensions of his client.").

[21] Wolfram, *supra* note 1, at 541-42.

[22] *Id*. In most jurisdictions, that which a lawyer communicates to a client is subject to the privilege, just as a client communication would be. *Id*. at 542.

[23] The position taken by Dean Wigmore was followed and adhered to by the Second, Sixth, Seventh, Ninth and Tenth Circuits. *See* Gregg F. LoCascio, *Reassessing Attorney-Client Privileged Advice in Patent Litigation*, 69 NOTRE DAME L.REV. 1203, 1207 n.23 (1994).

[24] WIGMORE, *supra* note 9, § 2292.

[25] The position taken by Judge Wyzanski was followed by the First, Third, Fourth, Fifth,

attorney-client privilege when:

(1) The asserted holder of the privilege is or ought to become a client;
(2) the person to whom the communication was made
  (a) is a member of the bar of a court, or his subordinate and
  (b) in connection with this communication is acting as a lawyer;
(3) the communication relates to a fact of which the attorney was informed
  (a) by his client
  (b) without the presence of strangers
  (c) for the purpose of securing primarily either
    (i) an opinion of law or
    (ii) legal services or
    (iii) assistance in some legal proceeding, and
  (d) not for the purpose of committing a crime or tort; and
(4) the privilege has been
  (a) claimed and
  (b) not waived by the client.[26]

The Restatement (Third) of the Law Governing Lawyers more briefly defines the attorney-client privilege as: (1) a communication; (2) made between privileged persons; (3) in confidence; (4) for the purpose of obtaining or providing legal assistance to the client.[27] Some courts have "treated the Wigmore and Restatement definitions as sufficiently similar to be somewhat interchangeable."[28]

1. Waiver

Attorney-client privilege can be waived; and waiver of the privilege is absolute, being "construed broadly against the party claiming the privilege."[29] At issue is whether waiver is triggered when confidential information is embedded in a document that is sent to a third party. Waiver can result from intentional voluntary disclosure, inadvertent disclosure, or the offensive use of

Eighth, Eleventh and District of Columbia Circuits. *See* LoCascio, *supra* note 23, at 1209 n.30.

[26] United States v. United Shoe Mach. Corp., 89 F. Supp. 357, 358-59 (D. Mass. 1950).

[27] RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 68 (2000). While seemingly simply stated, elements of the Restatement definition are further defined within other Restatement sections. *See* James N. Willi, *Proposal for a Uniform Federal Common Law of Attorney-Client Privilege for Communications with U.S. and Foreign Patent Practitioners*, 13 TEX. INTELL. PROP. L.J. 279, 289 (2005).

[28] *Id*. at 289; *see also* Pratt, *supra* note 8, at 152-53.

[29] Mackintosh & Angus, *supra* note 8, at 43.

otherwise privileged communications.[30] Some members of the legal community see disclosure of protected communications to a third party as the greatest threat to the protections offered by the attorney-client privilege.[31]

a. Voluntary Disclosure

"The client, not counsel, can voluntarily waive the privilege."[32] If a client willingly shares a privileged communication with a non-privileged person, "a court will feel free to find that, in this instance, the assurance of confidentiality was not important to the client, and that the general policy of free access by adversaries to all relevant evidence should prevail."[33] Some courts take the position that voluntary disclosure pursuant to a government subpoena constitutes only "limited waiver," retaining the disclosed communication's privileged status against other parties.[34] Other courts, however, find that voluntary disclosure to any non-privileged party constitutes waiver.[35]

---

[30] *Id*.; Wolfram, *supra* note 1, at 544.

[31] *See* Mackintosh & Angus, *supra* note 8, at 43.

[32] *Id*. at 42-43.

[33] Wolfram, *supra* note 1, at 544. Two or more parties with a common interest that "is the subject of confidential communications generally are allowed to share this information without losing the attorney-client privilege." ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw). It is felt that in litigation, parties with allied interests should be able to communicate and coordinate their positions so as to more effectively present their claims. *Id.*

[34] *See* Westinghouse Elec. Corp. v. Republic of the Philippines, 951 F.2d 1414, 1427-29 (3d Cir. 1991) (discussing how various courts have treated the theory of selective waiver); Diversified Indus. v. Meredith, 572 F.2d 596, 611 (8th Cir. 1977); Leonen v. Johns-Manville, 135 F.R.D. 94, 99 (D.N.J. 1990); Palmer v. Farmers Ins. Exch., 861 P.2d 895, 908-09 (Mont. 1993); *see also* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[35] Most federal courts view these acts as voluntary disclosure constituting a complete waiver of the attorney-client privilege. *See* United States v. Mass. Inst. of Tech., 129 F.3d 681, 685 (1st Cir. 1997); Genentech Inc. v. U.S. Int'l Trade Comm'n, 122 F.3d 1409, 1417 (Fed. Cir. 1997); *In re* Steinhardt Partners L.P., 9 F.3d 230, 235 (2d Cir. 1993); Westinghouse Elec. Corp., 951 F.2d at 1424-26 ; *In re* Martin Marietta Corp., 856 F.2d 619, 623-24 (4th Cir. 1988); Permian Corp. v. United States, 665 F.2d 1214, 1220-21 (D.C. Cir. 1981); Mackintosh & Angus, *supra* note 8, at 45 n.87. Although currently under attack by organizations as well as legislators, some federal programs encourage companies to self-report wrongdoing, or give mitigation credit to companies that make voluntary disclosures of privileged material. *See, e.g.*, "House Overwhelmingly Approves Bill to Limit DOJ Policy on Corporate Privilege Waivers," [Current Report] 23 Laws. Man. on Prof. Conduct (ABA/BNA) No. 24, at 604 (Nov. 28, 2007); "Counsel Group Assails Prosecution Policy Compelling Corporations to Waive Privileges," [Current Report] 16 Laws. Man. on Prof.

b. Inadvertent Disclosure

Opinion differs on whether attorney-client privilege is waived when there is inadvertent disclosure. Typically, when approaching the issue of inadvertent disclosure, one of three tests is applied: (1) "the strict responsibility test," where any disclosure, even inadvertent disclosures, waives attorney-client privilege; (2) "the subjective intent test," where inadvertent disclosure does not waive attorney-client privilege since waiver requires an intention to waive; or (3) "the balancing test," where waiver is determined based on an evaluation of circumstances.[36]

The strict responsibility test, which is the traditional test, was adhered to by Wigmore, putting the "risk of insufficient precautions on the client."[37] The rationale for the strict view is as follows: privilege acts as an obstacle to discovery of the truth; disclosure of privileged materials makes it impossible to achieve the benefits of privilege; therefore, "when the policy underlying the rule can no longer be served, it would amount to no more than mechanical obedience to a formula to continue to recognize it."[38] Some courts have favored this strict test because it forces self-regulatory behavior, and to do otherwise would be unfair to the party seeking to use the inadvertently disclosed communication.[39] It is also heralded as predictable and easy to apply.[40] Additionally, some find inadvertence "a euphemism for negligence, and, certainly . . . one is expected to pay a price for one's negligence."[41]

The subjective intent test, the most lenient approach, "holds that as long as the client did not intend to waive, the privilege remains intact, despite disclosure of the client's confidences to her adversary."[42] A court adopting

Conduct (ABA/BNA) No. 10, at 275 (June 7, 2000); *see also* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[36] *See* Mackintosh & Angus, *supra* note 8, at 43 n.58; ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[37] WIGMORE, *supra* note 9, at 633.

[38] Vincent S. Walkowiak, Sarah E. Lemons & Thomas J. Leach, *Loss of Attorney-Client Privilege Through Inadvertent Disclosure of Privileged Documents*, *in* ATTORNEY-CLIENT PRIVILEGE IN CIVIL LITIGATION 313, 316 (Vincent S. Walkowiak ed., 2004) (quoting United States v. Kelsey-Hayes Wheel Co., 15 F.R.D. 461, 465 (E.D. Mich. 1954)).

[39] *See* Walkowiak, Lemons & Leach, *supra* note 38, at 317 (citing Suburban Sew 'N Sweep, Inc. v. Swiss-Bernina, Inc., 91 F.R.D. 254 (N.D. Ill. 1981); *In re* Sealed Case, 877 F.2d. 976, 980 (D.C. Cir. 1989)).

[40] *See* Walkowiak, Lemons & Leach, *supra* note 38, at 317.

[41] *Id*. at 316 (quoting *In re* Standard Fin. Mgmt. Corp., 77 B.R. 324, 330 (D. Mass. 1987)).

[42] Walkowiak, Lemons & Leach., *supra* note 38, at 318.

this approach found it "the better-reasoned rule."[43] The essence of the court's rationale was that "'inadvertent production is the antithesis' of an intentional relinquishment of a known right and, if the privilege is for the welfare of the *client*, more than the *attorney's* negligence should be required before the *client* loses the privilege."[44] As with the strict responsibility test, the subjective intent test has the advantage of being reasonably predictable and easy to apply.[45] It is criticized, however, in that "it exalts subjective considerations over objective ones,"[46] and does little to encourage care of privileged documents.[47]

The most popular of the three tests is the balancing test, where courts weigh circumstances which surround an inadvertent disclosure to determine whether a loss of privilege should result.[48] Although its proponents acknowledge that it is more difficult to apply, "ultimately this approach is fairer to both parties and the policy of preserving the privilege for confidential communications as it focuses on the confidentiality aspect of the privilege."[49] Described as the middle ground, this approach is sometimes criticized for being uncertain and for giving too much discretion to the court.[50] When determining if privilege is retained, courts generally balance the reasonableness of precautions taken to prevent disclosure, the time taken to recognize the error, the scope of the production, the extent of the disclosure, and considerations of fairness and justice.[51] Although routinely presented as a multi-factor test, it has nevertheless been asserted that courts primarily concentrate on only two considerations, those being "the conduct of the client and lawyer claiming the privilege, and the prejudice to the party to whom the privileged material was disclosed should the court uphold the privilege despite disclosure."[52] Two main questions are implicated: "did the lawyer/client invoking the privilege really act in a careful manner we expect from someone truly concerned with guarding a confidence, both before and after the inadvertent disclosure?" and,

[43] Mendenhall v. Barber-Greene Co., 531 F. Supp. 951, 954 (N.D. Ill. 1982).

[44] Walkowiak, Lemons & Leach, *supra* note 38, at 318 (quoting Mendenhall, 531 F. Supp. at 955).

[45] *Id.* at 319.

[46] *Id.* at 318.

[47] *Id*. at 319.

[48] *Id.*

[49] *Id*. (quoting Kanter v.Superior Court, 253 Cal. Rptr. 810, 815 (Ct. App. 1988)).

[50] *Id.* at 321.

[51] *See* Mackintosh & Angus, *supra* note 8, at 45 n.87. The Chancery Court in Delaware notes that overall fairness must be "judged against the care or negligence with which the privilege is guarded." *In re* Kent County Adequate Public Facilities Ordinances Litigation Consolidated, C.A. No. 2921-VCN, at 5 (Del. Ch. April 7, 2008) 2008 WL 1851790.

[52] Walkowiak, Lemons & Leach, *supra* note 38, at 321.

"[w]ould it be fair to the person who has received the privileged information to try to make her expunge her knowledge of it from the litigation?"[53]

In December 2007, Senators Leahy and Specter introduced legislation in the United States Senate [S.2450] to create a new Federal Rule of Evidence 502.[54] The new rule attempted to resolve the disputes and conflicting decisions about the effect of inadvertent disclosure in federal court litigation. The House of Representatives approved the bill creating the new evidentiary rule by voice vote on September 8, 2008, which became law on September 19, 2008 when it was signed by the President of the United States.[55] The rule essentially provides that disclosure of privileged material does not result in the waiver of attorney-client privilege, as long as: (1) the disclosure is inadvertent; (2) the party responsible for the disclosure took reasonable steps to prevent disclosure; and (3) the party responsible for the disclosure took reasonable steps to correct the error after it occurred.[56] The rule also addresses the issue of scope of the waiver. When there is inadvertent disclosure, there is dispute among the courts as to whether the privilege is waived only as to those documents or communications that are inadvertently disclosed, or whether the waiver extends to all communications on the subject covered by the inadvertently disclosed communications.[57] The rule takes the position held by the majority

---

[53] *Id.*

[54] *See* Ralph Lindeman, *Leahy, Specter Introduce Bill to Create Evidence Rule to Prevent Privilege Waivers*, 23 Laws. Man. on Prof. Conduct (ABA/BNA) 646 (Dec. 26, 2007). Any proposed rule that would change an evidentiary privilege must be approved by Congress under the Rules Enabling Act, 28 U.S.C. § 2071. *Id.*

[55] *See* Ralph Lindeman, *House Gives Backing to New Court Rule to Prevent Accidental Privilege Waivers*, 24 Laws. Man. on Prof. Conduct (ABA/BNA) 496 (Sept. 17, 2008).

[56] There were five versions of Bill Number S.2450 for the 110th Congress. *See* http://thomas.loc.gov/cig-bin/thomas. The version which was signed into law, embodied in S.2450, provides as follows at Rule 502(b):

> Inadvertent Disclosure – When made in a Federal proceeding or to a Federal office or agency, the disclosure does not operate as a waiver in a Federal or State proceeding if:
>
> (1) the disclosure is inadvertent;
>
> (2) the holder of the privilege or protection took reasonable steps to prevent disclosure; and
>
> (3) the holder promptly took reasonable steps to rectify the error, including (if applicable) following Federal Rule of Civil Procedure 26(b)(5)(B).

Pub. L. No. 110-322, §1(b), 122 Stat. 3537 (2008). The Advisory Committee on Evidence Rules noted the following with respect to the amendment:

> The rule establishes a compromise between two competing premises. On the one hand, information covered by the attorney-client privilege or work product protection should not be treated lightly. On the other hand, a rule imposing strict liability for an inadvertent disclosure threatens to impose prohibitive costs for privilege review and retention, especially in cases involving electronic discovery.

Report of the Advisory Committee on Evidence Rules, Committee on Rules of Practice and

of courts,[58] which is that any waiver is limited to the actual material disclosed, and does not extend to other material which exists on the covered subject.[59] It is only when the waiver is intentional that it will extend to all related material on the same subject.[60]

While the new Federal Rule of Evidence 502 targets litigation in federal court,[61] state proceedings are also affected.[62] The new rule provides that an inadvertent disclosure first made in state court does not waive the privilege in federal court proceedings.[63] The new rule also provides that a federal court order that privilege is not waived extends the protection against inadvertent waiver to other federal and state court proceedings.[64]

Procedure of the Judicial Conference of the United States, Committee Note on Subdivision (b), May 15, 2006 (revised June 30, 2006).

[57] *See* Mackintosh & Angus, *supra* note 8, at 43 n.58 (citing *In re* Grand Jury Proceedings, 727 F. 2d 1352, 1356 (4th Cir. 1984)).

[58] *Id*.

[59] Rule 502(a), embodied in § 2450, provides as follows:

> Disclosure Made in a Federal Proceeding or to a Federal Office or Agency; Scope of Waiver – When the disclosure is made in a Federal proceeding or to a Federal office or agency and waives the attorney-client privilege or work-product protection, the waiver extends to an undisclosed communication or information in a Federal or State proceeding only if:
>
> (1) the waiver is intentional;
>
> (2) the disclosed and undisclosed communications or information concern the same subject matter; and
>
> (3) they ought in fairness to be considered together.

Pub. L. No. 110-322, §1(a), 122 Stat. 3537 (2008).

[60] *Id*.

[61] *See* Lindeman, *supra* note 54.

[62] Rule 502(f), embodied in § 2450, provides as follows:

> Controlling Effect of This Rule – Notwithstanding Rules 101 and 1101, this rule applies to State proceedings and to Federal court-annexed and Federal court-mandated arbitration proceedings, in the circumstances set out in the rule. And notwithstanding Rule 501, this rule applies even if State law provides the rule of decision.

Pub. L. No. 110-322, §1(a), 122 Stat. 3537 (2008).

[63] Rule 502(c), embodied in § 2450, provides as follows:

> Disclosure Made in a State Proceeding – When the disclosure is made in a State proceeding and is not the subject of a State-court order concerning waiver, the disclosure does not operate as a waiver in a Federal proceeding if the disclosure:
>
> (1) would not be a waiver under this rule if it had been made in a Federal proceeding; or
>
> (2) is not a waiver under the law of the State where the disclosure occurred.

Pub. L. No. 110-322, §1(a), 122 Stat. 3537 (2008).

[64] Rule 502 (d), embodied in § 2450, provides as follows:

> Controlling Effect of a Court Order – A Federal court may order that the privilege or

c. Offensive Use of Otherwise Privileged Communications

Waiver can also result from offensive use of what would otherwise be privileged communications. The offensive use doctrine comes into play when a party to a proceeding introduces an issue related to advice received from a lawyer, impliedly waiving the confidentiality of the communication.[65] For instance, in a New York case, it was determined that the privilege was waived when the defendant relied upon the adequacy of an internal investigation as a defense in a sexual harassment suit.[66] Similarly, in a Delaware case, the court found the privilege was waived when respondents relied on communications with attorneys to support a motion for a protective order to preclude depositions.[67] That said, however, courts differ on the application of this

protection is not waived by disclosure connected with the litigation pending before the court – in which event the disclosure is also not a waiver in any other Federal or State proceeding.

Pub. L. No. 110-322, §1(d), 122 Stat. 3537 (2008).

Recently, the U.S. District Court for the Eastern District of Pennsylvania had an opportunity to apply new Federal Rule of Evidence 502 when 812 privileged documents were included in plaintiff's document production of 78,000 e-mail messages during electronic discovery. Rhoads Industries, Inc. v. Bldg. Materials Corp. of Am., 254 F.R.D. 216 (E.D. Pa. 2008). Defense counsel notified plaintiff's counsel that apparently privileged documents had been produced, whereupon plaintiff's counsel immediately responded that any such disclosure was inadvertent and no privilege had been waived. *Id.* at 218. In considering the matter, the district court looked to the new rule, since subsection (c) calls for it to apply in proceedings commenced after its enactment and "insofar as is just and practicable, in all proceedings pending on such date of enactment." *Id.* The court began by looking to see if the producing party showed "at least minimal compliance" with the three factors in Section 502(b). *Id.* at 226. Concluding that minimal compliance was met since plaintiff took steps to prevent and rectify the inadvertent error, the court went on to note, however, that to some extent its efforts were not reasonable. *Id.* Since the matter of reasonableness was at issue, the court proceeded to the five factor test followed by the majority of courts. *Id.* Applying the traditional five factor test, the court found the first four factors favored the defendant, but the final factor favored the plaintiff. *Id.* Denying documents to the defendants would not be prejudicial to them since they had no right to the privileged documents. Concluding that the defendants did not carry the burden of proof as to the 812 e-mails, the court determined the privileged nature of the e-mails was not forfeited by the inadvertent disclosure. *Id*. at 227.

[65] *See* Mackintosh & Angus, *supra* note 8, at 43 n.57 (citing Chevron Corp. v. Pennzoil Co., 974 F.2d 1156, 1162 (9th Cir. 1992)). The "offensive use" doctrine comes into play when "privileged material is used as a 'sword' rather than as a 'shield.'" Vincent S. Walkowiak, *An Overview of the Attorney Client Privilege When the Client is a Corporation*, *in* ATTORNEY CLIENT PRIVILEGE IN CIVIL LITIGATION 1, 19 (Vincent S. Walkowiak ed., 2004).

[66] Brownell v. Roadway Package Sys. Inc., 185 F.R.D. 19, 25 (N.D.N.Y. 1999).

[67] *See In re* Kent County Adequate Pub. Facilities Ordinances Litig. Consol., C.A. No. 2921-VCN, 2008 WL 1851790, at *5 (Del. Ch. April 7, 2008).

doctrine. At issue may be a determination of when a communication actually has been introduced. Some courts find the privilege is waived when the protected information is integral to the outcome of issues in the lawsuit.[68] Other courts require the privileged material to be "outcome determinative" for there to be waiver.[69] Yet still other courts apply more liberal standards, determining that waiver should be found when: assertion of the privilege is the result of a party's affirmative act; the asserting party put the protected information at issue by making it relevant to the case; and application of the privilege would deny the adversary access to information vital to his defense.[70] Referred to as the Hearn Test, the latter standard has been criticized as being "too liberal and potentially chilling confidential attorney-client communications."[71]

Recently, the U.S. Court of Appeals for the Second Circuit invoked the remedy of mandamus to clarify the uncertainty surrounding the "at issue" waiver.[72] Agreeing with the critics of the Hearn Test, the Second Circuit took the position that it "cuts too broadly," noting that it "would open a great number of privileged communications to claims of at-issue waiver."[73] According to the Second Circuit, an assertion that information is relevant is not enough for waiver. The court determined that for there to be waiver, "a party must *rely* on privileged advice from counsel to make his claim or defense."[74] The court also noted that the issue of fairness underlies privilege waiver, which is a matter that is decided "on a case-by-case basis, and depends primarily on the specific context in which the privilege is asserted."[75]

---

[68] *See* Mortgage Guarantee & Title Co. v. Cunha, 745 A.2d 156, 159 (R.I. 2000); Metro. Ins. Co. v. Aetna Casualty & Surety Co., 730 A.2d 51, 60 (Conn. 1991); *see also* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[69] *See* Republic Ins. Co. v. Davis, 856 S.W.2d 158, 163 (Tex. 1993); *see also* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[70] *See* Hearn v. Rhay, 68 F.R.D. 574, 581 (E.D. Wash. 1975).

[71] ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[72] *See Second Circuit Clarifies Test for "At Issue" Privilege Waiver*, 24 LAWS. MAN. ON PROF. CONDUCT (ABA/BNA), 556 (Oct. 29, 2008).

[73] *In re* County of Erie, 546 F.3d 222, 229 (2d Cir. 2008).

[74] *Id.*

[75] *Id.* The court noted that the issue of unfairness only comes into play "when a party uses an assertion of fact to influence the decision maker while denying its adversary access to privileged material potentially capable of rebutting the assertion." *Id.* (quoting John Doe Co. v. United States, 350 F.3d 299, 306 (2d Cir. 2003)).

2. Crime-Fraud Exception

If a lawyer and client devise a criminal or fraudulent act, and then use the attorney-client privilege as a shield, "the administration of justice is not served."[76] Therefore, the attorney-client privilege is lost when a client either "consults a lawyer for the purpose, later accomplished, of obtaining assistance to engage in a crime or fraud or aiding a third person to do so," or "regardless of the client's purpose at the time of consultation, uses the lawyer's advice or other services to engage in or assist a crime or fraud."[77] The lawyer consulted need not be aware of the client's intent to use the lawyer's services to perpetrate a crime or fraud.[78] In addition to losing attorney-client privilege, the crime-fraud exception also bars work product immunity for a client.[79]

B. Work-Product Immunity

Lawyer work-product immunity has been described as "a broadened but flattened version of the attorney-client privilege."[80] It is broader because it encompasses almost everything a lawyer generates in preparing a case for litigation, not just confidential communications between the lawyer and client.[81] The immunity covers "mental impressions, conclusions, opinions, or legal theories of an attorney" that relate to litigation.[82] The immunity is flattened because the material must be prepared in anticipation of litigation.[83] Anticipated litigation is litigation that need not be imminent, but must be more than a "remote prospect."[84] Material protected as work product is not accessible "through the otherwise broad powers of pretrial discovery."[85] It extends a zone of privacy to preparations for litigation, preventing prepared

[76] ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[77] RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 82 (2000).

[78] *See* United States v. Chen, 99 F.3d 1495, 1504 (9th Cir. 1996); *In re* Grand Jury Proceedings, 102 F.3d 748, 752 (4th Cir. 1996); United States v. Neal, 27 F.3d 1035, 1048 (5th Cir. 1994).

[79] *See In re* Green Grand Jury Proceedings, 492 F.3d 976, 980 (8th Cir. 2007).

[80] Wolfram, *supra* note 1, at 542.

[81] *Id*. at 543. Federal Rule of Civil Procedure 26(b)(3) extends work product to material prepared "by or for another party or by or for that other party's representative (including the other party's attorney, consultant, surety, indemnitor, insurer, or agent)." FED. R. CIV. P. 26(b)(3).

[82] Mackintosh & Angus, *supra* note 8, at 42 (quoting FED. R. CIV. P. 26(b)(3)).

[83] Wolfram, *supra* note 1, at 543-44.

[84] *In re* Special Sept. 1978 Grand Jury (II), 640 F.2d 49, 65 (7th Cir. 1980).

[85] Wolfram, *supra* note 1, at 543.

material from being exploited by adversaries.[86] It is different from attorney-client privilege because a lawyer can disclose work product to persons not assisting the lawyer in trial preparation, without losing immunity status, as long as "the disclosure does not create a substantial risk of divulgence to an adversary in litigation."[87]

A distinction is made between "ordinary" work product and "opinion" work product.[88] Ordinary work product consists of raw factual information, while opinion work product consists of mental impressions, conclusions, opinions or legal theories.[89] There are some situations in which work product protection can be overcome by an opposing party. To overcome work product protection, an opposing party usually must demonstrate a substantial need for the work product materials.[90] However, with respect to "opinion" work product, this type of material "is discoverable, if at all, only upon a showing of compelling need."[91]

Work product protection may also be vitiated by a prima facie showing of a crime or fraud,[92] or in some instances, unethical conduct by a lawyer.[93]

---

[86] *See* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[87] Wolfram, *supra* note 1, at 544.

[88] *See In re* Green Grand Jury Proceedings, 492 F.3d 976, 980 (8th Cir. 2007); Palmer v. Farmers Ins. Exch., 861 P.2d 895, 908-09 (Mont. 1993).

[89] *In re* Green, 492 F.3d at 980; Palmer, 861 P.2d at 910.

[90] Fed. R. Civ. P. 26(b)(3), provides in part as follows:

> Trial Preparation: Materials
>
> (A) *Documents and Tangible Things.* Ordinarily, a party may not discover documents and tangible things that are prepared in anticipation of litigation or for trial by or for another party or its representative (including the other party's attorney, consultant, surety, indemnitor, insurer, or agent). But, subject to Rule 26(b)(4), those materials may be discovered if:
>
> (i) they are otherwise discoverable under Rule 26(b)(1); and
>
> (ii) the party shows that it has substantial need for the materials to prepare its case and cannot, without undue hardship, obtain their substantial equivalent by other means.
>
> (B) *Protection Against Disclosure.* If the court orders discovery of those materials, it must protect against disclosure of the mental impressions, conclusions, opinions, or legal theories of a party's attorney or other representative concerning the litigation.

FED. R. CIV. P 26(b)(3)(A)&(B).

[91] ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw); *see* Palmer v. Farmers Ins. Exch., 861 P.2d at 911.

[92] *See* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw); RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 93 (2000); *see also supra* notes

However, a lawyer's independent work-product privilege is considered as a separate matter. A lawyer may assert the work-product doctrine with regard to opinion work product even if the client has used the lawyer's services for fraudulent or criminal purposes, so long as the lawyer was unaware that the client was doing so.[94] Seeking the lawyer's advice about the consequences of past activities does not fall within the crime-fraud exception.[95] Nor does the exception apply when the client does not accomplish the crime or fraud, for that would "penalize a client for doing what the privilege is designed to encourage, consulting a lawyer for the purpose of achieving law compliance."[96]

C. Ethical Obligation to Maintain Client Confidentiality

The Model Rules of Professional Conduct [Model Rules], on which almost all states in the United States base their legal ethics rules,[97] call for information relating to the representation of a client to be held in confidence, with limited exceptions.[98] This duty of confidentiality applies to all information related to

76-79 and accompanying text.

[93] *See* Moody v. Internal Revenue Service, 654 F.2d 795, 800 (D.C. Cir. 1981); *see also* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[94] *See In re* Green Grand Jury Proceedings, 492 F.3d at 981 (The Eighth Circuit stated that "we hold, as have our sister circuits, that an attorney who is not complicit in his client's wrongdoing may assert the work product privilege with respect to his opinion work product.").

[95] *See* RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 82 cmt. e (2000); ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[96] RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 82 cmt. c (2000); *see* ABA/BNA, Lawyers Manual on Professional Conduct: Corporate Privilege/Confidentiality Practice Guides 91:2201, LMPC 91:2201 (Westlaw).

[97] With the adoption of the Model Rules format by Maine in 2009, California remains the only state whose legal ethics rules do not comport with the ABA Model Rule format. *Maine Becomes Penultimate Jurisdiction to Adopt Model Rules*, 25 Laws. Man. on Prof. Conduct (ABA/BNA) 135 (Mar. 18, 2009). However, while almost all states in the U.S. have adopted the Model Rules, lawyers are not provided with a uniform standard. Interpretational differences exist among the jurisdictions, as do differences in the text of some of the rules. *See* Louise L. Hill, *Electronic Communications and the 2002 Revisions to the Model Rules*, 16 ST. JOHN'S J. LEGAL COMMENT. 529, 531 (2002).

[98] MODEL RULES OF PROF'L CONDUCT R. 1.6(b) (2004). Pursuant to Model Rule 1.6, lawyers are permitted to:

> reveal information relating to the representation of a client to the extent the lawyer reasonably believes necessary:
>
> (1) to prevent reasonably certain death or substantial bodily harm;

the representation, whatever its source.[99] It has been noted that the Model Rules do "not put generally known information outside the boundaries of confidentiality."[100] While it remains the rule, this approach to public information has been criticized as "so stringent as to approach the unworkable and the unrealistic."[101]

The Model Rule exceptions attaching to the duty of confidentiality were significantly expanded in 2002 and 2003. As originally adopted in 1983, the Model Rules permitted lawyers to disclose information relating to the representation of a client in two instances: to prevent the client from committing a criminal act likely to result in imminent death or substantial bodily harm;[102] and to respond to allegations, or establish a claim or defense on behalf of the lawyer, in designated proceedings.[103] Added to these exceptions in 2002 were securing legal advice about compliance with the Rules,[104] and compliance with other law or a court order.[105] The exceptions were again expanded in 2003 to address financial injury when the lawyer's services had been, or were being used in its furtherance. To that end, disclosure was permitted to prevent a client from committing a crime or fraud reasonably certain to result in substantial injury to the financial interests or property of another;[106] and to prevent, mitigate or rectify substantial injury to the financial

> (2) to prevent the client from committing a crime or fraud that is reasonably certain to result in substantial injury to the financial interests or property of another and in furtherance of which the client has used or is using the lawyer's services;
> (3) to prevent, mitigate or rectify substantial injury to the financial interests or property of another that is reasonably certain to result or has resulted from the client's commission of a crime or fraud in furtherance of which the client has used the lawyer's services;
> (4) to secure legal advice about the lawyer's compliance with these Rules;
> (5) to establish a claim or defense on behalf of the lawyer in a controversy between the lawyer and the client, to establish a defense to a criminal charge or civil claim against the lawyer based upon conduct in which the client was involved, or to respond to allegations in any proceeding concerning the lawyer's representation of the client; or
> (6) to comply with other law or a court order.

Id.

[99] *See* RESTATEMENT (THIRD) OF THE LAW GOVERNING LAWYERS § 59 cmt. (b) (2000).

[100] Garwin, *supra* note 2, at 34.

[101] HAZARD & HODES, *supra* note 4, § 9.15, at 9-60 (Supp. 2005-2).

[102] MODEL RULES OF PROF'L CONDUCT R. 1.6(b)(1) (1983). This was modified in the recent Model Rule revisions to the prevention of "reasonably certain death or substantial bodily harm." *Id*. at R. 1.6(b)(1) (2003).

[103] *Id*. at R. 1.6(b)(2) (1983).

[104] MODEL RULES OF PROF'L CONDUCT R. 1.6(b)(2) (2002).

[105] *Id*. at R. 1.6(b)(4).

[106] MODEL RULES OF PROF'L CONDUCT R. 1.6(b)(2) (2003).

interests or property of another that is reasonably certain to result or which had resulted from the client's commission of a crime or fraud.[107]

The exceptions to the general prohibition against disclosure of client information in Model Rule 1.6 are permissive rather than mandatory. However, once Model Rule 1.6(b) permits a disclosure, other rules or law may require it. For instance, some ethics rules require disclosure to the extent it is permitted under Model Rule 1.6.[108] Breach of the obligation of confidentiality can subject a lawyer to professional discipline, with typical sanctions being reprimand, suspension or disbarment.[109] Occasionally, although not pursuant to the Model Rules, a client can also obtain damage recovery if a lawyer unjustifiably divulges confidential information that results in the client being harmed.[110]

III. EMERGING TECHNOLOGY AND CLIENT CONFIDENTIALITY

Over the years, lawyers have used available technology to communicate with clients. During much of the twentieth century, lawyers routinely spoke with clients on the telephone. Even though telephone company employees could eavesdrop on these land-line calls, which could also be intercepted by third parties, people had an expectation that these conversations would be private.[111] The Federal Wiretap Act reflected this expectation of privacy, prohibiting intentional interception of wire or electronic communications, and providing that interception does not waive any otherwise available privilege.[112]

A. Facsimile Transmissions

When facsimile transmission became affordable and widely used in the 1980's, it was not suggested that the mere use of a fax machine to transmit

---

[107] *Id*. at R. 1.6(b)(3).

[108] Model Rule 4.1(b) provides as follows:

> In the course of representing a client a lawyer shall not knowingly fail to disclose a material fact when disclosure is necessary to avoid assisting a criminal or fraudulent act by the client, unless disclosure is prohibited by Rule 1.6.

Id. at R. 4.1(b) (2009).

[109] *See* Wolfram, *supra* note 1, at 545.

[110] *Id*.

[111] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[112] The Federal Wiretap Act provides that "[n]o otherwise privileged wire, oral, or electronic communication intercepted in accordance with, or in violation of, the provisions of this chapter shall lose its privileged character." 18 U.S.C. § 2517(4) (2006). The Act also forbids the disclosure or use of unlawfully intercepted communications and bars the introduction into evidence of unlawfully intercepted conversations. *Id*. at § 2515.

confidential information would contravene an ethics rule.[113] In fact, most courts considering the matter found that transmission by fax did not alter the nature of protection afforded by privilege.[114] However, as with any other means of communication, lawyers were cautioned that they could not ignore their responsibility to maintain the confidentiality of client information when faxing material.[115] Noted was the fact that "careless use of a fax machine may result in inadvertent delivery of client information to the wrong person, triggering a dispute over availability of the attorney-client privilege and possible malpractice liability."[116]

B. Cordless Telephones

When cordless telephones began to be used, the expectation of privacy diminished. Using analog voice signals transmitted by radio-waves that broadcast in all directions, cordless telephone conversations could be picked up by mistake as well as intentionally monitored with relative ease.[117] Inadvertent interception occurred frequently with cordless phones, since using one was like operating a radio station, the broadcast of which could be received by anyone in range.[118] Due to this situation, Congress amended the Federal Wiretap Act in 1986 to exclude the radio portion of a cordless telephone conversation from the definition of "wire communication" and "electronic communication."[119] However, in 1994 these exceptions were removed from the statute,[120] making "legal protections afforded to cordless phone broadcasts identical to those protecting land-based calls."[121] It should be noted that some take the position that privacy is not assured on a cordless phone since federal law does not apply to mistakes, just intentional interceptions.[122]

---

[113] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[114] *Id*. "Although a misdirected fax may lose its privilege, no one argues that the use of a fax machine or the possibility of misdirection destroys any claim of privilege." David Hricik, *Confidentiality & Privilege in High-Tech Communications*, 60 TEX. B.J. 104, 110 (Feb. 1997).

[115] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[116] *Id.* at§ 55:401.

[117] *Id*.

[118] *See* Hricik, *supra* note 114, at 108.

[119] Electronic Communications Privacy Act of 1986, Pub. L. No. 99-508, § 101, 100 Stat. 1848, 1848-1849 (1986) [hereinafter ECPA] amending 18 U.S.C. §§ 2510 (1), (12).

[120] Communications Assistance for Law Enforcement Act, Pub. L. No. 103-414, § 202(a), 108 Stat. 4279 (1994).

[121] Hricik, *supra* note 114, at 108.

[122] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic

C. Cellular Telephones

The use of cellular telephones followed the cordless phone. With the advent of the cellular phone, radio signals were transmitted to a base station in a geographic area,[123] which were then transmitted via microwaves to the switch center of the cellular service provider, and transferred to local telephone service providers.[124] Capable of being intercepted by any receiver in the broadcast area capable of receiving cellular frequencies, Congress enacted legislation to make it a federal crime to intentionally intercept cellular communications.[125] In 1992, Congress also enacted legislation to prohibit the manufacture and importation of certain scanners primarily used to intercept cellular calls.[126] This notwithstanding, a "monitoring phenomenon" seemed to exist.[127] Due to this ease of monitoring, to protect their analog cellular calls, some lawyers used devices and services to provide protection against eavesdropping. Using scrambling, conversion or encryption techniques, lawyers sought to protect their calls from those who were unaware of the law, or chose to ignore it.[128]

Ethics opinions from various states considered the use of cordless or cellular phones by lawyers. To avoid a possible breach of confidentiality, many bar committees urged lawyers to use cordless or cellular phones with caution. It was suggested that lawyers warn those with whom they conversed that conversations via this technology were not secure and sensitive material should not be discussed.[129] Ethics opinions in several states indicated that communications conducted in this manner might not be considered confidential and might not be covered by the attorney-client privilege.[130] As a

---

Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw). Today, cordless telephones that use digital transmissions, or that encrypt the digital signal, "thwart casual hobbyists who eavesdrop using commercial scanners." *Id.*

[123] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw) (The geographic area is referred to as a "cell.").

[124] *Id.*

[125] *See* ECPA, *supra* note 119, amending 18 U.S.C. § 2511(1).

[126] 47 U.S.C. 302a (d); *see* Hricik, supra note 114, at 108.

[127] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[128] *Id.*

[129] *Id.*; *see* [1994] Nat'l Rep. Legal Ethics (Univ. Pub. Am.) Mass. Ethics Op. 94-5; NYCBA Comm. On Professional Ethics, Formal Op. 1994-11 (1994); WSBA Informal Op. 91-1 (1991); *see also* Hricik, *supra* note 114, at 112.

[130] *See* [1995] Nat'l Rep. Legal Ethics (Univ. Pub. Am.) N. C. Proposed RPC 215; [2003] LAWS. MAN. ON PROF. CONDUCT (ABA/BNA) Iowa Ethics Op. 90-44, 1001:3601; [1990] LAWS. MAN. ON PROF. CONDUCT (ABA/BNA) Ill. Ethics Op. 90-7, 901:3001.

result, lawyers shied away from using analog phones for client conversations.

With the advent of digital technology, concerns about interception of cellular telephone conversations diminished. Unlike analog service, digital cellular service turns voices into bits, and calls transmitted in digital format cannot be heard by simple radio frequency scanners.[131] Digital phones offered greater security than their predecessors, eliminating concern about widespread eavesdroppers. Although susceptible to interception by the sophisticated, with digital technology came an expectation of a greater degree of privacy.[132]

D. Internet Transmissions

The mid-1990's saw the emergence of the internet and e-mail as an integral form of communication, and not surprisingly, an integral part of legal practice.[133] Because of this, the legal profession has devoted considerable attention to this type of technology and its ramifications in the practice of law.[134] Although federal statutes prohibit intentional interception of e-mail,[135] lawyers disagreed about the propriety and malpractice risk of communicating confidential client information via unencrypted e-mail.[136] The result of this discourse was the recognition of a reasonable expectation of privacy in e-mail messages, and the determination that the interception of an electronic communication does not cause an otherwise privileged electronic communication to lose its privileged character.[137] This notwithstanding, many lawyers tended to avoid using e-mail for sensitive material and encrypted material, or employed some generally accepted security system when

---

[131] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[132] *Id*.

[133] *See* Joan C. Rogers, *Ethics, Malpractice Concerns Cloud E-mail, On-line Advice*, 12 LAWS. MAN. ON PROF. CONDUCT (ABA/BNA) 59 (1996).

[134] From the sender's computer, most internet e-mail goes through several routers before arriving at the intended password protected mailbox of the recipient. The routers, which are owned by various third parties, temporarily store and help distribute e-mail messages. *See* Hricik, *supra* note 114, at 113-14.

[135] As amended by the Electronic Communications Privacy Act, e-mail is protected from interception by the Federal Wiretap Act in that it is an electronic communication. ECPA, *supra*, note 119.

[136] *See* [1996] Nat'l Rep. Legal Ethics (Univ. Pub. Am.) Iowa Ethics Op. 96-1; [2003] LAWS. MAN. ON PROF. CONDUCT (ABA/BNA) S.C. Ethics Op. 94-27, 1001:7901; [1995] Nat'l Rep. Legal Ethics (Univ. Pub. Am.) N. C. Proposed RPC 215*; see also* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[137] *See* 18 U.S.C. § 2517(4) (2006).

communicating with clients.[138]

No authority exists which suggests that privilege is unavailable simply because a lawyer and client communicate via internet e-mail.[139] It has been argued that the federal statutory prohibitions against intercepting these communications render them "sufficiently private to satisfy the conditions for the attorney-client privilege to apply."[140] Also, it has been noted that "[f]ull use of all available technology to prevent interception is not required."[141] Generally, only steps that are reasonable under the circumstances are called for.

When the matter of mandatory encryption of e-mail was addressed by the American Bar Association in 1999, an ABA Committee concluded that a lawyer may communicate with a client via e-mail without encryption.[142] It reached this conclusion reasoning that the expectation of privacy for e-mail is the same as that for ordinary telephone calls, and the unauthorized interception of an electronic message is illegal. The ABA Committee noted, however, that unusual circumstances involving extraordinarily sensitive information might warrant enhanced security measures like encryption, just as ordinary telephones and other normal means of communication would be deemed inadequate to protect confidentiality in some situations.[143]

---

[138] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw). It has been noted that:

> A client with a sensitive issue to discuss is likely to be quite unhappy, and could well sue, if high-tech means of communication lead others to become aware of this discussion and if the client's lawyer failed to take adequate precautions or failed to warn the client of the potential risks. This is so even if 'privilege' as such is not lost.

Rogers, *supra* note 133, at 64 (quoting Peter Jarvis & Bradley Tellam, Electronic Ethics and Malpractice Issues, 5-5, Washington State Bar Seminar on Lawyers and the Internet (1995)).

[139] *See* Hricik, *supra* note 114, at 116.

[140] ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw).

[141] *Id*. (quoting MUELLER & L. KIRKPATRICK, MODERN EVIDENCE, § 5.13 (1995)).

[142] ABA Comm. on Ethics and Prof'l Responsibility Formal Op. 99-413 (1999).

[143] *Id.* In an ethics opinion of the Committee on Professional Ethics of the Delaware State Bar Association, it was determined that a lawyer may make communications in confidence when using e-mail or a cell phone absent extraordinary circumstances. The test proposed by the committee was whether the lawyer reasonably anticipated the possibility of interception and used the example of sharing e-mail accounts with another. To determine if an extraordinary circumstance exists, the committee suggested the lawyer determine if there is a significant risk of inadvertent disclosure, and if not, then the communication can generally be made in confidence using e-mail or a cell phone. Del. State Bar Ass'n Comm. on Prof'l Ethics, Op. 2001-2 (2001). More recently, in an ethics opinion of the Professional Ethics Commission of the Maine Board of Bar Overseers, it was determined that as a

E. Model Rule 4.4(b)

With the proliferation of electronic communications, the relative ease of transmission has resulted in an increase of inadvertent communications being disseminated. This matter is specifically addressed in Model Rule 4.4(b) and its commentary. The rule itself provides:

> A lawyer who receives a document relating to the representation of the lawyer's client and knows or reasonably should know that the document was inadvertently sent shall promptly notify the sender.[144]

Notification to the sender of an "errant" communication enables that person to take protective measures. The commentary to the rule specifically notes that additional steps to be taken by the lawyer, such as returning the document, as well as whether the privileged status of the document has been waived, are beyond the scope of the rule.[145]

IV. THE DISPUTE SURROUNDING METADATA

An issue associated with electronic communications that is receiving considerable attention relates to "metadata," which is hidden information in digital documents. As a general premise, metadata falls into categories, the first of which is data that is generated and stored in a document by the software used to create it.[146] Software generated metadata, sometimes referred to as system metadata, appears on the drafter's disk drives.[147] While it does not appear in the on-screen or printed version of a document, typically, it can be accessed relatively easily.[148] A second type of metadata, sometimes referred to as substantive metadata, is generated by the person who created the document.[149] This metadata can track the revision history of a document and can either appear in the on-screen or printed version of a document, or be

---

general matter, an attorney may utilize unencrypted e-mail without violating the lawyer's ethical obligation to maintain client confidentiality. The Commission went on to note, however, that some circumstances might require a more secure method of communication. Me. Prof'l Ethics Comm'n of the Bd. of Overseers of the Bar, Op. 194 (2007).

[144] MODEL RULES OF PROF'L CONDUCT R. 4.4(b) (2009).

[145] *Id.* at cmt. 2.

[146] *See* Martin Whittaker, *Speakers Examine Metadata Phenomenon and Explore Whether Lawyers Should Fear It*, 23 ABA/BNA Law. Manual on Professional Conduct 305 (June 13, 2007).

[147] *See* D. Md. Local R., *Suggested Protocol for Discovery of Electronically Stored Information* ("Suggested Protocol") at 25, www.mdd.uscourts.gov/localrules/localrules.html.

[148] *See* Whittaker, *supra* note 146, at 305. Often it can be found in the "file" menu under "properties." *Id.*

[149] *See id.*; *see also* Suggested Protocol, *supra* note 147.

hidden from view.[150] A third type of metadata, sometimes referred to as embedded metadata, is "inferred through a relationship to another document."[151] This metadata is data or content input by the user which is not typically visible in the output display, such as spread sheet formulas, hidden columns, linked files, database information or field codes.[152] Metadata does not appear in the final print-ready version of a final electronic document, but it can be easily accessed. It accompanies every Word document unless it is "scrubbed."[153] At issue is an electronic document, sent to a non-client, which may have confidential information available to a non-privileged viewer. Questions arise as to whether this destroys the privileged nature of the document, as well as how lawyers should deal with hidden data imbedded in documents they receive.

A. The Position of the American Bar Association

There is disagreement among the authorities regarding how lawyers should treat metadata. An ABA Formal Opinion released in 2006 indicates that a receiving lawyer is free to review and use embedded information contained in electronic documents.[154] Noting that the Model Rules do not specifically prohibit such practice, the ABA Committee found MR 4.4(b) [155] to be the most closely applicable rule, calling for the sole requirement of notice to the

---

[150] It is available through the "insert comment" and "track changes" functions of Word. *See* Whittaker, *supra* note 146, at 305-06.

[151] Williams v. Sprint/United Mgmt. Co., 230 F.R.D. 640, 647 (D. Kan. 2005).

[152] *See* Suggested Protocol, *supra* note 147, at 27.

[153] *See* Whittaker, *supra* note 146, at 305. Pennsylvania Bar Association Committee on Legal Ethics and Professional Responsibility Formal Opinion 2009-100 addresses the removal of metadata as follows:

> Corel WordPerfect Versions X3 and X4 permit a user to easily remove all or specific metadata. Microsoft Office products do not permit the easy removal of this information. Microsoft Office 2007 includes several different "Document Inspectors" that may be used to find and remove different kinds of hidden data and personal information. Some of these Inspectors are specific to individual Office programs. The Document Inspector displays different sets of Inspectors in Office Word 2007, Office Excel 2007, and Office PowerPoint 2007 to enable the user to find and remove hidden data and personal information that is specific to each of these programs. Users must be cautious, however, because there are many types of metadata and these processes may not remove all of the metadata.

Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Responsibility, Formal Op. 2009-100 n.3 (2009).

[154] ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 06-442 (2006) (discussing the review and use of metadata).

[155] MODEL RULE OF PROF'L CONDUCT R. 4.4(b) (2009).

sender[156] that the inadvertently sent information was received.[157] The ABA Committee observed that much metadata is inconsequential and that steps can be taken by the sender to limit the likelihood that metadata will be transmitted in electronic documents.[158] The ABA position, that a lawyer is free to look for hidden embedded data and use it to the advantage of the receiving lawyer's client, was contrary to the position taken previously in two New York State ethics opinions, discussed below.

B. The Position of the New York State Bar Association

In 2001, the Committee on Professional Ethics of the New York State Bar Association considered whether lawyers could use available technology to surreptitiously examine and trace electronic documents.[159] In reaching the conclusion that this would not be permissible, the Committee looked to New York's Disciplinary Rules which prohibit a lawyer from engaging in conduct "involving dishonesty, fraud, deceit or misrepresentation,"[160] and "conduct that is prejudicial to the administration of justice."[161] The Committee then reasoned:

> We believe that in light of the strong public policy in favor of preserving confidentiality as the foundation of the lawyer-client relationship, use of technology to surreptitiously obtain information that *may* be protected by the attorney-client privilege, the work product doctrine or that *may* otherwise constitute a "secret" of another lawyer's client would violate the letter and spirit of these Disciplinary Rules.[162]

Relying on this 2001 New York opinion, in 2004, the New York State Bar Association Committee on Professional Ethics again considered a matter

---

[156] ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 06-442 (2006) (the committee concluded that Rule 4.4(b)'s sole requirement of promptly notifying the sender was "evidence of the intention to set no other specific restrictions on the receiving lawyer's conduct. . .").

[157] *Id*. The committee noted, however, that:

> Whether the receiving lawyer knows or reasonably should know that opposing counsel's sending, producing, or otherwise making available an electronic document that contains metadata was "inadvertent" within the meaning of Rule 4.4(b), and is thereby obligated to provide notice of its receipt to the sender, is a subject that is outside the scope of this opinion.

Id. at 4.

[158] *Id.*

[159] N.Y. State Bar Ass'n Op. 749 (Dec. 14, 2001).

[160] N.Y. CODE OF PROF'L RESPONSIBILITY DR 1-102(A)(4) (2007).

[161] *Id*. at DR 1-102(A)(5).

[162] N.Y. State Bar Ass'n Op. 749, *supra* note 159, at 3 (emphasis added).

involving electronic documents.[163] Looking to the Disciplinary Rule that states a lawyer shall not '"knowingly' reveal a confidence or secret of a client,"[164] the Committee considered whether a lawyer who transmits documents that contain "metadata" reflecting client confidences or secrets violates this rule.[165] The New York Committee concluded that under their disciplinary rules, lawyers have a duty "to use reasonable care when transmitting documents by e-mail to prevent the disclosure of metadata containing client confidences or secrets."[166] As to what constitutes reasonable care, the Committee stated it will vary with the circumstances, including:

> the subject matter of the document, whether the document was based on a "template" used in another matter for another client, whether there have been multiple drafts of the document with comments from multiple sources, whether the client has commented on the document, and the identity of the intended recipients of the document.[167]

It was also noted that reasonable care may "call for the lawyer to stay abreast of technological advances and the potential risks in transmission in order to make appropriate decisions with respect to the mode of transmission."[168]

C. The Position of the Florida Bar Association

In September of 2006, a month following the release of the ABA opinion on metadata, the Florida Bar issued an opinion addressing the ethical duties of lawyers when sending and receiving electronic documents in the course of client representation.[169] Siding with the approach taken in New York, rather than that of the ABA, the Florida Bar determined that "a lawyer receiving an electronic document should not try to obtain information from metadata."[170] In considering this matter, the Florida Bar set forth the following obligations for lawyers when transmitting electronic documents:

> 1) It is the sending lawyer's obligation to take reasonable steps to safeguard the confidentiality of all communications sent by electronic means to other lawyers and third parties and to protect from other lawyers and third parties all confidential information, including information

---

[163] N.Y. State Bar Ass'n Op. 782 (Dec. 8, 2004).

[164] N.Y. CODE OF PROF'L RESPONSIBILITY DR 4-101(B) (2007).

[165] *See* N.Y. State Bar Ass'n Op. 782, *supra* note 163.

[166] *Id.* at 3.

[167] *Id*. at 2.

[168] *Id.* at 3.

[169] Fla. Bar Prof'l Ethics Comm., Op. 06-02 (2006), http://www.floridabar.org/tfb/tfbetopin.nsf/SearchView/ETHICS,+OPINION+06-2?opendocument.

[170] *Id*.

> contained in metadata, that may be included in such electronic communications.
>
> 2) It is the recipient lawyer's concomitant obligation, upon receiving an electronic communication or document from another lawyer, not to try to obtain from metadata information relating to the representation of the sender's client that the recipient knows or should know is not intended for the recipient. Any such metadata is to be considered by the receiving lawyer as confidential information which the sending lawyer did not intend to transmit.
>
> 3) If the recipient lawyer inadvertently obtains information from metadata the recipient knows or should know was not intended for the recipient, the lawyer must "promptly notify the sender."[171]

The Florida opinion, which did not address electronic documents in the context of discovery,[172] also noted that these obligations "may necessitate a lawyer's continuing training and education in the use of technology in transmitting and receiving electronic documents in order to protect client information."[173]

D. The Positions of Other Jurisdictions

When the legal community first began to consider how to handle metadata, two divergent points of view emerged. The ABA position, indicating a receiving lawyer is free to review and use imbedded information,[174] and the position taken by New York and Florida, indicating a receiving lawyer should not try to obtain information from metadata that the lawyer knows, or should know, was not intended for him.[175] While the committees disagreed about the receiving lawyer's responsibilities with respect to metadata, they did not disagree on the sending lawyer's responsibilities. It is the responsibility of the sending lawyer to take reasonable measures to avoid the disclosure of confidential information imbedded in electronic materials.[176]

As subsequent jurisdictions considered the metadata issue, some leaned toward the position taken by the ABA, some favored the approach taken by New York and Florida, while others employed their own variations.

---

[171] *Id.* (citation omitted).

[172] Specifically stating that it did "not address metadata in the context of documents that are subject to discovery under applicable rules of court or law," the opinion noted it did "not address the role of the lawyer acting as a conduit to produce documents in response to a discovery request." *Id.*

[173] *Id.*

[174] *See supra* notes 154-58 and accompanying text.

[175] *See supra* notes 159, 163, 169 and accompanying text.

[176] *See supra* notes 158, 166 and accompanying text.

Committees from Maryland[177] and Colorado[178] were inclined toward the ABA position, while committees from Arizona,[179] Alabama,[180] Maine[181] and New Hampshire[182] sided with the approach taken by New York[183] and Florida. The committee from the District of Columbia distinguished its approach, calling for actual knowledge of inadvertent disclosure before barring access to metadata.[184] The committee from Pennsylvania originally took a middle of the road approach, calling for lawyers who receive electronic information to use their own judgment in deciding whether to look for and use embedded information.[185] However, apparently upon reflection, the Pennsylvania Committee decided to "generally align" itself with the ABA position, "concluding that 'an attorney who receives [. . .] inadvertently transmitted information from opposing counsel may generally examine and use the metadata for the client's benefit without violating the Rules of Professional Conduct.'"[186]

1. Maryland

The Maryland State Bar Association Committee on Ethics was asked to consider whether an attorney who receives electronic documents containing metadata may view or use that metadata without first ascertaining whether the sending attorney inadvertently or intentionally included the material.[187] Viewing the matter from the perspective of electronic discovery, the Maryland Committee answered that question in the affirmative. A receiving lawyer may

[177] Md. State Bar Ass'n Comm. on Ethics, Op. 2007-09 (2007).

[178] Colo. Bar Ass'n Ethics Comm., Formal Op. 119 (2007), http://www.cobar.org/index.cfm/ID/386/subID/23789/CETH/.

[179] Ariz. State Bar Comm. on Rules of Prof'l Conduct, Ethics Op. 07-03 (2007), http://www.myazbar.org/Ethics/opinionview.cfm?id=695.

[180] Ala. Office of Gen. Counsel, Ethics Op. 2007-02 (2007), http://www.alabar.org/ogc/PDF/2007-02.pdf.

[181] Me. Prof'l Ethics Comm'n of the Bd. of Bar Overseers Op. 196 (2008), http://www.mebaroverseers.org/Ethics%20Opinions/Opinion%20196.htm.

[182] N.H. Bar Ass'n Ethics Comm. Op. 2008-2009/4 (2009).

[183] In March of 2008, a committee from the New York County Lawyers' Association Committee on Professional Ethics endorsed the position previously taken by the New York State Bar Association. N. Y. County Lawyers' Ass'n Comm. on Prof. Ethics Op. 738 (Mar. 24, 2008).

[184] D. C. Bar Ethics Op. 341 (2007).

[185] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2007-500 (2007).

[186] *Lawyers May Review and Use Metadata, Panel Advises in Second Look at Issue*, 25 Laws. Man. on Prof'l Conduct (ABA/BNA) 245 (May 13, 2009) (quoting Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Formal Op.2009-100 (2009)).

[187] Md. Ethics Op. 2007-09, *supra* note 177.

view and make use of metadata in electronic documents without first ascertaining whether the sender intended to include it.[188] With respect to a sending attorney's obligations, the Committee took the position that "the sending attorney has an ethical obligation to take reasonable measures to avoid the disclosure of confidential or work product materials," that might be embedded in documents.[189] This obligation is based primarily on Rule 1.1,[190] addressing lawyer competence, and Rule 1.6, addressing client confidentiality.[191] The Maryland Committee noted, however, that not every inadvertent disclosure of privileged or work product material would constitute a violation of Rule 1.1 and/or Rule 1.6. "[E]ach case would have to be evaluated based on the facts and circumstances applicable thereto."[192]

On December 1, 2006, amendments to the Federal Rules of Civil Procedure became effective, which created a set of rules to govern discovery of electronically stored information [ESI].[193] In response to these changes in the

---

[188] *Id.* The Maryland Committee noted that the Maryland Rules of Professional Conduct do not include Model Rule 4.4(b). *See Id.*

[189] *Id.*

[190] The Maryland Rule comports with ABA Model Rule 1.1, which provides:

> A lawyer shall provide competent representation to a client. Competent representation requires the legal knowledge, skill, thoroughness and preparation reasonably necessary for the representation.

MODEL RULES OF PROF'L CONDUCT R. 1.1 (2009); *see* MD. RULES OF PROF'L CONDUCT R. 1.1 (2005).

[191] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6(b) (2004). The Maryland Rule permits a lawyer to reveal information likely to result in substantial injury to the financial interest or property of another, and also permits revelation when the client's act is not only fraudulent, but criminal. *See* MD. RULES OF PROF'L CONDUCT R. 1.6 (2004).

[192] Md. Ethics Op. 2007-09, *supra* note 177.

[193] Fed. R. Civ. P. 16(b) was amended "to alert the court to the possible need to address the handling of discovery of electronically stored information early in the litigation if discovery is expected to occur" and "to include among the topics that may be addressed . . . any agreements that the parties reach to facilitate discovery by minimizing the risk of waiver of privilege or work-product protection." FED. R. CIV. P. 16 advisory committee's note (2006). Addressing the contents of the Scheduling Order which the judge must make, Rule 16(b) states that it may "provide for disclosure or discovery of electronically stored information" and "include any agreements the parties reach for asserting claims of privilege or of protection as trial-preparation material after information is produced." FED. R. CIV. P. 16(b)(3)(B)(iii)-(iv).

Amendments to Fed. R. Civ. P. 26(f) were made "to direct parties to discuss discovery of electronically stored information during their discovery-planning conference," hoping to "avoid later difficulties" and "make discovery more efficient." FED. R. CIV. P. 26 advisory committee's note (2006). Aware that "discovery difficulties can result from efforts to guard against waiver of privilege and work-product protection," the amendments also suggest that

rules, a joint bar-court committee in Maryland was formed, which developed a proposed protocol for use in cases which might involve ESI.[194] The purpose of

these issues be discussed. *Id.* Included in the discovery plan which the parties are instructed to make must be "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced" and "any issues about claims of privilege or of protection as trial-preparation materials, including-if the parties agree on a procedure to assert these claims after production-whether to ask the court to include their agreement in an order." FED. R. CIV. P. 26(f)(3)(C)&(D).

Fed. R.Civ. P. 34(a) was amended "to confirm that discovery of electronically stored information stands on equal footing with discovery of paper documents." FED. R. CIV. P. 34 advisory committee's note (2006). It provides in part as follows:

> (a) In General. A party may serve on any other party a request within the scope of Rule 26(b):
>
> (1) to produce and permit the requesting party or its representative to inspect, copy, test or sample the following items in the responding party's possession, custody, or control:
>
> (A) any designated documents or electronically stored information-including writings, drawings, graphs, charts, photographs, sound recordings, images and other data or data compilations-stored in any medium from which information can be obtained either directly or, if necessary, after translation by the responding party into a reasonably usable form;

FED. R. CIV. P. 34(a)(1)(A). The rule "is intended to be broad enough to cover all current types of computer-based information, and flexible enough to encompass future changes and developments." FED. R. CIV. P. 34 advisory committee's note (2006). Fed. R. Civ. P. 33(d) was amended to parallel Rule 34(a) "by recognizing the importance of electronically stored information." FED. R. CIV. P. 33 advisory committee's note (2006).

Addressing FED. R. CIV. P. 34(b), the advisory committee stated that the rule:

> permits the requesting party to designate the form or forms in which it wants electronically stored information produced . . . . In the written response to the production request that Rule 34 requires, the responding party must state the form it intends to use for producing electronically stored information if the requesting party does not specify a form or if the responding party objects to a form that the requesting party specifies . . . . The rule does not require a party to produce electronically stored information in the form in which it is ordinarily maintained, as long as it is produced in a reasonably usable form.

FED. R. CIV. P. 34 advisory committee's note (2006).

The 2006 amendments to the Federal Rules of Civil Procedure also acknowledged that "the routine alteration and deletion of information that attends ordinary use . . . may alter or destroy information, for reasons that have nothing to do with how that information might relate to litigation." FED. R. CIV. P. 37 advisory committee notes (2006). Therefore, a new rule was added which provides as follows:

> Absent exceptional circumstances, a court may not impose sanctions under these rules on a party for failing to provide electronically stored information lost as a result of the routine, good-faith operation of an electronic information system.

FED. R. CIV. P. 37 (e). *See infra* notes 213-14 and accompanying text.

[194] *See* Suggested Protocol for Discovery of Electronically Stored Information ("ESI") §

the proposed protocol is "to facilitate the just, speedy and inexpensive conduct of discovery involving ESI in civil cases, and to promote, wherever possible, the resolution of disputes regarding the discovery of ESI without court intervention."[195]

The proposed protocol states that whether or not ordered by the court, parties should conduct a conference to discuss discovery of ESI and report the results of the conference to the court.[196] Before the Fed. R. Civ. P. 26(f) Conference of Parties, counsel should discuss the exchange of information about ESI and advise their respective clients of "substantive principles governing the preservation of relevant or discoverable ESI while the lawsuit is pending,"[197] including "the extent to which Meta-Data, deleted data, or fragmented data, will be subject to litigation hold."[198] At the Conference of Parties, the scope, objections and form of production of ESI should be discussed. [199] If meta-data is to be produced, "[p]ost-production assertion, and preservation or waiver of, the attorney-client privilege, work product doctrine, and other privileges . . ." should be discussed, as well as "procedures under which ESI that contains privileged information or attorney work product should be immediately returned to the Producing Party if the ESI appears on its face to have been inadvertently produced or if there is prompt written notice of inadvertent production by the Producing Party."[200] As to the discoverability of metadata, the proposed protocol sets forth the following principles:

A. Meta-Data is part of ESI . . . .

1, *available at* www.mdd.uscourts.gov/localrules/localrules.html. The suggested protocol "is a working model that has not been adopted by the court but may be of assistance to counsel." *Id.*

[195] *Id*. at 3.

[196] *Id*. at 4.

[197] *Id*. at 8.

[198] *Id*. at 9. "[W]here Meta-Data, or data that has been deleted but not purged, is to be preserved," there should be instructions in the litigation hold notice regarding a method to preserve such data. *Id.* at 11.

[199] *Id.* at 17. Included in the discussion should be whether production will be in Native File or Static Image format. "'Native File' means ESI in the electric format of the application in which such ESI is normally created, viewed and/or modified." *Id.* at 4. "'Static Image' means a representation of ESI produced by converting a Native File into a standard image capable of being viewed and printed on standard computer systems." *Id.* Any party wanting to redact contents of a Native File for privilege should indicate that fact, but retain an original, unmodified file during the pendency of the case. *Id*. at 18. Also, the volume and cost of metadata production and review should be discussed. *Id.* at 19.

[200] *Id.* at 20. The Proposed Protocol notes that "[t]his provision is procedural and return of materials pursuant to this Protocol is without prejudice to any substantive right to assert, or oppose, waiver of any protection against disclosure." *Id.*

> B. Meta-Data may generally be viewed as either System Meta-Data, Substantive Meta-Data, or Embedded Meta-Data . . . . System Meta-Data is less likely to involve issues of work product and/or privilege.
>
> C. . . . Meta-Data, especially substantive Meta-Data, need not be routinely produced, except upon agreement of the requesting and producing litigants, or upon a showing of good cause in a motion filed by the Requesting Party . . . .
>
> D. If a Producing Party produces ESI without some or all of the Meta-Data that was contained in the ESI, the Producing Party should inform all other parties of this fact . . . .
>
> E. Embedded Meta-Data is generally discoverable and in appropriate cases . . . should be produced as a matter of course . . . .[201]

Not addressed are substantive issues related to metadata, such as a duty to preserve meta-data, its authenticity or its admissibility.[202]

2. Alabama

The Disciplinary Commission in Alabama was next to consider the matter of metadata in an ethics opinion. They raised the following questions:

> 1. Does an attorney have an affirmative duty to take reasonable precautions to ensure that confidential metadata is properly protected from inadvertent or inappropriate production via an electronic document before it is transmitted?
>
> 2. Is it unethical for an attorney to mine metadata from an electronic document he or she received from another party?[203]

The Alabama Commission gave both inquiries an affirmative response. As to the first question, the Commission based its answer on a lawyer's duty under Rule 1.6.[204]

---

[201] *Id.* at 12. Mindful of the cost that may be involved in removing metadata, the Principles also state that "upon agreement of the parties, the Court will consider entry of an order approving an agreement that a party may produce Meta-Data in Native Files upon the representation of the recipient that the recipient will neither access nor review such data." *Id.* at 27.

[202] *Id.*

[203] Ala. Ethics Op. 2007-02, *supra* note 180.

[204] *Id.* Alabama Rule 1.6(a) follows the ABA Model Rule, which provides that "[a] lawyer shall not reveal information relating to the representation of a client unless the client gives informed consent, the disclosure is impliedly authorized in order to carry out the representation or the disclosure is permitted by paragraph (b)." MODEL RULES OF PROF'L CONDUCT R. 1.6(a) (2009).

Calling for the exercise of reasonable care in taking reasonable precautions, the Commission noted that these "will, of course, vary according to the circumstances of each individual case."[205] Factors to be considered may include "steps taken by the attorney to prevent the disclosure of metadata, the nature and scope of the metadata revealed, the subject matter of the document, and the intended recipient."[206]

As to the second question, the Alabama Commission aligned its affirmative response with that of the New York position, finding "[a]bsent express authorization from a court, it is ethically impermissible for an attorney to mine metadata from an electronic document he or she inadvertently or improperly receives from another party."[207] However, the Commission distinguished situations involving electronic discovery, noting "that parties may be sanctioned for failing to provide metadata along with electronic discovery

---

[205] Ala. Ethics Op. 2007-02, *supra* note 180.

[206] *Id*. The Commission noted an attorney would need to exercise greater care when submitting documents to an opposing party than filing a pleading with a court: "[t]here is simply a much higher likelihood that an adverse party would attempt to mine metadata, than a neutral and detached court." *Id*. However, it has been noted that "[i]t is not just the opposing party with whom one shares an electronic document who can get access to a party's MS Word documents." Brian D. Zall, *Metadata: Hidden Information in Microsoft Word Documents and Its Ethical Implications*, 33 COLO. LAW. 53, 55 (2004). For instance, in the statewide electronic filing system of the Colorado State Courts, anyone with an account with the LexisNexis File & Serve service can access an original MS Word document, including metadata, when the MS Word document is uploaded to the Courts' website for conversion to PDF format. *Id*.

[207] Ala. Ethics Op. 2007-02, *supra* note 180. The Commission determined that the unauthorized mining of metadata to uncover confidential information would violate Rule 8.4, Misconduct, of the Alabama Rules of Professional Conduct, which provides:

> It is professional misconduct for a lawyer to:
>
> (a) Violate or attempt to violate the Rules of Professional Conduct, knowingly assist or induce another to do so, or do so through the acts of another;
> (b) Commit a criminal act that reflects adversely on the lawyer's honesty, trustworthiness or fitness as a lawyer in other respects;
> (c) Engage in conduct involving dishonesty, fraud, deceit or misrepresentation;
> (d) Engage in conduct that is prejudicial to the administration of justice;
> (e) State or imply an ability to influence improperly a government agency or official;
> (f) Knowingly assist a judge or judicial officer in conduct that is a violation of applicable Canons of Judicial Ethics or other law; or
> (g) Engage in any other conduct that adversely reflects on his fitness to practice law.

ALA. RULES OF PROF'L CONDUCT R. 8.4 (2008). The Model Rules of Professional Conduct are similar, but not identical, to Alabama Rule 8.4. Also, the Model Rule does not have Rule 8.4(g). *See* MODEL RULES OF PROF'L CONDUCT R. 8.4 (2009).

submissions."[208] The Commission cautioned that parties to litigation should seek direction from the court on whether to produce metadata during discovery.[209]

3. District of Columbia

When the District of Columbia Bar addressed metadata in electronic documents, it too distinguished "between electronic documents provided in discovery or pursuant to a subpoena from those electronic documents voluntarily provided by opposing counsel."[210] The D.C. Bar Legal Ethics Committee considered the metadata issue in a bifurcated fashion, analyzing the responsibilities of lawyers who send and receive electronic documents during discovery, separately from those who send and receive electronic documents outside the discovery context.

---

[208] Ala. Ethics Op. 2007-02, *supra* note 180. In support for this position, the Commission cited a case from Kansas and a case from Ohio. In the Kansas case, the defendant was ordered to disclose electronic documents in the form in which they were maintained. However, before providing the documents, the defendant scrubbed metadata from documents, allegedly to preclude the recovery of privileged and protected information. Defendant also locked data within spreadsheet cells before providing them to plaintiffs, allegedly to limit information in the spread sheets to that which was relevant to the underlying issues. Williams v. Sprint/United Mgnt Co., 230 F.R.D. 640, 646-47 (D. Kan. 2005). Regarding metadata, based on "emerging standards," the court in Kansas stated the following:

> [W]hen a party is ordered to produce electronic documents as they are maintained in the ordinary course of business, the producing party should produce the electronic documents with their metadata intact, unless that party timely objects to production of metadata, the parties agree that the metadata should not be produced, or the producing party requests a protective order. The initial burden with regard to the disclosure of metadata would therefore be placed on the party to whom the request or order to produce is directed. The burden to object to the disclosure of metadata is appropriately placed on the party ordered to produce its electronic documents as they are ordinarily maintained because that party already has access to the metadata and is in the best position to determine whether producing it is objectionable. Placing the burden on the producing party is further supported by the fact that metadata is an inherent part of an electric document, and its removal ordinarily requires an affirmative act by the producing party that alters the electronic document. (footnotes omitted)

Id. at 652. In the Ohio case, where plaintiffs sought sanctions for discovery abuse, among which was missing metadata, the court noted that in discovery, "people aren't allowed to go on a fishing expedition and at the same time they are certainly allowed to have material that may lead to relevant material." *In re* Telxon Corp. Sec. Litg., 2004 WL 3192729 *27 (N.D. Ohio 2004). Taking up on plaintiffs' suggestion that defendant withheld, or "improperly destroyed discoverable information," the court entered default judgment on liability issues against the defendant. *Id*. at *34-36.

[209] *See* Ala. Ethics Op. 2007-02, *supra* note 180.

[210] D.C. Bar Legal Ethics Comm., *supra* note 184.

Outside the context of discovery, the D.C. Bar Legal Ethics Committee sided with the generally held position on a sending lawyer's responsibilities.[211] According to the D.C. Bar Legal Ethics Committee, under Rule 1.6, a sending lawyer is obligated to take reasonable steps to maintain the confidentiality of documents, which "includes taking care to avoid providing electronic documents that inadvertently contain accessible information that is either a confidence or a secret" and "to employ reasonably available technical means to remove such metadata before sending the document."[212] However, the D.C. Bar Legal Ethics Committee took a different stance when it addressed the receiving lawyer's duty in a non-discovery context. While generally agreeing with New York and Alabama's position that Rule 8.4(c) is "implicated when a receiving lawyer wrongfully 'mines' an opponent's metadata," the D.C. Bar Legal Ethics Committee posited that "Rule 8.4 is implicated only when the receiving lawyer has an actual prior knowledge that the metadata was inadvertently provided."[213] Since the sending lawyer is obligated to avoid inadvertent production of metadata, "mere uncertainty by the receiving lawyer as to the inadvertence of the sender does not trigger an ethical obligation by the receiving lawyer to refrain from reviewing metadata."[214] Only when the receiving lawyer has actual knowledge that metadata was inadvertently sent is its review prohibited. According to the D.C. Bar Legal Ethics Committee, in this situation, "the receiving lawyer's duty of honesty requires that he refrain from reviewing the metadata until he has consulted with the sending lawyer to determine whether the metadata includes privileged or confidential information."[215]

With respect to electronic documents provided in discovery, the D.C. Bar Legal Ethics Committee noted that the Federal Rules of Civil Procedure provide steps to identify and address issues related to electronic discovery:

[P]arties are required to consult at the outset of a case about the nature of

---

[211] *Id.*

[212] *Id.* District of Columbia Rule 1.6(c)(2) permits a lawyer to reveal client confidences "to prevent the bribery or intimidation of witnesses, jurors, court officials, or other persons who are involved in proceedings before a tribunal if the lawyer reasonably believes" such acts will likely occur without revelation. Rule 1.6(h) applies the obligation of the Rule "to confidences and secrets learned prior to becoming a lawyer in the course of providing assistance to another lawyer." D.C. RULES OF PROF'L CONDUCT R. 1.6 (2004).

[213] D.C. Bar Legal Ethics Comm., *supra* note 184. District of Columbia Rule 8.4(c) follows the ABA Model Rule.

[214] D.C. Bar Legal Ethics Comm., *supra* note 184.

[215] *Id.* The opinion suggests that if the sending lawyer advises the receiving lawyer that "protected information is included in the metadata, then the receiving lawyer should comply with the instructions of the sender. The receiving lawyer may, however, reserve the right to challenge the claim of privilege and obtain an adjudication, where appropriate." *Id.*

> pertinent electronic documents in their possession and the manner in which they are maintained. This should include specific discussions as to whether a receiving lawyer wants to obtain the metadata, and if so, whether the sending party wishes to assert a claim of privilege as to some or all of the metadata.[216]

Focusing on applicable District of Columbia rules, the D.C. Bar Legal Ethics Committee noted that a lawyer shall not "obstruct another party's access to evidence or alter, destroy or assist another person to do so, if the lawyer reasonably should know that the evidence is or may be the subject of discovery or subpoena in any pending or imminent proceeding."[217] As far as the sending lawyer is concerned, "[b]ecause it is impermissible to alter electronic documents that constitute tangible evidence, the removal of metadata may, at least in some instances, be prohibited," leading to discovery sanctions and may under some circumstances constitute a crime.[218]

Looking next to the receiving lawyer, the D.C. Bar Legal Ethics Committee stated that "a receiving lawyer is generally justified in assuming that metadata was provided intentionally."[219] In fact, "when an electronic document constitutes tangible evidence, or potential tangible evidence, the receiving lawyer has an obligation competently and diligently to review, use and preserve the evidence."[220] It is only when the receiving lawyer has "actual knowledge that metadata containing protected information was inadvertently sent by the sending lawyer," that the metadata should not be reviewed "without first consulting with the sender and abiding by the sender's instructions."[221]

---

[216] *Id.* Federal Rule of Civil Procedure 26(b)(5)(B) also has a provision for "clawing back" a privileged document provided during discovery:

> If information is produced in discovery that is subject to a claim of privilege or of protection as trial-preparation material, the party making the claim may notify any party that received the information of the claim and the basis for it. After being notified, a party must promptly return, sequester, or destroy the specified information and any copies it has and may not use or disclose the information until the claim is resolved. A retrieving party may promptly present the information to the court under seal for a determination of the claim. If the receiving party disclosed the information before being notified, it must take reasonable steps to retrieve it. The producing party must preserve the information until the claim is resolved.

FED. R. CIV. P. 26(b)(5)(B).

[217] D.C. Bar Legal Ethics Comm., *supra* note 184 (citing D. C. RULES OF PROF'L CONDUCT R. 3.4(a)).

[218] *Id.* (citing D.C. RULES OF PROF'L CONDUCT R. 3.4 cmt. 4).

[219] *Id.*

[220] *Id.* Using an analogy to a fingerprint expert, the D.C. Bar Legal Ethics Committee notes that a lawyer "may consult with a computer expert to determine the means by which the metadata can be most fully revealed." *Id.*

[221] *Id.* The D.C. Bar Legal Ethics Committee notes that in such a situation "the receiving

4. Arizona

In November 2007, the Committee on the Rules of Professional Conduct of the State Bar of Arizona issued a *sua sponte* opinion on the metadata issue, "[g]iven the importance of the subject matter."[222] Identifying the relevant ethical rules as Rule 1.6(a),[223] Rule 4.4(b)[224] and Rule 8.4(a)-(d),[225] the Arizona Committee noted that when transmitting a communication, the sending lawyer "must take reasonable precautions to prevent the information from coming into the hands of unintended recipients."[226] Cautioning lawyers about the inclusion of comments on documents that are ultimately intended for transmission to opposing counsel, the Committee directed lawyers to use documents in '"clean' form and not a document that was used for another client."[227] Considering documents in litigation as a separate matter, the Arizona Committee stated that when "removing or restricting access to metadata," sending lawyers "must take care not to violate any duty of disclosure to which the lawyer or the lawyer's client is subject."[228]

When assessing the duty of a receiving lawyer, the Arizona Committee

lawyer is permitted to take protective measures to ensure that potential evidence is not destroyed and to preserve the right to challenge the claim that the information is privileged or otherwise not subject to discovery and obtain an adjudication on that point." *Id.*

[222] Ariz. State Bar Comm. on Rules of Prof'l Conduct, Ethics Op. 07-03 (2007), http://www.myazbar.org/Ethics/opinionview.cfm?id=695.

[223] ARIZ. RULES OF PROF'L CONDUCT R. 1.6(a) (2004) (following the ABA Model Rule). However, the Arizona Rule permits a lawyer to reveal the intention of a client to commit a crime and Rule 1.6(d)(5) applies only to "other law or a final order of a court or tribunal of competent jurisdiction directing the lawyer to disclose such information." *Compare id. with* MODEL RULES OF PROF'L CONDUCT R. 1.6(a) (2003).

[224] ARIZ. RULES OF PROF'L CONDUCT R. 4.4(b) (2004) (Arizona Rule 4.4(b) differs from the corresponding ABA Model Rule, in that it imposes an additional requirement on the lawyer who receives the inadvertently sent document to "preserve the status quo for a reasonable period of time in order to permit the sender to take protective measures."); *see* MODEL RULES OF PROF'L CONDUCT R. 4.4(b) (2003).

[225] ARIZ. RULES OF PROF'L CONDUCT R. 8.4 (following the ABA Model Rule). *Compare id. with* MODEL RULES OF PROF'L CONDUCT R. 8.4 (2009).

[226] Ariz. Ethics Op. 07-03, *supra* note 179 (citing ARIZ. R. PROF'L CONDUCT R. 1.6 cmt. 2). The Committee stated that what is '"reasonable' in the circumstances depends on the sensitivity of the information, the potential consequences of its inadvertent disclosure, whether further disclosure is restricted by statute, protective order, or confidentiality agreement, and any special instructions given by the client." *Id.*

[227] *Id.* The Committee further noted that sending lawyers should also "be aware that the electronic document may be received or distributed to a person who is not a lawyer and who therefore does not have the duties of a recipient lawyer with respect to such document." *Id.*

[228] *Id.* Effective January 1, 2008, the Arizona Rules of Civil Procedure include provisions relating to discovery and disclosure of ESI. *Id.*

noted that it "respectfully decline[d] to follow the ABA position" that a receiving lawyer is free to review and use embedded information contained in electronic documents.[229] Since "it may not be possible for the sending lawyer to be absolutely certain that all of the potentially harmful metadata has been 'scrubbed' from the document before it is transmitted electronically . . . the sending lawyer would be at the mercy of the recipient lawyer" should the ABA position be followed.[230] Instead, "reminded of the duty to take reasonable steps to prevent the inadvertent disclosure of confidential or privileged information . . . the recipient lawyer has a corresponding duty not to 'mine' the document for metadata that may be embedded therein."[231]

Just as the ABA looked to Rule 4.4(b) when analyzing this matter,[232] so did the Arizona Committee. However, Arizona's Rule 4.4(b) places a burden beyond mere notice to a sending lawyer that an inadvertent document was received. Under the Arizona Rule, a lawyer who receives an inadvertent document also must "preserve its status quo for a reasonable period of time in order to permit the sender to take protective measures."[233] The Committee points out, however, that it "expresses no opinion on whether any evidentiary privilege continues to exist once an inadvertent disclosure has occurred, or whether the lawyer has incurred civil liability as a result of such disclosure."[234]

5. Pennsylvania

The Pennsylvania Bar Association Committee on Legal Ethics and Professional Responsibility considered the matter of metadata in 2007, undertaking a review of the positions embraced by the bars in the various states.[235] Commenting that each of the different conclusions reached by the various states offered "a persuasive rationale," the Committee opined that it was "difficult to establish a rule applicable in all circumstances."[236] Therefore,

---

[229] *Id.; see supra* note 154 and accompanying text.

[230] Ariz. Ethics Op. 07-03, *supra* note 179.

[231] *Id*. The receiving lawyer is also cautioned not to "otherwise engage in conduct which amounts to an unjustified intrusion into the client-lawyer relationship that exists between the opposing party and his or her counsel." *Id.* However, "[a] lawyer who receives an electronic communication may attempt to discover the metadata that is embedded therein if he or she has the consent of the sender, or if such conduct is allowed by a rule, order, or procedure of a court or other applicable provision of law." *Id.*

[232] *See* MODEL RULES OF PROF'L CONDUCT 4.4(b) (2009).

[233] ARIZ. RULES OF PROF'L CONDUCT R. 4.4(b) (2004).

[234] Ariz. Ethics Op. 07-03, *supra* note 179.

[235] Due to the timing of the opinions, Arizona Ethics Op. 07-03 was not considered by the Pennsylvania Committee in 2007. *See* Pa. Bar Ass'n Comm. On Legal Ethics & Prof'l Resp. Formal Op. 2007-500 (2007).

[236] *Id*.

the Pennsylvania Committee took the position that "the final determination of how to address the inadvertent disclosure of metadata should be left to the individual attorney and his or her analysis of the applicable facts."[237]

The Pennsylvania Committee noted that there is no specific rule in Pennsylvania relating to inadvertently transmitted metadata, although the Committee considered Rules 1.6(a)[238] and 4.4(b),[239] along with selected commentary, in its analysis.[240] Noting that the "utilization of metadata by attorneys receiving electronic documents from an adverse party is an emerging problem," the Committee ultimately concluded that many factors will be involved in analyzing "the decision of how or whether a lawyer may use the information contained in metadata."[241] Included in those factors are the following:

- The judgment of the lawyer;

- The particular facts applicable to the situation;

- The lawyer's view of his or her obligations to the client under Rule of Professional Conduct 1.3, and the relevant Comments to this Rule;

- The nature of the information received;

- How and from whom the information was received;

- Attorney-client privilege and work product rules; and,

- Common sense, reciprocity and professional courtesy.[242]

---

[237] *Id.*

[238] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6(a) (2003). Pennsylvania Rule 1.6(a) follows the Model Rule language. *See* 42 PA. CONS. STAT. ANN. § 1.6(a) (2008).

[239] *See* MODEL RULES OF PROF'L CONDUCT R. 4.4(b) (2009). Pennsylvania Rule 4.4(b). follows the Model Rule language. *See* 42 PA. CONS. STAT. ANN. § 4.4(b) (2008).

[240] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2007-500 (2007).

[241] *Id.* The Committee noted that "[a]lthough a transmitting attorney has tools at his disposal that can minimize the amount of metadata contained in a document he or she is transmitting, those tools still may not remove all metadata." *Id.* For "metadata does not disappear with the click of a button." Daniel J. Siegel, *Scrub Your Documents! Removing Metadata Before E-mailing Can Help Maintain Client Confidences*, 68 THE PHILADELPHIA LAWYER 56, 57 (Fall 2005). It is suggested that lawyers should establish policies that address "under what circumstances electronic files may be sent to other counsel." *Id.* Should the transmission of electronic documents be approved, lawyers "should establish a procedure that assures that metadata is removed before a file is sent to opposing counsel or others, including the media." *Id.*

[242] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2007-500 (2007) (footnotes omitted). Pennsylvania Rule 1.3 follows Model Rule 1.3, which states that "[a] lawyer shall act with reasonable diligence and promptness in representing a client." MODEL RULES OF

Although recognizing that waiver of attorney-client privilege is a matter for judicial determination, the Pennsylvania Committee stated that "the inadvertent transmissions of such materials should not constitute a waiver of the privilege, except in the case of extreme carelessness or indifference."[243]

In 2009, The Pennsylvania Committee revisited its 2007 position on metadata, stating that its 2007 opinion "provided insufficient guidance to recipients of documents containing metadata and did not provide correlative guidance to attorneys who send such documents."[244] With respect to the sending lawyer, the Pennsylvania Committee looked to Rules 1.1[245] and 1.6,[246] and their commentary, noting that "[c]ompetence includes the knowledge and skill to secure appropriate protection for documents to ensure that information that would negatively affect the client's case is not provided to an opposing party by any means, including by inadvertently embedded metadata."[247] Recognizing that the primary burden of keeping client confidences lies with the sending lawyer, the committee reiterated that "an attorney sending electronic materials has a duty of reasonable care to remove unwanted metadata."[248]

When addressing the duties of the receiving lawyer, the Pennsylvania Committee stated that Rule 4.4(b)[249] "requires that a lawyer accessing metadata evaluate whether the extra-textual information was intended to be deleted or scrubbed from the document prior to transmittal."[250] The result of this evaluation "determines the course of action required."[251] If metadata is inadvertently sent, Rule 4.4(b) calls for the sender to be promptly notified.[252]

PROF'L CONDUCT R. 1.3 (2009).

[243] Pa. Ethics Op. 2007-500 (2007).

[244] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 n.3 (2009).

[245] *See* MODEL RULES OF PROF'L CONDUCT R. 1.1 (2009). Pennsylvania Rule 1.1 follows the ABA Model Rule. PA. CONS. STAT. ANN. § 1.1 (Comm. On Legal Ethics 2008).

[246] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6 (2003). Pennsylvania adds a Rule 1.6(d) which states that "[t]he duty not to reveal information relating to representation of a client continues after the client-lawyer relationship has terminated." 42 PA. CONS. STAT. ANN. § 1.6(d) (2008). Also, a lawyer may reveal information relating to the representation of a client that the lawyer reasonably believes necessary to "effectuate the sale of a law practice consistent with Rule 1.17." *Id.* at § 1.6(c)(6).

[247] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 n.3 (2009).

[248] *Id.*

[249] *See* MODEL RULE OF PROF'L CONDUCT R. 4.4(b) (2003); *see also supra* text accompanying note 144.

[250] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 (2009).

[251] *Id.*

[252] *Id.; see* MODEL RULE OF PROF'L CONDUCT R. 4.4(b) (2009); *see also supra* text accompanying note 144.

Focusing on the lawyer's duty to the lawyer's client, competent representation of the client under Rule 1.1 calls for the lawyer first to determine:

> whether the tribunal in which the matter is or will be proceeding may find an impropriety in the review or use of inadvertently transmitted metadata, or whether its use may unduly impact future dealings with opposing counsel, resulting in adverse consequences to the client. In such an instance, competent representation may require that the attorney refrain from disclosing or using the information. Conversely, if the inadvertently received material is beneficial to the client's case and can be viewed and/or used without adverse consequences, then Rule 1.1 may require that the attorney do so.[253]

Pursuant to Rule 1.4[254] on communication, "a lawyer has an obligation to keep the client fully appraised of important developments in the client's case so that the client may make informed decisions concerning the representation."[255] One such important event could be "potentially useful metadata."[256] Lawyers have a duty to advise their clients and respect a client's authority to control the objectives of the representation.[257] Even if the attorney judges the metadata is

---

[253] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 (2009).

[254] The Pennsylvania rule on communication comports with Model Rule 1.4 Communication, which provides that:

> (a) A lawyer shall:
>
> (1) promptly inform the client of any decision or circumstance with respect to which the client's informed consent, as defined in Rule 1.0(e), is required by these Rules;
>
> (2) reasonably consult with the client about the means by which the client's objectives are to be accomplished;
>
> (3) keep the client reasonably informed about the status of the matter;
>
> (4) promptly comply with reasonable requests for information; and
>
> (5) consult with the client about any relevant limitation on the lawyer's conduct when the lawyer knows that the client expects assistance not permitted by the Rules of Professional Conduct or other law.
>
> (b) A lawyer shall explain a matter to the extent reasonably necessary to permit the client to make informed decisions regarding the representation.

MODEL RULES OF PROF'L CONDUCT R. 1.4 (2009).

[255] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 (2009).

[256] *Id.*

[257] *Id.* The committee references a lawyer's duty "to respect the client's authority to control the objectives and means of pursuit under Rule 1.2." The Pennsylvania Rule comports with Model Rule 1.2, Scope of Representation and Allocation of Authority between Lawyer and Client, which provides in part as follows:

> (a) . . . a lawyer shall abide by a client's decision concerning the objectives of representation and, as required by Rule 1.4, shall consult with the client as to the

not useful to the client's case, "there will in most instances remain a duty to advise the client of the receipt of the metadata and the reason for nondisclosure."[258]

Continuing to focus on the duty of lawyers to their clients, the Pennsylvania Committee posited that "the lawyer's duty to the lawyer's own client trumps any theoretical responsibility to protect the right of confidentiality as between another lawyer and that lawyer's client."[259] As a general premise, a lawyer who receives inadvertently transmitted information, may "examine and use the metadata for the client's benefit without violating the [Rules of Professional Conduct]."[260] However, the receiving lawyer must determine whether the metadata can be used as a matter of substantive law; consider the potential effect on the client if the metadata is used; and consult with the client about the appropriate course of action.[261]

6. Colorado

The Ethics Committee of the Colorado Bar Association issued an opinion addressing metadata, setting forth obligations of sending and receiving lawyers who transmit electronic documents.[262] As an initial premise, the Colorado Committee asserted that "[t]he ultimate responsibility for control of metadata rests with the lawyers who send the electronic documents."[263]

Regarding Rule 1.6(a),[264] Rule 1.1,[265] and Rules 5.1 and 5.3,[266] the

> means by which they are to be pursued. A lawyer may take such action on behalf of the client as is impliedly authorized to carry out the representation.

MODEL RULES OF PROF'L CONDUCT R. 1.2(a) (2009).

[258] Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2009-100 (2009).

[259] *Id.*

[260] *Id.*

[261] *Id.*

[262] *See* Colo. Bar Ass'n Ethics Comm., Formal Op. 119 (2007).

[263] *Id.*

[264] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6(a) (2009). Generally speaking, the Colorado Rules of Professional Conduct follow the ABA Model Rules and Colorado's Rule 1.6(a) comports with the ABA Model Rule. COLO. RULES OF PROF'L CONDUCT R. 1.6 (2008). *But see infra* note 284, and accompanying text.

[265] See MODEL RULES OF PROF'L CONDUCT R. 1.1 (2009).

[266] Colorado Rules of Professional Conduct 5.1 and 5.3 follow the ABA Model Rules. Rule 5.1 requires that lawyers with managerial authority in law firms and associations "make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that all lawyers in the firm conform to the Rules of Professional Conduct." MODEL RULES OF PROF'L CONDUCT R. 5.1(a) (2003). Rule 5.3 requires that lawyers with managerial authority "make reasonable efforts to ensure that the firm has in effect measures giving reasonable assurance that [the conduct of nonlawyers employed by, retained by, or

Committee noted that:

> [A] Sending Lawyer must act competently to avoid revealing a client's Confidential Information, and to ensure that others at the Sending Lawyer's firm similarly avoid revealing a client's Confidential Information. This requires a Sending Lawyer to use reasonable care to ensure that metadata that contain Confidential Information are not disclosed to a third party.[267]

The obligations of the receiving lawyer are addressed by the Colorado Committee as two distinct issues, the initial issue being whether it is ethical for a receiving lawyer to review metadata.[268] To this inquiry the Colorado Committee gave an affirmative response. Siding with the positions taken by the ABA,[269] Maryland[270] and the District of Columbia,[271] rather than those of New York,[272] Arizona,[273] Alabama[274] and Florida,[275] the Colorado Committee concluded that generally, a receiving lawyer "may ethically search for and review metadata embedded in an electronic document that the Receiving Lawyer receives from opposing counsel or other third party."[276] The Colorado Committee arrived at this decision for three primary reasons. First, it opined that "there is nothing inherently deceitful or surreptitious about searching for metadata," so "[r]eferring to searching for metadata as 'mining' or

associated with a lawyer] is compatible with the professional obligations of the lawyer." MODEL RULES OF PROF'L CONDUCT R. 5.3(a) (2003). Pursuant to these rules, the Colorado Committee notes that "[a] supervising lawyer has a duty to make reasonable efforts to make sure that the lawyer's firm has appropriate technology and systems in place so that subordinate lawyers and nonlawyer assistants can control transmission of metadata." Colo. Ethics Op. 119, *supra* note 178.

[267] Colo. Bar Ass'n Ethics Comm., Formal Op. 119 (2007). What would constitute "reasonable care will depend on the facts and circumstances." *Id*. However, "[t]he duty to provide competent representation requires a lawyer to ensure that he or she is reasonably informed about the types of metadata that may be included in an electronic document or file and the steps that can be taken to remove metadata . . . ." *Id*.

[268] *Id*.

[269] *See* ABA Comm. on Ethics and Prof'l Responsibility Formal Op. 06-442 (2006) (Review and Use of Metadata).

[270] *See* Comm. on Ethics of Md. State Bar Ass'n Op. 2007-09 (2007).

[271] *See* D. C. Bar Ethics Op. 341 (2007).

[272] *See* N. Y. State Bar Ass'n Op. 749 (Dec. 14, 2001); N. Y. State Bar Ass'n Op. 782 (Dec. 8, 2004).

[273] *See* Ariz. State Bar Comm. on Rules of Prof'l Conduct Op. 07-03 (2007).

[274] *See* Ala. Office of Gen. Counsel, Ethics Op. 2007-02 (2007).

[275] *See* Fla. Bar Op. 06-02 (Sept. 15, 2006).

[276] Colo. Bar Ass'n Ethics Comm., Formal Op. 119 (2007).

'surreptitiously get[ting] behind' a document is, therefore, misleading."[277] Second, in many cases there is no confidential information in metadata. Third, "metadata [is] often of no import."[278]

The second issue the Colorado Committee considered was the appropriate response for a lawyer receiving metadata that appears to contain confidential information. In such a situation, the Colorado Committee indicated that the receiving lawyer "should assume that the Confidential Information was transmitted inadvertently."[279] The Colorado Committee stated that the receiving lawyer "must promptly notify the Sending Lawyer," and the lawyers may "discuss whether a waiver of privilege or confidentiality has occurred."[280] However, the "Receiving Lawyer's only duty upon viewing confidential metadata is to notify the Sending Lawyer. There is no rule that prohibits the Receiving Lawyer from continuing to review the electronic document or file and its associated metadata."[281] In contrast, in situations "where the Receiving Lawyer has prior notice from the sender of the inadvertent transmission of confidential metadata," the lawyer is prohibited from reviewing the material.[282] Rule 4.4(c) of the Colorado Rules is controlling, and provides:

> Unless otherwise permitted by court order, a lawyer who receives a document relating to the representation of the lawyer's client and who, before reviewing the document, receives notice from the sender that the document was inadvertently sent, shall not examine the document and shall abide by the sender's instructions as to its disposition.[283]

There is no comparable Model Rule to this provision.

---

[277] *Id.*

[278] *Id.*

[279] *Id.* This is the case, "unless the Receiving Lawyer knows that confidentiality has been waived." *Id.*

[280] *Id.* "If this is not possible, then the Sending Lawyer or the Receiving Lawyer may seek a determination from a court or other tribunal as to the proper disposition of the electronic documents or files, based on the substantive law or waiver." *Id.*

[281] *Id.* The Colorado Committee disagrees with the approach taken by the District of Columbia Committee, that a receiving lawyer must stop reviewing an electronic document when the receiving lawyer has actual knowledge that the sending lawyer did not intend to disclose confidential information. *Id.* Nor does the Colorado Committee agree with the position taken by the California Supreme Court in *Rico v. Mitsubishi Motors Corp.*, 42 Cal. 4th 807 (Cal. 2007), that a receiving lawyer must stop reviewing material when it becomes "reasonably apparent" that the disclosure of confidential information was not intended. Colo. Bar Ass'n Ethics Comm., Formal Op. 119 (2007).

[282] *Id.*

[283] COLO. RULES OF PROF'L CONDUCT R. 4.4(c) (2008).

7. Maine

The professional Ethics Commission of the Maine Board of Overseers of the Bar was asked by Bar Counsel to give an opinion concerning "the ethical duties of lawyers involving the transmission, retrieval and use of metadata embedded in documents which may reveal client confidences or other legally privileged information."[284] To that end the Maine Commission considered the duties of receiving and sending lawyers separately, concluding as follows:

> 1. Without authorization from a court, it is ethically impermissible for an attorney to seek to uncover metadata, embedded in an electronic document received from counsel for another party, in an effort to detect confidential information that should be reasonably known not to have been intentionally communicated.
>
> 2. A sending attorney has an ethical duty to use reasonable care when transmitting an electronic document to prevent the disclosure of metadata containing confidential information.[285]

With respect to the receiving lawyer, the Maine Commission sided with the New York position and characterized "purposefully seeking" to uncover confidential information of another party as "dishonest," striking "at the foundational principles that protect attorney-client confidences," which "prejudices the administration of justice."[286] With respect to the sending lawyer, the Commission followed "the consensus approach on the subject," calling for "reasonable measures" to be taken "to avoid the communication of confidential information, regardless of the mode of transmission."[287]

Addressing the scope of reasonable measures the sending lawyer should take, the Commission did not find it reasonable that an attorney should be "ignorant of the standard features and capabilities of word processing and other software used by that attorney, including their reasonably known capacity for transmitting certain types of data that may be confidential."[288] In fact, "a basic understanding of the existence of metadata embedded in electronic documents, the features of the software used by the attorney to generate the document and practical measures that may be taken to purge documents of sensitive metadata where appropriate to prevent the disclosure of confidential information" is called for in undertaking a lawyer's duty.[289]

---

[284] Me. Prof'l Ethics Comm'n of the Bd. of Bar Overseers Op. 196 (2008).

[285] *Id.* (footnotes omitted).

[286] *Id.*

[287] *Id.*

[288] *Id.* This, however, would not dictate the retention of a computer expert in routine work. *Id.*

[289] *Id.*

8. New Hampshire

Most recently, the Ethics Committee of the New Hampshire Bar Association considered the duties of lawyers with respect to metadata, outside the context of litigation.[290] The Committee determined that both sending lawyers and receiving lawyers "share ethical obligations to preserve confidential information relating to the representation of clients."[291] With respect to sending lawyers, there is a "duty to use reasonable care to guard against disclosure of metadata that might contain confidential information."[292] Looking to Rules 1.1,[293] 5.1 and 5.3[294] the Committee asserted that "lawyers should be reasonably informed about the types of metadata that may be included in documents when they are transmitted electronically and the steps that can be taken to remove it."[295]

With respect to lawyers who receive metadata from opposing counsel, the New Hampshire Committee determined that they "have an ethical obligation not to search for, review or use metadata containing confidential information that is associated with transmission of electronic materials from opposing counsel."[296] Any confidential information contained in electronic material is inadvertently sent, triggering Rule 4.4(b) obligations. New Hampshire Rule

---

[290] N.H. Bar Ass'n Ethics Comm. Op. 2008-2009/4 (2009).

[291] *Id.*

[292] *Id.*

[293] *See* MODEL RULES OF PROF'L CONDUCT R. 1.1 (2009). New Hampshire Rule 1.1 defines competence in detail, providing a list of requirements a lawyer must follow to achieve "legal competence." N.H. RULES OF PROF'L CONDUCT R. 1.1 (2008).

[294] *See* MODEL RULES OF PROF'L CONDUCT R. 5.1(a) (2009). New Hampshire Rules 5.1 & 5.3 impose a duty on "each" lawyer with managerial authority to emphasize that this is an obligation of all managers which cannot be delegated to one manager. *See* N.H. RULES OF PROF'L CONDUCT R. 5.1 & 5.3 (2008).

[295] N.H. Bar Ass'n Ethics Comm. Op. 2008-2009/4 (2009). The New Hampshire Committee noted the following:

> [A]s a result of rapid technological advances, some lawyers are generally unaware of the myriad of ways that client confidences may be disclosed in the form of metadata that accompanies electronic documents and files. However, unless lawyers obtain a reasonable understanding of the risks inherent in the use of technology in transmitting and receiving electronic materials that may contain confidential information, they risk violating their ethical obligations to clients. Of course, this does not mean that lawyers must necessarily purchase expensive computer software to ensure that metadata is removed or "scrubbed" from documents in all cases. In most circumstances, lawyers can limit the likelihood of transmitting metadata containing confidential information by avoiding its creation during document drafting or subsequently deleting it, as well as by sending a different version of the document without the embedded information through hard copy, scanned or faxed versions.

Id.

[296] *Id.* at 1.

4.4(b), Respect for Rights of Third Persons, varies from the Model Rule, and provides as follows:

> A lawyer who receives materials relating to the representation of the lawyer's client and knows that the material was inadvertently sent shall promptly notify the sender and shall not examine the materials. The receiving lawyer shall abide by the sender's instructions or seek determination by a tribunal.[297]

Regarding metadata, the New Hampshire Committee posits that "all circumstances, with the exception of express waiver and mutual agreement on review of metadata, lead to a necessary conclusion that metadata is 'inadvertently sent.'"[298] The New Hampshire Committee seems to champion a shared responsibility on both the sending and receiving lawyer to protect the attorney-client privilege. With respect to a receiving lawyer, "unless receiving lawyers have a sound basis to believe that the information was intentionally sent or there has been an express waiver of confidentiality, receiving lawyers should not take steps to review or to use metadata embedded in documents received from opposing counsel."[299]

V. PROPOSED TREATMENT OF METADATA

The last fifteen years have seen the proliferation of electronic communications within the practice of law. Adversaries exchange electronic documents on a routine basis and within the context of civil litigation, electronic discovery is commonplace. Because of the ease of electronic transmission and the volume of material being exchanged, it has not been unusual for a document, or material embedded in a document, to be inadvertently transmitted. What impact this has, along with the concomitant duties and responsibilities it brings to legal practitioners, is a matter of significant concern.

A. Responsibilities of Sending Lawyers

As jurisdictions consider the issues surrounding metadata, the tendency has been to distinguish between transmissions that are subject to discovery in litigation and those which are not. This is primarily because within the context of litigation, rules and procedures may require that certain metadata be produced, and failure to do so could subject lawyers to some type of sanction

---

[297] N.H. RULES OF PROF'L CONDUCT R. 4.4(b) (2008). New Hampshire's Rule 4.4(b) was amended in 2008 "to provide guidance to lawyers who receive confidential information from opposing counsel or third persons." N.H. Bar Ethics Op. 2008-2009, *supra* note 182, at 4.

[298] N.H. Bar Ass'n Ethics Comm. Op. 2008-2009/5 (2009).

[299] *Id.* at 6.

or censure. However, with respect to material that could be subject to privilege, there should be no distinction between the responsibility of a sending lawyer, whether during litigation or otherwise. Across the board, whether outside or within the context of discovery, a sending lawyer has a duty to use reasonable care when transmitting documents to prevent the disclosure of metadata containing information which could be subject to privilege. Not surprisingly, what constitutes reasonable care will vary with the circumstances. One factor that has been considered is whether the lawyer has stayed abreast of technological advances regarding the transmission of electronic information.[300]

1. Outside the Discovery Context

Both outside and within the context of discovery, the sending lawyer has a duty to use reasonable care to see that no material which could be subject to privilege is included in documents that are transmitted to a third party. However, ethical mandates indicate that outside of litigation, information relating to the representation of a client would also be included in this prohibition.[301] The Model Rules call for a lawyer to provide competent representation to a client,[302] and with limited exceptions, not to reveal information relating to the client's representation.[303] While described by some as "stringent," "unworkable and unrealistic,"[304] a lawyer's ethical duty to maintain client confidentiality is very broad. It has been posited that *no* imbedded information should accompany documents sent to anyone outside one's firm.[305] Couching commercial scrubbers as "cheap and effective," many feel they "should be considered essential equipment for fulfilling lawyers' duties of competence and care."[306]

It may be that a lawyer intends to include embedded information when transmitting an electronic document to a third party. Perhaps embedded data is included for a third party's review or perhaps costs associated with conversion of particular files are significant, and since most metadata is harmless, a decision is made to send a file in its native format.[307] In such situations, conscious decisions to include metadata are involved. If sending lawyers do

[300] *See supra* notes 168, 173, 268, 289 and accompanying text.

[301] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6(b) (2009).

[302] *See* FED. R. CIV. P 26(b)(3)(A)&(B).

[303] *See* MODEL RULES OF PROF'L CONDUCT R. 1.6(b) (2009).

[304] *See supra* note 4; *supra* note 101 and accompanying text.

[305] Whittaker, *supra* note 146, at 307 (emphasis added).

[306] *Id.* The effectiveness of commercial scrubbers is a matter on which different views are held. *See* Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Resp. Op. 2007-500 (2007); Siegel, *supra* note 242. Also, a forensic technologist can often retrieve information that has been scrubbed.

[307] *See, e.g.*, Suggested Protocol, *supra* note 147.

not intend to include metadata, it should be blocked or removed. Lawyers need to be mindful that it is inappropriate, if not dangerous, for a lawyer unintentionally to transmit information related to the representation of a client, especially that which could be considered sensitive. Lawyers must be aware of what is in the documents they transmit and how, or whether, embedded data can be accessed. Transmitting electronic documents to third parties that contain embedded information relating to the representation of a client could constitute a breach of a lawyer's ethical duty. Furthermore, in addition to being an ethical breach, lawyers might subject themselves to malpractice liability. "An attorney's failure to use the skill and knowledge ordinarily used by attorneys for communicating with or about a client could conceivably result in malpractice liability if the breach of duty proximately causes injury to the client."[308]

While a lawyer's ethical obligation to maintain the confidentiality of client information is clear, it is recognized that information may be mistakenly or inadvertently sent.[309] This can be the case even when reasonable care is used. To fulfill the lawyer's responsibility to exercise care to guard against such disclosure, lawyers should establish procedures to analyze, and where appropriate, cleanse, documents before sending files to a third party. Furthermore, should a sending lawyer determine that material was sent that should not have been, he or she should immediately notify the recipient of this fact and ask that remedial steps be taken. Such steps could include the immediate return of the information or its destruction.

2. Within the Context of Discovery

Both within and outside the context of discovery, a sending lawyer has the responsibility not to transmit information which could be subject to a claim of privilege. However, particularly within the context of discovery, a systematic removal of metadata may be both inappropriate and dangerous. Before removing metadata from a document that might be subject to discovery, sending lawyers must take care not to violate any duty of disclosure to which the lawyers or their clients are subject. The Model Rules specifically carve an exception to the lawyer's duty of confidentiality for compliance "with other law or a court order."[310]

Under mandates in new federal rules,[311] as well as under various state

[308] ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw) (citing R. MALLEN & J. SMITH, LEGAL MALPRACTICE, § 8.12, 18.2 (4th ed. 1996)).

[309] *See, e.g.*, MODEL RULES OF PROF'L CONDUCT R. 4.4 cmt. 2 (2009).

[310] MODEL RULES OF PROF'L CONDUCT R. 1.6(b)(6) (2009).

[311] *See* FED. R. CIV. P. 16(b)(3)(iii), (iv); FED. R. CIV. P. 26(f)(C), (D); FED. R. CIV. P. 34(a)(1)(A); FED. R. CIV. P. 37(e).

provisions,[312] mechanisms are outlined by which prior to document production, the parties discuss the production of electronic documents, including metadata, and the assertion of any claims of privilege, and challenges thereto. Anticipating that information may be mistakenly or inadvertently sent, included in such discussions may also be procedures for asserting any such claims after information has been provided. For those matters on which counsel cannot agree, adjudication typically is available. Removing, or failing to preserve, metadata prior to the implementation of any outlined procedures could lead to sanctions being imposed on counsel and amount to a breach of a duty. Before removing or blocking embedded information in documents that might be subject to discovery, counsel should obtain direction from the court, or mutually work out how to proceed. Metadata which is determined to be confidential, or trial preparation material, may be protected. However, material that would constitute tangible evidence must be produced. While fishing expeditions are not allowed in discovery, access to "material that may lead to relevant material" is allowed.[313]

B. Responsibilities of Receiving Lawyers

It is the responsibility of the sending lawyer to take reasonable measures to avoid the disclosure of information which could be subject to a claim of privilege or protection, and outside the context of discovery, information relating to client representation. This notwithstanding, even with the employment of measures that are reasonable, material containing this information can be mistakenly or inadvertently sent. Given the volume and incidence of the exchange of electronic documents in law practice today, such an instance is contemplated in the Model Rules, which call for the receiving lawyer to notify the sender of inadvertently sent documents.[314] While jurisdictions are in general accord as to the sending lawyers' ethical obligations in this regard, disparate views are taken on the receiving lawyers' responsibilities.

Some jurisdictions take the position that a receiving lawyer is free to review and use embedded information which is transmitted in an electronic document.[315] Especially within the context of discovery, it has been asserted that not only is this permissible, but a receiving lawyer has an obligation to

---

[312] *See, e.g.*, Suggested Protocol *supra* note 147; Ariz. Ethics Op. 07-03, *supra* note 179.

[313] *In re* Telxon Corp. Sec. Litg., 2004 WL 3192729 *27 (N.D. Ohio 2004).

[314] MODEL RULE OF PROF. CONDUCT R. 4.4(b) (2009).

[315] *See* ABA Formal Op. 06-442, *supra* note 174 and accompanying text; MODEL RULE OF PROF. CONDUCT R. 4.4(b) (2009); Md. Ethics Op. 2007-09, *supra* note 188 and accompanying text; Pa. Ethics Op. 2009-100, *supra* note 261 and accompanying text; Whittaker, *supra* note 261 at 305 and accompanying text; Colo. Ethics Op. 119, *supra* note 277 and accompanying text.

competently and diligently review this material.[316] Perhaps this follows from the premise that because the presence of embedded data in documents is well known,[317] sending a document with metadata raises a presumption that it was intentional. Or perhaps this follows from the premise that given the duty of sending lawyers to remove unintended embedded information, metadata which is sent should be presumed to be intentional. Other jurisdictions, conversely, take the position that a receiving lawyer should not try to obtain information from metadata.[318] To these divergent points of view are variations, one of which calls for there to be actual knowledge of inadvertent disclosure for review to be precluded.[319]

1. Within the Context of Discovery

Within the context of discovery, as with the sending lawyer, procedural or evidentiary rules suggest mechanisms which help chart the receiving lawyer's responsibilities with respect to embedded information in documents that are subject to discovery. Either by agreement of the parties, or court order, the receiving lawyer's access to embedded information should often be pre-determined. It may be that metadata which is sent is tangible evidence which the receiving lawyer is free, if not obligated, to carefully review. Then again, because of the cost involved in the conversion of some files, a party may send documents in Native format with metadata intact, when such information is deemed to be inconsequential or irrelevant. And yet again, because of the cost involved in the conversion of some files, the parties may agree that files will be sent in Native format with metadata intact, which the receiving lawyer will agree not to access.[320]

In those situations where no rule, protocol or agreement exists for the handling of electronic documents, direction should be sought from the court. If direction is not forthcoming from the court, it seems reasonable to infer that information, including metadata which is produced in discovery, should be presumed to have been intentionally provided. Therefore, a lawyer who receives a document that contains metadata should be free to view and use this

[316] *See* D.C. Ethics Op. 341, *supra* note 221 and accompanying text.

[317] *See* Hricik, *supra* note 114. There are many types of metadata, some of which may not be removed by conventional means. *See* Pa. Bar Ass'n Comm. on Legal Ethics & Prof'l Responsibility, Formal Op. 2009-100 (2009). However, lawyers generally are aware of the existence of metadata and the problems associated with it. *See* Hricik, *supra* note 114.

[318] *See* N.Y. State Bar Ass'n Op. 749, *supra* note 162 and accompanying text; Ala. Ethics Op. 2007-02, *supra* note 207 and accompanying text; Ariz. Ethics Op. 07-03, *supra* note 232 and accompanying text; Me. Ethics Op. 196, *supra* note 287 and accompanying text; N.H. Ethics Op. 2008-2009/4, *supra* note 297 and accompanying text.

[319] *See* D.C. Ethics Op. 341, *supra* notes 214-15 and accompanying text.

[320] *See* Suggested Protocol, *supra* note 147 §§ 8(A) & 11.

information. However, this presumption of intentional submission is rebuttable. If a sending lawyer notifies opposing counsel that protected information was inadvertently sent, it should not be examined by the receiving lawyer. Also, if upon reviewing metadata, the receiving lawyers know, or should know, that the metadata was not intended for them, review should stop, the material should be treated as protected information which was not meant to be sent and the sending lawyer should be notified. Requiring receiving lawyers to comply with a standard of "actual knowledge," rather than one of "reasonably knows or should know," is inappropriate. Such a standard works against the confidentiality doctrine itself. As to what the receiving lawyer should do next, the information at issue should be returned or sequestered, until resolution of the issue by the means employed in the jurisdiction, or by the means decided upon in that particular litigation.

2. Outside the Discovery Context

A sending lawyer is obligated to take reasonable precautions to avoid the inadvertent transmission of documents, including metadata. Because lawyers should be familiar with this duty,[321] it seems reasonable to presume that most metadata which is transmitted in a document is done so intentionally by the sending lawyer. Most embedded data is inconsequential.[322] Thus lawyers who receive documents with embedded information should be free to review it, if they so desire. However, if the receiving lawyer obtains notice from the sending lawyer that metadata was inadvertently provided, the information should not be reviewed by the receiving lawyer. Additionally, if upon review of the metadata, the receiving lawyers know, or should know, that the metadata was not intended for them, it should be treated as protected information which was not meant to be sent, and the receiver should notify the sender. As with the situation where discovery is ongoing, calling for receiving lawyers to comply with a standard of "actual knowledge," rather than one of "reasonably knows or should know," is inappropriate.

It has been noted that no lawyer intentionally transmits confidential information to a third party, so any confidential information included in metadata to an adversary would be inadvertent.[323] It has also been noted that

[321] A 2004 study revealed that 43% of respondents were aware of the existence of embedded data. Warnings about embedded data have been released since 2006. *See* David Hricik, *Mining for Embedded Data: Is it Ethical to Take Intentional Advantage of Other People's Failures?*, 8 N.C. J. L. & TECH. 231, 246 (2007).

[322] *See* ABA Comm. on Ethics and Prof'l Responsibility Formal Op. 06-442 (2006) (Review and Use of Metadata). It has been noted that while most metadata is harmless, some of it can be useful. Elizabeth W. King, *The Ethics of Mining for Metadata Outside of Formal Discovery*, 113 PENN STATE L. REV. 801, 807 (2009).

[323] *See* Hricik, *supra* note 321, at 246-47; N.H. Bar Ass'n Ethics Comm. Op. 2008-

since a receiving lawyer's decision to review metadata is an intentional act, and a sending lawyer's inclusion of confidential information an inadvertent one, a receiving lawyer's search for metadata would be a dishonest act, taking advantage of a sending lawyer's mistake.[324] This proposition, while interesting, would be better grounded if most of the information contained in metadata were confidential information or sensitive. The converse is true; most metadata does not fall into this category.[325] The review of metadata by a receiving lawyer should not be considered a dishonest act. Furthermore, permitting the review of metadata does not put confidentiality at risk.[326] Embedded information which is confidential is afforded protection since review is precluded once the lawyer knows, or should know, its character.

One suggested approach to the metadata issue, couched as "a proactive stance," is for the receiving lawyer to reserve the right to its review.[327] Just as lawyers use disclaimers related to legal advice, client representation, and the like, assertions related to embedded information could be used. Lawyers may want to represent that they "reserve the right to use whatever readily available tools and techniques are available to examine any and all documents" that are transmitted.[328]

C. Inadvertent Disclosure as Waiver of Attorney-Client Privilege

Various jurisdictions have examined the transmission of metadata in electronic documents with an eye toward the obligations that attach to lawyers who send and receive these communications. However, in addition to

2009/4 (2009).

[324] *See* Hricik, *supra* note 321, at 241, 247. It has also been asserted that "searching for metadata is unethical because it is an intentional intrusion into the attorney-client relationship and constitutes conduct that is dishonest and prejudicial to the administration of justice." King, *supra* note 322, at 828.

[325] The following are examples of metadata categories: author's name; author's initials; author's company or organization name; name of network server or hard disc where author saved document; other file properties and summary information; non-visible portions of OLE objects; names of previous document authors; document revisions; document versions; template information; hidden text; comments to documents; and time spent editing documents. *See* Zall, *supra* note 206, at 54; David Hricik & Robert Jueneman, *The Transmission and Receipt of Invisible Confidential Information*, 15 PROF. LAW. 18 (2004-05). Often metadata simply acts as a bookmark and directs a reader where to look, similar to a "post-it" on a paper document.

[326] Some feel that "[a] rule allowing receiving attorneys to search for metadata wrongfully favors the duty of diligence over the duty of confidentiality." King, *supra* note 322, at 833.

[327] Hricik & Jueneman, *supra* note 325, at 20.

[328] *Id.*

attending to the ethical obligations of lawyers is the poignant question of whether the transmission of metadata, containing confidential or trial preparation material, results in the waiver of any protection or privilege that might attach. The trend has been toward a resolution of non-waiver.

When considering whether inadvertent disclosure waives attorney-client privilege, most jurisdictions apply a balancing test.[329] Typically, balancing factors relating to the care taken by the client and sending lawyer both before and after transmission, as well as fairness to the recipient, a judicial determination is made. Attempting to resolve resulting conflicting decisions, new Federal Rule of Evidence 502 comes down on the side of non-waiver when disclosure is inadvertent and the sending lawyer acted with reasonable care.[330] The scope of waiver is also narrow under Rule 502. If it is determined that the privilege is waived for an inadvertently sent document, Rule 502 limits the extent of the waiver to the actual material disclosed, in lieu of extending the waiver to other material on the covered subject.[331] This tendency toward non-waiver should be reflected in the treatment of metadata.

When metadata is sent in a document, it should be presumed that it was done so intentionally. This follows from our understanding of the sending lawyer's responsibility to use reasonable care to see that unintended metadata is not transmitted. However, this presumption of intent is not absolute. As has been noted, no lawyer intentionally transmits confidential information to a third party, so if material that is subject to privilege or protection is sent, the presumption is rebutted and the transmission is considered to be an inadvertent one.[332] For the purpose of evaluating waiver of attorney-client privilege, when metadata which is confidential is included in a document, it should constitute an inadvertent disclosure, even if the electronically transmitted document which included the imbedded information was intentionally sent.

As with other technologies used for communication in the practice of law, lawyers cannot ignore their responsibility to maintain the confidentiality of client information when transmitting documents electronically. However, as with facsimile transmissions, e-mail, and the like,[333] there is no indication that

---

[329] *See* Walkowiak, Lemons & Leach, *supra* note 38, at 319, 321 and accompanying text; Mackintosh & Angus, *supra* note 8, at 45 n.87.

[330] *See* Lindeman, *supra* note 54, at 646; Lindeman, *supra* note 55, at 496 and accompanying text.

[331] *See* Mackintosh & Angus, *supra* note 8, at 43 n.58; Pub. L. No. 110-322, §1(b), 122 Stat. 3537 (2008); s*upra* note 59 and accompanying text.

[332] *See* Hricik, *supra* note 321, at 246-47; N.H. Bar Ass'n Ethics Comm. Op. 2008-2009/4 (2009); *supra* note 324 and accompanying text.

[333] *See* ABA/BNA Lawyers' Manual on Professional Conduct Electronic Communications Practice Guide, 55:401, LMPC 55:401 (Westlaw); David Hricik, *Confidentiality & Privilege in High-Tech Communications*, 60 TEX. B.J. 104, 110 (Feb.

a lack of expectation of privacy should be associated with the mere use of electronic document transmission. As a profession, we contemplate, if not expect, some documents to be mistakenly or inadvertently transmitted, particularly electronic ones. Ethics opinions, rules and protocols address these inadvertent transmissions of documents, providing procedures for lawyers to follow as remedial steps.[334] When evaluating whether inadvertent disclosure waives a privilege or protection, the profession has focused on the reasonable care that was used both before and after transmission. Lawyers must be aware of the perils associated with electronic transmission of documents and have a mechanism in place to guard against transmitting information unintentionally. Depending on the situation, to comply with the lawyer's duty to use reasonable care, this may necessitate the retention of a computer expert, especially in particularly sensitive situations. Just as unusual circumstances involving extraordinarily sensitive information might warrant enhanced security measures with e-mail,[335] circumstances may call for the retention of technological specialists with the transmission of certain electronic documents.

While the standard of reasonable care has been embraced when evaluating waiver of privilege with inadvertent communications, perhaps it is time for the profession to explore a more liberal approach. Continuing with the current tendency toward non-waiver, and emphasizing the duty owed to one's client, perhaps a standard similar to that used in the subjective intent test should be employed.[336] Concomitant with this, it should not go unnoticed that the policy behind privilege, at common law, was grounded on subjective considerations.[337] Since the privilege is for the welfare of the client, it has been noted that "more than the attorney's negligence should be required before the client loses the privilege."[338] Opponents to this point of view opine that this approach does little to encourage care of privileged documents.[339] However, this is not necessarily the case. Although privilege may not be lost when an intent standard is employed, the lawyer would still have an ethical duty to protect client information. Should this duty be breached, the lawyer would be subject to discipline, or liability, in certain circumstances. Such an

1997) and text accompanying note 114.

[334] *See, e.g.*, *supra* notes 144, 200, 215, & 217; MODEL RULES OF PROF'L CONDUCT R. 4.4(b) (2009); Suggested Protocol, *supra* note 147, §. 8(D); D.C. Ethics Op. 341, *supra* note 184; FED. R. CIV. P. 26(B)(5)(B), *supra* note 216 and accompanying text.

[335] *See* discussion *supra* note 143 and accompanying text.

[336] *See* Walowiak et al., *supra* note 38; *supra* text accompanying note 42; *see also supra* note 190.

[337] *See supra* note 20.

[338] Walkowiak, Lemons & Leach, *supra* note 38, at 318 (quoting Mendenhall v. Barber-Greene Co., 531 F. Supp. 951, 955 (N.D. Ill. 1982)).

[339] *See* Walowiak et al., *supra* note 38; *supra* text accompanying note 47.

approach would work to hold the lawyer in check, while protecting privilege for the client.

IV. CONCLUSION

It is aptly stated that "[t]echnology is wonderful, but failing to understand it can lead to disastrous results."[340] Sending lawyers, both within and outside the context of discovery, must guard against transmitting information that is protected. Be it within or outside discovery, there is no distinction between the duty owed a client as it relates to information which could be subject to privilege. There is a distinction, however, with how counsel should proceed in furthering this mandate, depending on the setting. Outside of litigation, the sending lawyer also has a duty to see that metadata relating to the representation of a client is not available to a third party, unless there is an intent to transmit this information. Accomplishing this typically will involve employing some type of technological means. However, within the litigation context, counsel must employ these technological means very cautiously. Before removing or blocking embedded information in documents that might be subject to discovery, counsel should obtain direction from the court, or arrive at a mutual agreement, as to how they both will proceed.

As to lawyers who receive embedded information in electronic documents, in light of the duty imposed upon the sending lawyer, it is reasonable to assume that imbedded data that is sent in an electronic file usually is intentional. Lawyers receiving such information should be free to review it, unless the lawyers know, or should know, that it was not intended for them. In the situation where material is sent which is not intended for the receiving lawyer, the sending lawyer should be notified of its transmission and the receiving lawyer should follow the sending lawyer's directions.

In situations where confidential material is mistakenly or inadvertently sent, such act of the sending lawyer should not, in and of itself, amount to a waiver of attorney-client privilege. Typically, the standard of reasonable care is employed when determining if privilege has been waived. However, since privilege is for the welfare of the client, perhaps the profession should reconsider whether a mistaken or inadvertent act on the part of the lawyer, or even lawyer negligence, should deprive the client of privilege. While the majority of jurisdictions call for reasonable care on behalf of the sender, along with fundamental fairness, when determining waiver of privilege with inadvertent transmissions, it seems that the fate of the client should not solely rest on the action of the lawyer. Instead, perhaps the intent to disclose should be the linchpin for waiver of privilege as it relates to the client, especially in light of today's digital document exchange. We anticipate mistakes with

[340] Hricik & Jueneman, *supra* note 327, at 20.

electronic transmissions, and implement rules to help protect client information. Perhaps we should take protection of the client a step further, so clients do not lose privilege because their lawyers make mistakes or act in a careless manner.

427

International Law in Cyberspace

Remarks
Harold Hongju Koh
Legal Advisor U.S. Department of State
USCYBERCOM Inter-Agency Legal Conference
Ft. Meade, MD
September 18, 2012

As prepared for delivery

Thank you, Colonel Brown, for your kind invitation to speak here today at this very important conference on "the roles of cyber in national defense." I have been an international lawyer for more than thirty years, a government lawyer practicing international law for more than a decade, and the State Department's Legal Adviser for nearly 3 ½ years. While my daily workload covers many of the bread and butter issues of international law—diplomatic immunity, the law of the sea, international humanitarian law, treaty interpretation—like many of you, I find more and more of my time is spent grappling with the question of how international law applies in cyberspace.

Everyone here knows that cyberspace presents new opportunities and new challenges for the United States in every foreign policy realm, including national defense. But for international lawyers, it also presents cutting-edge issues of international law, which go to a very fundamental question: *how do we apply old laws of war to new cyber-circumstances, staying faithful to enduring principles, while accounting for changing times and technologies?*

Many, many international lawyers here in the U.S. Government and around the world have struggled with this question, so today I'd like to present an overview of how we in the U.S. Government have gone about meeting this challenge. At the outset, let me highlight that the entire endeavor of applying established international law to cyberspace is part of a broader international conversation. We are not alone in thinking about these questions; we are actively engaged with the rest of the international community, both bilaterally and multilaterally, on the subject of applying international law in cyberspace.

With your permission, I'd like to offer a series of questions and answers that illuminate where we are right now – in a place where we've made remarkable headway in a relatively short period of time, but are still finding new questions for each and every one we answer. In fact, the U.S. Government has been regularly sharing these thoughts with our international partners. Most of the points that follow we have not just agreed upon internally, but made diplomatically, in our submissions to the UN Group of Governmental Experts (GGE) that deals with information technology issues.

I. International Law in Cyberspace: What We Know

So let me start with the most fundamental questions:

Question 1: *Do established principles of international law apply to cyberspace?*

Answer 1: **Yes, international law principles do apply in cyberspace.** Everyone here knows how cyberspace opens up a host of novel and extremely difficult legal issues. But on this key question, this answer has been apparent, at least as far as the U.S. Government has been concerned. Significantly, this view has not necessarily been universal in the international community. At least one country has questioned whether existing bodies of international law apply to the cutting edge issues presented by the internet. Some have also said that existing international law is not up to the task, and that we need entirely new treaties to impose a unique set of rules on cyberspace. But the United States has made clear our view that established principles of international law *do* apply in cyberspace.

Question 2: *Is cyberspace a law-free zone, where anything goes?*

Answer 2: **Emphatically no. Cyberspace is not a "law-free" zone where anyone can conduct hostile activities without rules or restraint.**

Think of it this way. This is not the first time that technology has changed and that international law has been asked to deal with those changes. In particular, because the tools of conflict are constantly evolving, one relevant body of law – international humanitarian law, or the law of armed conflict – affirmatively anticipates technological innovation, and contemplates that its existing rules will apply to such innovation. To be sure, new technologies raise new issues and thus, new questions. Many of us in this room have struggled with such questions, and we will continue to do so over many years. But to those who say that established law is not up to the task, we must articulate and build consensus around how it applies and reassess from there whether and what additional understandings are needed. Developing common understandings about how these rules apply in the context of cyberactivities in armed conflict will promote stability in this area.

That consensus-building work brings me to some questions and answers we have offered to our international partners to explain how both the law of *going to* war (*jus ad bellum*) and the laws that apply in conducting war (*jus in bello*) apply to cyberaction:

Question 3: *Do cyber activities ever constitute a use of force?*

428

Answer 3: **Yes**. **Cyber activities may in certain circumstances constitute uses of force within the meaning of Article 2(4) of the UN Charter and customary international law.** In analyzing whether a cyber operation would constitute a use of force, most commentators focus on whether the direct physical injury and property damage resulting from the cyber event looks like that which would be considered a use of force if produced by kinetic weapons. *Cyber activities that proximately result in death, injury, or significant destruction would likely be viewed as a use of force.* In assessing whether an event constituted a use of force in or through cyberspace, we must evaluate factors: including the context of the event, the actor perpetrating the action (recognizing challenging issues of attribution in cyberspace), the target and location, effects and intent, among other possible issues. Commonly cited examples of cyber activity that would constitute a use of force include, for example: (1) operations that trigger a nuclear plant meltdown; (2) operations that open a dam above a populated area causing destruction; or (3) operations that disable air traffic control resulting in airplane crashes. Only a moment's reflection makes you realize that this is common sense: if the physical consequences of a cyber attack work the kind of physical damage that dropping a bomb or firing a missile would, that cyber attack should equally be considered a use of force.

Question 4: *May a State ever respond to a computer network attack by exercising a right of national self-defense?*

Answer 4: **Yes**. **A State's national right of self-defense, recognized in Article 51 of the UN Charter, may be triggered by computer network activities that amount to an armed attack or imminent threat thereof.** As the United States affirmed in its 2011 International Strategy for Cyberspace, "when warranted, the United States will respond to hostile acts in cyberspace as we would to any other threat to our country."

Question 5: *Do jus in bello rules apply to computer network attacks?*

Answer 5: **Yes. In the context of an armed conflict, the law of armed conflict applies to regulate the use of cyber tools in hostilities, just as it does other tools. The principles of necessity and proportionality limit uses of force in self-defense and would regulate what may constitute a lawful response under the circumstances.** There is no legal requirement that the response to a cyber armed attack take the form of a cyber action, as long as the response meets the requirements of necessity and proportionality.

Question 6: *Must attacks distinguish between military and nonmilitary objectives?*

Answer 6: Yes. The *jus in bello* principle of distinction applies to computer network attacks undertaken in the context of an armed conflict. The principle of distinction applies to cyber activities that amount to an "attack" – as that term is understood in the law of war – in the context of an armed conflict. As in any form of armed conflict, the principle of distinction requires that the intended effect of the attack must be to harm a legitimate *military* target. We must distinguish military objectives – that is, objects that make an effective contribution to military action and whose destruction would offer a military advantage – from civilian objects, which under international law are generally protected from attack.

Question 7: *Must attacks adhere to the principle of proportionality?*

Answer 7: **Yes. The *jus in bello* principle of proportionality applies to computer network attacks undertaken in the context of an armed conflict.** The principle of proportionality prohibits attacks that may be expected to cause incidental loss to civilian life, injury to civilians, or damage to civilian objects that would be excessive in relation to the concrete and direct military advantage anticipated. Parties to an armed conflict must assess what the expected harm to civilians is likely to be, and weigh the risk of such collateral damage against the importance of the expected military advantage to be gained. In the cyber context, this rule requires parties to a conflict to assess: (1) the effects of cyber weapons on both military and civilian infrastructure and users, including shared physical infrastructure (such as a dam or a power grid) that would affect civilians; (2) the potential physical damage that a cyber attack may cause, such as death or injury that may result from effects on critical infrastructure; and (3) the potential effects of a cyber attack on civilian objects that are *not* military objectives, such as private, civilian computers that hold no military significance, but may be networked to computers that are military objectives.

Question 8: *How should States assess their cyber weapons?*

Answer 8: **States should undertake a legal review of weapons, including those that employ a cyber capability.** Such a review should entail an analysis, for example, of whether a particular capability would be *inherently indiscriminate*, *i.e.,* that it could not be used consistent with the principles of distinction and proportionality. The U.S. Government undertakes at least two stages of legal review of the use of weapons in the context of armed conflict – first, an evaluation of new weapons to determine whether their use would be *per se* prohibited by the law of war; and second, specific operations employing weapons are always reviewed to ensure that each particular operation is also compliant with the law of war.

Question 9: *In this analysis, what role does State sovereignty play?*

Answer 9: **States conducting activities in cyberspace must take into account the sovereignty of other States, including outside the context of armed conflict.** The physical infrastructure that supports the internet and cyber activities is generally located in sovereign territory and subject to the jurisdiction of the territorial State. Because of the interconnected, interoperable nature of cyberspace, operations targeting networked information infrastructures in one country may create effects in another country. Whenever a State contemplates conducting activities in cyberspace, the sovereignty of other States needs to be considered.

Question 10: *Are States responsible when cyber acts are undertaken through proxies?*

Answer 10: **Yes**. **States are legally responsible for activities undertaken through "proxy actors," who act on the State's instructions or under its direction or control.** The ability to mask one's identity and geography in cyberspace and the resulting difficulties of timely, high-confidence attribution can create significant challenges for States in identifying, evaluating, and accurately responding to threats. But putting attribution problems aside for a moment, established international law does address the question of proxy actors. States are legally responsible for activities undertaken through putatively private actors, who act on the State's instructions or under its direction or control. If a State exercises a sufficient degree of control over an ostensibly private person or group of persons committing an internationally wrongful act, the State assumes responsibility for the act, just as if official agents of the State itself had committed it. These rules are designed to

ensure that States cannot hide behind putatively private actors to engage in conduct that is internationally wrongful.

II. International Law in Cyberspace: Challenges and Uncertainties

These ten answers should give you a sense of how far we have come in doing what any good international lawyer does: applying established law to new facts, and explaining our positions to other interested lawyers. At the same time, there are obviously many more issues where the questions remain under discussion. Let me identify three particularly difficult questions that I don't intend to answer here today. Instead, my hope is to shed some light on some of the cutting-edge legal issues that we'll all be facing together over the next few years:

Unresolved Question 1: **How can a use of force regime take into account all of the novel kinds of *effects* that States can produce through the click of a button?**

As I said above, the United States has affirmed that established *jus ad bellum* rules do apply to uses of force in cyberspace. I have also noted some clear-cut cases where the physical effects of a hostile cyber action would be comparable to what a kinetic action could achieve: for example, a bomb might break a dam and flood a civilian population, but insertion of a line of malicious code from a distant computer might just as easily achieve that same result. As you all know, however, there are other types of cyber actions that do not have a clear kinetic parallel, which raise profound questions about exactly what we mean by "force." At the same time, the difficulty of reaching a definitive legal conclusion or consensus among States on when and under what circumstances a hostile cyber action would constitute an armed attack does not automatically suggest that we need an entirely new legal framework specific to cyberspace. Outside of the cyber-context, such ambiguities and differences of view have long existed among States.

To cite just one example of this, the United States has for a long time taken the position that the inherent right of self-defense potentially applies against *any* illegal use of force. In our view, there is no threshold for a use of deadly force to qualify as an "armed attack" that may warrant a forcible response. But that is not to say that any illegal use of force triggers the right to use any and all force in response – such responses must still be *necessary* and of course *proportionate*. We recognize, on the other hand, that some other countries and commentators have drawn a distinction between the "use of force" and an "armed attack," and view "armed attack" – triggering the right to self-defense – as a subset of uses of force, which passes a higher threshold of gravity. My point here is not to rehash old debates, but to illustrate that States have long had to sort through complicated *jus ad bellum* questions. In this respect, the existence of complicated cyber questions relating to *jus ad bellum* is not in itself a new development; it is just applying old questions to the latest developments in technology.

Unresolved Question 2: **What do we do about "*dual-use* infrastructure" in cyberspace?**

As you all know, information and communications infrastructure is often shared between State militaries and private, civilian communities. The law of war requires that civilian infrastructure not be used to seek to immunize military objectives from attack, including in the cyber realm. But how, exactly, are the *jus in bello* rules to be implemented in cyberspace? Parties to an armed conflict will need to assess the potential effects of a cyber attack on computers that are *not* military objectives, such as private, civilian computers that hold no military significance, but may be networked to computers that are valid military objectives. Parties will also need to consider the harm to the civilian uses of such infrastructure in performing the necessary proportionality review. Any number of factual scenarios could arise, however, which will require a careful, fact-intensive legal analysis in each situation.

Unresolved Question 3: **How do we address the problem of *attribution* in cyberspace?**

As I mentioned earlier, cyberspace significantly increases an actor's ability to engage in attacks with "plausible deniability," by acting through proxies. I noted that legal tools exist to ensure that States are held accountable for those acts. What I want to highlight here is that many of these challenges – in particular, those concerning attribution – are as much questions of a technical and policy nature rather than exclusively or even predominantly questions of law. Cyberspace remains a new and dynamic operating environment, and we cannot expect that all answers to the new and confounding questions we face will be *legal* ones.

These questions about effects, dual use, and attribution are difficult legal and policy questions that existed long before the development of cyber tools, and that will continue to be a topic of discussion among our allies and partners as cyber tools develop. Of course, there remain many other difficult and important questions about the application of international law to activities in cyberspace – for example, about the implications of sovereignty and neutrality law, enforcement mechanisms, and the obligations of States concerning "hacktivists" operating from within their territory. While these are not questions that I can address in this brief speech, they are critically important questions on which international lawyers will focus intensely in the years to come.

And just as cyberspace presents challenging new issues for lawyers, it presents challenging new technical and policy issues. Not all of the issues I've mentioned are susceptible to clear legal answers derived from existing precedents – in many cases, quite the contrary. Answering these tough questions within the framework of existing law, consistent with our values and accounting for the legitimate needs of national security, will require a constant dialogue between lawyers, operators, and policymakers. All that we as lawyers can do is to apply in the cyber context the same rigorous approach to these hard questions that arise in the future, as we apply every day to what might be considered more traditional forms of conflict.

III. The Role of International Law in a "Smart Power" Approach to Cyberspace

This, in a nutshell, is where we are with regard to cyberconflict: We have begun work to build consensus on a number of answers, but questions continue to arise that must be answered in the months and years ahead. Beyond these questions and answers and unresolved questions, though, lies a much bigger picture, one that we are very focused on at the State Department. Which brings me to my final two questions:

Final Question 1: *Is international humanitarian law the only body of international law that applies in cyberspace?*

Final Answer 1: **No. As important as international humanitarian law is, it is *not* the only international law that applies in cyberspace.**

430

Obviously, cyberspace has become pervasive in our lives, not just in the national defense arena, but also through social media, publishing and broadcasting, expressions of human rights, and expansion of international commerce, both through online markets and online commercial techniques. Many other bodies of international and national law address those activities, and how those different bodies of law overlap and interact with the laws of cyber conflict is something we will all have to work out over time.

Take human rights. At the same time that cyber activity can pose a threat, we all understand that cyber-communication is increasingly becoming a dominant mode of expression in the 21st century. More and more people express their views not by speaking on a soap box at Speakers' Corner, but by blogging, tweeting, commenting, or posting videos and commentaries. The 1948 Universal Declaration of Human Rights (UDHR) – adopted more than 70 years ago – was remarkably forward-looking in anticipating these trends. It says: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas *through any media and regardless of frontiers*." (emphasis added) In short, all human beings are entitled to certain rights, whether they choose to exercise them in a city square or an internet chat room. This principle is an important part of our global diplomacy, and is encapsulated in the Internet Freedom agenda about which my boss, Secretary Clinton, has spoken so passionately.

You all know of this Administration's efforts not just in the areas of cyberconflict, but also in many other cyber areas: cybersecurity, cybercommerce, fighting child pornography and other forms of cybercrime. stopping intellectual property piracy, as well as promoting free expression and human rights. So the cyberconflict issues with which this group grapples do not constitute the whole of our approach to cyberspace; they are an important part – but only a part –of this Administration's broader "smart power" approach to cyberspace.

What I have outlined today are a series of answers to cyberspace questions that the United States is on the record as supporting. I have also suggested a few of the challenging questions that remain before us, and developments over the next decade will surely produce new questions. But you should not think of these questions and answers as just a box to check before deciding whether a particular proposed operation is lawful or not. Rather, these questions and answers are part of a much broader foreign policy agenda, which transpires in a broader framework of respect for international law.

That leads to my Final Question for this group: *Why should U.S Government lawyers care about international law in cyberspace at all?*

The Answer: **Because compliance with international law *frees* us to do more, and do more legitimately, in cyberspace, in a way that more fully promotes our national interests. Compliance with international law in cyberspace is part and parcel of our broader "smart power" approach to international law as part of U.S. foreign policy.**

It is worth noting two fundamentally different philosophies about international law. One way to think about law, whether domestic or international, is as a straitjacket, a pure constraint. This approach posits that nations have serious, legitimate interests, and legal regimes restrict their ability to carry them out. One consequence of this view is that, since law is just something that constrains, it should be resisted whenever possible. Resisting so-called "extensions" of the law to new areas often seems attractive: because, after all, the old laws weren't built for these new challenges anyway, some say, so we should tackle those challenges without the legal straitjacket, while leaving the old laws behind.

But that is *not* the United States Government's view of the law, domestic or international. We see law not as a straitjacket, but as one great university calls it when it confers its diplomas, a body of "wise restraints that make us free." International law is not purely constraint, it frees us and empowers us to do things we could never do without law's legitimacy. If we succeed in promoting a culture of compliance, we will reap the benefits. And if we earn a reputation for compliance, the actions we *do* take will earn enhanced legitimacy worldwide for their adherence to the rule of law.

These are not new themes, but I raise them here because of they resonate squarely with the strategy we have been pursuing in cyberspace over the past few years. Of course, the United States has impressive cyber-capabilities; it should be clear from the bulk of my discussion that adherence to established principles of law does not prevent us from using those capabilities to achieve important ends. But we also know that we will be safer, the more that we can rally other States to the view that these established principles *do* impose meaningful constraints, and that there is already an existing set of laws that protect our security in cyberspace. And the more widespread the understanding that cyberspace follows established rules – and that we live by them – the stronger we can be in pushing back against those who would seek to introduce brand new rules that may be contrary to our interests.

That is why, in our diplomacy, we do not whisper about these issues. We talk *openly and bilaterally* with other countries about the application of established international law to cyberspace. We talk about these issues *multilaterally*, at the UN Group of Governmental Experts and at other fora, in promoting this vision of compliance with international law in cyberspace. We talk about them *regionally*, as when we recently co-sponsored an ASEAN Regional Forum event to focus the international community's attention on the problem of proxy actors engaging in unlawful conduct in cyberspace. Preventing proxy attacks on us is an important interest, and as part of our discussions we have outlined the ways that existing international law addresses this problem.

The diplomacy I have described is not limited to the legal issues this group of lawyers is used to facing in the operational context. These issues are interconnected with countless other cyber issues that we face daily in our foreign policy, such as cybersecurity, cyber-commerce, human rights in cyberspace, and public diplomacy through cybertools. In all of these areas, let me repeat again, *compliance with international law in cyberspace is part and parcel of our broader smart power approach to international law as part of U.S. foreign policy*. Compliance with international law – and thinking actively together about how best to promote that compliance – can only free us to do more, and to do more legitimately, in the emerging frontiers of cyberspace, in a way that more fully promotes our U.S. national interests.

Thank you very much.

HARVARD INTERNATIONAL LAW JOURNAL

FEATURE ARTICLE
DECEMBER 2012

Online
Volume 54

International Law in Cyberspace: The Koh Speech and Tallinn Manual Juxtaposed

Michael N. Schmitt[1]

In 2011, the White House issued the *International Strategy for Cyberspace*, which noted that "[t]he development of norms for state conduct in cyberspace does not require a reinvention of customary international law, nor does it render existing international norms obsolete. Long-standing international norms guiding state behavior—in times of peace and conflict—also apply in cyberspace."[2] However, the document cautioned

[1] *Chairman and Professor, Department of Law, United States Naval War College. Professor Schmitt is also Honorary Professor at Durham University in the United Kingdom and former Dean of the Marshall Center in Germany. From 2009–2012, he served as Director of the NATO Cooperative Cyber Defence Centre of Excellence's Tallinn Manual project. The views expressed in this article are those of the author in his personal capacity and do not necessarily reflect those of the United States government.*

[2] THE WHITE HOUSE, INTERNATIONAL STRATEGY FOR CYBERSPACE: PROSPERITY, SECURITY, AND OPENNESS IN A NETWORKED WORLD 9 (2011), *available at*

that the "unique attributes of networked technology require additional work to clarify how these norms apply and what additional understandings might be necessary to supplement them."[3]

On September 18, 2012, State Department Legal Adviser Harold Koh took an important step towards publically elucidating the U.S. positions on how international law applies to cyberspace.[4] At a conference sponsored by United States Cyber Command (USCYBERCOM), Mr. Koh offered brief answers to what he labeled the "fundamental questions" on the issue. He also identified several "unresolved questions" with which the United States would likely be forced to grapple in the future. Since the speech had been fully cleared in the interagency process, it can be viewed as reflecting the U.S. Government's views on the issues, not just those of Mr. Koh or the State Department.

The timing of the speech was propitious. Less than three weeks earlier, NATO's Cooperative Cyber Defence Centre of Excellence (CCD COE) had released a draft the long-awaited *Tallinn Manual*, due for formal publication in early 2013.[5] The Manual is the product of a three-year project sponsored by the Centre in which an "International Group of Experts" examined, *inter alia*, the very issues cited in the Koh Speech, supra note 4. Participants included distinguished legal academics and practitioners, supported by a team of technical experts.[6] USCYBERCOM, the

---

http://www.whitehouse.gov/sites/default/files/rss_viewer/international_strategy_for_cyberspace.pdf

[3] *Id.*

[4] Harold Honhgu Koh, Legal Advisor of the Dep't of State, International Law in Cyberspace, Address to the USCYBERCOM Inter-Agency Legal Conference (Sept. 18, 2012), *available at* http://www.state.gov/s/l/releases/remarks/197924.htm. A footnoted version of the Koh Address is forthcoming on the Harvard International Law Journal Online. 54 Harv. Int'l L.J. Online 1 (Forthcoming, 2012). The United States has participated in meetings of the U.N. "Group of Governmental Experts" on cyber issues. It provided a paper on the U.S. position which was largely appended to the 2011 report issued by the Group. United Nations, *Developments in the Field of Information and Telecommunications in the Context of International Security 31* (2011), http://www.un.org/disarmament/HomePage/ODAPublications/DisarmamentStudySeries/PDF/DSS_33.pdf (hereinafter GGE Report). Of note was the U.S. acceptance of the applicability of the *jus ad bellum* and *jus in bello* to activities in cyberspace. *Id.* at 35–37. Note that the U.S. submission was, on matters of law, somewhat less detailed than the Koh speech and not draw significant attention beyond the expert community.

[5] TALLINN MANUAL ON THE INTERNATIONAL LAW APPLICABLE TO CYBER WARFARE (Michael N. Schmitt, gen. ed., forthcoming Cambridge University Press 2013), http://www.ccdcoe.org/249.html.

[6] Although numerous members of the group were serving in senior posts in their countries, all participated in their personal capacity.

International Committee of the Red Cross, and NATO each provided an observer who participated actively throughout the project, albeit in a non-voting capacity.

The *Tallinn Manual* consists of "rules" adopted unanimously by the International Group of Experts that are meant to reflect customary international law, accompanied by "commentary" that delineates their legal basis and highlights any differences of opinion among the Experts as to their interpretation in the cyber context. A select group of peer reviewers offered comments on the various drafts, as did a number of states that were willing to informally and unofficially do so. The author served as Director of the Project.

The relative congruency between the U.S. Government's views, as reflected in the Koh speech, and those of the International Group of Experts is striking. This confluence of a state's expression of *opinio juris* with a work constituting "the teachings of the most highly qualified publicists of the various nations" significantly enhances the persuasiveness of common conclusions. [7] Of course, the limited differences that exist as to particular points of law render the respective positions on those points somewhat less compelling.

This article serves two purposes. First, it functions as a concordance between the positions articulated in the Koh speech and those found in the *Tallinn Manual.* The comparison is particularly apropos in light of the parallels in their content. Second, drawing on the *Tallinn Manual*, the article provides analytical granularity as to the legal basis for the positions proffered in the Koh Speech, supra note 4. In doing so, it usefully catalogues the various competing interpretive perspectives. The article is crafted around Mr. Koh's "Questions and Answers," which are reordered topically and set forth at the beginning of each section.

I. APPLICABILITY OF INTERNATIONAL LAW

"Do established principles of international law apply to cyber space? Yes, international law principles do apply in cyberspace."

"Is cyber-space a 'law-free' zone where anything goes? Cyberspace is not a "law-free" zone where anyone can conduct hostile activities without rules or restraints."

"Do *jus in bello* rules apply to computer network attack? Yes. In the context of an armed conflict, the law of armed conflict applies to regulate the use of cyber tools in hostilities, just as it does other tools. The principles of necessity

---

[7] Statute of the International Court of Justice, art. 38(1)(d), June 26, 1945, 59 Stat. 1055.

and proportionality limit uses of force in self-defense and would regulate what may constitute a lawful response under the circumstances."[8]

The *sine qua non* issue for the *Tallinn Manual* was whether international law applies to cyber activities at all, for absent an affirmative response the project would have been pointless. In unanimously agreeing that it does, the International Group of Experts adopted precisely the same position as the U.S. Government on each of the answers set forth above. Since their work was focused on cyber conflict, the Experts took particular note of the International Court of Justice's *Nuclear Weapons* Advisory Opinion.[9] In that case, the Court had to consider whether the prohibition on the use of force found in Article 2(4) and related articles of the U.N. Charter (elements of the *jus ad bellum*) governed the use of nuclear weapons.[10] It opined that they "apply to any use of force, regardless of the weapons employed."[11] Applying this normative logic analogously, the Experts concluded, "the mere fact that a computer (rather than a more traditional weapon, weapon system, or platform) is used during an operation has no bearing on whether that operation amounts to a 'use of force'. Similarly, it has no bearing on whether a State may use force in self-defence."[12]

The International Group of Experts espoused the same view with regard to the *jus in bello* (international humanitarian law), which the International Court of Justice did not hesitate to apply to nuclear weapons. It began by highlighting the well-known pronouncement in the 1907 Hague Convention IV Regulations that "the right of belligerents to adopt means of injuring the enemy is not unlimited."[13] The Court then turned to the "cardinal" international humanitarian law principles of distinction and unnecessary suffering as baselines for analyzing the legality of the use of nuclear weapons.[14] This approach confirmed that, for the Court, international humanitarian treaty and customary law that predated the fielding of nuclear weapons governs their

---

[8]Koh Speech, *supra* note 4, at 2–4

[9] Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226 (hereinafter, Nuclear Weapons).

[10] *Id.*; U.N. Charter art. 2, para. 4; art. 51; art. 42. The text of the first two articles is set forth in the text accompanying sec. II and III, respectively. Article 42 provides, in relevant part, that "[s]hould the Security Council consider that measures provided for in Article 41 would be inadequate or have proved to be inadequate, it may take such action by air, sea, or land forces as may be necessary to maintain or restore international peace and security."

[11] Nuclear Weapons, *supra* note 8, para. 39.

[12] Tallinn Manual, *supra* note 5, para. 1 of commentary accompanying chapeau to ch. II.

[13] Nuclear Weapons, *supra* note 8, para. 77, *citing* Convention (IV) Respecting the Laws and Customs of War on Land and its annex: Regulations concerning the Laws and Customs of War on Land, art. 22, Oct. 18, 1907, 36 Stat. 2277.

[14] Nuclear Weapons, *supra* note 8, para. 78.

employment. The Experts found no reason to deviate from this position in the cyber context.

Like the International Court of Justice in the *Nuclear Weapons* Advisory Opinion, the International Group of Experts also emphasized the Martens Clause's relevance to cyber operations.[15] The clause, which first appeared in the 1899 Hague Convention II, finds its contemporary expression in the 1977 Additional Protocol I to the 1949 Geneva Conventions: "[i]n cases not covered by this Protocol or by other international agreements, civilians and combatants remain under the protection and authority of the principles of international law derived from established custom, from the principles of humanity and from the dictates of public conscience."[16] For over a century, therefore, it has been well accepted that a lack of directly applicable treaty law does not create an international humanitarian law-free zone. Indeed, international humanitarian law's requirement for a legal review of weapons prior to fielding, which is discussed below, confirms the fact that cyber weapons, as with other new weapons, are subject to preexisting law.[17]

Simply put, the Experts rejected any characterization of cyberspace as a distinct domain subject to a discrete body of law. The fulcrum of their conclusion was the fact that a person located at a particular place uses tangible cyber infrastructure to conduct cyber activities.[18] Application of international law to cyber activities is accordingly a matter of identifying the relevant legal principles that bear on the person, place, object, or type of activity in question.

Although there was no dissent over international law's applicability, it became clear during the project's proceedings that interpretation of international law norms in the cyber context can be challenging. For instance, crafting a consensus understanding of how international humanitarian law's definition of "attacks" applies to cyber operations proved arduous. The Experts also discovered that applying international law principles to cyberspace raises many of the same controversies that attend their application on land, at sea, or in the air. The best illustration of this reality concerns the dispute over "war-sustaining" military objectives. The debates on both issues are discussed below. In light of these are similar challenges, the Experts involved in drafting the *Tallinn Manual* would emphatically agree with Mr. Koh's assertion that

---

[15] *Id.*; Tallinn Manual, *supra* note 5, R. 20 cmt. 10.
[16] Convention (II) with Respect to the Laws and Customs of War on Land, pmbl., July 29, 1899, 22 Stat. 1803; Protocol Additional (I) to the Geneva Conventions of 12 August 1949, and relating to the Protection of Victims of International Armed Conflicts (Additional Protocol I), art. 1(2), June 8, 1977, 1125 U.N.T.S. 3.
[17] Additional Protocol I, art. 36; Tallinn Manual, *supra* note 5, R. 48.
[18] Tallinn Manual, *supra* note 5, para. 2 of commentary accompanying chapeau to Part 1.

"we must articulate and build consensus around how [international law] applies and reassess from there whether and what additional understandings are needed."[19]

II. THE USE OF FORCE

"**Do cyber activities ever constitute a use of force? Cyber activities may in certain circumstances constitute uses of force within the meaning of Article 2(4) of the UN Charter and customary international law."**

"How can a use of force regime take into account all of the novel kinds of *effects* that States can produce through the click of a button? Unresolved."[20]

International law's prohibition of the use of force is set forth in Article 2(4) of the U.N. Charter: "All Members [of the United Nations] shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any State, or in any other manner inconsistent with the Purposes of the United Nations."[21] The article undoubtedly represents a norm of customary international law.[22]

In *jus ad bellum* analyses, the notion of "use of a force" is often confused that of "armed attack." The former bears on whether an action violates international law as codified in Article 2(4). By contrast, act(s) that cross the armed attack threshold found in Article 51 of the U.N. Charter (and customary international law) concern a target-state's entitlement to respond defensively with its own kinetic or cyber use of force. Moreover, while the use of force prohibition only applies to the acts of states (or those attributable to states under the law of state responsibility), the right of self-defense arguably encompasses attacks mounted by nonstate actors.[23]

Although it is incontrovertible that the prohibition on the use of force applies to cyber operations, the question remains as to when such operations amount to uses of force, such that they are prohibited absent one of the two recognized exceptions to

---

[19] Koh Speech, *supra* note 4, at 3, 7.
[20] *Id.*
[21] U.N. Charter, art. 2, para. 4.
[22] Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), 1986 I.C.J. 14, (hereinafter, Nuclear Weapons) at para. 188–90.
[23] See Tallinn Manual, s*upra* note 5, R. 10 cmt. 5, R. 13 cmt. 16, for a discussion setting forth the specifics of this issue.

the prohibition (self-defense and mandate or authorization by the Security Council).[24] For the U.S. Government, the physical effects of a cyber operation are the key. In particular, Mr. Koh asserted that "[c]yber activities that proximately result in death, injury, or significant destruction would likely be viewed as a use of force."[25] For him, it is a matter of common sense: "if the physical consequences of a cyber attack work the kind of physical damages that dropping a bomb or firing a missile would, that cyber attack should equally be considered a use of force." Mr. Koh goes on to suggest that "[i]n assessing whether an event constituted a use of force in or through cyberspace, we must evaluate factors: including the context of the event, the actor perpetrating the action (recognizing challenging issues of attribution in cyberspace), the target and location, effects and intent, among other possible issues."[26]

The International Group of Experts came to a similar conclusion regarding physical effects in the *Tallinn Manual*: "A cyber operation constitutes a use of force when its scale and effects are comparable to non-cyber operations rising to the level of a use of force."[27] For the Experts, "[a]cts that injure or kill persons or damage or destroy objects are unambiguously uses of force,"[28] so long as the effects are not trivial in nature and the cyber operations have been carried out by, or are attributable to, a state. The Experts were even more categorical than Mr. Koh, who cautiously noted that such acts "would likely be viewed" as uses of force, and suggested that factors such as those mentioned above would have to be evaluated when making the use of force determination.[29] Despite the minor difference in confidence level, the U.S. Government and the International Group of Experts would likely come to the same conclusions in specific cases. For instance, Mr. Koh cited cyber operations triggering a nuclear plant meltdown, opening a dam upriver from a populated area, and disabling air-traffic control as examples of uses of force.[30] The Experts discussed these very examples during their sessions.

As pointed out in the speech, some cyber incidents lack a clear kinetic parallel. Most noteworthy are those involving cyber operations that do not result in physical damage or injury. With regard to these incidents, the Experts took the position that "[a] use of force need not involve the employment of military or other armed forces by the

---

[24] U.N. Charter, arts. 42 & 51.
[25] Koh Speech, *supra* note 4, at 4.
[26] *Id.*
[27] Tallinn Manual, s*upra* note 5, R. 11. Note that the phrase "scale and effects" is drawn from Nicaragua, *supra* note 20, para. 195. Although the International Court of Justice used it there with reference to the "armed attack" standard of Article 51, the International Group of Experts also found it a useful approach with respect to evaluating potential uses of forces.
[28] Tallinn Manual, s*upra* note 5, R. 11 cmt. 8.
[29] Koh Speech, *supra* note 4, at 4.
[30] *Id.*

State in question."[31] As support, they pointed to the *Nicaragua* case, in which the International Court of Justice held that although merely funding guerrillas who were conducting hostilities against another State did not reach the use of force threshold, arming and training them did.[32] The holding suggests that an act need not have immediate physical consequences to comprise a use of force.

While this may be so, the dilemma of how to determine where the use of force threshold lies in cases not involving physical harm remains unresolved. Given the absence of a definitive threshold, the International Group of Experts adopted an approach that seeks to determine the probability that States (and others) will characterize a cyber operation as a use of force. They identified eight key non-exclusive factors likely to be considered on a case-by-case basis during such assessments.[33]

Of these, the most significant is "severity". Indeed, as noted, a cyber operation that results in damage, destruction, injury, or death is "highly likely to be considered a use of force" irrespective of the other factors.[34] Those other factors include: immediacy (the speed with which consequences manifest), directness (the causal relation between a cyber operation and its consequences), invasiveness (the degree to which a cyber operation intrudes into targeted systems), measurability of the effects, military character of the cyber operation, extent of State involvement, and presumptive legality (acts not expressly prohibited by international law).[35] Depending on the circumstances, additional factors like the prevailing political environment, whether the operations portend imminent military force, the attacker's identity, the attacker's cyber operations track record, and the nature of the target could also prove influential.[36] Based on the aforementioned factors and the *Nicaragua* judgment, the Experts concluded, for example, that providing an organized armed group with malware to be used against another State would constitute a use of force, whereas merely providing sanctuary to that group would, for a majority of the Experts, not rise to that level.[37] Ultimately every determination depends on a holistic assessment of the incident in light of the attendant circumstances.

---

[31] Tallinn Manual, s*upra* note 5, R. 11 cmt. 4.
[32] Nicaragua, *supra* note 18, para. 228.
[33] Tallinn Manual, s*upra* note 5, R. 11 cmt. 9.
[34] *Id.*
[35] *Id.*
[36] Tallinn Manual, s*upra* note 5, R. 11 cmt. 10.
[37] Tallinn Manual, s*upra* note 5, R. 11 cmts. 4,5.

III. SELF-DEFENSE

"May a State ever respond to a computer network attack by exercising a right of national self-defense? Yes. A State's national right of self-defense, recognized in Article 51 of the UN Charter, may be triggered by computer network activities that amount to an armed attack or imminent threat thereof."[38]

Article 51 of the United Nations Charter sets forth the right of self-defense: "Nothing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs against a Member of the United Nations, until the Security Council has taken the measures necessary to maintain international peace and security." In his speech, Mr. Koh reiterated the U.S. position on self-defense against a cyber armed attack, one that had previously been announced in the *International Strategy for Cyberspace*: "when warranted, the United States will respond to hostile acts in cyberspace as we would to any other threat to our country."[39]

The *Tallinn Manual* is in accord. It provides that "[a] State that is the target of a cyber operation that rises to the level of an armed attack may exercise its inherent right of self-defense."[40] The Experts and the US Government agree that cyber operations that kill or seriously injure individuals or cause serious damage to objects qualify as armed attacks. Defensive actions are, as with kinetic actions, subject to the requirements of necessity, proportionality, imminency, and immediacy.[41]

The question remains, however, as to when a cyber operation amounts to an armed attack. On this point, the U.S. Government and the International Group of Experts part ways. The government is of the view that "the inherent right of self-defense potentially applies against *any* illegal use of force.… [T]here is no threshold for a use

---

[38] Koh Speech, *supra* note 4, at 4.

[39] The *International Strategy for Cyberspace* notes: "The development of norms for State conduct in cyberspace does not require a reinvention of customary international law, nor does it render existing international norms obsolete. Long-standing international norms guiding State behavior—in times of peace and conflict—also apply in cyberspace." International Strategy for Cyberspace, *supra* note 3, at 9.

[40] Tallinn Manual, s*upra* note 5, R. 13.

[41] *Id.*, R. 13– 5. *See also* Nicaragua, s*upra* note 20, paras. 176, 194; Nuclear Weapons, s*upra* note 8, para. 41; Oil Platforms (Iran v. U.S.), 2003 I.C.J. 161, at paras. 43, 73–74, 76; Judgment of the International Military Tribunal Sitting at Nuremberg, Germany (Sept. 30, 1946), *in* 22 THE TRIAL OF GERMAN MAJOR WAR CRIMINALS: PROCEEDINGS OF THE INTERNATIONAL MILITARY TRIBUNAL SITTING AT NUREMBERG, GERMANY (1950), at 435 (referring to the *Caroline* formula).

of deadly force to qualify as an 'armed attack' that may warrant a forcible response."[42] This has long been its official position.[43]

No member of the International Group of Experts agreed that an armed attack was nothing more than a use of force, *sans plus*. Instead, they endorsed the International Court of Justice's requirement to "distinguish the most grave forms of the use of force (those constituting an armed attack) from other less great forms."[44] In other words, whereas all armed attacks are uses of force, not all uses of force are armed attacks. Whether a cyber use of force qualifies as an armed attack depends on its "scale and effects."[45]

Uncertainty as to what those scale and effects are plagued the *Tallinn Manual* deliberations. The Experts observed, for instance, that the International Court of Justice differentiated a "mere frontier incident" from an armed attack,[46] but later opined that an attack on a single warship might qualify as an armed attack.[47] Such inexplicable distinctions obfuscated their attempt to identify practicable legal thresholds.

Most of the Experts adopted "serious death, injury, damage, or destruction" as the apposite effects-based threshold for armed attack.[48] However, several argued that the severity of a cyber operation's effects was of greater relevance in qualifying it as an armed attack than their physical nature. For instance, although a massive cyber operation against the economy might cause no physical harm, the magnitude of its economic impact would better justify characterizing the operation as an armed attack than would limited physical damage.[49] In their opinion, it was incongruent, and therefore contrary to the object and purpose of the right to self-defense, to characterize the latter as an armed attack, and not the former. The other Experts were willing to entertain the prospect of States eventually accepting this interpretive approach, but believed that it presently represented *lex ferenda*, not *lex lata*.[50]

The International Group of Experts also examined the possibility of a State being targeted by multiple cyber operations, none of which alone rise to the level of an armed attack. May "pinprick attacks" be amalgamated for the purpose of finding an

---

[42] Koh Speech, *supra* note 4, at 7.
[43] *See, e.g.,* Abraham D. Sofaer, *International Law and the Use of Force*, *in* 82 AMERICAN SOCIETY OF INTERNATIONAL LAW PROCEEDINGS 420, 422 (1988). Sofaer was at the time the State Department's Legal Adviser.
[44] Tallinn Manual, *supra* note 5, R. 13 cmt. 6, *citing* Nicaragua, s*upra* note 20, para. 191.
[45] *Id.*, para. 195.
[46] *Id.*
[47] Oil Platforms, s*upra* note **Error! Bookmark not defined.**, paras. 57, 61.
[48] Tallinn Manual, *supra* note 5, R. 13, para. 6.
[49] Tallinn Manual, *supra* note 5, R. 13, para. 9.
[50] *Id.*

Unknown
Deleted: *Id.*, para. 195

armed attack? The Experts agreed that pursuant to the "accumulation of effects" theory, combining effects to meet the armed attack threshold is appropriate so long as the cyber operations are conducted by the same attacker (or attackers operating in concert), are related in terms of objective, and satisfy the requisite scale and effects threshold.[51]

As reflected in the Koh speech, the US Government maintains that self-defense is permissible in the face of an imminent attack. Most members of the International Group of Experts also took the view that international law allows for "anticipatory self-defense". Accordingly, the *Tallinn Manual* notes that "[t]he right to use force in self-defense arises if a cyber armed attack occurs or is imminent. It is further subject to a requirement of immediacy."[52] By this approach, a State need not take the first "cyber hit" before acting to defend itself.

The devil is in the details. Some of the Experts who acknowledged the existence of a right of anticipatory self-defense adopted a strict temporal approach, one grounded in Secretary of State Daniel Webster's famed 19th Century assertion during the *Caroline* incident that the right of self-defense only applies when the "necessity of self-defense is instant, overwhelming, leaving no choice of means, and no moment for deliberation."[53] For these Experts, the legality of defensive actions taken anticipatorily is to be gauged by reference to the time that passes between the act in question and the pending armed attack that necessitated it.

However, the majority of the Experts were of the view that "a State may act in anticipatory self-defense against an armed attack, whether cyber or kinetic, once the attacker is clearly committed to launching an armed attack and the victim-State will lose its opportunity to effectively defend itself unless it acts."[54] For them, "[t]he critical question is not the temporal proximity of the anticipatory defensive action to the perspective armed attack, but whether a failure to act at that moment would

---

[51] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 13 cmt. 8. *See also* YORAM DINSTEIN, WAR AGGRESSION, AND SELF-DEFENCE 206 (5th ed., 2011).
[52] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 15. *See also* DEREK W. BOWETT, SELF-DEFENCE IN INTERNATIONAL LAW 188–189 (1958). *But see* IAN BROWNLIE, INTERNATIONAL LAW AND THE USE OF FORCE BETWEEN STATES 275–8 (1963); YORAM DINSTEIN, WAR AGGRESSION AND SELF DEFENCE 203–204 (5th ed. 2011). Imminency refers to the defensive measures taken before an armed attack has occurred, while immediacy refers to those taken following an armed attack.
[53] Letter from Daniel Webster to Lord Ashburton (Aug. 6, 1842), *reprinted in* 2 INTERNATIONAL LAW DIGEST 412 (John Bassett Moore ed., 1906).
[54] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 15 cmt. 4.

reasonably be expected to result in that State being unable to defend itself effectively when that attack actually starts."[55]

The International Group of Experts flatly rejected the notion of "preventive" self-defense. An act amounts to preventive self-defense if undertaken when the prospective cyber attacker either lacks the capability to conduct an armed cyber attack or, despite possessing the capability, has not yet formed an intention to carry one out.[56] Since cyber armed attacks are relatively easy to mount, it is the latter requirement that is the most likely to bar the taking of defensive actions.

A critical issue in light of the ease with which devastating cyber attacks can sometimes be mounted is whether non-State actors, such as terrorist groups, are capable of launching a cyber armed attack as a matter of law. Although the Koh speech did not directly address the issue, the U.S. government had previously taken the position that they were.[57] It is well accepted that the actions of a non-State actor may under limited circumstances be attributed to a State such that the victim-State may respond in self-defense against the State sponsor. The International Court of Justice made this point in the *Nicaragua* judgment when it stated that the notion of armed attack includes "'the sending by or on behalf of a State of armed bands, groups, irregulars or mercenaries, which carry out acts of armed force against another State of such gravity as to amount to' (*inter alia*) an actual armed attack conducted by regular forces, 'or its substantial involvement therein'."[58]

The more difficult question is whether a non-State actor's cyber operations that are not attributable to a State can nevertheless qualify as an armed attack justifying a defensive response at the level of a use of force against that non-State actor. The majority of the International Group of Experts were of the view that such attacks can so qualify, assuming the operations are conducted by an organized group (rather than isolated individuals), generate consequences of the requisite scale and effects, and are directed against a State[59] They based their conclusion on the reaction of the international community to major terrorist attacks, especially those of 9/11. States treated the terrorist attacks as armed attacks that could be responded to in self-defense despite the fact that the State support for the terrorists fell well below the

---

[55] *Id.*

[56] Some of the Experts adopted the position that a State that lacks the capability may nevertheless be deemed to possess it at the point when the defending State will not be able to defend itself effectively unless it acts immediately. Even in such cases, the State acquiring the means in question must have decided to use it before the right of self-defense matures.

[57] GGE Report, *supra* note **Error! Bookmark not defined.**, at 36.

[58] Nicaragua, *supra* note **Error! Bookmark not defined.**, para. 195.

[59] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 11 cmt. 5.

Nicaragua threshold, or was non-existent altogether.[60] The Experts rejected the approach adopted by the International Court of Justice in the *Wall* opinion and the *Armed Activities in the Congo* judgment.[61] In those cases, the Court seemingly took the position that some nexus with a State at the *Nicaragua* level is required before the group's actions can be deemed an armed attack.

IV. THE *JUS IN BELLO*

"Must attacks distinguish between military and nonmilitary objectives? Yes. The *jus in bello* principle of distinction applies to computer network attacks undertaken in the context of an armed conflict."

"Must attacks adhere to the principle of proportionality? Yes. The *jus in bello* principle of proportionality applies to computer network attacks undertaken in the context of an armed conflict."

"What do we do about 'dual-use infrastructure' in cyberspace? Unresolved."

"How should States assess their cyber weapons? States should undertake a legal review of weapons, including those that employ a cyber capability."[62]

At the heart of international humanitarian law lies the principle of distinction.[63] Codified in Additional Protocol I, it requires that "the Parties to the conflict shall at all times distinguish between the civilian population and combatants and between civilian objects and military objectives and accordingly shall direct their operations only against military objectives."[64] The applicability of this principal to cyber

---

[60] See discussion of this issue in Michael N. Schmitt, *Responding to Transnational Terrorism under the Jus ad Bellum: A Normative Framework*, in INTERNATIONAL LAW AND ARMED CONFLICT: EXPLORING THE FAULTLINES 157, 165–168 (Michael N. Schmitt & Jelena Pejic eds., 2007).
[61] Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory, Advisory Opinion, 2004 I.C.J. 136, para. 139 (July 9); Armed Activities on the Territory of the Congo (Dem. Rep. Congo v. Uganda), 2005 I.C.J.168, paras. 146–147 (Dec. 19). at.
[62] Koh Speech, *supra* note **Error! Bookmark not defined.**, at 5–8.
[63] Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 48. The principle derives from that set forth in the 1868 St. Petersburg Declaration: "the only legitimate object which States should endeavour to accomplish during war is to weaken the military forces of the enemy." Declaration Renouncing the Use, in Time of War, of Explosive Projectiles Under 400 Grammes Weight, Nov. 29/Dec. 11, 1868, 18 Martens Nouveau Recueil (ser. 1) 474.
[64] Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 48. That the principle is customary in nature is beyond question. For instance, the International Court of Justice has noted, "States must never make civilians the object of attack and must consequently never use

operations was acknowledged in the Koh speech and confirmed in the *Tallinn Manual*.[65]

International humanitarian law operationalizes the general principle of distinction by prohibiting attacks against specified protected persons and objects, imposing restrictions on how attacks may be conducted, and setting a limit on the extent of incidental harm to civilians and civilian objects that may be caused during an attack.[66] For instance, it is prohibited to attack civilians who are not directly participating in hostilities or civilian objects that have not been transformed into military objectives through either use or purpose.[67] What is of particular importance is that many of the rules governing the conduct of hostilities are framed in terms of "attacks". This term, which must be distinguished from the term "armed attack" in the *jus ad bellum* context, was the focus of great attention during the drafting of the Manual.

Additional Protocol I to the 1949 Geneva Conventions defines "attacks" as "acts of violence against the adversary, whether in offence or in defence."[68] The Experts unanimously agreed that although cyber operations are not violent in the sense of releasing kinetic energy, the term attack should logically be interpreted as extending to non-kinetic actions having violent *consequences*, specifically injury to or death of persons or damage to or destruction of objects.[69] However, they were sharply split as to whether the notion of attack included acts having consequences falling below that threshold. After three years of vigorous debate, the majority of the Experts adopted an interpretation that characterizes "interference with functionality" as damage to an object if "restoration of functionality requires replacement of physical components."[70] Most of these Experts would also characterize interference with functionality that necessitates re-installation of the operating system as damage.[71]

Consider the significance of this interpretation; cyber operations directed against civilian computer systems do not violate the prohibition on attacking civilian objects unless they qualify as an attack by virtue of their consequences. The paradigmatic case is a cyber psychological operation (PSYOP) that involves denial of services, but causes no physical damage. Similarly, the incidental effects of a cyber attack against a lawful military objective need not be considered when assessing proportionality

---

weapons that are incapable of distinguishing between civilian and military targets." Nuclear Weapons, *supra* note **Error! Bookmark not defined.**, para. 78.

[65] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 31.

[66] *See, e.g.,* Additional Protocol I, *supra* note **Error! Bookmark not defined.**, pt. IV, sec. I.

[67] *Id.,* arts. 51 & 52; I INTERNATIONAL COMMITTEE OF THE RED CROSS, CUSTOMARY INTERNATIONAL HUMANITARIAN LAW, (Jean-Marie Henckaerts & Louise Doswald-Beck eds., 2005), Rs. 1 & 7.

[68] Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 49(1).

[69] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 13 cmt. 3.

[70] *Id.*, R. 30.

[71] Id., R. 30 cmt. 11.

(discussed below) or the requirement to minimize civilian harm if those effects do not include physical damage or interference with functionality.[72]

The Experts also struggled with the importunate controversy over the meaning of the term "military objectives." As discussed, international humanitarian law requires an attacker to distinguish between military objectives and civilian objects; attacks are permissible only against the former. Civilian objects are defined in the negative as objects which do not qualify as military objectives.[73] Military objectives are "those objects which by their nature, location, purpose or use make an effective contribution to military action and whose total or partial destruction, capture or neutralization, in the circumstances ruling at the time, offers a definite military advantage."[74] For example, a military computer network and a civilian server used to transmit military data are both military objectives.

The United States, unlike most other States, takes the position that the aforementioned definition of military objectives encompasses not only objects that are "war-fighting and war-supporting," but also those that are "war-sustaining," such as oil-production facilities in a country that relies on oil export profits to finance its war effort.[75] Inability to agree on whether the concept of military objectives extends to war-sustaining objects has significant implications with respect to cyber operations because such targets tend to be especially vulnerable to cyber attack. The majority of the International Group of Experts rejected the US position, which was defended by a vocal minority.

---

[72] *See infra* sec. IV on proportionality. The requirement to take precautions in attack is codified for States Party in Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 57. Summarized, that article provides that an attacker must take all feasible measures to avoid collateral damage. Such measures include weapons, tactics, and target selection, as well as taking steps to verify the target and providing warnings when reasonable to do so. *See also* Customary International Humanitarian Law Study, *supra* note 67, Rs. 14–21.

[73] Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 52(1).

[74] *Id.*, art. 52(2).

[75] War-fighting refers to military equipment, such as military cyber attack systems. War-supporting objects are exemplified by a factory that produces war-fighting equipment. War-sustaining generally refers to economic targets, the destruction or neutralization of which would deprive the enemy of funds needed to carry on the war effort effectively. The most current US military international humanitarian law manual, the *Commander's Handbook on the Law of Naval Operations*, substitutes the phrase "war-fighting or war-sustaining capability" for "military action." U.S. NAVY/U.S. MARINE CORPS/U.S. COAST GUARD, THE COMMANDER'S HANDBOOK ON THE LAW OF NAVAL OPERATIONS, NWP 1-14M/MCWP 5-12.1/COMDTPUB P5800.7A para. 8.2 (2007).

As noted by Mr. Koh, international humanitarian law's rule of proportionality applies to cyber attacks conducted during an armed conflict.[76] The rule of proportionality is codified in Additional Protocol I.[77] As replicated in the *Tallinn Manual*, it provides that "[a] cyber attack that may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated is prohibited."[78] The rule is extraordinarily difficult to apply in practice because it requires a comparative evaluation of two dissimilar entities: collateral damage and military advantage.

A number of issues as to application of the rule to cyber operations posed challenges for the Experts. In particular, the Experts had to determine the types of harm to civilians and civilian objects that constitutes as collateral damage in the proportionality analysis. They agreed that the standard of harm qualifying a cyber operation as an attack applies equally when identifying collateral damage. For the majority, this means that an attacker need only consider civilian death, injury, damage, or destruction during a cyber attack on a lawful military objective; inconvenience, irritation, stress, or fear do not bear on the proportionality assessment. Moreover, the majority of the Experts agreed that the mere loss of data does not amount to collateral damage unless the loss interferes with the functionality of the civilian object in question. The same logic would hold true with regard to effects on civilians and civilian objects qualifying as damage in the context of the separate requirement to minimize collateral damage during an otherwise lawful attack.[79]

An issue that sometimes arises in discussions of the rule of proportionality is whether an attack's indirect effects count as collateral damage. The issue is especially relevant with regard to cyber operations because the interconnectivity of cyber infrastructure heightens the likelihood that an attack against a military objective might have bleed over effects into civilian systems. The International Group of Experts agreed that collateral damage is not limited to the direct effects of a cyber attack (the effects experienced by the target system). Instead, they adopted a foreseeability test in which

---

[76] Koh Speech, *supra* note **Error! Bookmark not defined.**, at 4.
[77] Additional Protocol I, *supra* note 15, arts. 51(5)(b) & 57(2)(iii). *See also* Second Protocol to the Hague Convention of 1954 for the Protection of Cultural Property in the Event of Armed Conflict, art. 7, Mar. 26, 1999, 2253 U.N.T.S. 212; Protocol (to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects) on Prohibitions or Restrictions on the Use of Mines, Booby-Traps and Other Devices, art. 3(3), Oct. 10, 1980, 1342 U.N.T.S. 168; *Id.* as amended on May 3, 1996, art. 3(8), 2048 U.N.T.S. 133; Customary International Humanitarian Law Study, *supra* note 67, rule 14.
[78] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 51.
[79] *Id.*, Rs. 52–58; Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 57; Customary International Humanitarian Law Study, *supra* note 67, Rs. 15–21.

any foreseeable collateral effects on civilian systems have to be factored into the proportionality calculation.[80]

Mr. Koh highlighted the existence of widespread "dual-use infrastructure" in cyberspace, referring to cyber infrastructure that is shared by military and civilian users. He asserted that shared use raises issues as to the applicability of the proportionality rule and the rule prohibiting the use of civilian objects in order to shield military objectives from attack.[81] By contrast, the International Group of Experts did not find dual-use cyber infrastructure to be uniquely problematic as a matter of law. On the contrary, while the targeting of dual-use infrastructure can be complex, the same is true as to attacks on other dual-use targets like airfields, railheads, electrical networks, and communication systems.[82]

Mr. Koh's reference to shielding merits clarification. The prohibition only applies to civilians and a limited number of specified civilian objects, such as hospitals. [83] It is not expressly prohibited to use civilian objects as such. In any event, civilian cyber infrastructure would, as a practical matter, generally need to be "used" to effectively shield military transmissions. Once that occurs, the shielding issue becomes moot since "[a]n object used for both civilian and military purposes—including computers, computer networks, and cyber infrastructure—is a military objective."[84]

To the extent that civilians or civilian objects (that are not being used for military ends) are harmed during an attack on dual-use cyber infrastructure, the harm factors into the proportionality assessment and the determination of whether precautionary measures have to be taken in order to minimize collateral damage. The International Group of Experts identified two problematic situations in this regard.

---

[80] Their position appears to have been adopted by the United States. U.S. Commander's Handbook, *supra* note 75, para. 8.11.4 (stating in the context of cyber operations that indirect effects of an attack may be one of the factors included when weighing anticipated incidental injury or death to protected persons).
[81] Koh Speech, *supra* note **Error! Bookmark not defined.**, at 8.
[82] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 39.
[83] Additional Protocol I, *supra* note **Error! Bookmark not defined.**, art. 51(7); Customary International Humanitarian Law Study, *supra* note 67, R. 97. *See also* Statute of the International Criminal Court, art. 8(2)(b)(xxiii), July 17, 1998, 2187 U.N.T.S. 90. As to prohibitions on using particular categories of persons or objects as shields, *see* Convention (III) Relative to the Treatment of Prisoners of War, art. 23, Aug. 12, 1949, 75 U.N.T.S. 135; Convention (IV) Relative to the Protection of Civilian Persons in Time of War, art. 28, Aug. 12, 1949, 75 U.N.T.S. 287; Additional Protocol I, *supra* note**Error! Bookmark not defined.**, art. 12(4).
[84] Tallinn Manual, *supra* note **Error! Bookmark not defined.**, R. 39; *see also* Hague IV Regulations, *supra* note 12, art. 27; Additional Protocol I, *supra* note 15, art 52(2); Customary International Humanitarian Law Study, *supra* note 47, at 32.

First, it is sometimes impossible to identify the parts of a dual-use network over which military transmissions pass. The Experts concluded that in such cases the entire network qualifies as a military objective, much like a road network in which only certain roads are used by the enemy.[85] Second, the Experts struggled with the use of social networks for military purposes. In recent conflicts, Twitter, Facebook, and other social media have been used to transmit military information. The Experts agreed that such use would transform those facets of the social media networks that are used for military purposes into military objectives. [86] However, the entire networks would not be subject to direct attack. They also emphasized that the rule of proportionality and the requirement to take precautions in attack would provide the social networks a degree of protection. And, of course, the issues of targetability, proportionality, and precautions only arise when the consequences of the cyber operations are such that the operations qualify as attacks.[87]

Although the International Group of Experts disagreed with the assertion that the law governing dual-use cyber infrastructure is unresolved, the Experts concurred with Mr. Koh's view that cyber weapons should be subject to a legal review.[88] This requirement has been codified in Article 36 of Additional Protocol I, which the International Group of Experts, going further than Mr. Koh, believed reflective (in part) of customary international law.[89] For the purposes of the *Tallinn Manual*, the Experts defined cyber weapons as any "cyber device, materiel, instrument, mechanism, equipment, or software used, designed, or intended to be used to conduct a cyber attack."[90]

A unique aspect of cyber weapons is that they are sometimes developed for immediate operational use without going through the standard development, acquisition, and review cycle. For instance, military cyber operators may discover a vulnerability in the enemy's cyber infrastructure and immediately develop malware capable of exploiting it. The Experts took the position that in such cases the lawyer who provides advice to the commander of the unit employing the malware is responsible for conducting the legal review.[91] Similarly, if significant changes are made to a previously reviewed cyber weapon, further legal review by the commander's

---

[85] Tallinn Manual, *supra* note 5, R. 39 cmt. 3.
[86] *Id.*, R. 39 cmt. 4.
[87] *Id.*
[88] Tallinn Manual, *supra* note 5, R. 48(a).
[89] The Experts were only willing to characterize the requirement to review means of warfare (i.e., weapons), not methods of warfare (i.e. tactics), as customary. Consequently, Rule 48(b) of the *Tallinn Manual* applies only to States Party to Additional Protocol I. Tallinn Manual, *supra* note 5, R. 48(b).
[90] Tallinn Manual, *supra* note 5, R. 41 cmt. 2. The reference to cyber attack is to an attack in the jus in bello sense (Rule 30), rather than armed attack as that term as used in the jus ad bellum (Rule 13).
[91] *Id.,* R. 48 cmt. 8.

lawyer is required before it may be employed.[92] However, minor changes that do not significantly alter the operational effect of a cyber weapon do not require a formal legal review.[93]

A legal review of a cyber weapon considers, *inter alia,* whether:

(i) it is, in its normal or intended circumstances of use, of a nature to cause superfluous injury or unnecessary suffering; (ii) it is by nature indiscriminate; (iii) its use is intended or may be expected to breach law of armed conflict rules pertaining to the environment to which the State is Party; and (iv) there is any *ad hoc* provision of treaty or customary international law that directly addresses it.[94]

The process would normally include a review of the technical description of the cyber weapon, as well as consideration of its likely targets, the desired effect on the targets for which it has been designed, the dynamic by which the effects will be achieved, the likely scope of the effects, and the cyber weapon's precision when striking targeted cyber infrastructure.

V. SOVEREIGNTY

"In this analysis, what role does State sovereignty play? . . . States conducting activities in cyberspace must take into account the sovereignty of other States, including outside the context of armed conflict."[95]

The Koh speech dealt with a number of issues beyond the *jus ad bellum* and *jus in bello*, including sovereignty and State responsibility. The U.S. Government's position on sovereignty mirrors that of the International Group of Experts, which found that "no State may claim sovereignty over cyberspace *per se*" and that "States may exercise sovereign prerogatives over any cyber infrastructure located on their territory, as well as activities associated with that cyber infrastructure."[96] Sovereignty is "the right [within a State's territory] to exercise . . ., to the exclusion of any other State, the functions of a State."[97] Those functions include the right to exercise legal and regulatory control over cyber infrastructure located on its territory. Territorial sovereignty also affords protection to cyber infrastructure under international law irrespective of whether it is owned privately or by the government.

---

[92] *Id.*
[93] *Id.,* R. 48 cmt. 9.
[94] *Id.,* R. 48 cmt. 10 (citations omitted).
[95] Koh Speech, *supra* note 4, at 6.
[96] *Id.,* R. 1 cmt. 1.
[97] Island of Palmas (Neth. v. U.S.), 2 R.I.A.A. 829, 838 (Perm. Ct. Arb. 1928).

In the exercise of its sovereign prerogatives, a State may shut down access to the Internet, so long as doing so complies with international human rights and telecommunications law. In particular, the Experts observed that "[t]he fact that cyber infrastructure located in a given State's territory is linked to the global telecommunications network cannot be interpreted as a waiver of its sovereign rights over that infrastructure."[98] Although they also agreed that a cyber operation violates a State's sovereignty if physical damage is caused to cyber infrastructure located in its territory, no consensus was reached as to whether the mere placement of malware causing no physical damage (as with malware designed to monitor activity) amounts to a violation.[99]

Sovereignty is the basis for the exercise of jurisdiction (the authority of a State to prescribe, enforce, and adjudicate) in international law. Consistent with general jurisdictional precepts, the *Tallinn Manual* provides that "a State may exercise its jurisdiction: (a) [o]ver persons engaged in cyber activities on its territory; (b) [o]ver cyber infrastructure located on its territory: and (c) [e]xtraterritorially, in accordance with international law."[100]

Two forms of territorial jurisdiction are especially significant in the cyber context—subjective and objective.[101] When a cyber operation has been initiated within a State's territory the state has subjective jurisdiction, irrespective of where the effects occur. Objective territorial jurisdiction grants a State jurisdiction over cyber operations initiated outside its territory mounted against cyber infrastructure within the territory.[102] The Experts recognized certain other potential bases for the exercise of extraterritorial jurisdiction over cyber activities. These, depending on the circumstances, include the nationality of the perpetrator (active personality), the nationality of the victim (passive personality), national security (protective principle), and violation of a universal norm of international law (universal jurisdiction).[103] The confluence of the various grounds for jurisdiction means that multiple States sometimes enjoy jurisdiction over a particular cyber incident.[104]

Sovereignty creates not only rights, but obligations. Accordingly, the *Tallinn Manual* provides that "[a] State shall not knowingly allow the cyber infrastructure located in its territory or under its exclusive governmental control to be used for acts that adversely and unlawfully affect other States."[105] This principle is well established in international

---

[98] Tallinn Manual, *supra* note 5, R. 1 cmt. 10.
[99] Tallinn Manual, *supra* note 5, R. 1 cmt. 6.
[100] *Id.,* R. 2.
[101] *Id.,* R. 1 cmt. 6.
[102] *Id.*
[103] *Id.,* R. 1 cmt. 8.
[104] *Id.,* R. 1 cmt. 9.
[105] *Id.,* R. 5.

law. In its very first case, *Corfu Channel*, the International Court of Justice held that a State may not "allow knowingly its territory to be used for acts contrary to the rights of other States."[106] The holding was consistent with that in the celebrated *Trail Smelter* case, in which the arbitral tribunal noted that "under the principles of international law. . . no State has the right to use or permit the use of its territory in… a manner as to cause injury. . . in or to the territory of another or the properties or persons therein, when the case is of serious consequence. . . ."[107]

The obligation unquestionably attaches whenever the cyber operations in question are underway and the State knows of them. For instance, a State would be obligated to take feasible measures to end cyber attacks launched by a terrorist group from its territory against other States. The duty extends to situations in which only private entities, such as Internet service providers, are capable of taking remedial action. In such cases, the State must act to compel those entities to do so. [108] The Experts differed with regard to the principle's application to prospective acts.[109] Whereas some were of the view that a State must take reasonable measures to ensure the harmful cyber activities are not carried out from its territory, others suggested that no such duty exists in international law.

During an international armed conflict, the law of neutrality governs these situations. Drawn in great part from the 1907 Hague Conventions and now customary in character, the law of neutrality balances the rights and obligations of neutral and belligerent States during armed conflicts.[110] Certain of its rules are especially germane to cyber operations.

First, "[t]he exercise of belligerent rights by cyber means in neutral territory is prohibited."[111] The prohibition on actions by parties to a conflict would encompass both conducting cyber operations from neutral territory and taking remote control of cyber infrastructure located in that territory and using it to conduct belligerent cyber operations.[112] Second, "[a] neutral State may not knowingly allow the exercise of belligerent rights by the parties to the conflict from cyber infrastructure located in its

---

[106] Corfu Channel Case (U.K v. Alb.) 1949 I.C.J. 4, 22.
[107] Trail Smelter Case (U.S. v. Can.), 3 R.I.A.A. 1905, 1965 (1941).
[108] Tallinn Manual, *supra* note 5, R. 5 cmt. 9.
[109] Id., R. 5 cmt. 7.
[110] *See generally* Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons in Case of War on Land, Oct.18, 1907, 36 Stat. 2310 [hereinafter Hague Convention V]; Convention (XIII) Concerning the Rights and Duties of Neutral Powers in Naval War, Oct. 18, 1907, 36 Stat. 2415.
[111] Tallinn Manual, *supra* note 5, rule 92. This rule is based on Hague Convention V, *supra* note 109, arts. 2 & 3, and Hague Convention XIII, *supra* note 84, arts. 2 & 5.
[112] Tallinn Manual, *supra* note 5, R. 92 cmt. 2.

territory or under its exclusive control."[113] The Experts agreed that an exception to this rule applies in the case of "public, internationally and openly accessible networks, such as the internet."[114] Should a neutral State decide to impose restrictions on the use of such a network, it must do so impartially.[115] Third, "[i]f a neutral State fails to terminate the exercise of belligerent rights on its territory, the aggrieved party to the conflict may take such steps, including by cyber operations, as are necessary to counter that conduct."[116] Before a belligerent may act pursuant to this rule, the violation of neutral territory involved must be "serious."[117]

VI. STATE RESPONSIBILITY

"Are States responsible when cyber acts are undertaken through proxies? . . . Yes. States are legally responsible for activities undertaken through 'proxy actors,' who act on the State's instructions or under its direction or control."

"How do we address the problem of attribution in cyberspace?" Unresolved.[118]

The *Tallinn Manual* includes a number of rules drawn from the law of State responsibility, which the U.S. Government and the International Group of Experts agreed applies in cyberspace. In great part, they reflect relevant aspects of the International Law Commission's Articles on State Responsibility.[119] Although the Articles are not hard law, the document, which the General Assembly adopted in 2001, was considered by the Experts to accurately capture the customary international law of state responsibility.[120]

---

[113] Tallinn Manual, *supra* note 5, R. 93 (based on Hague Convention V, *supra* note 109, art. 5).
[114] *Id.,* R. 93 cmt. 3.
[115] *Id.,* R. 93 cmt. 3 (citing Hague Convention V, *supra* note 84, art. 9).
[116] *Id.,* R. 94.
[117] *Id.,* R. 94 cmt. 3; *see also* INTERNATIONAL INSTITUTE OF HUMANITARIAN LAW, SAN REMO MANUAL ON INTERNATIONAL LAW APPLICABLE TO ARMED CONFLICTS AT SEA (Louise Doswald-Beck ed., 1995), *available at http://www.icrc.org/ihl.nsf/FULL/560?OpenDocument*, R. 22. For a belligerent to act, the conduct must also "represent an immediate threat to the security of the aggrieved party and there must be no feasible and timely alternative to taking action on neutral territory." Tallinn Manual, *supra* note 5, R. 94 cmt. 4.
[118] Koh Speech, *supra* note 4, at 6, 8.
[119] International Law Commission, Responsibility of States for Internationally Wrongful Acts, G.A. Res. 56/83 annex, U.N. Doc. A/RES/56/83 (Dec. 12, 2001) [hereinafter Articles of State Responsibility].
[120] A three-year research project sponsored by the NATO Cooperative Cyber Defence Centre of Excellence will examine the subject of State responsibility for cyber operations in much greater depth. The author will serve as director of the project.

Under international law, "[a] State bears international legal responsibility for a cyber operation attributable to it and which constitutes a breach of an international obligation."[121] The obligation may derive from either treaty or customary international law, and its breach can consist of an omission or commission.[122]

Certain acts are self-evidently attributable to a State. Any wrongful act or omission undertaken by organs of the State, including *ultra vires* acts performed in an apparently official capacity, are automatically attributable to that State.[123] Similarly, acts or omissions of persons or entities authorized to act with governmental authority are attributable to the State granting that authority.[124] In the cyber context, the most common example is that of private Computer Emergency Response Teams authorized to defend government cyber infrastructure and networks.[125]

The Koh speech narrowed in on attribution of the cyber activities of non-State actors. According to the Articles on State Responsibility, "[t]he conduct of a person or group of persons shall be considered an act of a State under international law if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct."[126] The International Group of Experts noted the lack of agreement as to the precise level of control necessary for attribution of a non-State actor's cyber operations to a State.[127] Although acknowledging that an "overall control" test finds some support in the International Criminal Tribunal for the Former Yugoslavia's Appeals Chamber judgment in *Tadić*,[128] the majority of the Experts, drawing on International Court of Justice jurisprudence, took the position that a State must have "effective control" over non-State actors for attribution to occur.[129] To reach the higher threshold, the State "needs to have issued specific instructions or directed or controlled a particular operation. . . .

---

[121] Tallinn Manual, *supra* note 5, R. 6. *See also* Articles of State Responsibility, *supra* note 118, art. 2.
[122] Tallinn Manual, *supra* note 5, R. 6 cmt. 8.
[123] Articles of State Responsibility, *supra* note 94, art. 4(1)–(2). On *ultra vires* acts, see *id.,* commentary accompanying art. 4. *Accord* Tallinn Manual, *supra* note 5, R. 6 cmt. 6.
[124] Articles of State Responsibility, *supra* note 92, art. 5. *Accord* Tallinn Manual, *supra* note 5, R. 6 cmt. 8.
[125] Tallinn Manual, *supra* note 5, R. 6 cmt. 8.
[126] Articles of State Responsibility, *supra* note 118, art. 8.
[127] Tallinn Manual, *supra* note 5, R. 6 cmt. 10.
[128] Prosecutor v. Tadić, Case No. IT-94-1-A, Appeals Chamber Judgment ¶¶ 131, 145 (Intl'l Crim. Trib. for the Former Yugoslavia July 15, 1999).
[129] Nicaragua, *supra* note 18, ¶ 115; Application of the Convention on the Prevention and Punishment of the Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro.), 2007 I.C.J. 43¶¶ 399–405. The Experts noted that the Tadić decision bore on the issue of whether an international armed conflict was underway, rather than State responsibility *per se*. *See* Tallinn Manual, *supra* note 5, R. 6 cmt. 11.

Merely encouraging or otherwise expressing support for the independent acts of non-State actors does not meet the…threshold."[130] As an example, State A would bear State responsibility for cyber operations conducted by a non-State group against State B if A provided cyber target data and the malware necessary to carry out the operations.

In his speech, Mr. Koh pointed to the "ability to mask one's identity and geography in cyberspace and the resulting difficulties of timely, high-confidence attribution." There are two facets to this issue. First, although the ability of an advanced cyber power to accurately identify the originator of a cyber operation is significantly greater than realized by the general public,[131] in certain cases tracing an operation to a State may be problematic. Second, it can sometimes be difficult to link a State to cyber operations conducted by a non-State actor. Cognizant of these challenges, the International Group of Experts offered guidelines designed to inform the process of determining whether an act or omission may be attributed to a State as a matter of law. In their view, "[t]he mere fact that a cyber operation has been launched or otherwise originates from governmental cyber infrastructure is not sufficient evidence for attributing the operation to that State but is an indication that the State in question is associated with the operation."[132]Moreover, "[t]he fact that a cyber operation has been routed via the cyber infrastructure located in a State is not sufficient evidence for attributing the operation to that State."[133] However, as noted by Mr. Koh, the hurdles tend to be technical or policy-oriented in nature, rather than legal.

VII. CONCLUDING OBSERVATIONS

It is hardly a jurisprudential epiphany to assert that international law applies fully formed to activities in cyberspace. This is particularly so once it is grasped that cyber activities involve individuals using tangible objects in physical domains that have long been subject to international law's normative architecture. It is quite remarkable, therefore, that it has taken States so long to state the obvious, and that the international legal community seemed to struggle so mightily with a rather straightforward issue.

In fact, the International Group of Experts who drafted the *Tallinn Manual* found no relevant body of law that was inapplicable to cyber activities. Be that as it may, the

---

[130] Tallinn Manual, *supra* note 5, R. 6 cmt. 11.
[131] "Potential [cyber] aggressors should be aware that the United States has the capacity to locate them and hold them accountable…." Armed Forces Press Service, *Panetta Spells Out DOD Roles in Cyberdefense* (Oct. 11, 2012), http://www.defense.gov//News/NewsArticle.aspx?ID=118187 (quoting Leon Panetta).
[132] Tallinn Manual, *supra* note 5, R. 7.
[133] *Id.*, R. 8.

unique nature of cyber activities, in particular the fact that they may have devastating results without causing physical injury or damage, can lead to interpretive uncertainty. The Koh speech and the *Tallinn Manual* are but initial forays into the demanding process of exploring how the extant norms of international law will apply in cyberspace. But the long overdue journey has at least finally begun.

INTERNATIONAL LAW STUDIES
PUBLISHED SINCE 1901
U.S. NAVAL WAR COLLEGE

The Geography of Cyber Conflict: Through a Glass Darkly

Ashley Deeks

89 INT'L L. STUD. 1 (2013)

Volume 89 2013

The Geography of Cyber Conflict: Through a Glass Darkly

Ashley Deeks[*]

I. INTRODUCTION

Imagine an Israeli Air Force jet is shot down in international airspace just outside Turkish airspace. Imagine further that the Israel Defense Forces (IDF) and Israeli intelligence services quickly ascertain with a high level of confidence that a Hezbollah cell located in Turkey was responsible for the shoot-down. Israel now confronts a difficult question: having suffered an armed attack, may it use force in self-defense against a non-state actor in the territory of a state with which it is not in an armed conflict and that was not the author of the attack?

In previous work, I have argued that Israel may only take action in Turkish territory against Hezbollah if it has Turkish consent or if it determines that Turkey is unwilling or unable to suppress the threat posed by Hezbollah.[1] This "unwilling or unable" test, which has analogical roots in

* Associate Professor of Law, University of Virginia School of Law.

1. *See generally* Ashley Deeks, *"Unwilling or Unable": Toward a Normative Framework for Extraterritorial Self-Defense*, 52 VIRGINIA JOURNAL OF INTERNATIONAL LAW 483 (2012).

the law of neutrality,[2] serves as an attempt to balance the security of the State that suffered the attack (the "victim State") against the sovereignty of the State from which the non-State actor launched the attack (the "territorial State"). The test also reflects the international community's interest in reducing to the lowest level feasible inter-State conflict (and State uses of force in self-defense).

Imagine now that the IDF learns that its air force's command and control center is being severely compromised electronically and has begun to send faulty coordinates to all of the IDF's military aircraft, including those currently airborne. As a result of the cyber attack,[3] the IDF loses communications with two of its jets, which crash into the Mediterranean Sea. Israel has a high level of confidence that several servers in Turkey are the source of the ongoing attack; additionally, the offending code behind the attack has Hezbollah's digital fingerprints on it and Israel has intelligence that Hezbollah has been trying for several years to conduct just such an attack. Assuming that Israel has the technological capacity to disable the Turkish servers currently routing the attack and believes that such an action is the only way to stop this attack, may Israel disable those Turkish servers (using cyber or kinetic tools)? What, if anything, must it do first?

This article argues that the "unwilling or unable" test applies to this scenario as well, although the issues facing Israel and Turkey in the two scenarios are different in important ways. Other scholars have suggested that the "unwilling or unable" test is relevant in the cyber context,[4] but no

---

2. J.M. SPAIGHT, WAR RIGHTS ON LAND 482 (1911) ("[W]here the neutral cannot or will not enforce its rights, then the belligerent is fully entitled to prevent the violation permitted by the neutral redounding to his disadvantage.").

3. This paper uses the phrase "cyber attacks" generically to refer to acts that "alter, disrupt, deceive, degrade, or destroy computer systems or networks or the information and/or programs resident in or transiting these systems or networks," TECHNOLOGY, POLICY, LAW AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES 1 (WILLIAM OWENS, KENNETH DAM, & HERBERT LIN, EDS., 2009) [hereinafter OWENS ET AL.]. A particular cyber attack may or may not constitute a use of force or armed attack, as those terms are used in the *jus ad bellum* sense.

4. *See* Duncan Hollis, *Why States Need an International Law for Information Operations*, 11 LEWIS & CLARK LAW REVIEW 1023, 1050 (2007) ("International law contemplates that the [State injured by information operations] would notify the State from whose territory it believes the IO originated and request that State put a stop to it. The requested State is expected to comply with such requests . . . . Only if the requested State is unable or unwilling to stop the IO can the aggrieved State take counter-measures (or perhaps exercise a right of self-defense against the requested State . . . ."); George Walker, *Information Warfare and Neutrality*, 33 VANDERBILT JOURNAL OF TRANSNATIONAL LAW 1079, 1199 (2000)

scholar has analyzed how a State actually should or would apply that test, how the test's application will differ in the cyber and non-cyber contexts and what that divergence teaches us about conflict in the cyber realm. This paper addresses those three issues, focusing on situations in which the cyber activity rises to the level of a cyber armed attack (rather than cyber activities that fall below that threshold). At the same time, it highlights the particular importance of State practice in adopting and expounding the use of the "unwilling or unable" test in the cyber context. Indeed, news reports suggest that the United States is wrestling mightily to determine when it is appropriate and lawful to take cyber action outside the boundaries of its own networks.[5] Establishing State-to-State expectations about what types of cyber activities will trigger what types of responses will provide important incentives for ostensibly neutral States to take steps to protect their computer networks while minimizing the likelihood of inter-State misunderstandings that lead to unnecessary conflict in the cyber or non-cyber realms.[6]

Part II describes the "unwilling or unable" test, including relevant factors that States should use in assessing whether another State has met that test. Part III applies those factors to the cyber context. Part IV considers how the U.S. government may be approaching these issues. Part V concludes.

II. THE "UNWILLING OR UNABLE" TEST

In the wake of the September 11 attacks, the United States concluded that it was in an international armed conflict with Al Qaeda, a non-State actor. Perhaps the most controversial aspect of this claim was the implicit argument that the United States therefore could use force against members of Al Qaeda anywhere they appeared. This concept resulted in the much-

("The 'means at a neutral's disposal' principle should be the test for a neutral's duty for belligerents' IW [information warfare] incursions; the neutral should be held to apply means at its disposal to detect and repel these incursions. Such being the case, the correlative right of a belligerent aggrieved by IW incursions should be that the belligerent may take such actions as are necessary in the territory of a neutral that is unable (or perhaps unwilling) to counter enemy IW force activities making unlawful use of that territory, a principle from the law of naval warfare.").

5. Ellen Nakashima, *Pentagon Proposes More Robust Role for Its Cyber-Specialists*, WASHINGTON POST, Aug. 9, 2012.

6. *See* OWENS ET AL., *supra* note 3, at 318 (explaining rationales behind legal regimes that regulate the development and use of certain kinds of weapons).

maligned idea of the "global war on terror." The United States later took care to clarify that its international armed conflict claim did not mean that it would use force in all countries in which members of Al Qaeda appeared. Rather, the United States asserted that it would only use force in those countries that either gave the United States consent to do so or were "unwilling or unable" to suppress the threat itself.[7] Nor is the United States the only State to employ the "unwilling or unable" test when evaluating the legality of using force against non-State actors in another State's territory. Israel, Russia and Turkey all have cited the test in recent years.[8] Scholars, too, have described the "unwilling or unable" test as the applicable test in this situation,[9] though some contest that the test has any status in international law.[10]

---

7. John B. Bellinger III, Legal Adviser, U.S. Department of State, Address at the London School of Economics: Legal Issues in the War on Terrorism (Oct. 31, 2006); Harold Koh, Legal Adviser, U.S. Department of State, Address at the Annual Meeting of American Society of International Law (Mar. 25, 2010).

8. *See* Deeks, *supra* note 1, at 486–87 (listing other States' claims).

9. *See, e.g.*, NOAM LUBELL, EXTRATERRITORIAL USE OF FORCE AGAINST NON-STATE ACTORS 42 (2010) (reciting the "unwilling or unable" test as the correct test for determining when a victim State may take measures against non-State actors in the territorial State); YORAM DINSTEIN, WAR, AGGRESSION, AND SELF-DEFENSE 217 (3d ed. 2001) ("Extraterritorial law enforcement is a form of self-defence, and it can be undertaken by Utopia against armed bands or terrorists inside Arcadian territory, in response to an armed attack unleashed by them from that territory. Utopia is entitled to enforce international law extraterritorially only when Arcadia is unable or unwilling to prevent repetition of that armed attack."); Carsten Stahn, *Terrorist Acts as "Armed Attack": The Right to Self-Defense, Article 51 (1/2) of the UN Charter, and International Terrorism*, 27 FLETCHER FORUM OF WORLD AFFAIRS JOURNAL 35, 47 (2003); Greg Travalio & John Altenburg, *Terrorism, State Responsibility, and the Use of Military Force*, 4 CHICAGO JOURNAL OF INTERNATIONAL LAW 97, 116 (2003) ("[S]hould a State be unwilling or unable to prevent its territory from being used as a sanctuary or base of operations by a transnational terrorist organization, a State threatened with an imminent attack by such an organization may . . . engage in a self-defense use of force to deal with this threat."); Alberto Coll, *The Legal and Moral Adequacy of Military Responses to Terrorism*, 81 AMERICAN SOCIETY INTERNATIONAL LAW PROCEEDINGS 297, 305 (1987) ("[O]nce it becomes reasonably evident that the harboring State is unable or unwilling to act, the injured State should be free to use the minimum of force required to stop the terrorist threat."); Ian Brownlie, *International Law and the Activities of Armed Bands*, 7 INTERNATIONAL & COMPARATIVE LAW QUARTERLY 712, 732 (1958) ("Military action across a frontier to suppress armed bands, which the territorial sovereign is unable or unwilling to suppress, has been explained in terms of legitimate self-defense on a limited number of occasions in the present century."); Tatiana Waisberg, *Colombia's Use of Force in Ecuador Against a Terrorist Organization*, 12 ASIL Insights (2008), *available at* http://www.asil.org/insights080822.cfm ("State practice and the UN Security Council's

Although this test plays a significant role in regulating the geography of an armed conflict (or the geographic location of a State's response to an armed attack), its precise substantive and procedural content remains unclear. Must the victim State request assistance from the territorial State before using force against the non-State actor in the territorial State? By what standards should the victim State evaluate the territorial State's proposed means to address the threat and its capacity to do so? In the context of an armed conflict, what level of threat justifies taking action, using the "unwilling or unable" theory? International law currently does not answer these questions.

As complicated as an "unwilling or unable" inquiry may be in the non-cyber context, it becomes even more complicated in the cyber context. First, it is far easier to employ the cyber infrastructure of third States for hostile ends than it is to employ the physical territory of those States to commit conventional hostile acts. States and non-State actors that are engaged in armed conflicts or that are intent on committing armed attacks tend to operate from a single State or from a limited number of States, by virtue of cost, politics, logistics and terrain. In contrast, those same States and non-State actors are able to employ the cyber infrastructure of a much larger number of third States in forcible pursuit of their goals. Second, the difficulty of attribution in the cyber context is well-known.[11] As a result, there will be many situations in which the victim State can ascertain that a third country's servers are being used for hostile purposes but be unable to identify with certainty the actual authors of the attacks. In some cases, the victim State may not even be able to identify the geographic origin of a given cyber attack.[12] This stands in contrast to kinetic activities outside the cyber context, where the victim State often is able to identify the authors of the armed attacks and their locations, using well-established intelligence and investigatory resources. Third, the increased anonymity of cyberspace

---

actions after the September 11 attacks may, however, indicate a trend toward recognizing that a State that suffers large-scale violence perpetrated by non-State actors located in another State has a right to use force in self-defense when . . . that other State proves unwilling or unable to reduce or eliminate the source of the violence.").

10. *See* Kevin Jon Heller, *The Unwilling or Unable Standard for Self-Defense*, OPINIO JURIS, (Sept. 17, 2011), http://opiniojuris.org/2011/09/17/the-unwilling-or-unable-standard-for-self-defense-against-non-State-actors/ (rejecting "unwilling or unable" test as customary international law on basis that there is insufficient State practice and evidence of *opinio juris*).

11. *See, e.g.*, Jack L. Goldsmith, *The New Vulnerability*, NEW REPUBLIC (June 7, 2010).

12. OWENS ET AL., *supra* note 3, at 294.

may mean growth in the number of actors that seek to use cyber attacks. The deterrence that accompanies the fear of getting caught is reduced because the chance of being held accountable is lower.

Before turning to the cyber scenarios in which a State will need to employ the "unwilling or unable" test, however, it is important to clarify several assumptions in this article. First, this piece assumes that cyber attacks that produce effects similar to those of kinetic military actions will constitute "armed attacks" that trigger the victim State's right of self-defense.[13] Second, it assumes that non-State actors may be authors of armed attacks, even when those attacks are not attributable to a State.[14] Third, it assumes that, in the context of an international armed conflict, it would not violate the law of neutrality for a neutral State to allow a belligerent State to use, or not prevent it from using, its public internet and communications networks as a conduit for a cyber attack.[15] It assumes, however, that neutrality law would prohibit a neutral State from allowing a State or non-State actor to use its tangible computer equipment or operating systems, including servers, to host those attacks.[16] This means that a victim State, in responding to

13. *See, e.g.*, Michael Schmitt, *Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework*, 37 COLUMBIA JOURNAL OF TRANSNATIONAL LAW 885, 929 (1998–99) (discussing possibility that computer network attacks could constitute "armed attacks"); Nils Melzer, *Cyberwarfare and International Law* 13, UNIDIR Resources (2011) (stating that cyber operations have the qualitative capacity to qualify as an armed attack within the meaning of UN Charter Article 51).

14. *See* Deeks, *supra* note 1, at 492–93 (describing three schools of thought on this question).

15. This follows from Hague Convention (V) Respecting the Rights and Duties of Neutral Powers and Persons during War on Land, art. 8. Oct 18, 1907, 36 Stat. 2310, 1 Bevans 654 [hereinafter Hague V], which States that a neutral power is not required to "forbid or restrict the use on behalf of the belligerents of telegraph or telephone cables or of wireless telegraphy apparatus belonging to it or to companies or private individuals."

16. *Id.* Allowing particular servers within the neutral State to host attacks is more closely akin to allowing a belligerent to move munitions of war across neutral territory or furnishing military supplies to a belligerent, which Hague V would prohibit. This seems to be the approach taken by the U.S. Department of Defense in 1999 in its Assessment of International Legal Issues in Information Operations. That document provides, "[U]se of a nation's communications networks as a conduit for an electronic attack would not be a violation of its sovereignty . . . . A transited State would have somewhat more right to complain if the attacking State obtained unauthorized entry into its computer systems as part of the communications path to the target computer. It would be even more offended if malicious logic directed against a target computer had some harmful effect against the transited State's own equipment, operating systems, or data." *See also* TALLINN MANUAL, Rule 92 (Michael Schmitt ed., forthcoming 2013); Eric Jensen, *Sovereignty and Neutrality in*

a cyber armed attack against it, would not violate Article 2(4) simply by directing its response through a third State's public communications channels. The victim State would trigger Article 2(4), however, if it damaged a server hosted in that third State.[17] Fourth, it assumes that the victim State will be able to direct its response in a manner consistent with both the *jus ad bellum* and the laws of armed conflict, including the principles of distinction and proportionality.[18] Finally, it assumes that, as a matter of policy, the victim State will conduct its responses to a cyber armed attack in the cyber realm, although there is no legal requirement that it do so.[19]

There are at least three scenarios in which a State that has suffered a cyber attack may seek to take responsive (forcible) action in a third State's territory and therefore will need to assess the third State's willingness and ability to take action to address that cyber attack. First, a State may be fighting another State in an international armed conflict, where the State's opponent has launched a cyber attack from a third State's territory. In international armed conflict, the laws of neutrality apply. Assuming that the third State is neutral in the international armed conflict, the laws of neutrality require the neutral State to prevent its territory from being used by a belligerent as a place from which to launch attacks.[20] If a belligerent nevertheless is initiating or conducting cyber attacks against another belligerent using the cyber infrastructure of a neutral State, the neutral State must

---

Cyber Conflict, 35 FORDHAM INTERNATIONAL LAW JOURNAL 815, 826–27 (2012) (arguing that the law of neutrality would require a neutral State to prevent a belligerent from initiating or facilitating an attack within neutral territory, but not to prevent the mere passage of malware or malicious code over its public cyber infrastructure); Melzer, *supra* note 13, at 20 (reasoning that neutral States can be expected to prevent belligerents from conducting "cyber hostilities" from within neutral territory but not the "routing of belligerent cyber operations through their publicly accessible communications infra-structure"); *but see* Davis Brown, *A Proposal for an International Convention to Regulate the Use of Information Systems in Armed Conflict*, 47 HARVARD INTERNATIONAL LAW JOURNAL 179, 210 (2006) (arguing that even allowing the transit of malicious code over a State's public internet infrastructure would violate that State's neutrality obligations).

17. Such an act would be akin to severing a telephone wire in a State, interrupting general telecommunications in that State.

18. Whether this requires the victim State to identify with certainty the nature and identity of the cyber attacker is not clear.

19. The U.S. cyber security strategy preserves the right to respond kinetically to a cyber armed attack. However, given the level of caution with which the U.S. government seems to be proceeding in crafting doctrine for cyber responses, it seems reasonable to assume that using kinetic force against a cyber armed attack, particularly in a third State's territory, would occur only in an extreme case.

20. Hague V, arts. 2, 4; TALLINN MANUAL, *supra* note 16, at Rule 93.

make efforts to terminate that use. If the neutral State is unwilling or unable to stop that belligerent, the belligerent's opponent may take forcible measures within the neutral State to do so.[21]

Second, a State may be in a non-international armed conflict, fighting against a non-State actor whose operations are primarily based either within that State or within a foreign State's territory. The non-State actor may undertake cyber actions during that conflict that utilize systems located in foreign States. In this case, one may reason by analogy to the law of neutrality to assert that the State fighting the non-international armed conflict may take measures in that foreign State to suppress the non-State actor's cyber attacks where the foreign State is unwilling or unable to do so itself.[22] The United States appears to believe that this is the appropriate test to apply in the context of kinetic armed conflicts against non-State actors that transcend a single State's borders.[23] It is not clear whether a victim State could respond forcibly to *any* cyber uses of force emanating from the third State, or if the victim State only could respond forcibly to those cyber uses of force that rise to the level of a cyber armed attack.[24]

---

21. *See* SPAIGHT, *supra* note 2, at 482; John Norton Moore, *Legal Dimensions of the Decision to Intercede in Cambodia*, 65 AMERICAN JOURNAL OF INTERNATIONAL LAW 38, 51 (1971) ("It is well established in customary international law that a belligerent Power may take action to end serious violations of neutral territory by an opposing belligerent when the neutral Power is unable to prevent belligerent use of its territory . . . ."); TALLINN MANUAL, *supra* note 16, at Rule 94.

22. Melzer, *supra* note 13, at 21 ("Strictly speaking, the law of neutrality applies only in international armed conflict. Arguably, however, the pragmatic logic of its core principles has already found its way into the practice of non-international armed conflicts as well."); International Committee of the Red Cross Official Statement of 8 March 2001 to the United Nations High Commissioner for Refugees Global Consultations on International Protection ("It is the ICRC's view that [Hague Convention V] can also be applied by analogy in situations of non-international armed conflicts, in which combatants either from the government side or from armed opposition groups have fled into a neutral State.").

23. Koh, *supra* note 7; John Brennan, Assistant to the President for Homeland Security and Counterterrorism, Address at Harvard Law School: Strengthening Our Security by Adhering to Our Values (Sept. 16, 2011), http://www.whitehouse.gov/the-press-office/2011/09/16/remarks-john-o-brennan-strengthening-our-security-adhering-our-values-an.

24. The United States generally asserts that virtually all uses of force constitute armed attacks that trigger a State's right of self-defense. *See* William H. Taft IV, *Self-Defense and the* Oil Platforms *Decision*, 29 YALE JOURNAL OF INTERNATIONAL LAW 295, 299 (2004) (rejecting idea that attacks must rise to certain level of severity in order to qualify as armed attacks). Many States disagree with that position, however, and thus would have to confront how to respond to a use of cyber force short of an armed attack, launched by a non-

Third, a State that is not fighting an ongoing non-international armed conflict nevertheless may suffer a cyber armed attack from a non-State actor (or face an imminent threat thereof). This armed attack would trigger the victim State's right of self-defense under Article 51 of the U.N. Charter.[25] The victim State would then have to assess whether it was necessary to use force in self-defense against that non-State actor and, secondarily, whether it was necessary to use force *in that particular foreign State* against that non-State actor. If the territorial State is both willing and able to suppress the threat posed by that actor, it would not be necessary (and therefore would not be lawful) for the victim State to use force within the territorial's borders.

In each of these three scenarios, the victim State will be required to examine whether the territorial State can and will take action to halt or mitigate the attacks affecting the victim State. Although the test itself has traction in international law, its lack of substantive and procedural content makes it harder to apply and less legitimate as a restraining force in international relations. I previously suggested five principles, drawn from historical practice, that would help guide the test's application. These principles include the requirements that the victim State (1) prioritize cooperation or consent with the territorial State, rather than unilateral use of force; (2) ask the territorial State to address the threat and give it adequate time to respond; (3) reasonably assess the territorial State's capacity and control within the relevant region; (4) reasonably asses the territorial State's proposed

---

State actor extraterritorially. Those States might conclude that, absent an armed attack that triggers Article 51, the States cannot take any action in response that would violate Article 2(4) of the Charter. In practice, the United States actually may be imposing policy constraints on itself that bring it closer to that position. In his Harvard speech, John Brennan noted that "[b]ecause we are engaged in an armed conflict with al-Qa'ida, the United States takes the legal position that . . . we have the authority to take action against al-Qa'ida and its associated forces without doing a separate self-defense analysis each time." However, he also stated, "In practice, the U.S. approach to targeting in the conflict with al-Qa'ida is far more aligned with our allies' approach than many assume. This Administration's counterterrorism efforts outside of Afghanistan and Iraq are focused on those individuals who are a threat to the United States, whose removal would cause a significant – even if only temporary – disruption of the plans and capabilities of al-Qa'ida and its associated forces. Practically speaking, then, the question turns principally on how you define 'imminence.'" Brennan, *supra* note 23.

25. In the context of scenarios 2 and 3, "any action taken against [non-State actors] may raise issues about violating the sovereignty of that nation and its rights and obligations with respect to terrorist operations from or through its territory." OWENS ET AL., *supra* note 3, at 274.

means to suppress the threat; and (5) evaluate its prior interactions with the territorial State. Part III takes up these factors and applies them in the context of cyber attacks.

III. APPLYING THE TEST'S FACTORS TO CYBER ATTACKS

A. Preference for consent or cooperation

In the ideal situation, a victim State will approach the territorial State and inform the latter of the fact of the imminent or actual armed attack and its reasons for believing that the attacker is employing the victim State's infrastructure to commit the attacks. It then will seek consent to take action (whether forcible or not) to suppress the attacks emanating from the territorial State's computer systems. When it acts pursuant to and consistent with that consent, the victim State will not violate Article 2(4) of the Charter or the customary principle of non-intervention. Examples of such consent are not hard to find, particularly outside the cyber realm: Iraq previously allowed Turkey to use force in Iraq against a Kurdish terrorist group (the Kurdistan Workers' Party), and the United States reportedly is using force in Somalia and Yemen against members of Al Qaeda and associated forces with the consent of those governments.[26] Even if the territorial State is reluctant to let the victim State operate alone in its computer systems, there may well be opportunities for the two States to work cooperatively to suppress the threat.

This approach has several advantages. First, it minimizes the chance of cyber clashes between the victim and territorial States, and reduces the likelihood that those States find themselves working at cross-purposes against the cyber attacker. Second, this type of cooperation has the potential to enhance the victim State's own operations, to the extent that the territorial State has a deeper knowledge of its own computer systems, relationships with private sector companies whose computers the attacker may be using to facilitate the attack, and relevant information about past penetrations into the victim (or territorial) State's systems. Third, this cooperation and the corresponding information that it receives from the territorial State may help the victim State limit the collateral damage from its response, a

26. Scott Shane, *Yemen Sets Terms of a War on Al Qaeda*, NEW YORK TIMES, Dec. 4, 2010, at 1; U.S. Department of State Cable 09 NAIROBI 1057 ("Somalia TFG Prime Minister Worried About Rival") (Somalia).

constant concern in the cyber context.[27] The advantages of cooperation here may be more modest than in the more traditional context in which non-State actors conduct physical attacks against the victim State, however. In that context, local knowledge about terrain, terrorist camp locations and politics may prove particularly helpful in addressing the kinetic threats posed by terrorist or rebel groups. One disadvantage to obtaining consent or cooperation is temporal: in many cyber cases, a State may need (or wish) to respond to an ongoing attack immediately, leaving no time to seek a cooperative approach with the territorial State. One way to mitigate this temporal concern, while also promoting cooperation between the territorial and victim States, would be to negotiate consent agreements in advance.[28] In these agreements, the territorial State could provide advance consent to victim State operations in the former's cyber networks when certain triggers are met.

At the same time, the anonymity of cyber activity and the ease with which an actor may cover his tracks may reduce the victim State's overall incentives to seek any type of consent or cooperation from the territorial State before penetrating its cyber systems. In the cyber context, the victim State's actions in redress are less likely to come to light and, even if they do, it is easy for the victim State credibly to deny that it was the actual actor in that case.[29] In the non-cyber context, it is difficult (though not impossible) for a victim State physically to penetrate and use force in a territorial State without being detected. For example, an international investigation into the Cheonan incident (in which North Korea torpedoed a South Korean Navy ship) readily revealed Korean markings on the torpedo fragments.[30] In addition, the territorial State may be reluctant to cooperate with the host State

27. *See, e.g.*, Nakashima, *supra* note 5 (discussing U.S. concerns that actions in another country's networks could result in unintended consequences, including the disruption of civilian networks).

28. By way of precedent, the United States has negotiated a number of bilateral agreements relating to operations and ship-boarding to suppress the movement of narcotics and weapons of mass destruction. The latter set of agreements provides advance consent for either party to board a vessel flagged to the other party if the vessel is suspected of carrying illicit shipments of weapons of mass destruction. *See, e.g.*, Emma L. Belcher, *The Proliferation Security Initiative: Lessons for Using Nonbinding Agreements*, COUNCIL ON FOREIGN RELATIONS SPECIAL REPORT (July 2011).

29. *See* OWENS ET AL., *supra* note 3, at 81 (noting that most cyber attacks are inherently deniable).

30. Letter from the Permanent Representative of the Republic of Korea to the United Nations Addressed to the President of the Security Council (S/2010/281), June 4, 2010.

for national security reasons, particularly where the territorial State does not want to disclose information about its networks, systems and technology.

From a legal perspective, obtaining consent is an ideal way to avoid having to answer the host of difficult legal questions that currently attach to offensive and defensive uses of cyber tools. Action pursuant to consent also makes it less important that the victim State have a firm sense of who the author of the attacks is, because the territorial State is less likely to challenge the victim State's actions. From a political and military perspective, however, the costs of acting without seeking territorial State consent appear far lower than in the non-cyber context.

B. Request to address the cyber attack

Assume that the territorial State has not affirmatively consented to the victim State's use of cyber (or kinetic) tools to suppress the cyber threat emanating from the territorial State, perhaps because it is concerned about allowing the victim State to access its computer networks. At this point, the most direct way for a victim State to assess the territorial State's willingness and ability is to ask it to terminate the threat. Not only will this clearly put the territorial State on notice of the cyber attack, but it also will place an onus on the victim State to share relevant intelligence about the attack. If the territorial State responds by providing a plan for suppressing the attack, the victim State then has the basic information it needs to begin to assess the territorial State's willingness and ability to act.

Governments almost certainly will demand a caveat to this requirement, however. Where the victim State believes that the territorial State is colluding with the author of the cyber attack or will tip off the cyber attacker, the victim State should not be obligated to ask the territorial State before taking measures in the territorial State. This is a serious concern with States such as Russia and China, which are reported to use civilian proxies to conduct cyber attacks.[31] It is a particular concern in the cyber context because a hostile actor tipped off by the territorial State easily may divert its attacks through a different third State. Doing so in the non-cyber context takes time and money and poses significant logistical challenges.

31. Paul Rosenzweig, *From Worms to Cyber War*, DEFINING IDEAS, Dec. 9, 2011, http://www.hoover.org/publications/defining-ideas/article/102401 (describing Russian "cyber patriots"); David E. Sanger, John Markoff & Thom Shanker, *U.S. Plans Attack and Defense in Web Warfare*, NEW YORK TIMES, Apr. 28, 2009, at A1.

Creating too robust a caveat to the requirement to request assistance, however, will erode the balance that the "unwilling or unable" test strikes by putting a heavy finger on the scales in favor of security over sovereignty.

Even where the victim State is not concerned about a link between the territorial State and the hostile cyber actors, this factor magnifies complications that already exist in the non-cyber context. For the victim State, a requirement that it inform the territorial State about the cyber attacks it is suffering is not onerous. However, if the territorial State seeks additional information about those attacks—Are you sure they are coming from our territory? How do you know? What cyber tools do you have that can detect that, and how reliable are they?—the victim State may be hesitant to reveal its technological capacities.[32] Consider the territorial State's point of view as well. If the victim State simply asks it to suppress the threat, without seeking information about how the territorial State will do so, the territorial State may willingly comply, without having to reveal its cyber tools to the victim State. If the victim State seeks technological details about how the territorial State plans to proceed (which it reasonably might do to assure itself that the attacks will stop), the territorial State may be loath to reveal those details.[33] In the non-cyber context, it is far more likely that States will have adequate intelligence about each other's military hardware and capabilities. In the cyber context, the political relationship between the victim and territorial States—and, concomitantly, their willingness to share intelligence and technology—becomes highly predictive of how the victim State will proceed.

C. Good faith assessment of territorial State control and capacity

In the non-cyber world, when analyzing a territorial State's ability to suppress the threat, a victim State should assess what level of control the territorial State has over the geographic area from which the attacks are emanating. Conventional attacks plotted and launched from within a capital city may be far easier to detect, locate and suppress than attacks launched from remote jungles far from any town. A related question goes to the capacity of the territorial State's law enforcement and military officers, and

32. *See* Matthew Waxman, *Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)*, 36 YALE JOURNAL OF INTERNATIONAL LAW 421, 425 (2011) ("[N]o governments speak in much detail about their cyberwarfare capabilities and strategies at this point.").

33. As States garner increasing amounts of intelligence on each other's capabilities, this concern may diminish.

whether there are any reasons that those actors would not be able (or willing) to act against the non-State actors. In the cyber context, the question becomes how technologically sophisticated is the territorial State? While it is possible that one or more hostile actors is physically present in the territorial State, it is more likely that those committing cyber attacks against the victim State are present only electronically in the territorial State. Stopping those attacks, therefore, depends both on the capacity of the territorial State's cyber gurus and on the attacker's level of technical sophistication.

There is not enough publicly available information to gauge how often a victim State is likely to encounter a territorial State that is technically unable to defeat a cyber attack against the victim State. Some reports suggest that cyber is the great equalizer, allowing States with far weaker conventional militaries to take on those with traditionally strong conventional militaries.[34] Others assert that the cyber capacities of States such as the United States, Russia and China far exceed those of most other States.[35] Putting aside the objective capabilities of a particular territorial State, the secondary question of how much the victim State knows about the territorial State's capabilities remains a tricky one as long as cyber-capacities remain closely-held secrets. Publications such as *Jane's Defence Weekly* (as well as a State's domestic intelligence reports and the fact that States such as the United States may have provided weapons and training to the territorial State in question) make it relatively easy to ascertain what a State's kinetic capabilities are. In the cyber context, though, it will be particularly challenging for a victim State to assess the control and capacity of another State with which it does not have a close relationship already.

D. Good faith assessment of the territorial State's proposed means to suppress threat

Closely related to an assessment of the territorial State's capacity and control is a good faith assessment of the proposed means by which the territorial State will suppress the threat. The victim State must assess those pro-

34. Waxman, *supra* note 32, at 451, 455 (noting that "some experts assess that the United States is currently strong relative to others in terms of offensive capabilities" but also that "some States that are developing offensive cyber-warfare capabilities (such as North Korea, according to many experts) are non-status-quo powers or aspiring regional powers").

35. Leon Panetta, U.S. Secretary of Defense, Remarks on Cybersecurity (Oct. 11, 2012) ("It's no secret that Russia and China have advanced cyber capabilities."), http://www.defense.gov/transcripts/transcript.aspx?transcriptid=5136.

posed actions objectively. Even if the victim State would prefer to act itself, it should accept the territorial State's proposed approach if a "reasonable State" would believe that the approach will accomplish the victim State's core goal of suppressing the attack or threat of imminent attack.

In the non-cyber context, weeks may elapse between the time a territorial State proffers an operational plan and the time it executes it. In contrast, there will be almost immediate feedback on the success or failure of the territorial State's efforts to suppress the cyber threat. This makes it even more reasonable to defer to the territorial State's plan in the first instance, unless the ongoing attack against the victim State is so significant that there is no time for trial and error.

Establishing a preference for the territorial State's proposal is not without costs. Assume the territorial State proposes simply to shut off the server that is hosting the attacks against the victim State. Assume further that, if the territorial State permitted the victim State to address the threat itself, the victim State could stop the attack in a way that would allow it to continue to gather intelligence about the attacker. Should we continue to favor the territorial State's reasonable plan, even where doing so may force the victim State to lose some modicum of intelligence about its attacker? Probably so, though reasonable minds may disagree. What if the territorial State's plan is reasonable but is likely to result in some level of collateral damage, while the victim State has a high level of certainty that its plan would produce no such damage? In that case, the victim State would have at least a credible argument that the territorial State was "unable" to suppress the threat in a responsible way. Difficult questions such as these abound.

E. Prior interactions with the territorial State

Finally, in assessing the territorial State's proposed means to address the threat itself, the victim State should consider past interactions with the territorial State. Has the territorial State previously suppressed threats (conventional or cyber) emanating from its territory? Has the territorial State revealed a level of technical competence in the past that should give the victim State comfort that its proposed approach will work this time? Is the territorial State one from which cyber attacks consistently emanate, or is this an unusual incident? The more historically reliable and responsive the territorial State is, the less justification the victim State will have if it choos-

es to take action itself, and the more difficult it will be for the victim State to defend its actions if they come to light.

IV. ADVANCING CYBER LAW?

The United States has asserted that it will treat hostile acts in cyberspace as it would "any other threat to our country" and that it reserves the right to employ a military response, "as appropriate and consistent with applicable international law."[36] In other statements, the United States has made clear that in the non-cyber context, international law allows the United States to use force against non-State actors in another State's territory only when the territorial State has consented or is unwilling or unable to suppress the threat.[37] This—coupled with multiple news reports about internal debates within the U.S. government on cyber questions[38]—suggests that the United States is attempting to reason by analogy from existing international law governing the *jus ad bellum*. These reports also suggest that the United States is attempting to craft appropriate, and apparently highly restrictive, operational rules of the road in the cyber sphere. One news report stated that U.S. officials are focused on "concerns that action in another country's networks could violate international law, upset allies or result in unintended consequences, such as the disruption of civilian networks."[39] The article further reported that the U.S. Department of Defense has developed "strict conditions governing when military cyber-specialists could take action out-

36. THE WHITE HOUSE, U.S. INTERNATIONAL STRATEGY FOR CYBERSPACE 14 (May 1, 2011); Harold Koh, "International Law in Cyberspace," USCYBERCOM Interagency Legal Conference, Sept. 18, 2012.

37. Brennan, *supra* note 23 ("The United States does not view our authority to use military force against al-Qa'ida as being restricted solely to 'hot' battlefields like Afghanistan. Because we are engaged in an armed conflict with al-Qa'ida, the United States takes the legal position that—in accordance with international law—we have the authority to take action against al-Qa'ida and its associated forces without doing a separate self-defense analysis each time. And as President Obama has stated on numerous occasions, we reserve the right to take unilateral action if or when other governments are unwilling or unable to take the necessary actions themselves.").

38. *See, e.g.*, Nakashima, *supra* note 5; Ellen Nakashima, *Cyber-Intruder Sparks Massive Federal Response and Debate Over Dealing with Threats*, WASHINGTON POST, Dec. 8, 2011.

39. Nakashima, *supra* note 5. It is not clear whether those contemplated U.S. actions would be forcible or would consist of actions short of force (such as non-forcible countermeasures).

side U.S. networks" and that those conditions "are so stringent that the new capability to go outside military boundaries might never be used."[40]

In some ways this U.S. process is puzzling. In the face of such legal and technological uncertainty, one might expect a country with extensive cyber capabilities to take a minimalist approach to legal compliance, at least until the international community formulated certain common understandings about how to approach cyber warfare. Indeed, in the non-cyber context, the U.S. Government has done less hand wringing about using force extra-territorially, even though the manifestation of that force is far more public. Why is the United States working so hard to find the law and apply it in the cyber realm, where violations of Article 2(4) would be both legally uncertain and difficult to detect?

There are at least five factors that may explain why the United States has been edging cautiously toward a relatively constraining legal regime (one that in all likelihood will be a unilateral approach for some time to come). First, there often is an inherent institutional instinct in the U.S. government to anchor novel legal situations in existing bodies of law and practice, and to reason by analogy. This is, after all, the approach the Obama administration took toward detainee habeas cases.[41] There, the government determines (and asks courts to affirm) that someone is a combatant based on functional analogies between that person's activities and the activities of a member of a State's armed forces. Particularly where the analogies are quite reasonable (as they are between kinetic and cyber activities), it often is easier to draw from existing rules than to craft new ones from whole cloth. Additionally, U.S. government lawyers know that other governments are

40. *Id.* (noting that shutting down a server in another country likely would require Presidential permission). *See also* David Sanger, John Markoff & Thom Shanker, *U.S. Steps Up Effort on Digital Defenses*, New York TIMES, Apr. 27, 2009 (stating that President Bush personally authorized penetration by the U.S. military of a computer in Iraq to lure Al Qaeda members into an ambush); Ellen Nakashima, *Pentagon Considers Preemptive Strikes as Part of Cyber-Defense Strategy*, WASHINGTON POST, Aug. 28, 2010 (reporting on internal U.S. government debate about when the United States may go into foreign cyberspace and take preemptive action).

41. Respondent's Memorandum Regarding the Government's Detention Authority Relative to Detainees Held at Guantanamo Bay, In re: Guantanamo Bay Litigation, Mar. 13, 2010 (stating that the President has the authority under the 2001 Authorization for Use of Military Force to detain those persons whose relationship to Al Qaeda or the Taliban would, in appropriately analogous circumstances in a traditional international armed conflict, render them detainable).

likely to use those existing rules as a starting point from which to evaluate U.S. action.[42]

As a related matter, the U.S. culture surrounding the use of force and the conduct of armed conflicts has grown increasingly legalistic in the past ten years. While the United States always has been conscious of the legal role that the UN Charter plays in regulating uses of force, the past decade has found lawyers playing a particularly prominent role in structuring government decision making in this area.[43] A robust interagency process within the National Security Council ensures a forum for voices (such as those from the State Department) that are concerned about the diplomatic and reputational impacts of cyber activities that are seen as unlawful or illegitimate. And a perennial interest in being seen as following the rule of law renders unappealing an approach that ignores legal constraints entirely.[44]

Third, the United States is keenly aware of the ongoing controversy about its geographic approach to the U.S. conflict with Al Qaeda and associated forces.[45] The notion that the United States takes a forward-leaning approach to using force in third States with which it is not in conflict remains uncomfortable and legally contentious for many States. It follows that the United States would be similarly attuned to the far greater number of States that may (advertently or inadvertently) host cyber attacks against it, and to the almost-certain controversies that would follow from its uses of cyber (or kinetic) force in those States, absent a robust and well-articulated legal defense of those actions. Developing cautious standards through a cautious process is one way to establish that defense and to place other States on notice of its contents.

---

42. Matthew Waxman suggests that this is not the only approach that the United States might have taken. Waxman, *supra* note 32, at 453 (noting that it might be "in the United States' strategic interest to legally *delink* cyber-activities from armed force instead of defining force by reference to effects").

43. For a discussion of the role of international law in the Cuban Missile Crisis, *see* ABRAM CHAYES, THE CUBAN MISSILE CRISIS (1974). For the lawyers' role in the past ten years, *see* JACK GOLDSMITH, POWER AND CONSTRAINT xv (referring to "faceless executive-branch lawyers" micromanaging national security decisions).

44. Brennan, *supra* note 23 (describing one of the core values of the United States as "adhering to the rule of law").

45. *Id.* ("An area in which there is some disagreement is the geographic scope of the conflict. The United States does not view our authority to use military force against al-Qa'ida as being restricted solely to 'hot' battlefields like Afghanistan. . . . Others in the international community—including some of our closest allies and partners—take a different view of the geographic scope of the conflict, limiting it only to the 'hot' battlefields.").

Even assuming these three propositions are true, this does not explain why the United States has not chosen to adopt freedom of action in cyberspace—at least for now, while the law is very unclear and it remains difficult to attribute a particular cyber action to any particular actor. That is, if the United States felt that it were justified in responding to a particular incoming attack—even one with origins in a friendly and technologically advanced State—why would it not simply respond to the attack in that friendly State and then deny knowledge of the response? One answer seems to lie in concerns about cyber collateral damage.[46] Past efforts to dismantle particular websites have resulted in unexpected disruptions of servers in various countries. For instance, when the U.S. military dismantled a Saudi web site in 2008, it inadvertently disrupted over 300 servers, including in Texas, Saudi Arabia and Germany.[47] The high likelihood of collateral damage (and the concomitant likelihood that such damage becomes public) may place significant pressure on a country such as the United States to set a prudentially high bar for using cyber force in other States' territories.[48]

Finally, reciprocity concerns echo loudly in the ears of U.S. policymakers and lawyers. Even though the United States rarely will find itself being accused by other States of being unwilling or unable to suppress a particular cyber threat, the United States should be interested in prioritizing consent wherever possible, to create an expectation that other States affected by cyber attacks emanating from the United States will approach the U.S. government in the first instance, before taking unilateral action against U.S. cyber infrastructure.[49] This is particularly true because the United States is viewed as a major source of cyber attacks, cyber exploitations and botnets.[50] It is not in the U.S. interest to allow other States to claim that there is a legal black hole regarding cyber uses of force or to be able to

46. *See* OWENS ET AL., *supra* note 3, at 121–26 (describing difficulty in calculating accurately collateral damage from a cyber attack and describing damage assessment techniques for cyber attacks as "primitive").

47. Nakashima, *Preemptive Strikes*, *supra* note 40.

48. Note that this is true even if the United States is in an international armed conflict with State X and wishes to use cyber force against computers located within State X. Even that activity, which does not implicate the "unwilling or unable" test, may lead to collateral damage in third States.

49. It seems much more likely that a State would contemplate using unilateral cyber force against the United States than using unilateral kinetic force against it.

50. Jack Goldsmith, *Cybersecurity Treaties: A Skeptical View* 7, HOOVER INSTITUTE, media.hoover.org/documents/FutureChallenges_Goldsmith.pdf (last visited Oct. 28, 2012).

claim that the "unwilling or unable" test has no substantive or procedural content.

V. CONCLUSION

The "unwilling or unable" test remains a relevant proposition when a victim State suffers a cyber armed attack that is launched from the territory of a non-hostile State. Depending on the kinds of cyber activities that States treat as violating a neutral State's obligations and those that they treat as rising to the level of an armed attack, the international community will employ the test more or less frequently. The nature of cyber attacks—including the speed at which they occur—places pressure on the victim State to conduct both a rapid and accurate assessment of the territorial State's capabilities and political disposition. Cyber attacks also place pressure on the territorial State to reveal some of its technological capacity if it wishes to avoid having the victim State act in its stead. The relationship between the territorial and victim States will play an outsized role in the outcome of the "unwilling or unable" inquiry. Yet this inquiry stands between the victim State and a "global cyberwar on terror," and must be taken seriously.

\\jciprod01\productn\V\VLR\56-3\VLR315.txt unknown Seq: 1 15-DEC-11 12:39

2011]

CYBER OPERATIONS AND THE *JUS AD BELLUM* REVISITED

MICHAEL N. SCHMITT*

OVER a decade ago, Professor John Murphy and I presented papers at a U.S. Naval War College conference on Computer Network Attack and International Law.[1] At the time, I was an Air Force officer assigned to the College and he an eminent scholar at Villanova University School of Law. Professor Murphy generously and graciously took me under his wing and has been a cherished mentor ever since. It was a particular pleasure to serve as the Naval War College's 2007-2008 Charles H. Stockton Professor, a post which Professor Murphy held with great distinction in 1980-1981. Over the years, it has also been my good fortune to become close friends with him. Therefore, it is a tremendous honor for me to contribute to this effort to mark Professor Murphy's long and distinguished service to our nation, the international law community, and Villanova University. I do so by revisiting the topic that began our friendship: cyber operations.

I. INTRODUCTION

In April and May 2007, Estonia was victimized by massive computer network attacks.[2] The incident began with rioting incited by ethnic Russian cyber agitators in response to the government's decision to move a Soviet war memorial from the center of Tallinn to a military cemetery on the outskirts of the capital. Subsequent actions included direct cyber attacks against Estonian targets, including government and commercial Internet infrastructure and information systems such as the those of the President, Prime Minister, Parliament, State Audit Office, ministries, political parties, banks, news agencies, and Internet service providers. They involved denial of service (DoS), distributed denial of service (DDoS), defacement, and destruction.

* Chairman, International Law Department, US Naval War College. The views expressed in this Article are those of the author in his personal capacity and do not necessarily represent those of the US Navy or any other US government entity. This Article benefitted from the generous support of the National Research Council of the National Academies. It is based in part on Michael N. Schmitt, *Cyber Operations in International Law: The Use of Force, Collective Security, Self-Defense, and Armed Conflicts*, *in* DETERRING CYBERATTACKS: INFORMING STRATEGIES AND DEVELOPING OPTIONS FOR U.S. POLICY 151 (2010), portions of which are reproduced with permission.

1. The proceedings of the 1999 conference were published as 76 INTERNATIONAL LAW STUDIES: COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW (Michael N. Schmitt & Brian O'Donnell eds., 2002).

2. For an excellent discussion of the attacks, see ENEKEN TIKK ET AL., INTERNATIONAL CYBER INCIDENTS: LEGAL CONSIDERATIONS 14-33 (2010).

Because Estonia had invested heavily in networking following independence, the attacks proved devastating. By 2007, the country relied on information services for everything from banking and filing tax returns to paying for parking and public transportation. Internet services covered all of Estonia, with half the population enjoying access from their homes.

Most of the attacks emanated from outside the country, principally Russia. Their origin was also traced to at least 177 other countries.[3] Initially, they came from private IP addresses, although experts tracked a number to Russian government institutions. It remains uncertain whether the latter were launched with the government's knowledge. As the cyber attacks unfolded, they became increasingly sophisticated, evidencing considerable organization and command and control. While various pro-Russian activist groups apparently executed some of the second-wave operations, there is no firm evidence that the Russian government either conducted or orchestrated them.

The impact of the cyber assault proved dramatic; government activities such as the provision of state benefits and the collection of taxes ground to a halt, private and public communications were disrupted, and confidence in the economy plummeted. But what was the legal character of the incident?

In the aftermath of the Second World War, the international community crafted a new normative scheme in the form of the United Nations Charter, which includes both a prohibition on the use of force in international relations and a system for enforcing the proscription. Today, the Charter, together with related customary international law norms,[4] governs how and when force may be employed by States.

This Article explores the contemporary international law governing cyber operations. In particular, it asks three questions relevant to the Charter scheme governing the use of force in international relations:

1) When does a cyber operation constitute a wrongful "use of force" in violation of Article 2(4) of the United Nations Charter and customary international law?;
2) When does a cyber operation amount to a "threat to the peace, breach of the peace, or act of aggression," such that the Security Council may authorize a response thereto?; and
3) When does a cyber operation constitute an "armed attack," such that the victim-state may defend itself, even kinetically, pursuant to the right of self-defense set forth in Article 51 of the UN Charter and customary international law?

3. *See* Charles Clover, *Kremlin-Backed Group Behind Estonia Cyber Blitz*, FIN. TIMES, Mar. 11, 2009, http://www.ft.com/cms/s/0/57536d5a-0ddc-11de-8ea3-0000779fd2ac.html#axzz1DtPlzO27.

4. For a further discussion of customary international law, see *infra* note 13 and accompanying text.

The attacks against Estonia, similar ones against Georgia during its armed conflict with Russia in 2008,[5] and the thousands of others directed against government, corporate, and private systems worldwide on a daily basis aptly demonstrate the reality, immediacy, and scale of the threat. It is one well-recognized by states. The May 2010 U.S. National Security Strategy cites cyber security threats as "one of the most serious national security, public safety, and economic challenges we face as a nation."[6] Similarly, the analysis and recommendations on NATO's new Strategic Concept prepared by a group of distinguished experts led by former U.S. Secretary of State Madeleine Albright singled out "cyber assaults of varying degrees of severity" as one of the three likeliest threats the NATO Allies will face in the next decade.[7]

Unfortunately, the existing legal norms do not offer a clear and comprehensive framework within which states can shape policy responses to the threat of hostile cyber operations. In particular, international law as traditionally understood departs at times from what the international community would presumably demand in the cyber context. To some extent, this divergence can be accommodated through reasonable interpretation of the relevant norms. Where it cannot, the law would seem to require attention, either through treaty action or through the development of new understandings of the prevailing legal concepts.[8]

II. CYBER OPERATIONS AS A "USE OF FORCE"

The United Nations Charter, in Article 2(4), states that "[a]ll Members [of the United Nations] shall refrain in their international relations from the threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations." Despite the reference to territorial integrity and political independence, it is now widely understood that the prohibition applies to any use of force not otherwise permitted by the terms of the Charter, specifically uses of force authorized by the Security Council and defensive operations, each discussed separately below.[9]

5. *See* TIKK ET AL., *supra* note 2, at 66-90 (describing cyber attacks against Georgia during dispute with Russia over South Ossetia).

6. THE WHITE HOUSE, NATIONAL SECURITY STRATEGY 27 (2010) [hereinafter 2010 NATIONAL SECURITY STRATEGY], *available at* http://www.whitehouse.gov/sites/default/files/rss_viewer/national_security_strategy.pdf.

7. *See* N. Atlantic Treaty Org. [NATO], *NATO 2020: Assured Security; Dynamic Engagement* 17 (May 17, 2010) [hereinafter *NATO 2020*], *available at* http://www.nato.int/strategic-concept/expertsreport.pdf. The others are an attack by a ballistic missile and strikes by international terrorist groups. *See id.* (listing most probable threats of coming decade).

8. For book length treatment of these issues, see INTERNATIONAL LAW STUDIES: COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW, *supra* note 1; THOMAS C. WINGFIELD, THE LAW OF INFORMATION CONFLICT: NATIONAL SECURITY LAW IN CYBERSPACE (2000); 64 A.F. L. REV. (2009) (dedicating edition to cyberlaw).

9. In its original form, the draft Charter contained no reference to territorial integrity or political independence, and their subsequent inclusion was controver-

Article 2(4) was revolutionary in its extension to threats. Of course, only those threats of a use of force that would otherwise be unlawful qualify.[10] For instance, threatening destructive defensive cyber attacks against another state's military infrastructure if that state unlawfully mounts cross-border operations would not breach the norm. However, threats of destructive cyber operations against another state's critical infrastructure unless that state cedes territory would do so.

The prohibition applies only to an explicit or implied communication of a threat; its essence is coercive effect. It does not reach actions which simply threaten the security of the target state, but which are not communicative in nature. Thus, the introduction into a state's cyber systems of vulnerabilities that are capable of destructive activation at some later date would not constitute a threat of the use of force unless their presence is known to the target state and the originating state exploits them for some coercive purpose.[11]

It is generally accepted that the prohibition on the threat or use of force represents customary international law.[12] Resultantly, it binds all states regardless of membership in the United Nations. Article 38 of the Statute of the International Court of Justice (ICJ) defines customary law as "general practice accepted as law."[13] It requires the coexistence of state practice and *opinio juris sive necessitatis*, a belief that the practice is engaged in, or refrained from, out of a sense of legal obligation (rather than practical or policy reasons).

Although simple in formulation, the norm is complex in substantive composition. It poses two key questions: "What is a use of force?" and "To whom does the prohibition apply?" Both bear heavily on the legality of cyber operations, which did not exist when the UN Charter was adopted by states in 1945. The difficulty of applying a legal provision that did not contemplate a particular type of operation is apparent.

Finally, it must be borne in mind that neither Article 2(4) nor its customary counterpart is remedial in nature. Rather, they merely set a threshold for breach of international law. The nature of the response to a

sial. The "other manner" language was inserted to make clear that their inclusion was not meant to limit the reach of the provision. *See* DOC. 1123, I/8, 6 U.N.C.I.O. DOCS. 65 (1945); DOC. 885, I/1/34, 6 U.N.C.I.O. DOCS. 387 (1945); DOC. 784, I/1/27, 6 U.N.C.I.O. DOCS. 336 (1945).

10. This point was made by the International Court of Justice in *Legality of the Threat or Use of Nuclear Weapons*, Advisory Opinion, 1996 I.C.J. 226, ¶ 47 (July 8) ("[I]f the use of force itself in a given case is illegal—for whatever reason—the threat to use such force will likewise be illegal.").

11. Although a threat must be coercive in some sense, there is no requirement that a specific "demand" accompany the threat.

12. *See* Military and PARAMILITARY Activities in and Against Nicaragua (*Nicaragua*) (Nicar. v. U.S.), 1986 I.C.J. 14, 98-101 (June 27).

13. Statute of the International Court of Justice, art. 38, 1977 I.C.J. Acts & Docs. 61. On customary law, see Yoram Dinstein, *The Interaction Between Customary International Law and Treaties*, *in* 322 COLLECTED COURSES OF THE HAGUE ACADEMY OF INTERNATIONAL LAW 243 (2006).

wrongful use of force is instead determined by the law of state responsibility, the scope of authority of the Security Council, and the law of self-defense. Each is addressed below.

A. *Uses of Force*

Do cyber operations constitute a "use of force" as that phrase is understood in relation to the prohibition? The interpretive dilemma is that the drafters of the Charter took a cognitive shortcut by framing the treaty's prohibition in terms of the *instrument* of coercion employed—force. Thus, the norm did not outlaw economic and political coercion, but disallowed military force, at least absent an express Charter exception. Yet, it is seldom the instrument employed, but instead the *consequences* suffered, that matter to states. At the time the Charter was drafted an instrument based-approach made sense, for prior to the advent of cyber operations the consequences that states sought to avoid usually comported with instrument-based categories. Cyber operations do not fit neatly into this paradigm because although they are "non-forceful" (that is, non-kinetic), their consequences can range from mere annoyance to death. Resultantly, as the Commander of U.S. Cyber Command noted during his confirmation hearings, policy makers must understand that "[t]here is no international consensus on a precise definition of a use of force, in or out of cyberspace. Consequently, individual nations may assert different definitions, and may apply different thresholds for what constitutes a use of force."[14]

That the term "use of force" encompasses resort to armed force by a state, especially force levied by the military is self-evident. Armed force thus includes kinetic force—dropping bombs, firing artillery, and so forth. It would be no less absurd to suggest that cyber operations that generate consequences analogous to those caused by kinetic force lie beyond the prohibition's reach, than to exclude other destructive non-kinetic actions, such as biological or radiological warfare. Accordingly, cyber operations that directly result (or are likely to result) in physical harm to individuals or tangible objects equate to armed force, and are therefore uses of force. For instance, those targeting an air traffic control system or a water treatment facility clearly endanger individuals and property. But cyber operations are usually mounted without causing such consequences, as illustrated by the case of Estonia. Are such operations nonetheless barred by the use-of-force prohibition?

The starting point for any interpretive endeavor in law is the treaty text in question.[15] In this regard, note that the adjective "armed" does not

14. Staff of S. Comm. on Armed Services, Advance Questions for Lieutenant General Keith Alexander, USA, Nominee for Commander, United States Cyber Command from the United States' Armed Services Committee (Apr. 15, 2010), www.senate.gov/~armed_services/statemnt/2010/04%20April/Alexander%2004-15-10.pdf.

15. According to the Vienna Convention on the Law of Treaties, "[a] treaty shall be interpreted in good faith in accordance with the ordinary meaning to be

appear with reference to "force" in Article 2(4). By contrast, the Charter preamble cites the purpose of ensuring that "armed force shall not be used, save in the common interest." Similarly, the Charter excludes "armed force" from the non-forceful measures the Security Council may authorize under Article 41 and mentions planning for armed force with regard to forceful Article 42 measures.[16] And the Charter only allows forceful defensive actions in the face of an "armed attack."[17] This textual distinction suggests an interpretation of "force" that is broader in scope than the common understanding of the term.

When text is ambiguous, recourse may be had to "the preparatory work of [a] treaty and the circumstances of its conclusion."[18] The Charter's *travaux preparatoires* indicate that during the drafting of the instrument a proposal to extend the reach of Article 2(4) to economic coercion was decisively defeated.[19] A quarter century later, the issue again arose during proceedings leading to the UN General Assembly's Declaration on Friendly Relations.[20] The question of whether "force" included "all forms of pressure, including those of a political or economic character, which have the effect of threatening the territorial integrity or political independence of any State" was answered in the negative.[21] Whatever force is, then, it is not economic or political pressure. Therefore, a cyber operation that involves such coercion is definitely not a prohibited use of force. Psychological cyber operations (assuming they are non-destructive) intended solely to undermine confidence in a government or economy illustrate such actions.

---

given to the terms of the treaty in their context and in light of its object and purpose" which can be gleaned from the text, "including its *preamble* and annexes . . . ." Vienna Convention on the Law of Treaties, art. 31(1)-(2), May 23, 1969, 1155 U.N.T.S. 331, 340 (emphasis added). The United States is not a party to the Vienna Convention, but treats most of its provisions as reflective of customary international law.

16. *See* U.N. Charter art. 46 (referencing planning for armed force). "Plans for the application of armed force shall be made by the Security Council with the assistance of the Military Staff Committee." *Id.*

17. U.N. Charter art. 51.

18. Vienna Convention on the Law of Treaties, *supra* note 15, art. 32, 1155 U.N.T.S. at 340.

19. *See* DOC. 2, G/7(e)(4), 3 U.N.C.I.O. DOCS. 251, 253-54 (1945). Economic coercion, which typically involves trade sanctions, must be distinguished from "blockade," which has the effect of cutting off trade, but employs military force to do so. It has historically been accepted that imposition of a blockade is an "act of war."

20. *See* Declaration on Principles of International Law Concerning Friendly Relations and Co-operation Among States in Accordance with the Charter of the United Nations, G.A. Res. 2625 (XXV), U.N. GAOR, 25th Sess., Supp. No. 28, U.N. DOC. A/RES/8082 (Oct. 24, 1970).

21. U.N. GAOR Special Comm. on Friendly Relations, U.N. DOC. A/AC.125/SR.114 (1970); *accord* Rep. of the Special Comm. on Principles of Int'l Law Concerning Friendly Relations and Co-operation Among States, U.N. GAOR, 24th Sess., Supp. No. 19, U.N. DOC. A/7619 (1969). The draft declaration contained text tracking that of U.N. Charter Article 2(4).

Suggestions to limit "force" to "armed force," or even the force required to amount to an "armed attack," were likewise rejected during the proceedings.[22] This seemed to indicate that "force" was not coterminous with "armed force," thereby strengthening the significance of the absence of the term "armed" in Article 2(4). In the *Case Concerning Military and Paramilitary Activities in and Against Nicaragua*,[23] the ICJ expressly characterized certain actions that were non-kinetic in nature as uses of force:

> [W]hile the arming and training of the contras can certainly be said to involve the threat or use of force against Nicaragua, this is not necessarily so in respect of all the assistance given by the United States Government. In particular, the Court considers that the mere supply of funds to the contras, while undoubtedly an act of intervention in the internal affairs of Nicaragua . . . does not in itself amount to a use of force.[24]

The determination that a use of force can embrace acts, like arming or training guerillas, which fall short of armed force leaves open the possibility that non-physically destructive cyber operations may fall within the term's ambit. The threshold for a use of force must therefore lie somewhere along the continuum between economic and political coercion on the one hand and acts which cause physical harm on the other.

Unfortunately, unequivocal state practice in characterizing particular cyber attacks as (or not as) uses of force is lacking. In part, this is because the Article 2(4) prohibition extends solely to acts of states, and very few states have definitively been identified as the initiator of a cyber operation that might amount to a use of force. Moreover, states may well hesitate to label a cyber operation as a use of force out of concern that doing so would escalate matters or otherwise destabilize the situation. Therefore, one can only speculate as to future state practice regarding the characterization of cyber operations.

Over a decade ago, this author identified a number of factors that would likely influence assessments by states as to whether particular cyber operations amounted to a use of force.[25] They are based on a recognition that while states generally want to preserve their freedom of action (a motivation to keep the threshold high), they equally want to avoid any harmful consequences caused by the actions of others (a motivation to keep the threshold low). Thus, states will seek to balance these conflicting objectives through consideration of factors such as those set forth below. The approach has generally withstood the test of time.

22. *See* U.N. GAOR Special Comm. on Friendly Relations, U.N. DOC. A/AC.125/SR.114 (1970).

23. *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14 (June 27).

24. *Id.* ¶ 228.

25. *See* Michael N. Schmitt, *Computer Network Attack and Use of Force in International Law: Thoughts on a Normative Framework*, 37 COLUM. J. TRANSNAT'L L. 885, 914-16 (1999).

1) *Severity*: Consequences involving physical harm to individuals or property will alone amount to a use of force. Those generating only minor inconvenience or irritation will never do so. Between the extremes, the more consequences impinge on critical national interests, the more they will contribute to the depiction of a cyber operation as a use of force. In this regard, the scale, scope, and duration of the consequences will have great bearing on the appraisal of their severity. Severity is self-evidently the most significant factor in the analysis.

2) *Immediacy*: The sooner consequences manifest, the less opportunity states have to seek peaceful accommodation of a dispute or to otherwise forestall their harmful effects. Therefore, states harbor a greater concern about immediate consequences than those that are delayed or build slowly over time.

3) *Directness*: The greater the attenuation between the initial act and the resulting consequences, the less likely states will be to deem the actor responsible for violating the prohibition on the use of force. Whereas the immediacy factor focused on the temporal aspect of the consequences in question, directness examines the chain of causation. For instance, the eventual consequences of economic coercion (economic downturn) are determined by market forces, access to markets, and so forth. The causal connection between the initial acts and their effects tends to be indirect. In armed actions, by contrast, cause and effect are closely related—an explosion, for example, directly harms people or objects.

4) *Invasiveness*: The more secure a targeted system, the greater the concern as to its penetration. By way of illustration, economic coercion may involve no intrusion at all (trade with the target state is simply cut off), whereas in combat the forces of one state cross into another in violation of its sovereignty. The former is undeniably not a use of force, whereas the latter always qualifies as such (absent legal justification, such as evacuation of nationals abroad during times of unrest). In the cyber context, this factor must be cautiously applied. In particular, cyber exploitation is a pervasive tool of modern espionage. Although highly invasive, espionage does not constitute a use of force (or armed attack) under international law absent a nonconsensual physical penetration of the target state's territory, as in the case of a warship or military aircraft which collects intelligence from within its territorial sea or airspace. Thus, actions such as disabling cyber security mechanisms to monitor keystrokes would, despite their invasiveness, be unlikely to be seen as a use of force.

5) *Measurability*: The more quantifiable and identifiable a set of consequences, the more a state's interest will be deemed to have been affected. On the one hand, international law does not view economic coercion as a use of force even though it may cause significant suffering. On the other, a military attack that causes only a limited degree of destruction clearly qualifies. It is difficult to identify or quantify the harm caused

by the former (e.g., economic opportunity costs), while doing so is straightforward in the latter (X deaths, Y buildings destroyed, etc).

6) *Presumptive legitimacy*: At the risk of oversimplification, international law is generally prohibitory in nature. In other words, acts which are not forbidden are permitted; absent an express prohibition, an act is presumptively legitimate.[26] For instance, it is well accepted that the international law governing the use of force does not prohibit propaganda, psychological warfare, or espionage. To the extent such activities are conducted through cyber operations, they are presumptively legitimate.

7) *Responsibility*: The law of state responsibility (discussed below) governs when a state will be responsible for cyber operations. But it must be understood that responsibility lies along a continuum from operations conducted by a state itself to those in which it is merely involved in some fashion. The closer the nexus between a state and the operations, the more likely other states will be inclined to characterize them as uses of force, for the greater the risk posed to international stability.

The case of the Estonian cyber attacks can be used to illustrate application of the approach. Although they caused no deaths, injury, or physical damage, the attacks fundamentally affected the operation of the entire Estonian society. Government functions and services were severely disrupted, the economy was thrown into turmoil, and daily life for the Estonian people was negatively affected. The consequences far exceeded mere inconvenience or irritation. The effects were immediate and, in the case of confidence in government and economic activity, wide-spread and long-term. They were also direct, as with the inability to access funds and interference with the distribution of government benefits. Since some of the targeted systems were designed to be secure, the operations were highly invasive. While the consequences were severe, they were difficult to quantify, since most involved denial of service, rather than destruction of data. Although political and economic actions are presumptively legitimate in use-of-force terms, these operations constituted more than merely pressuring the target state. Instead, they involved intentionally frustrating governmental and economic functions. Taken together as a single "cyber operation," the incident arguably reached the use-of-force threshold. Had Russia been responsible for them under international law, it is likely that the international community would have (or should have) treated them as a use of force in violation of the UN Charter and customary international law.

26. In *The Case of the S.S. "Lotus"*, the Permanent Court of International Justice famously asserted that "[t]he rules of law binding upon States . . . emanate from their own free will as expressed in conventions or by usages generally accepted as expressing principles of law and established in order to regulate the relations between these co-existing independent communities or with a view to the achievement of common aims." S.S. "Lotus" (Fr. v. Turk.), 1927 P.C.I.J. (ser. A) No. 10, at 18 (Sept. 7).

The criteria are admittedly imprecise, thereby permitting states significant latitude in characterizing a cyber operation as a use of force, or not. In light of the increasing frequency and severity of cyber operations, a tendency towards resolving grey areas in favor of finding a use of force can be expected to emerge. This state practice will over time clarify the norm and its attendant threshold.

B. *Applicability of the Prohibition*

By its own express terms, Article 2(4) applies solely to Members of the United Nations. As discussed, the prohibition extends to non-Members by virtue of customary law. That is the limit of applicability. Non-state actors, including individuals, organized groups, and terrorist organizations, cannot violate the norm absent a clear relationship with a state. Their actions may be unlawful under international and domestic law, but not as a violation of the prohibition on the use of force. Thus, in the Estonian case, and barring any evidence of Russian government involvement, none of those individuals or groups conducting the operations violated the Article 2(4) prohibition. But when can the conduct of individuals or groups be attributed to a state, such that the state is legally responsible for their actions? The law of state responsibility governs such situations.[27]

Obviously, states are legally responsible for the conduct of their governmental organs or entities.[28] This principle extends to unauthorized acts.[29] Accordingly, any cyber operation rising to the level of an unlawful use of force will entail responsibility on the part of the state when launched by its agents, even when they are acting *ultra vires*.

The fact that a state did not itself conduct the cyber operations at hand does not mean that it escapes responsibility altogether. States are also responsible for "the conduct of a person or group of persons . . . if the person or group of persons is in fact acting on the instructions of, or under the direction or control of, that State in carrying out the conduct."[30] The ICJ addressed the degree of control necessary for attribution in the *Nicaragua* case. There the Court considered attribution of the acts of the Nicaraguan Contras (a rebel group supported by the United States) to the United States, such that the United States would be responsible for breaches of international humanitarian law (IHL) committed by the group. While finding the United States responsible for its own "planning, direction and support" of the Contras,[31] the court limited responsibility

27. This law is set forth, in non-binding form, in the International Law Commission's Draft Articles on Responsibility of States for Internationally Wrongful Acts, *in* Report of the Int'l Law Comm'n, 53d Sess., Apr. 23-June 1, July 2-Aug. 10, 2001, UN DOC. A/56/10; GAOR, 56th Sess., Supp. No. 10 (2001) [hereinafter Draft Articles on State Responsibility].

28. *See id.* art. 4, at 44.

29. *See id.* art. 7, at 44.

30. *See id.* art. 8, at 45.

31. *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 86 (June 27).

for the Contra actions to those in which the United States exercised "*effective control* of the military or paramilitary operations in the course of which the alleged violations were committed."[32] Mere support for their activities did not suffice.

The Appeals Chamber of the International Criminal Tribunal for the Former Yugoslavia (ICTY) took a different tack in the *Prosecutor v. Tadic*[33] case, where it held that the authority of the government of the Federal Republic of Yugoslavia over the Bosnia Serb armed groups "required by international law for considering the armed conflict to be international was *overall control* going beyond the mere financing and equipping of such forces and involving also participation in the planning and supervision of military operations."[34] It is essential to note that although the tribunal expressly rejected the higher *Nicaragua* threshold of effective control, the technical legal issue was not state responsibility, but rather the nature of the armed conflict. Thus, while *Tadic* brings *Nicaragua* into question by proffering a lower threshold, it does not necessarily supplant the effective control test.[35] It remains unclear whether effective control, overall control, or some other test governs in international law, although the ICJ has twice reaffirmed its version.[36]

In the cyber context, then, states will be responsible for violating the prohibition on the use of force to the extent that they either direct private individuals or groups to conduct the operations or are heavily involved in them. Determinations will be made on a case-by-case basis by looking to the extent and nature of involvement by the state with the group and in the particular operations.

Even if conduct is not attributable to a state as under its control, it will nevertheless "be considered an act of that State . . . if and to the extent that the State acknowledges and adopts the conduct in question as its own."[37] The ICJ addressed this situation in the *Case Concerning United States Diplomatic and Consular Staff in Tehran*,[38] which involved seizure of the U.S. Embassy by Iranian militants in 1979. The Iranian government was uninvolved in the initial seizure, but later passed a decree that accepted and maintained the occupation of the embassy. According to the ICJ, "[t]he approval given to [the occupation of the Embassy] by the

32. *Id.* ¶ 115 (emphasis added); *id.* ¶ 109.

33. Prosecutor v. Tadic, Case No. IT-94-1-A, Appeals Chamber Judgment (Int'l Crim. Trib. for the Former Yugoslavia July 15, 1999).

34. *Id.* ¶ 145.

35. Although, the court in dicta suggested the test is also suitable for application to issues of state responsibility. *See id.*, ¶¶ 116-37.

36. *See, e.g.*, Application of Convention on Prevention and Punishment of Crime of Genocide (Bosn. & Herz. v. Serb. & Montenegro), Judgment, 2007 I.C.J. 91, 391-92 (Feb. 26); Armed Activities on the Territory of the Congo (*Congo*) (Dem. Rep. Congo v. Uganda), Judgment, 2005 I.C.J. 168, ¶ 160 (Dec. 19).

37. Draft Articles on State Responsibility, *supra* note 27, art. 11, at 45.

38. United States Diplomatic and Consular Staff in Teheran (U.S. v. Iran), 1980 I.C.J. 3 (May 24).

Ayatollah Khomeini and other organs of the Iranian State, and the decision to perpetuate them, translated continuing occupation of the Embassy and detention of the hostages into acts of that State."[39]

It should be cautioned that mere expressions of approval do not suffice for attribution; rather, the state must somehow subsequently embrace the actions as its own, for instance, by tangibly supporting their continuance, failing to take actions to suppress them, or otherwise adopting them. Adoption may either be express, as in the *Hostages* case, or implied, as when a state engages in conduct that undeniably constitutes adoption. In the Estonian case, had Russia publicly encouraged further attacks, it would have borne responsibility not only for the subsequent attacks, but also those in the initial wave.

A state may also be held responsible for the effects of unlawful acts of private individuals or groups on its territory when it fails to take reasonably available measures to stop such acts in breach of its obligations to other states. In this situation, its violation is of the duty owed to other states, but its responsibility extends to the effects of the act itself. Applying this standard in the *Hostages* case, the ICJ found that the Iranian government failed to take required steps to prevent the seizure of the U.S. Embassy or regain control over it, placing Iran in breach of its international obligation to safeguard diplomatic premises.[40] The key to such responsibility lies in the existence of a separate legal duty to forestall the act in question, and an ability to comply with said duty. The ICJ articulated this principle in its very first case, *Corfu Channel*,[41] where it held that every state has an "obligation not to allow knowingly its territory to be used for acts contrary to the rights of other States."[42] Of the many obligations that states owe to each other, ensuring their territory is not a launching pad for the use of force or armed attacks (see discussion below) against other states certainly ranks among the most important. The fact that a use of force consists of cyber operations rather than traditional armed force would not diminish the responsibility of the state involved.

Finally, consider a situation in which the effects of a cyber operation extend beyond the targeted state. This is an especially relevant scenario in the cyber context, for networking and other forms of interconnectivity mean that a cyber use of force by State A against State B may have consequences in State C that would rise to the level of a use of force if directed against C. The causation of such effects would not amount to a violation of Article 2(4) vis-à-vis C. Article 2(4)'s requirement that Members "refrain in their international *relations*"[43] from the use of force implies an element of purposely engaging in some action in respect of another speci-

39. *Id.* ¶ 74.

40. *See id.* ¶¶ 76-78.

41. Corfu Channel (U.K. v. Alb.), 1949 I.C.J. 4 (Apr. 9).

42. *Id.* at 22.

43. U.N. Charter art. 2(4) (emphasis added).

fied state. Inadvertent effects caused in states other than the target state do not constitute a form of "international relations."

However, even if the state did not intend such effects, it is clear that it bears responsibility for them. As noted in the Draft Articles of State Responsibility, "[t]here is an internationally wrongful act of a State when conduct consisting of an action or omission: (a) [i]s attributable to the State under international law; and (b) [c]onstitutes a breach of an international obligation of the State."[44] In the envisaged case, since State A conducted the cyber operation, the action is directly attributable to it. Further, the wrongful use of force against B would constitute a breach of A's international obligation to refrain from the use of force. That the intended "victim" was B matters not. The criterion has been met once the breach of an international obligation has occurred. This is so even if the effects in C were unintended. As noted in the International Law Commission's Commentary to the relevant article:

> A related question is whether fault constitutes a necessary element of the internationally wrongful act of a State. This is certainly not the case if by "fault" one understands the existence, for example, of an intention to harm. In the absence of any specific requirement of a mental element in terms of the primary obligation, it is only the act of a State that matters, independently of any intention.[45]

C. *Remedies for Violation*

In the event of state responsibility for an unlawful act, the victim-state is entitled to reparation, which can take the form of restitution, compensation, or satisfaction.[46] With regard to cyber operations amounting to a use of force, compensation could be claimed for any reasonably foreseeable physical or financial losses. A state may also take any responsive actions that neither amount to a use of force nor breach an existing treaty or customary law obligation. As an example, a state may choose to block incoming cyber transmissions emanating from the state that has used force against it.

---

44. Draft Articles of State Responsibility, *supra* note 27, art. 2, at 43.

45. JAMES CRAWFORD, THE INTERNATIONAL LAW COMMISSION'S ARTICLES ON STATE RESPONSIBILITY: INTRODUCTION, TEXT AND COMMENTARIES 84 (2002).

46. *See* Draft Articles on State Responsibility, *supra* note 27, arts. 34-37, at 52. Restitution is reestablishing "the situation which existed before the wrongful act was committed." *Id.* art. 35, at 52. Compensation is covering any financially assessable damage not made good by restitution. *See id.* art. 36, at 52. Satisfaction is "an acknowledgement of the breach, an expression of regret, a formal apology or another appropriate modality" that responds to shortfalls in restitution and compensation when making good the injury caused. *Id.* art. 37, at 52.

Additionally, the victim-state may take "countermeasures" in response to a use of force.[47] Countermeasures are "measures which would otherwise be contrary to the international obligations of the injured State *vis-à-vis* the responsible State if they were not taken by the former in response to an internationally wrongful act by the latter in order to procure cessation and reparation."[48] They are distinguished from retorsion, which is the taking of unfriendly but lawful actions, such as the expulsion of diplomats.

The wrong in question has to be on-going at the time of the countermeasures, since their purpose is not to punish or provide retribution, but instead to compel the other party to desist in its unlawful activities.[49] Countermeasures must be proportionate to the injury suffered,[50] and the victim-state is required to have called on the state committing the wrong to refrain from the conduct (and make reparations if necessary), or, in the case of acts emanating from its territory, take measures to stop them.[51] Unlike collective self-defense (discussed below), countermeasures may only be taken by the state suffering the wrong.[52]

Countermeasures involving cyber operations would be particularly appropriate as a response to a cyber use of force, although the strict limitations placed on countermeasures weaken their viability in situations demanding an immediate reaction. On the other hand, it would be improper to respond with a cyber operation that rose to the level of a use of force, for "[c]ountermeasures shall not affect . . . the obligation to refrain from the threat or use of force as embodied in the Charter of the United Nations."[53] Responses amounting to a use of force are only permissible when falling within the two recognized exceptions to the prohibition on the use of force—action authorized by the Security Council and self-defense.

Although the limitation of countermeasures to non-forceful measures is widely accepted, in a separate opinion to the ICJ's *Case Concerning Oil Platforms*[54] judgment, Judge Simma argued for what might be labeled "self-defense lite" in the face of an "unlawful use of force 'short of' an armed attack . . . within the meaning of Article 51."[55] For Judge Simma, such "defensive military action 'short of' full scale self-defence" is of a "more limited range and quality of response" than that which is lawful in re-

47. *See id.* art. 49(1), at 56; *see also* Gabcikovo-Nagymaros Project (Hung. v. Slovk.) 1997 I.C.J. 7, 55-56 (Sept. 25); *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 249 (June 27).

48. Report of the Int'l Law Comm'n, *supra* note 27, at 324.

49. *See* Draft Articles on State Responsibility, *supra* note 27, art. 52(3)(a), at 57-58.

50. *See id.* art. 51, at 57.

51. *See id.* art. 52(1), at 57.

52. *See Nicaragua* (Nicar. v. U.S.), *1986 I.C.J. 14*, ¶¶ 211, 252.

53. Draft Articles on State Responsibility, *supra* note 27, art. 50(1)(a), at 57.

54. *Oil Platforms* (Iran v. U.S.), 2003 I.C.J. 161 (Nov. 6).

55. *Id.* at 331 (separate opinion of Judge Simma).

sponse to an armed attack in the self-defense context. The key difference with classic self-defense is that Judge Simma would exclude collective actions.[56] Reduced to basics, he is arguing for normative acceptance of forceful countermeasures.

The core problem with the approach is that it posits a tiered forceful response scheme. However, because the intensity of a defensive response is already governed, as will be discussed below, by the principle of proportionality, all that is really occurring is a relaxation of the threshold for engaging in forceful defensive actions. Such an approach is counter-textual, for the combined effect of Articles 2(4) and 51 of the UN Charter is to rule out forcible responses by states against actions other than "armed attacks." Nevertheless, acceptance of such an approach by states would be significant in the cyber context because by it cyber operations, which themselves would be a use of force under Article 2(4), may be launched in reaction to a cyber use of force that did not rise to the level of an armed attack under Article 51.

III. AUTHORIZATION BY THE SECURITY COUNCIL

Pursuant to Article 39 of the UN Charter, the Security Council is empowered to determine that a particular situation amounts to a "threat to the peace, breach of the peace or act of aggression." When it does, the Council "shall make recommendations, or decide what measures shall be taken in accordance with Articles 41 and 42, to maintain or restore international peace and security." Articles 41 and 42 set forth, respectively, non-forceful and forceful options for responding to such situations.

The scope of the phrase "threat to the peace, breach of the peace or act of aggression" has been the subject of much attention in international law. Breach of the peace would seemingly require the outbreak of violence; cyber operations harming individuals or property would reasonably qualify, but whether those falling short of this level would do so is uncertain. As to aggression, in 1974 the General Assembly adopted a resolution in which it characterized "aggression" as ranging from the "use of armed force" and blockade to allowing one's territory to be used by another state to commit an act of aggression and sending armed bands against another state.[57] A cyber operation causing significant physical harm in another state would certainly rise to this level; whether others would is unclear.

This ambiguity is essentially irrelevant in light of the "threat to the peace" criterion. Little guidance exists on those acts which qualify, although they must be conceptually distinguished from activities constituting threats of the use of force in contravention of Article 2(4). In *Tadic*, the ICTY opined that a threat to the peace should be assessed with regard to the Purposes of the United Nations delineated in Article 1 and the Prin-

56. *See id.* at 331-33.

57. Definition of Aggression, G.A. Res. 3314 (XXIX), Annex art. 3, U.N. DOC. A/RES/3314 (Dec. 14, 1974).

ciples set forth in Article 2.[58] This is a singularly unhelpful proposition, since said Purposes and Principles include such intangibles as developing friendly relations and solving social problems.

In fact, a finding that a situation is a threat to the peace is a political decision, not a legal one. It signals the Security Council's willingness to involve itself in a particular matter. There are no territorial limits on situations which may constitute threats to the peace, although they logically tend to be viewed as those which transcend borders, or risk doing so. Nor is there a limitation to acts conducted by or at the behest of states; for instance, the Council has repeatedly found transnational terrorism to be a threat to the peace.[59] No violence or other harmful act need have occurred before the Council may make a threat to the peace determination. Most importantly, since there is no mechanism for reviewing threat to the peace determinations, the Council's authority in this regard is unfettered. Simply put, a threat to the peace is whatever the Council deems it to be. This being so, the Council may label any cyber operation a threat to the peace (or breach of peace or act of aggression), no matter how insignificant.

Once it does, the Security Council may, under Article 41, authorize measures "not involving the use of armed force" necessary to maintain or restore international peace and security. Article 41 offers a number of examples, including "complete or partial interruption of economic relations and of rail, sea, air, postal, telegraphic, radio or other means of communication." Interruption of cyber communications would necessarily be included. An interruption could be broad in scope, as in blocking cyber traffic to or from a country, or surgical, as in denying a particular group access to the Internet. Any other cyber operations judged necessary would likewise be permissible. Given the qualifier "armed force," operations resulting in physical harm to persons or objects could not be authorized pursuant to Article 41.

Should the Council determine that Article 41 measures are proving ineffective, or if before authorizing them it decides that such measures would be fruitless, it may, pursuant to Article 42, "take such action by air, sea, or land forces as may be necessary to maintain or restore international peace and security." The reference to operations by "air, sea, or land forces" plainly contemplates forceful military action, although a Security Council resolution authorizing the use of force will typically be framed in

58. *See* Prosecutor v. Tadic, Case No. IT-94-1-I, Decision on the Defence Motion for Interlocutory Appeal on Jurisdiction, ¶ 29 (Int'l Crim. Trib. for the Former Yugoslavia Oct. 2, 1995).

59. *See, e.g.*, S.C. Res. 1618, U.N. DOC. S/RES/1618 (Aug. 4, 2005); S.C. Res. 1611, U.N. DOC. S/RES/1611 (July 7, 2005); S.C. Res. 1530, U.N. DOC. S/RES/1530 (Mar. 11, 2004); S.C. Res. 1516, U.N. DOC. S/RES/1516 (Nov. 20, 2003); S.C. Res. 1465, U.N. DOC. S/RES/1465 (Feb. 13, 2003); S.C. Res. 1450, U.N. DOC. S/RES/1450 (Dec. 13, 2002); S.C. Res. 1440, U.N. DOC. S/RES/1440 (Oct. 24, 2002); S.C. Res. 1438, U.N. DOC. S/RES/1438 (Oct. 14, 2002); S.C. Res. 1377, U.N. DOC. S/RES/1377 (Nov. 12, 2001).

terms of taking "all necessary measures." To the extent that military force can be authorized, it is self-evident that cyber operations may be as well. It would be lawful to launch them alone or as an aspect of a broader traditional military operation. The sole limiting factors would be the requirement to comply with other norms of international law, such as the IHL prohibition on attacking the civilian population,[60] and the requirement to restrict operations to those within the scope of the particular authorization or mandate issued by the Council. Article 42 actions are not limited territorially or with regard to the subject of the sanctions. For example, it would undoubtedly be within the power of the Council to authorize cyber attacks against transnational terrorist groups (e.g., in order to disrupt logistics or command and control). It is important to emphasize that the measures only extend to restoring peace if breached, or maintaining it when threatened. No authority exists for taking punitive measures.

Pursuant to Article 25 of the Charter, UN Members "agree to accept and carry out the decisions of the Security Council in accordance with the present Charter." This obligation applies even in the face of conflicting domestic or international legal obligations.[61] Consequently, if the Council ordered restrictions, for example, on cyber communications, individual states would be obligated to abide by them and ensure, to the extent feasible, their enforcement on their territory. How they do so is not the concern of the Council, so long as its decision is respected.

Since the United Nations does not itself control cyber networks or have the capability to mount cyber operations, it would have to rely on states to effectuate any cyber related resolutions. Originally, it was envisioned that the Security Council would have dedicated forces at its disposal to conduct Article 42 operations pursuant to "special agreements" with contributing countries.[62] Such arrangements have never been executed. The Council has instead relied upon authorizations granted to individual states, ad hoc coalitions of states, security organizations such as NATO, or UN forces consisting of troop contributions from its Members. State practice has established that no obligation exists for states to provide military forces or finance specific operations that have been authorized. Therefore, if the Council were to endorse specific defensive or offensive cyber operations under Article 42, it would be wholly dependent on the willingness of states to provide the necessary cyber assets and forces to execute them.

Finally, it must be recalled that the entire UN collective security system depends on the readiness of the five Permanent Members of the Security Council (P5) to allow for action by refraining from exercise of their

60. *See* Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts arts. 48, 51, 52, June 8, 1977, 1125 U.N.T.S. 3, 25-27 [hereinafter AP I].

61. *See* U.N. Charter art. 103.

62. *Id.* art. 43.

veto right.[63] In light of Russia's and China's presence on the Council (cyber operations regularly emanate from their territory), this limitation may well prove the greatest obstacle to effective UN action in the face of those cyber operations which would in some fashion endanger international stability.

IV. SELF-DEFENSE

The second recognized exception to the prohibition on the use of force is the right of states to take forceful actions to defend themselves. This customary international law right is codified in Article 51 of the UN Charter. In relevant part, it provides that "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defense if an armed attack occurs against a Member of the United Nations, until the Security Council has taken measures necessary to maintain international peace and security." The article is the *conditio sine qua non* of the Charter, for although Articles 41 and 42 provide Member States some degree of protection from attack, their provisions rely upon implementation by the Security Council. Article 51 represents an essential safeguard in the event the collective security mechanism fails (or proves insufficiently timely), for it provides a means of defense requiring no Security Council approval. In practice, the right of self-defense has proven the principal means by which states ensure their security.

The right of self-defense bears solely on the remedies available to the victim of an armed attack, since all such attacks are "uses of force" in the context of Article 2(4) and customary law, with their legality determined by reference to those norms. By contrast, the issue in self-defense is the lawfulness of a forceful defensive response (including its nature, intensity, duration, and scope) that would otherwise constitute an unlawful use of force by a state. This being so, it has no bearing on passive cyber defenses, which merely block attacks; all such defenses are lawful. It is only in the case of active defenses, whether kinetic or cyber in nature, that the law of self-defense comes into play by directly imposing physical costs on the group or state launching an attack.[64]

Further, states alone enjoy the right of self-defense. Private entities, such as a corporation that has been subjected to a hostile cyber attack, cannot respond pursuant to the law of self-defense regardless of its severity. Their responses would be governed by domestic and international criminal law norms. However, cyber attacks against a state's nationals may

63. *See id.* art. 27(3).

64. Note that one of the recommendations of the experts in the *NATO 2020* report was that "NATO should plan to mount a fully adequate array of cyber defence capabilities, including passive and active elements." *NATO 2020*, *supra* note 7, at 45. It should be noted, however, that passive defenses must nevertheless comport with other aspects of international law. *See, e.g.*, Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (*Wall*), Advisory Opinion, 2004 I.C.J. 136 (July 9).

sometimes qualify as an armed attack on the state itself; there is no requirement in international law that state property or organizations be targeted. In such a case, the state may respond forcefully in self-defense should it choose to do so.

A. *Armed Attack*

The key text in Article 51, and the foundational concept of the customary law right of self-defense, is "armed attack." But for an armed attack, states enjoy no right to respond forcefully to a cyber operation directed against them, even if that operation amounts to an unlawful use of force. This dichotomy was intentional, for it comports with the general presumption permeating the Charter scheme against the use of force, especially unilateral action. In the *Nicaragua* case, the ICJ acknowledged the existence of this gap between the notions of use of force and armed attack when it recognized that there are "measures which do not constitute an armed attack but may nevertheless involve a use of force" and distinguished "the most grave forms of the use of force from other less grave forms."[65] Recall that the court specifically excluded the supply of weapons and logistical support to rebels from the ambit of armed attack, but noted that such actions might constitute uses of force.[66] Simply put, all armed attacks are uses of force, but not all uses of force qualify as armed attacks.

As a result of the gap, the remedies for a use of force not meeting the threshold of armed attack are limited to lawful, non-forceful actions and countermeasures or recourse to the Security Council. What this means in practical terms is that, absent Security Council authorization, a state subjected to a use of force may not respond in kind unless the use of force rises to the level of an armed attack. In light of the difficulties of identifying the source of a cyber operation, this cautious two-tiered system is especially appropriate in the cyber context. It is important to emphasize, however, that once it is established that an armed attack has occurred, no authorization from the Security Council is necessary before defensive actions, including those involving destructive cyber operations, may be mounted.

Consistent with the use of force prohibition, the Charter drafters elected an instrument-based approach to articulating the right of self-defense. And as with that norm, the intent was to preclude certain consequences (in this case, a premature forceful reaction by a state threatened with harm that would itself threaten community stability), while nevertheless allowing states to react forcefully when the consequences justified as much. But, again, the possibility of devastating consequences caused by a non-kinetic cyber attack was obviously not considered during the drafting

65. *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶¶ 191, 210 (June 27); *accord Oil Platforms* (Iran v. U.S.), 2003 I.C.J. 161, 186-87 (Nov. 6).

66. *Nicaragua*, 1986 I.C.J. 14, ¶ 195.

process. Had it been, the drafters would surely have allowed for defense in the face of the severe consequences that can be caused by such attacks.

There is a problem in extending the notion of armed attack to address cyber attack operations of this magnitude. The facts that the use of force language in Article 2(4) is not qualified by the term "armed" and that the phrase "use of force" has been authoritatively interpreted as not necessarily implying a kinetic action allow for interpretive leeway, and the resulting application of the seven factors set forth above. By contrast, the phrase "armed attack" tolerates little interpretive latitude.

Clearly, an armed attack includes kinetic military force. Applying the consequence-based approach, armed attack must also be understood in terms of the effects typically associated with the term "armed." The essence of an armed operation is the causation, or risk thereof, of death of or injury to persons or damage to or destruction of property and other tangible objects. Therefore, while an armed attack need not be carried out through the instrument of classic military force, its consequences (or likely consequences but for successful defensive action) must be analogous to those resulting from its employment. A cyber operation that does not risk these results may qualify as an unlawful use of force, but will not comprise an armed attack permitting forceful defensive action.

In light of the grave consequences that cyber operations can cause without physically harming persons or objects, this interpretation may seem wholly unsatisfactory. Nevertheless, it is the extant law. It must be acknowledged that states victimized by massive cyber attacks, similar to or more aggravated than those suffered by Estonia, may choose to treat them as justifying a forceful response. If state practice along these lines became widespread and well-accepted, the Article 51 norm would shift accordingly through the natural process by which existing international law remains current. For the moment, that has not occurred.

Cyber operations that accompany military action otherwise constituting an armed attack have no bearing on the nature of the attack. For instance, cyber attacks would likely be conducted against enemy command and control or air defense systems as an element of a broader military operation. They can be responded to forcefully, regardless of whether they independently qualify as an armed attack, because they are a component of the overall military action. Similarly, cyber operations that are part of a lawful military response to an armed attack are obviously permissible so long as they comply with IHL, such as the prohibition on attacking civilians or civilian objects.[67] On the other hand, cyber operations need not accompany classic military operations. A cyber attack standing alone will comprise an armed attack when the consequence threshold is reached. Equally, states subjected to an armed attack may elect to respond solely with cyber operations.

67. *See* AP I, *supra* note 60, arts. 48, 51, 52, 1125 U.N.T.S. at 25-27.

In the *Nicaragua* case, the ICJ noted that not all attacks qualify as armed attacks, citing the case of "a mere frontier incident."[68] According to the court, an armed attack must exhibit certain "scale and effects." Unfortunately, the court failed to prescribe criteria by which to resolve whether an attack meets the armed attack threshold. Not only has this proposition been fairly criticized, but in the *Oil Platforms* case the court itself admitted that the mining of even a single ship could amount to an armed attack giving rise to the right of self-defense.[69] Consequently, by contemporary international law, qualitative indicators of attack (death, injury, damage, or destruction) are more reliable in identifying those actions likely to be characterized as an armed attack than quantitative ones (number of deaths or extent of destruction). So long as a cyber operation is likely to result in the requisite consequences, it is an armed attack.

With regard to cyber operations, it must be cautioned that the mere destruction or damage (alteration) of data would not suffice. Were it to, the armed attack threshold would be so reduced that the vast majority of cyber operations would qualify as armed attacks. Rather, to comport with the accepted understanding of "armed attack," the destruction of or damage to the data would have to result in physical consequences, as in causing a generator to overheat and catch fire or rendering a train or subway uncontrollable such that it crashed. Destruction of data designed to be immediately convertible into tangible objects, like banking data, could also be reasonably encompassed within the scope of "armed attacks." But the destruction of or damage to data, standing alone, would not rise to the level of an armed attack.

It is sometimes argued that a cyber operation directed against a nation's military capability necessarily constitutes an armed attack. If the attack is physically destructive, there is no question that this is so. But the mere fact that cyber operations "compromise the ability of units of the DOD to perform DOD's mission" does not alone suffice.[70] Only when non-destructive cyber operations indicate that an attack is imminent ("preparing the battlefield") or represent the first step in an attack that is underway (as in bringing down an air defense radar network to facilitate penetration of enemy airspace) are forceful actions in self-defense permissible. Obviously, it may be difficult to determine whether a particular cyber operation against military assets is either an indication or a component of attack; yet, that is a practical problem which does not affect the norm itself. As with the challenge of identifying an attacker or determin-

68. *Nicaragua (*Nicar v. U.S.*)*, 1986 I.C.J. 14, ¶ 195.

69. *See Oil Platforms* (Iran v. U.S.), 2003 I.C.J. at 195-96; *see also* YORAM DINSTEIN, WAR, AGGRESSION AND SELF-DEFENCE 194-96 (4th ed. 2005); William H. Taft IV, *Self-Defense and the* Oil Platforms *Decision*, 29 YALE J. INT'L L. 295, 300 (2004).

70. NAT'L RESEARCH COUNCIL, TECHNOLOGY, POLICY, LAW, AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES 245 (William A. Owens, Kenneth W. Dam & Herbert S. Lin eds., 2009).

ing when attack is imminent (discussed below), the legal issue is whether the defender's conclusion is reasonable in the circumstances.

Finally, a cyber use of force by State A against State B may generate "bleed-over" effects in State C. This situation does not, as noted earlier, constitute a use of force against C, although A would nevertheless be responsible for the consequences caused. However, if the effects in C rise to the level of those qualifying as an armed attack, C may respond in self-defense against State A, even though C was not the intended target of the attack.[71]

The distinction arises from the fact that while the use of force prohibition solely pertains to the issue of whether there has been a particular violation of international law, the law of self-defense addresses whether a victim-state enjoys the right to employ force to protect itself. It would be incongruous to suggest that a state was barred from acting defensively when subjected to such effects. From its perspective (the correct vantage point in interpreting the law of self-defense), what matters is deterring or stopping the harmful actions; the intention of the actor is but a secondary consideration. Of course, the defensive actions must meet the criteria of self-defense set forth below, in particular the requirement that a forceful response be "necessary." Because C was not the intended target of the attack, it may suffice to simply notify A that it is suffering effects from the attack on B and demand that A takes steps to arrest them.

B. *Anticipatory Self-Defense*

Textually, Article 51 addresses only those situations where an armed attack is underway. Nevertheless, it is well-accepted that a state need not sit idly by as the enemy prepares to attack; instead, a state may defend itself once attack is "imminent."[72] The generally accepted standard of imminency was articulated in the nineteenth century by Secretary of State

71. As the right of self-defense extends to armed attacks by non-state actors, an identical conclusion would apply to actions they undertake against one state having effects in another.

72. Acceptance of the standard is not universal. For instance, Professor Yoram Dinstein argues against its existence, suggesting instead that such actions are better seen as "interceptive self-defense." He notes that "an interceptive strike counters an armed attack which is in progress, even if it is still incipient: the blow is 'imminent' and practically 'unavoidable.'" DINSTEIN, *supra* note 69, at 191. It might also be noted that whereas the notion of *armed* attack was interpreted with fidelity to the Charter text, this article accepts an interpretation of self-defense which runs contrary to the precise text of the UN Charter. The apparent inconsistency can be justified in a number of ways. Note that Article 51 refers to the "inherent" right of self-defense, which has been interpreted as either pre-existing (and thereby maintained in the Charter) or as inherent in the illogic of requiring States to suffer a potentially devastating strike before acting in self-defense. Additionally, Article 2 of the Definition of Aggression resolution provides that the first use of force is merely prima facie evidence of an act of aggression. *See* Definition of Aggression, *supra* note 57, art. 2. As such, it contemplates the possibility of a first use which does not qualify as an armed attack and which, therefore, can only be justified in terms of anticipatory self-defense.

Daniel Webster following the famous *Caroline* incident. In correspondence with his British counterpart regarding an incursion into U.S. territory to attack Canadian rebels during the Mackenzie Rebellion, Webster opined that the right of self-defense applied only when "the necessity of that self-defense is instant, overwhelming, and leaving no moment for deliberation."[73] Although the incident actually had nothing to do with actions taken in anticipation of attack (the attacks in question were ongoing), Webster's formulation has survived as the classic expression of the temporal threshold for anticipatory defensive actions;[74] indeed, the Nuremberg Tribunal cited the *Caroline* case with approval.[75]

Following the events of September 11th, 2001, the United States suggested that a new self-defense paradigm was needed. As President Bush noted in his 2002 National Security Strategy:

> For centuries, international law recognized that nations need not suffer an attack before they can lawfully take action to defend themselves against forces that present an imminent danger of attack. Legal scholars and international jurists often conditioned the legitimacy of pre-emption on the existence of an imminent threat—most often a visible mobilization of armies, navies, and air forces preparing to attack.
>
> We must adapt the concept of imminent threat to the capabilities and objectives of today's adversaries. Rogue states and terrorists do not seek to attack us using conventional means. They know such attacks would fail. Instead, they rely on acts of terror and, potentially, the use of weapons of mass destruction—weapons that can be easily concealed, delivered covertly, and used without warning.[76]

His conclusion was that the "greater the threat, the greater is the risk of inaction—and the more compelling the case for taking anticipatory action to defend ourselves, even if uncertainty remains as to the time and place of the enemy's attack."[77] The United States has maintained this approach to the present.[78]

---

73. Letter from Daniel Webster, U.S. Sec'y of State, to Lord Ashburton, British Special Minister (Aug. 6, 1842), *reprinted in* 2 JOHN BASSETT MOORE, A DIGEST OF INTERNATIONAL LAW 412 (1906).

74. *See, e.g.*, THOMAS M. FRANCK, RECOURSE TO FORCE: STATE ACTION AGAINST THREATS AND ARMED ATTACKS 97 (2002).

75. *See* International Military Tribunal (Nuremberg), Judgment and Sentences, Oct. 1, 1946, *reprinted in* 41 AM. J. INT'L L. 172, 205 (1947).

76. THE WHITE HOUSE, THE NATIONAL SECURITY STRATEGY OF THE UNITED STATES OF AMERICA 15 (2002).

77. *Id.*

78. *See, e.g.*, THE WHITE HOUSE, THE NATIONAL SECURITY STRATEGY OF THE UNITED STATES OF AMERICA 18 (2006). The Obama National Security Strategy does not expressly adopt the doctrine of pre-emption, but nor is it rejected. It specifi-

Despite being characterized by some as revolutionary, even unlawful, the pre-emption doctrine represented a reasonable accommodation to the changed circumstances cited by the President. Indeed, it is arguable that the approach represented a *de minimus* departure from existing law. The underlying premise of anticipatory self-defense is that to effectively defend themselves, states must sometimes act before an aggressive blow falls. Traditionally, a standard requiring temporal proximity to the armed attack had been employed to assess the need. The underlying intent of the standard was to allow as much opportunity as possible for non-forceful measures to work in alleviating the crisis. Yet, as correctly noted in the National Security Strategy, the *modus operandi* of terrorists is to strike without warning, thereby denuding the opportunity the victim-state has to anticipatorily defend itself.

In such circumstances, the most reasonable accommodation of the law of self-defense to both the changed threat and to international law's rebuttable presumption against the legality of using force lies in restricting the victim-state from acting forcefully in self-defense until the point at which its window of opportunity to mount an effective defense is about to close. The imminency criterion should therefore not be measured by reference to the moment of armed attack, but rather with regard to the point at which a state must act defensively, lest it be too late.[79]

The "last feasible window of opportunity" standard must not be interpreted as permitting *preventive* strikes, that is, those against a prospective attacker who lacks either the means to carry out an attack or the intent to do so. The fact that an overtly hostile state is capable of launching cyber attacks—even devastating ones—does not alone entitle a potential victim to act defensively with force. Such hostility must mature into an actual decision to attack. The decision may be evidenced by, for example, preparatory cyber operations amounting to a demonstration of "hostile intent."[80] Moreover, the circumstances must be such that the pending attack has to be responded to immediately if the victim-state is to have any reasonable hope of fending it off. Consider a state's introduction of cyber vulnerabilities into another state's critical infrastructure. Such an action might amount to a use of force, but the victim-state may not react forcefully until it reasonably concludes that: 1) its opponent has decided to

cally reserves the right to act unilaterally. *See* 2010 NATIONAL SECURITY STRATEGY, *supra* note 6, at 22.

79. For a fuller discussion, see Michael N. Schmitt, *Responding to Transnational Terrorism under the* Jus ad Bellum*: A Normative Framework*, 56 NAVAL L. REV. 1, 16-19 (2008).

80. The U.S. Standing Rules of Engagement define hostile intent as "[t]he threat of imminent use of force against the United States, US forces, or other designated persons or property. It also includes the threat of force to preclude or impede the mission and/or duties of US forces, including the recovery of US personnel or vital USG property." CHAIRMAN OF THE JOINT CHIEFS OF STAFF INSTRUCTIONS, CJCSI 3121.01B, STANDING RULES OF ENGAGEMENT/STANDING RULES FOR THE USE OF FORCE FOR U.S. FORCES, at A-4(2005).

actually exploit those vulnerabilities; 2) the strike is likely to generate consequences at the armed attack level; and 3) it must act immediately to defend itself. Until arriving at these conclusions, the victim-state's response would be limited to non-forceful measures, including countermeasures, and referral of the matter to the Security Council.

Although transnational terrorism represents the obvious justification for the approach, cyber operations present many of the same challenges to application of the traditional temporal criterion. Like terrorism, cyber operations are typically launched without any warning that attack is imminent. The time between launch of an operation and impact is measured in seconds at most, thereby often depriving the victim of an opportunity to foil the initial attack as it is unfolding; viable defenses could resultantly be limited to passive measures, such as firewalls and antivirus software. Moreover, although the immediate severity of a cyber armed attack may not reach the level of attacks with weapons of mass destruction, cyber operations have the potential, because of networking, to affect many more individuals and activities. In light of these realities, an approach centering on a state's opportunity to defend itself is no less suitable in the context of cyber operations than in that of terrorism. Cyber or kinetic operations designed to foil an attack which has been approved, and which qualifies as an armed attack, would therefore be lawful when it reasonably appears that failure to act promptly will deprive the target State of any viable opportunity to defend itself.

C. *Criteria for Engaging in Self-Defense*

Actions in self-defense must meet two legal criteria—necessity and proportionality. The ICJ acknowledged both in the *Nicaragua* case, and later confirmed them in its *Oil Platforms* judgement.[81] Necessity requires that there be no reasonable option other than force to effectively deter an imminent attack or defeat one that is underway. This does not mean that force need represent the only available response; it merely requires that defense necessitate actions that are forceful in nature as a component of an overall response, which may well also include non-forceful measures such as diplomacy, economic sanctions, or law enforcement measures.

Proportionality, by contrast, addresses the issue of how much force is permissible once it is deemed necessary. The criterion limits the scale, scope, duration, and intensity of the defensive response to that which is required to neutralize a prospective attack or repel one that is underway. It does not restrict the amount of force used to that employed in the armed attack, since more force may be needed to successfully conduct a defense; of course, less may suffice. In addition, there is no requirement that the defensive force be of the same nature as that constituting the armed attack. Cyber operations may be responded to with kinetic opera-

81. *See Oil Platforms* (Iran v. U.S.), 2003 I.C.J. 161, 183, 196-98 (Nov. 6); *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 194 (June 27).

tions and vice versa. The point of reference is the need to effectively defend oneself, not the character of the armed attack.

The key to the necessity analysis in the cyber context is the existence, or lack thereof, of alternative, non-forceful courses of action. Should passive cyber defenses be adequate to thwart a cyber armed attack, forceful defensive measures would be disallowed. Similarly, if active cyber operations not rising to the level of force are adequate to deter armed attacks (prospective or ongoing), forceful alternatives, whether cyber or kinetic, would be barred. However, when non-forceful measures alone cannot reasonably be expected to defeat an armed attack and prevent subsequent ones, destructive cyber and kinetic operations are permissible under the law of self-defense.

Any forceful defensive cyber or kinetic operations must equally be proportionate. The victim of a cyber armed attack does not have carte blanche to conduct its cyber or kinetic defense. Rather, the extent and nature of its response are limited to ensuring the victim-state is no longer subject to attack. The requirement should not be overstated. It may be that the source of the cyber armed attack is relatively invulnerable to cyber operations. This would not preclude kinetic or cyber defensive operations against other targets in an effort to compel the attacker to desist, although they must be scaled to that purpose.

D. *Evidentiary Issues*

Identification of an "attacker" poses particular problems in the cyber context. For instance, it is possible to "spoof" the origin of attack; the lone indication of where an attack originated from, or who launched it, may be an IP address or other machine discernable data. And the speed by which cyber operations proceed dramatically compresses the time available to make such determinations. How certain must the target state be as to the identity of its attacker before responding in self-defense?

Although international law sets no specific evidentiary standard for drawing conclusions as to the originator of an armed attack, a potentially useful formula was contained in the U.S. notification to the Security Council that it was acting in self-defense when it launched its October 2001 attacks against the Taliban and al-Qaeda in Afghanistan. There, U.S. Ambassador Negroponte stated that "my Government has obtained clear and compelling information that the Al-Qaeda organization, which is supported by the Taliban regime in Afghanistan, had a central role in the attacks."[82] NATO Secretary General Lord Robertson used the same language when announcing that the attacks of 9/11 fell within the ambit of

82. Permanent Rep. of the United States of America to the U.N., Letter dated 7 October 2001 from the Permanent Rep. of the United States of America to the United Nations Addressed to the President of the Security Council, U.N. DOC. S/2001/946 (Oct. 7, 2001) [hereinafter U.S. Rep. Letter].

the collective defense provisions of Article V of the North Atlantic Treaty.[83]

"Clear and compelling" is a threshold higher than the preponderance of the evidence (more likely than not) standard used in certain civil and administrative proceedings and lower than criminal law's "beyond a reasonable doubt." In essence, it obliges a state to act reasonably, that is, in a fashion consistent with the normal state practice in same or similar circumstances. Reasonable states neither respond precipitously on the basis of sketchy indications of who has attacked them nor sit back passively until they have gathered unassailable evidence. So long as the victim-state has taken reasonable steps to identify the perpetrator of an armed attack, cyber or kinetic, and has drawn reasonable conclusions based on the results of those efforts, it may respond forcefully in self-defense. That the state in fact drew the wrong conclusion is of no direct relevance to the question of whether it acted lawfully in self-defense.[84] Its responses are assessed as of the time it took action, not ex post facto.

Although the temporal aspect cannot be ignored, the time available to make the determination is merely one factor bearing on the reasonableness of any conclusion. In particular, automatic "hack-back" systems that might involve a response amounting to a use of force are neither necessarily lawful nor unlawful. Their use must be judged in light of many factors, such as the reliability of the determination of origin, the damage caused by the attack, and the range of available response options.

An analogous standard of reasonableness would apply in the case of anticipatory self-defense against an imminent cyber attack. International law does not require either certainty or absolute precision in anticipating another state's (or non-state actor's) future actions. Rather, it requires reasonableness in concluding that a potential attacker has decided to attack and wields the capability to carry out said attack, and that it must act defensively in anticipation of the attack lest it lose the opportunity to effectively defend itself. States could not possibly countenance a higher threshold, for such a standard would deprive them of a meaningful right of self-defense.

Admittedly, ascertaining a possible adversary's intentions in the cyber environment is likely to be demanding. Aside from the difficulties of accurately pinpointing identity discussed above, it will be challenging in the context of anticipatory self-defense to identify the purpose behind a particular cyber operation. For instance, is a cyber probe of a state's air de-

83. *See* NATO Sec'y Gen. Lord Robertson, Statement at NATO Headquarters (Oct. 2, 2001), *available at* http://www.nato.int/docu/speech/2001/s011002a.htm.

84. Note by way of analogy to international criminal law, that pursuant to the Statute of the International Criminal Court, a mistake of fact is grounds for excluding criminal responsibility when the mistake negates the mental element required by the crime. *See* Rome Statute of the International Criminal Court, art. 32(1), July 17, 1998, 2187 U.N.T.S. 90, 108.

fense designed merely to gather intelligence or instead to locate vulnerabilities in anticipation of an attack which is about to be launched? Obviously, such determinations must be made contextually, considering factors such as the importance of the matter in contention, degree of political tensions, statements by military and political leaders, military activities like deployments, exercises and mobilizations, failed efforts to resolve a contentious situation diplomatically, and so forth. The speed with which the defender may have to make such an assessment to effectively defend itself further complicates matters. Despite the factual and practical complexity, the legal standard is clear; a state acting anticipatorily in self-defense must do so reasonably. In other words, states in the same or similar circumstances would react defensively.

When a state asserts that it is acting in self-defense, it bears the burden of proof. In the *Oil Platforms* case, the ICJ noted that the United States had failed to present evidence sufficient to "justify its using force in self-defense"[85] Specifically, it could not demonstrate that Iran was responsible for a 1987 missile attack against an oil tanker sailing under U.S. flag or the 1988 mining of a U.S. warship during the Iran-Iraq "tanker war," to which the United States responded by attacking Iranian oil platforms. The court rejected evidence offered by the United States which was merely "suggestive," looking instead for "direct evidence" or, reframed, "conclusive evidence."[86] "Clear and compelling" evidence would meet these requirements. Thus, states responding to a cyber armed attack must be prepared to present evidence of this quality as to the source and nature of an impending attack, while those acting in anticipation of an attack must do likewise with regard to the potential attacker's intent and capability.

E. *Collective Responses*

Unlike countermeasures, defensive actions may be collective. This possibility is explicitly provided for in Article 51's reference to "individual or collective self-defense." Collective self-defense may be mounted together by states which have all been attacked or individually by a state (or states) which has not, but comes to the defense of another. Although the basic norm is clear in theory, it is complex in application. As noted in the Group of Experts' Report on the new NATO Strategic Concept, "there may well be doubts about whether an unconventional danger—such as a cyber attack or evidence that terrorists are planning a strike—triggers the collective defence mechanisms of Article 5 [the North Atlantic Treaty implementation of Article 51]."[87]

The mere fact of an armed attack allows for collective defensive action; no authorization from the Security Council is necessary. But there are legal limits on the exercise of the right. In the *Nicaragua* case, the ICJ

85. *Oil Platforms* (Iran v. U.S.), 2003 I.C.J. at 189.
86. *Id.* at 189-90, 194.
87. *NATO 2020*, *supra* note 7, at 20.

suggested that only the victim-state is empowered to determine whether an armed attack has occurred, and it must request assistance before others act on its behalf.[88] Absent such a determination and request, collective actions would themselves amount to unlawful uses of force, and, depending on their nature, even armed attacks (paradoxically, against the state launching the initial armed attack). These requirements are designed to prevent states from claiming to act in collective self-defense as a subterfuge for aggression.

Given the practical difficulties of identifying a cyber operation's originator, this is a sensible limitation. It must be noted that some distinguished commentators challenge the strict application of these requirements. They argue that in cases where the collective self-defense actions occur outside the territory of the victim-state, other states may be entitled to act on the basis of their own right to ensure their security. The right arguably derives from a breach of the duty to refrain from armed attack that the state initiating the armed attack bears.[89] This latter scenario is particularly germane in the cyber context since the effects of cyber armed attacks could easily spread through networks, thereby endangering states other than those which are the intended target. The prevailing view is nevertheless that there must be a request from the victim-state before the right of collective self-defense matures.

In many cases, a pre-existing treaty contemplates collective self-defense. Article 52(1) of the UN Charter provides that "nothing in the present Charter precludes the existence of regional arrangements or agencies for dealing with such matters relating to the maintenance of international peace and security as are appropriate for regional action . . . ." Despite the reference to "regional" arrangements, the agreements need not be limited to states in a particular region or to actions occurring in a defined area. Such arrangements may take multiple forms. For instance, bilateral and multilateral mutual assistance treaties typically provide that the Parties will treat an armed attack against one of them as an armed attack against all.[90] As a practical matter, the effectiveness of collective self-defense provisions usually depends on the willingness of the treaty partners to come to each other's aid. A state that does not see collective self-defensive action as in

88. *See Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 199 (June 27). The Court reiterated this position in the *Oil Platforms* case of 2003. *See Oil Platforms* (Iran v. U.S.), 2003 I.C.J. at 188.

89. *See* DINSTEIN, *supra* note 69, at 270. This was the position adopted in Judge Jenning's dissent in *Nicaragua*. *See Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. at 544-46 (dissenting opinion of Judge Sir Robert Jennings).

90. For instance, the Japan-United States mutual defense treaty provides that "[e]ach Party recognizes that an armed attack against either Party in the territories under the administration of Japan would be dangerous to its own peace and safety and declares that it would act to meet the common danger in accordance with its constitutional provisions and processes." Treaty of Mutual Cooperation and Security Between Japan and the United States of America, U.S.–Japan, art. V, Jan. 19, 1960, 11 U.S.T. 1632.

its national interest may be expected to contest characterization of a cyber operation as an armed attack.

Military alliances based on the right to engage in collective self-defense also exist, the paradigmatic example being the North Atlantic Treaty Organization (NATO). Pursuant to Article V of the treaty, Member States

> agree that an armed attack against one or more of them in Europe or North America shall be considered an attack against them all and consequently they agree that, if such an armed attack occurs, each of them, in exercise of the right of individual or collective self-defence recognised by Article 51 of the Charter of the United Nations, will assist the [Member State] or [Member States] so attacked by taking forthwith, individually and in concert with the other [Member States], such action as it deems necessary, including the use of armed force, to restore and maintain the security of the North Atlantic area.[91]

The benefit of alliances is that they generally involve a degree of advanced planning for combined operations in the event of armed attack, and, as with NATO, military structures are often set up to coordinate and direct military operations. Preplanning and the existence of collective mechanisms for managing joint and combined action are especially valuable with regard to defending against cyber attacks. However, like mutual assistance treaties, alliance arrangements are subject to the reality that they are composed of states, which can be expected to act pursuant to their own national interests. In the case of NATO, for instance, decisions to act are taken by consensus in the North Atlantic Council; a single Member State can therefore block NATO collective action. Indeed, had the cyber operations against Estonia risen to the level of an armed attack, it is not altogether certain that NATO would have come to its defense militarily, especially in light of Russia's place in the European security environment and the countervailing commitments of NATO allies elsewhere, especially Afghanistan and Iraq.

F. *State Sponsorship of Attacks by Non-State Actors*

The issue of state sponsorship of cyber operations was addressed earlier in the context of the responsibility of states for uses of force by non-state actors. There the question was, When does a state violate the use of force prohibition by virtue of its relationship with others who conduct cyber operations? However, the issue of state sponsorship in the self-defense context is much more momentous. It asks when may forceful defensive actions, even kinetic ones, be taken against a state which has not engaged in cyber operations, but which has "sponsored" them? In other

91. North Atlantic Treaty, art. 5, Apr. 4, 1949, 63 Stat. 2241, 34 U.N.T.S. 243, 246.

words, when is an armed attack attributable to a state such that the state may be treated as if it had itself launched the attack?

Until the transnational attacks of September 11, 2001, the generally accepted standard was set forth in the *Nicaragua* case. There the ICJ stated that

> an armed attack must be understood as including not merely action by regular forces across an international border, but also "the sending by or on behalf of a State of armed bands, groups, irregulars or mercenaries, which carry out acts of armed force against another State of such gravity as to amount to" (inter alia) an actual armed attack conducted by regular forces, "or its substantial involvement therein."[92]

The court noted that the activities involved should be of a "scale and effects" that would equate to an armed attack if carried out by the state's military. Thus, "acts by armed bands where such attacks occur on a significant scale" would qualify, but "a mere frontier incident would not."[93]

By this standard, attribution requires (1) acts qualifying as an armed attack and (2) that the state dispatched the non-state actors or was substantially involved in the operations. As noted earlier, the ICTY took a more relaxed view of the degree of control necessary, accepting "overall control" as sufficient.[94] The events of 9/11 brought the issue of threshold to light in a dramatic way. Assistance provided by the Taliban to al-Qaeda met neither the *Nicaragua* nor *Tadic* standards, since the Taliban merely provided sanctuary to al-Qaeda. The cyber analogy would be doing nothing to put an end to the activities of cyber "terrorists" or other malicious hackers operating from a state's territory when it is within its capability, legal and practical, to do so.

Even though there was seemingly no legal basis for attribution to Afghanistan, when the Coalition responded with armed force against both al-Qaeda and the governing Taliban, no objection was raised. On the contrary, the Security Council condemned the Taliban "for allowing Afghanistan to be used as a base for the export of terrorism by the Al-Qaida network and other terrorist groups and for providing safe haven to Usama Bin laden, Al-Qaida and others associated with them."[95] It seems that the international community had lowered the normative bar of attribution measurably. While the underlying operations must still amount to an armed attack, it is arguable that today much less support is required for attribution than envisaged in either *Nicaragua* or *Tadic*. Far from being counter-legal, this process of reinterpretation is natural; understandings of

92. *Nicaragua* (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 195 (citation omitted).

93. *Id.*

94. It must be emphasized that the legal issue involved in that case was not attribution of an armed attack, but rather the existence of an international armed conflict.

95. S.C. Res. 1378, pmbl., U.N. DOC. S/RES/1378 (Nov. 14, 2001).

international legal norms inevitably evolve in response to new threats to the global order. In that cyber operations resemble terrorism in many regards, states may equally be willing to countenance attribution of a cyber armed attack to a state which willingly provides sanctuary to non-state actors conducting them.

G. *Armed Attacks by Non-State Actors*

Although most cyber operations are launched by individuals such as the anti-Estonian "hacktivists," concern is mounting about the prospect that transnational terrorist organizations and other non-state groups will turn to cyber operations as a means of attacking states.[96] The concern is well-founded. Al-Qaeda computers have been seized that contain hacker tools, the membership of such groups is increasingly computer-literate, and the technology to conduct cyber operations is readily available. In one case, a seized al-Qaeda computer contained models of dams, a lucrative cyber attack target, and the computer programs required to analyze them.[97]

International lawyers have traditionally, albeit not universally, characterized Article 51 and the customary law of self-defense as applicable solely to armed attacks mounted by one state against another. Violent actions by non-state actors fell within the criminal law paradigm. Nonetheless, the international community treated the 9/11 attacks by al-Qaeda as armed attacks under the law of self-defense. The Security Council adopted numerous resolutions recognizing the applicability of the right of self-defense.[98] International organizations such as NATO and many individual states took the same approach.[99] The United States claimed the right to act forcefully in self-defense,[100] and no state objected to the assertion. Lest this approach be dismissed as simply an emotive reaction to the horrific attacks of 9/11, it must be noted that when Israel launched operations into Lebanon in response to Hezbollah's 2006 terrorism, the

96. *See* 2010 NATIONAL SECURITY STRATEGY, *supra* note 6, at 27; *NATO 2020*, *supra* note 7, at 17.

97. *See* CLAY WILSON, CONG. RESEARCH SERV., RL32114, COMPUTER ATTACK AND CYBER TERRORISM: VULNERABILITIES AND POLICY ISSUES FOR CONGRESS, 11-13 (2003).

98. *See, e.g.*, S.C. Res. 1373, U.N. DOC. S/RES/1373 (Sept. 28, 2001); S.C. Res 1368, U.N. DOC. S/RES/1368 (Sept. 11, 2001).

99. *See, e.g.*, Org. of American States, *Terrorist Threat to the Americas*, OAS DOC. RC.24/RES.1/01 (Sept. 21, 2001); Brendan Pearson, *PM Commits to Mutual Defence,* AUSTRALIAN FIN. REV., Sept. 15, 2001, at 9; Press Release, NATO, Statement by the North Atlantic Council (Sept. 12, 2001), *available at* http://www.nato.int/docu/pr/2001/p01-124e.htm.

100. "In response to these attacks, and in accordance with the inherent right of individual and collective self-defence, United States armed forces have initiated actions designed to prevent and deter further attacks on the United States. These actions include measures against Al-Qaeda terrorist training camps and military installations of the Taliban regime in Afghanistan." U.S. Rep. Letter, *supra* note 82.

international community again seemed to accept a country's right to defend itself against armed attacks mounted by non-state actors.[101]

Despite acceptance by states of the premise that non-state actors may qualify as the originators of an armed attack, the ICJ seems to have taken a step backwards in two post-9/11 cases. In the advisory opinion on the *Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory*[102] and the *Armed Activities on the Territory of the Congo*[103]case, the ICJ refrained from considering claims of self-defense against actions by non-state actors, noting that no assertion had been made that the relevant actions were imputable to a state.[104] Although the court's reasoning was nuanced and fact-specific, it has nevertheless been widely criticized as inattentive to contemporary understandings of the relevant law. In particular, in the *Wall* case, three judges expressly departed from the majority's approach on the bases that it ignored the fact that Article 51 makes no mention of the originator of an attack (while Article 2(4) specifically addresses uses of force by states) and that the Security Council had deliberately treated terrorist attacks as armed attacks in the aftermath of the 9/11.[105]

The court's hesitancy to embrace the notion of armed attack by non-state actors is understandable in light of the risk of abuse. States might well apply it to engage in robust military operations against groups in situations in which law enforcement is the more normatively appropriate response. For instance, significant concerns have been raised regarding counterterrorist operations occurring outside an armed conflict mounted in states which do not consent to them. Such concerns are likely to be even more acute in relation to cyber operations, which are conducted not by armed members of groups resembling classic military forces, but rather by cyber experts equipped with computers. Nevertheless, as a matter of law, states seem comfortable with applying the concept of armed attacks to situations involving non-state actors. Should such groups launch cyber attacks meeting the threshold criteria for an armed attack, states would likely respond within the framework of the law of self-defense.

The point that the attacks must meet the threshold criteria cannot be overemphasized. There is no state practice supporting extension of the concept to the actions of isolated individuals, such as hacktivists or patri-

101. *See generally* Michael N. Schmitt, *"Change Direction" 2006: Israeli Operations in Lebanon and the International Law of Self-Defense*, 29 MICH. J. INT'L L. 127 (2008). Many commentators and States saw the actions as violating the proportionality criterion discussed above.

102. Legal Consequences of the Construction of a Wall in the Occupied Palestinian Territory (*Wall*), Advisory Opinion, 2004 I.C.J. 136 (July 9).

103. Armed Activities on the Territory of the Congo (*Congo*) (Dem. Rep. Congo v. Uganda), 2005 I.C.J. 168, ¶ 160 (Dec. 19).

104. *See Congo* (Dem. Rep. Congo v. Uganda), Judgment, 2005 I.C.J. 168, 217 (Dec. 19); *Wall*, Advisory Opinion, 2004 I.C.J. 136, ¶ 139 (July 9).

105. *See Wall*, 2004 I.C.J. 136, ¶ 6 (separate declaration of Judge Buergenthal); *id.* ¶ 33 (separate opinion of Judge Higgins); *id.* ¶ 35 (separate opinion of Judge Koojmans).

otic hackers. Further, the cyber operations must be severe enough to qualify as armed attacks, that is, they have to result in damage to or destruction of property or injury to or death of individuals. Finally, as the debate over minor border incursions demonstrates, it is uncertain whether attacks which meet the aforementioned threshold, but are not of significant scale, would qualify. As an example, a cyber attack that caused a single plant's generator to overheat, thereby temporarily interrupting service until it could be repaired, would presumably not, by the more restrictive standard, qualify as an armed attack. Rather, it would be the cyber equivalent of a border incursion.

H. *Cross-Border Operations*

When armed attacks by non-state actors emanate from outside a state, may that state take defensive actions against its perpetrators in the territory of the state where they are based? This question has been raised recently in the context of unmanned aerial vehicle strikes against terrorists in Pakistan and elsewhere. It is no less pertinent to situations involving cyber armed attacks launched by non-state actors from abroad.

It is indisputable that one state may employ force in another with the consent of the territorial state. For instance, a state may grant others the right to enter its territory to conduct counterterrorist operations, as often occurs in Pakistan, or a state embroiled in an internal conflict with insurgents may request external assistance in restoring order, as with International Security Assistance Force (ISAF) operations in Afghanistan or United States Forces (USF) in Iraq. A State subjected to an armed attack, whether cyber or kinetic, could, with the acquiescence of the territorial state, equally launch cyber defensive operations into the state from which the attacks emanated.

The legal dilemma arises when operations are conducted without territorial state approval. By the principle of sovereignty (and the derivative notion of territorial integrity), a state enjoys near absolute control over access to its territory. In affirmation, the UN General Assembly has cited the use of force by a state on the territory of another as an act of aggression.[106] Yet, the right of states to use force in self-defense is no less foundational. When terrorists or insurgents seek sanctuary in a state other than that in which they are conducting operations, they bring the territorial state's right of sovereignty into conflict with the victim-state's right of self-defense.

Fortunately, international law does not require an either-or resolution when norms clash. Instead, it seeks to balance them by fashioning a compromise which best achieves their respective underlying purposes. In this case, such a balance would ensure that the territorial state need not suffer unconstrained violations of its sovereignty, but nor would the victim-state have to remain passive as non-state groups attack it with impunity from

106. *See* Definition of Aggression, *supra* note 57, art. 3(a).

abroad. The resulting compromise is as follows. The victim-state must first demand the territorial state fulfill its legal duty to ensure actions on or from its territory do not harm other states and afford the territorial state an opportunity to comply.[107] If that state subsequently takes effective steps to remove the threat, then penetration of its territory by the victim-state, whether kinetically or by cyber means, is impermissible. But if the territorial state fails to take appropriate and timely action, either because it lacks the capability to conduct the operations or simply chooses not to do so (e.g., out of sympathy for the non-state actors or because its domestic laws preclude action), the victim-state may act in self-defense to put an end to the non-state actor's attacks. It matters not whether the actions are kinetic or cyber in nature, as long as they comply with the principles of proportionality and necessity.

V. FAULT LINES IN THE LAW

The legal analysis set forth above should strike most readers as unsatisfactory. Clear fault lines in the law governing the use of force have appeared because it is a body of law that predates the advent of cyber operations. The normative scheme made sense when close congruity existed between the coercive instruments of international relations, particularly military force, and their effects. To the extent one state disrupted order in the international community, it usually did so by using force to harm objects and persons. Resultantly, instrument-based normative shorthand (use of *force, armed* attack, and *armed* conflict) was employed as a means of precluding those effects (death, injury, destruction, and damage) which were perceived as most disruptive of community stability, and as most threatening to state security. Debates such as whether actions short of military operations are uses of force or whether minor border incursions qualify as armed attacks demonstrate that the foundational concerns were actually consequence-based, for both reflect recognition that the instrument-based approach is not perfectly calibrated.

The advent of cyber operations threw the instrument-based approach into disarray by creating the possibility of dramatically destabilizing effects caused by something other than kinetic actions. They weakened the natural congruency between the normative shorthand employed in the law governing resort to force and those consequences which the law sought to avoid as disruptive. Conceptually, the qualitative scheme, by which prohibitions were expressed in terms of types of activities (use of the military and other destructive instruments as distinguished from non-destructive ones), no longer sufficed to preclude those effects about which states had become most concerned. A non-kinetic, non-destructive means of generating effects which states cannot possibly countenance now existed;

107. On the duty to police one's own territory, see Corfu Channel (U.K. v. Alb.), 1949 I.C.J. 4 (Apr. 9).

the qualitative shorthand no longer tracked the quantitative concerns of states.

The prohibition on the use of force has proven somewhat adaptable to this new reality because it has long been understood to extend beyond the application of kinetic force. Thus, it is reasonable to employ the criteria suggested in this Article to identify situations in which non-kinetic actions will result in quantitatively unacceptable, and therefore prohibited, consequences. The UN Charter mechanism for Security Council-based responses to threats to the peace, breaches of the peace, and acts of aggression is likewise adaptable because by it threats to the peace include, simply put, whatever the Council wishes.

Evidence of disquiet abounds. In a recent report by the National Academy of Science, examples of armed attack included "cyberattacks on the controlling information technology for a nation's infrastructure that had a significant impact on the functioning of that infrastructure (whether or not it caused immediate large-scale death or destruction of property)" and "a cyberattack against the stock exchanges that occurs repeatedly and continuously, so that trading is disrupted for an extended period of time (e.g., days or weeks)."[108] As a matter of law, they would likely qualify as uses of force, but not, by a strict interpretation of the self-defense norm, as armed attacks (or as initiating an armed conflict). The problem is that most states would surely treat them as such. In other words, the National Academy report has misconstrued the law, but accurately identified probable state behavior.

When state expectations as to the "rules of the game" deviate from those that actually govern their actions, new norms can emerge. One method by which this can occur is through new treaty law. However, it is highly unlikely that any meaningful treaty will be negotiated to govern cyber operations in the foreseeable future. The greatest obstacle is that those states which are most vulnerable to cyber operations tend to be those which are also most capable of conducting them. Such tension will cause such states to hesitate before agreeing to prohibitions designed to protect them which may also definitively limit their freedom of action. This is especially so in light of the nascent nature of cyber warfare and the lack of experience of most states in these operations. In international relations, states are often comfortable with a degree of vagueness.

Much more likely is the emergence of new understandings of the existing treaty law which are responsive to the realities of cyber operations. While only subsequent treaty action can technically alter a treaty's terms, state practice can inform their interpretation over time. A well-known example involves veto action by permanent members of the Security Council. The UN Charter provides that a binding resolution of the Council requires the affirmative vote of all five permanent members.[109] However,

108. NAT'L RESEARCH COUNCIL, *supra* note 70, at 254-55.
109. *See* U.N. Charter art. 27(3).

state practice has been to treat the provision as blocking action only when a member of the "P5" vetoes a proposed resolution. This counter-textual interpretation is now accepted as the law.[110] The recent extension of the notion of armed attack to actions by non-state actors similarly illustrates normative evolution prompted by shifting state expectations.

In due course, similar evolution in how the concept of armed attack is understood should be anticipated, as states increasingly accept the proposition that armed attacks must be judged qualitatively *and* quantitatively. Consequences will remain the focus of concern, but they will be assessed both in terms of their nature and as to their impact on affected states. In this regard, the seven criteria proffered above in the use of force context can serve as useful indicators of whether states are likely to characterize particular cyber operations as armed attacks (or as initiating an armed conflict), and thus suggest the probable vector of the law. However, for the moment the existing law remains intact; it will be left to states to articulate the expectations and engage in practices that can serve to fuel the normative process necessary to transform *lex ferenda* into *lex lata*.[111]

110. *See* Bruno Simma, Stefan Brunner & Hans-Peter Kaul, *Article 27, in* 1 THE CHARTER OF THE UNITED NATIONS: A COMMENTARY 476, 493-98 (Bruno Simma ed., 2d ed. 2002). The veto principle does not apply to votes on procedural matters.

111. The law as it should be and the law that is, respectively.

The Law of Cyber-Attack

Oona A. Hathaway, Rebecca Crootof, Philip Levitz, Haley Nix, Aileen Nowlan, William Perdue & Julia Spiegel*

Cyber-attacks have become increasingly common in recent years. Capable of shutting down nuclear centrifuges, air defense systems, and electrical grids, cyber-attacks pose a serious threat to national security. As a result, some have suggested that cyber-attacks should be treated as acts of war. Yet the attacks look little like the armed attacks that the law of war has traditionally regulated. This Article examines how existing law may be applied—and adapted and amended—to meet the distinctive challenge posed by cyber-attacks. It begins by clarifying what cyber-attacks are and how they already are regulated by existing bodies of law, including the law of war, international treaties, and domestic criminal law. This review makes clear that existing law effectively addresses only a small fraction of potential cyber-attacks. The law of war, for example, provides a useful framework for only the very small number of cyber-attacks that amount to an armed attack or that take place in the context of an ongoing armed conflict. This Article concludes that a new, comprehensive legal framework at both the domestic and international levels is needed to more effectively address cyber-attacks. The United States could strengthen its domestic law by giving domestic criminal laws addressing cyber-attacks extra-territorial effect and by adopting limited, internationally permissible countermeasures to combat cyber-attacks that do not rise to the level of armed attacks or that do not take place during an ongoing armed conflict. Yet the challenge cannot be met by domestic reforms alone.

* Gerard C. and Bernice Latrobe Smith Professor of International Law, Yale Law School; law clerk, Judge Mark Kravitz (D. Conn.); J.D., Yale Law School, 2012; J.D., Yale Law School, 2012; J.D., Yale Law School, 2012; Associate, Arnold & Porter LLP; J.D. Candidate, Yale Law School, and MPA Candidate, Woodrow Wilson School, Princeton University, respectively. We thank Sara Solow, Elizabeth Nielsen, Chelsea Purvis, Saurabh Sanghvi, and Teresa Miguel for their assistance in preparing this Article. We thank participants in the George Washington University School of Law Symposium on the Future of Cyber-Warfare and in the Harvard National Security Journal and Harvard University National Security & Law Association Annual Symposium for their very helpful comments and suggestions.

International cooperation will be essential to a truly effective legal response. New international efforts to regulate cyber-attacks must begin with agreement on the problem—which means agreement on the definition of cyber-attack, cyber-crime, and cyber-warfare. This would form the foundation for greater international cooperation on information sharing, evidence collection, and criminal prosecution of those involved in cyber-attacks—in short, for a new international law of cyber-attack.

INTRODUCTION

In 2010, Iran's nuclear program ground to a halt, the subject of a sophisticated attack that sent centrifuges spinning wildly out of control. The weapon? Stuxnet, a computer "worm" that appears to have many authors from around the world and was likely tested by Americans and Israelis at the Israeli Dimona complex in the Negev desert.[1]

A few months later, a so-called "distributed denial of service" attack took the entire population of Burma off the Internet immediately preceding the country's first national election in twenty years.[2] Observers suspect that the military junta in Burma coordinated the attack to shut down the Internet and thereby restrict the free flow of information,[3] but American public officials

1. The seeds for this attack were apparently sown well before 2010. The worm was first detected in 2008, when it infected networks around the world. It did no damage to most systems. At first, it was assumed that the attack, which appeared to target nuclear facilities in Iran, was not successful. Yet, in the fall of 2010, reports spread that Iran's uranium enriching capabilities had been diminished. *The Stuxnet Worm: A Cyber-Missile Aimed at Iran?*, ECONOMIST BABBAGE BLOG (Sept. 24, 2010, 1:32 PM), http://www.economist.com/blogs/babbage/2010/09/stuxnet_worm; *see also* Jonathan Fildes, *Stuxnet Worm 'Targeted High-Value Iranian Assets,'* BBC NEWS (Sept. 23, 2010, 6:46 AM), http://www.bbc.co.uk/news/technology-11388018; William J. Broad et al., *Israeli Tests Called Crucial in Iran Nuclear Setback*, N.Y. TIMES, Jan. 16, 2011, at A1. Stuxnet is the first computer virus known to be capable of specifically targeting and destroying industrial systems such as nuclear facilities and power grids. Fildes, *supra*.

2. *Burma Hit by Massive Net Attack Ahead of Election*, BBC NEWS (Nov. 4, 2010, 3:33 PM), http://www.bbc.co.uk/news/technology-11693214.

3. *See id.*

have resisted blaming the attack on the government, even as they have criticized the election.[4]

In the summer of 2011, evidence emerged of a long-suspected government-sanctioned cyber-attack program in China. In late August, a state television documentary aired on the government-run China Central Television appeared to capture an in-progress distributed denial of service attack by China's military on a Falun Gong website based in Alabama.[5] This revelation followed on the heels of a report by the McAfee cyber-security company suggesting that a "state actor"—widely believed to be China—had engaged in a years-long cyber-attack program aimed at a range of governments, U.S. corporations, and United Nations groups.[6]

What law governs these attacks? Some have referred to these and similar attacks as "cyber-warfare," suggesting that the law of war might apply.[7] Yet the attacks look little like the armed conflict that the law of war traditionally regulates. And if they are "warfare," does that mean that victims of such attacks might claim the right to use conventional force in self-defense—potentially legally authorizing Iran, for example, to respond to Stuxnet with a physical attack?

This Article examines these questions and, in the process, offers new insights into how existing law may be applied—and adapted and amended—to meet the distinctive challenge posed by cyber-attacks. It does so in two principal ways. First, the Article clarifies what cyber-attacks are and how they relate to existing bodies of law, including the law of war;[8] recent international

4. *See, e.g.*, Barack Obama & Michelle Obama, Remarks by the President and the First Lady in Town Hall with Students in Mumbai, India (Nov. 7, 2010), *available at* http://www.whitehouse.gov/the-press-office/2010/11/07/remarks-president-and-first-lady-town-hall-with-students-mumbai-india; Barack Obama, Statement by President Obama on Burma's November 7 Elections (Nov. 7, 2010), *available at* http://www.whitehouse.gov/the-press-office/2010/11/07/statement-president-obama-burmas-november-7-elections.

5. Ellen Nakashima & William Wan, *China's Denials on Cyberattacks Undercut*, WASH. POST, Aug. 24, 2011, at A12.

6. David Barboza & Kevin Drew, *Security Firm Sees Global Cyberspying*, N.Y. TIMES, Aug. 3, 2011, at A11. This was not the first suggestion of a program of cyber-attacks on private and government actors by China. Computer attacks on Google that originated in China were believed to be part of a broader political and corporate espionage effort and prompted Google to withdraw from the Chinese market. Ariana Eunjung Cha & Ellen Nakashima, *Google Attack Part of Vast Campaign; Targets Are of Strategic Importance to China, Where Scheme Is Thought to Originate*, WASH. POST, Jan. 14, 2010, at A1.

7. *See, e.g.*, RICHARD A. CLARKE & ROBERT K. KNAKE, CYBER WAR: THE NEXT THREAT TO NATIONAL SECURITY AND WHAT TO DO ABOUT IT 6 (2010); Stephen Dycus, *Congress's Role in Cyber Warfare*, J. NAT'L SECURITY L. & POL'Y 155, 162 (2010) ("Cyber warfare, as that term is used here, refers to conflicts that utilize cyber or electronic weapons either offensively or defensively, or both."); *Understanding Cyber Warfare*, LAWS.COM, http://cyber.laws.com/cyber-warfare (last visited Apr. 18, 2012).

8. For simplicity's sake, this Article refers collectively to *jus in bello* and *jus ad bellum* as the "law of war."

efforts to directly regulate cyber-attacks; international bodies of law that may be used to indirectly regulate cyber-attacks; and domestic criminal law.

Second, the Article demonstrates how existing law is deficient and what needs to be done to improve it. Although such bodies of law do offer some tools for responding to cyber-attacks, these tools are far from complete or adequate. The law of war, for example, provides a useful legal framework for regulating only the very small slice of cyber-attacks that amount to an armed attack or that take place in the context of an ongoing armed conflict. Other existing legal frameworks—both domestic and international—offer equally fragmentary assistance in policing cyber-attacks through law. Examining existing law leads to a clear conclusion: a new, comprehensive legal framework is needed to address cyber-attacks.

The terms "cyber-attack," "cyber-warfare," and "cyber-crime" are frequently used with little regard for what they are meant to include. This lack of clarity can make it all the more difficult to design a meaningful legal response. We therefore begin this Article in Part I by defining these terms. This may seem a mundane task, but it is a critical starting point for any reform effort. To that end, we define "cyber-attack" as "any action taken to undermine the functions of a computer network for a political or national security purpose." We also explain the difference between "cyber-attacks," "cyber-warfare," and "cyber-crime," and describe three common forms of cyber-attacks: distributed denial of service attacks, planting inaccurate information, and infiltration of a secure computer network.[9]

In Part II, we turn to examining how the law of war might govern cyber-attacks. We parse the way the law of war, most of which was developed at a time when cyber-attacks were inconceivable, applies to this new zone of conflict. We first apply *jus ad bellum*—the law governing a state's right to resort to armed force—to cyber-attacks. We conclude that most cyber-attacks do not rise to the level of an armed attack and thus do not justify the use of armed force in response. "Cyber-warfare" is thus a term properly used only to refer to the small subset of cyber-attacks that do constitute armed attacks or that occur in the context of an ongoing armed conflict. This definition is crucial because it limits the application of the "war" framework to those actions that actually constitute "war" as a matter of international law. With the scope of cyber-warfare clear, we then explore how *jus in bello*—the law governing conduct in an armed conflict—applies.

9. This definition differs from that currently applied by the U.S. Cyber Command, which uses the term "Cyber Attack" to apply only to attacks that cause physical damage to property or injury to persons. E-mail from Gary D. Brown, Col. U.S. Airforce, Staff Judge Advocate, U.S. Cyber Command to author (May 15, 2012 10:07AM) (on file with author). Our terminology allows differentiation between those attacks that are covered by the law of armed conflict (which we call cyber-warfare) and those that violate the norm of nonintervention but are not covered by the law of armed conflict (which we call cyber-attack).

Because the law of war regulates only a small subset of cyber-attacks, in Part III we examine other existing legal regimes that could regulate cyber-attacks. These include (1) the law of countermeasures, which governs how states may respond to international law violations that do not justify uses of force in self-defense; (2) international agreements and other cooperative efforts to directly regulate cyber-attacks; (3) international agreements that regulate means or locations of cyber-attacks, including telecommunications, aviation, space, satellites, and the sea; and (4) U.S. criminal law regulating cyber-attacks. We conclude that, as with the law of war, these existing bodies of law effectively address only a small part of the problem—leaving many harmful cyber-attacks unregulated and uncontrolled by either domestic or international law.

Finally, in Part IV we consider how the problem of cyber-attacks might be more effectively addressed, offering recommendations for both domestic and international reforms. At the domestic level, states may expand the extraterritorial reach of domestic criminal law and develop plans for the deployment of customary countermeasures in response to cyber-attacks. Yet an effective solution to this global challenge cannot be achieved by individual states acting alone. It will require global cooperation. We therefore outline the key elements of a cyber-treaty—namely, codifying clear definitions of cyber-warfare and cyber-attack and providing guidelines for international cooperation on evidence collection and criminal prosecution—that would provide a more comprehensive and long-term solution to the emerging threat of cyber-attacks.

I.
WHAT IS A CYBER-ATTACK?

The first challenge in evaluating how domestic and international law might be used to address cyber-attacks is to determine the nature and scope of the problem we face. Activities in cyberspace defy many of the traditional categories and principles that govern armed conflict under the law of war. This Part first offers a precise definition of "cyber-attack." This step is not only necessary to the legal analysis that follows, but it also fills a gap in the existing literature, which often uses the term without clarifying what it is meant to include and exclude. We then offer three categories of activities that fall within this definition, illuminating the extraordinary range of activities that fall under even a carefully constructed and limited definition of "cyber-attacks." This serves as a prelude to an analysis of what portion of cyber-attacks are governed by the law of war and other existing bodies of law.

A. Defining "Cyber-Attack"

For well over a decade, analysts have speculated about the potential consequences of a cyber-attack. The scenarios—ranging from a virus that

scrambles financial records or incapacitates the stock market,[10] to a false message that causes a nuclear reactor to shut off[11] or a dam to open,[12] to a blackout of the air traffic control system that results in airplane crashes[13]—anticipate severe and widespread economic or physical damage. While none of these scenarios has thus far occurred, numerous cyber-incidents occur regularly.[14] Nevertheless, there is no settled definition for identifying these incidents as cyber-attacks,[15] much less as cyber-warfare. The absence of a shared definition has made it difficult for analysts from different countries to develop coordinated policy recommendations and for governments to engage in coordinated action. Hence the technical project of defining cyber-attack is an important first step toward addressing the growing threat posed by cyber-attacks. After describing some existing definitions, we offer a definition that effectively encompasses the activity that lies at the heart of the concerns raised by cyber-attacks.[16]

1. Existing Conceptions of Cyber-Attack

Existing definitions of "cyber-attack" and related terms vary widely. Perhaps one of the most widely cited definitions comes from government security expert Richard A. Clarke, who defines cyber-war as "actions by a nation-state to penetrate another nation's computers or networks for the purposes of causing damage or disruption."[17] Similarly, former NSA and CIA director Michael Hayden has spoken of cyber-war as the "deliberate attempt to disable or destroy another country's computer networks."[18] These definitions, however, do not distinguish between a cyber-crime, cyber-attack, and cyber-war.[19] As a result, they are open to a dangerously broad application of the war framework in the cyber context.[20] In addition, Clarke's definition is too narrow

10. Duncan B. Hollis, *Why States Need an International Law for Information Operations*, 11 LEWIS & CLARK L. REV. 1023, 1042 (2007).

11. Vida M. Antolin-Jenkins, *Defining the Parameters of Cyberwar Operations: Looking for Law in All the Wrong Places?*, 51 NAVAL L. REV. 132, 140 (2005).

12. Barton Gellman, *Cyber Attacks by al Qaeda Feared; Terrorists at Threshold of Using Internet as Tool of Bloodshed, Experts Say*, WASH. POST, June 27, 2002, at A1.

13. U.S. GEN. ACCOUNTING OFFICE, AIR TRAFFIC CONTROL: WEAK COMPUTER SECURITY PRACTICES JEOPARDIZE FLIGHT SAFETY (May 1998).

14. *See, e.g.*, *infra* Part I.B (providing recent examples of cyber-attacks).

15. As distinct from cyber-crime. *See infra* Part I.B.

16. In Part IV of this Article, we explore methods by which the U.S. government and other governments can adopt the proposed definition or a similar, uniform definition.

17. CLARKE & KNAKE, *supra* note 7, at 6; *see, e.g.*, *More Than Firewalls: Three Challenges to American Cyber Security*, ASYMMETRIC THREAT (Aug. 2011), http://asymmetricthreat.net/docs/snapshot2011_08.pdf (citing Clarke's definition); *Understanding Cyber Warfare*, *supra* note 7.

18. Tom Gjelten, *Extending the Law of War to Cyberspace*, NAT'L PUB. RADIO (Sept. 22, 2010), http://www.npr.org/templates/story/story.php?storyId=130023318 (last visited Apr. 18, 2012).

19. *See infra* Part I.A.3 for a discussion of the importance and mechanics of distinguishing between the concepts of cyber-attack and cyber-crime.

20. *See infra* Part II.A for a detailed exploration of *jus ad bellum* as it applies to cyber-attacks.

in one respect: it limits the definition to attacks perpetrated by *nation-states*, thereby excluding entirely plausible scenarios in which attacks are carried out by non–state actors.

Technical experts have proposed more limited definitions. For example, in his famous and prescient 1995 work on information warfare, Martin Libicki limits cyber-warfare to semantic attacks—digital assaults that cause systems to seem to operate normally, when in fact they generate "answers at variance with reality."[21] This approach excludes the broad range of potential threats to a country's national security that target cyber-infrastructure but do not meet the requirements of a semantic attack. These threats have the same capacity to inflict harm on computer systems or networks, and thus any definition of cyber-attack that excludes them is necessarily incomplete.

There have been two particularly prominent government-led efforts to understand the scope of the threat posed by cyber-attacks, one by the U.S. government and the other by the Shanghai Cooperation Organization—a security cooperation group composed of China, Russia, and most of the former Soviet Central Asian republics, as well as observers including Iran, India, and Pakistan. Perhaps not surprisingly, they have arrived at very different understandings of the problem.

Shortly after establishing the United States Cyber Command, the Joint Chiefs of Staff published a lexicon in 2011 for military use in cyber-operations, which included the first official military definition of cyber-attack. It defines a cyber-attack as:

> A hostile act using computer or related networks or systems, and intended to disrupt and/or destroy an adversary's critical cyber systems, assets, or functions. The intended effects of cyber attack are not necessarily limited to the targeted computer systems or data themselves—for instance, attacks on computer systems which are intended to degrade or destroy infrastructure or C2 capability. A cyber attack may use intermediate delivery vehicles including peripheral devices, electronic transmitters, embedded code, or human operators. The activation or effect of a cyber attack may be widely separated temporally and geographically from the delivery.[22]

A key feature of this approach is that it limits "cyber-attacks" to those hostile acts that are intended to harm critical cyber systems—thus restricting the definition based on the objective of the attack.[23]

---

21. MARTIN C. LIBICKI, WHAT IS INFORMATION WARFARE? 77 (1995).

22. Gen. James E. Cartwright, Memorandum for Chiefs of the Military Servs., Commanders of the Combatant Commands, Dirs. of the Joint Staff Directories on Joint Terminology for Cyberspace Operations 5 (Nov. 2011).

23. Alternative views of cyber-attack and cyber-warfare preceded this announcement in policy circles in the United States. In 2001, the Congressional Research Service defined cyber-warfare as "warfare waged in cyberspace. It can include *defending* information and computer networks, *deterring* information attacks, as well as *denying* an adversary's ability to do the same. It can include *offensive*

The Shanghai Cooperation Organization, by contrast, has adopted a more expansive means-based approach to cyber-attacks. The Organization has "express[ed] concern about the threats posed by possible use of [new information and communication] technologies and means for the purposes [sic] incompatible with ensuring international security and stability in both civil and military spheres."[24] It defines an "information war" as "mass psychologic[al] brainwashing to destabilize society and state, as well as to force the state to take decisions in the interest of an opposing party."[25] Moreover, it identifies the dissemination of information harmful to "social and political, social and economic systems, as well as spiritual, moral and cultural spheres of other states" as one of the main threats to information security.[26]

Hence the Shanghai Cooperation Organization appears to have adopted an expansive vision of cyber-attacks that includes the use of cyber-technology to undermine political stability. Commentators fear that this definition represents an effort to justify censorship of political speech on the Internet.[27] This concern is particularly salient in light of recent government efforts to suppress political organizing using new media in Iran, Egypt, and elsewhere.[28]As the Internet is increasingly utilized as a forum for exchange of ideas and political organization, such suppression threatens human rights.

---

information operations mounted against an adversary, or even *dominating* information on the battlefield." STEVEN A. HILDRETH, CONG. RESEARCH SERV., CRS REPORT FOR CONGRESS: CYBERWARFARE 16 (2001). In 2009, the U.S. National Research Council, an independent organization in Washington, D.C., defined cyber-attack as "deliberate actions to alter, disrupt, deceive, degrade, or destroy computer systems or networks or the information and/or programs resident in or transiting these systems or networks." NAT'L RESEARCH COUNCIL, TECHNOLOGY, POLICY, LAW, AND ETHICS REGARDING U.S. ACQUISITION AND USE OF CYBERATTACK CAPABILITIES 1 (William A. Owens et al. eds., 2009) [hereinafter NRC REPORT].

24. AGREEMENT BETWEEN THE GOVERNMENTS OF THE MEMBER STATES OF THE SHANGHAI COOPERATION ORGANIZATION ON COOPERATION IN THE FIELD OF INTERNATIONAL INFORMATION SECURITY, 61ST PLENARY MEETING (Dec. 2, 2008) [hereinafter SHANGHAI COOPERATION AGREEMENT]. The distinction between this interpretation and that of the United States is understandable in light of Matthew Waxman's analysis of strategic differences in the cyber-attack context. As Waxman notes, "major state actors in this area are likely to have different views on legal line drawing because they perceive a different set of strategic risks and opportunities." Matthew C. Waxman, *Cyber-Attacks and the Use of Force: Back to the Future of Article 2(4)*, 36 YALE J. INT'L L. 421, 458–59 (2011).

25. SHANGHAI COOPERATION AGREEMENT, Annex I, at 209.

26. *Id.* at 203.

27. *See, e.g.*, Tom Gjelten, *Seeing the Internet as an 'Information Weapon,'* NAT'L PUB. RADIO (Sept. 23, 2010), http://www.npr.org/templates/story/story.php?storyId=130052701; *see also infra* Part I.A.2.e.

28. *See, e.g.*, Saeed Kamali Dehghan, *Iran Clamps Down on Internet Use*, GUARDIAN (Jan. 5, 2012), http://www.guardian.co.uk/world/2012/jan/05/iran-clamps-down-internet-use; Matt Richtel, *Egypt Halts Most Internet and Cell Service, and Scale of Shutdown Surprises Experts*, N.Y. TIMES, Jan. 29, 2011, at A13; Sal Gentile, *Gadhafi Regime "Turns Off the Tap" on Libya's Internet*, *Live Blog: Libya Revolts*, PBS (Mar. 4, 2011, 6:46 PM), http://www.pbs.org/wnet/need-to-know/the-daily-need/libya-revolts-a-live-blog/7679/.

The distance between these two government-led understandings of cyber-attacks demonstrates the importance of specifying a clear definition of the problem to be faced. The next Subsection takes on this task.

2. Recommended Definition

In this Article, we adopt a narrow definition of cyber-attack, one meant to focus attention on the unique threat posed by cyber-technologies:

> A cyber-attack consists of any action taken to undermine the functions of a computer network for a political or national security purpose.

This Subsection discusses each aspect of this definition to explain the reasoning behind the language and to clarify which activities it encompasses.

a. "A cyber-attack . . ."

Implicit in this term is the requirement that the conduct must be active: either offense or active defense.[29] Active defense includes "electronic countermeasures designed to strike attacking computer systems and shut down cyber-attacks midstream."[30] Governments are likely to employ both active and passive defenses—and the two are often designed to work in tandem[31]—but the passive defense cannot on its own amount to a cyber-attack.[32]

b. ". . . consists of any action taken . . ."

A cyber-attack may be carried out by means of *any* action—hacking, bombing, cutting, infecting, and so forth—but to be a cyber-attack it must aim to undermine or disrupt the function of a computer network. In this respect, this Article adopts the U.S. objective-based approach rather than the means-based approach of the Shanghai Cooperation Organization.

Warfare may be classified on the basis of the means of attack. For example, warfare may be classified as kinetic (conventional, physical) warfare, biological warfare, chemical warfare, nuclear warfare, intelligence-based warfare, network-based warfare,[33] or guerilla warfare. Warfare may also be

29. Measures of passive defense against cyber-attacks, such as virus scanning software or firewalls, are outside the scope of this definition.

30. JEFFREY CARR, INSIDE CYBER WARFARE 46 (2010).

31. Active defense may be triggered by passive activities. For example, a routine virus scan that identifies a virus and then eliminates it switches from passive (scanning) to active (elimination).

32. The U.S. government currently utilizes both active and passive defenses. *See* U.S. DEP'T OF DEF., DEPARTMENT OF DEFENSE STRATEGY FOR OPERATING IN CYBERSPACE (2011) [hereinafter DOD STRATEGY].

33. This is distinct from "network warfare," which is defined as "the employment of Computer Network Operations (CNO) with the intent of denying adversaries the effective use of their computers, information systems, and networks, while ensuring the effective use of our own computers, information systems, and networks." NRC REPORT, *supra* note 23, at 165. Network-based warfare is any type of warfare that utilizes networks. Note a similar distinction between intelligence-based warfare (which describes the means) and information warfare (which describes the objective).

defined by its objective. "Objective" here means the direct target, rather than the long-range purpose, of the action. Examples include information warfare, psychological warfare, command and control warfare,[34] electronic warfare, and economic warfare.

Because we define cyber-attack according to its objective ("to undermine the functions of a computer network for a political or national security purpose"), any means may be used to accomplish a cyber-attack. Defining cyber-attack by objective rather than means is superior for three reasons.

First, and most important, this type of definition is simply more intuitive. Using a computer network in Nevada to operate a predator drone for a kinetic attack in Pakistan is not a cyber-attack; rather, it is technologically advanced conventional warfare. Using a regular explosive to sever the undersea network cables that carry the information packets between continents, on the other hand, is a cyber-attack.[35] This view is consistent with that offered by the U.S. Department of Defense, which has identified kinetic attack as a strategy in cyber-offensive operations.[36]

Second, the objective-based approach is logical. Warfare traditionally functions in four domains—land, air, sea, and space—each of which is addressed by one of the full-time armed services.[37] With the rise of cyber-warfare, strategists have identified a fifth domain: cyberspace.[38] In response, the United States has created the U.S. Cyber Command, a subdivision of the joint services Strategic Command.[39] Although the Cyber Command is not a

---

34. "Command and control warfare" includes any attack meant to interfere with the enemy's capacity to command and control its troops. *See* GEORGE J. STEIN, INFORMATION ATTACK: INFORMATION WARFARE IN 2025, at 2 (1996), *available at* http://csat.au.af.mil/2025/volume3/vol3ch03.pdf. The Department of Defense defines command and control as

> [t]he exercise of authority and direction by a properly designated commander over assigned and attached forces in the accomplishment of the mission. Command and control functions are performed through an arrangement of personnel, equipment, communications, facilities, and procedures employed by a commander in planning, directing, coordinating, and controlling forces and operations in the accomplishment of the mission.

Joint Publication 1-02, Department of Defense Dictionary of Military and Associated Terms, Joint Chiefs of Staff (Nov. 2010), *available at* http://www.dtic.mil/ doctrine/new_pubs/jp1_02.pdf.

35. *See* Antolin-Jenkins, *supra* note 11, at 138 ("[K]inetic weapons are certainly part of the cyberwar arsenal.").

36. CHAIRMAN OF THE JOINT CHIEFS OF STAFF, U.S. DEP'T OF DEFENSE, NATIONAL MILITARY STRATEGY FOR CYBERSPACE OPERATIONS 15 (2006). A National Research Council report on "cyber offensive operations" excluded kinetic attacks on computer networks for the purposes of the report, but acknowledged that such attacks were realistic forms of cyber-attack. NRC REPORT, *supra* note 23, at 12–19.

37. Space is difficult to assign to the Army, Navy, or Air Force, but its proper classification is outside the scope of this paper.

38. *See* DOD STRATEGY, *supra* note 32, at 5; *War in the Fifth Domain*, ECONOMIST (July 1, 2010), http://www.economist.com/node/16478792. The Joint Chiefs of Staff identify cyberspace as one of the "global commons," along with international waters, air space, and space. JOINT CHIEFS OF STAFF, NATIONAL MILITARY STRATEGY OF THE UNITED STATES OF AMERICA 5 (2004).

39. William H. McMichael, *DoD Cyber Command Is Officially Online*, ARMYTIMES (May 22, 2010, 9:20 AM), http://www.armytimes.com/news/2010/05/military_cyber_command_052110/;

unique service, it coordinates the functional operations of the Army, Navy (and Marines), and Air Force. The armed services are traditionally organized by domain rather than by platform. The Army's function is to control land, not to drive tanks and fire land-based artillery; the Navy's function is to control the seas, not to operate boats and ships; and the Air Force's function is to control the skies, not to fly planes and drop bombs. Each service has access to whatever tools and weapons it deems necessary to control its domain: planes, boats, missiles, artillery, computer networks, and so forth. By the same logic, Cyber Command's mission is not to utilize computer networks for any objective, but to defend the ability to operate in cyberspace by any means.[40]

Third, an objective-based approach avoids unnecessarily limiting Internet speech, thereby avoiding the serious risks posed by a means-based definition. By encompassing any activity that uses cyber-technology and jeopardizes stability, a means-based understanding of cyber-warfare can be used to constrain the expression of free speech and political dissent online.[41] The Shanghai Cooperation Organization's definition may have been designed to be means-based in part for this reason.[42]

c. ". . . to undermine the function . . ."

The objective of a cyber-attack must be to undermine the *function* of a computer network. A computer network may be compromised in many different ways. Syntactic attacks disrupt a computer's operating system, causing the network to malfunction.[43] Examples include "worms, viruses, [and] Trojan horses."[44] The incident in Burma, discussed in the opening to this Article, constituted a syntactic attack. In contrast, semantic attacks preserve the operating system but compromise the accuracy of the information it processes and to which it reacts.[45] As a result, "[a] system under semantic attack operates and will be perceived as operating correctly, . . . but it will generate answers at variance with reality."[46] The Stuxnet attack described above was, in part, a semantic attack because the nuclear plant appeared to be operating normally even as it was malfunctioning.[47]

---

see Thom Shanker, *Cyberwar Chief Calls for Secure Computer Network*, N.Y. TIMES, Sept. 24, 2010, at A1.

40. *See* DOD STRATEGY, *supra* note 32, at 5 ("[T]reating cyberspace as a domain is a critical organizing concept for DoD's national security missions. This allows DoD to organize, train, and equip for cyberspace as we do in air, land, maritime, and space to support national security interests.").

41. *See* Gjelten, *supra* note 27.

42. *See id.*; SHANGHAI COOPERATION AGREEMENT, *supra* note 24.

43. Antolin-Jenkins, *supra* note 11, at 139.

44. *Id.*

45. *Id.* at 140.

46. LIBICKI, *supra* note 21, at 77.

47. Cyber-attacks need not be limited to syntactic or semantic attacks. The U.S. cyber-operation in Iraq discussed below, for example, was neither syntactic nor semantic. Nevertheless, it

By contrast, neither cyber-espionage nor cyber-exploitation constitutes a cyber-attack because these concepts do not involve altering computer networks in a way that affects their current or future ability to function.[48] For example, in 2003, a security breach created numerous leaks of sensitive information from U.S. Department of Defense computers, which occurred over several months.[49] The Department has acknowledged that the majority of such incidents—collectively referred to as "Titan Rain"—were orchestrated by China as a method of cyber-espionage.[50] Another recent example of cyber-espionage occurred when hackers operating from China copied data from Google and other major Internet technology companies in 2010. The alleged purpose of the prolonged security breach ranged from theft of intellectual property to unlawful surveillance of human rights activists.[51] Subsequent developments imply that at least one purpose of the attack—dubbed "Operation Aurora"—was to monitor U.S. government officials' emails.[52] More recently, the Department of Defense admitted that it suffered one of its worst cyber-espionage leaks in March 2011, when foreign hackers gained access to over 24,000 Pentagon files.[53] Meanwhile, the extent to which the United States is conducting similar activities is unknown.[54]

---

constitutes a cyber-attack under this Article's definition, as it did "undermine the function" of the secure email system by causing it to send an email from an unauthorized user.

48. This Article adopts the following definition of cyber-espionage: "[T]he science of covertly capturing e-mail traffic, text messages, other electronic communications, and corporate data for the purpose of gathering national-security or commercial intelligence." Seymour M. Hersh, *The Online Threat: Should We Be Worried About a Cyber War?* NEW YORKER (Nov. 1, 2010), http://www.newyorker.com/reporting/2010/11/01/101101fa_fact_hersh?. The former director of the Central Intelligence Agency (CIA) emphasizes that cyber-espionage does not fall under the umbrella of cyber-warfare, likely because the U.S. government—like many other governments—routinely engages in espionage over communications networks. Gjelten, *supra* note 18. Notably, the National Research Council draws a similar line. It distinguishes what it calls cyber-exploitation—which includes actions that merely gather information from the cyber-domain and is therefore related to, if perhaps somewhat broader than, cyber-espionage—from cyber-attack because "[t]he [law of armed conflict] presumes that a clear distinction can be drawn between the use of force and espionage, where espionage is avowedly not a use of force." NRC REPORT, *supra* note 23, at 22, § 1.6.

49. CLAY WILSON, CONG. RESEARCH SERV., BOTNETS, CYBERCRIME, AND CYBERTERRORISM: VULNERABILITIES AND POLICY ISSUES FOR CONGRESS 14 (2008).

50. *Id.*

51. *A New Approach to China*, OFFICIAL GOOGLE BLOG (Jan. 12, 2010, 3:00 PM), http://googleblog.blogspot.com/2010/01/new-approach-to-china.html; *see also* James Glanz & John Markoff, *State's Secrets Day 7; Vast Hacking by a China Fearful of the Web*, N.Y. TIMES, Dec. 4, 2010, at A1.

52. *See, e.g.*, Amir Efrati & Siobhan Gorman, *Google Mail Hack Is Blamed on China*, WALL ST. J., June 2, 2011, at A1; Wyatt Andrews, *China Google Hacker's Goal: Spying on U.S. Govt*, CBS NEWS (June 2, 2011), http://m.cbsnews.com/fullstory.rbml?catid=20068474&feed_id=0&videofeed=36 (last visited Apr. 19, 2012).

53. Thom Shanker & Elisabeth Bumiller, *After Suffering Damaging Cyberattack, the Pentagon Takes Defensive Action*, N.Y. TIMES, July 15, 2011, at A6.

54. *See* Jack Goldsmith, *What Is the Government's Strategy for the Cyber-Exploitation Threat?*, LAWFARE BLOG (Aug. 10, 2011, 10:58 PM), http://www.lawfareblog.com/2011/08/what-is-the-government%E2%80%99s-strategy-for-the-cyber-exploitation-threat/ (last visited Apr. 19, 2012).

Although all of these incidents of cyber-espionage compromised the security of a computer network for the purpose of carrying out a military objective,[55] they did not "undermine the function" of a computer system and thus were not cyber-attacks as defined here. To "undermine the function" of a computer system, an actor must *do more than passively observe a computer network or copy data*, even if that observation is clandestine. The actor must affect the operation of the system either by damaging the operating system or by adding false, misleading, or unwelcome information. Such activities may be criminal—as acts of corporate or political cyber-espionage—but they are not cyber-*attacks*. In this respect, our definition reflects a common distinction between espionage and attacks in more traditional settings.

d. ". . . of a computer network . . ."

A cyber-attack must target a computer network, where a computer network is defined as a system of computers and devices connected by communications channels. Frequently, this connection exists over the Internet, but there are also numerous closed networks, such as the secure networks employed by agencies of the U.S. government.

It is important to bear in mind that computers are now everywhere. The concept of a computer encompasses more than a simple desktop or laptop; it also includes the devices that control elevators and traffic lights, regulate pressure on water mains, and are ubiquitous in appliances such as cell phones, televisions, and even washing machines.[56] The potential for widespread damage from a cyber-attack grows in tandem with the spread of computers to nearly every facet of human activity.

e. ". . . for a political or national security purpose."

A political or national security purpose distinguishes cyber-attack from simple cyber-crime. Any aggressive action taken by a state actor in the cyber-domain necessarily implicates national security and is therefore a cyber-attack (where the action satisfies all the other elements of the definition), whether or not it rises to the level of cyber-warfare. A cyber-crime committed by a non–state actor for a political or national security purpose is a cyber-attack. On the other hand, a cyber-crime that is not carried out for a political or national security purpose, such as most instances of Internet fraud, identity theft, and intellectual property piracy, does not fit this final element of a "cyber-attack" and is therefore mere cyber-crime.

55. *See* Michael Joseph Gross, *Enter the Cyber-Dragon*, VANITY FAIR, Sept. 2011, *available at* http://www.vanityfair.com/culture/features/2011/09/chinese-hacking-201109 (detailing these and other successful hacks of public and private systems).

56. CLARKE & KNAKE, *supra* note 7, at 70–74.

There are at least two important reasons for excluding nonpolitical cyber-crimes (that is, cyber-crimes not carried out for a political or national security purpose) from the definition of cyber-attack. First, such activities, while troubling, do not raise the same legal questions as activities that might breach public international law. The actions of the Kremlin Kids, private hackers who allegedly shut down the Georgian Internet during Russia's invasion of South Ossetia,[57] invoke legal doctrines surrounding state responsibility and terrorism[58] in a way that the actions of Onel de Guzman, a student who was suspected of infecting tens of millions of computers in 2000 with the destructive but undirected "love bug virus,"[59] do not. Second, a cleaner delineation between cyber-attacks that present threats to national security and purely private cyber-crime will clarify ownership of cyber-security needs among various government departments.

A political or national security purpose also denotes the public nature of the cyber-attacks without limiting the definition to state actors. This is important because, due to their low cost and the relative invulnerability of non–state actors to in-kind retribution, cyber-attacks are a particularly attractive weapon for terrorists and other non–state actors.[60] Because non–state actors may execute or may be the victim of cyber-attacks, the purpose, rather than the actor, must distinguish a cyber-attack from a simple cyber-crime. This definition does not distinguish between state and non–state actors. Rather, it identifies a legal framework that is compatible with existing law of war and international law distinctions between non–state and state actors.

Although this distinction is notable, it is not without risks. There is always a danger that cyber-regulations may be applied against individuals using technology for legitimate political dissent, which necessarily has a political purpose. While the First Amendment protects dissent in the United States, the use of cyberspace regulations to suppress dissent is a serious possibility in

57. *See* Noah Shachtman, *Kremlin Kids: We Launched the Estonian Cyber War*, WIRED (Mar. 11, 2009, 12:45 PM), http://www.wired.com/dangerroom/2009/03/pro-kremlin-gro/ (last visited Apr. 19, 2012); *infra* Part I.B.1.

58. The line drawn between simple cyber-crime and cyber-attack by private individuals is analogous to the line drawn between violent crime and terrorism. *See* 18 U.S.C. § 2331(1)(B) (2006) (defining international terrorism according to its apparent political intentions); BLACK'S LAW DICTIONARY 1611 (9th ed. 2009) (defining terrorism as using violence "as a means of affecting political conduct").

59. Mark Landler, *A Filipino Linked to 'Love Bug' Talks About His License to Hack*, N.Y. TIMES, Oct. 21, 2000, at C1.

60. *See* NRC REPORT, *supra* note 23, at 20, §1.4 (on low cost); *id.* at 41 (on limited applicability of deterrence by threat of in-kind response); DOD STRATEGY, *supra* note 32, at 3 (discussing the power of small groups to cause significant harm due to the low barriers to entry for cyber-activity); Shanker & Bumiller, *supra* note 53 (noting that while most major efforts to penetrate military computer networks are still orchestrated by large rival nations, the technical expertise is certain to migrate to rogue states and non–state actors).

countries that do not have the same liberal democratic traditions.[61] Internet regulations in China are a troubling testament to this fact.[62] As a foreign policy matter, the United States must ensure that any proposed domestic legislation (which may serve as a model for other countries) or international regime (which may be susceptible to multiple readings) clearly maintains online space for legitimate dissent while strengthening the legal tools to combat and punish cyber-attacks.[63] This definition seeks to keep legitimate dissent out of the category of cyber-attack by specifying that a cyber-attack's objective must be to undermine the function of a computer network. It would not include, for example, computer-based efforts to organize political protests.

The definition offered here adopts the objective-based approach taken by the U.S. government, but it adds a "purpose" requirement that enables policy-makers to distinguish between mere cyber-crime and cyber-attacks. Such a distinction is crucial to domestic and international efforts to implement cyber-security, because it more effectively tailors the legal approach to the threat posed and focuses resources on true national security threats.

3. Cyber-Attack, Cyber-Crime, and Cyber-Warfare Compared

We summarize our definition of "cyber-attack" and the distinctions between "cyber-attack," "cyber-crime," and "cyber-warfare" in Table 1 and Figure 1.

61. *See, e.g.*, Gjelten, *supra* note 27 (on Chinese and Russian efforts to control communication on the Internet).

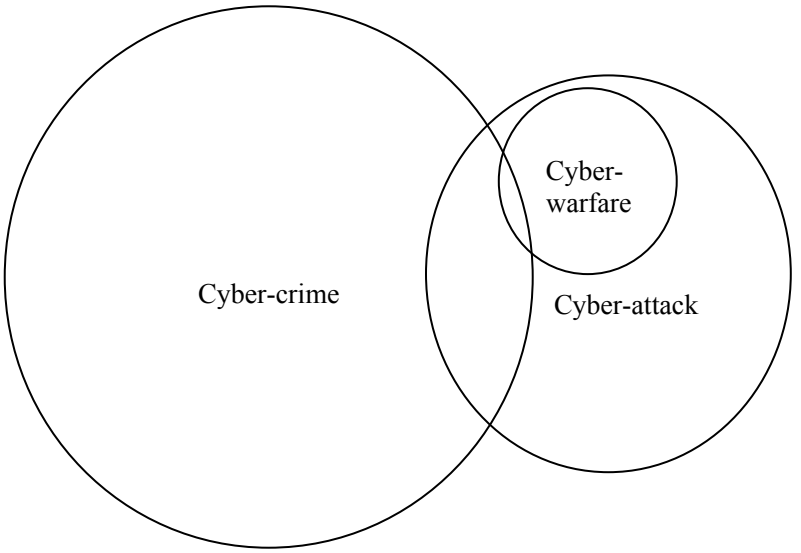
62. China has also been embroiled in cyber-conflict with private entities as well—namely, Google and Yahoo. Since the early 2000s, the U.S.-based companies have been criticized for their cooperation with the Chinese government, both in policing internal dissidents and in censoring external information of a political nature. *See Yahoo 'Helped Jail China Writer*,*'* BBC NEWS (Sept. 7, 2005, 8:18 AM), http://news.bbc.co.uk/2/hi/4221538.stm; *Google Censors Itself for China*, BBC NEWS (Jan. 25, 2006, 8:45 AM), http://news.bbc.co.uk/2/hi/technology/4645596.stm. Pressure from the Chinese government for such cooperation comes in response to activity it labels as "cyber-attacks"—the dissemination of information that undermines civil and military stability. *See* SHANGHAI COOPERATION AGREEMENT, *supra* note 24.

63. The White House's recent strategy paper on cyberspace addresses the danger that efforts to reduce cyber-attacks could stifle free speech. It notes that "the ability to seek, receive, and impart information and ideas through any medium and regardless of frontiers has never been more relevant" and urges that "exceptions to free speech in cyberspace must also be narrowly tailored." OFFICE OF THE PRESIDENT, INTERNATIONAL STRATEGY FOR CYBERSPACE: PROSPERITY, SECURITY, AND OPENNESS IN A NETWORKED WORLD 5 (2011) [hereinafter WHITE HOUSE CYBERSPACE STRATEGY]. Protecting fundamental freedoms and privacy is one of the White House's seven high-level policy priorities for cyberspace, *id.* at 23–24, and one of the three law enforcement policy priorities is to "[f]ocus cybercrime laws on combating illegal activities, not restricting access to the internet," *id.* at 20.

TABLE 1: Essential characteristics of different cyber-actions

	Type of cyber-action		
	Cyber-attack	*Cyber-crime*	*Cyber-warfare*
Involves only non–state actors		√	
Must be violation of criminal law, committed by means of a computer system		√	
Objective must be to undermine the function of a computer network	√		√
Must have a political or national security purpose	√		√
Effects must be equivalent to an "armed attack," or activity must occur in the context of armed conflict			√

FIGURE 1: Relationship between cyber-actions



In order to understand cyber-attack, it is important to appreciate the distinctions between cyber-attack and cyber-crime. Cyber-crime is a broad concept analytically distinct from cyber-attack. While, as with the concept of cyber-attack, there is no universally recognized definition of cyber-crime,[64]

64. *See, e.g.*, Sarah Gordon & Richard Ford, *On the Definition and Classification of Cybercrime*, 2 J. COMPUTER VIROLOGY 13, 13 (2006) ("Despite the fact that the word 'Cybercrime' has entered into common usage, many people would find it hard to define the term precisely."); Sylvia

there are aspects of cyber-crime that are broadly recognized. In particular, cyber-crime is generally understood as the use of a computer-based means to commit an illegal act. One typical definition describes cyber-crime as "any crime that is facilitated or committed using a computer, network, or hardware device."[65] Cyber-crime, unlike the definition of cyber-attack proposed in this Article, is thus often defined by its means—that is, a computer system or network. As such, cyber-crime encompasses a very broad range of illicit activity. Among the priorities of the Department of Justice and FBI units addressing cyber-crime are fraudulent practices on the Internet, online piracy, storage and sharing of child pornography on a computer, and computer intrusions.[66] Unlike cyber-attacks, cyber-crimes need not undermine the target computer network (though in some cases they may do so), and most do not have a political or national security purpose. Finally, like all crimes, but unlike cyber-attacks, cyber-crimes are generally understood to be committed by individuals, not states.[67] While the distinction between cyber-crime and cyber-attack is important, we acknowledge that it often will not be readily apparent at the moment of the cyber-event whether it is one or the other (or both)—in part because the identity and purpose of the actor may not be apparent. Such uncertainty counsels in favor of an immediate response that would be appropriate to either cyber-crime or a cyber-attack.

Most cyber-crimes do not also constitute cyber-attack or cyber-warfare, as depicted in Figure 1. An act is only a cyber-crime when a non–state actor commits an act that is criminalized under domestic or international law.

Mercado Kierkegaard, *International Cybercrime Convention*, IGI GLOBAL, http://www.igi-global.com/viewtitlesample.aspx?id=7486 (last visited Apr. 6, 2012) ("[T]here is still no accepted definition of what really constitutes cybercrime."); *see also* DEBRA LITTLEJOHN SHINDER, SCENE OF THE CYBERCRIME: COMPUTER FORENSICS HANDBOOK 16 (Ed Tittel ed., 2002) ("[T]he definition of computer crime under state law differs, depending on the state.").

65. Gordon & Ford, *supra* note 64, at 14. In addition, some proposed definitions are broad enough to include not only all crimes committed by means of a computer, but also any crime in any way involving a computer as a means or a target. *See, e.g.*, SHINDER, *supra* note 64, at 17 (referring to the Tenth United Nations Congress on the Prevention of Crime and Treatment of Offenders' broad definition of "computer-related crime," as compared to its narrower, means-based definition of "computer crime").

66. *See generally* COMPUTER CRIME AND INTELLECTUAL PROPERTY SECTION, CRIMINAL DIVISION, U.S. DEP'T OF JUSTICE, PROSECUTING COMPUTER CRIMES (2d ed. 2010); *Cyber Crime*, FBI, http://www.fbi.gov/about-us/investigate/cyber (last visited Apr. 21, 2012). The Council of Europe Convention on Cybercrime, similarly, covers a broad range of criminal activity committed by means of a computer, including "action directed against the confidentiality, integrity and availability of computer systems, networks and computer data as well as the misuse of such systems, networks and data." Convention on Cybercrime, Council of Europe, E.T.S. No. 185, pmbl., Nov. 23, 2001 (entered into force July 1, 2004), *available at* http://conventions.coe.int/Treaty/EN/Treaties/Html/185.htm [hereinafter Cybercrime Convention].

67. Therefore, under our definition, while public officials may commit cyber-crimes while acting outside the scope of their authority, the actions of states, even if unlawful, are not considered to be crimes as such.

Consider the following three scenarios, each of which includes a cyber-crime that is not a cyber-attack:

First, a non–state actor commits an illegal act for a political or national security purpose by means of a computer network but does *not* undermine that network. For example, an individual might commit a cyber-crime by expressing political dissent over the Internet where that dissent is illegal under domestic law. Similarly, an individual might commit a cyber-crime by hacking into a major bank's records with a national security or political purpose but without undermining the bank's system in the process.

Second, a non–state actor commits an illegal act by means of a computer network—and undermines a computer network—but not for a political or national security purpose. Again consider the bank data hacker, who now manages to undermine the bank's online account system but whose only purpose is economic gain. This, too, would constitute a cyber-crime, but not a cyber-attack or cyber-warfare.

Third, a non–state actor is engaged in illicit activity using a computer or network but does not undermine the function of a computer network and does not operate with a political or national security purpose. A person who transfers child pornography, for example, would commit a cyber-crime but not a cyber-attack, both because the actions do not undermine the function of a computer network and because he or she is not motivated by a political or national security purpose.

As shown in Figure 1, just as some cyber-crimes are neither cyber-attacks nor cyber-warfare, some cyber-attacks are neither cyber-crimes nor cyber-warfare. Two scenarios fall into this cyber-attack-only category. The first includes attacks carried out by a state actor, outside the context of an armed conflict, provided its effects do not rise to the level of an armed attack. An example is the attack by the Chinese government on the Falun Gong website in 2011.[68] Note that such attacks must still satisfy all elements of the cyber-attack definition, including undermining the function of a computer network for a political or national security purpose. As noted above, however, any act by a state actor automatically satisfies the political or national security purpose requirement.

The second cyber-attack-only scenario includes attacks by non–state actors that do not rise to the level of an armed attack and which do not constitute a cyber-crime, either because they have not been criminalized under national or international law or because they do not use computer-based means. Practically speaking, it is unlikely for a private actor to purposefully[69] undermine the function of a computer network without also violating the law,

68. *See* Nakashima & Wan, *supra* note 5.

69. Because a cyber-attack must be "for a political or national security purpose," the only actions falling into this category would be purposeful.

but such gaps in the criminal law are conceptually possible. It is furthermore worth noting that a large majority of cyber-attacks would likely involve computer-based means, though such means are not necessary to cyber-attack under the definition proposed here.

While cyber-activity may constitute only cyber-crime or only cyber-attack, a substantial proportion of cyber-crimes are also cyber-attacks. The overlapping area between cyber-crime and cyber-attack seen in Figure 1 occurs when a non–state actor commits an illegal act by means of a computer network, undermines a computer network, *and* has a political or national security purpose. The consequences of this act would not rise to the level of an armed attack, or the activity would also constitute cyber-warfare. Note also that a state committing this very same act would not fall within this overlap, since only a non–state actor can commit a cyber-crime. Take, for example, a hypothetical group of individuals who hacked into the U.S. State Department's server and shut it down out of disdain for the U.S. government. This instance would fall within the overlap between cyber-crimes and cyber-attacks given that a non–state actor committed the act, for a political or national security purpose, and it undermined a computer network.

Cyber-warfare is distinctive among the three cyber-categories considered here in that cyber-warfare *must* also constitute a cyber-attack. The overlapping area between cyber-attack and cyber-warfare (but not cyber-crimes) in Figure 1 includes two types of attacks. The first type includes attacks carried out by any actor in the context of an armed conflict, provided those actions could not be considered cyber-crimes, either because they do not constitute war crimes, or do not employ computer-based means, or both. The second type includes attacks carried out by a state actor, which produce effects equivalent to those of a conventional armed attack. Note that this use of force may be either lawful or unlawful; because the actor is a state actor, even unlawful actions do not necessarily constitute "cyber-crime."

Cyber-warfare can also constitute both cyber-attack and cyber-crime. The area of intersection between all three circles in Figure 1 includes two types of attacks carried out by a non–state actor. First, it includes attacks in the context of an existing armed conflict that undermine the function of a computer network for a political or national security purpose, violate the criminal law (for example, war crimes), and were committed by means of a computer system or network. Second, it includes attacks that produce effects equivalent to those of a conventional armed attack, undermine the function of a computer network for a political or national security purpose, and are violations of the criminal law committed by means of a computer system or network.

As summarized in Table 1 and Figure 1, then, a cyber-attack may be carried out by state or non–state actors, must involve active conduct, must aim to undermine the function of a computer network, and must have a political or national security purpose. Some cyber-attacks are also cyber-crimes, but not all

cyber-crimes are cyber-attacks. Cyber-warfare, on the other hand, always meets the conditions of a cyber-attack. But not all cyber-attacks are cyber-warfare. Only cyber-attacks with effects equivalent to those of a conventional "armed attack," or occurring within the context of armed conflict, rise to the level of cyber-warfare. We say more about when this condition is met in Part II below.

B. Recent Cyber-Attacks

There are a variety of activities that fall within this Article's definition of cyber-attacks. The following examples of recent cyber-incidents—though far from exhaustive—demonstrate the variety and scope of recent cyber-attacks. They also introduce the wide-ranging challenges to regulating such attacks.

1. Distributed Denial of Service Attacks

Distributed Denial of Service ("DDOS") attacks have been the most prevalent form of cyber-attack in recent years. In these attacks, coordinated botnets—collections of thousands of "zombie" computers hijacked by insidious viruses—overwhelm servers by systematically visiting designated websites. The attack in Burma, described above, was a DDOS attack, as was the attack on a Falun Gong website inadvertently aired on China Central Television. There are several other recent examples of such attacks—a few of which we describe here to provide a sense of the varied ways in which such attacks may be carried out.

After controversially moving a Soviet-era war memorial in April 2007, the densely wired[70] Republic of Estonia suffered a DDOS attack. Such attacks often cause mere inconvenience, but this one nearly had life threatening consequences—the emergency line to call for an ambulance or a fire truck was out of service for an hour.[71] Allegedly executed by networks of hackers,[72] authorities never officially attributed the attack to a state, but some suspect Russia's involvement due to the sophistication and scale of the attack.[73]

A similar fate befell Georgia in the summer of 2008, when the country found itself unable to communicate with the outside world over the Internet as Russian forces invaded South Ossetia.[74] Despite early speculation that the

70. Estonia has one of the highest network saturation rates in the world. CLARKE & KNAKE, *supra* note 7, at 13.

71. *Newly Nasty: Defences Against Cyberwarfare Are Still Rudimentary. That's Scary*, ECONOMIST (May 24, 2007), http://www.economist.com/node/9228757?story_id=9228757 (last visited Apr. 19, 2012).

72. Specifically, a youth movement (funded by the Russian government) later claimed responsibility for the attack. Shachtman, *supra* note 57.

73. Jeffrey T. G. Kelsey, Note, *Hacking into International Humanitarian Law: The Principles of Distinction and Neutrality in the Age of Cyber Warfare*, 106 MICH. L. REV. 1427, 1429 (2008).

74. *The Threat from the Internet: Cyberwar: It Is Time for Countries to Start Talking About Arms Control on the Internet*, ECONOMIST (July 1, 2010), http://www.economist.com/node/16481504 (last visited Apr. 19, 2012).

Russian government had planned the incident, it now appears likely that the government simply stood by as private hackers openly orchestrated the attack.[75]

Russians are certainly not the only source of DDOS attacks. In July 2009, a number of government and commercial websites in the United States and South Korea were shut down by a DDOS attack. Although South Korea quickly blamed North Korea,[76] the United States was more circumspect.[77] There remain some questions about where the attack originated. This serves to illustrate a common problem for cyber-attacks in general and DDOS attacks in particular: by enlisting unsuspecting computers from around the world, botnets spin a web of anonymity around the attacker or attackers, making accurate attribution uniquely difficult.

2. Planting Inaccurate Information

Another form of cyber-attack is a semantic attack, in which the attacker surreptitiously inputs inaccurate information in a computer system. More sophisticated than the DDOS attack, a semantic attack causes the computer system to appear to operate normally, even as it fails.[78]

In 1999, for example, the United States developed a plan to feed false target data into the Serbian air defense command network, inhibiting Serbia's ability to target NATO aircraft.[79] This attack would have exploited the increasing reliance on computer networks that characterizes modern warfare. In the end, NATO forces abandoned the plan due to legal concerns about collateral damage.[80]

The Israeli Air Force employed a similar strategy on September 6, 2007 during its air strike against a nuclear facility in Syria. Israeli planes arrived undetected at their targets because of an earlier cyber-attack that compromised the Syrian air-defense system. The exact method of attack is unknown, but Israel apparently fed false messages to the radars, causing them to show clear skies on the night of the strike.[81]

Because these cyber-attacks frequently accompany and facilitate conventional attacks, attribution is less problematic. The difficulty here is in

75. Brian Krebs, *Report: Russian Hacker Forums Fueled Georgia Cyber Attacks*, WASH. POST SECURITY FIX BLOG (Oct. 16, 2008, 3:15 PM), http://voices.washingtonpost.com/securityfix/2008/10/report_russian_hacker_forums_f.html.

76. Malcolm Moore, *North Korea Blamed for Cyber Attack on South Korea*, TELEGRAPH (July 8, 2009), http://www.telegraph.co.uk/news/worldnews/ asia/southkorea/5778176/North-Korea-blamed-for-cyber-attack-on-South-Korea.html.

77. Officials anonymously leaked qualified reports of U.S. suspicions that the attack emerged in North Korea. *U.S. Eyes N. Korea for 'Massive' Cyber Attacks*, MSNBC.COM (July 9, 2009, 3:31 AM), http://www.msnbc.msn.com/id/31789294/ns/technology_and_science-security.

78. LIBICKI, *supra* note 21, at 77.

79. William M. Arkin, *The Cyber Bomb in Yugoslavia*, WASH. POST (Oct. 25, 1999), http://www.washingtonpost.com/wp-srv/national/dotmil/arkin.htm.

80. Kelsey, *supra* note 73, at 1434–35.

81. CLARKE & KNAKE, *supra* note 7, at 1–9.

identifying when a cyber-attack has occurred, since the disruption remains hidden until its kinetic sequel.

3. Infiltrating a Secure Computer Network

Once an attacker infiltrates a secure computer network she can execute a variety of actions.[82] For example, the Stuxnet attack, in addition to being a semantic attack, targeted the secure computer networks at Iranian nuclear facilities for the purpose of disrupting the function of the nuclear facility.

Such an attack does not always destroy the computer network or the infrastructure it controls. In 2003, shortly before the invasion of Iraq, the United States infiltrated the Iraqi Defense Ministry email system to contact Iraqi officers with instructions for a peaceful surrender. The messages apparently worked: American troops encountered abandoned military equipment arranged in accordance with the email.[83] This cyber-attack was a "Command and Control Attack"—a term that includes any attack meant to interfere with the enemy's capacity to command and control its troops.

These incidents demonstrate that attacks need not arrive over the Internet, but may instead involve infiltrating separate, secure networks. These networks may include not only desktops and laptops, but the ubiquitous and unseen computing systems, such as industrial control systems, that facilitate modern life. Together, these examples also illustrate the growing number of cyber-attacks and the diversity of their forms and scope that make the project of crafting a legal approach to them all the more challenging. The next Part examines when a cyber-attack rises to the level of "cyber-warfare" governed by the law of war—and when and how that law allows states to respond to such attacks.

II.
LAW OF WAR AND "CYBER-WARFARE"

Although the term "cyber-warfare" has become part of common parlance, few have aimed to examine closely the scope of cyber-activity that might be governed by the law of war. In this Part, we aim to fill this gap by examining when a cyber-attack constitutes an armed attack under *jus ad bellum* and thus can be accurately considered "cyber-warfare." We also examine how the laws governing conduct in the course of war—known as *jus in bello*—might apply to cyber-attacks. We do not attempt a detailed application of *jus ad bellum* and *jus in bello* to cyber-attacks, because such inquiries are intensely fact specific. Instead, we lay out the general types of cyber-attacks that would be governed

82. For reasons explained above, cyber-espionage—stealing rather than planting information—is not included in most definitions of cyber-attack. *See supra* text accompanying notes 43–46.

83. CLARKE & KNAKE, *supra* note 7, at 9–10.

by the law of war and note how an attack's cyber-based nature complicates the traditional law of war analysis. We conclude that while the law of war provides useful guidelines for addressing some of the most dangerous forms of cyber-attack, the law of war framework ultimately addresses only a small slice of the full range of cyber-attacks.[84] Cyber-warfare is only a part of a much larger problem.

It is worth noting at the outset that applying the existing law of war framework to cyber-attacks is extraordinarily challenging. The key treaties governing conduct in war, the Geneva Conventions, were last revised in the wake of World War II. Nothing was further from the minds of the drafters of the Geneva Conventions than attacks carried out over a worldwide computer network. One unanticipated challenge is how to address attacks that have little or no direct physical consequences, but that nonetheless cause real harm to national security. Perhaps for this reason, thus far no state has claimed that a cyber-attack constitutes an "armed attack" giving rise to a right of self-defense under Article 51 of the U.N. Charter. Nor has any state argued that cyber-attacks generally constitute a prohibited use of force. The fact that such attacks are increasing in number and scope, however, suggests that there is a growing need for states to reach a consensus as to when a cyber-attack constitutes an armed attack or use of force. In the absence of agreement, the increase in attacks heightens the possibility that states might respond to a cyber-attack with conventional military means.[85] The rise in attacks also creates a more pressing need for a more comprehensive legal framework to regulate activities—such as those causing widespread economic damage—that would not be governed by the law of war.[86]

---

84. Practitioners and scholars are divided on how easily the law of war can be applied to cyber-attacks. The Handbook guiding Navy, Marine, and Coast Guard operations, discussing information operations, states that "[l]egal analysis of intended wartime targets requires traditional law of war analysis." DEP'T OF THE NAVY, COMMANDER'S HANDBOOK ON THE LAW OF NAVAL OPERATIONS, § 8.11.1 (2007) [hereinafter COMMANDER'S HANDBOOK]. Some scholars argue that "[t]he law of war targeting principles of military necessity, proportionality, and unnecessary suffering govern all uses of force, whatever means employed." Sean Watts, *Combatant Status and Computer Network Attack*, 50 VA. J. INT'L L. 391, 425 (2010); *see also* Michael N. Schmitt, *Wired Warfare: Computer Network Attack and the* Jus in Bello, *in* COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW 187, 195 (Michael N. Schmitt & Brian T. O'Donnell eds., 2002) (arguing that existing norms remain intact, although a computer network attack offers new means to target nonmilitary objectives); Eric Talbot Jensen, *Unexpected Consequences from Knock-On Effects: A Different Standard for Computer Network Operations?*, 18 AM. U. INT'L L. REV. 1145 (2003) (arguing that no new legal framework is necessary).

85. This is not mere speculation. The Department of Defense issued a report in late 2011 in which it declared that the United States reserves the right to respond to cyber-attacks using "all necessary means—diplomatic, informational, military, and economic." DEP'T OF DEFENSE, CYBERSPACE POLICY REPORT: A REPORT TO CONGRESS PURSUANT TO THE NATIONAL DEFENSE AUTHORIZATION ACT FOR FISCAL YEAR 2011, SECTION 934, at 2 (2011).

86. Others argue that the law of war as it currently stands is insufficient and in need of revision in light of cyber-attacks. *See* Hollis, *supra* note 10, at 1027–28; Davis Brown, *A Proposal for an*

We turn first to the most vital question under *jus ad bellum*—when would a cyber-attack rise to the level of an armed attack justifying self-defense under Article 51 of the U.N. Charter? As indicated in Table 1 above, we conclude that the best test of when a cyber-attack is properly considered cyber-warfare is whether the attack results in physical destruction—sometimes called a "kinetic effect"—comparable to a conventional attack. Arriving at this conclusion requires examining not only the Charter's text—which is quite general and vague—but also the meaning given to that text by state practice and interpretation over time. Because an armed conflict has never begun solely as a result of a cyber-attack, there is no state practice on what cyber-attacks justify an armed response. Accordingly, the legal analysis here is necessarily speculative.

We turn next to applying the law of war once armed conflict has commenced, or *jus in bello*, to cyber-warfare. This body of law is less speculative, as there have been documented incidents of cyber-attacks in the context of an armed conflict. Even so, it is challenging to apply even widely accepted core *jus in bello* principles of proportionality and distinction to cyber-warfare. These challenges illustrate the importance of commencing an international dialogue on these issues to bring clarity to existing law of war principles in this context. They also demonstrate that the law of war alone cannot address the new challenges posed by cyber-attacks.

A. Jus ad Bellum

What law governs states' right to resort to armed force in self-defense against cyber-attacks? To answer this question, we proceed in three steps. First, we outline the general prohibition on the use or threat of force in international relations contained in Article 2(4) of the U.N. Charter. Second, we discuss the exceptions to that prohibition for collective security operations and self-defense, paying particular attention to when a cyber-attack would justify resort to self-defense. Finally, we close by explaining the customary international law requirements of *jus ad bellum* necessity and proportionality and by detailing the limitations and problems of applying *jus ad bellum* requirements to cyber-attacks. We conclude that states may only use defensive armed force in response to a cyber-attack if the effects of the attack are equivalent to those of a conventional armed attack.

1. Governing Legal Principles: Prohibition on Use of Force and Intervention in Internal Affairs

Article 2(4) of the U.N. Charter provides that member states "shall refrain in their international relations from the threat or use of force against the

---

International Convention to Regulate the Use of Information Systems in Armed Conflict, 47 HARV. INT'L L.J. 179 (2006).

territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."[87] This prohibition is complemented by a customary international law norm of nonintervention, which prohibits states from interfering in the internal affairs of other states.[88] The International Court of Justice ("ICJ") has held that, where the interference takes the form of a use or threat of force, the customary international law norm of nonintervention is coterminous with Article 2(4).[89]

The precise scope of the international prohibition on the threat or use of force has been the subject of intense international and scholarly debate. Weaker states and some scholars have argued that Article 2(4) broadly prohibits not only the use of armed force, but also political and economic coercion. Nonetheless, the consensus is that Article 2(4) prohibits only armed force.[90]

Discussions about cyber-attacks have the potential to reignite debates over the scope of Article 2(4).[91] Because it is much less costly to mount cyber-attacks than to launch conventional attacks, and because highly industrialized states are generally more dependent upon computer networks and are more vulnerable to cyber-attacks, cyber-attacks may prove to be a powerful weapon of the weak. This change in the cost structure of offensive capabilities may both increase the likelihood of cyber-attacks and change the political valence of different interpretations of Article 2(4)'s scope. Stronger states may begin to favor more expansive readings of Article 2(4) that prohibit coercive activities like cyber-attacks.[92]

---

87. U.N. Charter art. 2, para. 4.

88. *See* G.A. Res. 37/10, U.N. Doc. A/RES/37/10 (Nov. 15, 1982); G.A. Res. 25/2625, U.N. Doc. A/RES/25/2625 (Oct. 24, 1970).

89. Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), 1986 I.C.J. 14, para. 209 (June 27) ("[A]cts constituting a breach of the customary principle of nonintervention will also, if they directly or indirectly involve the use of force, constitute a breach of the principle of non-use of force in international relations."). It is possible, however, that to the extent cyber-attacks do not constitute a use of force, they may nevertheless violate the customary international law norm of nonintervention, as discussed below.

90. Daniel B. Silver, *Computer Network Attack as a Use of Force Under Article 2(4) of the United Nations Charter*, *in* COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW 73, 80–82 (Michael N. Schmitt & Brian T. O'Donnell eds., 2002). The principal arguments for the prevailing view are: (1) that Article 2(4) was conceived against a background of efforts to limit unilateral recourse to armed force, not economic and political coercion; (2) that the *travaux preparatoires* show that the San Francisco Conference rejected a proposal that would have extended Article 2(4) to include economic sanctions; and (3) that the ICJ has held that financing armed insurrection does not constitute force, indicating that other economic measures that are even less directly related to armed violence would not constitute prohibited force either. *Id.* at 81. There remains some ambiguity, however, as to the extent to which Article 2(4) prohibits nonmilitary physical force, such as flooding, forest fires, or pollution. *Id.* at 82–83.

91. *See* Waxman, *supra* note 24.

92. Walter Sharp has advocated that the United States make precisely this kind of strategic interpretive move, arguing that a broad array of coercive cyber-activities should fall within Article 2(4)'s prohibition. WALTER GARY SHARP, SR., CYBERSPACE AND THE USE OF FORCE 129–33 (1999).

Cyber-attacks may also violate the customary international law norm of nonintervention, as defined by a growing record of state practice and opinion juris. First, states generally do not engage in cyber-attacks openly, but rather try to hide their responsibility by camouflaging attacks through technical means[93] and by perpetrating the attacks through non–state actors with ambiguous relationships to state agencies.[94] As Thomas Franck has observed, "[l]ying about facts . . . is the tribute scofflaw governments pay to international legal obligations they violate."[95] In other words, the very fact that states attempt to hide their cyber-attacks may betray a concern that such attacks may constitute unlawful uses of force. Second, when states acknowledge that they have been victims of cyber-attack, they and their allies tend to denounce and condemn the attacks.[96] Third, in its common approach to cyber-defense, NATO has indicated that cyber-attacks trigger states parties' obligations under Article 4 of the NATO treaty,[97] which applies only when "the territorial integrity, political independence or security of any of the Parties is threatened."[98] The invocation of this provision strongly suggests that NATO member states believe that cyber-attacks violate the customary norm of nonintervention or a related international law norm.[99] Still, as the next Subsection explains, the fact that a cyber-attack is unlawful does not necessarily mean that armed force can be used in response.

2. Exceptions for Collective Security and Self-Defense

Article 2(4)'s blanket prohibition on the nonconsensual use or threat of force is subject to two exceptions: actions taken as part of collective security operations and actions taken in self-defense.

The first exception falls under Article 39 of the U.N. Charter. Article 39 empowers the Security Council to "determine the existence of any threat to the peace, breach of the peace, or act of aggression, and [to] make

93. *See* Matthew J. Sklerov, *Solving the Dilemma of State Responses to Cyberattacks: A Justification for the Use of Active Defenses Against States Who Neglect Their Duty to Prevent*, 201 MIL. L. REV., Fall 2009, at 1, 74–75.

94. *See, e.g.*, CARR, *supra* note 30, at 29 ("Hacking attacks cloaked in nationalism are not only not prosecuted by Russian authorities, but they are encouraged through their proxies, the Russian youth associations, and the Foundation for Effective Policy.").

95. Thomas M. Franck, *Legitimacy After Kosovo and Iraq*, *in* INTERNATIONAL LAW AND THE USE OF FORCE AT THE TURN OF CENTURIES: ESSAYS IN HONOUR OF V.D. DEGAN 69, 73 (Vesna Crnić-Grotić & Miomir Matulović eds., 2005).

96. *See, e.g.*, Ian Traynor, *Russia Accused of Unleashing Cyberwar to Disable Estonia*, GUARDIAN (May 16, 2007), http://www.guardian.co.uk/world/2007/may/17/topstories3.russia (detailing the reactions by Estonian, EU, and NATO officials to a cyber-attack on Estonia).

97. *NATO Agrees Common Approach to Cyber Defence*, EURACTIV.COM (Apr. 4, 2008), http://www.euractiv.com/en/infosociety/nato-agrees-common-approach-cyber-defence/article-171377.

98. North Atlantic Treaty, art. 4, Apr. 4, 1949, 63 Stat. 2241, 34 U.N.T.S. 243.

99. As noted below, however, NATO does not believe that cyber-attacks rise to the level of armed attacks justifying self defense. *See NATO Agrees Common Approach to Cyber Defence*, *supra* note 97; *infra* Part II.A.2.

recommendations, or decide what measures shall be taken . . . to maintain or restore international peace and security."[100] The Security Council may employ "measures not involving the use of armed force"[101] and authorize "action by air, sea, or land forces."[102] Collective security operations under Article 39 can be politically difficult, however, because they require authorization by the often deadlocked or slow-moving Security Council.

The second exception to Article 2(4) is codified in Article 51, which provides that "[n]othing in the present Charter shall impair the inherent right of individual or collective self-defence if an armed attack occurs."[103] Lawful self-defense can be harder to define and identify than lawful collective security operations. Indeed, in many armed conflicts, both sides claim to be acting in self-defense, and the international debates tend to focus on factual and political disputes rather than legal doctrine.[104] It is clear, however, that the critical question determining the lawfulness of self-defense is whether or not an armed attack has occurred. A cyber-attack must rise to the level of an armed attack for a state to respond lawfully under Article 51.[105]

The term "armed attack" is linguistically distinct from several other related terms in the U.N. Charter and has been interpreted to be substantively narrower than them.[106] For example, there may be acts that violate Article 2(4)'s prohibition on the use or threat of force that do not rise to the level of an armed attack, and hence do not trigger the right of self-defense under Article 51. The ICJ has indicated that cross-border incursions that are minor in their "scale and effects" may be classified as mere "frontier incident[s]" rather than "armed attacks."[107] Instead, to qualify as armed attacks sufficient to justify a

100. U.N. Charter art. 39.

101. *Id.* art. 41.

102. *Id.* art. 42.

103. *Id.* art. 51. For example, the White House's recent cyberspace strategy paper includes the right of self-defense as one of the norms that should guide conduct in cyberspace. WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 10.

104. CHRISTINE GRAY, INTERNATIONAL LAW AND THE USE OF FORCE 95–96 (2d ed. 2004).

105. *See, e.g.*, WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 14 ("When warranted, the United States will respond to hostile acts in cyberspace as we would to any other threat to our country. All states possess an inherent right to self-defense, and we recognize that certain hostile acts conducted through cyberspace could compel actions under the commitments we have with our military treaty partners.").

106. *See* Yoram Dinstein, *Computer Network Attacks and Self-Defense*, *in* COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW 99, 100–01 (Michael N. Schmitt & Brian T. O'Donnell eds., 2002).

107. Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 195 (June 27); *cf.* Definition of Aggression, G.A. Res. 29/3314, Annex, art. 2, U.N. Doc. A/RES/29/3314 (Dec. 14, 1974) [hereinafter Definition of Aggression] (determining that "[t]he first use of armed force by a State in contravention of the Charter shall constitute *prima facie* evidence of an act of aggression although the Security Council may . . . conclude that a determination that an act of aggression has been committed would not be justified in the light of other relevant circumstances, including the fact that the acts concerned or their consequences are not *of sufficient gravity*" (emphasis

response under Article 51, attacks must constitute "most grave forms of the use of force."[108] Where they may not resort to defensive force under Article 51 (because an attack does not rise to the level of an "armed attack"), states may be permitted to respond with retorsions or nonforceful countermeasures within carefully proscribed legal limits.[109] As described in more detail in Part III.A, such countermeasures might include responses in cyberspace.[110]

In scholarly debates over the application of *jus ad bellum* to cyber-attacks, three leading views have emerged to determine when a cyber-attack constitutes an armed attack that triggers the right of armed self-defense: the instrument-based approach, the target-based approach, and our preferred approach: the effects-based approach.[111]

One scholar has given the moniker "instrument-based" to the classical approach to the armed attack inquiry.[112] Under this view, a cyber-attack alone will almost never constitute an armed attack for purposes of Article 51 "because it lacks the physical characteristics traditionally associated with military coercion"—in other words, because it generally does not use traditional military weapons.[113] This approach treats a cyber-attack as an armed

---

added)). Scholars generally agree that there is a gap between the prohibition on the use of force and the right of self-defense. *See, e.g.*, Dinstein, *supra* note 106, at 99, 100–01.

108. Military and Paramilitary Activities in and Against Nicaragua (Nicar. v. U.S.), 1986 I.C.J. 14, ¶ 191 (June 27).

109. Retorsions are lawful unfriendly acts made in response to an international law violation by another state; countermeasures are acts that would be unlawful if not done in response to a prior international law violation. U.N. Int'l Law Comm'n Draft Articles on Responsibility of States for Internationally Wrongful Acts, Rep. of the Int'l Law Comm'n, U.N. GAOR, 53d Sess., Supp. No. 10, U.N. Doc. A/56/10 (2001), at 31, 80 [hereinafter Draft Articles]. *See infra* Part III.A for a more detailed discussion of countermeasures.

110. *See* OFFICE OF GEN. COUNSEL, DEP'T OF DEF., AN ASSESSMENT OF INTERNATIONAL LEGAL ISSUES IN INFORMATION OPERATIONS (1999), *reprinted in* COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW 459, 484–85 [hereinafter DOD MEMO] (Michael N. Schmitt & Brian T. O'Donnell eds., 2002) ("If the provocation is not considered to be an armed attack, a similar response will also presumably not be considered to be an armed attack.").

111. Once a state has been the victim of an armed attack, a further question arises as to against whom the state can respond. Where the armed attack is perpetrated by a state, this question is easily answered—self-defense may be directed against the perpetrating state. However, cyber-attacks may be perpetrated by non–state actors or by actors with unclear affiliations with state security agencies. Although some scholars argue that cyber-attacks (and conventional attacks) must be attributable to a perpetrating state in order for the victim state to take defensive action that breaches another state's territory, others—drawing on traditional jurisprudence on self-defense—argue that states possess the right to engage in self-defense directly against non–state actors if certain conditions are met. *See* Jordan J. Paust, *Self-Defense Targetings of Non-State Actors and Permissibility of U.S. Use of Drones in Pakistan*, 19 J. TRANSNAT'L L. & POL'Y 237, 238–39 (2010) ("The vast majority of writers agree that an armed attack by a non-state actor on a state, its embassies, its military, or other nationals abroad can trigger the right of self-defense addressed in Article 51 of the United Nations Charter, even if selective responsive force directed against a non-state actor occurs within a foreign country.").

112. Michael N. Schmitt, *Computer Network Attack and the Use of Force in International Law: Thoughts on a Normative Framework*, 37 COLUM. J. TRANSNAT'L L. 885, 909 (1999); *see also* Hollis, *supra* note 10, at 1041.

113. Hollis, *supra* note 10, at 1041.

attack only if it uses military weapons. Bombing computer servers or Internet cables could meet the requirements of an armed attack, for example, if the strike was of sufficient gravity.

The text of the U.N. Charter provides some support for the instrument-based approach, since Article 41 characterizes the "complete or partial interruption of . . . telegraphic, radio, and other means of communication" as a "measure[] not involving the use of armed force."[114] The U.N. General Assembly's Definition of Aggression also implicitly supports the instrument-based view: it lists a number of acts that would constitute "aggression" under Article 39—a broader category than armed attack under Article 51—and all of them involve military weapons or force.[115] NATO has also signaled its agreement with this view; its new common approach to cyber-defense establishes that a cyber-attack will obligate member states to "consult" with one another under Article 4 of the NATO treaty, but a cyber-attack will not constitute an armed attack that obligates member states to assist one another under Article 5 of the treaty.[116]

The chief advantage of the instrument-based approach is simplicity of application, since uses of military weapons and force are relatively easy to identify. However, because cyber-attacks have the potential to cause catastrophic harm without employing traditional military weapons, most scholars have rejected the instrument-based approach to defining armed attacks as dangerously outdated.

Recognizing the fundamental inability of the instrument-based approach to account for harms not caused by conventional means, the target-based approach classifies as an armed attack any cyber-attack that targets a sufficiently important computer system.[117] The primary aim of this approach is to determine when a cyber-attack portends imminent harm sufficient to justify the use of anticipatory self-defense in response.[118]

While the target-based approach has the benefit of allowing for aggressive protection of critical national systems, it broadly sanctions forceful self-defense, increasing the likelihood that cyber-conflicts will escalate into more

---

114. U.N. Charter art. 41.

115. *See* Definition of Aggression, *supra* note 107, art. 3.

116. North Atlantic Treaty, *supra* note 98, arts. 4, 5, 63; *NATO Agrees Common Approach to Cyber Defence*, *supra* note 97.

117. Walter Sharp, the leading proponent of this approach, argues that a cyber-attack constitutes an armed attack, and would grant the target the right to use force in self-defense whenever it penetrates any critical national infrastructure system, regardless of whether it has yet caused any physical destruction or casualties. SHARP, *supra* note 92, at 129–30; *see also* Sean M. Condron, *Getting it Right: Protecting American Critical Infrastructure in Cyberspace*, 20 HARV. J.L. & TECH. 403, 415-16 (2007) (advocating a similar approach); Eric Talbot Jensen, *Computer Attacks on Critical National Infrastructure: A Use of Force Invoking the Right of Self-Defense*, 38 STAN. J. INT'L L. 207, 208–09 (2002) (same).

118. Hollis, *supra* note 10, at 1041 n.73.

destructive conventional armed conflicts.[119] A cyber-attack need only penetrate a critical system to justify a conventional military response that could start a physical, kinetic war. This approach could undermine the security of the international community by making war much more likely.

Finally, the effects-based approach classifies a cyber-attack as an armed attack based on the gravity of its effects. Steering a middle course between the instrument- and target-based views, the effects-based approach is the most promising and most widely accepted approach. Different versions of the effects-based approach may measure that gravity by reference to any of a variety of factors, from the sheer severity of the harm to the length of the causal chain between the cyber-attack itself and the ultimate harm. But all versions of this approach share a common orientation towards the inquiry.

The problem with the effects-based approach, however, lies in articulating ex ante what types of effects justify self-defense.[120] Consider, for example, an attack on an air traffic control system, an attack that disables a regional electrical power grid, an attack on the New York Stock Exchange or national financial networks, or the 2007 cyber-attack on prominent Estonian websites. Which of these cyber-attacks, if any, have effects large enough to be considered armed attacks justifying the use of defensive force in response? All of these attacks may cause small- or large-scale civilian deaths and infrastructure damage, but it would be difficult for the aggressor country to predict the outcome of any individual attack. Different versions of the effects-based approach may reach different conclusions for each of these examples.

Professor Michael Schmitt, the best-known proponent of the effects-based approach for determining when a cyber-attack should be considered an armed attack, argues that a cyber-attack's effects should be measured by reference to six factors: (1) severity: the type and scale of the harm; (2) immediacy: how quickly the harm materializes after the attack; (3) directness: the length of the causal chain between the attack and the harm; (4) invasiveness: the degree to which the attack penetrates the victim state's territory; (5) measurability: the degree to which the harm can be quantified; and (6) presumptive legitimacy: the weight given to the fact that, in the field of cyber-activities as a whole, cyber-attacks constituting an armed attack are the exception rather than the rule.[121] These factors are illuminating, but they call for such a wide-ranging

119. *See* Sklerov, *supra* note 93, at 56 n.352 (criticizing the target-based approach for encouraging escalation and advocating an effects-based approach).

120. This difficulty is aggravated by the reality that the "indirect effects" of cyber-attacks are often "more consequential" than the immediate ones. NRC REPORT, *supra* note 23, at 19.

121. Schmitt, *supra* note 112, at 914–15; *see also* Sean P. Kanuck, *Recent Development: Information Warfare: New Challenges for Public International Law*, 37 HARV. INT'L L.J. 272, 290 (1996) ("Each suspect activity could be reviewed for its effects on other states, and sanctioned accordingly.").

inquiry that they may not provide sufficient guidance to decision makers.[122] In other words, different analysts applying this version of the effects-based approach might plausibly classify all or none of the examples listed above as armed attacks.

Daniel Silver, former General Counsel of the CIA and National Security Agency, argues instead that the key criterion determining when a cyber-attack constitutes an armed attack is the severity of the harm caused. A cyber-attack justifies self-defense "only if its foreseeable consequence is to cause physical injury or property damage and, even then, only if the severity of those foreseeable consequences resembles the consequences that are associated with armed coercion."[123] Under this test, a cyber-attack on the air traffic control system causing planes to crash would be regarded as an armed attack because it is foreseeable that such an attack would cause loss of life and substantial property damage. But a cyber-attack on a website or mere penetration of a critical computer system generally would not, unless it caused physical injury or property damage. A cyber-attack on financial systems presents a harder case for this approach—the analysis would depend on whether the attack was found to have caused substantial damage to property.

It is important to note that the purpose of the attack is already accounted for in the definition of cyber-attack recommended herein: the attack must have been committed for a political or national security purpose. Therefore a cyber-attack that has unforeseen national security consequences would not be considered a cyber-attack, much less cyber-warfare.

This final version of the effects-based approach provides the best balance between enabling states to adequately respond to catastrophic cyber-attacks and preventing states from resorting to armed force too easily. The test defines a small core of harmful cyber-attacks that rise to the level of an armed attack.[124] It also focuses the armed attack analysis on a limited set of criteria—particularly severity and foreseeability.[125]

---

122. *See* Silver, *supra* note 90, at 89 (claiming that "examination of [Schmitt's] criteria suggests that virtually any event of [computer network attack] can be argued to fall on the armed force side of the line"); *see also* Jason Barkham, *Information Warfare and International Law on the Use of Force*, 34 N.Y.U. J. INT'L L. & POL. 57, 85–86 (2001) (criticizing Schmitt's use of presumptive legitimacy as a criterion, as well as Schmitt's assumption that policy makers will be able to engage in a thorough factual inquiry when responding to cyber-attacks).

123. Silver, *supra* note 90, at 90–91.

124. *See id.* at 92.

125. The Department of Defense has signaled its approval of this approach. *See* DOD MEMO, *supra* note 110, at 483 (arguing "the consequences are likely to be more important than the means used," and providing examples of cyber-attacks that would cause civilian deaths and property damage).

3. Ad Bellum *Necessity and Proportionality*

A state's use of armed force in response to a cyber-attack must not only conform with U.N. Charter and customary international law limits on the use of armed force, but it must also comply with the *jus ad bellum* principles of necessity and proportionality under customary international law. The principle of necessity requires that force must be used only as a last resort, when peaceful means, such as a diplomatic settlement, cannot achieve the state's overall aim.[126] Proportionality extends this logic, prohibiting force if the overall scope and intensity of force is excessive in relation to the state's actual or imminent danger.[127] The United States has acknowledged that these principles apply to military responses to cyber-attacks.[128]

While principles of necessity and proportionality are clear, applying those principles to state responses to cyber-attacks is challenging. Evaluating whether an invocation of self-defense complies with the principles of necessity and proportionality is difficult and fact intensive even for conventional attacks, and cyber-attacks present hard new questions. For example, cyber-attacks rising to the level of armed attacks may require decision makers to devise ways of measuring harm to computer networks and its indirect effects against more conventional kinds of harm in order to determine what would constitute a lawful response.

Applying the existing *jus ad bellum* framework in the context of cyber-attacks is challenging. Moreover, the framework only applies to the small subset of cyber-attacks that are addressed by Security Council resolutions or that constitute an armed attack, giving rise to a right of self-defense under Article 51. As a result, only a small number of cyber-attacks are properly considered "cyber-warfare," to which the laws of war apply. Part III of this Article explores other international legal regimes that may help to regulate cyber-attacks that do not fall within these narrow boundaries. First, however, the following Section describes the legal framework governing cyber-attacks during an ongoing armed conflict.

---

126. *See, e.g.*, R.Y. Jennings, *The* Caroline *and* McLeod *Cases*, 32 AM. J. INT'L L. 82, 89 (1938) (quoting Secretary of State Daniel Webster's letter to his British counterpart concerning the *Caroline* incident as follows: "It must be shown that admonition or remonstrance to the persons on board the *Caroline* was impracticable, or would have been unavailing . . . but that there was a necessity, present and inevitable, for attacking her . . . .").

127. *See* Robert D. Sloane, *The Cost of Conflation: Preserving the Dualism of* Jus ad Bellum *and* Jus in Bello *in the Contemporary Law of War*, 34 YALE J. INT'L L. 47, 108–09 (2009) ("*Ad bellum* proportionality is . . . parasitic on *ad bellum* necessity. . . . An act is *ad bellum* disproportionate if the same *ad bellum* objective sought by force clearly could have been achieved by diplomacy or another nonviolent strategy at a roughly comparable, or even moderately greater, cost.").

128. *See* WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 14 ("[W]e will exhaust all options before military force whenever we can; will carefully weigh the costs and risks of action against the costs of inaction; and will act in a way that reflects our values and strengthens our legitimacy, seeking broad international support whenever possible.").

B. Jus in Bello

Although a stand-alone cyber-attack has never instigated an armed conflict, cyber-attacks have been used in wars in response to traditional provocations or to prepare the way for an imminent conventional attack. This Section examines the relationship between traditional *jus in bello* requirements and cyber-attacks employed in the course of conventional armed conflicts. The novel conditions of cyberspace can pose challenges to applying *jus in bello* principles of necessity, proportionality, distinction, and neutrality. Because cyber-attacks are often not immediately lethal or destructive and may cause only temporary incapacity of network systems, it may be hard to evaluate whether a cyber-attack is proportional. It can also be difficult to distinguish between combatants, civilians directly participating in hostilities, civilians engaged in a continuous combat function, and protected civilians in the context of cyber-attacks. Finally, the ease of masking the source of a cyber-attack makes enforcement of neutrality duties complicated and expensive. We briefly address each challenge in turn.

1. In Bello *Necessity*

Although the necessity of a cyber-attack may be difficult to evaluate, this difficulty arises from line-drawing debates that did not originate in cyber-warfare and are not unique to *in bello* cyber-attacks. *In bello* necessity relates to the concrete military advantage to be gained from a specific hostile act. An individual cyber-attack may be unnecessary if it does not advance the military's objective.[129] While cyber-attacks must be necessary to be lawful, evaluating their *in bello* necessity does not present novel challenges.

2. In Bello *Proportionality*

The *in bello* proportionality requirement prohibits "[a]n attack which may be expected to cause incidental loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof, which would be excessive in relation to the concrete and direct military advantage anticipated."[130] To

129. In contrast, the *ad bellum* necessity analysis helps determine if nonforcible measures to abate a threat are inadequate, excusing an otherwise unlawful use of force.

130. Protocol Additional to the Geneva Conventions of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol 1), art. 51(5)(b), June 8, 1977, 1125 U.N.T.S. 3 [hereinafter Protocol Additional I]; *see also id.* art. 85(3)(b). An indiscriminate attack, defined by *excessive effect*, is not to be confused with an attack that does not discriminate amongst civilian and military objectives, which is defined by *objective*, and is prohibited by article 85(3)(a). *See infra* Part II.B.3. Some scholars argue that, given the ability to avoid civilian casualties or damage to property and achieve the same military advantage, a state *must* do so. *See* DIMITRIOS DELIBASIS, THE RIGHT TO NATIONAL SELF-DEFENSE IN INFORMATION WARFARE OPERATIONS 268 (2007) (arguing that the "unmatched accuracy" of information warfare "practically nullifies the element of chance embodied in all military entanglements"); Dakota S. Rudesill, *Precision War and Responsibility: Transformational Military Technology and the Duty of Care Under the Laws of War*, 32 YALE J. INT'L

conduct a *jus in bello* proportionality analysis, a military decision maker must weigh potential civilian casualties, destruction of civilian property, and the loss of indispensable civilian items against the benefit of achieving a military objective.[131]

Due to the nature of harm they inflict, the proportionality of cyber-attacks poses unique challenges. It can be difficult to evaluate whether an attack would be proportional according to the relevant categories of "loss of civilian life, injury to civilians, damage to civilian objects, or a combination thereof," as the typical direct effects of cyber-attacks may be nonlethal or temporary, yet severe.[132] In particular, how should the temporary incapacity of critical systems be evaluated?[133] A cyber-attack that effectively stops the transmission of information through the Internet might merely inconvenience the populace, but it might also have more severe consequences. For example, it might cause hospitals to be unable to communicate vital information, leading to loss of life. An ex ante *in bello* proportionality analysis for a DDOS attack may therefore carry a much greater degree of uncertainty than would a conventional attack. An *in bello* proportionality analysis requires anticipating the probable consequences of an action, but additional uncertainty will make that analysis much more difficult in the cyber context. As a result, cyber-attacks may change the weight given to temporary consequences, and may force states to confront more uncertainty than they typically face in making decisions about the legality of planned attacks.

3. Distinction

The principle of distinction—which requires states to distinguish civilian and military personnel and restrict attacks to military objectives[134]—presents

L. 517, 535 (2007) (arguing that the United States might be held to heightened standard of care due to advances in military technology).

131. Protocol Additional I, *supra* note 130, arts. 51(5)(b), 54, 57(2)(a)(iii). After deciding that the target is a military objective, the elements of the balancing test include "target selection, the means and methods chosen for the military strike, the lack of negligence in the execution of the military strike, and the determination of what constitutes the military advantage of a particular military strike." Randy W. Stone, *Protecting Civilians During Operation Allied Force: The Enduring Importance of the Proportional Response and NATO's Use of Armed Force in Kosovo*, 50 CATH. U. L. REV. 501, 522 (2001).

132. Protocol Additional I, *supra* note 130, art. 57(2)(a)(iii).

133. Similar questions arise in debates around nonlethal deployments of biological and chemical weapons, such as riot agents. *See* James D. Fry, *Gas Smells Awful: U.N. Forces, Riot-Control Agents, and the Chemical Weapons Convention*, 31 MICH. J. INT'L L. 475 (2010); Mirko Sossai, *Drugs as Weapons: Disarmament Treaties Facing the Advances in Biochemistry and Non-Lethal Weapons Technology*, 15 J. CONFLICT & SECURITY L. 5 (2010).

134. Louise Doswald-Beck, *Some Thoughts on Computer Network Attack and the International Law of Armed Conflict*, *in* COMPUTER NETWORK ATTACK AND INTERNATIONAL LAW 163, 166 (Michael N. Schmitt & Brian T. O'Donnell eds., 2002). Distinction also imposes responsibilities on combatants to identify themselves in order to facilitate distinction on the battlefield and to receive the protections that are due to combatants. *See* Watts, *supra* note 84, at 438–39. States also have a duty to facilitate distinction: "The application of this duty requires that personnel and

another legal challenge.[135] Under this principle, military commanders must employ weapons that can target accurately and must use this capability to distinguish between civilian and military objectives.[136] By extension, the law of war prohibits *in bello* cyber-attacks that are uncontrollable, unpredictable, or do not discriminate between civilian and military objectives.[137] Furthermore, Additional Protocol I prohibits attacks that deny the civilian population indispensable objects, such as food or water supplies.[138]

There are situations where the principle of distinction is easily applied to cyber-attacks. For example, a cyber-attack that targets a military air traffic control system and only causes a troop transport to crash would comply with the principle of distinction.[139] Other cyber-attacks would clearly violate the principle of distinction—for example, an attack on the civilian banking sector or on hospitals, museums, or places of worship.[140] Cyber-attacks against the networks that manage these targets, like any other attack on these objects, would be unlawful.[141]

Such cases are easy, but cyberspace offers many much more difficult ones. The distinction analysis will often be complicated in the context of a cyber-attack because the likely targets are used by a multiplicity of actors at once. Ninety-five percent of military communications use civilian networks at some stage,[142] so it is possible that civilian networks could be attractive military targets.[143] Because much of cyberspace is dual use—used by both the

---

equipment directly engaged in information warfare be located in facilities whose attack by kinetic weapons would not result in excessive collateral damage." Brown, *supra* note 86, at 192.

135. *See* DELIBASIS, *supra* note 130, at 274 (arguing that information warfare will likely run afoul of distinction and proportionality); Kelsey, *supra* note 73, at 1431 (arguing that cyber-attacks will often violate the principles of distinction and neutrality).

136. *See* Jensen, *supra* note 84, at 1154. The ICJ has found that nuclear weapons may violate international humanitarian law if they cannot be used in a manner that distinguishes between civilians and military objectives. Legality of the Threat or Use of Nuclear Weapons, Advisory Opinion, 1996 I.C.J. 226, ¶ 78 (Jul. 8).

137. Military objectives are targets that meet two criteria: they serve a military purpose and their incapacitation conveys a definite advantage. Protocol Additional I, *supra* note 130, art. 52(2). For example, the first missile strikes of Operation Desert Storm in 1991 targeted Iraqi radar stations. Kanuck, *supra* note 121, at 282. On distinction, *see* Doswald-Beck, *supra* note 134, at 165–71; Brown, *supra* note 86, at 195 (comparing malicious code, which is indiscriminate, to biological weapons). Schmitt also argues that indiscriminate weapons are unlawful, including in that category not only cyber-attacks that cannot distinguish civilian and military objects, but also those which cannot be limited to a military objective. Schmitt, *supra* note 84, at 201 (citing Protocol Additional I, *supra* note 130, art. 51(4)).

138. Protocol Additional I, *supra* note 130, art. 54(2).

139. Schmitt, *supra* note 84, at 196 ("Military equipment and facilities . . . are clearly military objectives.").

140. *See, e.g.*, Protocol Additional I, *supra* note 130, art. 85(4)(d).

141. Schmitt, *supra* note 84, at 200; Brown, *supra* note 86, at 199.

142. Antolin-Jenkins, *supra* note 11, at 133.

143. Jensen later argues that, given that military use of civilian infrastructure makes it a legitimate military target, the U.S. government has a duty to protect civilian networks from cyber-

military and civilians—upholding the distinction requirement in cyberspace can be more challenging than it is in a conventional context.

a. Who May Lawfully Be Targeted in a Cyber-Attack?

Under the law of war, only three categories of individuals may be lawfully targeted: combatants, civilians directly participating in hostilities, and civilians acting in a continuous combat function. Civilians lose their right not to be targeted to the extent that they "take a direct part in hostilities."[144] Furthermore, under customary international law affirmed by the International Committee of the Red Cross, civilians who adopt a continuous combat function may also be targeted.[145] These rules are familiar in the post-9/11 context. Yet the unique characteristics of civilian contributions to and participation in cyber-attacks threaten to blur the line between direct participation, continuous combat function, and other types of involvement in the execution of hostilities.[146]

The civilian designer of a weapons system has traditionally not been treated as a direct participant in hostilities. However, the programmer who works with military intelligence may tweak the code to enable the attack, right up until the moment of the attack.[147] The actions of such a civilian—particularly of a civilian who regularly engages in such activity—could be considered a "continuous function [that] involves the preparation, execution, or command of acts or operations amounting to direct participation in hostilities."[148] As a result, civilians involved in cyber-attacks might be regarded as performing tasks that alter their status under the law of war, rendering them lawful targets of a counterattack.[149]

b. Who May Lawfully Carry Out a Cyber-Attack?

In addition to the question of who may be targeted in a cyber-attack, the principle of distinction restricts how states constitute their cyber-fighting

attacks. Eric Talbot Jensen, *Cyber Warfare and Precautions Against the Effects of Attacks*, 88 TEX. L. REV. 1533 (2010).

144. Protocol Additional I, *supra* note 130, art. 51(3).

145. INT'L COMM. OF THE RED CROSS, INTERPRETIVE GUIDANCE ON THE NOTION OF DIRECT PARTICIPATION IN HOSTILITIES UNDER INTERNATIONAL HUMANITARIAN LAW 16 (2009), *available at* http://www.icrc.org/eng/assets/files/other/icrc_002_0990.pdf [hereinafter ICRC, INTERPRETIVE GUIDANCE].

146. *See id.* at 37 (noting the challenge that private contractors and civilian employees pose to the definition of direct participation due to "geographic and organizational closeness").

147. Watts, *supra* note 84, at 429.

148. ICRC, INTERPRETIVE GUIDANCE, *supra* note 145, at 34.

149. Geoffrey S. Corn, *Unarmed but How Dangerous? Civilian Augmentees, the Law of Armed Conflict, and the Search for a More Effective Test for Permissible Civilian Battlefield Functions*, 2 J. NAT'L SEC. L. & POL'Y 257, 286-87 (2008). Although the principle that a civilian who directly participates in hostilities or who adopts a continuous combat function may be lawfully attacked is not in dispute, the status of a civilian who provides indispensable, contemporaneous assistance in cyber-attacks remains unresolved.

forces.[150] A state that sponsors use of force by civilians may be placing those civilians outside the protections they enjoy under the law of armed conflict, and may be undermining the principle of distinction between combatants and civilians.[151]

Despite the legal consequences, there are many reasons to think states will be tempted to use civilians in the cyber context. First, civilians may possess technical expertise that governments do not. Second, by using civilians to carry out cyber-attacks, states can mask their own involvement in such operations.[152] For example, Nashi—a pro-Kremlin youth group started by Vladimir Putin—has taken responsibility for the 2007 cyber-attacks against Estonia.[153] It has been alleged that Russian business owners fund Nashi to carry out cyber-attacks favored by the Russian government. The business owners "ingratiate themselves with the regime," and the Russian government may plausibly deny involvement in the attack.[154]

A former Special Assistant for Law of War Matters of the Judge Advocate General, Lieutenant Colonel Geoffrey S. Corn, argues that the current direct participation test is outdated.[155] He offers a new functional discretion test to determine who may carry out a cyber-attack based on whether "the exercise of discretion associated with this function [will] implicate [law of war] compliance."[156] Operating within a command relationship is the dispositive criterion for combatant status "because members of the armed forces are subject to responsible command, and they operate within a military hierarchy

150. Watts, *supra* note 84, at 420.

151. *See* DELIBASIS, *supra* note 130, at 281. The allocation of responsibilities for cyber-warfare has been examined by the U.S. armed forces—the recently declassified Air Force cyberspace operations document explains that National Guard members may train for, but not carry out, cyber-attacks. *See* U.S. AIR FORCE, CYBERSPACE OPERATIONS: AIR FORCE DOCTRINE DOCUMENT 3-12, at 29 (2010), *available at* http://www.fas.org/irp/doddir/usaf/afdd3-12.pdf. Even though the United States has launched a new Cyber Command, the details of responsibility for defending against a cyber-attack are still being worked out. *See* Jim Garamone, *Official Details DOD Cybersecurity Environment*, AM. FORCES PRESS SERV. (Oct. 20, 2010), http://www.defense.gov/News/NewsArticle.aspx?ID=61356 ("Government and private officials are grappling with basics such as what constitutes a cyber attack and who has responsibility to defend against threats."). The DoD strategy emphasizes partnering with the private sector to encourage innovation, incremental improvements, and workforce development, but says little about the nature of those collaborations. *See* DOD STRATEGY, *supra* note 32, at 10–11.

152. States that do so may not only deny those civilians the protections due to civilians under the laws of war, but may also be guilty of perfidy themselves. *See* Protocol Additional I, *supra* note 130, art. 37.

153. *See* Hollis, *supra* note 10, at 1024–25 (describing the attacks against Estonia); Shachtman, *supra* note 57.

154. Shachtman, *supra* note 57.

155. *Cf. supra* note 146 and accompanying text.

156. Corn, *supra* note 149, at 287. Corn emphasizes the importance of distinction and law of war compliance, for regular forces and for paramilitaries. *Id.* at 264–65. This functional test is different from Schmitt's consequences test, which focuses on whether the cyber-attack would cause foreseeable death, injury, or destruction.

involving training, discipline, and unitary loyalty."[157] Corn argues that only individuals subject to command authority should be able to exercise discretion that could result in a law of armed conflict violation, because the actions of those individuals are within a command and discipline structure that can prevent and punish violations.[158] Under this reasoning, states may not employ civilian contractors to carry out activities where they will exercise discretion that implicates the law of armed conflict.

4. Neutrality

A final challenge in evaluating the legality of an *in bello* cyber-attack is the fact that a cyber-attack may appear to originate, or may actually originate, from a neutral state.[159] A state may be neutral, either permanently, such as Switzerland, or for the duration of a specific conflict.[160] The principle of neutrality includes both rights and responsibilities: "The principal right of the neutral nation is that of inviolability; its principal duties are those of abstention and impartiality. Conversely, it is the duty of a belligerent to respect the former and its right to insist upon the latter."[161]

Scholars hold differing views regarding neutral states' obligations to guard against the use of their facilities by belligerents. Some argue that neutral states are not obligated to stop belligerents from using their communications facilities, but they may not help belligerents build such facilities.[162] Others argue that neutral states that are unable or unwilling to stop an unlawful attack originating from their territory, including their information systems, may lawfully be targeted for the purpose of stopping the unlawful attack.[163] They claim that states have an obligation not only to refrain from committing cyber-attacks themselves, but also "not to allow knowingly [their] territory to be used for acts contrary to the rights of other States."[164]

---

157. Corn, *supra* note 149, at 287; *see also* Brown, *supra* note 86, at 191 (arguing that only armed forces should carry out cyber-attacks). *But see* SUSAN W. BRENNER, CYBERTHREATS: THE EMERGING FAULT LINES OF THE NATION STATE 199 (2009) (arguing that the rationale for excluding civilians was to protect them from retaliatory attack, but since civilian infrastructure is very likely to be attacked in cyber-warfare, this rationale for excluding civilians from combat is less persuasive).

158. Corn, *supra* note 149, at 261.

159. *See* BRENNER, *supra* note 157, at 131–32 (noting the difficulty of identifying attackers in the cyber-threat context); *see also* Brown, *supra* note 86, at 208 (on rights and responsibilities of neutrality).

160. *See* George K. Walker, *Information Warfare and Neutrality*, 33 VAND. J. TRANSNAT'L L. 1079, 1141–42 (2000) (discussing neutrality and information warfare).

161. COMMANDER'S HANDBOOK, *supra* note 84, ¶ 7.2 (noting also that "[t]his customary law has, to some extent, been modified by the Charter of the United Nations").

162. *See* Doswald-Beck, *supra* note 134, at 176.

163. *See* DELIBASIS, *supra* note 130, at 284; COMMANDER'S HANDBOOK, *supra* note 84, ¶ 7.3.

164. Corfu Channel Case (U.K. v. Albania) (Merits), 1949 I.C.J. 4, 22 (Apr. 9). *See, e.g.*, Sklerov, *supra* note 93, at 43.

Certain characteristics of cyber-attacks make the evaluation of the principle of neutrality unusually complex. Cyber-attacks may harness zombie computers located in one country to harm networks in another country—without the knowledge of any individual, much less the government—by masking their origin through a series of servers and computers.[165] Such cyber-attacks pose challenges to analysis under the principle of neutrality for two reasons. First, a country may not know its computers are being used for a cyber-attack, and it therefore may not know its neutrality is threatened. Second, the principle of neutrality determines lawful responses to attacks based on the identity of the origin country. Consequently, the inability to attribute attacks to a certain state impedes the neutrality analysis.[166] However, it is also possible that political uncertainty about lawful responses to cyber-attack may be masquerading as an inability to attribute attacks; further clarity around the legal framework governing cyber-attacks may reduce barriers to attribution. While the political problems of attribution might contribute to the apparent difficulties of attribution, the possibility remains that a country may not know attacks are emanating from its borders.

The existing law of war framework—both *jus ad bellum* and *jus in bello*—provides some guidance, albeit incomplete and imperfect, for states seeking to determine the scope of permissible offensive and defensive cyber-attacks. But it does not regulate the vast majority of cyber-attacks. Most cyber-attacks do not rise to the level of an armed attack or take place in the context of an armed conflict. Consequently, they do not implicate the law of war. Yet this does not necessarily mean that these cyber-attacks are unregulated. As the next Part shows, there are a variety of other legal frameworks that fill some of the gaps left by the law of war framework.

III.
OTHER LEGAL FRAMEWORKS GOVERNING CYBER-ATTACKS

There are several existing legal frameworks in addition to the law of war that explicitly or implicitly regulate cyber-attacks. We begin with what is potentially the most important such framework—the international law of countermeasures, which regulates how states may respond to international law violations that do not rise to the level of an armed attack justifying self-defense. Next, we outline the international legal regimes that directly regulate some elements of cyber-attacks. We then describe international legal regimes that indirectly govern some cyber-attacks by regulating the means through which those attacks are conducted. Finally, we examine U.S. domestic laws that could be used to address some cyber-attacks.

165. Goldsmith, *supra* note 54, at 10–12.

166. Shanker & Bumiller, *supra* note 53 ("Officials say the main challenge for the United States in a retaliatory cyberoperation is determining the attacker.").

These other bodies of law offer victims of cyber-attacks useful tools for responding to attacks. Yet each individual tool has significant limits. Even taken together, the legal framework is piecemeal and incomplete. This should come as no surprise: much of the law that applies to cyber-attacks was not designed for this purpose. This Part sets the stage for reflections on legal reforms that would enable domestic and international law to more effectively regulate cyber-attacks.

A. Countermeasures

The customary international law of countermeasures governs how states may respond to international law violations that do not rise to the level of an armed attack justifying self-defense—including, implicitly, cyber-attacks. The Draft Articles on State Responsibility define countermeasures as "measures that would otherwise be contrary to the international obligations of an injured State *vis-à-vis* the responsible State, if they were not taken by the former in response to an internationally wrongful act by the latter in order to procure cessation and reparation."[167]

The international law of countermeasures does not define when a cyber-attack is unlawful. Indeed, the Draft Articles do not directly address cyber-attacks at all. The law simply provides that when a state commits an international law violation, an injured state may respond with a countermeasure.[168] As explained above, some cyber-attacks that do not rise to the level of an armed attack nonetheless violate the customary international law norm of nonintervention.[169] These violations of international law may entitle a harmed state to use countermeasures to bring the responsible state into compliance with the law.

The Draft Articles lay out the basic customary international law principles regulating states' resort to countermeasures.[170] The Draft Articles provide that countermeasures must be targeted at the state responsible for the prior wrongful act and must be temporary and instrumentally directed to induce the responsible state to cease its violation.[171] Accordingly, countermeasures cannot

---

167. Draft Articles, *supra* note 109, at 128. Traditionally, these acts were termed "reprisals," but this report follows the Draft Articles in using the more modern term "countermeasures." Reprisals now predominantly refer to forceful belligerent reprisals. *Id.*

168. States thus resort to countermeasures at their own risk. If the use of countermeasures does not comply with the applicable international legal requirements, the state may itself be responsible for an internationally wrongful act. *Id.* at 130.

169. *See supra* Part II.A.1.

170. Countermeasures are distinct from retorsions. Retorsions are acts that are unfriendly but lawful, such as limiting diplomatic relations or withdrawing from voluntary aid programs, and they always remain a lawful means for a State to respond to a cyber-attack or other international legal violation.

171. Draft Articles, *supra* note 109, at 129. Accordingly, the law of countermeasures does not specify how states may respond to international law violations by non–state actors. However, international law violations by non–state actors often lead to international law violations by states. For

be used if the international law violation has ceased. Countermeasures also can never justify the violation of fundamental human rights, humanitarian prohibitions on reprisals, or peremptory international norms, nor can they excuse failure to comply with dispute settlement procedures or to protect the inviolability of diplomats.[172]

Before resorting to countermeasures, the injured state generally must call upon the responsible state to cease its wrongful conduct, notify it of the decision to employ countermeasures, and offer to negotiate a settlement.[173] However, in some situations, the injured state "may take such urgent countermeasures as are necessary to preserve its rights."[174] Countermeasures need not necessarily be reciprocal, but reciprocal measures are favored over other types because they are more likely to comply with the requirements of necessity and proportionality.[175]

Under the customary law of countermeasures, an attacking state that violates its obligation not to intervene in another sovereign state through a harmful cyber-attack may be subject to lawful countermeasures by the injured state. Such countermeasures might go beyond "passive defenses" that aim to repel cyber-attacks (such as firewalls), and constitute "active defenses," which attempt to disable the source of an attack.[176] Active defenses—if properly designed to meet the requirements of necessity and proportionality—might be considered a form of "reciprocal countermeasures," in which the injured state ceases obeying the same or a related obligation to the one the responsible state violated (in this case, the obligation of nonintervention).

Before a state may use active defense as a countermeasure, however, it must determine that an internationally wrongful act caused the state harm and identify the state responsible, as well as abide by other restrictions.[177] The countermeasures must be designed, for example, to induce the wrongdoing state to comply with its obligations. The Draft Articles also have detailed provisions regarding when acts committed by non–state agents may be

---

example, if a non–state actor launches an attack on state *A* from state *B*'s territory and state *B* is unwilling or unable to stop it, state *B* may violate an international law obligation to prevent its territory from being used for cross-border attacks. *See, e.g.*, Corfu Channel Case (U.K. v. Albania) (Merits), 1949 I.C.J. 4, 22 (Apr. 9) (holding that states are obligated "not to allow knowingly its territory to be used for acts contrary to the rights of other States"). In the cyber-attack context, a state may commit an international law violation by allowing harmful cyber-attacks to be launched from its territory. *See* Sklerov, *supra* note 93, at 62–72.

172. Draft Articles, *supra* note 109, at 131.

173. *Id.* at 135.

174. *Id.*

175. *Id.* at 129.

176. DoD has recently made clear that it employs such "active cyber defense" to "detect and stop malicious activity before it can affect DoD networks and systems." DOD STRATEGY, *supra* note 32, at 7.

177. Draft Articles, *supra* note 109, at 129–34.

attributed to a state—for instance, when the state aids and assists the act with knowledge of the circumstances.[178]

While countermeasures provide states with a valuable tool for addressing cyber-attacks that do not rise to the level of an armed attack, countermeasures are far from a panacea. First and foremost, they require the identity of the attacker and the computer or network from which the attack originates to be accurately identified. Second, in order for a countermeasure to be effective, the targeted actor must find the countermeasure costly—ideally costly enough to cease its unlawful behavior. If the target can easily relocate its operations, as is often possible in the cyber context, the countermeasure may not impose a significant cost on the actor responsible for the attack. For this reason, countermeasures are likely to be more effective against state actors and less effective against non–state actors. Finally, it can be difficult to design a countermeasure that injures only the actor that perpetuated the legally wrongful attack. In particular, a countermeasure that disables a computer or network may very well cause harm to those who have little or nothing to do with the unlawful attacks. This could have the perverse effect of making the state injured by the original attack a perpetrator of an unlawful attack against those who simply happen to share a network with the actor that generated the original attack, or whose computers were used as pawns without its knowledge or acquiescence. Together, these challenges can lead a system that relies too heavily on active countermeasures to spin out of control. As a result, the customary law of countermeasures offers only a partial answer to the problem of cyber-attacks. We thus turn next to other international legal regimes that directly regulate cyber-attacks.

B. International Legal Regimes That Directly Regulate Cyber-Attacks

While no comprehensive international legal framework currently governs all cyber-attacks, a patchwork of efforts provides some tools the United States and other countries can employ to control this growing threat. This Section surveys legal mechanisms created by the United Nations, NATO, the Council of Europe, the Organization of American States, and the Shanghai Cooperation Organization to directly regulate cyber-attacks. While both the Council of Europe and the Organization of American States have taken actions relating to cyber-crime—a category of activity that overlaps in part with cyber-attacks, as noted above—the increased computer network protection and regulations are also relevant to efforts to combat cyber-attacks. Collectively, these organizational measures demonstrate a growing interest in addressing this issue through common legal frameworks. Yet these efforts have thus far fallen short of establishing a rigorous legal framework that can effectively govern all cyber-attacks.

178. *Id.* at 65.

1. The United Nations

There has been only limited U.N. action on the issue of cyber-security. The U.N. General Assembly has passed several related resolutions.[179] These resolutions, however, are vague and have not required any specific action by U.N. members.[180]

In August 1999, the United Nations sponsored an international meeting of experts in Geneva to better grasp the security implications of emerging information technologies.[181] A follow-up General Assembly resolution in 2002 called for further consideration and discussion of "information security."[182] The resolution also called for a new study of international informational security issues,[183] but little action resulted.[184] The United Nations sponsored a two-phase summit in 2003 and 2005 called the World Summit on the Information Society, but again with little concrete result.[185]

The United Nations did take a step forward in July 2010, when government cyber-security specialists from fifteen countries—including major

179. These resolutions have been based on the ongoing agenda item: "Developments in the field of information and telecommunications in the context of international security." *See, e.g.*, G.A. Res. 58/32, U.N. Doc. A/RES/58/32 (Dec. 8, 2003); G.A. Res. 59/61, U.N. Doc. A/RES/59/61 (Dec. 3, 2004); G.A. Res. 60/45, U.N. Doc. A/RES/60/45 (Jan. 6, 2006); G.A. Res. 61/54, U.N. Doc. A/RES/61/54 (Dec. 19, 2006); G.A. Res. 62/17, U.N. Doc. A/RES/62/17 (Jan. 8, 2008); G.A. Res. 63/37, U.N. Doc. A/RES/63/37 (Jan. 9, 2009); G.A. Res. 64/25, U.N. Doc. A/RES/64/25 (Jan. 14, 2010).

180. This is equally true of the General Assembly's two related resolutions on the Creation of a Global Culture of Cybersecurity and the Protection of Critical Informational Infrastructures, G.A. Res. 58/199, U.N. Doc. No. A/RES/58/199 (Jan. 30, 2004), and Creation of a Global Culture of Cybersecurity and Taking Stock of National Efforts to Protect Critical Information Infrastructures, G.A. Res. 64/211, U.N. Doc. No. A/RES/64/211 (Mar. 17, 2010).

181. G.A. Res. 54/49, at 2, U.N. Doc. A/RES/54/49 (Dec. 23, 1999).

182. *Id.* ¶ 1. The resolution called upon Member States to:

> promote further at multilateral levels the consideration of existing and potential threats in the field of information security, as well as possible measures to limit the threats emerging in this field . . . [and] . . . [i]nvite[ed] all Member States to continue to inform the Secretary-General of their views and assessments on the following questions:
>
> (a) General appreciation of the issues of information security;
>
> (b) Definition of basic notions related to information security, including unauthorized interference with or misuse of information and telecommunications systems and information resources . . . .

Id. ¶¶ 1-3.

183. *Id.* ¶ 4.

184. Similar exhortations appear in subsequent resolutions. *See* G.A. Res. 58/32, *supra* note 179, ¶ 4; G.A. Res. 59/61, *supra* note 179, ¶ 4; G.A. Res. 60/45, *supra* note 179, ¶ 4; G.A. Res. 61/54, *supra* note 179, ¶ 4; G.A. Res. 62/17, *supra* note 179, ¶ 4; G.A. Res. 63/37, *supra* note 179, ¶ 4; G.A. Res. 64/25, *supra* note 179, ¶ 4.

185. *See* WORLD SUMMIT ON THE INFORMATION SOCIETY: GENEVA 2003–TUNIS 2005, http://www.itu.int/wsis/index.html (last visited Apr. 21, 2012) (compiling conference documents and follow-up documents, including annual "outcome documents"); G.A. Res. 60/252, ¶ 11, U.N. Doc. A/RES/60/252 (Apr. 27, 2006) ("*Urges* Member States, relevant United Nations bodies and other intergovernmental organizations, as well as non-governmental organizations, civil society and the private sector, to contribute actively, inter alia by initiating actions, where appropriate, to the implementation and follow-up of the outcomes of the Geneva and Tunis phases of the Summit.").

cyber-powers like the United States, China, and Russia—submitted a set of recommendations to the U.N. Secretary-General as "an initial step towards building the international framework for security and stability that these new technologies require."[186] The recommendations called for

i. Further dialogue among States . . . ;
ii. Confidence-building, stability and risk reduction measures . . . including exchanges of national views on the use of [information and communication technologies] in conflict;
iii. Information exchanges on national legislation and national information and communications technologies security strategies and technologies, policies and best practices;
iv. Identification of measures to support capacity-building in less developed countries;
v. Finding possibilities to elaborate common terms and definitions . . . .[187]

Though vague, these recommendations represent real progress in overcoming a long impasse between the United States and Russia over how to address cyber-security issues.[188] The cooperation may even suggest possibilities for a future multilateral treaty under the auspices of the United Nations, which Russia has been advocating for some time.[189] At present, however, the role of the United Nations with respect to cyber-security remains largely limited to discussions and information sharing.

2. NATO

NATO recently began to address the threat of cyber-attacks. NATO did little in response to the 2007 cyber-attack on Estonia, laying bare that it "lacked both coherent cyber doctrine and comprehensive cyber strategy."[190] On the heels of that attack,[191] NATO held its first meeting—the 2008 Bucharest

186. U.N. Secretary-General, *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, 4, U.N. Doc. A/65/201 (July 30, 2010).

187. *Id.* at 8.

188. Historically Russia and the United States have expressed conflicting views on cyber-security as it relates to sovereignty and political dissent as well as international cooperation. *See, e.g.*, TIM MAURER, CYBER NORM EMERGENCE AT THE UNITED NATIONS: AN ANALYSIS OF THE ACTIVITIES AT THE UN REGARDING CYBER-SECURITY 1, 17, 25, 27, 47 (2011) (describing the contrasting views of the two countries).

189. John Markoff, *Step Taken to End Impasse over Cybersecurity Talks*, N.Y. TIMES, July 17, 2010, at A7.

190. Rex B. Hughes, *NATO and Cyber Defence: Mission Accomplished?*, ATLANTISCH PERSPECTIEF (Apr. 2009), *available at* http://www.atlcom.nl/site/english/nieuws/wp-content/Hughes.pdf.

191. This followed an October 2007 meeting of NATO defense ministers during which they called for the development of a NATO cyber-defense policy. *NATO Opens New Centre of Excellence on Cyber Defence*, N. ATL. TREATY ORG. NEWS (May 14, 2008), http://www.nato.int/docu/update/2008/05-may/e0514a.html.

Summit—to formally address cyber-attacks. This summit prompted the creation of two new NATO divisions focused on cyber-attacks: the Cyber Defence Management Authority and the Cooperative Cyber Defence Centre of Excellence.[192]

The Cyber Defence Management Authority aims to centralize cyber-defense capabilities across NATO members. Although little information is publicly available, the Authority is believed to possess "real-time electronic monitoring capabilities for pinpointing threats and sharing critical cyber intelligence in real-time," with the goal of eventually becoming an operational war room for cyber-defense.[193]

The Cooperative Cyber Defence Centre of Excellence aspires to "advance the development of long-term NATO cyber defence doctrine and strategy."[194] The North Atlantic Council, however, retains control of NATO cyber-policy and defense.[195] Despite strong pressure from Eastern European countries, cyber-attacks are still considered to activate only Article 4 of the NATO treaty, which calls upon members to "consult together" in cases of cyber-attacks, but does not bind them to "assist" each other, as would be required under Article 5.[196]

Although NATO's creation of these two divisions signifies concrete progress and recognition of the need for a more coherent cyber-strategy, concerns persist that "these teeth may not be sufficiently sharp to ward off any mischievous cyber bears or other e-adversaries seeking to compromise or destroy NATO digital assets deployed in either the Euro-Atlantic community or the 'near abroad.'"[197] NATO's cyber-plans and capabilities are still nascent.

3. Council of Europe

The Council of Europe[198] has taken the most direct and concrete approach to regulating a subset of the cyber-security problem—in particular, cyber-crime—of any international organization to date. As the first international treaty on crimes committed using the Internet and other computer networks, the

192. Hughes, *supra* note 190. This is NATO's tenth COE, and is the only one focused solely on defending against and countering cyber-attacks. *See* Scott J. Shackelford, *Estonia Three Years Later: A Progress Report on Combating Cyber Attacks*, J. INTERNET L., Feb. 2010, at 22.

193. Hughes, *supra* note 190.

194. *Id.*

195. *Defending The Networks: The NATO Policy on Cyber Defence*, N. ATL. TREATY ORG. (2011), http://www.nato.int/nato_static/assets/pdf/pdf_2011_09/20111004_110914-policy-cyberdefence.pdf ("The NATO Policy on Cyber Defence reiterates that any collective defence response is subject to decisions of the North Atlantic Council.").

196. North Atlantic Treaty, *supra* note 98, arts. 4, 5; *see also NATO Agrees Common Approach to Cyber Defence*, *supra* note 97 ("The competencies of the [Cyber Defence Management Authority] will fall exclusively on Article 4 of the North Atlantic Treaty.").

197. Hughes, *supra* note 190.

198. Founded in 1949, the Council of Europe aims to promote cooperation amongst its forty-seven European member states.

2001 Council of Europe Convention on Cybercrime ("Cybercrime Convention") promulgated "a common criminal policy aimed at the protection of society against cybercrime," primarily through legislation and international cooperation.[199] The United States ratified the Convention in 2006.[200]

Cyber-attacks implicate the Cybercrime Convention's offenses relating to "confidentiality, integrity, and availability of computer data and systems"—particularly illegal access, data interference, and system interference.[201] These rules, however, do not appear to apply to government actions, whether taken for law enforcement or national security purposes.[202] For example, Article 2 of the Convention requires that states adopt "legislative and other measures . . . to establish as criminal offenses under [their] domestic law, when committed intentionally, the access to the whole or any part of a computer system *without right*."[203] The Convention's accompanying "explanatory report" clarifies that the "without right" caveat allows for classic legal defenses, such as self-defense or necessity, but also "leaves unaffected conduct undertaken pursuant to lawful government authority"—including acts to "maintain public order, protect national security or investigate criminal offences."[204] This suggests, as Duncan Hollis and others have argued, that the Convention negotiators were aware of state interests in using cyber-attacks and sought to draft the agreement to permit such governmental action.[205]

Nonetheless, the Cybercrime Convention may still impose limited constraints on the execution of cyber-attack operations by ratifying countries. Parties to the Convention have agreed to "co-operate with each other . . . to the widest extent possible for the purposes of investigations or proceedings

199. Cybercrime Convention, *supra* note 64, pmbl.; *see also* Rasha AlMahroos, *Phishing for the Answer: Recent Developments in Combating Phishing*, 3 I/S: J. L. & POL'Y FOR INFO. SOC'Y 595, 613 (2008) ("The Council of Europe's Convention on Cybercrime . . . is the first and only international treaty that deals explicitly with cybercrime.").

200. The convention allows members of the Council of Europe and other states that participated in its elaboration (among them the United States) to join the Convention. Cybercrime Convention, *supra* note 66, at ch. IV; Declan McCullagh & Anne Broache, *Senate Ratifies Controversial Cybercrime Treaty*, CNET, (Aug. 4, 2006, 11:25 AM), http://news.cnet.com/Senate-ratifies-controversial-cybercrime-treaty/2100-7348_3-6102354.html. As of January 2012, thirty countries have ratified the Convention on Cybercrime, and another sixteen have signed but have not yet ratified it (including Australia, Japan, and South Africa). *Convention on Cybercrime*, TREATY OFFICE, COUNCIL OF EUROPE, http://conventions.coe.int/Treaty/Commun/ChercheSig.asp?NT=185&CM=8&DF=28/10/2010&CL=ENG (last visited Apr. 21, 2012).

201. Cybercrime Convention, *supra* note 66, arts. 2, 4, 5.

202. *See* Arie J. Schaap, *Cyber Warfare Operations: Development and Use Under International Law*, 64 A.F. L. REV. 121, 171 (2009) ("However, [the Cybercrime Convention's] rules do not apply to government activities, whether for law enforcement or national security purposes."); Hollis, *supra* note 10, at 1052 ("[The Cybercrime Convention's] rules, however, do not apply to government activities, whether for law enforcement or national security purposes.").

203. Cybercrime Convention, *supra* note 66, art. 2 (emphasis added).

204. Council of Eur., *Convention on Cybercrime: Explanatory Report*, 109th Sess., ¶ 38 (Nov. 8, 2001), *available at* http://conventions.coe.int/Treaty/en/Reports/Html/185.htm.

205. Hollis, *supra* note 10.

concerning criminal offences related to computer systems and data."[206] Although not explicit, this agreement to cooperate could limit the extent to which parties to the Convention could conduct cyber-attacks against other state parties, since that would undermine the overall intent of the agreement. It is unclear, however, what consequences or repercussions would result from such a breach of the Convention's intent and purpose by a state party.

For these reasons, the Convention—the most developed international legal framework directly regulating cyber-attacks—addresses only a portion of the overall challenge. It is limited, in particular, both by its failure to regulate most attacks by state parties and by its largely regional membership. Yet it offers a starting point for designing a comprehensive international framework for regulating unlawful cyber-attacks.

4. Organization of American States

The Organization of American States ("OAS"), representing thirty-five states from the Americas,[207] only recently began taking preliminary action to regulate cyber-attacks. In April 2004, the OAS approved a resolution stating that member states should "evaluate the advisability of implementing the principles of the Council of Europe Convention on Cybercrime (2001)" and should "consider the possibility of acceding to that convention."[208] The OAS also adopted a "Comprehensive Inter-American Cybersecurity Strategy," which aims, among other things, to adopt "cybercrime policies and legislation that will protect Internet users and prevent and deter criminal misuse of computers and computer networks, while respecting the privacy and individual rights of Internet users."[209] To this end, the OAS agreed to deploy an Experts Group that will "provide technical assistance to Member States in drafting and enacting laws that punish cybercrime, protect information systems, and prevent the use of computers to facilitate illegal activity."[210] These experts only offer guidance; the OAS is not promulgating a set of uniform laws with which member states can combat cyber-crime and cyber-attacks.

At a January 2010 meeting, the OAS Working Group on Cyber-Crime recommended that members that had not already done so establish state bodies for investigating and prosecuting cyber-crimes and adopt domestic legislation

206. Cybercrime Convention, *supra* note 66, art. 23.

207. The OAS aims for its member states to achieve "an order of peace and justice, to promote their solidarity, to strengthen their collaboration, and to defend their sovereignty, their territorial integrity, and their independence." Charter of the Organization of American States art. 1, *available at* http://www.oas.org/dil/treaties_A-41_Charter_of_the_Organization_of_American_States.htm.

208. Organization of American States, AG/RES. 2040 (XXXIV-O/04), at ch. IV, ¶ 8 (June 8, 2004), *available at* http://www.oas.org/juridico/english/ga04/agres_2040.htm.

209. Organization of American States, AG/RES. 2004 (XXXIV-O/04), at app. A, (June 8, 2004), *available at* http://www.oas.org/XXXIVGA/english/docs/approved_documents/adoption_strategy_combat_threats_cybersecurity.htm.

210. *Id.*

criminalizing cyber-crime and enabling international cooperation to investigate and prosecute such crimes.[211] The Working Group pledged to review the progress made in implementing these measures at its next meeting.[212] The OAS has thus begun a useful regional conversation on joint strategies for battling the portion of cyber-attacks that constitute cyber-crime, but it has not yet developed a more active program for addressing cyber-attacks more generally.

5. Shanghai Cooperation Organization

The Shanghai Cooperation Organization, an intergovernmental mutual security organization founded in 2001 by China, Kazakhstan, Kyrgyzstan, Russia, Tajikistan, and Uzbekistan,[213] has taken significant preliminary steps toward cooperation in the cyber-security area. In its Yekaterinburg Declaration of June 16, 2009, "[t]he SCO member states stress[ed] the significance of the issue of ensuring international information security as one of the key elements of the common system of international security."[214] The Organization presents a possible center of gravity in international legal action on cyber-attacks. As explained above,[215] the Organization has thus far adopted an expansive understanding of cyber-attacks that includes the use of cyber-technology to undermine political stability. As such, it represents a model that is likely to be at odds with that of Western Europe and the United States, which have sought to avoid regulations of cyber-activities that may interfere with the expression of political dissent.[216]

As this Section demonstrates, international efforts to regulate cyber-attacks are still at an embryonic stage. With the possible exception of the Council of Europe's Convention on Cybercrime, most international agreements have not proceeded beyond the stage of discussing future strategies. Nonetheless, the widespread efforts demonstrate increasing interest in establishing a set of transnational regulations to address cyber-attacks. The diversity of approaches taken by these organizations also demonstrates that the central challenge—at least initially—will be defining the scope of the activity that should be addressed in an international agreement. Before we outline our recommendations for future efforts at directly regulating cyber-attacks, however, we first must complete the existing legal picture by outlining the

211. Organization of American States, Sixth Meeting of the Working Group on Cyber-Crime, Recommendations, Jan. 21–22, 2010, OEA/Ser.K/XXXIV, CIBER-VI/doc.4/10 rev. 1, ¶¶ 1–2, *available at* http://www.oas.org/juridico/english/cyb_VIrec_en.pdf.

212. *Id.* ¶ 17.

213. These six countries are the only members of the SCO, though others are able to participate as observer states, dialogue partners, and guest attendees. More information on the SCO can be found here: http://www.fmprc.gov.cn/eng/topics/sco/t57970.htm.

214. CONSULATE GEN. OF UZB. IN N.Y.C., YEKATERINBURG DECLARATION OF THE HEADS OF THE MEMBER STATES OF THE SHANGHAI COOPERATION ORGANISATION, (July 9, 2009), http://www.uzbekconsulny.org/news/572/.

215. *See supra* text accompanying notes 24–27.

216. MAURER, *supra* note 188.

international regimes that indirectly regulate cyber-attacks as well as the domestic laws that address cyber-attacks.

C. International Legal Regimes That Indirectly Regulate Cyber-Attacks

Several international legal frameworks are not directly aimed at cyber-attacks but nonetheless regulate means that may be used in or may be a focus of a cyber-attack. These include, most notably, the international law governing telecommunications, aviation, space, and the law of sea.[217] These legal regimes were largely formed prior to the emergence of cyber-attacks and therefore do not expressly regulate or prohibit cyber-attacks. Instead, these "means-based" frameworks can be used to address a cyber-attack only if the attack employs the particular means regulated by the agreement.[218] Hence the international regimes that indirectly regulate cyber-attacks provide a patchwork of laws that are likely to apply to only a small portion of harmful cyber-attacks.

1. Telecommunications Law

Cyber-attacks that involve international wire or radio frequency communications may be subject to telecommunications law. Modern international telecommunications law is regulated by the International Telecommunications Union, the leading U.N. agency that establishes

217. While a number of countries have recognized Internet access as a human right, we do not discuss it here, due to its diffuse and currently unenforceable status. *See, e.g.*, David Meyer, *European 'Internet Freedom' Law Agreed*, ZDNET (Nov. 5, 2009, 1:11 PM), http://www.zdnet.co.uk/news/networking/2009/11/05/european-internet-freedom-law-agreed-39860587/. It therefore would not offer an alternate governing legal framework for cyber-attack with any practical significance. Moreover, the right to access the Internet does not implicate one of the key elements of our proposed cyber-attack definition: a national security or political purpose.

218. *See* Richard W. Aldrich, *The International Legal Implications of Information Warfare*, AIRPOWER J., Fall 1996, at 99, 109, *available* at http://www.au.af.mil/au/awc/awcgate/au/aldrich.pdf ("[M]ost of the law to which legal scholars are looking for guidance was developed, in many cases, decades before information warfare concepts were envisioned."); Barkham, *supra* note 122, at 95–96 (discussing existing treaty regimes that could be used to regulate information warfare); Dimitrios Delibasis, *State Use of Force in Cyberspace for Self-Defence: A New Challenge for a New Century*, PEACE, CONFLICT & DEV.: AN INTERDISC. J., Feb. 2006, at 1; BRYAN W. ELLIS, THE INTERNATIONAL LEGAL IMPLICATIONS AND LIMITATIONS OF INFORMATION WARFARE: WHAT ARE OUR OPTIONS? 3–4 (Apr. 10, 2001) (USAWC Strategy Research Project) (explaining how a network attack may implicate existing international telecommunications law); Schaap, *supra* note 202, at 160–70 (discussing other treaties and conventions that could impact cyber warfare operations, including the International Outer Space Law, International Telecommunications Law, and International Aviation Law); Scott J. Shackelford, *From Nuclear War to Net War: Analogizing Cyber Attacks in International Law*, 27 BERKELEY J. INT'L L. 191, 250–51 (2009) ("To the extent that cyber attacks are below the threshold of an armed attack, provisions of space law, nuclear non-proliferation, UNCLOS, and communications law, all have a role to play in crafting a functioning legal regime."); David Willson, *A Global Problem: Cyberspace Threats Demand an International Approach*, ISSA J., Aug. 2009, at 12, *available* at http://www.issa.org/Library/Journals/2009/August/Willson-A%20Global%20Problem.pdf; William Yurcik, *Information Warfare: Legal & Ethical Challenges of the Next Global Battleground*, *in* PROCEEDINGS OF THE SECOND ANNUAL ETHICS AND TECHNOLOGY CONFERENCE 1 (1997).

multinational standards for information and communication technology.[219] The Union's stated aim is "the preservation of peace and the economic and social development of all States . . . by means of efficient telecommunications services."[220] The International Telecommunications Union enacts rules known as Administrative Regulations, which are treaties that bind all member parties; Radio Regulations, which also bind all parties; as well as nonbinding Telecommunications Standards.[221] The Union regulates the use of radio and telecommunication technologies in order to distribute them to member states in an efficient and equitable manner—for example, through developing methods of assigning rights to radio spectrums.[222]

International Telecommunication regulations may be used to address cyber-attacks that make use of electromagnetic spectrum or international telecommunications networks. For instance, broadcasting stations from one nation may not interfere with broadcasts of other states' services on their authorized frequencies.[223] Member states may cut off any nonstate private telecommunications that "may appear dangerous to the security of the State or contrary to its laws, to public order or to decency"[224] or suspend international telecommunication services "either generally or only for certain relations and/or for certain kinds of correspondence, outgoing, incoming or in transit, provided that it immediately notifies such action to each of the other Member States through the Secretary-General."[225] Member states also must regulate against "harmful interference"[226] that "endangers the functioning of a radionavigation service or of other safety services or seriously degrades, obstructs or repeatedly interrupts a radiocommunication service"[227] and pursue all possible measures to ensure the secrecy of international correspondence,

219. CHARLES H. KENNEDY & M. VERONICA PASTOR, AN INTRODUCTION TO INTERNATIONAL TELECOMMUNICATIONS LAW 30–33 (1996). The International Telecommunications Convention is the founding charter that established the ITU. The ITU first began in 1865 as the International Telegraph Union and was founded in order to universalize telegraph services among mostly European nations. *Id.* at 30–32. It is based in Geneva, Switzerland, and its membership includes 193 member states and more than seven hundred sector members and associates. *About ITU*, INT'L COMM. UNION, http://www.itu.int/net/about/index.aspx (last visited Apr. 21, 2012). The full text of the Convention is available at *Basic Texts of ITU*, INT'L COMM. UNION, http://www.itu.int/net/about/basic-texts/index.aspx (last visited Apr. 21, 2012).

220. Constitution of the International Telecommunications Union, pmbl., Dec. 22, 1992, *available at* http://itu.int/net/about/basic-texts/index.aspx [hereinafter ITU Constitution]; *see also* International Telecommunications Convention, Nov. 6, 1982, U.N. Doc. 26559.

221. KENNEDY & PASTOR, *supra* note 219, at 33.

222. More information about the agency's work is available at *Committed to Connecting the World*, INT'L COMM. UNION, http://www.itu.int/en/pages/default.aspx (last visited Apr. 21, 2012); *see also The ITU Mission: Bringing the Benefits of ICT to All the World's Inhabitants*, INT'L COMM. UNION, http://www.itu.int/net/about/mission.aspx (last visited Apr. 21, 2012).

223. ITU Constitution, *supra* note 220, art. 45.

224. *Id.* art. 34.

225. *Id.* art. 35.

226. *Id.* art. 6.

227. *Id.* annex. (definition of "harmful interference").

unless such secrecy would contravene their domestic laws or international conventions.[228]

Despite the above restrictions, international telecommunications law does not specifically prohibit the use of telecommunications for military purposes, such as cyber-attacks. Article 48 states that "Member States retain their entire freedom with regard to military radio installations."[229] The article requests that states limit such use: "Nevertheless, these installations must, so far as possible, observe . . . the measures to be taken to prevent harmful interference."[230] The International Telecommunications Union cautions against "harmful interference," but it allows for military transgressions of these regulations—without requiring a reporting mechanism or otherwise limiting its use. This exception might include within its scope cyber-attacks and possibly even cyber-warfare. In addition to this military exception, the International Telecommunication Union provisions have a second important limitation as a legal framework for regulating cyber-attacks: violations of Union rules and regulations have only limited repercussions, given that the Union possesses little enforcement or punitive capacity.[231]

2. Aviation Law

Cyber-attack operations that target or interfere with nonmilitary aviation could implicate three major aviation regulations: the 1944 Chicago Convention on International Civil Aviation (Chicago Convention),[232] the 1971 Montreal Convention for the Suppression of Unlawful Acts Against Civil Aviation

228. *Id.* art. 37.

229. *Id.* art. 48(1).

230. *Id.* art. 48(2).

231. The International Telecommunication Union's main "regulatory" body originally was the International Frequency Registration Board (IFRB), which was formed "to manage the [radio frequency] spectrum internationally and to solve arising problems in a neutral manner." Wladyslaw Moron, *Radio Regulations Board (RRB): 'Its Place, Role and Functioning in the ITU*,*'* INT'L TELECOMM. UNION (Mar. 1, 2010), http://www.itu.int/ITU-R/information/promotion/e-flash/4/article7.html (last visited Apr. 21, 2012). Its founders envisioned it as a "cross between the Federal Communication Commission and the International Court of Justice." *Id.* (internal quotation marks omitted). This board, however, was never empowered to uphold its adjudicatory visions. *Id.* In 1994, the Radio Regulations Board subsumed the IFRB, aiming to act as an "independent interpreter and mediator" when dealing with noncompliance and sometimes conflicting interests of member states. *Id.* Even the Board, however, does not have full regulatory authority, since it can only issue recommendations when cases of "harmful interference" arise. *The International Telecommunication Union (ITU): Structure*, ENCYCLOPEDIA OF THE NATIONS, http://www.nationsencyclopedia.com/United-Nations-Related-Agencies/The-International-Telecommunication-Union-ITU-STRUCTURE.html#b (last visited Apr. 21, 2012). Furthermore, ITU resolutions are not considered legally binding. *See* STEPHEN GOROVE, DEVELOPMENTS IN SPACE LAW: ISSUES AND POLICIES 49 (1991) ("While states generally abide by ITU resolutions, they are not legally bound by them.").

232. Convention on International Civil Aviation, Dec. 7, 1944, 61 Stat. 1180 [hereinafter Chicago Convention].

(Montreal Convention),[233] and the 1988 Protocol for the Suppression of Unlawful Acts of Violence at Airports Serving International Civil Aviation (Montreal Protocol).[234] For example, the disruption of air traffic control, the modification of flight passenger lists, or the addition of a name to a country's no-fly list all exemplify cyber-attacks that implicate aviation law.[235]

The 1944 Chicago Convention created a specialized U.N. agency tasked with coordinating and regulating international air travel.[236] It also established a set of rules on airspace, aircraft, navigation, registration, and safety.[237] The Convention stipulates that all states must show "due regard for the safety of navigation of civil aircraft."[238] Cyber-attack operations that target civilian flights, if launched by a government against another actor, could run counter to this Convention's safeguard against interference with civilian flights. Such an operation would also run afoul of the 1984 amendment against using weapons targeting a civil aircraft in flight.[239] However, the Convention does allow a member state to derogate from the Convention's obligations during war or state emergencies,[240] so long as the state "notifies the fact to the Council."[241]

The Montreal Convention outlines as unlawful specific conduct that could jeopardize the safety of civil aviation.[242] Article 1 states that a person commits a crime if he or she intentionally and unlawfully does or attempts to do a series of acts that would render an aircraft incapable of flight or would seriously endanger the safety of the aircraft while in flight, including through "destroy[ing] or damag[ing] air navigation facilities or interfer[ing] with their operation, . . . or communicat[ing] information which he [or she] knows to be false, thereby endangering the safety of an aircraft in flight."[243] This agreement would not seem to restrict any cyber-attack operations unless it rendered an aircraft unable to fly (for example, by interfering with the aircraft's operating system) or endangered the safety of an aircraft in flight (for example, interfering with air traffic control communication or other aspects of aircraft navigation).

---

233. Convention for the Suppression of Unlawful Acts Against the Safety of Civil Aviation, Sept. 23, 1971, 24 U.S.T. 564 [hereinafter Montreal Convention].

234. Protocol for the Suppression of Unlawful Acts of Violence at Airports Serving International Civil Aviation, Feb. 24, 1988, 1589 U.N.T.S. 474 [hereinafter Montreal Protocol].

235. Schaap, *supra* note 202, at 166.

236. Chicago Convention, *supra* note 232, arts. 43, 44. The agency is called the International Civil Aviation Organization. *Id.*

237. *Id.* pt. I.

238. *Id.* art. 3(d).

239. This 1984 amendment to the Chicago Convention "reaffirm[s] the principle of non-use of weapons against civil aircraft in flight." Protocol Relating to an Amendment to the Convention on International Civil Aviation, pmbl., May 10, 1984, 23 I.L.M. 705.

240. Chicago Convention, *supra* note 232, art. 89.

241. *Id.*

242. Montreal Convention, *supra* note 233.

243. *Id.* art. 1.

The Montreal Protocol extended the legal framework from civil aircraft in flight to "acts of violence which endanger or are likely to endanger the safety of persons at airports . . . or which jeopardize the safe operation of such airports."[244] Article 2 states that a person commits a crime if he or she intentionally and unlawfully does or attempts to do any of the following while using a device, substance, or weapon:

> (a) performs an act of violence against a person at an airport serving international civil aviation which causes or is likely to cause serious injury or death; or
>
> (b) destroys or seriously damages the facilities of an airport serving international civil aviation or aircraft not in service located thereon or disrupts the services of the airport,
>
> if such an act endangers or is likely to endanger safety at that airport.[245]

This Protocol thereby prohibits any cyber-attacks that could undermine safety at an international airport, such as tampering with no-fly lists, passenger manifests, or an airport's computer network system.

3. Law of Space

Given that computer-operated satellites are integral to international telecommunications and military operations, cyber-attacks could implicate space law. Multiple scholars have proposed that treaties on outer space, the moon, and damage caused by space objects, as well as satellite regulations, could be used to regulate cyber-attacks.[246] Treaties related to the damage caused by space objects[247] or the moon[248] are clearly inapplicable to cyber-

244. Montreal Protocol, *supra* note 234, pmbl.

245. *Id.* art. 2.

246. Aldrich, *supra* note 218, at 20–24; Delibasis, *supra* note 218, at 15–17 (discussing how the law of space is applicable to cyber-warfare); LAWRENCE T. GREENBERG ET AL., INFORMATION WARFARE AND INTERNATIONAL LAW 8–9 (1998) ("Space law, though, leaves ample room for information warfare."); Hollis, *supra* note 10, at 1051 ("[B]ecause information infrastructures frequently use outer space to relay communications or collect data, space law may affect [information operations]."); Schaap, *supra* note 202, at 160–69 (discussing international outer space law, international telecommunications law, and international aviation law as legal regimes that states should consider in developing cyber-warfare operations).

247. The Convention on International Liability for Damage Caused by Space Objects lays out a set of procedures for determining state liability for activities in outer space. Article 2 states that "[a] launching State shall be absolutely liable to pay compensation for damage caused by its space object on the surface of the Earth or to aircraft in flight." Convention on International Liability for Damage Caused by Space Objects, art 2, Mar. 29, 1972, 24 U.S.T. 2389, 961 U.N.T.S. 187. The Convention defines damage as "loss of life, personal injury or other impairment of health; or loss of or damage to property of States or of persons, natural or juridical, or property of international intergovernmental organizations." *Id.* art 1. It is unlikely, however, that the definition of damage or of space object would apply to cyber-attacks.

248. The Moon Treaty grants the international community jurisdiction over all heavenly bodies, including the orbits around such bodies. Agreement Governing the Activities of States in Outer Space, on the Moon and Other Celestial Bodies, Dec. 5, 1979, 1363 U.N.T.S. 53. The treaty refers to

attacks as we have defined them, and therefore we do not discuss them here. Instead, we focus on satellite regulations and the Treaty on Principles Governing the Activities in the Exploitation and Use of Outer Space. We conclude, however, that these treaties also have little promise for the regulation of cyber-attacks.

The 1967 Outer Space Treaty provides for the free exploration of space and prohibits the use of space for particular destructive purposes.[249] It stipulates that

> States Parties to the Treaty undertake not to place in orbit around the Earth any objects carrying nuclear weapons or any other kinds of weapons of mass destruction, install such weapons on celestial bodies, or station such weapons in outer space in any other manner.
>
> The moon and other celestial bodies shall be used by all States Parties to the Treaty exclusively for peaceful purposes.[250]

The Outer Space Treaty expressly permits certain military uses of space, such as earth orbit military reconnaissance satellites, remote sensing satellites, military global positioning systems, and space-based aspects of an antiballistic missile system.[251] Because cyber-attacks are unlikely to cause mass destruction of the kind contemplated in the treaty, it is unlikely that cyber-attacks could be properly characterized as prohibited by the treaty.[252]

Satellite regulations offer another potential avenue for cyber-attack regulation. The Agreement Relating to the 1971 International Telecommunications Satellite Organization ("Telecommunications Satellite Organization")[253] and the Convention of the 1979 International Maritime Satellite Organization ("Maritime Satellite Organization")[254] contain "peaceful purpose" provisions applicable to satellites similar to the Outer Space Treaty.

---

the "common heritage of mankind," reflecting a belief that all nations should share equitably in benefits derived from resources on the moon and other celestial bodies. *Id.* art. 11(1). The treaty also underscores that the moon should be used exclusively for "peaceful purposes." *Id.* art. 3. Beyond this principle, however, the treaty offers little concrete means by which cyber-warfare could be regulated. Furthermore, the countries and organizations mainly engaged in space exploration, such as the United States, the European Union, Russia, China, and Japan, have not ratified the treaty. As of January 1, 2011, only thirteen states had ratified and four signed the Moon Treaty. U.N. Office for Outer Space Affairs, Comm. on the Peaceful Uses of Outer Space, *Status of International Agreements Relating to Activities in Outer Space*, U.N. Doc. ST/SPACE/11/Rev.2/Add/3 (2011).

249. Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, Including the Moon and Other Celestial Bodies, Jan. 27, 1967, 18 U.S.T. 2410, 610 U.N.T.S. 205.

250. *Id.* art. 4.

251. Shackelford, *supra* note 218, at 219.

252. Celestial bodies "refers only to natural bodies, such as the moon, asteroids, and planets, not to man-made satellites," the main means in outer space by which cyber-warfare could be conducted. Aldrich, *supra* note 218, at 20.

253. Agreement Relating to the International Telecommunications Satellite Organization, "INTELSAT," Aug. 20, 1971, 23 U.S.T. 3813 [hereinafter Telecommunications Satellite Agreement].

254. Convention of the International Maritime Satellite Organization London, Sept. 3, 1976, 31 U.S.T. 1 [hereinafter INMARSAT].

However, despite the fact that satellites are likely to have a role in cyber-attacks, these treaties have little utility in regulating attacks. The Telecommunications Satellite Organization, which initially formed as an intergovernmental body mandated to "carry forward . . . the design, development, construction, establishment, operation and maintenance of the space segment of the global commercial telecommunications satellite system,"[255] was privatized in 2000.[256] Similarly, the Maritime Satellite Organization has largely ceased to represent intergovernmental interests.[257] Consequently, neither organization is well situated to promulgate public regulations related to cyber-attacks.

4. Law of the Sea

The 1982 United Nations Convention on the Law of the Sea ("UNCLOS")—particularly articles 19, 109, and 113—tangentially implicates cyber-attack operations at sea.[258] The article 19 obligation, which allows a vessel the right of innocent passage through another nation's territorial sea so long as its activities are not "prejudicial to the peace, good order or security of the coastal State,"[259] is widely accepted as customary international law.[260] Activities prohibited by article 19 include

255. Telecommunications Satellite Agreement, *supra* note 253, art. 2.

256. To "promote a more competitive global satellite services market," the Telecommunications Satellite Organization became a private company in 2000 named "Intelsat." U.S. GOV'T ACCOUNTABILITY OFFICE, TELECOMMUNICATIONS: INTELSAT PRIVATIZATION AND THE IMPLEMENTATION OF THE ORBIT ACT 1 (2004).

257. The Maritime Satellite Organization, originally founded as a nonprofit international organization to establish a maritime satellite communications network, changed its name to "International Mobile Satellite Organization" when it began to provide services to aircraft and portable users. JONATHAN HIGGINS, SATELLITE NEWSGATHERING 247–48 (2d ed. 2007). In 1999, the organization divided into two separate parts: most converted into a commercial company, and a small group became the intergovernmental regulatory body, the International Mobile Satellite Organization (IMSO). *Id.* at 248. Through a private-public partnership, the IMSO oversees certain public satellite safety and security communication services provided by Inmarsat satellites.

258. United Nations Convention on the Law of the Sea, art. 19, Dec. 10, 1982, 1833 U.N.T.S. 3 [hereinafter UNCLOS]. The United States has not ratified the Convention on the Law of the Sea, even though it has been abiding by the Convention since President Regan's 1983 Statement of Oceans Policy, and it signed the 1994 Agreement Relating to Implementation of Part XI. Nonetheless, many of the provisions of the Convention are considered binding on the United States and other countries as customary international law. Div. for Ocean Affairs and the Law of the Sea, *Table Recapitulating the Status of the Convention and of the Related Agreements, as at 20 September 2011*, http://www.un.org/Depts/los/reference_files/status2010.pdf; Senator Richard G. Lugar, The Law of the Sea Convention: The Case for Senate Action, Address at the Brookings Institution (May 4, 2004), *available at* http://www.brookings.edu/speeches/2004/0504energy_lugar.aspx (discussing the United States abiding by the Law of the Sea Convention).

259. UNCLOS, *supra* note 258, art. 19(1).

260. RÜDIGER WOLFRUM, FREEDOM OF NAVIGATION: NEW CHALLENGES (2008), *available at* http://www.itlos.org/fileadmin/itlos/documents/statements_of_president/wolfrum/freedom_navigation_080108_eng.pdf.

> (a) any threat or use of force against the sovereignty, territorial integrity or political independence of the coastal State, or in any other manner in violation of the principles of international law embodied in the Charter of the United Nations;
>
> . . .
>
> (c) any act aimed at collecting information to the prejudice of the defence or security of the coastal State;
>
> (d) any act of propaganda aimed at affecting the defence or security of the coastal State;
>
> . . .
>
> (k) any act aimed at interfering with any systems of communication or any other facilities or installations of the coastal State . . . .[261]

These regulations, particularly part (k), could be read to prohibit cyber-attacks that make use of computer systems on vessels that are at sea.

Similarly, article 109 stipulates that all states should cooperate in suppressing unauthorized broadcasting from the high seas.[262] UNCLOS defines "unauthorized broadcasting" as "the transmission of sound radio or television broadcasts from a ship or installation on the high seas intended for reception by the general public contrary to international regulations, but excluding the transmission of distress calls."[263] The prohibition could extend to a cyber-attack that compromises the computer network that operates a ship's broadcast system.[264] Similarly, article 113 requires states to put in place "laws and regulations necessary" to punish willful damage to submarine cables, including damage caused by a cyber-attack.[265] Thus, by prohibiting actions that undermine the functioning of communications systems at sea, these provisions provide some minimal legal protections against cyber-attacks that occur on or originate from the high seas.

Together, international law governing telecommunications, aviation, space, and the sea provide potentially effective tools for addressing forms of cyber-attack within specific contexts. Yet this patchwork of regulations fails to provide a complete or coherent mechanism for addressing all forms of cyber-attacks. Given the limits of current international law, the following Section considers how U.S. domestic law might be used to address cyber-attacks.

---

261. UNCLOS, *supra* note 258, art. 19(2).

262. *Id.* art. 109(1).

263. *Id.* art. 109(2).

264. *Id.* art. 109(3). In particular, article 109(3) states that prosecution may occur in "the court of: (a) the flag State of the ship; (b) the State of registry of the installation; (c) the State of which the person is a national; (d) any State where the transmissions can be received; or (e) any State where authorized radio communication is suffering interference." *Id.*

265. *Id.* art. 113.

D. U.S. Domestic Law

Domestic law offers an important tool for combating cyber-attacks, including those that cross international borders. Because many cyber-attacks are also cyber-crimes,[266] domestic criminal law is particularly relevant. Unfortunately, existing domestic law largely fails to directly address the novel modern challenges posed by cyber-attacks,[267] and is severely limited by its lack of extraterritorial reach.

Although there is no U.S. federal statute that directly criminalizes cyber-attacks,[268] there is extensive federal criminal law that offers an important legal tool for addressing cyber-attacks.[269] At the federal level, criminal laws address fraud involving devices, computers, or email;[270] malicious interference in communications lines, stations, or systems;[271] electronic communication interception;[272] illicit access to electronic communications and records;[273] and recording of dialing, routing, addressing, and signaling information.[274]

The majority of existing criminal laws bearing on cyber-attack do not apply extraterritorially—that is, they do not reach criminal activity occurring outside the United States.[275] There are, however, some exceptions to that

---

266. *See supra* Part I.A.3 and Figure 1.

267. *See, e.g.*, Sklerov, *supra* note 93, at 6 ("Unfortunately, state responses to cyberattacks are governed by an anachronistic legal regime that impairs a state's ability to defend itself.").

268. As this Article went to print, several cyber-security bills had been proposed but none passed. *See* Brendan Sasso, *Senate Dems Modifying Cybersecurity Bill to Pick Up GOP Votes*, HILLICON VALLEY (May 6, 2012), http://thehill.com/blogs/hillicon-valley/technology/225607-senate-dems-revamping-cybersecurity-bill-; Ellen Nakashima, *On Hill, Imagining a Cyberattack on New York*, WASH. POST CHECKPOINT WASHINGTON (Mar. 9, 2012), http://www.washingtonpost.com/blogs/checkpoint-washington/post/officials-use-nyc-blackout-scenario-to-sell-senators-on-cyber-attack-legislation/2012/03/09/gIQA9Z530R_blog.html.

269. In addition to criminal liability, there have been proposals for the use of tort law against cyber-attackers or intermediaries who negligently facilitate cyber-attack. *See, e.g.*, Michael D. Scott, *Tort Liability for Vendors of Insecure Software: Has the Time Finally Come?*, 67 MD. L. REV. 425 (2008); Jay P. Kesan & Carol M. Hayes, Mitigative Counterstriking: Self-Defense and Deterrence in Cyberspace 1, 31–32 (2011) (unpublished research paper) (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1979857). Such proposals face a number of serious challenges, however, including attribution and jurisdictional problems, *id.* at 30, and, for intermediaries, causation problems and a virtual "tax on technophobia, punishing those who do not know enough about protecting their personal computers," *id.* at 32. Moreover, if software designers were held liable for leaving their products vulnerable to cyber-attack, software costs could increase substantially. *Id.*

270. 18 U.S.C. §§ 1029, 1030, 1037 (2006). 18 U.S.C. § 1030 is the codification of the Computer Fraud and Abuse Act.

271. *Id.* § 1362.

272. *Id.* §§ 2510–2522.

273. *Id.* §§ 2701–2712.

274. *Id.* §§ 3121–3127.

275. There is generally a presumption against extraterritorial application of federal law. *See* United States v. Cotten, 471 F.2d 744, 750 (9th Cir. 1973). Nevertheless, "Congress has the authority to enforce its laws beyond the territorial boundaries of the United States," and may do so by evidence of its intent as gauged through statutory interpretation. EEOC v. Arabian Am. Oil Co., 499 U.S. 244, 248 (1991) (internal citations omitted). In certain cases, extraterritorial reach may also be extended without explicit or implied congressional authorization based on detrimental effects in the United

general rule. For example, the criminal statute banning access device fraud, as amended by the USA PATRIOT Act of 2001, provides that

> [a]ny person who, outside the jurisdiction of the United States, engages in any act that, if committed within the jurisdiction of the United States, would constitute an offense under . . . this section, shall be subject to the fines, penalties, imprisonment, and forfeiture provided in this title if—
>
> 1. the offense involves an access device issued, owned, managed, or controlled by a[n] . . . entity within the jurisdiction of the United States; and
> 2. the person transports, delivers, conveys, transfers to or through, or otherwise stores, secrets, or holds within the jurisdiction of the United States, any article used to assist in the commission of the offense or the proceeds of such offense or property derived therefrom.[276]

The statute banning computer fraud was likewise amended as part of the USA PATRIOT Act to provide for extraterritorial applicability.[277] Both of these statutes may serve as useful models for extending extraterritorial application to other domestic laws related to cyber-attack.

In addition, several recent legislative efforts tackle pieces of the cyber-attack threat not addressed by U.S. criminal law. These include the Cybersecurity Enhancement Act;[278] the Executive Cyberspace Authorities Act of 2010;[279] the Rockefeller-Snowe Cybersecurity Act;[280] the International Cyberspace and Cybersecurity Coordination Act of 2010;[281] and the Protecting Cyberspace as a National Asset Act of 2010.[282]

The most widely discussed of these efforts has been the Protecting Cyberspace as a National Asset Act, cowritten by Senators Lieberman, Collins, and Carper, which was introduced in the Senate and the House in June 2010.[283] The bill builds on the military's recent establishment of the U.S. Cyber

---

States. *See* United States v. Muench, 694 F.2d 28, 33 (2d Cir. 1982) ("The intent to cause effects within the United States . . . makes it reasonable to apply to persons outside United States territory a statute which is not expressly extraterritorial in scope.").

276. 18 U.S.C. § 1029 (2006); U.S. DEP'T OF JUSTICE, PROSECUTING COMPUTER CRIMES, *supra* note 65, at 94, 115.

277. 18 U.S.C. § 1030 (2006) ("[T]he term 'protected computer' [to which this statute applies] means a computer . . . which is used in interstate or foreign commerce or communication, including a computer located outside the United States that is used in a manner that affects interstate or foreign commerce or communication of the United States."); U.S. DEP'T OF JUSTICE, PROSECUTING COMPUTER CRIMES, *supra* note 66, at 5.

278. H.R. 4061, 111th Cong. (2010).

279. H.R. 5247, 111th Cong. (2010).

280. S. 773, 111th Cong. (2009).

281. S. 3193, 111th Cong. (2010).

282. S. 3480, 111th Cong. (2010); H.R. 5548, 111th Cong. (2010).

283. S. 3480; H.R. 5548.

Command[284] by proposing the establishment of two new administrative bodies: (1) an Office of Cyberspace Policy in the White House, charged with developing and coordinating a national strategy to increase the security and resiliency of cyberspace; and (2) a National Center for Cybersecurity and Communications within the Department of Homeland Security, designed to "enable automated and continuous monitoring of any information collected" and "use [of] the information to enhance the risk-based security of the Federal information infrastructure."[285] The bill also addresses a wide range of related cyber-security matters, including definitions and federal information security management provisions.[286]

The bill sparked a vigorous debate over the proper role of the government in regulating cyberspace. Opponents dubbed the proposed regulation the "kill switch bill," seeing it as an effort to grant the president emergency powers over certain Internet communications.[287] However, had it passed into law, the bill would likely have established more checks on the president's power to respond to cyber-emergencies than currently exist.[288] Its authors amended and reintroduced the bill, but it has made little progress toward a vote on the Senate floor.[289] It has since been superseded by alternative proposals, none of which has yet won the approval of Congress.[290]

This debate offers an important lesson for advocates of cyber-attack regulation: any future law must clearly indicate what activities are to be covered, place a high and transparent bar on emergency measures, and address well-founded concerns that efforts to strengthen cyber-security might simultaneously weaken free and open access to modern technology for those engaging in political speech and organizing.

Other domestic legal efforts to address cyber-attacks are either based in criminal law or have focused on developing U.S. defensive capabilities. However, none of the recent legislative efforts that might strengthen defensive capacity against cyber-attack have yet been made into law. Moreover, the existing domestic law framework is insufficient for addressing the larger global

284. McMichael, *supra* note 39.

285. S. 3480; H.R. 5548.

286. S. 3480; H.R. 5548.

287. *See* Emelie Rutherford, *Senate Committee OKs Cybersecurity Bill on Majority Leader's Radar*, DEFENSE DAILY (June 25, 2010), http://findarticles.com/p/articles/mi_6712/is_61_246/ai_n54561980/ (last visited Apr. 22, 2012). The bill has since been reintroduced with changes meant to prevent the government from using a "kill switch" to shut off Internet service as a political tool. *Id.*; *see also* Diane Bartz, *Reid Pushes US Republicans for Cybersecurity Bill*, REUTERS (July 27, 2011, 5:09 PM), http://www.reuters.com/article/2011/07/27/congress-cybersecurity-idUSN1E76Q1M320110727.

288. *See, e.g.*, Rutherford, *supra* note 287 (describing congressional "frustration . . . that people have misconstrued a provision related to the president's emergency powers to take over communications networks" when "[t]he president already has this authority, . . . and the bill would restrict when he can use it").

289. *See id.*; Bartz, *supra* note 287.

290. *See supra* note 268.

problem.[291] In particular, the lack of extraterritorial reach of most criminal laws that apply to cyber-attacks severely limits their ability to reach those initiating such cyber-attacks, who are often located outside the United States. The next Part offers recommendations for remedying the substantial limitations of both the domestic and international legal frameworks for addressing cyber-attack.

IV.
NEW LAW FOR CYBER-ATTACKS

Cyber-attacks present a new and growing threat—one that current international and domestic laws are not yet fully prepared to meet. The law of war offers a basis for responding only to those cyber-attacks that amount to an armed attack or that take place in the context of an ongoing armed conflict. Other existing international legal frameworks offer only embryonic or piecemeal protection. U.S. domestic law, though potentially a powerful tool for battling cyber-attacks, has not yet addressed the challenge directly, and what remedies exist are in many cases restricted by jurisdictional limits.

To begin to fill the gaps in existing law, we propose legal reform on both domestic and international levels.[292] Our recommended domestic law reforms are twofold. First, the United States should add extraterritorial applicability to criminal laws bearing on cyber-attack. Second, the United States should utilize limited countermeasures, as appropriate, to combat cyber-attacks that do not rise to the level of armed attacks under the law of war.

These domestic measures will address elements of the problem, but getting at the root of the global cyber-attack challenge will require international solutions. We therefore recommend an international cyber-treaty with two central aims. First, such an agreement should provide a definition of cyber-attacks and cyber-warfare that limits the cyber-attacks to which states may respond with force. Second, the treaty should empower states to cooperate in evidence collection and criminal prosecution of individuals involved in transnational cyber-attacks. While this second aim will likely be a longer-term project, it offers the only truly effective solution to the inherently international problem of cyber-attacks.

---

291. *See* Andy Johnson & Kyle Spector, *Deterring Cyber War: A U.S.-Led Cybersecurity Summit*, THIRD WAY 3 (Oct. 2010), *available at* http://content.thirdway.org/publications/343/Third_Way_Idea_Brief_-_Deterring_Cyber_War-A_US-Led_Cybersecurity_Summit.pdf (last visited Apr. 22, 2012).

292. We focus here on potential legal reforms. In addition to legal reform, government should coordinate with the private sector to address cyber-attack threats. Indeed, the Obama administration has recognized that "[e]nsuring the resilience of our networks and information systems requires collective and concerted national action that spans the whole of government, in collaboration with the private sector and individual citizens." WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 13. The U.S. Department of Defense has also suggested that there may be a need for "incentives or other measures . . . to promote private sector participation." DOD STRATEGY, *supra* note 32, at 9. The legal reforms outlined here are meant to compliment such cooperative measures, not substitute for them.

A. Battling Cyber-Attacks at Home

1. Extend the Extraterritorial Reach of Domestic Law

As noted above, a number of existing and proposed domestic laws may play a role in combating cyber-attacks, including numerous criminal statutes regulating harmful cyber-activity outside the context of armed conflict.[293] It is important to recall that domestic criminal law alone cannot regulate cyber-attacks because not all cyber-attacks are defined as cyber-crimes. But many cyber-attacks—including those involving non–state actors and computer-based means—are also cyber-crimes that fall within the ambit of domestic criminal law.[294] Unfortunately, only a small number of existing criminal laws that might govern cyber-attacks explicitly provide for extraterritorial reach.[295]

To remedy this limitation, legislators could amend domestic criminal statutes to give them extraterritorial reach. If other states reciprocate by making their own criminal statutes pertaining to cyber-attacks extraterritorial as well, this could greatly increase global enforcement.[296] Indeed, increased domestic enforcement through extraterritorial application will be much more successful and legitimate if it takes place in concert with the creation of an international treaty that establishes basic shared standards regarding cyber-attacks.

Even if domestic criminal laws that apply to cyber-attacks extend across borders, jurisdictional hurdles will likely hamper enforcement by any individual state. It may be difficult, for example, for the United States to gain custody of accused cyber-criminals operating abroad, particularly if they are not U.S. citizens or operate in countries that do not have extradition treaties with the United States. Thus, strengthened extradition relationships around the world would complement increased extraterritorial application of domestic law. Though dramatic improvement in extradition relationships may not be immediately feasible given that extradition treaties, which are negotiated on a bilateral basis, take substantial time and effort to negotiate and pass, such relationships could help effectuate the prosecution of many crimes resulting from increasing globalization including drug, weapon, and human trafficking, and transnational white-collar crime.[297] Thus, the United States should prioritize the development of these relationships moving forward.

Further, the United States, and the global community in general, should endeavor to explicitly criminalize aspects of cyber-attacks that fall outside the

---

293. *See supra* Part III.D.

294. *See supra* Part I.A.3.

295. *See supra* Part III.D.

296. This extraterritorial reach would not regulate cyber-actions taken by governments but rather those of individuals and other non–state actors.

297. *See generally* John T. Soma, et al., *Transnational Extradition for Computer Crimes: Are New Treaties and Laws Needed?*, 34 HARV. J. ON LEGIS. 317 (1997) (discussing the limitations of current extradition treaties and proposing potential reforms).

scope of existing domestic or international law, including the law of war. In the present absence of an international cyber-crime agreement, it is possible for the United States to more effectively counter cyber-attacks through carefully crafted and narrowly framed domestic law.

2. Countermeasures in Response to Cyber-Attacks

Although the international law of countermeasures has played a minimal role in legal debates around cyber-attacks thus far, it nonetheless offers an extremely useful legal framework for states seeking to respond to a cyber-attack. The United States and other countries interested in regulating cyber-attacks should begin to develop a policy defining the types of countermeasures legally and strategically appropriate for different types of cyber-attacks.

As noted in the discussion of *jus ad bellum* above, the vast majority of cyber-attacks do not rise to the level of an armed attack.[298] But armed self-defense is not the only manner in which states can respond to cyber-attacks. Provided that the initial cyber-attack violates an international obligation of the perpetrating state, the victim state is entitled under customary international law to employ necessary and proportional countermeasures designed to induce the perpetrating state to resume compliance with international norms and to stop conducting (or allowing) cyber-attacks from its territory.[299]

While active defense is the most commonly discussed type of countermeasure that might be employed in response to a cyber-attack, it is only one option among many.[300] A key limit to a legally permissible countermeasure is that it must be proportional to the injury suffered by the victim state.[301] Moreover, countermeasures must be designed to enable a return to the status quo ante, in which both the perpetrating and victim states comply with their relevant legal duties towards one another.[302] Countermeasures must be temporary so that once the cyber-attacks stop, the countermeasure may stop as well and normal international relations may resume.[303]

The Draft Articles on State Responsibility express a preference for reciprocal countermeasures, but this is not a requirement.[304] Still, the closer the relationship between the breach and countermeasure the more likely the countermeasure is to be proportional and therefore lawful.[305] The United States should consider in advance what international obligations it has toward likely cyber-aggressor states that it might lawfully revoke in case of an unlawful

298. *See supra* Part II.A.
299. *See* Draft Articles, *supra* note 109, art. 49.
300. *See* Sklerov, *supra* note 93, at 2 n.5 (comparing active and passive defenses).
301. *See* Draft Articles, *supra* note 109, art. 51.
302. *See id.* art. 49(1).
303. *See id.* art. 49.
304. *See id.* pt. 3, ch. 2, cmt., ¶ 5.
305. *Id.*

cyber-attack. Indeed, the United States could develop a policy regarding the types of countermeasures legally available in response to particular types of cyber-attacks.

B. A Cyber-Attack Treaty

Changes in domestic law and policy, such as adding extraterritorial reach to criminal laws and planning for the use of countermeasures, are valuable legal responses to the threat of cyber-attack. Yet to truly address the cyber-attack challenge, international coordination will be necessary.[306] The scope of the problem is global, and the solution must be as well. As the U.S. Department of Defense has explained, "cyberspace is a network of networks that includes thousands of ISPs [internet service providers] across the globe; no single state or organization can maintain effective cyber defenses on its own."[307]

The United States has developed a Cyberspace Strategy that emphasizes working "with like-minded states to establish an environment of expectations, or norms of behavior, that ground foreign and defense polices and guide international partnerships."[308] While the development of international norms is useful, it will not provide governments and private actors with the clarity of a codified definition of cyber-attack or written guidelines on how states should respond to certain types of challenges. For this reason, we recommend that the international community create a multilateral agreement with two central features. First, it must offer a shared definition of cyber-crime, cyber-attack, and cyber-warfare.[309] Second, it should offer a framework for more robust international cooperation in information sharing, evidence collection, and criminal prosecution of those participating in cross-national cyber-attacks. That framework should be attentive to the challenges of overcriminalization, maintaining room for individuals to use the Internet and related technologies to engage in lawful dissent.[310]

---

306. As discussed in Part III.B, there have already been several efforts to create a cyber-attack treaty. *See* CLARKE & KNAKE, *supra* note 7, at 268–71 (arguing for a Cyber War Limitations Treaty); *cf.* JACK GOLDSMITH, CYBERSECURITY TREATIES: A SKEPTICAL VIEW (2011) (offering a skeptical take on the possibility of a cyber-security treaty). Russia has for some time been proposing a treaty banning cyber-attack. *See, e.g.*, John Markoff & Andrew E. Kramer, *U.S. and Russia Differ on Treaty for Cyberspace*, N.Y. TIMES, June 28, 2009, at A1 ("Russia favors an international treaty along the lines of those negotiated for chemical weapons and has pushed for that approach at a series of meetings . . . and in public statements . . . ."). Yet the shape of the agreement proposed here is quite different—it begins with securing a shared agreement on the activity meant to be prohibited.

307. DOD STRATEGY, *supra* note 32, at 9.

308. WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 9. The United States is currently prepared to build bilateral and multilateral partnerships, to work with regional organizations, and to collaborate with the private sector. *See id.* at 12.

309. It is worth noting again that cyber-attacks that do constitute use of force under the law of war are already covered by *jus in bello* principles, which may be more clearly defined over time in the cyber-attack context through state practice. *See also supra* Part II.B.

310. *See* WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 7.

1. Define Cyber-Attack and Cyber-Warfare

The first aim of a cyber-attack treaty regime should be to develop a shared definition of cyber-attack, cyber-crime, and cyber-warfare. These definitions could serve as the foundation for domestic criminal legislation targeting cyber-attacks and cyber-crime as well as more extensive international cooperation. A similar strategy has been used, for example, in the international effort to battle bribery: the OECD Bribery Convention provides a definition of bribery that state parties then integrate into national legislation forbidding the practice.[311] Under the Bribery Convention, "signatories pledged to criminalize and prosecute the bribery of foreign public officials."[312] The thirty-eight state parties have then used that shared definition as the basis for domestic implementing legislation.[313]

We have proposed a definition of cyber-attack that would include any action taken to undermine the function of a computer network for a political or national security purpose. An appropriate definition of cyber-crime would include any violation of criminal law by non–state actors, committed by means of a computer system. Finally, cyber-warfare should be defined as a cyber-attack that causes physical injury or property damage comparable to a conventional armed attack.

States could adopt a clear definition of cyber-attack, cyber-crime, and cyber-warfare in the context of a comprehensive binding treaty, nonbinding declaration, or through independent agreements in anticipation of more broad-based future cooperation. Even a stand-alone nonbinding defining declaration could provide an important starting point for future cooperation if it provides a definition that is later incorporated into a more comprehensive international treaty.[314] Such a document could offer much-needed clarity on when cyber-

311. Convention on Combating Bribery of Foreign Public Officials in International Business Transactions, Dec. 18, 1997, 37 I.L.M. 4 (1998) [hereinafter OECD Bribery Convention]. Under the Convention, the "Offense of Bribery of Foreign Public Officials" is defined as

> intentionally to offer, promise or give any undue pecuniary or other advantage, whether directly or through intermediaries, to a foreign public official, for that official or for a third party, in order that the official act or refrain from acting in relation to the performance of official duties, in order to obtain or retain business or other improper advantage in the conduct of international business.

Id. art. 1(1).

312. *Developments in the Law – Extraterritoriality*, 124 HARV. L. REV. 1226, 1285 (2011); *see* OECD Bribery Convention, *supra* note 311, art 1(1) ("Each Party shall take such measures as may be necessary to establish that [bribery] is a criminal offence under its law.").

313. *OECD Anti-Bribery Convention: National Implementing Legislation*, ORG. FOR ECON. CO-OPERATION & DEV., http://www.oecd.org/document/30/0,3746,en_2649_34859_2027102_1_1_1_1,00.html (last visited Apr. 21, 2012). Unfortunately, it appears that few countries have actually been enforcing the domestic antibribery provisions. *See Developments in the Law*, *supra* note 312, at 1285.

314. The idea that a nonbinding, defining declaration can provide a basis for negotiating a subsequent binding treaty is illustrated by the successful U.N. effort to criminalize torture and other cruel, inhuman, or degrading treatment. Before the Convention Against Torture was adopted by the U.N. General Assembly in 1984, the General Assembly adopted the Declaration Against Torture. Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment, Dec.

attacks amount to an armed conflict that warrants self-defense,[315] and could offer a common reference point for subsequent domestic criminal legislation.

Even an agreement limited to common definitions will likely face challenges.[316] In particular, it will be necessary to bridge fairly substantial divides between the United States and other leading cyber-powers that have a more expansive view of what activity ought to be criminalized through international cooperation, including some forms of legitimate political dissent.[317] As noted earlier, Russia and other members of the Shanghai Cooperation Organization have been promoting an international agreement banning cyber-attack for some time,[318] but their focus differs greatly from that of the United States and much of Europe in the cyber-attack arena.[319] A key challenge of this first stage agreement will thus be to find common ground with major cyber-powers without expanding the definition of cyber-attack in ways that would quell free speech and democratic political organization.

2. International Cooperation on Evidence Collection and Criminal Prosecution

Once states develop a shared definition of cyber-attacks, cyber-crime, and cyber-warfare, the next step is more extensive cooperation among states on

---

10, 1984, 1465 U.N.T.S. 85 [hereinafter CAT]; CHRIS INGELSE, THE UN COMMITTEE AGAINST TORTURE: AN ASSESSMENT 73 (2001). The Declaration described consensus on key elements of the definition of torture. These included "the infliction of severe physical or mental pain or suffering," intentional infliction of pain and suffering, the action or sanction of a public official, and conduct that serves a proscribed purpose, "such as obtaining information or a confession." *Id.* at 70. The Declaration provided much of the substance that later was incorporated into the Convention Against Torture, which has been ratified by 149 states, including the United States. *See* Status, Convention Against Torture and Other Cruel, Inhuman or Degrading Treatment or Punishment (last visited Nov. 8, 2011)**.** In fact, the Swedish draft of the Convention, which formed the basis of the negotiations, used the exact text of the definition of torture from the Declaration. INGELSE, *supra*, at 74. Unfortunately, the draft Sweden submitted to the 34th Session, E/CN.4/1285, is not available on the U.N. Documents database.

315. The White House predicts that shared understanding about norms of acceptable cyber-behavior will bring "predictability to state conduct, helping prevent the misunderstandings that could lead to conflict." WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 9. As a result, the strategy commits the United States to take the lead in building consensus on norms of cyber-behavior. *Id.* at 18.

316. Indeed, some have suggested that a successful treaty may be nearly impossible to achieve, at least in the short term. *See, e.g.*, Waxman, *supra* note 24, at 425–26 ("[N]ot only do certain features of cyber-activities make international legal regulation very difficult, but major actors also have divergent strategic interests that will pull their preferred doctrinal interpretations and aspirations in different directions, impeding formation of a stable international consensus."); GOLDSMITH, *supra* note 306, at 12 ("This paper has argued that the fundamental clash of interests concerning the regulation of electronic communications, the deep constraints the United States would have to adopt to receive reciprocal benefits in a cybersecurity treaty, and the debilitating verification problems will combine to make it unfeasible to create a cybersecurity treaty that purports to constrain governments."). For a dissenter's view on the appropriate international response to cyber-attack, see Duncan B. Hollis, *An E-SOS for Cyberspace*, 52 HARV. INT'L L.J. 373 (2011) (arguing for a duty to assist cyber-threat victims, rather than regulation of bad cyber-actors).

317. *See supra* text accompanying notes 21–26.

318. *See* Markoff & Kramer, *supra* note 306.

319. *See supra* text accompanying notes 24–26.

information sharing, evidence collection, and criminal prosecution of those involved in cyber-attacks. A useful starting point for building such a treaty is the Council of Europe Convention on Cybercrime, described in Part III.B.3, which provides for harmonized regulation of a wide range of cyber-crimes. This treaty remains largely limited to Europe (though the United States has ratified the agreement) and it does not address all cyber-attacks that a comprehensive agreement would ideally regulate.[320] Nonetheless, it provides a framework from which a more comprehensive agreement might begin.

Building on this framework, the new agreement should require parties to pass domestic laws banning the cyber-attack-related conduct prohibited under the treaty, so as to harmonize laws across states. The agreement could begin with information-sharing, layering on additional mechanisms for fostering cooperation in identifying and stopping the sources of cyber-attacks through criminal law enforcement agencies. International cooperation in information sharing could be an extremely valuable complement to other regulation of cyber-attacks.[321]

Member states could agree to share access to cyber-related information with other member states. That information would not be available to nonmembers or to states that fail to comply with the treaty's core obligations. Offering privileged access to information to member states in good standing would provide states with an incentive to participate in and comply with the treaty regime.[322]

Finally, consistent with the Tunis Commitment[323] and Agenda,[324] a treaty could encourage more-technologically-developed countries to assist less-developed ones in responding to shared cyber-threats. As the recent White House Cyberspace Strategy memo observed,

> Enhancing national-level cybersecurity among developing nations is of immediate and long-term benefit [to the United States and all nations], as more states are equipped to confront threats emanating from within their borders and in turn, build confidence in globally interconnected

320. *Convention on Cybercrime, Chart of Signatures and Ratifications*, COUNCIL OF EUR. (last visited Apr. 21, 2012), http://conventions.coe.int/Treaty/Commun/ChercheSig.asp?NT=185&CL=ENG. Canada, Japan, and South Africa are the other non-European signatories, but the United States is the only one of the four that has ratified the Convention. *Id.*

321. Information sharing in this context was endorsed by a group of experts from countries as diverse as the United States, China, and Russia in a 2010 report to the U.N. Secretary-General. *Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security*, *supra* note 186, at 8.

322. This proposal aims to harness the power of outcasting to build a strong treaty regime. *See* Oona Hathaway & Scott J. Shapiro, *Outcasting: The Enforcement of Domestic and International Law*, 121 YALE L. J. 252 (2011).

323. *Tunis Commitment*, WORLD SUMMIT ON THE INFO. SOC'Y (Nov. 18, 2005), http://www.itu.int/wsis/docs2/tunis/off/7.html (last visited Apr. 21, 2012).

324. *Tunis Agenda for the Information Society*, WORLD SUMMIT ON THE INFO. SOC'Y (Nov. 18, 2005), http://www.itu.int/wsis/docs2/tunis/off/6rev1.html (last visited Apr. 21, 2012).

> networks and cooperate across borders to combat criminal misuse of information technologies. It is also essential to cultivating dynamic, international research communities able to take on next-generation challenges to cybersecurity.[325]

Because any country's cyber-security can be compromised by its allies' security gaps, a collective attempt to prevent cyber-attacks must include efforts to improve the defenses of partner countries as well.[326]

Another challenge to any comprehensive cyber-attack treaty is the difficulty of verifying where cyber-attacks originate.[327] Uncertainty in tracing and attributing a cyber-attack "makes retaliation for breach much harder for any president or general to order."[328] Yet while verification of a cyber-attack's origin is difficult, even those who have expressed skepticism about the short-term feasibility of a cyber-treaty acknowledge that it is not impossible. As Jack Goldsmith has put it, "Sometimes traceback and related forensic tools can provide good-enough attribution."[329] Indeed, while negotiations on the treaty are underway, states should continue a parallel technical effort to enhance their capacities to trace the source of cyber-attacks.

As General Keith Alexander, chief of the new U.S. Cyber Command, explained earlier this year when reopening negotiations with Russia on this issue, "We do have to establish the lanes of the road" for what cyber-activities governments can and cannot pursue.[330] Establishing those lanes is the necessary first step to addressing the challenge of cyber-attacks. Only once they are in place will verification challenges become salient.

CONCLUSION

The emergence of Stuxnet in 2010 heralded a new era for cyber-attacks. Although its damage was apparently limited to the Iranian nuclear program it was designed to attack, it revealed how vulnerable even nation-states are to cyber-attacks. Indeed, by the time it was discovered, Stuxnet had wormed its way into computer networks around the world.

Cyber-attacks on vital infrastructure are already becoming widespread. Cyber-security professionals report that the computer infrastructure has become more vulnerable even in just a year.[331] And yet, while the threat of cyber-

---

325. WHITE HOUSE CYBERSPACE STRATEGY, *supra* note 63, at 15.

326. Shanker & Bumiller, *supra* note 53 (noting that the United States' allies are "all over the map" on cyber-security issues, according to James Lewis, an expert on computer network warfare at the Center for Strategic and International Studies).

327. *See* GOLDSMITH, *supra* note 306, at 10–12.

328. *Id.* at 11.

329. *Id.* at 10.

330. Siobhan Gorman, *U.S. Backs Talks on Cyber Warfare*, WALL ST. J., June 4, 2010, at A3.

331. Mark Clayton, *Security Lags Cyberattack Threats in Critical Industries, Report Finds*, CHRISTIAN SCI. MONITOR (Apr. 20, 2011), http://www.csmonitor.com/USA/2011/0420/Security-lags-cyberattack-threats-in-critical-industries-report-finds (citing a global survey of 200 computer security

attacks has rapidly grown, the response has not kept pace. This Article has shown that both the U.S. government and the international community have thus far largely failed to update legal frameworks that might respond to cyber-attacks. To face new and growing threats, governments continue to rely on limited and piecemeal bodies of law not designed to meet the challenge of cyber-attacks.

It is past time to begin a conversation about the scope of the threat posed by cyber-attacks and the best ways to meet it. The United States should expand the reach of domestic law abroad and develop a system for utilizing limited countermeasures where appropriate to respond to certain types of cyber-attacks. Yet the United States is restricted in what it can accomplish alone. Cyber-attacks are often transnational—designed by authors in multiple countries, run through networks across the world, and used to undermine computer systems in countries where those designing the attack have never set foot. This global threat may only be effectively met by a global solution—by the international community working together to design a new law for cyber-attacks.

professionals working in critical infrastructure industries, "In the Dark: Crucial Industries Confront Cyberattacks").

Fordham International Law Journal

Volume 30, *Issue* 3 2006 *Article* 9

When to Push the Envelope: Legal Ethics, the Rule of Law, and National Security Strategy

Peter Margulies*

*

When to Push the Envelope: Legal Ethics, the Rule of Law, and National Security Strategy

Peter Margulies

Abstract

This Article argues for pushing the envelope when three conditions are met: (1) the executive engages in dialogue with other players, either before the fact or through timely ex post ratification; (2) pushing the envelope will generate a net positive aggregate of institutional consequences, viewed from an intermediate and long-term perspective; and (3) pushing the envelope harmonizes executive policy with evolving international or domestic norms. When these conditions are met, the lawyer for the executive should recommend the action, even if it appears inconsistent with the letter of existing law. While acting gives both the lawyer and her client "dirty hands," a failure to act may expose the United States to even greater risk. When the executive is unable or unwilling to meet all of these conditions, however, approving the proposed action places the lawyer in ethical peril. Part I of this Article discusses the adverse effects of detention policies on legal ethics and the integrity of the justice system. Part II uses a broader lens to describe costs to the United States' credibility and reputation. Part III sets out the test for pushing the envelope, and discusses two examples from history: Lend-Lease and the Cuban Missile Crisis. The goal of this Article is to show that legal ethics in national security strategy must reject absolutes. A blind aggrandizement of executive power will pose ethical and policy problems. A risk-averse position that avoids pushing the envelope, however, can also pose dangers. Judgment, not a categorical approach, is necessary to discern the most prudent path.

WHEN TO PUSH THE ENVELOPE: LEGAL ETHICS, THE RULE OF LAW, AND NATIONAL SECURITY STRATEGY

*Peter Margulies**

Lawyers in national security matters face a perennial dilemma. On the one hand, an unyielding respect for the letter of the law does not mix well with national security strategy. Courts have long recognized that a doctrinaire absolutism about legal commands cannot accommodate the fluidity of foreign policy.[1] Moreover, a preoccupation with clean hands may prevent the politician from making difficult choices that ensure survival.[2] On the other hand, lawyers and other policymakers in the national security realm must also uphold core legal principles and preserve the integrity of legal institutions. Too often, lawyers in national security crises have skewed this calculus toward expediency, without paying sufficient attention to abiding values.[3]

This loss of equipoise is especially acute where, as in the case of Guantanamo, policies entail detention without trial. U.S. history has shown that regimes of mass detention undermine the legal system's values. A number of sorry episodes, most notably

---

* Professor of Law, Roger Williams University. I thank Laura Corbin for her enterprising and resourceful research assistance, and John Barrett, Bruce Green, David Luban, and participants at a workshop at Roger Williams Law School for comments on a previous draft.

1. *See* Youngstown Sheet and Tube Co. v. Sawyer, 343 U.S. 579, 640 (1952) (Jackson, J., concurring) (recommending "scope and elasticity afforded by what seems to be reasonable" in construction of executive power, and cautioning against "rigidity dictated by a doctrinaire textualism"); *cf.* Peter Margulies, *Judging Terror in the "Zone of Twilight": Exigency, Institutional Equity, and Procedure After September 11*, 84 B.U. L. REV. 383, 402-16 (2004) (outlining pragmatic approach to separation of powers questions in law and terrorism cases).

2. *See generally* Michael Walzer, *Political Action: The Problem of Dirty Hands*, *in* TORTURE: A COLLECTION 61-75 (Sanford Levinson ed., 2004) (describing the dilemma between governing innocently and making ethnical choices for the good of the nation).

3. *See* Jose Alvarez, *Torturing the Law*, 37 CASE W. RES. J. INT'L L. 175, 215-21 (2006). *See generally* Kathleen Clark, *Ethical Issues Raised by the OLC Torture Memorandum*, 1 J. NAT'L SEC. L. & POL'Y 455 (2005) (finding that the lawyers that drafted the Torture Memos failed to provide candid legal advice or inform their client about the risks of its actions); George C. Harris, *The Rule of Law and the War on Terror: The Professional Responsibilities of Executive Branch Lawyers in the Wake of 9/11*, 1 J. NAT'L SEC. L. & POL'Y 409 (2005) (arguing that the lawyers who drafted the Torture Memos failed to discharge their professional obligations).

the internment of Japanese-Americans in World War II,[4] demonstrate that detentions develop an institutional momentum that undermines accountability, fairness, and equality. This Article offers a framework for determining when pushing the envelope in national security crises is justifiable as a matter of law and legal ethics.

This Article argues for pushing the envelope when three conditions are met: (1) the executive engages in dialogue with other players, either before the fact or through timely *ex post* ratification; (2) pushing the envelope will generate a net positive aggregate of institutional consequences, viewed from an intermediate and long-term perspective; and (3) pushing the envelope harmonizes executive policy with evolving international or domestic norms. When these conditions are met, the lawyer for the executive should recommend the action, even if it appears inconsistent with the letter of existing law. While acting gives both the lawyer and her client "dirty hands," a failure to act may expose the United States to even greater risk. When the executive is unable or unwilling to meet all of these conditions, however, approving the proposed action places the lawyer in ethical peril.

Part I of this Article discusses the adverse effects of detention policies on legal ethics and the integrity of the justice system. Part II uses a broader lens to describe costs to the United States' credibility and reputation. Part III sets out the test for pushing the envelope, and discusses two examples from history: Lend-Lease and the Cuban Missile Crisis. The goal of this Article is to show that legal ethics in national security strategy must reject absolutes. A blind aggrandizement of executive power will pose ethical and policy problems. A risk-averse position that avoids pushing the envelope, however, can also pose dangers. Judgment, not a categorical approach, is necessary to discern the most prudent path.

I. *NATIONAL SECURITY, DETENTION, AND LAWYERS' ETHICS*

Detention of perceived national security threats outside the traditional confines of the criminal justice system strains the eth-

4. *See* Korematsu v. United States, 323 U.S. 214, 219 (1944).

ics of government lawyers. Detention may be necessary in exigent situations.[5] Nevertheless, the charged atmosphere of national security advice and litigation can cast legal ethics as a luxury that the attorney can ill afford. In this context, institutional and ideological factors can erode compliance with ethical norms.

Institutional factors include the familiar collective action problem of the "race to the bottom." While government lawyers do not bill by the hour, they do compete for power, prestige, and influence.[6] In the national security arena, government lawyers compete for influence on decision-makers by signaling their willingness to tolerate conduct that is close to the line of legality.[7] In the short term, a lawyer who tells a decision-maker what that senior official wants to hear receives even more attention. This attention generates more prestigious assignments. In addition, the lawyer gets to see her recommendations played out in actual government policy.[8]

---

5. *See* Hamdi v. Rumsfeld, 542 U.S. 507, 538 (2004) (authorizing detention of suspected terrorist apprehended in "theatre of war," subject to procedural protections); *cf.* Margulies, *supra* note 1 (discussing appropriateness of narrowly tailored detention); Stephen I. Vladeck, Note, *The Detention Power*, 22 YALE L. & POL'Y REV. 153, 155 (2004) (expressing doubt about authority of executive to detain individuals without express congressional approval).

6. Competition is often a beneficial phenomenon, sharpening the competence of the participants and producing goods that match consumer needs. Allocating power, prestige, and influence through other criteria, including status, race, or ethnicity, has obvious downsides. However, competition can also have a negative effect on public goods, including the overall integrity of the system. "Market failure" of this kind is a compelling rationale for regulation of competition, generally, and lawyers' ethics in particular. *Cf.* Lucian Arye Bebchuk & Christine Jolls, *Managerial Diversion and Shareholder Wealth*, 15 J.L. ECON. & ORG. 487, 489-90 (1999) (discussing importance of regulation to ensure transparency in executive compensation); George C. Triantis, *Organizations as Internal Capital Markets: The Legal Boundaries of Firms, Collateral, and Trusts in Commercial and Charitable Enterprises*, 117 HARV. L. REV. 1102, 1116 (2004) (noting need to regulate self-interest of corporate managers).

7. Comparable problems led to the abdication of corporate lawyers' gate keeping role in the Enron debacle. *See* JOHN C. COFFEE, JR., GATEKEEPERS: THE PROFESSIONS AND CORPORATE GOVERNANCE 205-06 (2006); Peter Margulies, *Lawyers' Independence and Collective Illegality in Government and Corporate Misconduct, Terrorism, and Organized Crime*, 58 RUTGERS L. REV. 939 (2006); Milton C. Regan, Jr., *Ethics in Corporate Representation: Teaching Enron*, 74 FORDHAM L. REV. 1139, 1146-47 (2005); William H. Simon, *The Post-Enron Identity Crisis of the Business Lawyer*, 74 FORDHAM L. REV. 947, 948-49 (2005).

8. Effective ties between attorney and client also play a role. Lawyers often turn to public service because they feel inspired by a particular public official. This was certainly true, for example, of lawyers who served Presidents Roosevelt, Kennedy, and Reagan. The approval of these charismatic figures may attain a special import for the lawyer, overwhelming ethical scruples. Moreover, lawyers want to be seen by clients whom

Ideological allegiances also play a major role in this process. For many lawyers in the current Administration, any tensions with legal ethics are at best hiccups that distract from the main mission: restoring the power of the Presidency.[9] This view has a compelling origin story: a narrative attributed to the Framers, in which the President wields virtually untrammeled power in foreign affairs. Supposed constraints on the President in domestic or international law are suspect. The problem is that this origin story of executive power badly distorts the Framers' words, actions, and intent. While the Framers recognized that in emergencies the President had certain institutional advantages, they also recognized the need for collaboration between the branches of Government.[10] In addition, they understood the importance of treaty obligations and other authority under international law.[11] Ideological champions of presidential power can, however, dismiss these critiques as the carps and cavils of the uninitiated.

Once the policy universe includes extraordinary detention regimes, institutional momentum takes over.[12] The lack of accountability becomes seductive. The new solution goes off in search of problems to solve.[13] As Twain said, "[g]ive someone a

---

they admire, respect, and depend on for career advancement as "getting with the program."

Lawyers do not have to go this route. Indeed, finding equipoise between achieving the client's goals and upholding the integrity of the system is a central responsibility for the lawyer. However, the urgency of national security matters makes this balance very difficult to maintain. *See* David Luban, *Liberalism, Torture, and the Ticking Bomb*, 91 VA. L. REV. 1425, 1452-61 (2005); W. Bradley Wendel, *Legal Ethics and the Separation of Law and Morals*, 91 CORNELL L. REV. 67, 80-85 (2005); W. Bradley Wendel, *Professionalism as Interpretation*, 99 NW. U.L. REV. 1167, 1171-74 (2005).

9. *See, e.g.*, John C. Yoo, *War and the Constitutional Text*, 69 U. CHI. L. REV. 1639, 1642 (2002).

10. *See generally* Curtis A. Bradley & Martin S. Flaherty, *Executive Power Essentialism and Foreign Affairs*, 102 MICH. L. REV. 545 (2004) (critiquing view that Vesting Clause of U.S. Constitution grants President unfettered power over foreign affairs).

11. *See generally* Martin S. Flaherty, *History Right?: Historical Scholarship, Original Understanding, and Treaties as "Supreme Law of the Land,"* 99 COLUM. L. REV. 2095 (1999) (outlining a historical approach to international law and treaties based on the intent of the Framers of the Constitution).

12. *See* JONATHAN SIMON, GOVERNING THROUGH CRIME: HOW THE WAR ON CRIME TRANSFORMED AMERICAN DEMOCRACY AND CREATED A CULTURE OF FEAR 29 (2007) (arguing that proliferation of anti-crime legislation since the 1960's reflected availability of sweeping government authority in this area, more than substantive priority of crime over other social problems such as poverty or environmental depredation).

13. Means for implementing a policy often influence identification and analysis of

hammer and they look for a nail."[14] In this environment, a spectrum of legal ethics problems emerge from lawyers' eagerness to justify the new approach. For example, lawyers risk counseling their policymaker clients to engage in illegal acts or target minorities. Lawyers also may display a lack of candor with courts. The Article addresses each issue in turn.

A. *Assisting the Client's Illegal Acts*

Under the rules governing legal ethics, a lawyer may not knowingly counsel or assist the client in committing an illegal act.[15] The reason for this is simple: our legal system places a high value on lawyers, but regards accomplices more dimly. Unfortunately, national security strategy places attorneys in tension with this mandate.[16] National security strategy may clash with international law, as perceived national interests conflict with the international legal structure.[17] In addition, the executive branch may find it desirable to act inconsistently with the will of Congress. Unless the President has power under Article II of the Constitution to take the action, the lawyers' approval of the act will upset the orderly scheme of separation of powers, which describes the President's power as weakest when he acts in defiance of the legislature.[18]

The roots of the Bush Administration's disregard for law stem from an episode, Iran-Contra, where the Reagan Administration disregarded both a federal statute and international norms. Much of the Administration's view that it is not only per-

---

the underlying problem. *See, e.g.*, JAMES G. MARCH & JOHAN P. OLESEN, REDISCOVERING INSTITUTIONS: THE ORGANIZATIONAL BASIS OF POLITICS 13 (1989) (noting that a "solution [in public policy terms] . . . is an answer actively looking for a question").

14. *See* Alan Dershowitz, *Tortured Reasoning*, *in* TORTURE: A COLLECTION, *supra* note 2, at 257, 271.

15. A lawyer may not "[c]ounsel a client to engage, or assist a client, in conduct that the lawyer knows is criminal or fraudulent." MODEL RULES OF PROF'L CONDUCT R. 1.2(d) (2003); *see* MODEL CODE OF PROF'L RESPONSIBILITY DR 7-102(A)(7) (1983).

16. *See generally* Alvarez, *supra* note 3 (arguing that lawyers feel the need to "torture the law" to insulate themselves from liability); Clark, *supra* note 3 (describing the Government's reliance on the inaccurate characterization of the Bybee Memorandum, written by the Office of Legal Counsel, for drafting interrogation policies); Margulies, *supra* note 7.

17. *Cf.* Jed Rubenfeld, *Unilateralism and Constitutionalism*, 79 N.Y.U. L. REV. 1971 (2004) (analyzing both complementarity and conflict between international law and national self-government).

18. *See* Youngstown Sheet and Tube Co. v. Sawyer, 343 U.S. 579, 637-39 (1952) (Jackson, J., concurring).

mitted, but virtually required to disregard otherwise applicable law flows from the Minority Report issued by Republican members of the House Select Committee investigating the Reagan Administration's attempts to both provide weapons to the Iranians and undermine the Nicaraguan Government.[19] Doing the latter involved violating the Boland Amendment, a federal statute that barred aid to the Contra rebel group.[20] It also involved significant tension with international law norms that forbid the unjustified use of military force by one State against another through nongovernmental surrogates or the United States' own military forces.[21] By aiding the Contras, and then lying about it to Congress, the Reagan Administration turned its back on the *Youngstown* framework and cost itself credibility at home and abroad.

The present Administration's lawyers have engaged in even more problematic behavior with respect to Model Rule of Professional Conduct 1.2. Consider the problematic stance on international law adopted by John Yoo, Jay Bybee, and the other authors of the so-called "Torture Memos." Administration lawyers articulated a narrow definition of torture wholly at odds with the spirit and logic of international law,[22] thus giving United States personnel a virtual license to mistreat detainees.[23] In addition, this

---

19. *See* STEPHEN DYCUS ET AL., NATIONAL SECURITY LAW 482-86, 505 (3d ed. 2002); *see also* Jane Mayer, *The Hidden Power: A Secret Architect of the War on Terror*, NEW YORKER, July 3, 2006, at 44, 49 (noting role of long-time Cheney aide David Addington in the Minority Report and policy in the Bush Administration); Jeffrey Rosen, *Power of One: Bush's Leviathan State*, NEW REPUBLIC, July 24, 2006, at 8-10 (discussing origins of Bush Administration view in Iran-Contra affair).

20. *See* DYCUS ET AL., *supra* note 19, at 491-93.

21. *See generally* Military and Paramilitary Activities (Nicar. v. United States), 1986 I.C.J. 14 (June 27). Ironically, more recent events, including the State manipulation of private death squads in the former Yugoslavia and the international community's revulsion at the involvement of the Taliban in supporting al-Qaeda, have arguably led to a broader test for determining a State's responsibility for the actions of private groups. *See* Prosecutor v. Tadic, Case No. IT-94-1-A, Judgment, ¶ 137 (July 15, 1999); *cf.* Vincent-Joel Proulx, *Babysitting Terrorists: Should States Be Strictly Liable for Failing to Prevent Transborder Attacks?*, 23 BERKELEY J. INT'L L. 615, 630-41 (2005) (arguing that broader standard is appropriate to encourage State diligence).

22. *See* United Nations Convention Against Torture and Other Cruel, Inhuman, or Degrading Treatment or Punishment ("CAT"), Dec. 10, 1984, 112 Stat. 2681, 1465 U.N.T.S. 85 (defining torture as "any act by which severe pain or suffering, whether physical or mental, is intentionally inflicted on a person" for purposes of punishment, intimidation, discrimination, or extracting a confession).

23. Memorandum for Alberto R. Gonzales, Counsel to the President, Re: Standards of Conduct for Interrogation under 18 U.S.C. §§ 2340-2340A (Aug. 1, 2002), *in*

cramped definition was inconsistent with the purpose and meaning of the War Crimes Act, which relies on international standards.[24]

The Administration ran into further trouble with its national security wiretapping policy.[25] This policy violated specific provisions in the Foreign Intelligence Surveillance Act ("FISA"), which require the government to seek a warrant within three days of beginning surveillance and allow fifteen days of warrantless surveillance during a war.[26] By opining that the President could act inconsistently with FISA, the Administration's lawyers again ran afoul of the venerable *Youngstown* framework.

While gray areas are common in national security law, issues such as torture and warrantless wiretapping also feature some reasonably clear boundaries. In isolated, highly exigent situations, government conduct that crosses the line may be difficult to condemn categorically.[27] Lawyers who consistently advise conduct that straddles the boundary, however, blur the distinction between attorney and accomplice.[28]

MARK DANNER, TORTURE AND TRUTH: AMERICA, ABU GHRAIB, AND THE WAR ON TERROR 115, 145 (2004) [hereinafter Memo for Alberto Gonzales] (arguing that federal statute criminalizing practice of torture "must be construed as not applying to interrogations undertaken pursuant to [the President's] Commander-in-Chief authority"). *But see* Sanford Levinson, *Contemplating Torture, in* TORTURE: A COLLECTION, *supra* note 2, at 23, 28-30 (criticizing analysis in torture memos); *cf.* Peter Margulies, *Beyond Absolutism: Legal Institutions in the War on Terror*, 60 U. MIAMI L. REV. 309, 313 (2006) (arguing for pragmatic focus on interaction of norms and institutions in torture debate).

24. *See* John Yoo & Robert J. Delahunty, U.S. Dep't of Justice, Office of Legal Counsel, *Application of Treaties and Laws to Al Qaeda and Taliban Detainees, in* DYCUS ET AL., *supra* note 19, at 47-48 (Supp. 2005-2006) (citing then-current 18 U.S.C. § 2441(c)(3)(2005)) (war crimes included violations of common Article 3 of the Geneva Convention, such as torture and cruel, inhuman, or degrading treatment). In the recently enacted Military Commissions Act, Congress diluted the provision dealing with Common Article 3 by specifying that only "grave breaches," not mere violations were prohibited. *See* 18 U.S.C. § 2441(d)(2007). Congress retained the prohibition on torture and cruel and inhuman treatment. *See id.*

25. *See generally* ACLU v. NSA, 438 F. Supp. 2d 754 (E.D. Mich. 2006) (striking down part of NSA wiretapping policy).

26. Foreign Intelligence Surveillance Act, Pub. L. No. 95-11, 92 Stat. 1753 (1978) (codified at 50 U.S.C. §§1801-1811).

27. *See* Margulies, *supra* note 23. *But see* Kim Lane Scheppele, *Hypothetical Torture in the "War on Terrorism,"* 1 J. NAT'L SEC. L. & POL'Y 285 (2005) (critiquing facile and frequent use of "ticking bomb" scenario to justify torture).

28. *See* Margulies, *supra* note 7; *cf.* Peter Margulies, *The Virtues and Vices of Solidarity: Regulating the Roles of Lawyers for Clients Accused of Terrorist Activity*, 62 MD. L. REV. 173 (2003) (discussing ethical risks for criminal defense lawyers).

B. *Targeting Based on Race*

Another disturbing aspect of lawyering on matters of detention is reliance on stereotypes and profiling. Generalizations about citizens and immigrants have been a mainstay of national security policy since World War I.[29] In contrast, rules against "bias or prejudice" in the practice of law are of recent origin.[30] An unduly rigid application of ethical restrictions on bias might chill lawyering even where ethnicity, religion, or national origin was one criterion among many. A more robust interpretation of the ethical rules, however, would promote liberty, equality, and accountability. In addition, decreasing reliance on stereotypes in decisions about arrest, detention, and deportation would promote efficiency in law enforcement and national security policy.

The ethical strictures against lawyers' "words or conduct" that manifest bias echoes clear prohibitions in international law.[31] For example, the International Covenant on Civil and Political Rights ("ICCPR") bars discrimination on the basis of nationality, race, religion, and other factors, and imposes duties on States to implement this prohibition. While the United States has limited the legal force of the ICCPR, the overarching principle of non-discrimination commands wide respect. In an increasingly interdependent world, equality is a good for its own sake. Moreover, on the international stage, a commitment to the principle of equality also encourages reciprocity by countries and communities that might otherwise be suspicious of each

29. *See* OFFICE OF THE INSPECTOR GEN., U.S. DEP'T OF JUSTICE, THE SEPTEMBER 11 DETAINEES: A REVIEW OF THE TREATMENT OF ALIENS HELD ON IMMIGRATION CHARGES IN CONNECTION WITH THE INVESTIGATION OF THE SEPTEMBER 11 ATTACKS (June 2003) (critiquing criteria used to detain aliens after September 11); *cf.* Natsu Taylor Saito, *Symbolism Under Siege: Japanese American Redress and the "Racing" of Arab Americans as "Terrorists,"* 8 ASIAN L.J. 1 (2001) (exploring common themes between Japanese-American internment and post-September 11 measures); Peter Margulies, *Above Contempt?: Regulating Government Overreaching in Terrorism Cases*, 34 SW. U. L. REV. 449 (2005) (discussing the dangers of a monolithic paradigm in times of national emergency).

30. *See* MODEL RULES OF PROF'L CONDUCT R. 8.4 cmt.3 (2003).

31. *See* International Covenant on Civil & Political Rights ("ICCPR"), Dec. 19, 1966, 999 U.N.T.S. 171; Daniel Moeckli, *The Selective "War on Terror": Executive Detention of Foreign Nationals and the Principle of Non-Discrimination*, 31 BROOK. J. INT'L L. 495, 515-19 (2006); *cf.* Diane F. Orentlicher, *Criminalizing Hate Speech in the Crucible of Trial: Prosecutor v. Nahimana*, 21 AM. U. INT'L L. REV. 557, 570-73 (2006) (discussing legal options regarding state adherence to provisions of international agreements on hate speech and problems with prosecuting hate speech in international tribunals).

other's motives.[32] This element of reciprocity is important because politicians throughout the world often act against minority communities under the guise of national security.[33] Legal advice that limits resort to this gambit will promote a positive brand of reciprocity, as well as a more focused national security strategy that concentrates on genuine threats.

Unfortunately, government lawyers addressing national security policies that rely on stereotypes have failed to consider international law norms barring discrimination. For example, after September 11, the Administration engaged in a round-up of undocumented aliens from Middle Eastern and South Asian countries.[34] There is no evidence, however, that legal advice to the Bush Administration on measures such as the round-up considered international law norms or the impact of such actions on public opinion abroad. While the Administration has engaged in negotiations with certain countries regarding the detention of their nationals at Guantanamo, these negotiations have been ad hoc, without the bedrock of principle that would comply with the spirit of international norms and persuade international audiences of the United States' good faith. Administration lawyers have also defended measures such as the immigration round-up as justified exercises of executive authority.[35] While the content of the Administration lawyers' arguments has not affirmatively promoted stereotypes, one can argue that the deference to policies based on stereotypes, not particularized proof, nonetheless manifests bias.

Government lawyers dealing with national security issues have been similarly unconstrained by provisions of U.S. law that require particularized suspicion and equal treatment under the law. During the Civil War, military authorities used the suspension of habeas corpus to detain thousands of citizens, some ap-

---

32. *Cf.* Catherine Powell, *The Role of Transnational Norm Entrepreneurs in the U.S. "War on Terrorism,"* 5 THEORETICAL INQUIRIES L. 47, 72 (2004) (stressing importance of dialogue between transnational non-governmental organizations and United States on human rights issues).

33. *See* BONNIE HONIG, DEMOCRACY AND THE FOREIGNER 81 (2001) (discussing persecution of immigrant dissidents).

34. *See* DAVID COLE, ENEMY ALIENS (2003). *See generally* Letti Volpp, *The Citizen and the Terrorist*, 49 UCLA L. REV. 1575 (2002) (describing the marginalization of particular communities after September 11).

35. At least one court has agreed. *See, e.g.*, Turkmen v. Ashcroft, 2006 U.S. Dist. Lexis 39170 (E.D.N.Y. June 14, 2006).

propriately (particularly early in the conflict), but others on charges so nebulous that the detainees' jailers could not even recall them when asked.[36] During World War I, the Administration rounded up dissidents, many of them Jews, like the anarchist Emma Goldman,[37] and imprisoned or deported many. The Government's actions during World War II, however, present the most troubling case for both policy and lawyers' ethics.

The legal defense of the internment policy hinged on a stereotype of Japanese-Americans as insidious and inscrutable security risks.[38] Ironically, the narrative shaped by government lawyers acknowledged the discrimination that Japanese-Americans had frequently faced in the United States, but then leveraged that history of discrimination to paint Japanese-Americans as a resentful and cloistered minority eager to exact their revenge on U.S. interests. Based on the government lawyers' skillful advocacy, the Supreme Court accepted this argument in *Hirabayashi v. United States*.[39] Once accepted by the Court, these arguments formed part of the backdrop for the Court's decision in *Korematsu v. United States*[40] that upheld a statute that criminalized resistance to the forced evacuation of Japanese-Americans from their homes.

One aspect of a learned profession like the law is that practitioners should learn from their mistakes. Unfortunately, the present Administration has not taken that lesson to heart. As episodes like the post-9/11 round-up show, lawyers for the Bush Administration have seen fit not to learn from the mistakes of the past, but to repeat them.

C. *Lack of Candor With the Tribunal*

Another disturbing effect of detention policies has been the lack of candor thereby promoted among lawyers charged with defending the Government. Candor with the tribunal has been among the most important ethical dictates of the lawyer. Without candor the adversary system cannot function. For this rea-

36. *Cf.* MARK NEELY, THE FATE OF LIBERTY 52-65 (1991) (discussing examples of abusive confinement during the Civil War).

37. *See* PETER IRONS, JUSTICE AT WAR 15 (1983).

38. Lawyers honed this strategy despite their private doubts. *Id.*

39. 320 U.S. 81 (1943). For background on the case, see IRONS, *supra* note 36, at 249-50.

40. 323 U.S. 214 (1944).

son, ethical rules prohibit the lawyer from making false statements of law or fact to a tribunal, or "failing to correct" false statements previously made.[41] Unfortunately, episodes of detention—sometimes involving illustrious U.S. lawyers—have revealed a lack of compliance with this ethical norm.

Deception is a perpetual risk because governments often resort to detention when evidence is murky or nonexistent. Conceding this lack of evidence can produce embarrassment, shame, and legal liability. Locked in a race to the bottom, lawyers turn to exaggeration, fabrication, and concealment as winning strategies.

The most salient example of lack of candor occurred during the litigation in the Supreme Court concerning the Japanese-American internment during World War II. The litigation of the *Korematsu* case involved a number of celebrated lawyers, including the Secretary of War Henry Stimson, Assistant Secretary John McCloy and Assistant Attorney General Herbert Wechsler, a professor at Columbia who subsequently drafted the Model Penal Code, co-wrote a pioneering casebook on federal courts, and authored a profoundly influential article on "neutral principles" in constitutional law. An important feature of this litigation was the report prepared by General John DeWitt for the War Department setting out the basis for the government's policy ("Report" or "DeWitt Report"). DeWitt's Report was the fulcrum for the lawyers' lack of candor.

DeWitt made a number of damning claims in the Report, including the charged assertion that the government had documented Japanese-American radio transmissions from the West Coast to the forces of the Japanese Empire. DeWitt also asserted that threats of violence by whites prevented the voluntary movement of Japanese-Americans from the West Coast to less sensitive areas further east. The Report stated that the failure of a voluntary program of evacuation and resettlement justified the evacuation and internment policy.[42] Both of these claims were false. An investigation by the Federal Bureau of Investigation ("FBI")

---

41. *See* MODEL RULES OF PROF'L CONDUCT, Rule 3.3 (2005). The Model Code of Professional Responsibility, an earlier codification of ethical rules first adopted in 1908 and still in force in a majority of states, has comparable rules. *See* MODEL CODE OF PROF'L RESPONSIBILITY, DR 7-102(A)(5) (2005) (prohibiting the lawyer from knowingly making a false statement of law or fact).

42. *See* IRONS, *supra* note 36, at 294-95; *cf.* Joseph Margulies, Evaluating Crisis Gov-

failed to find any documented instances of radio transmissions.[43] Indeed, the FBI concluded that these claims were fabrications.[44] In addition, the facts wholly failed to demonstrate that violence against Japanese-Americans would have doomed a voluntary program.[45]

Fidelity to ethical rules requiring candor with the tribunal would have required a clear distinction between the discredited claims in the DeWitt Report and the facts as stated in the government's brief. A straightforward disavowal of the Report would have been the action most appropriate for avoiding any misapprehension on the part of the Court. Several Justice Department lawyers wished to take this step. Wechsler, however, at McCloy's urging, decided that a less precise caveat was appropriate.[46]

Wechsler drafted a footnote that said the following: "We have specifically recited in this brief facts relating to the justification for the Evacuation, of which we ask the Court to take judicial notice, and we rely upon the . . . [DeWitt] Report only to the extent that it relates to such facts."[47] This cryptic footnote, however, failed to fulfill the Justice Department lawyers' duty under the ethical rules. First, the footnote failed to adequately identify those portions of the DeWitt Report that the lawyers knew to be false. This invited confusion on the part of the Court, particularly since the Court had already relied on the Report in the *Hirabayashi* case.[48] Second, on its own terms, the footnote was faulty. While the Justice Department asked the Court to take judicial notice of the report's accuracy on the matter of violence against Japanese-Americans, the facts did not provide the clarity and certainty that judicial notice demands.[49]

---

ernment, 40 CRIM. L. BULL. 627, 638-39 (2004) (discussing problematic basis for U.S. imposition of martial law in Hawaii during World War II).

43. *See id.* at 291.

44. *See id.*

45. *See id.* at 294-95, 299-300.

46. *See* Norman Silber & Geoffrey Miller, *Toward "Neutral Principles" in the Law: Selections from the Oral History of Herbert Wechsler*, 93 COLUM. L. REV. 854, 886-90 (1993).

47. *See* IRONS, *supra* note 36, at 290-91.

48. *See id.* at 291 (including the assertion in *Hirabayashi* that the "opportunity for espionage and sabotage" justified the evacuation); *cf.* Korematsu v. United States, 584 F.Supp. 1406, 1417-19 (N.D. Calif. 1984) (granting writ of *coram nobis* vacating Korematsu's conviction, based on government's reliance on a misleading factual record).

49. *See id.* 299-300.

Ultimately, the Court in *Korematsu* relied on the DeWitt Report as a basis for upholding the statute criminalizing failure to report to a government facility.[50] In *Ex Parte Endo*,[51] issued on the same day, the Court granted the secret wish of the Justice Department's lawyers and held that the executive lacked statutory authority to detain a concededly loyal Japanese-American.[52] While this decision effectively ended the internment program,[53] the *Korematsu* holding has retained its impact in U.S. history and culture. No case better demonstrates the pressures that undermine lawyers' fidelity to ethical rules in national security matters, and the ill effects yielded by government lawyers' failure to internalize ethical norms.

Lawyers in the Bush Administration have also acted in a fashion that suggests a lack of candor on issues of detention. Consider the case of Brandon Mayfield, a Portland lawyer whom the FBI arrested in May 2004 as a material witness in the investigation of the Madrid train bombing.[54] The FBI had examined a third-hand version of a fingerprint found at the scene, and matched that print with Mayfield.[55] Prosecutors submitted an affidavit asserting that Spanish authorities agreed with the fingerprint analysis of the FBI.[56] Based on the affidavit, a judge approved a covert search of Mayfield's residence and Mayfield's detention for seventeen days as a material witness. Federal authorities released Mayfield after they conceded that the fingerprints did not match.[57] In fact, FBI agents knew from the start that Spanish authorities disagreed with the FBI's fingerprint analysis.[58] If the prosecutor knew or came to know about this

50. Korematsu v. United States, 323 U.S. 214, 219 n.2, 223-24 (1944).

51. 323 U.S. 283, 294 (1944); *cf.* Patrick O. Gudridge, *Remember* Endo?, 116 HARV. L. REV. 1933 (2003) (discussing significance of case).

52. *See* Jane B. Baron & Julie Epstein, *Is Law Narrative?*, 45 BUFF. L. REV. 141, 160 (1997) (noting Wechsler's recollection that, in *Endo*, the Justice Department lawyers "lost and were delighted to lose").

53. *See* Gudridge, *supra* note 50, at 1933.

54. *See* Sarah Kershaw & Eric Lichtblau, *Spain Had Doubts Before U.S. Held Lawyer in Blast*, N.Y. TIMES, May 26, 2004, at A1.

55. *Id.*; *cf.* Darryl K. Brown, *Rationalizing Criminal Defense Entitlements: An Argument From Institutional Design*, 104 COLUM. L. REV. 801, 823-24 (2004) (noting surprisingly weak reliability of fingerprint evidence).

56. *See* OFFICE OF THE INSPECTOR GEN. ("OIG"), U.S. DEP'T OF JUSTICE, A REVIEW OF THE FBI'S HANDLING OF THE BRANDON MAYFIELD CASE (March 2006), *available at* http://www.usdoj.gov/oig/special/s0601/final.pdf [hereinafter OIG REP.].

57. *See* Kershaw & Lichtblau, *supra* note 53.

58. The FBI's focus on Mayfield, who was entirely unconnected to the Madrid at-

disagreement, making a contrary assertion in the affidavit or failing to alert the judge upon discovering the discrepancy between the facts and the pleadings submitted would be a violation of the duty of candor.[59] The government recently settled a subsequent civil suit by Mayfield, agreeing to pay him US$2 million.[60]

Government lawyers have also discounted the need for candor in the extensive litigation surrounding Jose Padilla, a U.S. citizen whom the government detained for three and a half years as an alleged enemy combatant. In *Padilla v. Hanft*,[61] Judge J. Michael Luttig of the U.S. Court of Appeals for the Fourth Circuit, an ideological soul-mate of the Administration, told the sad story of government lawyers' slippery strategy in arguing for the legality of Padilla's detention. Those lawyers almost certainly knew that if the Fourth Circuit agreed, the government would moot the dispute by charging Padilla with criminal violations, rather than face Supreme Court review.

Judge Luttig responded with a blistering opinion that called the government's good faith into question. After observing that the government was "steadfastly maintaining that [Padilla's detention] . . . was imperative in the interest of national security,"[62] Luttig noted the peculiar coincidence that the governments' fil-

---

tacks, may have stemmed from evidence relating to Mayfield's religion and professional associations: Mayfield was a practicing Muslim, had called the head of a local Islamic organization, and had represented a terrorism defendant in a completely unrelated family law case. *But see* OIG REP., *supra* note 55, at 267, 270 (claiming that Mayfield's religious affiliations and legal experience had no impact on law enforcement decisions).

59. While the OIG cleared Department of Justice ("DoJ") attorneys of any wrongdoing, see OIG REP., *supra* note 55, the Report does not explain how an alert prosecutor could have failed to ask Federal Bureau of Investigation ("FBI"). *See* Daniel Richman, *Prosecutors and Their Agents, Agents and Their Prosecutors*, 103 COLUM. L. REV. 749 (2003) (discussing complex relationship between agents and prosecutors). *See generally* MODEL RULES OF PROF'L CONDUCT R. 3.8 cmt.1 (2005) ("a prosecutor has the responsibility of a minister of justice and not simply that of an advocate"); Bruce A. Green & Fred C. Zacharias, *Regulating Federal Prosecutors' Ethics*, 55 VAND. L. REV. 381, 439-41 (2002) (advocating for uniform regulation of federal prosecutors through Congressional action). If Mayfield *had* turned out to be factually guilty, the misrepresentations in the affidavit could have been considered both material and entered into in bad faith, thus requiring exclusion of evidence obtained through a search authorized in reliance on the affidavit. *See, e.g.*, Franks v. Delaware, 438 U.S. 154, 164-72 (1978) (discussing requirements for accuracy in affidavits supporting warrant applications).

60. *See* Eric Lichtblau, *U.S. Will Pay $2 Million to Lawyer Wrongly Jailed*, N.Y. TIMES, Nov. 30, 2006, at A18.

61. 432 F.3d 582 (4th Cir. 2005), *overruled on other grounds*, 126 S. Ct. 978 (2006).

62. *Id.* at 584.

ing of criminal charges occurred just two days before the Administration's brief in support of Padilla's continued detention was due in the Supreme Court.[63] Luttig inferred from these facts that the government had filed the indictment to avoid Supreme Court review of Padilla's detention.[64] When conducting litigation with these stakes, so "imbued with significant public interest,"[65] the Court continued, the government should not engage in forum shopping.[66]

The Administration's tactics, the court observed, had serious institutional consequences for the Administration's credibility in the war on terror—consequences that the government, as well as its legal advisors, had underestimated in their hurried search for an expedient solution to their litigation dilemma.[67] The Court deplored the government's switching of stories regarding the basis for Padilla's detention and subsequent indictment, from a lurid plan to obtain a "dirty bomb" that would spew radiation to a more mundane effort to aid Muslim fighters in Bosnia and Chechnya.[68] The Court added the chilling view that the government had, through its dance of expedience, both damaged its own claims to detention in truly exigent circumstances and "left . . . the impression that Padilla may have been held for . . . years . . . by mistake."[69]

The ramping up of plans to try high-level terrorism suspects before military commissions will only exacerbate issues of candor with the tribunal. The Military Commissions Act ("MCA") requires that tribunals categorically exclude evidence based on torture.[70] It also mandates the exclusion of evidence obtained by

---

63. *See id.*

64. *See id.* at 585.

65. *Id.*

66. *See id.*

67. *See* 432 F.3d at 587.

68. *See id.* at 584.

69. *Id.* at 587. National security cases where the government has brought criminal charges display problems similar to the lack of candor described in the text, including the withholding of exculpatory evidence. *See generally* Koubriti v. United States, 336 F. Supp. 2d 676 (E.D. Mich. 2004) (granting Government's motion to vacate convictions on grounds because the prosecutor failed to disclose views of government experts); United States v. Wilson, 289 F. Supp. 2d 801 (S.D. Tex. 2003) (finding that court found that prosecutors may have willfully deceived the court by stating that defendant lacked government authorization for many of his activities).

70. *See* Military Commissions Act ("MCA"), Pub. L. No. 109-481, 10 U.S.C. § 948r(b)(2006). Congress enacted the MCA after the Supreme Court struck down the

methods other than torture if that evidence is not reliable.[71] Prosecuting attorneys in these cases will face enormous pressure to conceal, minimize, or misrepresent the methods used.[72] In high-profile criminal cases, such pressure is often present.[73] In the terrorism prosecutions, where the stakes include the release of individuals who appear by any calculation to be dangerous to the United States, the pressure will be very difficult to withstand.

II. *AGENCY COSTS OF DETENTION GONE AWRY: DAMAGE TO REPUTATION, CREDIBILITY, AND LEGITIMACY*

In addition to the adverse effects on lawyers' ethics and the integrity of the legal system, a misconceived detention regime can generate an array of agency costs. As defined here, agency costs are costs borne by an entity, including a country such as the United States, by the decisions of the entity's leaders. The regime of detention established after September 11 has injured U.S. interests by impairing the perceived legitimacy of U.S. actions on a global scale and weakening the system of international governance in which the United States has a principal stake.

A. *Legitimacy*

Commentators have long argued that one of the greatest attributes of the United States is its "soft power"—its ability to persuade and influence other countries through cultural, social, and political strength, without the use of force.[74] This soft

---

President's order establishing military commissions. *See* Hamdan v. Rumsfeld, 126 S. Ct. 2749 (2006); *cf.* Martin S. Flaherty, *More Real than Apparent: Separation of Powers, the Rule of Law, and Comparative Executive "Creativity" in* Hamdan v. Rumsfeld, CATO SUP. CT. REV. 51 (2005-2006) (analyzing *Hamdan*).

71. *See* 10 U.S.C. § 948r(c).

72. *See* Peter Margulies, *The Military Commissions Act, Coerced Confessions, and the Role of the Courts*, 26 CRIM. JUST. ETHICS (forthcoming 2007), *available at* http://ssrn.com/sol3/papers.cfm?abstract_id=954415 (arguing that courts should use statutory interpretation, supervisory authority, and construction of fundamental constitutional rights to address issues of coercion).

73. *See, e.g.*, New York v. Wise, 752 N.Y.S.2d 837, 845-46 (Sup. Ct. N.Y. Co. 2002) (granting prosecution motion to vacate convictions in New York's infamous Central Park Jogger case, based on government's disregard of pervasive and material inconsistencies in the alleged "confessions" of the defendants, and discovery of new DNA evidence indicating that another individual had committed the crime in question).

74. *See* JOSEPH S. NYE, JR., THE PARADOX OF AMERICAN POWER: WHY THE WORLD'S ONLY SUPERPOWER CAN'T GO IT ALONE 35 (2002) (arguing that a preemptive approach by the United States will result in the loss of "important opportunities for cooperation in the solution of global problems such as terrorism").

power hinges on perceptions that the United States acts fairly in international relations.[75] The observance of human rights and international humanitarian norms is a central element in such perceptions of fairness. When the United States signals that it takes such human rights norms seriously, the world responds, as it did during the founding of the United Nations. Pressure on the human rights front also contributed to the collapse of Communist regimes in Russia and Eastern Europe. By the same token, U.S. defection from international law norms can discredit those who seek adherence to such norms around the world, and bolster regimes that seek to oppress their own people. Ironically, a regime that disregards such norms may eventually trigger a revolution, as the case of Iran demonstrates.[76] Such drastic shifts do not serve the interests of the United States.

A studied disregard of international law also foregoes valuable opportunities to collaborate with other countries and international organizations on improvements in the international law system. International law may well be unduly idealistic in some respects, without sufficient regard for the prerogatives of individual States and the need for flexibility in the conduct of foreign affairs. If this is true, the soundest strategy is to work to reform international law by making international agreements and tribunals more sensitive to such concerns.[77] Unilateral repeal, modification, or disregard of international law short-circuits this process, impeding dialogue and legal innovation within international law. A lawyer giving advice needs to consider these opportunity costs.

---

75. *See generally* Harold Honju Koh, *On American Exceptionalism*, 55 STAN. L. REV. 1479 (2003) (exploring the idea of "American exceptionalism," or the idea that the United States is qualitatively superior to other developed nations due to its unique origins, national credo, historical evolution, and distinctive political and religious institutions).

76. *See* STEPHEN KINZER, OVERTHROW: AMERICA'S CENTURY OF REGIME CHANGE FROM HAWAII TO IRAQ 196 (2006) (observing that attempts by the United States to dislodge democratically elected regimes in Iran and elsewhere exacerbated anti-U.S. sentiment).

77. *See* Jane E. Stromseth, *New Paradigm of the* Jus Ad Bello*?*, 38 GEO. WASH. INT'L L. REV. 561, 571-72 (2006) (discussing challenges to the current paradigm on the use of force in international law, while suggesting that appropriate changes are possible within international law framework); Allen S. Weiner, *The Use of Force and Contemporary Security Threats: Old Medicine for New Ills?*, 59 STAN. L. REV. 415, 421-26 (2006) (arguing that definitions of the permissible use of force can be adopted to address the threat of terrorism).

B. *The Challenge of Finding an Exit Strategy*

The detention of suspected terrorists at Guantanamo also lacks a clear exit strategy.[78] Without a clear exit path, the Government loses control over the behavior of detainees. In addition, the lack of an exit path makes it more likely that the experience of detention will radicalize detainees, making them more likely to engage in violence if they ever are released.

Controlling any detained population is difficult with both sticks and carrots. In prison, inmates have a clear exit—they leave when their term is up. Often, they can get time off for good behavior, or at least avoid serving additional time by refraining from criminal conduct while in prison. In other facilities without fixed terms of confinement, such as psychiatric facilities, expert diagnoses hasten release. At Guantanamo, however, these incentives do not operate, since the government appears to wish to hold many of the detainees indefinitely. Accordingly, interrogators have far fewer incentives to provide to detainees for obtaining worthwhile information or ensuring good behavior.[79] The experience of being held indefinitely may also be a self-fulfilling prophecy. This experience can instill militancy where none had been present, producing an implacable foe of the United States. If radicalized individuals are eventually released, preventing violence in the future may be difficult.

Finally, the peculiar status of Guantanamo as a facility for foreign nationals also triggers international law obligations accepted by the United States that impede an exit strategy. Even if the United States wished to release the great bulk of the detainees at Guantanamo, the Convention Against Torture ("CAT")[80] would create obstacles. Under CAT, the United States cannot release detainees to a country where they are likely to be tortured. Unfortunately, the government's labeling of the detainees as terrorists increases the likelihood of bad treatment if they are returned to their country of origin. Many States practice tor-

78. *See* Diane Marie Amann, *Guantanamo*, 42 COLUM. J. TRANSNAT'L L. 263 (2004); Gerald L. Neuman, *Closing the Guantanamo Loophole*, 50 LOY. L. REV. 1, 44-53 (2004).

79. See Tim Golden, *The Battle for Guantanamo*, N.Y. TIMES MAGAZINE, Sept. 17, 2006, at 65-66.

80. *See* CAT, Dec. 10, 1984, 112 Stat. 2681, 1465 U.N.T.S. 85; *cf.* STEPHEN H. LEGOMSKY, IMMIGRATION AND REFUGEE LAW AND POLICY 1145-64 (4th ed. 2005) (analyzing non-*refoulement* obligation under CAT); Robert M. Chesney, *Leaving Guantanamo: The Law of International Detainee Transfers*, 40 U. RICH. L. REV. 657, 670-85 (2006).

ture, and the fact that a State has signed and ratified CAT is no guarantee of a commitment to avoid this practice (as U.S. citizens have discovered in the wake of revelations about Guantanamo and Abu Ghraib). Without assurances, release of detainees to their countries of origin violates international law. The result is that Guantanamo is Humpty Dumpty in reverse. When Humpty Dumpty shattered into pieces, it was impossible to put him back together. Here, in contrast, since Guantanamo has been established, it will be difficult to take it apart. A broader commitment to understanding and working within international law would have prompted greater caution in setting up Guantanamo. Since Administration policymakers and their attorneys disparaged international law, however, they were ill-situated to provide this valuable advice.

C. *Torture and Institutional Momentum*

Finally, authorizing conduct in tension with international law triggered institutional drift toward coercive interrogation as the norm, rather than the exception. Here, too, the government's policymakers and counselors were less than prescient. It is true that then White House Counsel Alberto Gonzales noted that permitting "alternative methods" of interrogation could adversely affect "military culture."[81] Gonzales's response to this concern—that the military would not back-slide from commitments to the Geneva Conventions because President Bush had "directed" them to adhere to those principles[82]—fails to recognize the mixed messages about international law conveyed by the Bush Administration. The signals sent by Gonzales himself, that the Geneva Conventions were "quaint" and "obsolete"[83] helped pave the way for the abuses at Abu Ghraib and Guantanamo. As the race to the bottom dynamic predicts, people resort to coercion when norms are ambiguous and coercive methods are expedient.[84] Coercion moves from the exception to de-

81. *See* Alberto Gonzales, *Decision re Application of the Geneva Convention on Prisoners of War to the Conflict with Al Qaeda and the Taliban*, *in* DYCUS ET AL., *supra* note 19, at 52, 54 (Supp. 2005-2006) ("a determination that . . . [the Geneva Convention] does not apply to al-Qaeda and the Taliban could undermine U.S. military culture which emphasizes maintaining the highest standards of conduct").

82. *See id.* at 55-56.

83. *Id.* at 53.

84. *See* Margulies, *supra* note 23, at 313; Louis M. Seidman, *Torture's Truth*, 72 U. CHI. L. REV. 881, 893 (2005).

fault rule. This process crowds out alternatives, such as building rapport between captor and captive, that seasoned professionals view as more effective.[85]

D. *Summary*

In sum, the institutional consequences of the President's decisions on detention and interrogation boxed the Bush Administration into largely unproductive policies. While the then White House Counsel Gonzales wrote that declining to apply the Geneva Conventions "[p]reserves flexibility" and "holds open options,"[86] the opposite is true. Coercive interrogation and arbitrary detentions at Guantanamo in fact imposed significant opportunity costs on the United States, hampering a transition to more productive and legally defensible methods.

III. *PUSHING THE ENVELOPE JUSTIFIED*

While the dangers of proceeding with a detention regime are plain, cautionary tales have a downside. Sometimes pushing the envelope is a necessary course for lawyers advising the President on national security strategy. In such cases, lawyers may appropriately advise the President to violate existing law in a fashion that is consistent with the rules of legal ethics. Such advice, however, must meet three conditions. First, the lawyers and decision-makers must display what I call a "dialogic disposition," entailing an open exchange of views before the fact or within a reasonable time with international bodies, Congress, or the courts. Second, the lawyers must consider the intermediate and long-term institutional consequences of their advice. Third, the advice must harmonize government policies with evolving norms of international or domestic law. This section discusses these criteria, and offers as examples two national security decisions with significant ramifications for international law: Roosevelt's Lend-Lease program and the Kennedy Administration's successful effort to defuse the Cuban Missile Crisis.

---

85. Moreover, the MCA accelerates the institutionalization of coercive interrogation. The MCA permits the introduction into evidence in military commissions of evidence obtained by coercion before December 30, 2005, as long as that evidence is "reliable." *See* Military Commissions Act, 10 U.S.C. § 948r (2006). Unless courts interpret the MCA to exclude such evidence, further damage will result to the United States' credibility. *See* Margulies, *supra* note 71.

86. *See* Gonzales, *supra* note 80, at 53.

A. *Dialogic Disposition*

A dialogic disposition is the first element of decisions for pushing the national security envelope. It is important for three reasons. First, dialogue is crucial for a civic humanist view that values participation for its own sake. Dialogue between lawyers, policymakers, and other relevant institutions or audiences, including Congress and international organizations, allows a multiplicity of players to offer their views as active contributors to debate. Second, it assures that a decision will be more accurate and well-founded, with a given approach exposed to light from a range of possible perspectives that counteract biases and individual agendas. Third, a decision made through dialogue is more likely to be a tailored use of power, since policymakers appreciate that tailored decisions are easier to justify.

A dialogic disposition can entail ratification after the fact.[87] A commitment to seek such ratification, however, should be part of the original decision. Moreover, an effort to secure ratification should follow the original decision in a reasonable period of time, typically six months or less. Attempts at ratification that are forced on a decision-maker and post-date the decision by a substantially longer period cannot really count as dialogue.[88]

Timely post-hoc ratification is also appropriate from a legal ethics perspective. The ethics rules permit advocates to seek good-faith modifications of existing law.[89] Since legal advisors contemplate that external audiences and institutions will have to ratify a policy, they have incorporated the transparency that the

87. *See* Oren Gross, *Chaos and Rules: Should Responses to Violent Crisis Always be Constitutional?* YALE L.J. 1011, 1108 (2003) (discussing Jefferson's view).

88. For this reason, the Bush Administration's belated efforts to secure approval for its policies on detention, coercive interrogation, and national security surveillance through the Military Commissions Act of 2006 do not meet the dialogic disposition criterion. These late entries responded to court decisions and media disclosures. The Bush Administration also clarified its views on torture in response to public pressure. *See* Memorandum from Daniel Levin, Assistant Attorney General to James B. Comey, Deputy Attorney General (Dec. 30, 2004), http://www.usdoj.gov/olc/18usc23402340a2.htm (last visited Apr. 25, 2007) (discussing legal standards applicable under 18 U.S.C. §§ 2340-2340A). The Levin Memorandum categorically rejects the use of torture. *See id.* at 1 (declaring that "Torture is abhorrent to both American law and values and international norms"). However, this categorical rejection seems inconsistent with the Levin Memorandum's claim that its conclusions regarding treatment of detainees are identical with the conclusions drawn by the earlier memos, despite the narrower definition of torture those memos advance. *Id.* at n.8.

89. *See* MODEL RULES OF PROF'L CONDUCT R. 1.2, 8.4 (1983).

ethics rules demand. While pushing the envelope may still create tension with the ethics rules, this tension is ultimately productive, leading to changes in the law that the rules recognize as both inevitable and desirable.

B. *Consideration of Institutional Consequences*

Legal advisers should also consider the institutional consequences of particular decisions. At their best, lawyers grasp not only legal doctrine but how institutions work. The rules of legal ethics encourage lawyers to offer advice on non-legal consequences.[90] Prudent legal advice should point out not only the benefits if a proposed action is successful, but also the risk of error in estimating the likelihood of success. Particularly when success hinges on the convergence of variables, the risk of error may be high. Lawyers with blind spots engendered by ideology, aspirations for career advancement, or intoxication with making an impact, may systematically overestimate the probability of success.

In the case of a proposal involving detention of national security risks outside normal channels, history provides a clear account of the risks. As we have seen, these consequences can include the erosion of the legal system's integrity, lawyers' ethics, and the ideal of equality. Such programs, as tempting as they may seem when first proposed, have substantial opportunity costs. When they involve violations of international law, they undermine the United States' credibility, and make forging international consensus more difficult. In addition, a decision to approve detention that challenges international law norms can also be difficult to reverse. Lawyers advising decision-makers must assess the difficulty of exiting from a policy, once it becomes counterproductive.

At the same time, lawyers must assess the consequences of failing to take action. When national security crises such as the destruction of railways and bridges in Maryland linking Washington, D.C. to the North emerged at the start of the Civil War, adherence to the letter of the law would have risked the entire structure of democracy and self-government. President Abraham Lincoln argued persuasively that here the long-term view argued for some temporary curtailing of habeas corpus, asking

90. *See id.* R. 2.1; Katyal, *supra* note 69, at 120-21.

in his message to Congress, "are all the laws but one [habeas] to go unexecuted, and the government itself go to pieces, lest that one be violated?"[91] One can view Lincoln's suspension of habeas corpus at this place and time as a narrowly tailored response to an existential threat, in which the institutional costs of inaction outweighed the costs of decisive measures to contain the Maryland insurrection.[92]

Moreover, Lincoln was discerning in the timing of the suspension, holding it in abeyance until after the Maryland legislature considered a secession vote. While others, including General Winfield Scott had urged the arrest of secessionist Maryland legislators, Lincoln thought better of this strategy. First, Lincoln noted that the legislators had a "clearly legal right to assemble."[93] Lincoln also noted that the remedy of suspension would only complicate the challenging political situation, observing that, "we can not permanently prevent their action. If we arrest them, we can not long hold them as prisoners; and when liberated, they will immediately re-assemble, and take their action."[94] Through clear-headed insight into institutional consequences, Lincoln appreciated that suspension of habeas corpus, whatever its virtues in dealing with the precarious military situation in the early days of the Civil War, would never be a solution to the political challenges faced by his Administration.[95]

Unfortunately, this insight did not prevent Lincoln's Administration, with his tacit or active consent, from using suspension of habeas corpus as an expedient through the rest of the conflict.[96] Through the rest of the war, Lincoln's administrators and generals used habeas corpus readily to arrest and detain approximately 13,000 people,[97] often without any official indication that these individuals were disloyal or plotting violence.[98]

---

91. *See* NEELY, *supra* note 35, at 12.

92. *See* DANIEL FARBER, LINCOLN'S CONSTITUTION 16-17 (2003); Frank J. Williams, *Abraham Lincoln and Civil Liberties: Then and Now—The Southern Rebellion and September 11*, 60 N.Y.U. ANN. SURV. AM. L. 463, 466 (2004).

93. *See* NEELY, *supra* note 35, at 6.

94. *See id.* at 7.

95. *See id.* ("Suspending the write of habeas corpus was not originally a political measure, and it would never become primarily political.").

96. *See* Sanford Levinson, *Constitutional Norms in a State of Permanent Emergency*, 40 GA. L. REV. 699, 718 (2006).

97. *See* NEELY, *supra* note 35, at 23.

98. *See id.* at 20-21.

Indeed, the ready availability of arrest and detention, more than any conduct by those actually arrested and detained, accounted for its use. Lincoln seemed to exhibit little interest in stopping these adverse institutional consequences, even when they provided fodder for his opponents.[99]

C. *Harmonization With Evolving Norms*

Policymakers at certain crucial junctures in U.S. history have defied the letter of the law to promote equality, dignity, and nonaggression. A purposive style of interpretation drives these decisions, premised on the goals served by constitutional or international law. Examples include: the protection of human dignity in the Emancipation Proclamation,[100] safety from aggression, as in President Franklin D. Roosevelt's Lend-Lease program, or the use of tailored and limited force to prevent a wider conflict, as in the Kennedy Administration's approach to the Cuban Missile Crisis. While each of these measures could also claim a pragmatic justification, each represented a conscious break with the past. Moreover, despite the tension triggered with existing norms, each decision vindicates the rule of law.

The law must reckon with evolving norms because societies and circumstances change. Sometimes values integral to the founding of an entity become submerged under the weight of popular fears, sectarian interests, or bureaucratic in-fighting. In such situations, a return to first principles is essential. Both Lincoln and Frederick Douglass, for example, understood that the struggle against slavery was about reconciling the ideals of the Declaration of Independence with an evolving vision of the Constitution's."[101] The renewed founding embodied in this return

---

99. *See id.* at 18 ("the impact of the [subsequent arrest of legislators] on later Maryland elections is difficult to determine, but they were more likely harmful than helpful to the Administration's cause by supplying an issue to the opposition").

100. *See* Sanford Levinson, *Was the Emancipation Proclamation Constitutional? Do We/ Should We Care What the Answer Is?*, 2001 U. ILL. L. REV. 1135, 1142-43; *see also* Michael Stokes Paulsen, *The Emancipation Proclamation and the Commander in Chief Power*, 40 GA. L. REV. 807, 814-23 (2006) (defending Emancipation Proclamation as legitimate exercise of presidential authority in time of war).

101. *See* MILNER S. BALL, THE WORD AND THE LAW 146-49 (1993) (discussing Frederick Douglass' views); MARK NEELY, THE LAST BEST HOPE OF EARTH: ABRAHAM LINCOLN AND THE PROMISE OF AMERICA 154 (1995) (discussing the rhetorical strategies at play in

to first principles is not a rejection of the rule of law, but a necessary step in affirming the rule of law's continued relevance.

Both legal ethics and international law recognize the importance of change in the law. Model Rule of Professional Conduct Rule 1.2 permits lawyers to challenge existing law, not only directly through litigation, but also indirectly through legal advice to groups engaging in civil disobedience. International law, elaborates and augments core principles in the formation of customary international law. While fundamental norms, such as the prohibition on torture, are *jus cogens* and therefore inviolable, other values and applications flow from an accretional process that reflects actions by States and tribunals, as well as surveys of the landscape by learned students of the process. Moreover, although detecting emerging norms is not always easy, the Supreme Court has indicated that courts have the competence to ascertain emerging international norms.[102] If courts have this power, lawyers certainly have the aptitude to make similar calls.

D. *Lend-Lease*

As one example of a national security decision that meets the above criteria, consider the Lend-Lease program. In Lend-Lease, President Franklin D. Roosevelt agreed, prior to the United States' entry into World War II, to send U.S. destroyers to Britain in exchange for a commitment by the British to lease bases in the Caribbean to the United States. Roosevelt made the agreement with British Prime Minister Winston Churchill without prior consultation with Congress.[103] Despite the utility of the agreement in holding Nazi Germany at bay, Lend-Lease was inconsistent with both statutory and international law.

---

the Gettysburg Address); Peter Margulies, *Progressive Lawyering and Lost Traditions*, 73 TEX. L. REV. 1139, 1177-78 n.228 (1995).

102. *See generally* Sosa v. Alvarez-Machain, 542 U.S. 692 (2004) (discussing process of adjudication under Alien Tort Statute, which allows plaintiffs to seek relief in United States courts for violations of the "law of nations").

103. *See* ROBERT DALLEK, FRANKLIN D. ROOSEVELT AND AMERICAN FOREIGN POLICY 1932-45, at 246-47, 256-60 (1979) (discussing chronology of Lend-Lease agreement, including post-agreement approval and appropriations by Congress); *cf.* ROBERT H. JACKSON, THAT MAN: AN INSIDER'S PORTRAIT OF FRANKLIN D. ROOSEVELT 93-103 (John Q. Barrett ed., 2003) (providing account by Roosevelt's Attorney General, later Supreme Court Justice). Technically, the agreement with Churchill that preceded congressional authorization is called the "destroyer deal." *Id.* at 81-82. The author uses the term "Lend-Lease" throughout for the reader's convenience.

Then Attorney General Robert Jackson's opinion supporting the Lend-Lease program does not resolve these inconsistencies. Federal statutes passed by an isolationist Congress barred the conveyance of material "essential" to U.S. defense. In addition, both international law and a federal statute (the Espionage Act) prohibited provision of material by the supposedly neutral United States to a belligerent. On the question of whether the destroyers were essential to United States defense, Jackson basically changed the subject, arguing that the leasing of British bases would be a net security plus. On the statutory and international obligations that neutral status imposed on the United States, Jackson argued that the United States could not send material to a belligerent that had been built expressly to assist that party but could send material built for another purpose.

Neither of Jackson's arguments stands up to scrutiny. On the question of whether material was "essential," while Jackson's argument about net benefits is resourceful, there was no evidence that Congress contemplated aggregating costs and benefits as Jackson outlined. There was, however, ample evidence that Congress wished to avoid moves that might yield foreign entanglements. On the implications for neutrality of sending material to belligerents, Jackson's distinction between material built for that purpose and material built for another purpose but subsequently converted into aid to a belligerent seems sophistic at best.[104]

Viewed in this stark light, Jackson's opinion appears to counsel the willful evasion, if not outright defiance, of both international and domestic law. Jackson clings tenuously to the vine of subjective intent, arguing in essence that his belief compensates for the lack of reasonable support for his position. Lack of support also permits an inference that the lawyer, particularly a well-placed government lawyer, with access to all advice, did not actually believe that the action was lawful. Jackson also remained silent while Roosevelt, not in an actual court but in the

104. *See* Aaron Xavier Fellmeth, *A Divorce Waiting to Happen: Franklin Roosevelt and the Law of Neutrality, 1935-41*, 3 BUFF. J. INT'L L. 413, 473-80 (1996-1997); *cf.* U.S. Attorney General, Acquisition of Naval and Air Bases in Exchange for Over-Age Destroyers, 39 Op. Att'y Gen. 484, 486-88 (1940) (relying on *United States v. Curtiss-Wright Export Corp.*, 299 U.S. 304, 319-22 (1936)). The reliance of Jackson on *Curtiss-Wright*, which also supplied the clincher for Yoo's arguments in the Torture Memos, suggests that Jackson was less than confident in his statutory arguments.

court of public opinion during an election year, was at best less than candid about the existence of an agreement with Churchill.

Three factors make this stance appropriate. First, Roosevelt insisted on a far-reaching debate within his administration on the legality of the program, actively encouraging skepticism about whether Lend-Lease would be consistent with Congress's commands. Roosevelt also sought congressional authorization within six months of the agreement (although after the election). After powerful public statements by Roosevelt, including his memorably simple comparison of Lend-Lease with the loan of a garden hose to a neighbor whose house was on fire, Congress gave its approval.

Second, Jackson's position is clearly consistent with evolving international norms. The law of neutrality, for example, seemed also painfully irrelevant to the crucible of World War II. It did nothing, for example, to prevent the wholesale slaughter that the Nazis were preparing for Jews and other groups. In this sense, the law of neutrality obstructed realization of ideals that are essential in a legitimate world order, including freedom from force and want.[105] Indeed, the announcement of the Atlantic Charter by Roosevelt and Churchill demonstrated that both men were committed to a new international regime prior to the U.S.'s entry into the war.[106] After the war, the formation of the United Nations renewed commitments to comprehensive norms such as the prohibition of aggression. Fulfilling this vision required resistance to the Axis Powers, with their dreams of global domination. Roosevelt and Jackson grasped that fact, even as they discounted the letter of the law.

Third, institutional consequences support Roosevelt and Jackson's view. Disclosure of the full scope of the Lend-Lease program during the 1940 election could have roiled the voters, and encouraged posturing by congressional leaders, as well as a possible filibuster by committed isolationists. In this charged environment, the initiative may have withered on the vine, with significant adverse effects on British confidence and the war effort. By controlling the timing, Roosevelt waited until the moment

105. *See* DALLEK, *supra* note 102, at 8-9 (discussing formation of Roosevelt's views as Assistant Secretary of the Navy in Wilson Administration); *cf.* JACKSON, *supra* note 103, at 103 (describing law of neutrality as "obsolete.").

106. *See id.* at 282-85.

was ripe politically. In the process, Roosevelt and Jackson got some grime on their hands. Their maneuvering, however, avoided a greater ethical failure, the crime of doing nothing.[107]

E. *The Cuban Missile Crisis*

As another example that meets the criteria for pushing the envelope, consider the Cuban Missile Crisis faced by the Kennedy Administration. President Kennedy and his advisors, including his brother and Attorney General, Robert Kennedy, were caught between the U.S. military's pressure for an aggressive policy and the prohibition in international law of the unjustified use of force. The response formulated by Robert Kennedy, aided by a small phalanx of elite legal advisors, almost certainly violated the letter of international law, but it also avoided a larger conflict. Moreover, the purposive approach adopted by President Kennedy also reflected an effort to take law seriously that current national security strategists would do well to emulate.

As students of national security policy know, the crisis began when the Kennedy Administration learned in October, 1962 that Soviet nuclear missiles placed in Cuba were offensive in nature, designed for a first strike on U.S. cities. The military wished to attack Cuba to destroy the threat. There were, however, two significant problems with the attack option. First, it risked all-out nuclear war. Second, it would have violated international law.

The second problem arose because an attack would not meet the test of The Caroline, Daniel Webster's framing of a State's right to use force in self-defense,[108] or of Article 51 of the U.N. Charter,[109] which arguably codifies the customary international law principle articulated in Webster's letter. Under this test, a State can use force to prevent an imminent attack. Since it was not clear that the Cubans would use the missile imminently—or indeed at all—the United States could not meet this test.

---

107. *See* Walzer, *supra* note 2, at 61-75.

108. Secretary of State Daniel Webster wrote that a nation may use force in SELF DEFENSE when there exists "[a] necessity of self-defence, instant, overwhelming, and leaving no choice of means and no moment for deliberation." *See* Daniel Webster, Letter from Daniel Webster to Henry Fox (Apr. 24, 1841), *in* THE PAPERS OF Daniel Webster (Alfred S. Konefsky & Andrew J. King eds., 1982).

109. *See* U.N. Charter art. 51.

Tellingly, Administration lawyers considered not merely the abstract principle, but the institutional consequences that would have emanated from the use of force in this context. Robert Kennedy warned that other States and history itself would perceive the United States unfavorably if we emulated the aggression of the Axis Powers during World War II. He insisted that an attack would amount to "Pearl Harbor in reverse."[110]

To reconcile legal concerns with the need to take decisive action that would lead to removal of the Soviet missile, the lawyers articulated a "quarantine" argument justifying a limited naval blockade of Cuba.[111] The blockade also appeared inconsistent with international law barring the unjustified use of force.[112] While a limited blockade targeting Soviet vessels was a more proportionate response to the threat posed by the missiles than an all-out attack, the blockade nevertheless involved the application of military power against another sovereign State. Moreover, if the threat posed by the missiles was not imminent, than any use of force was unjustified under international law.[113]

The quarantine approach ultimately fares better under the test for pushing the envelope. First, a purposive approach to international law suggests that the quarantine approach harmonizes effectively with evolving norms. Article 2(4) of the U.N. Charter prohibits the use of force only "against the territorial integrity or political independence of any state, or in any manner inconsistent with the purposes of the United Nations."[114] One can argue that this qualifying language permitted U.S. policymakers a small window for the limited blockade that President Kennedy imposed.[115] Moreover, the quarantine approach

---

110. *See* ROBERT F. KENNEDY, THIRTEEN DAYS: A MEMOIR OF THE CUBAN MISSILE CRISIS 9 (1971) (recounting Kennedy's passing a note to his brother, the President, after listening to arguments for an air attack on Cuba, that said, "I now know how Tojo felt when he was planning Pearl Harbor"); Evan Thomas, *Bobby at the Brink*, NEWSWEEK, Aug. 14, 2000, at 49, 51.

111. *See* Richard N. Gardner, *Future Implications of the Iraq Conflict: Neither Bush nor the "Jurisprudes,"* 97 AM. J. INT'L L. 585, 587-88 (2003).

112. *See* U.N. Charter art. 2, ¶ 4.

113. *See* ABRAM CHAYES, THE CUBAN MISSILE CRISIS: INTERNATIONAL CRISES AND THE ROLE OF LAW 25-40 (1974).

114. *See* U.N. Charter art. 2, ¶ 4.

115. *See* Louis Henkin, *Comment*, *in* CHAYES, *supra* note 113, at 149, 152-53. Henkin's skepticism about a broader authorization for "anticipatory self-defense" under Article 51 lends credibility to his measured approval of the quarantine approach. *Id.* at 150.

represented a determined effort by the world's greatest power to limit force and promote a negotiated outcome. On the level of practice and symbolism, the self-restraint practiced by the United States nurtured values at the heart of international law.

Second, the United States showed a dialogic disposition throughout the crisis. The Administration focused intently on gaining assent from our allies in the region, through the Organization of American States ("OAS").[116] President Kennedy also submitted the problem to the United Nations, which knew of the quarantine but took no action. Kennedy consulted in this fashion, although matters were exigent, time was short, and the United States faced the single gravest crisis of the post-World War II period.[117] The ultimate resolution of the crisis, which also hinged on an unspoken bargain by the United States to remove missiles from Turkey that threatened Russia,[118] similarly demonstrates this commitment to dialogue.[119]

Finally, institutional consequences were manageable, at least compared with alternatives. A limited blockade authorized by the OAS provided legal and political cover for the Administration, and placed the Soviet Union on the defensive in the court of international public opinion. Limiting the use of force reduced—although it did not eliminate—the prospect of nuclear war. In contrast, doing nothing about the missiles would have eroded political support for the Administration, and generated momentum for a more extreme military response.[120] A straightforward swap of Soviet missiles in Cuba for U.S. missiles in Turkey would have allowed Russia to claim control of the interna-

116. *See* CHAYES, *supra* note 113, at 41-68.

117. *See id.* at 3 (quoting sources suggesting that President Kennedy believed the risk of nuclear war ranged from thirty-three to fifty percent).

118. *See id.* at 94-100.

119. The decision of the United States, with approval of the United Nations, to intervene militarily in Afghanistan after September 11 presents an even stronger case for legality. *See* Kenneth Watkin, *Controlling the Use of Force: A Role for Human Rights Norms in Contemporary Armed Conflict*, 98 AM. J. INT'L L. 1, 4-7 (2004); *cf.* Mark A. Drumbl, *Victimhood in Our Neighborhood: Terrorist Crime, Taliban Guilt, and the Asymmetries of the International Legal Order*, 81 N.C. L. REV. 16-35 (2002) (arguing intervention could be viewed under international law as legitimate use of self-defense against armed attack).

120. *See* CHAYES, *supra* note 113, at 31 (noting that legal advisor from the State Department "could not counsel passivity"); *cf.* GRAHAM T. ALLISON, ESSENCE OF DECISION: EXPLAINING THE CUBAN MISSILE CRISIS 58 (1971) (discussing reasons for rejecting "do nothing" option).

tional strategic agenda in a fashion that could have prejudiced U.S. interests.[121] Particularly since the do-nothing, straightforward swap options commanded virtually no support among the significant players, pushing the envelope with the quarantine approach was the most appropriate response for legal advisors to the President.

CONCLUSION

National security lawyers face a challenging task. They regularly encounter situations where pushing the envelope of international or domestic law seems expedient, desirable, or even necessary. In some cases, particularly those involving the authorization of regimes of detention or interrogation, resisting this temptation is typically the best way to serve the client. History tells us, in the *Korematsu* litigation and in the Bush Administration's establishment of Guantanamo, that pushing the envelope in this area can have deeply problematic results. Discounting or disregarding international and domestic norms can erode the integrity of the legal system, lawyers' ethics, and the credibility of the United States around the world.

A national security lawyer, however, cannot rigidly oppose pushing the envelope. While such a course should never be entered into lightly, necessity may dictate taking this path. The lawyer's guideposts in this uncertain realm, where legal doctrine and statecraft meet, should be the importance of dialogue, institutional consequences, and harmonization with evolving norms. Decisions such as the Emancipation Proclamation, Lend-Lease, and the response to the Cuban Missile Crisis meet these criteria. The United States, and arguably the world, benefited from lawyers and policymakers who pushed the envelope in those exigent circumstances. More recent events, such as the U.S. military intervention in Afghanistan after September 11, are cut from the same cloth. Knowing when to push the envelope is the central responsibility of the national security lawyer; this Article has offered some modest ground rules for the effort.

---

121. *See* ALLISON, *supra* note 118, at 58-59.