Cutter IT Journal

The Journal of
Information Technology Management

Vol. 19, No. 1
January 2006

"Since it is private industry that owns and operates the vast majority of what constitutes the Internet, it is therefore industry's responsibility to demonstrate leadership in the fight to secure cyberspace."

— Larry Clinton,
Guest Editor

Securing Cyberspace:

Is It Time to Rethink Our Strategy?

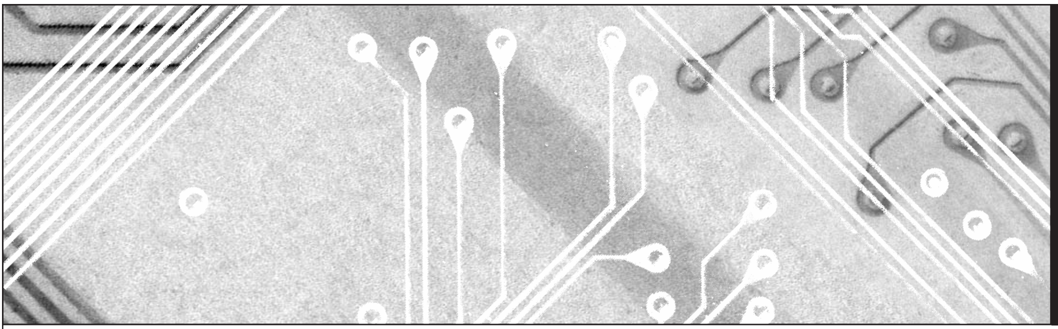
The Government Must Step In to Secure Cyberspace

National security demands a heightened emphasis on corporate information security. Companies that assume that cyber issues are low priority are whistling past their own future graveyard.

Cyber Security Really Is Not an IT Problem

Despite the obvious IT aspect of cyber security, it is the business managers, not the technologists, who must do the real heavy lifting to protect the Internet infrastructure and the information it carries.

CUTTER CONSORTIUM



About Cutter IT Journal

Part of Cutter Consortium's mission is to foster the debate of, and dialogue on, the business technology issues challenging enterprises today, to help organizations leverage IT for competitive advantage and business success. Cutter's philosophy is that most of the issues that managers face are complex enough to merit examination that goes beyond simple pronouncements. Founded in 1987 as *American Programmer* by Cutter Fellow Ed Yourdon, *Cutter IT Journal* is one of Cutter's key venues for debate.

The monthly *Cutter IT Journal* and its weekly companion *Cutter IT E-Mail Advisor* offer a variety of perspectives on the issues you're dealing with today. Armed with opinion, data, and advice, you'll be able to make the best decisions, employ the best practices, and choose the right strategies for your organization.

Unlike academic journals, *Cutter IT Journal* doesn't water down or delay its coverage of timely issues with lengthy peer reviews. Each month, our expert Guest Editor delivers articles by internationally known IT practitioners that include case studies, research findings, and experience-based opinion on the IT topics enterprises face today — not issues you were dealing with six months ago, or those that are so esoteric you might not ever need to learn from others' experiences. No other journal brings together so many cutting-edge thinkers or lets them speak so bluntly on issues such as:

- Proving the value of IT ROI
- The give and take of offshore outsourcing
- The value of high-quality data
- The evolution of agile project management
- When and how to kill a dying project

Cutter IT Journal subscribers consider the Journal a "consultancy in print" and liken each month's five or six articles exploring a single topic to the debates they participate in at the end of a day at a conference — with every participant in the conversation forcefully expressing his or her viewpoint, backing it up with data and real-life experience.

Every facet of IT — application integration, security, portfolio management, and testing, to name a few — plays a role in the success or failure of your organization's IT efforts. Only *Cutter IT Journal* and the *Cutter IT E-Mail Advisor* deliver a comprehensive treatment of these critical issues and help you make informed decisions about the strategies that can improve IT's performance.

Cutter IT Journal is unique in that it is written by IT professionals — people like you who face the same challenges and are under the same pressures to get the job done. The *Journal* brings you frank, honest accounts of what works, what doesn't, and why.

Competition remains intense in the high-tech world. Budget cutbacks and short deadlines have become the norm. You need advice and experience you can rely on. Put your IT concerns in a business context. Discover the best ways to pitch new ideas to executive management. Ensure the success of your IT organization in an economy that encourages outsourcing and intense international competition. Avoid the common pitfalls and work smarter while under tighter constraints. You'll learn how to do all this and more when you subscribe to *Cutter IT Journal*.

Opening Statement

by Larry Clinton

In the call for papers for *Cutter IT Journal*'s edition on cyber security, we asked, "Is it time to change our strategy?" Apparently, the answer is a resounding "yes."

Nearly every article we received (enough to necessitate a second edition on this topic to be published in August) called for fairly fundamental shifts in strategy to address a problem that most evidence suggests is growing like a malignant tumor — and which we may understand far less how to treat.

We received submissions from senior US government officials, major universities, think tanks, and corporations spanning Europe, Asia, and North America. The articles dealt with a wide array of issues, ranging from the macro concerns regarding how terrorists might use cyber warfare to inflict severe physical pain on large populations, down to more micro issues such as newly emerging forms of spam over wireless and VoIP technology. However, despite the breadth of contributors and topics covered, virtually every author essentially echoes the view offered in our lead article by Bob Stephan, US Department of Homeland Security's (DHS) Assistant Secretary for Infrastructure Protection: "In today's world, the traditional security paradigm is shifting to encompass the unique challenges presented by the digital landscape."

In addition to the notion that a new paradigm would be needed to address the lack of cyber security, there was broad consensus, although not unanimity, on several key points:

- The problem is alarming and growing.
- The very nature of the Internet, and the constant change surrounding its maturation, makes the problem of securing cyberspace substantially more difficult.
- The problem cannot be viewed simply as a "tech" problem, although that is how it has been and continues to be largely characterized.
- We don't have a solution yet, but when we come up with one, it will need to be multifaceted and international.

While there is no one solution offered — indeed the authors in this issue identify a multitude of problems at various levels of abstraction — each of the articles does offer concrete advice to guide the reader in addressing the issues raised herein.

WE NEED TO BE TOGETHER — BUT WE'RE NOT

Over the past few years, we have heard an ever-louder drumbeat in the popular press, and from some politicians, for stricter government regulation of the Internet as a response to the problem of information security. My own organization, the Internet Security Alliance, takes a somewhat different view by maintaining that since it is private industry that owns and operates the vast majority of what constitutes the Internet, it is therefore industry's responsibility to demonstrate leadership in the fight to secure cyberspace. This philosophy should not be confused with a laissez-faire approach that expects the invisible hand of the market to magically remedy the serious problems we are facing. Quite the contrary: the inherent characteristics of the Internet demand a far more intensive and creative effort to ameliorate the situation than will likely be accomplished via regulatory fiat.

In fact, when I asked DHS Assistant Secretary Stephan to contribute to this edition of *Cutter IT Journal*, I did so because I knew he would reflect the critical perspective that the cyber security issues we all face cannot be addressed unilaterally by either government or industry. We need the sort of new, creative government-industry partnership Stephan speaks of in his article. Unfortunately, we are only at the preliminary steps of creating such a functional working relationship.

There is a great deal to commend in Stephan's article and his work at DHS to increase the profile of cyber security issues and integrate them into the broad effort to assure homeland security. Stephan correctly points out that causing cyber mischief has never been easier, or more threatening, and that the traditional models for predicting incidents is no longer reliable in the current environment. He argues that we must create a coordinated risk management approach to bring coherence to the cyber security efforts of the US federal government, local governments, international partners, and industry worldwide. He outlines the approach DHS is currently suggesting via the draft *National Infrastructure Protection Plan* (NIPP), which is currently being circulated to interested parties for comment.

Stephan and his DHS colleagues are, quite appropriately, motivated by the vital need to ensure national defense. Perhaps it is unfortunate, but the common good is not what motivates the private sector when it comes to allocating resources for the portions of the cyber infrastructure it must defend. Even the most patriotic of corporations make investment decisions based on their corporate self-interest, and hence no plan for constructing a comprehensive cyber defense can be complete if it does not directly address the economic motivations affecting industry.

WHAT WE DON'T KNOW IS A LOT

Our second article, by Shari Lawrence Pfleeger, Rachel Rue, Jay Horwitz, and Aruna Balakrishnan of the RAND Corporation, tees up this issue nicely. In their abstract, the authors began with the troubling, but accurate, observation that "even when the technology to secure cyberspace is available and reasonably priced, individuals and corporations don't always use it."

In their article, Pfleeger et al. review the available literature to see what we actually know about the economics of cyber security. The short answer is, "not all that much." They note that little effort has gone into creating a method to accurately account for the direct and indirect costs of cyber incidents. As a result, both industry and government are left without a solid basis for making informed decisions.

The authors also report on a series of studies of interest to people (like myself) who believe that until the dynamism of the market is interwoven systematically into an incentive program to encourage private-sector investment in cyber defense, we're not going to get very far. For example, they cite studies on the positive influence of information sharing as well as studies that suggest creative mechanisms such as the use of "vulnerability markets" (similar to the market in pollution credits in the US) or more aggressive use of cyber insurance. We expect to review some of these topics in greater detail in our second installment later this year.

PLUG AND PLAY — NOT

The Software Engineering Institute's Carol Woody moves us from the realms of broad public policy paradigms and research needs down to the actual operational issues organizations face in maintaining secure systems. Like her RAND colleagues, she begins with a troubling observation: "Great authorization and authentication will not save a system from poorly planned implementation and sustainment. ... After implementation is when the security problems will show up. The developers move on, and the operational and sustainment support staff do the best they can."

The major problem Woody focuses on is the increasingly complex security management issues generated by organizations' reliance on COTS software. Organizations like to think that COTS components are "mature, stable, and adhere to well-recognized industry standards." She suggests that the reality of systems built using such components is more of "a Rube Goldberg mix of glue code."

Adding to the problem, Woody argues, is the virtually constant changes necessitated by the various vendors' individualized update schedules. Updates can trigger a complete restructuring of the system, placing the organization in a "damned if you do and damned if you don't" situation. If you apply the updates, you have to constantly monitor their interaction (and overhaul your system if it "breaks"); if you don't, you may face security problems and even the loss of maintenance support.

Even more insightful is Woody's analysis of how typical management issues — reliance on current ROI models, the trend toward outsourcing, and organizational structures geared to legitimate but non-security concerns — complicate an already dicey problem.

PHISHING, PHARMING, SPAM, SPIM, AND SPIT

Our last three pieces bring us directly into operational issues at very concrete levels.

Lee Imrey of the US Department of Justice provides some compelling math that explains why phishing has become one of the fastest-growing crimes of the past year.

He also notes almost the exact same problem Woody identified in the previous article — namely, the irresistible force of time in our fascinating and rapidly changing industry and how difficult that makes things. He observes, "Unfortunately, the detective measures currently on the market face the same challenges that antivirus vendors have faced for years. ... By the time any solution gets to market, there is a group of dedicated, well-funded, and highly motivated professionals working to defeat that solution." As a result, "Entire industries are chasing their tails in trying to resolve this problem."

In keeping with the notion of new paradigms, Imrey suggests that we would be better off thinking of phishing not as the problem to be solved, but as a symptom of a larger issue, namely the reliance on old techniques for authenticating users' identities when more enlightened and effective techniques may be near at hand.

Accompanying Imrey's article is a sidebar offered by M. Vidyasagar of Tata Consultancy Services in India. Vidyasagar offers a case study of how his large and growing corporation addressed the need to rethink its security model as part of a fully integrated Enterprise Management Program to address Tata's rapidly evolving business and technological needs. A significant aspect of this security overhaul was the implementation of just the sort of time-synchronized two-factor authentication Imrey advocates.

Finally, Charalampos Patrikakis of the National Technical University of Athens, Greece, and Anastasios Pallas of the University of Paisley, Scotland, give us a glimpse at the future of spam. It's not a pretty picture. As if it weren't bad enough that the ubiquitous problem of spam is continuing to grow both in size and cost, now the scourge is mutating. As Imrey reported with regard to phishing, spam (and its ugly stepsisters SPIM and SPIT) always seems to be one step ahead of the sheriff. They write, "Although antispam legislation is advancing, the spam phenomenon is always one step ahead, mutating into new forms of annoyance over mobile communications, instant messaging, Internet telephony (e.g., VoIP), and spoofing of Internet search engines."

The authors provide not only a concise description of the emerging spam mutants but also a wide-ranging list of countermeasures that might be deployed to fend off the pests. Interestingly, the authors conclude their analysis by noting how difficult the regulation of spam has proved to be, and in their last line they ask: "Isn't it time to rethink our strategy?"

I think this is where we started.

So in sum, what do we learn from this edition of *Cutter IT Journal*? My answer is: "Quite a bit, but not nearly enough." Essentially, we provide a brief survey course outlining the public policy inadequacies we face, the research inadequacies we face, the inappropriate managerial structures we have to deal with, and the evolution of mechanisms to undermine our ability to use the Internet as we desire and, potentially worse, to undercut the public's confidence in it.

The problems we touch on here are daunting, but not unsolvable. If we can build a personal computer, if we can figure out how to put one of these miraculous machines on virtually everyone's desk, if we can connect them all worldwide and come up with an instantaneous and user-friendly system for global communications — then we can solve these problems.

But we need a lot of new thinking. There is a lot of work to be done. Be sure to join us in August as we do more of this new thinking and more hard work on cyber security.

Larry Clinton is the COO of the Internet Security Alliance (ISAlliance), a leader in advocating market-based systems for improving information security. In the US Congress, Mr. Clinton served as cochair of the US congressionally appointed CISWG on market incentives, which developed recommendations to encourage better corporate security without federal mandates. Mr. Clinton testified before Congress on this program in April 2005. He also sits on the board of the National Cyber Security Partnership (NCSP), the Internet Education Foundation, and the US Congressional Internet Caucus Advisory Committee, and chairs the NCSP Committee on Incentives for Improved Corporate Security. In addition to publishing and testifying on cyber issues, Mr. Clinton has appeared on C-SPAN, MSNBC, and CNBC to discuss information security. Prior to joining ISAlliance, Mr. Clinton was with the US Telecom Association (USTA) for 12 years, including the last six as a VP. Before joining USTA, Mr. Clinton was a Legislative Director in the House of Representatives and consulted for a variety of industries.

Mr. Clinton can be reached at ISAlliance, 2500 Wilson Blvd., Arlington, VA 22201, USA; E-mail: lclinton@isalliance.org.

Cyber Risk Management: The Need for Effective Public and Private Partnership

by Robert B. Stephan

UNDERSTANDING THE CYBER THREAT

In today's world, the traditional security paradigm is shifting to encompass the unique challenges presented by the digital landscape. IT and cyber infrastructure are critical to national economies and homeland security. Cyber systems support everything from food distribution to financial transactions to national security. This reliance on cyber systems extends across the public sector to the private sector and individual citizens.

Cyber infrastructure has no boundaries, and threats often do not conform to traditional models. Because of increasing dependencies on cyber systems, the consequences of a successful cyber attack on individuals, businesses, and/or government agencies could be devastating. The US Department of Homeland Security (DHS) believes that the implementation of a risk management approach to assess risk, prioritize assets, and execute protective measures is critical to securing the nation's cyber infrastructure.

Individual businesses, the industries they operate in, and government agencies depend on cyber infrastructure in different capacities, from reliance on the Internet to the use of automated digital control systems.[1] Most businesses rely on the Internet and e-mail as a primary form of communication, while others rely on the Internet operationally to transact business with customers and suppliers, such as taking and placing orders and/or monitoring inventories. In addition, control systems are widely used across the private sector to operate chemical plant processes, refine petroleum products, produce and distribute electricity, transport oil and gas, and process food. Government agencies utilize control systems to treat drinking and waste water, manage locks and dams, and process mail.

A cyber attack can launch from anywhere in the world through the Internet and potentially have mass effect with little or no warning. Cyber infrastructure is not limited by geographical, organizational, or political boundaries, nor is it limited to only one machine, network, or organization. Attack tools and methodologies are becoming more sophisticated, easier to use, and widely available from point-and-click Web sites. Five to 10 years ago, these capabilities were resident in only the most advanced economies. Today, the proliferation of these sophisticated and readily available tools and methods places the capability to execute a successful cyber attack within reach of malicious actors with low to moderate computer skills and economic resources.

For example, the Sapphire worm (aka "Slammer") hit the cyber infrastructure on 25 January 2003. It exploited a vulnerability in Microsoft SQL server, multiplied itself rapidly, and soon spread to networks around the globe. Within 10 minutes, Sapphire had infected 75,000 machines [3]. The primary impact was network saturation due to bandwidth demands, which caused significant disruption to financial institutions (ATMs failed) and transportation (airline flights were cancelled). Worldwide, productivity losses cost the public and private sectors an estimated US $950 million-$1.2 billion [4]. Although Microsoft created and distributed a patch to prevent disruptions from the Sapphire worm, it still had a significant impact because system owners and operators did not install the patch consistently across their networks.

A malicious actor might have numerous motives for promulgating a cyber attack, including criminal intent for economic gain or damage, economic or international espionage, revenge, publicity, and/or a desire to reduce

[1] The draft *National Infrastructure Protection Plan* defines a control system as "an interconnection of components (designed to maintain operation of a process or system) connected or related in such a manner as to command, monitor, direct, or regulate itself or another system.... Examples include the management of the electric power grid using supervisory control and data acquisition (SCADA) systems within the energy sector and process control systems (PCS) that control the timing and volume of chemical processes within the chemical sector" [6].

the public's confidence. Of particular concern is an attacker's intent to inflict physical consequences through a cyber attack (such as an attack on control systems) or through a combination of cyber and physical attacks.

Malicious actors that attack cyber infrastructure range from individual hackers to criminal organizations to nation states. Some of the most dangerous attackers are those who are considered insiders to the organization they are attacking. These insiders have not only authorized access to systems, but also the institutional knowledge of how to be most effective in an attack capacity. In March 2002, a disgruntled employee who had recently quit an international financial services company planted malicious code that deleted 10 billion data files. The attack affected over 1,300 of the company's servers in the US and cost the firm approximately $3 million to repair the damage and reconstruct the files [5].

The traditional threat model of relying on historical incidents to predict the likelihood of future attacks is not as accurate for calculating cyber attacks. In partnership with the intelligence community and law enforcement agencies, DHS is working to more fully understand the cyber threat both domestically and overseas. The department is working to prevent damage or exploitation, as well as supporting the prosecution of those malicious actors who attempt to damage or exploit US cyber infrastructure. As threats continue to proliferate and mitigation strategies become more complex, there is increased attention on securing cyber infrastructure. With a multitude of individuals and organizations in the public and private sectors striving to secure the country's cyber infrastructure, there are opportunities for coordination and leadership among the various stakeholders. DHS is leveraging the growing momentum to help drive public- and private-sector efforts toward this common goal.

UNDERSTANDING CYBER DEPENDENCIES, VULNERABILITIES, AND CONSEQUENCES

In addition to understanding the potential threat landscape as it applies to cyber infrastructure, it is also important to examine and understand the realities of its dependencies and vulnerabilities in anticipating possible consequences. The responsibilities for cyber security cover the entire lifecycle of IT products, including producers, suppliers, and users. For example, product developers build security management capabilities into their products, and all users — including governments, companies, and individuals, irrespective of their scope and scale — may select the options that represent their desired security posture, taking into account risks, vulnerabilities, and mitigation costs. While IT industry security professionals understand their cyber security responsibilities, much of the broader user community may not recognize or fulfill *their* responsibilities.

A failure in energy, transportation, telecommunications, or water service can bring an organization's operations to a halt. In short order, such a failure can cascade to other infrastructures, affecting them as much or more. Physical infrastructure dependencies are known and accepted because they are prevalent across all businesses and agencies. While cyber infrastructure dependencies are an integral part of business operations, they are not as apparent, and thus are often overlooked or underestimated beyond the IT community. These dependencies are demonstrated through reliance on the Internet, commercial software, and "trusted connections."[2]

Insiders have not only authorized access to systems, but also the institutional knowledge of how to be most effective in an attack capacity.

In an effort to improve performance, business and government agencies rely on commercial software to streamline important business functions. Software supports many processes, from executing financial transactions, to storing personal data, to controlling the production of hazardous chemicals. However, software with unknown security flaws and software that is not properly patched can pose serious risks to an organization.

As companies and agencies strive to increase productivity, revenue, and profit, basic business operations are being outsourced. From payroll to marketing to managed security services, businesses and governments are increasing their dependence on other organizations to perform these basic business functions. This dependence on outsourcing usually relies on cyber infrastructure to manage data and information and introduces

[2] Trusted connections (i.e., system interconnections) are defined as the direct connection of two or more IT systems owned by separate organizations [1].

additional risks to business and government through trusted connections. Entities may have strong internal policies and controls in place to secure *their* cyber systems, but they cannot ensure the security of cyber systems and software of their outsourced services [2].

IMPLEMENTING A CYBER RISK MANAGEMENT APPROACH FOR CRITICAL INFRASTRUCTURE PROTECTION

How can companies, governments, and individuals work together to address the multitude of cyber threats and vulnerabilities and demonstrate real progress in securing our cyber infrastructure? The answer to this question requires a dynamic risk-based approach that takes into account the unique consequences, vulnerabilities, and threats of cyberspace. Analyzing risk to manage security priorities is one of the DHS's primary goals. As DHS Secretary Michael Chertoff noted in a recent speech, "Risk management must guide our decision making as we examine how we can best organize to prevent, respond, and recover from an attack."[3]

> **Entities may have strong internal policies and controls in place to secure *their* cyber systems, but they cannot ensure the security of cyber systems and software of their outsourced services.**

Consequences, vulnerabilities, and threats are the components of risk. While risk assessments should be conducted in all businesses, sectors, and agencies, many existing risk methodologies have either been limited in scope (they often do not include cyber assets or cyber components of physical assets) or not widely implemented. A flexible, widely used risk methodology that addresses cyber security vulnerability is necessary to inform the security practices applied to our physical infrastructure and cyber infrastructure.

Given today's highly interconnected environment, a common approach is required to unify the protective activities within public, private, and international entities and secure cyber infrastructure. Such an approach will allow for risk comparison across assets and more efficient allocation of resources, as well as efficient protection of cyber infrastructure. Recognizing this need, DHS adopted and is promoting a risk-based approach to protecting critical US infrastructure.

In response to Homeland Security Presidential Directive 7 (HSPD-7), "Critical Infrastructure Identification, Prioritization, and Protection," DHS drafted the *National Infrastructure Protection Plan* (NIPP), which details how the public and private sectors will work together to identify, prioritize, and conduct risk assessments of the 17 critical infrastructure and key resource (CI/KR) sectors [6]. DHS's risk-based approach is described in detail in the NIPP draft, which provides the unifying structure for the integration of CI/KR protection efforts into a single national program. It sets forth a Risk Management Framework for public- and private-sector security partners to work together to produce a comprehensive, systematic, and rational assessment of national or sector risk, which drives CI/KR risk reduction activities. The NIPP includes a cross-sector cyber element that is a component of each sector and recognizes the IT Sector specifically as one of the 17 CI/KR sectors.

The cornerstone of the NIPP is the Risk Management Framework (see Figure 1), which establishes the process for combining consequence, vulnerability, and threat information to assess risk.

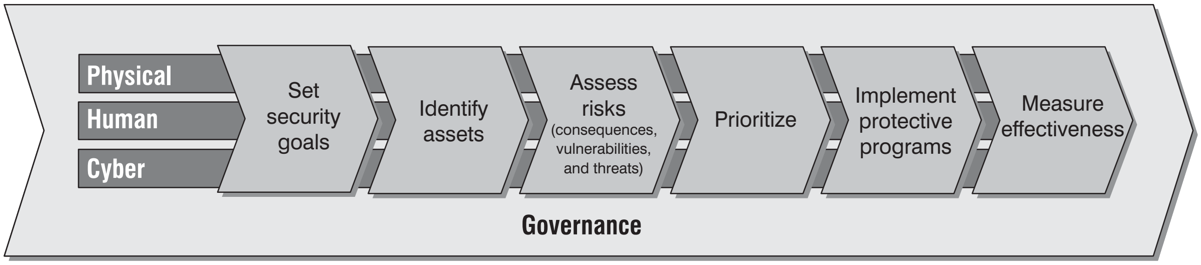


Figure 1 — The NIPP Risk Management Framework.

[3] Remarks by Secretary Michael Chertoff at the George Washington University Homeland Security Policy Institute on 16 March 2005.

THE NIPP RISK MANAGEMENT FRAMEWORK

The NIPP framework is composed of six specific activities:

1. **Set security goals.** Define specific outcomes, conditions, end points, or performance targets that collectively represent an effective security posture.
2. **Identify assets.** Develop an inventory of the individual assets and systems that make up the nation's CI/KR, some of which may be located outside the US, and collect information on them, including dependencies, interdependencies, and reliance on cyber systems.
3. **Assess risks.** Determine which assets and systems are critical by calculating risk, combining potential direct and indirect consequences of an attack (including dependencies and interdependencies associated with each identified asset), known vulnerabilities to various potential attack vectors, and general or specific threat information.
4. **Prioritize.** Aggregate and order assessment results to present a comprehensive picture of national CI/KR risk in order to establish protection priorities and provide the basis for planning and the informed allocation of resources.
5. **Implement protective programs.** Select appropriate protective measures or programs and allocate funding and resources designed to address targeted priorities.
6. **Measure effectiveness.** Incorporate metrics and other evaluation procedures at the national and sector levels to measure progress and assess effectiveness of the national CI/KR protection program.

A common approach is needed to assess risk so that protection priorities can be set across the CI/KR sectors. The first step toward achieving this common approach is to establish common definitions and analysis of the basic factors of risk:

1. Consequence analysis
2. Vulnerability assessment
3. Threat analysis

When the three basic factors of risk (consequence, vulnerability, and threat) are combined, they form the risk associated with an asset, system, or network (i.e., the potential for loss of or damage to an asset or system). As risk assessments are completed across assets, systems, or sets of assets or systems, the results are prioritized to help identify where risk reduction activities are most needed and to determine what protective actions should be targeted first.

The highly distributed and interconnected nature of cyber infrastructure (both physically and logically) requires that cyber protective actions and programs be implemented both within and across sectors. DHS is committed to identifying and supporting a variety of protective initiatives and fostering international cooperation to secure the nation's cyber infrastructure.

As mentioned above, the responsibilities for securing cyber infrastructure are dispersed and include both the producers and users of the infrastructure. DHS's efforts to implement the NIPP framework recognize that duality by addressing the IT Sector responsibility and the cross-sector cyber element that applies to all sectors under the NIPP.

IT Sector

The IT Sector, also referred to as the "IT Industrial Base," is composed of the producers of hardware, software, and IT services. DHS is working collaboratively with our private- and public-sector partners in the IT Sector through the IT Sector Coordinating Council and Government Coordinating Council to determine their approach and criteria for each step of the framework. As with all infrastructure sectors, private stakeholder participation in the process is essential to developing and implementing an efficient and effective IT Sector-Specific Plan (SSP). Furthermore, the Internet has been identified as a key resource, as all sectors rely upon and utilize the Internet to varying degrees. It is made up of assets from both the IT and Telecommunications Sectors, and the availability of the service is the responsibility of these two sectors.

Cross-Sector Cyber Element

While the producers of cyber infrastructure are addressed in the IT Sector, the cross-sector cyber element of the NIPP focuses on consumers of cyber infrastructure, including CI/KR sectors and their associated security partners. Each sector is responsible for securing its cyber infrastructure. The NIPP draft addresses cyber security and the cross-sector cyber element of CI/KR protection across all 17 sectors. The NIPP also addresses specific cyber responsibilities for sector security partners, processes, and initiatives to reduce cyber risk, and provides milestones and metrics to measure progress on enhancing the nation's protection of its cyber infrastructure. The 17 CI/KR SSPs will further detail risk reduction strategies related to their respective critical cyber infrastructure.

CONCLUSION

All public- and private-sector organizations should develop and implement a cyber risk management strategy to reduce the risk to the nation's cyber infrastructure. This strategy should include the following three components:

1. **Identify cyber assets, systems, and networks.** A process should be defined and implemented to identify cyber assets and cyber elements of physical assets of potential sector, regional, or national importance. Cyber assets represent a variety of hardware and software components, including business and control systems, networking equipment, database servers and software, and security systems. The process for identifying cyber assets should be scalable, distributable, and repeatable to ensure that it is practical and efficient and provides accurate results.
2. **Assess cyber risk.** Consequences, vulnerabilities, and threats should be identified and analyzed to assess risk. Potential consequences should include those that result from reliance on cyber assets. Vulnerability assessments can be conducted on cyber assets using a variety of approaches, methodologies, or criteria. Threat analysis should address those scenarios that are of highest concern.
3. **Implement protective programs to reduce risk.** Organizations should make decisions to implement protective programs based on their risk assessments and their desired security posture. While some risk may be acceptable, appropriate and effective protective measures will be necessary to balance risk and associated costs.

An organization's cyber risk mitigation strategy should be realistic and actionable, with stakeholders fully engaged in the implementation. The NIPP framework is flexible enough to allow individual organizations to tailor it to meet their requirements. By securing portions of our cyber infrastructure across multiple organizations and the 17 sectors, the overall infrastructure will become more resilient.

No one can protect the entire cyber infrastructure alone. DHS will continue to partner with state, territorial, tribal, local, and international governments, businesses, industries, and sectors to mitigate the risk associated with cyber consequences, vulnerabilities, and threats. DHS applauds the efforts of businesses and government agencies thus far and encourages them to continue partnering with their Sector-Specific Agencies (SSAs) and respective coordinating councils. Together, all infrastructure stakeholders can reduce risk and improve the overall security of the national cyber infrastructure.

REFERENCES

1. Grance, Tim, Joan Hash, Steven Peck, Jonathan Smith, and Karen Korow-Diks. *Security Guide for Interconnecting Information Technology Systems*. NIST Special Publication 800-47. National Institute of Standards and Technology, August 2002 (http://csrc.nist.gov/publications/nistpubs/800-47/sp800-47.pdf).

2. Goetz, Eric et al. *Survey and Analysis of Security Issues in the U.S. Banking and Finance Sector.* Institute for Security Technology Studies at Dartmouth College, September 2003 (www.ists.dartmouth.edu/analysis/secfin0903.pdf).

3. Moore, David, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver. "The Spread of the Sapphire/Slammer Worm." Cooperative Association for Internet Data Analysis, 2003 (www.caida.org/outreach/papers/2003/sapphire/sapphire.html).

4. Musil, Steven. "Week in Review: Worm's Wrath." *CNET News.com*, 7 February 2003 (http://news.com.com/Week+in+review+Worms+wrath/2100-1083_3-983720.html?tag=st.ref.goo).

5. Randazzo, Marisa Reddy, Michelle Keeney, Eileen Kowalski, Dawn Cappelli, and Andrew Moore. *Insider Threat Study: Illicit Cyber Activity in the Banking and Finance Sector*. US Secret Service and CERT Coordination Center/SEI, August 2004 (www.cert.org/archive/pdf/bankfin040820.pdf).

6. US Department of Homeland Security (DHS). *National Infrastructure Protection Plan* (draft). DHS, 2 November 2005 (www.fas.org/irp/agency/dhs/nipp110205.pdf).

Robert B. Stephan is the Assistant Secretary for Infrastructure Protection at the US Department of Homeland Security. He is responsible for the department's efforts to define the nation's critical infrastructure and key resources and coordinate risk-based strategies and protective measures to secure them from terrorist attack.

Mr. Stephan can be reached at Bob.Stephan@dhs.gov.

SURVEY SAYS ...

Investing in Cyber Security: The Path to Good Practice

by Shari Lawrence Pfleeger, Rachel Rue, Jay Horwitz, and Aruna Balakrishnan

RISKS AND CONSEQUENCES

According to the recent *Technology Assessment: Cyber Security for Critical Infrastructure Protection* conducted by the US Government Accountability Office (GAO):

> Since the early 1990s, increasing computer interconnectivity — most notably growth in the use of the Internet — has revolutionized the way that our government, our nation, and much of the world communicate and conduct business. While the benefits have been enormous, this widespread interconnectivity also poses significant risks to the government's and our nation's computer systems and, more important, to the critical operations and infrastructures they support. The speed and accessibility that create the enormous benefits of the computer age, if not properly controlled, allow unauthorized individuals and organizations to inexpensively eavesdrop on or interfere with these operations from remote locations, for mischievous or malicious purposes including fraud or sabotage. [37]

While security analysts understand a system's vulnerability to potential cyber attacks fairly well, the consequences of such attacks to companies, business sectors, and nations are largely unknown. To date, research on the economic consequences of cyber attacks has been limited, dealing primarily with microanalyses of the direct impacts of attacks on a particular organization. Many organizations recognize the significant potential of a cyber attack's effects to cascade from one computer or business system to another, but there have been no significant efforts to develop a methodology to account for both direct and indirect costs. Without such a methodology, governments and businesses are hard-pressed to make informed decisions about how much to invest in cyber security and how to invest each dollar most effectively.

Such investment decisions require the answer to at least two questions: (1) What is the likelihood of an attack?; and (2) What are the likely consequences of an attack? In this article, we explore some answers to these questions. We begin by addressing the data available to inform decisions about cyber security investment. Next, we look at what research tells us about the tradeoffs between investment and protection. We end with a discussion of where research is headed and what help businesses can expect in the short and long term from these research results.

DATA REALITIES

Many organizations assume that the likelihood of a cyber attack is reasonably high and may increase over time. Anecdotal evidence suggests that even among the many organizations that have taken steps to detect and prevent attacks, some have experienced significant incidents nevertheless. To provide a more realistic picture of the nature and number of cyber incidents, several surveys have been conducted in the last few years to capture information about security attacks and protection. The following are among the most well known:

- First conducted in 2002, the annual Australian Computer Crime and Security Survey (ACC)[1] is based on information provided by Australia's federal, state, and territorial law enforcement agencies and AusCERT.[2] It solicits data from large organizations about computer network attacks and computer misuse trends in Australia.
- The UK Department of Trade and Industry has administered seven Information and Security Breaches Surveys (ISBS)[3] since 1991. They report on Internet use, dependence on information technology, and computer security incidents at UK businesses.
- The annual CSI/FBI Computer Crime and Security Survey[4] polls computer security practitioners in US corporations, government agencies, financial institutions, medical institutions, and universities that have joined the Computer Security Institute (CSI) or have

[1] See www.auscert.org.au/render.html?it=2001.

[2] See www.auscert.org.au.

[3] See www.infosec.co.uk/files/DTI_Survey_Report.pdf.

[4] See www.usdoj.gov/criminal/cybercrime/FBI2005.pdf.

attended a CSI seminar or workshop. The survey addresses computer usage, attacks, and actions taken in response to security incidents.

In addition, there are global surveys within particular sectors, such as the Deloitte Touche Tohmatsu Global Security Survey (GSS).[5] The third GSS, administered in 2005, solicited input from chief security officers and security management teams of financial services industry organizations worldwide, asking their perceptions of how one organization's information security compares to its counterparts' security.

These surveys paint a mixed picture of the security landscape. The ACC reports a decrease in attacks of all types. On the other hand, the ISBS survey found that the percentage of UK businesses experiencing attacks has increased by one-third over the last two years, and 43% of the CSI member organizations surveyed have experienced increases in the rate of attacks from 2003 to 2004. At the same time, the Deloitte survey found that the rate of financial sector security breaches in the US has remained roughly the same for the past year. The variation in these reports may derive from the different populations being surveyed; the surveys represent different countries, different sectors, different degrees of sophistication about security matters, and bias in the pool of respondents. Moreover, most are convenience surveys, so the population represented by the respondents is unclear, making it difficult to generalize the results.

Another significant problem is the lack of standards in defining, tracking, and reporting security incidents and attacks. Surveys ask variously about the incidence of "electronic attacks" (ACC); "virus encounters" and "virus disasters" (the ICSA Labs Eighth Annual Computer Virus Prevalence Survey);[6] "total number of electronic crimes or network, system, or data intrusions" (CSI/FBI); "security incidents," "accidental security incidents," "malicious security incidents," and "serious security incidents" (ISBS); "any form of security breach" (GSS); "unauthorized use of computer systems" (CSI/FBI); and "incidents that resulted in an unexpected or unscheduled outage of critical business systems" (Ernst & Young Global Information Security Survey [EY]).[7] It would be difficult to find two surveys whose results are strictly comparable. Thus, much of the reported evidence is categorized differently from one study to another, and the answers are based on respondents' opinions, interpretations, or perceptions, not on consistent capture and analysis of solid empirical data.

Understanding the sources of attacks is similarly problematic. For example, the ACC survey reports that the rate of insider attacks has remained constant. However, Deloitte claims that, in financial services, the majority of attacks come from the inside, and the rate is rising. The EY survey emphasizes the rising threat of insider attacks as well. Several other surveys note that sources of attacks are unknown in a significant percentage of cases.

Across surveys, there is general agreement about which attacks are most serious. Viruses, Trojan horses, worms, and malicious code pose significant threats; most sectors also fear insider misuse and abuse of access. Phishing[8] is a relatively new and growing concern, and such attacks have increased dramatically over the past two years.

In addition to number and kind of attack, surveys often ask about effect, particularly in terms of cost. Significant variations exist in this category as well. For example, the ICSA survey reports a sharp (25%) increase in the cost of recovering lost or damaged data. However, the ACC, EY, and CSI/FBI surveys find a decrease in total damage from attacks, even though this cost is increasing for some kinds of attacks (such as unauthorized access and theft, noted by CSI/FBI). Twenty-five percent of responding organizations report financial loss to CSI/FBI, and 56% report operational losses.

Once again, the reasons for these variations are partly attributable to disparities in the pools of respondents. However, a more significant problem, acknowledged both by survey respondents and administrators, is the difficulty in detecting and measuring both direct and indirect costs from security breaches. There are neither accepted definitions of loss nor standard, reliable methods to measure it. For example, the ICSA 2004 survey notes that "respondents in our survey historically underestimate costs by a factor of 7 to 10."

[5] See www.deloitte.pt/dtt/research/0,1015,sid%253D1000%2526cid%253D85452,00.html.

[6] ICSA Labs, a division of TruSecure Corporation, has administered a survey about viruses since 1994. More information is available at www3.ca.com/Solutions/Collateral.asp?CID=41607&ID=156.

[7] See www.ey.com/global/download.nsf/Austria/2004_global_info_sec_survey/$file/2004_Global_Information_Security_Survey_2004.pdf.

[8] Phishing is the practice of directing Internet users to a fake Web site by using authentic-looking e-mail in an attempt to steal passwords, financial, or personal information, or to introduce malicious code.

There are several areas in which the various surveys do, in fact, reach consensus. For example, many surveys indicate that formal security policies and incident response plans are important. In addition, lack of education and training of staff, both within the IT security team and throughout the organization, appears to be a major obstacle to improved security. More generally, a poor "security culture" (in terms of awareness and understanding of security issues and policies) is often reported to be a problem. Thus, survey respondents report that regular testing and updating of security procedures, combined with practices that increase staff awareness, are very important. Unfortunately, there is very little quantitative evidence supporting these views, plausible as they may be.

The various survey results highlight another gap in our understanding of security investments. It is unclear how much organizations have invested in security protection, prevention, and mitigation; in addition, we do not know how they make investment decisions or measure the effectiveness of their security investments. Inputs required for such decision making — such as rates and severity of attacks, cost of damage and recovery throughout the enterprise, and actual cost of security measures of all types — are not known with any accuracy. Neither is it clear whether traditional measures, such as return on investment or internal rate of return, are the best ones to use in assessing security effectiveness. Simple questions, such as how much more security an extra dollar buys a company, go unanswered.

To address this situation, the Bureau of Justice Statistics at the US Department of Justice will soon administer the first large-scale, carefully designed and sampled cyber security survey. It will provide the first official US statistics on the extent and consequences of cyber crime against US businesses. Planned for early 2006, the results should be publicly available by the end of the year.

DATA NEEDS

To understand how to improve decision making about the economics of cyber security, we must first examine what data is needed and how it could be used. Ideally, a data source should provide information to support the following tasks.

Managing Resources to Monitor and Address Cyber Incidents

Survey data can inform resource allocation decisions ranging from government resources for monitoring cyber crime to industrial resources for sensing attempts at system penetration. Trend data about cyber incidents, including records from incident response teams, can support more effective strategic planning. Such data can then be used to help in:

- Understanding current best practices
- Evaluating existing regulations and standards and formulating new ones
- Choosing the most effective measures of effectiveness for resource allocation
- Choosing organizational structures to facilitate efficient use of resources
- Understanding current and future trends: types and frequency of attack, probable and possible consequences for each type of attack, targeted sectors and businesses, and motives for and intended consequences of attacks

Implementing Standards and Guidelines

Trend data about vulnerabilities and attacks can suggest areas that new or updated standards and guidelines might address. For example, the Common Vulnerabilities and Exposures (CVE)[9] list uses standardized names for vulnerabilities, which allows them to be cross-referenced and catalogued. The application of such standards allows organizations to search for common problems and possible solutions. In turn, standard classification and naming of vulnerabilities, types of attack, and techniques used in attacks can permit analysis that suggests best practices involving the most cost-effective technologies, policies and procedures, and organizational structures and processes.

Involving the Insurance Industry

The insurance industry may play a growing role in securing cyberspace. For example, the Basel II agreement[10] allows businesses to decrease their financial reserves in exchange for sharing information about

[9] See http://cve.mitre.org. To date, 115 organizations from industry, government, and academia worldwide have declared that 186 network security products or services are or will be CVE-compatible.

[10] The Basel Committee on Banking Supervision, a group of central bankers and banking regulators, proposed a new international capital standard, published in June 2004. The proposal, known as Basel II, will be implemented from year-end 2006. Basel II is a comprehensive framework for regulatory capital and risk management. It represents a major revision of the international standard on bank capital adequacy that was introduced in 1988, aligning the capital measurement framework with sound contemporary practices in banking, improvements in risk management, and enhanced financial stability.

cyber vulnerabilities and agreeing to comply with minimum standards. Credible survey data could be used to set policy terms and standards for insurability against cyber attacks.

Benchmarking Infrastructure Protection Efforts

Much of every nation's critical infrastructure depends heavily on information technology. Repeated and coordinated surveys could be used to compare the cyber security postures of different parts of the infrastructure and to measure the rate of improvement. Survey respondents could use the results to benchmark their own efforts.[11] In addition, benchmarks support:

- Analysis of trends in the frequency and, especially, severity of attacks and consequent losses
- Determination of best practices for addressing current and changing vulnerabilities
- Regular updating of standards

Benchmarking Government Systems

Repeated surveys could help benchmark government performance compared with the private sector. For example, data might show whether private-sector organizations are being attacked more or less than public-sector organizations and whether the private or public sector is more successful in defending against attacks.

Suggesting or Enforcing International Agreements

Survey data can support law enforcement officials in negotiating agreements about cyber crime. For example, data about the origins and nature of cyber incidents, particularly their strategic and financial impacts, could help identify where agreements are needed. Surveys could capture data regarding:

- Sources and targets identified by location
- Vulnerabilities requiring international agreement and enforcement

Measuring Effectiveness

Government plays a role in educating and encouraging both businesses and citizens to strengthen cyber security. Trends from the survey data could provide feedback on the effectiveness of such campaigns. Also, survey information on which types of businesses are using which types of prevention and mitigation strategies can highlight the most effective techniques, supporting decision making about research and financial investment. In turn, these types of data can influence:

- The perception and empirical measurement of effectiveness of security strategies of all types
- Development and dissemination of good metrics
- The perceived and actual effects of regulations and standards, and enforcement thereof
- The perceived and actual effect of both public- and private-sector education strategies

RESEARCH NEEDS AND RESEARCH REALITIES

To make effective decisions about economic investment in cyber security, we need more than data; we also need research into motivation, action, and relationships. By its nature, this research must be multidisciplinary, and it is often conducted within and across the boundaries of engineering, business, and arts and science schools.

Historically, cyber security has been the domain of engineers. However, social scientists have recently brought new insight to the problem, with strong arguments for richer economic analysis. For example, University of Cambridge researcher Ross Anderson makes the case that studying economic incentives is as important as studying the underlying technology [3]. Other researchers argue that cyber security is a core business concern [31] and that market incentives are fundamental to shaping relevant public policy [1].

But the literature is immature, as both the paucity of empirical analysis and the lack of agreement on findings reveal. Nevertheless, researchers bring economics to bear in four related streams of inquiry:

1. **Software quality.** What forces create vulnerable software and systems?
2. **Market interventions.** Which market-based and regulatory interventions are designed to spread risk, diminish information asymmetry, and better align stakeholder incentives?
3. **Enterprise decision making.** What are the various approaches to investing in cyber security?
4. **Evaluation.** How useful are different evaluative techniques and criteria?

We will discuss each of these in turn.

[11] The RAND Corporation has performed similar activities; see information about European security information-sharing initiatives at www.iaac.org.uk.

Software Quality

Software is generally held to be a complex product likely to have defects that create vulnerability to attack. Recently, researchers have begun to examine the contextual factors — such as time to market and shareholder value — that lead software vendors to invest (or underinvest) in quality.

- **Patches.** Some researchers have demonstrated vendors' incentives to release products early and make repairs later using patches [4]. Other research suggests that it is socially beneficial to release fixes quickly [36].
- **Disclosure timing.** Questions remain about when and whether vendors should disclose newfound vulnerabilities. Researchers are at odds about the details of specifying the vulnerabilities in a formal model [5, 10].
- **Stock-price effects.** Given the corporate mandate to create shareholder value, researchers are looking to the stock market for insight on vendor behavior. The results are mixed. For example, Katherine Campbell and her colleagues at the University of Maryland found limited evidence of a negative market reaction to public announcements of security breaches [9]. By contrast, Carnegie Mellon researchers describe the negative change in market valuation that results immediately after disclosure of a software failure [35].
- **Vulnerability reduction.** Researchers have also questioned the value of actively searching for vulnerabilities. Again, the results are mixed. Andy Ozment, a student of Ross Anderson's, drew no clear conclusion from the data [27], and Eric Rescorla, founder of RTFM, Inc., found no relationship between software quality and the effort to remove vulnerabilities [28].

Market Interventions

Researchers are formally examining the impact of information sharing, market mechanisms, and new approaches to insurance and liability to assess their effectiveness. The effects of these approaches are as yet inconclusive.

Information-Sharing Programs

In connection with work on software quality, researchers are examining the economic consequences of and motivations for sharing security information. Bruce Schneier, a frequent commentator on security issues, looks at full disclosure in the context of "inevitable vulnerability" and proposes efforts to shrink the window of exposure by prompting vendors to act quickly [32]. By drawing on the literature of trade associations and research joint ventures, Lawrence Gordon and his colleagues at the University of Maryland illustrate a theoretical approach to analyzing the voluntary Information Sharing and Analysis Centers (ISACs) formed by business sectors to advise the US government about homeland security issues [21]. The University of Pittsburgh's Esther Gal-Or and Carnegie Mellon's Anindya Ghose go a step further, using a formal model of ISACs to show that security investments strategically complement information sharing [16]. These results provide a context for understanding when ISACs can be effective. Overall, researchers have found that information-sharing programs can have a significant and positive effect on security.

Researchers have begun to examine the contextual factors — such as time to market and shareholder value — that lead software vendors to invest (or underinvest) in quality.

Market Mechanisms

Anderson draws parallels between cyber security and environmental pollution; both involve investment by one group that benefits another, a phenomenon that is called a "negative externality" [2]. For this reason, researchers have proposed creating "vulnerability markets" using transferable security credits, so that more vulnerabilities in one product can balance fewer in other products (much as the market in pollution credits works in the US today) [7, 8]. Stuart Schechter of MIT Lincoln Laboratory proposes using vulnerability markets to benchmark security strength by rewarding bug discoveries [29]. He argues that such a market will yield information that can be used to improve the software development process.

Ozment questions this proposal by comparing the mechanism to auctions and pointing out their inherent limitations [26]. Purdue University's Karthik Kannan and Carnegie Mellon's Rahul Telang also reject market-based mechanisms because of higher expected user losses than passive systems (such as Computer Emergency Response Teams) and — even worse — incentives for misuse by monopolists of an application or operating system [23].

Liability Reform

Some researchers recommend stricter liability requirements for software companies, ending the use of end-user license agreements that obviate vendor responsibility. For example, Schneier makes the case for tightening liability on software manufacturers [31], while Informed Security founder Adam Shostack calls for more subtle use of vendor liability requirements to create better signaling of product quality [33].

Research has also examined the benefits of shifting liability through insurance contracts. Jay Kesan and his colleagues at the University of Illinois' Information Trust Institute explain the limitations of traditional insurance when applied to cyber security, citing a lack of good data, overpricing, and excessive exclusions to skirt moral hazard risks [24]. Further research has identified additional challenges to a healthy insurance market in the form of interdependent risk, which decreases the benefit of risk diversification. This condition results from widespread incentives to undersecure and the prevalence of dominant software packages, where a single exploitation can affect a large population of systems [6, 25].

> **Researchers are now proposing new metrics to address the challenges of cost assignment and to gain more useful insight into the problem.**

Enterprise Decision Making

Companies face tough decisions about cyber security investments. Recognizing that most information infrastructure is controlled by wide-ranging corporations, economic research has provided tools — ranging from accepted methods instantiated in accessible tools to esoteric methods not yet widely available — to support analysis of corporate investment in cyber security.

Gordon and his University of Maryland colleagues provide a useful framework to help companies think about tradeoffs between investing indirectly in cyber insurance and directly in security countermeasures [22]. Furthermore, Gordon and coauthor Martin Loeb offer a systematic way to incorporate qualitative information into the investment analysis through a prioritizing approach called the Analytical Hierarchy Process to elicit user preferences [18], and their book explains how to perform return-on-investment calculations [19]. Similarly, James Conrad, a doctoral student at the University of Idaho, proposes using Monte-Carlo simulation to support midlevel managers in forecasting uncertainties in security investment decisions [12].

On the other hand, Stanford's Kevin Soo Hoo [34] and Fariborz Farahmand and his colleagues at Purdue [15] suggest more expansive frameworks using formal decision analysis to determine the amount to invest based on the likelihood of intrusion estimation. And Tulane's Huseyin Cavusoglu and his coauthors provide a detailed model to help companies select the best security architecture based on observed intrusions and derived likelihood of attack [11].

Evaluation

Consensus is yet to form on which evaluative criteria are most useful. Researchers and practitioners have looked to the literature on investment analysis in general and security investment in particular for guidance on how to make investment decisions in cyber security. Unfortunately, no method has emerged as a "gold standard" in theory or practice. Initial calls to discard the heuristic of fear-uncertainty-and-doubt were replaced by insistence on return on security investment (ROSI) analysis [17]. However, ROSI is itself highly contentious. Although consistent with other corporate investment decisions, ROSI and related concepts of internal rate of return and net present value are considered to be inappropriate frameworks for this kind of analysis. Gordon and Loeb contend that ROSI does not reveal the true economic rate of return and leads to the wrong investment objectives [20]. Furthermore, Cavusoglu et al. suggest that ROSI is frustrated by the need to assign costs to poorly defined outcomes [11].

Researchers are now proposing new metrics to address the challenges of cost assignment and to gain more useful insight into the problem. For example, Farahmand et al. consider using damage assessment across predefined categories as an evaluative framework [14]. Schechter [30] introduces cost-to-break (i.e., the effort required to invade a system) as a measure of security strength [29]. The twin metrics of cost-to-break and security strength work within an evaluative model that considers the effort required by an attacker to gain access to a system. Schechter offers this measure to improve predictions about the amount of risk faced by a system. The University of Milan's Marco Cremonini and Zucchetti Group's Patrizia Martini take a similar approach in channeling the perspective of the attacker. They use this vantage point to measure impacts and thereby benchmark appropriate levels of investment based on an attacker's potential "return-on-attack" [13].

THE WAY FORWARD

The economics of cyber security is an emerging field. The first *Workshop on the Economics of Information Security* was held in Berkeley, California, USA, in 2002, and *IEEE Security & Privacy* devoted a special issue to the topic in January 2005.[12] Both the public and private sectors are eager for better data, better understanding, and better methods for using resources wisely in protecting critical products and services and providing assurance that software will work as expected. We suggest that the most effective way forward involves three steps:

1. Continued interdisciplinary research into the motivations, methods, and opportunities for wisely investing in cyber security
2. Active participation and cooperation by government, industry, and academia, including involvement in high-quality surveys and analysis
3. Sharing of information about cyber security incidents, using vehicles that protect consumers, corporations, and markets but enhance our understanding of the nature, volume, and cost of attacks

In particular, *Cutter IT Journal* readers can keep abreast of cyber security economics issues by participating in the *Fifth Annual Workshop on the Economics of Information Security* in Cambridge, UK,[13] by participating in surveys and studies to better understand the nature and extent of cyber incidents, by sharing information with researchers and colleagues to enable business sectors to take a coordinated approach to preventing and mitigating attacks, and by applying appropriate business measures to balance security investments with other requests for corporate resources. We can all be active players in improving our understanding of cyber security economics by monitoring cyber incidents and responses, soliciting and using standard terminology and measures, and by sharing data whenever possible.

REFERENCES

1. Alderson, David, and Kevin J. Soo Hoo. *The Role of Economic Incentives in Securing Cyberspace*. Center for International Security and Cooperation (CISAC), Stanford University, November 2004.

2. Anderson, Ross. "Unsettling Parallels Between Security and the Environment." Paper presented at the *Workshop on the Economics of Information Security*, Berkeley, California, USA, May 2002.

3. Anderson, Ross. "Why Information Security Is Hard — An Economic Perspective." In *Proceedings of the 17th Annual Computer Security Applications Conference*. ACM, 2001.

4. Arora, Ashish, Jonathan P. Caulkins, and Rahul Telang. "Sell First, Fix Later: Impact of Patching on Software Quality." Working paper, Heinz School, Carnegie Mellon University, January 2005.

5. Arora, Ashish, Rahul Telang, and Hao Xu. "Optimal Policy for Software Vulnerability Disclosure." Paper presented at the *Third Annual Workshop on the Economics of Information Security*, Minneapolis, Minnesota, USA, May 2004.

6. Bohme, Rainer. "Cyber-Insurance Revisited." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

7. Camp, L. Jean, and Catherine Wolfram. "Pricing Security." Paper presented at the *CERT Information Survivability Workshop*, Boston, MA, October 2000.

8. Camp, L. Jean, and Catherine Wolfram. "Pricing Security: A Market in Vulnerabilities." In *Economics of Information Security*, edited by L. Jean Camp and Stephen Lewis. Springer-Kluwer, 2004.

9. Campbell, Katherine, Lawrence A. Gordon, Martin P. Loeb, and Lei Zhou. "The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market." *Journal of Computer Security*, Vol. 11, No. 3, March 2003, pp. 431-448.

10. Cavusoglu, Hasan, Huseyin Cavusoglu, and Srinivasan Raghunathan. "Emerging Issues in Responsible Vulnerability Disclosure." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

11. Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. "A Model for Evaluating: IT Security Investments." *Communications of the ACM*, Vol. 47, No. 7, July 2004.

12. Conrad, James R. "Analyzing the Risks of Information Security Investments with Monte-Carlo Simulations." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

13. Cremonini, Marco, and Patrizia Martini. "Evaluating Information Security Investments from Attacker's Perspective: The Return-on-Attack." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

[12] See www.computer.org/portal/site/security.

[13] See www.cl.cam.ac.uk/users/twm29/WEIS06 for more details.

14. Farahmand, Fariborz, Shamkant B. Navathe, Gunter P. Sharp, and Philip H. Enslow. *Assessing Damages of Information Security Incidents and Selecting Control Measures: A Case Study Approach.* Georgia Institute of Technology, 2005.

15. Farahmand, Fariborz, Shamkant B. Navathe, Gunter P. Sharp, and Philip H. Enslow. "A Management Perspective on Risk of Security Threats to Information Systems." *Information Technology and Management*, Vol. 6, No. 2-3, April 2005, pp. 203-225.

16. Gal-Or, Esther, and Anindya Ghose. "The Economic Incentives for Sharing Security Information." 2004 revision of a paper presented at the *Second Annual Workshop on the Economics of Information Security*, College Park, Maryland, USA, May 2003.

17. Geer, Daniel E. "Making Choices to Show ROI." *Secure Business Quarterly*, Vol. 1, No. 2, Fourth Quarter 2001.

18. Gordon, Lawrence A., and Martin P. Loeb. "Evaluating Information Security Investments Using the Analytical Hierarchy Process." *Communications of the ACM*, Vol. 48, No. 2, February 2005, pp. 78-83.

19. Gordon, Lawrence A., and Martin P. Loeb. *Managing Cyber Security Resources*, McGraw-Hill, 2005.

20. Gordon, Lawrence A., and Martin P. Loeb. "Return on Information Security Investments: Myths vs. Realities." *Strategic Finance*, Vol. 84, No. 5, November 2002.

21. Gordon, Lawrence A., Martin P. Loeb, and William Lucyshyn. "An Economics Perspective on the Sharing of Information Related to Security Breaches: Concepts and Empirical Evidence." Paper presented at the *Workshop on the Economics of Information Security*, Berkeley, California, USA, May 2002.

22. Gordon, Lawrence A., Martin P. Loeb, and Tashfeen Sohail. "A Framework for Using Insurance for Cyber-Risk Management." *Communications of the ACM*, Vol. 46, No. 3, March 2003.

23. Kannan, Karthik, and Rahul Telang. "Market for Software Vulnerabilities? Think Again." *Management Science,* Vol. 51, No. 5, May 2005, pp. 726-740.

24. Kesan, Jay P., Ruperto P. Majuca, and William J. Yurcik. "CyberInsurance as a Market-Based Solution to the Problem of Cybersecurity — A Case Study." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

25. Ogut, Hulisi, Nirup Menon, and Srinivasan Raghunathan. "Cyber Insurance and IT Security Investment: Impact of Interdependent Risk." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

26. Ozment, Andrew. "Bug Auctions: Vulnerability Markets Reconsidered." Paper presented at the *Third Annual Workshop on the Economics of Information Security*, Minneapolis, Minnesota, USA, May 2004.

27. Ozment, Andrew. "The Likelihood of Vulnerability Rediscovery and the Social Utility of Vulnerability Hunting." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

28. Rescorla, Eric. "Is Finding Security Holes a Good Idea?" Paper presented at the *Third Annual Workshop on the Economics of Information Security*, Minneapolis, Minnesota, USA, May 2004.

29. Schechter, Stuart. "Computer Security Strength & Risk: A Quantitative Approach." Ph.D. thesis, Harvard University, 2004.

30. Schechter, Stuart. "Quantitatively Differentiating System Security." Paper presented at the *Workshop on the Economics of Information Security*, Berkeley, California, USA, May 2002.

31. Schneier, Bruce. "Computer Security: It's the Economics, Stupid." Paper presented at the *Workshop on the Economics of Information Security*, Berkeley, California, USA, May 2002.

32. Schneier, Bruce. "Full Disclosure and the Window of Exposure." *Crypto-gram Newsletter*, 15 September 2000 (www.schneier.com/crypto-gram-0009.html).

33. Shostack, Adam. "Avoiding Liability: An Alternative Route to More Secure Products." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

34. Soo Hoo, Kevin J. *How Much Is Enough? A Risk-Management Approach to Computer Security*. CISAC, Stanford University, August 2000.

35. Telang, Rahul, and Sunil Wattal. "Impact of Software Vulnerability Announcements on the Market Value of Software Vendors: An Empirical Investigation." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

36. Thursby, Maria, and Dmitri Nizovtsev. "Economic Analysis of Incentives to Disclose Software Vulnerabilities." Paper presented at the *Fourth Annual Workshop on the Economics of Information Security*, Cambridge, Massachusetts, USA, June 2005.

37. US Government Accountability Office (GAO). *Technology Assessment: Cybersecurity for Critical Information Infrastructure Protection*. GAO-04-321. GAO, May 2004.

Shari Lawrence Pfleeger is a senior researcher at the RAND Corporation, specializing in software engineering, cyber security, and information technology policy. She is team leader for a multidisciplinary team of researchers from RAND, the University of Virginia, MIT's Lincoln Laboratory, Dartmouth's Tuck Business School, and the George Mason University School of Law, funded by the US Department of Homeland Security through the Institute for Information Infrastructure Protection. The team is investigating the economics of cyber security from the national, enterprise, and technology perspectives. Rachel Rue is a RAND mathematician, and Aruna Balakrishnan is a RAND research assistant. Jay Horwitz is a doctoral student in public policy at the Pardee RAND Graduate School. More information about this project can be found at www.thei3p.org.

Securely Sustaining Software-Intensive Systems

by Carol Woody

INTRODUCTION

Great authorization and authentication will not save a system from a poorly planned implementation and sustainment. Shepherding a software-intensive system through project development to implementation is just the beginning of the saga. Modern software-intensive systems face the confusing and multifaceted challenge of sustainment. After implementation is when the security problems will show up. The developers move on, and the operational and sustainment support staff do the best they can.

As today's systems become increasingly reliant on COTS software, the issues surrounding sustainment become even more complex. The risks of ignoring these issues can potentially undermine not only security but also the stability, usability, and longevity of systems in the field. Organizations use COTS products as system components because they believe that commercial products are mature, stable, and adhere to well-recognized industry standards — and will thus yield robust, maintainable systems. This is a myth. The reality indicates more of a Rube Goldberg mix of "glue code" that links the pieces and parts into a "working" structure. To make matters worse, change is a constant requirement, as each vendor provides regular updates on an individualized schedule, and these updates must be applied within a prescribed time frame. Vendors also release emergency updates in response to identified security threats. Failure to apply the updates may result in security exposure as well as the loss of maintenance support, but applying them may require a complete restructuring of the system to return the pieces to a working whole. The result is a system in constant change. Poorly managed change further increases opportunity for security problems.

The growing popularity of service-based architectures and the use of common services shared across many systems linked in system-of-systems environments further challenge sustainment capabilities. While this architecture approach provides greater flexibility in design and development, extensive use of common services forces greater responsibility on sustainment functions for ensuring security. Shared common services must function as stable, well-established communication mechanisms. In order to maintain these common services as technology and system requirements change, version control and backward compatibility for software upgrades become critical needs. Common services can become excellent conduits for security compromise throughout an infrastructure. In many cases, the common services are the security controls for the infrastructure, handling authentication and authorization functions. Sustainment of these services requires careful planning and control.

Systems cannot be constructed to eliminate security risk, which is an ever-changing challenge. However, systems can be built incorporating mechanisms that assist with — instead of ignoring — the need to recognize, resist, and recover from security problems when they do occur. Decisions by the system owner on things such as licensing, ongoing product and user support, infrastructure upgrades, and hardware standards contribute to security risk in sustainment efforts. These may disagree with operational management decisions on intrusion detection, firewall restrictions, hardware support, and software patch management. This article discusses the linkages between development and sustainment that impact security and offers an approach that can enhance planning for security during implementation and sustainment. It also includes lessons learned from applying the planning approach in practice.

WHAT IS SOFTWARE SUSTAINMENT?

Until recently, software maintenance was considered sufficient long-term support for an implemented system, but not anymore. The *IEEE Standard Glossary of Software Engineering Terminology* defines "software maintenance" as:

> The process of modifying a software system or component after delivery to correct faults, improve performance or other attributes, or adapt to a changed environment. [2]

However, this definition ignores a broad range of efforts that must be addressed to keep a fielded system fully functional, including configuration management, help-desk support, user and support staff training, COTS product management, documentation management, and technology refresh. Too frequently, these critical operational areas are hidden and poorly planned, which results in underfunding and limitations for support resources. Software sustainment requires addressing the processes, procedures, people, material, and information required to support, maintain, and operate all software aspects of a system.

Industry sources have asserted that the typical IT organization spends 20% of its IT dollars on developing new capability and 80% on maintaining and operating the existing infrastructure [1]. However, it appears the organizational planning is focused completely on the initial 20% while the rest of the activities are expected to transition smoothly into the ongoing operation with minimal effort. This delay strategy is supported by the prevailing ROI financial perspective, which views dollars to be spent some time in the future as of lower value [3] and, hence, of lesser concern. In reality, decisions made today frame the options that will or will not be available in the future. Also, the separation of development resources from the operational support responsibilities, while useful for effective management control, minimizes the level of operational knowledge readily available to developers.

An organization's ability to sustain a system is greatly influenced by the decisions made during the design, development, and deployment of a system and its components. In addition, existing operations standards and procedures, developed based on experience with previously fielded systems, cannot be applied unilaterally to every system under development. Many systems, especially modernization and reengineering efforts, change the way the organization is using (and abusing) technology.

WHY SHOULD OPERATIONS CARE ABOUT SECURITY FOR SYSTEMS IN DEVELOPMENT?

Each system implementation represents a security risk to the operational environment, which is where security problems will appear. The transition to operations (T2O) should be a partnership planned well in advance by a team with representatives from development, the user community, and operational support. This team must compare the needs of the new system to the current environment to identify critical gaps and options for closing them. Too frequently, installation appears more like a race to a finish line once the system has completed integration testing. T2O planning should include a security risk assessment and a vulnerability assessment as part of the standard process. In addition, both of these types of assessment should be scheduled periodically throughout the life of the fielded system.

Planning is particularly critical if the sustainment effort is handled through a contract with a support organization. The contract may already be in place when a new system is ready for implementation. To be effective, the sustainment effort must provide mechanisms for assembling all the details of support, including security, needed for each system. Training for sustainment personnel should be considered, and, at a minimum, accurate functional and technical documentation is needed. Funding for both of these activities must be provided as part of the implementation planning.

Those tasked with handling sustainment and operations support, which may be two separate groups, must understand the physical components of the system and determine if there are unique characteristics that may change the current infrastructure footprint. In addition, those supporting the fielded system must develop an appreciation for the critical business processes the system supports and the organizational dependencies on availability and recovery options. These may not fit neatly into how the current production support environment is functioning, and adjustments to the infrastructure must be planned.

THE SECURITY RISK ASSESSMENT

Security risk assessments should be performed during development to prepare for implementation and periodically throughout the life of a fielded system. The security risk assessment should include the following steps:

- Define the target system implementation.
- Determine security attributes.
- Identify threats.
- Identify risks.
- Develop a protection plan to address operational gaps.

In defining the target system, each specific component of the operational system must be clearly described. This description includes the hardware and software composition; the data stored, processed, and transmitted by each component; and where each component will live in the current infrastructure. The description should include planned security controls such as certificates, encryption, intrusion detection, transaction logging, and any other means of allowing each component to recognize, resist, and recover from a compromise.

Determining the system's security attributes enables the organization to identify what is important about the processes and critical information on the target system that requires security protection. The attributes describe the confidentiality, integrity, and availability requirements expected for the system.

Threat identification describes ways in which the system would likely be targeted for compromise. Accidental problems will occur, resulting in expected system actions with content mistakes and unexpected sequences. In addition, attackers will be attempting to create unexpected situations deliberately devised to subvert the system, with unusual sequences of actions and content designed to trigger compromising results. The attacker can be a user or an external source. These scenarios, often referred to as abuse cases, may be considered in system design when architecture and interface design selections are made. However, as operational experience with the system increases, the understanding of threats improves, and the system security as implemented is usually no longer adequate. Additional threats include limitations in current organizational policies and practices as well as third-party arrangements and outsourcing exposure.

For a threat to become a risk, there must be an organizational impact should the threat be realized, such as lost revenue, fines, or lost productivity. Identifying risk involves assigning the level of impact to each threat identified for the system. The potential increase of impact for cascading threats should also be evaluated.

The team performing the security risk assessment will develop a protection plan to address risks that the organization cannot ignore. This plan can include policy changes, operational changes, system development changes, restrictions on acceptable technology use, monitoring requirements, or any other effort that will establish a way to recognize, resist, and recover from a high-impact situation. Part of this planning process must include assigning responsibility for each activity and establishing a tracking process as part of the system implementation to make sure protection mechanisms are in place before the system is put into operation.

UNIQUE SUSTAINMENT CHALLENGES FOR COTS COMPONENTS

Vendor licensing management and support must be consistent with the planned rollout and ongoing needs of the system users. The development effort frequently uses a limited license set, and full production licensing needs are delayed to defer the cost outlay. If implementation planning has not included vendor licensing negotiations, there may be some last-minute surprises in the operating costs of the system. For security, maintenance support is a critical component of the vendor license. Patches for security vulnerabilities are distributed as part of the maintenance support. The termination of support for a product that is still fielded can leave the infrastructure vulnerable to subsequent security problems linked to that software or hardware component.

Someone has to be responsible for keeping a sufficient number of licenses current. If licensing management is already assigned within the organization, coordination with additional vendors may not require significant added effort. If this is not the case, a mechanism for tracking where licenses are deployed (which can be as simple as a spreadsheet of usage information) must be established and maintained as updated versions are fielded.

Vendor management for fielded systems can become a major security issue if the lines of communication between the organization and the vendor are unclear. What kind of vendor support will be available? How will vendor upgrades be handled? Will documents, files, or e-mail notifications be sent to an assigned individual, or will someone need to check the vendor's Web site for posted changes? Large organizations with a wide range of fielded COTS products may have standard procedures for handling these issues, and the acquisition process should include negotiation with the vendor to follow these. If COTS products are relatively limited in the operational environment, vendor management functions may need to be established.

Will each vendor upgrade require development support to ensure the components of the system continue to work together? Will the sustainment organization have sufficient capability to perform this testing, and who will handle problems if something does not check out? The necessary resources must be planned and allocated in advance to allow for effective security management. Delays in applying vendor patches to installed products represent security vulnerabilities to an organization's technology infrastructure, but patches can break systems. Organizations must establish mechanisms for addressing these issues in advance.

WHEN DOES A SYSTEM TRANSITION TO SUSTAINMENT?

A fielded system is not necessarily ready for sustainment. Pilot test sites may run for a long time as functionality is finalized and prepared for broader organizational use. However, operational control should begin as soon as the system enters the organizational infrastructure. Operational control of the physical boxes and the software running on them to ensure appropriate hardening is critical. Unprotected machines and software can be compromised in minutes.

Separation of the operational environment from the development and testing space may not be clearly defined for systems in transition. This can quickly become an area of high security risk. Developers building the systems should not have uncontrolled access to the production environment. Organizational policies and procedures are needed to ensure an appropriate level of protection for the information assets.

A FEW ADDITIONAL THINGS TO CONSIDER

If sustainment and operational support are outsourced, will security management be handled based on your organizational requirements, or will you have to adhere to the standard operating procedures of the sustainment organization? Will the level of security control the outsourcer provides be sufficient for the information assets you need to protect? What restrictions are in place to control subcontracting? Will the contracted resources be prepared to proactively address software fixes, vendor patches, and technology upgrades, or is the system only changed when something breaks? All of these issues must be addressed during the contracting process.

SOME LESSONS LEARNED

Sustainment organizations are not well versed in security and reliability of software and systems. Traditionally their focus has been on configuration management of systems with a very stable software base, and component failure involved the replacement of hardware parts.

Sustainment organizations typically provide software upgrades once a year. This may be sufficient if there is a mechanism for critical security patches to be deployed in a timely fashion. Adjustments to this upgrade cycle will incur additional support costs. It is important to have a mechanism in place for determining criticality.

Developers do not sufficiently understand the security risk of widely used development tools. In one instance, systems engineers wanted to use FTP for remote staging of production input files. However, operational management specifically forbade the use of this utility as a matter of policy because of the high number of security compromises reported against this product with tracking organizations such as US-CERT (www.us-cert.gov). Until a team was assembled to perform a security risk assessment, the organization had no forum for bringing the systems engineers and operational management together to discuss the reasons for the policy and identify other alternatives potentially acceptable to both groups. Once team members discussed the issue, they selected another approach that could accomplish the same result while improving sustainment security.

WHAT YOUR ORGANIZATION SHOULD BE DOING

The security challenges for systems in sustainment are growing, and the risk of ignoring them is extremely high. The most effective means of addressing these challenges requires a team of participants from software development, production operational support, and sustainment working together to plan for a successful production implementation. As part of the planning process, a security risk assessment must be performed[1] to pinpoint potential security problems, identify required protection activities, and assign responsibility for timely completion of each activity.

REFERENCES

1. Boehm, Barry W. *Software Engineering Economics*. Prentice Hall PTR, 1981.

2. Institute of Electrical and Electronics Engineers (IEEE). *IEEE 610.12 — IEEE Standard Glossary of Software Engineering Terminology*. IEEE Computer Society Press, 1990.

3. Viennear, Robert L. "The Present Value of Software Maintenance." *Journal of Parametrics*, Vol. 15, No. 1, 25 April 1995, pp. 18-36.

Carol Woody is a senior member of the technical staff at the Software Engineering Institute (SEI) at Carnegie Mellon University. Dr. Woody's research is focused on ways of improving software, system, and system-of-systems design, development, and sustainment that enhance the security and sustainability of fielded systems. Her recent publications include the technical note "Eliciting and Analyzing Quality Requirements: Management Influences on Software Quality Requirements."

Prior to joining the SEI, Dr. Woody was a strategic technology planner for New York City. She has 25 years of project management and systems development experience in large, complex organizations. She holds a BS in mathematics from The College of William & Mary, an MBA with distinction from Wake Forest University, and a Ph.D. in information systems from Nova Southeastern University, where she was elected to Upsilon Phi Epsilon, the international honor society for computing and information disciplines. She is a member of the IEEE Distinguished Visitors Program and serves on the steering committee for the IEEE software engineering portal (SEOnline).

Dr. Woody can be reached at cwoody@sei.cmu.edu.

[1] For the implementation of a family of systems that are closely linked, consider performing a security risk assessment both on the individual systems and on the group to identify potential cascading risks.

Authentication Mechanisms You Can Bank On

by Lee Imrey

In October 2004, BBC News reported that 2,000 British account holders lost £4.5 million in the previous 12 months, according to the British Association for Payment Clearing Services (APACS). From October 2004 to October 2005, the number of new phishing sites detected by the Anti-Phishing Working Group increased from 1,142 to more than 4,300, while the number of reported incidents increased to over 15,000.

Verifying the accuracy of such claims is difficult. The average phishing Web site is online for less than a week before being shut down. The lack of credible actuarial data on losses renders any estimate subject to dispute. However, what evidence there is suggests that the prevalence of phishing attacks, and their financial impact, is increasing at prodigious speed. Potential losses were estimated to range from half a billion dollars to nearly five times that figure, according to a *CNET News.com* article published in late 2004 [1].

IDENTIFYING YOUR CUSTOMER

These numbers speak to the growing problem of reliably identifying parties involved in business transactions. This includes business-to-business (B2B) transactions as well as business-to-consumer (B2C) transactions. While consumer-to-consumer (C2C) transactions are also at risk, as seen in Internet auction and classified-ad sales, losses due to fraud in these transactions currently represent a small fraction of total losses. Furthermore, C2C transactions are frequently facilitated or enabled by a third-party business, such as eBay or PayPal. B2B transactions are generally validated through authentication-and-encryption processes enabled by cross-certification of specified "gateway" machines at each business, or sometimes through use of a "shared secret," which is similar to a password, but generally with explicit requirements regarding length, character set, and frequency of changes.

The remaining category, B2C transactions, represents an increasing threat to online trade, placing substantial corporate funds at risk, as well as compromising the financial security of customers, both online and offline. It is hard to generate a reliable estimate of the number of customers and amount of money at risk, as few companies are willing to provide hard statistics on their avoidable losses unless required by mandate. However, recent legislation such as the California Senate Bill 1386 requires public notice of breaches in corporate databases, subject to certain conditions. Such new legislation may have been instrumental in the decision of the California Health and Human Services (HHS) agency to notify the media about a recent breach of a database containing the names, addresses, phone numbers, Social Security numbers, and dates of birth of 1.4 million California citizens. It is sobering to think how many similar breaches may have occurred in recent years without notification of the victims and potential victims of identity theft.

> **It is sobering to think how many breaches may have occurred in recent years without notification of the victims and potential victims of identity theft.**

IDENTIFYING YOUR VICTIM

One way that criminals can profitably leverage such information is through phishing expeditions. Phishing is a relatively new, and increasingly successful, form of online fraud that exploits customer trust in the corporations with which they conduct financial transactions. This trust is clearly implied by the willingness of most people to place the majority of their money in a bank, trusting that the bank will manage and secure those

DISCLAIMER: The opinions expressed herein represent those of the author and do not necessarily represent, nor are they intended to represent, those of the author's employer.

funds in a fiscally responsible manner. Financial institutions have generally done so, providing customers with timely and reliable access to their deposited funds. Unfortunately, they have also provided access to those who have fraudulently represented themselves as those same legitimate customers.

Early online consumer fraud consisted of a criminal representing himself to a corporation as a legitimate consumer, purchasing goods or services using a stolen credit card number, taking delivery of the goods using a disposable address (such as a commercial mailbox or mail service), and then disappearing, leaving the credit card holder with a significant debt and a damaged credit rating. While financial institutions generally release credit card holders from the majority of the debt, it can take a substantial investment of time on the part of the injured party to repair the damage to her credit rating. In fact, reports suggest that it takes an average of 40 work hours as well as the personal investment of close to $1,000 in charges.

To commit these crimes, the criminal had to have access to certain information about the targeted party, including her full name, credit card number, and credit card expiration date. The expiration date is a recent innovation, added as an extra check that a potential customer actually had the card in her possession at the time of purchase. Most recently, vendors have begun requiring buyers to provide the "card verification value" (CV2), a three-digit number printed on the back of the credit card. This allegedly serves to confirm that the buyer is in possession of the actual card at the time of purchase. However, this strategy suffers from the same inherent weakness as using the credit card number or expiration date to authenticate the customer's identity. Each piece of information can be recorded or duplicated without the cardholder's knowledge.

Clearly, if a criminal has access to a cardholder's name, credit card number, expiration date, CV2, and billing address, he can impersonate the legitimate cardholder in a telephone or Internet transaction. In fact, no matter how many unique identifiers we require for online purchases, those identifiers can be "harvested," in some cases from the very databases that they are being authenticated against (as in the cases of identity theft being perpetrated with the aid of an Experian insider last year).

VICTIMS HELPING THE CRIMINALS

Furthermore, criminal organizations have discovered an alternate method of harvesting information, known as phishing. As mentioned earlier, this is a fairly new technique for fleecing the naïve. The name is derived from the concept of "fishing" the Web for victims.

The principle behind phishing is simple — just ask the cardholder or account holder for the desired information directly. By sending legitimate-appearing e-mails, purportedly from the customer's financial institution (or other trusted businesses such as eBay or PayPal), criminals are able to request directly from the customer all the information they need to access customer accounts. The e-mails range from the inept (e.g., misspelling the name of the bank!) to the professional and persuasive. Some of the latter actually use graphics directly from the Web site of the financial institution and exploit obscure vulnerabilities in Web browsers to make it appear that the customer has established a legitimate connection to his bank. Given all of the recent news items about hackers, Internet crime, and online fraud, it's not surprising that less Net-savvy customers would promptly comply with the instructions in alleged security alerts. Unfortunately, when they are asked to verify their account information to prevent themselves from being similarly victimized, they proceed to inadvertently enable the very crimes they are trying to avoid.

Recent thefts of vast databases of personal information (the hard drive stolen from the Tri-West Healthcare Alliance, a Russian cellular phone company's client database, California HHS, Experian, etc.) will likely lead to further exploitation. The use of specific pieces of information will be used to "validate" the legitimacy of phishing exploits. Customers will receive an e-mail asking them to verify their account number, home address, and phone number using the three-digit number written on the back of their credit card. Many customers will assume that the e-mail is legitimate because, after all, the sender already had the credit card number. They will not connect the information provided with a recent compromise of a credit card processing facility, nor will they understand the significance of the three-digit number they provide as the last key criminals need to defraud their credit card company.

Alternatively, some customers will be willing to provide limited information in response to phishing attempts on the grounds that the information they provide is insufficient to access their accounts. However, they are making the implicit assumption that the information they provide is being sent to their own bank, and that, on the off chance it ends up in the wrong hands, it is all the information to which the unintended recipient will have access. Unfortunately, due to data aggregation with databases available on the black market, the phisher will likely have already collected all the corollary information, needing only the information requested in the e-mail to exploit the target's account.

HOW MUCH IS AT STAKE?

So how many people will fall for this trickery? The unfortunate reality is that it doesn't really matter. As long as some small percentage of account holders fall victim to this deception, it will continue to escalate. This is a simple function of the high profits associated with credit card fraud and identity theft, the minimal marginal cost of collecting and using personal data on cardholders, the comparatively small chance of apprehension, and the minimal deterrent effect of punishment for this type of white-collar crime. Taking a page out of John Allen Paulos's excellent book *A Mathematician Reads the Newspaper* [2], we can make a reasonable estimate of the profitability of this criminal enterprise.

Let's begin by making several conservative assumptions. These numbers are intended to be within an order of magnitude, but I make no guarantee as to their accuracy. If you disagree with one or more of these assumptions, feel free to substitute your own. (If you have data to substantiate your assertion, please contact me at the e-mail address at the end of this article, and I will include any revision in future work.)

$10,000	Cost of stolen database of consumer credit information on the black market
1,000,000	Number of records of consumer credit information/database
100,000	Number of phishing e-mails sent/day
0.1%	Percentage of phishing recipients to respond (1:1,000 response rate based on the assertion that there's one born every minute)
1	Average number of days for recipients to respond (based on the assumption that if the recipient does respond at all, she will do so rapidly)
50%	Percentage of respondents exploited (based on the assumption that some respondents will not be attractive prospects for one reason or another)
$10,000	Average profit per exploited respondent

Based on these rough estimates, the potential profit from a phishing expedition using the data from one database would be in the neighborhood of $4,990,000. Assuming that half of that money is paid off to associates, and that half of the remaining funds are lost laundering the illicit gains, this still represents a tax-free gain of over $1 million. Not bad for two weeks.

For any readers who feel that this is a generous estimate of potential profits and that the response rate would be significantly lower, consider the marginal cost of sending out more spam. If the response rate is only one-tenth of the estimate provided above, the perpetrators need only send out 10 times as many phishing invitations. The marginal cost of sending them, using one of the five primary spamming networks existing today, is inconsequential. Indeed, what cost there is will primarily be absorbed by legitimate Internet users, who will pay for the increased bandwidth utilization of these undesired transmissions.

WHAT CAN WE DO?

Many experts have weighed in on this problem in numerous forums. Technology solution companies, ranging from large organizations such as Microsoft and Cisco to small security startups, have offered numerous tools to prevent phishing. Self-proclaimed "phishing experts" have appeared worldwide, all of whom offer their own unique perspective on this dilemma. Law enforcement organizations seek appropriate means to apprehend and punish the perpetrators.

Yet these circles of interest offer little help and direction for the average businessperson. These groups typically focus on the trends and the magnitude of the phishing problem without offering concrete solutions. This problem is not going to go away, and it is unrealistic to expect businesses to shut their doors in defeat, regardless of the magnitude and severity of the problem.

BUYING THE SILVER BULLET?

Some technology companies offer hardware and software tools that their marketing teams claim will prevent phishing. However, it doesn't seem that any have

(Text continues on page 28.)

SECURITY AND IDENTITY MANAGEMENT AT TATA CONSULTANCY SERVICES

by M. Vidyasagar

Tata Consultancy Services Ltd. (TCS) is the oldest and largest IT company in India. Founded in 1968, TCS has recently seen an enormous growth spurt. While TCS had always been the standard bearer for India's IT industry, the company had grown quite slowly until a few years ago, when the pace changed significantly, due in part to a much publicized and extremely successful IPO.

Until about 10 years ago, TCS had gotten by on the traditional models of security — simple, nonelectronic identification cards for physical security, enforced by trained guards, and logical security provided by whatever password mechanisms came with our computer systems. Interconnection between networks was not ubiquitous as it is now, and personnel traveling from one site to another on business did not have too much trouble with physical access, because the guards could be instructed to recognize identification cards issued by other locations. However, TCS personnel often had little or no access to network resources from their base location when traveling.

Sometime during the last 10 years, physical access control gradually evolved into some sort of card-based electronic security in every location. However, this happened at different points of time in different locations, and there was little homogeneity in these systems, and certainly no integration. By this point, physical access for personnel traveling to other locations became a little more complicated, because the access control mechanisms were not integrated.

There were other problems. It was not only the size of the organization that was dictating communication needs — the industry itself was changing so that everyone expected access to ever larger amounts of information, with ever decreasing latency. This started to present a serious problem in information security.

THE TIMES THEY ARE A-CHANGIN'

All this had to undergo pretty significant changes in order to accommodate the spurt in growth that TCS was planning. We had a much larger number of locations in India now, and we had also opened development centers in South America, Hungary, and China, apart from the existing locations in the US, the UK, and elsewhere.

It was no longer possible, or advisable, to treat security and identity in the comfortable but "old world" fashion to which we were accustomed. We needed to put in place technologies that would help this new, much larger organization deal with its size and ensure its security.

What this all boiled down to is that we needed to create a well-managed, information-rich environment in which all TCS stakeholders could collaborate in real time and to use this environment as a basis for the next generation of enterprise management within TCS. It quickly became obvious that we would have to take a very long-term view of our needs and evolve a solid, future-proof foundation on which the organization could run. We called this foundation the Enterprise Management Program (EMP).

The EMP takes an integrated view, touching on everything from IT infrastructure to collaborative knowledge management tools and processes. For the remainder of this piece, I will focus only on those aspects of EMP that relate to security.

ENABLING THE BUSINESS TO RUN SECURELY

The security stream of the EMP effort was set up to benchmark TCS's security processes with the international best and to equip the work environment at TCS with the best technologies available to ensure a secure and reliable workplace. There are several subareas within the security stream where significant work has been done and continues to be done. I will highlight some of them here.

Identity Management and User Provisioning

The identity management group has created a centralized system for identity creation and management. All possible workflows pertaining to identity within the organization — new employee creation, an employee leaving or retiring, location change (aka "transfer"), role change, and so on — have been provided for. The system allows accounting and auditing of access rights and also offers a dual control check to enable a double sign-off for access rights changes if needed.

SECURITY AND IDENTITY MANAGEMENT AT TATA CONSULTANCY SERVICES (continued)

Password management, including the ability to securely communicate passwords for new users, is available, as is the synchronization of user provisioning on various systems (including Ultimatix [TCS's intranet portal] and the Lotus Notes system).

The identity management system uses smart cards and public key infrastructure (PKI); all associates receive a digital identity in the form of a digital certificate. The certificate is stored on a smart card, which is the primary device for establishing user credentials.

Access Control

The EMP includes a comprehensive model of access control, using PKI for nonrepudiation of accesses to key resources. Single sign-on (SSO) and directory services are implemented, and the associated directory services component also provides a globally scalable repository of users' business privileges. Resource owners grant or revoke access, and the process is automated using the identity management system. Some of these rights may be delegated if necessary.

Single Sign-On

EMP also implements an SSO product across the board to alleviate the need to remember multiple passwords (which, as we all know, is a leading cause of weak passwords). The system provides for multiple and alternative methods for sign-on, centralized credential stores, and various settings for synchronization options and local cache expiration. User session inactivity timeouts, failed login attempt tracking and response (such as user account disablement), real-time logging of all authentication and sign-on events, and detailed audit-related reports are available. Most importantly, the system supports multiple authentication options, including strong passwords, smart cards, and biometrics.

Other Features

In addition to the above, the EMP security stream provides for a security framework within which a central security operations center monitors all network- and security-related activity. There are also detailed incident management procedures defined to identify, prioritize, and track all incidents at various levels and to coordinate incident response activities.

SECURITY: INTEGRAL TO THE ENTERPRISE

The EMP has been a major effort, spanning more than three years of work by a talented team consisting of 27 very experienced professionals who owned various "streams," including security. More than anything, the EMP effort has shown us how important it is to take a holistic view of the enterprise and all of its needs. In the process, we learned that security is not an "add-on" but part of the design of any IT-related endeavor today.

M. Vidyasagar is Executive VP (Advanced Technology) of TCS and is based in Hyderabad, India. At TCS, Dr. Vidyasagar's charter is to set up and develop the Advanced Technology Centre (ATC), which works on cutting-edge technologies of relevance to the IT industry. At present, ATC is involved in three distinct activities: life sciences, e-security, and open source/Indian languages. The e-security division has developed solutions for end-to-end network security and is the originator of the Dhruvam, an enterprise PKI solution. This group has also supplied digital certificates and/or security solutions to more than 30 prestigious clients, including the Bombay Stock Exchange, the Reserve Bank of India, and the Department of Company Affairs.

Dr. Vidyasagar holds a doctorate in electrical engineering from the University of Wisconsin, USA. He is the author of 10 books and more than 130 technical papers. Dr. Vidyasagar has won several awards in recognition of his research, including the "Distinguished Service Citation" from the University of Wisconsin. He was named an IEEE Fellow at age 35, one of the youngest to receive this honor. He is also a Fellow of the Indian Academy of Sciences, the Indian National Science Academy, the Third World Academy of Sciences, and the Indian National Academy of Engineering.

Dr. Vidyasagar can be reached at Tata Consultancy Services, No. 1, Software Units Layout, Madhapur, Hyderabad 500 081, India; Tel: +91 40 5567 3001; Fax: +91 40 5567 2222; E-mail: m.vidyasagar@tcs.com.

(Text continued from page 25.)

offered a compelling and sustainable solution. If they had, major world financial institutions would gladly have invested in it, as that would be a clear market differentiator in today's economic climate.

Unfortunately, the detective measures currently on the market face the same challenges that antivirus vendors and spam-prevention companies have faced for years. Each of these segments is trying to market a plug-and-play "solution" to mitigate a threat caused by an intelligent adversary. By the time a solution gets to market, there is a group of dedicated, well-funded, and highly motivated professionals working to defeat that solution.

What we need is a better way to authenticate the individual.

Having seen the amount of money at stake, it is no surprise that as soon as a technology is developed to detect and prevent a particular type of phishing attack, these adversaries develop new and creative means of fulfilling their criminal goals. Many attempts at addressing the problem of phishing are futile against newer attack methodologies. In fact, some have the unfortunate side effect of aggravating the problem or creating a new one. For instance, the use of Bayesian filters to detect word patterns indicative of spam or phishing attempts have caused many spammers and phishers to send their communications as images. These are both harder to filter and have a greater impact on our limited bandwidth.

Entire industries are chasing their tails in trying to resolve this problem. The tools and techniques for identifying phishing attacks are costly to develop, implement, and maintain. The hardware, software, and personnel costs associated with deployment of such tools in large and complex environments are billed directly against the bottom line of the institution or are passed on to the customers in the form of increased fees. And as I've said, as soon as any solution is implemented, there is a substantial chance that new techniques of attack will be discovered, restarting the vicious cycle.

While this suggests a gloomy picture, let's consider changing our perspective. Perhaps the current focus on phishing itself is not the best approach to solving the problem.

SHOOTING AT THE RIGHT TARGET

So many companies are focused on phishing as the problem to be solved. It would be more productive to recognize phishing as the symptom of a systemic flaw — the lack of effective identification and authentication, which results in the crimes of impersonation and identity theft.

Why is this type of crime increasing so dramatically? As the information systems on which our society depends are becoming more integrated with our business processes and with each other, we are becoming more dependent on the services they provide. As accurate identification of the individual requesting these services is a prerequisite for appropriate allocation of resources, one might presume that as we place increasing reliance on the information systems behind our business processes, we would strengthen the authentication processes to ensure that access to them is only granted to appropriate personnel. However, in authenticating people, we continue to use the same personal information that we have used for decades. Unfortunately, it is no longer enough.

As a society, we have to reduce our reliance on presumably unique, person-specific information as an adequate form of authentication. As soon as our information-processing devices exceeded our own capacity to store and retrieve person-specific information on a large scale, we should have seen that any form of authentication based on rote memorization would no longer be sufficient. Computers are better at it than we are, and computers will work for criminals as willingly as for law-abiding citizens. What we need is a better way to authenticate the individual.

A PROMISING APPROACH

So how can we ensure that the correct individual is authenticating the use of a credit card, or the transfer of funds, or the signing of a contract? Two-factor authentication is one promising approach.

Two-factor authentication, also referred to as strong authentication, is based on the combination of a physical device, such as a key, with a piece of information, such as a password. The key is *something you have*, which you would presumably miss if it were lost or stolen. The password is *something you know*. By requiring something you know in conjunction with something you have, you mitigate the risk of losing the key. While you will not be able to gain access to resources without the key, anyone who gains possession of the key will be unable to capitalize on it, as he would lack the corresponding password. Some forms of two-factor authentication may substitute a personal attribute, such as a fingerprint or a voiceprint, for the key. This is referred to as *something you are*.

Two-factor authentication is not a new idea. In fact, it has been in use worldwide for the past two decades. Most banks now issue ATM cards, which allow the cardholder to access her bank account remotely by using one of over a million ATMs scattered around the globe. Clearly, effective authentication is a business requirement. This requirement is met by issuing the account owner an ATM card that stores a unique code. This card, *something you have*, can be used in conjunction with an assigned or user-selected PIN, a four-digit code. This code is *something you know*. The combination provides adequate authentication for both banks and customers to rely upon for financial transactions.

One known weakness with the two-factor authentication method used by ATMs is its susceptibility to a replay attack. That is, since the PIN is unchanging, someone who can intercept the transmission between the point-of-sale device (such as an ATM card scanner at a drug store) and the authentication-and-funds-authorization services could potentially "replay" the communications sequence and withdraw an equal sum from the cardholder's account. This weakness can be addressed through the use of one-time keying (OTK). The ATM card could contain a processor capable of computing temporary codes based on a time function and a unique key. If the ATM card's internal clock is synchronized with the clock on an authentication server, with protected access to a corresponding key, the authentication server could verify the authenticity of the unique, nonrepeating code, and thus verify possession of the card. This mitigates the risk of a replay attack, as any code used is only valid for a short time.

We Can't Afford Not To

This type of user authentication, though available from vendors for some years, has not been utilized by financial organizations for customer authentication due to economic infeasibility. However, the price point of time-synchronized two-factor authentication in an acceptable form, such as a credit card, has decreased to the point where it is economically infeasible for such organizations *not* to consider it. More to the point, it is now a viable technology and would result in significant long-term savings for financial institutions and their customers.

Now that the cost of two-factor authentication (on a customer basis) is less than 10% of the provisioning cost of setting up a new customer account, and the cost of the supporting technological infrastructure is substantially less than the annual losses written off to phishing and related fraud, corporations should be responsible for insulating their customers from fraudulent scams. It is in the customer's best interest and the economy's best interest.

Moreover, it is in the best interest of financial organizations to invest in the security of their customers, for that is an investment in their own future. As customers recognize the importance of credible authentication mechanisms to their financial security, the use of effective two-factor authentication by financial institutions will initially be a significant business differentiator. As such devices become more prevalent, it will become a business requirement.

REFERENCES

1. "Good News: 'Phishing' Scams Net *Only* $500 Million." *CNET News.com*, 29 September 2004 (http://news.com.com/Good+news+Phishing+scams+net+ionlyi+500+million/2100-1029_3-5388757.html).

2. Paulos, John Allen. *A Mathematician Reads the Newspaper*. Anchor, 1996.

Lee Imrey is an Information Security Architect with the US Department of Justice (DOJ). He writes information security policy to protect critical and classified information, manages the deployment of emerging security technologies, and supports the compliance program to ensure that programs and policies support the department mission. He also works with various government agencies and NGOs to drive common practices and technological procedures consistent with information security best practices.

Before joining the DOJ, Mr. Imrey worked for the International Information Systems Security Certification Consortium, Inc. (ISC)^2 as instructor, editor, and technical manager. He has taught information security to hundreds of students worldwide and continues to contribute time to various professional programs. He is a member of the ASIS Information Technology Security Council, as well as the ISSA Committee on Professional Ethics, of which he was chairperson from 2002 to 2004. He holds several industry certifications, including Certified Information Systems Security Professional, Certified Information Systems Auditor, GIAC Certified Firewall Analyst, and ASIS International Certified Protection Professional. He is completing the MBA program at the University of Chicago Graduate School of Business.

Mr. Imrey can be reached at lee.imrey@usdoj.gov.

Are We Ready to Face Next-Generation Spam?

by Charalampos Z. Patrikakis and Anastasios A. Pallas

There is no official definition of spam, yet most people know it when they see it. We consider spam to be advertisements or, more generally, all the e-mails that we do not want to appear in our mailbox. Formally, spam has three defining characteristics: it uses electronic means such as e-mail, is unsolicited, and comes in bulk.

The two official definitions that best describe spam are unsolicited bulk e-mail (UBE) and unsolicited commercial e-mail (UCE). UBE is the quintessence of spam: messages sent to a large number of recipients who have either not requested to receive them, or have requested to receive mail only once and later asked not to be included in the list of recipients. Any message received after such a request is also UBE. UCE, on the other hand, refers specifically to commercial messages trying to promote products, services, or companies. UBE therefore covers more kinds of spam, such as political spam, frauds, and malicious e-mails.

The term UCE is most frequently used in the US, where national legislation distinguishes this type of spam from the rest. For instance, the most important US spam law — the CAN-SPAM[1] Act of 2003 [5] — introduces a number of requirements for UCE, such as including opt-out instructions, displaying the sender's physical postal address, and labeling the e-mails. Likewise, it also prohibits the use of deceptive subject lines and false e-mail headers, the transmission of UCE after a recipient's objection within 10 business days, the harvesting of e-mail addresses, and dictionary attacks.[2]

According to the final amendments of the European Directive for e-Privacy 2002/58/EC [8], commercial e-mails within the European Economic Area (EEA) are not allowed without recipients' prior consent. But although antispam legislation is advancing, the spam phenomenon is always one step ahead, mutating into new forms of annoyance over mobile communications, instant messaging, Internet telephony (e.g., VoIP), and spoofing of Internet search engines.

CURRENT COUNTERMEASURES

Apart from the exasperation it introduces into the everyday life of Internet users, spam also produces extra costs for businesses, well hidden in direct and indirect costs. Most global spam research shows that spam ranged from 40% to 60% of total e-mails for the year 2004. Some pessimistic studies indicate that rates are between 60% and 75% of total e-mails for the year 2005, if one includes viruses and other unwanted content [13, 19]. Some believe that the situation is improving, due mainly to technological antispamming efforts (filtering, sender authentication, etc.) and law enforcement [15]. Nevertheless, the amount of spam is still unacceptably high.

Not all unsolicited e-mail is spam, of course. Many companies and organizations — including banks, brokerage firms, insurance companies, and universities — depend heavily on e-mail communication to contact their customers or members and to provide information about their services. Still, there is no easy and direct way for business and person-to-person communication to be distinguished from spam. As a result, antispam measures embrace a variety of different approaches. At the conceptual level, we can divide the antispamming methods into the following categories:

- **The technological approach** — tries to eliminate spam by blocking it using computers' software and/or hardware

[1] We should note that, in this case, the term SPAM does not refer to spam messages but rather to (Non-)**S**olicited **P**ornography **A**nd **M**arketing.

[2] A dictionary attack is a technique for trying to guess an authentication mechanism or other secret by running through a list of likely possibilities, often a list of words from a dictionary (see http://en.wikipedia.org/wiki/Dictionary_attack).

- **The legal approach** — emphasizes legislation that prevents the action of spamming and imposes penalties for those who transgress the laws
- **The social approach** — focuses on educating e-mail users about the harm of spam, how their behavior contributes to the problem, and how they can change their everyday routine to reduce the associated risks

We can further distinguish technology-based antispam methods based on where they reside — the server side (i.e., on the mail transport agent of an ISP or on dedicated outsourced hosting providers) or the client side (i.e., the end user's computer). Yet another viewpoint is to categorize technological antispam methodologies into three complementary approaches [12]:

1. Preventing spammers from harvesting e-mail addresses
2. Blocking in the sending stage
3. Blocking at the receiving stage

Preventing Spammers from Harvesting E-Mail Addresses

Users can employ a number of practices to prevent their e-mail addresses from being included in spammers' lists. One tactic for hiding an e-mail address from spammers' robots is to replace the @ symbol and use instead the word "at" or the equivalent HTML entry "&#64;". Likewise, one can replace "." with the word "dot" or "&#46;" (e.g., myname@mycompany.com can be written as "myname at mycompany dot com" or "myname&#64;mycompany&#46;com"). Another important issue is to avoid posting to Web newsgroups, forums, and other untrusted services on the Web with your private or company e-mail. Instead, use a free-hosting e-mail address from Yahoo or Hotmail. Protecting your e-mail address seems to be the most effective countermeasure available [18].

Blocking in the Sending Stage

It would be ideal if we could block spam in the sending stage, but so far the only effective measure here is to keep up to date on security, be protected against spyware intrusions, and regularly check SMTP and proxy relays in order to prevent potential spammers from transforming your server into a zombie.

Blocking at the Receiving Stage

Blocking at the receiving stage (i.e., antispam filtering) is an effective approach, although it too contains some flaws. It is done primarily through two means: sender authentication checks and content analysis. The first involves targeting the sender of the e-mail and establishing the trustedness of the route. Domain Name System Blacklists (DNSBLs), which contain abused e-mail servers like SMTP and proxy relays along with SPF (a new standard for Sender Policy Framework), make it easier to identify spoofs and distinguish authentic messages from forgeries.

Tests that are based on content are mainly classified into message header tests (let us exclude here sender authentication), tests on Spam URL Real-time Block Lists (SURBLs), fingerprint analysis tests or distributed checksums (like DCC), lexical analysis tests, and, finally, statistical analysis tests (like Bayesian filtering). It is amazing that SpamAssassin 3.0.x. has 234 different tests (including sender authentication) that are based on the header, representing 39% of the total 600 tests performed. As far as the effectiveness of SURBLs is concerned, the rates of spam identification are quite high, reaching up to 87% (i.e., 87% of spam contained a URL that was listed in a SURBL) [3, 24].

Having said this, it appears that Bayesian filtering is one of the most prominent methods of spam blocking (at the receiving stage), and almost all antispam vendors have adopted it. This kind of filtering relies on adaptive filtering algorithms, which have several noteworthy benefits. They can be customized to individual users' characteristic spam and legitimate messages; they can generate a filter automatically from a body of categorized messages rather than requiring human effort to explicitly develop rules; and they can be implemented in a very few number of lines of code [14]. Bayesian filtering works surprisingly well, and due to its personalization characteristics, spammers have a very difficult time adapting to it. However, Bayesian-based antispam filters require some attention to work well. This mainly has to do with the training of the filter, which requires the processing of a vast number of messages. We should also keep in mind that antispam content filtering introduces significant CPU workload and delays, especially for large volumes of e-mails.

Wanted: A Holistic Approach

So which one of these antispam methods is the best? The only accurate answer would be: none alone. Each one has advantages and shortcomings, and none of them is 100% effective at blocking spam while maintaining an acceptable false positive rate. Even the Bayesian statistical analysis method (undoubtedly the current best-of-breed approach to spam control) is not perfect.

Most of the methods in use today adopt a technological approach to the spam problem. What is required is an approach that encompasses technology along with policy-based decision making and user education. Any approach to controlling spam obviously needs support from the major stakeholders such as ISPs, and it should be aligned with existing antispam laws. We believe proposals that encompass multilayered spam filtering technologies are more likely to be effective against the spam phenomenon.

SPAM GOES WIRELESS

Spammers' activity clearly demonstrates that they are trying to take advantage of every low-cost broadcast medium available. A mobile phone is something personal, and the last thing you want is inconvenience and solicitations on your phone. For spammers, however, it is just another, even more powerful, medium for promoting their "goods," and we can argue that it is actually the leading direct marketing media available today.

Spam in mobile communications can take the form of the well-known SMS message, MMS message, or even e-mail, which is the favorite application of users of Web-enabled mobile phones [1]. A significant percentage of spam on mobile phones asks the user to call a phone number, which may be a premium-rate service or simply a message of a commercial nature that is considered threatening or intrudes upon a user's privacy. Yet another phenomenon of mobile spam is to send messages in the form of a virus that is harmful or attempts to change the handset settings. These mobile viruses, which are able to affect critical components of the device and effectively make it unusable, have increased rapidly since their first occurrence in June 2004. In 2005, there were more than 100 such viruses [9].

Unfortunately, a recent study shows that both consumers and mobile operators expect mobile spam to become a critical issue in the next two years [4], and they worry about the possible impact of mobile spam on consumers. The good news, however, is that both US and European spam laws [5, 8] already cover mobile communication, thus providing a legal protection mechanism.

In the same study [4], consumers worldwide indicated that mobile spam negatively impacts their opinion of a mobile network operator and that they would rather change their provider than apply for a new cell phone number. In addition, consumers perceive mobile marketing messages from mobile operators as mobile spam, which opens up a controversial issue regarding the legitimacy of operators sending their customers messages for services, competitions, and news. The European Directive for e-Privacy 2002/58/EC [8] requires companies to obtain the consent of the recipient before they send commercial messages. However, there are certain exceptions to this opt-in rule; a company is allowed to use information acquired during the customer's purchase of a mobile device and/or service to promote similar products or services. Since the company has informed the customer about the possible use of personal data for such purposes (a statement that is either well hidden in the contract or is considered as de facto and overlooked), it can use this information to send "legitimate" spam messages. Stopping these messages requires the customer to explicitly opt out (which the law states he may do at any time free of charge), but doing so is not really convenient though mobile phones.

Tracking a mobile user's location is another sensitive topic, as the ability to know the exact location of a user at all times raises the specter of "Big Brother." Privacy concerns are mainly the consequence of the low security levels of mobile and wireless networks, which increases the fears of further m-commerce-related spam activities. This is actually the same concern previously expressed about e-commerce, now exacerbated by the fear of extensive tracking and profiling of customers in the mobile world.

What's the difference for spammers? They now have the ability to "offer" unsolicited location-based services to the target group they want to reach. Of course, their actions imperil legitimate mobile marketing campaigns, which can be beneficial for all players, as they can offer consumers personalized information based on time, location, and interest [23].

SPAM IN INSTANT MESSAGING (SPIM)

SPIM is a new flavor of spam that has appeared in instant messaging (IM) systems such as ICQ and MSN Messenger. Spammers gather information from a user's directory, sign on with misleading names, and send unsolicited messages. Likewise, in a situation identified as "messenger spam," spammers exploited a security vulnerability of Windows "Messenger Service" designed for system and networks administrators to send notifications over a Windows network. As a result, annoying pop-ups were opening on computers connected to the Internet, displaying security warnings and threats instructing the user to follow a certain URL to work around the "problem." Microsoft, confirming this threat, issued a security workaround for all Windows operating systems [3, 16, 17].

But what makes SPIM unique? It's the speed of action, since IM, unlike e-mail, is a real-time communication suite. SPIM can be carried out in the form of an attack (e.g., a computer worm) whose speed can be measured in a few minutes and that can spread much faster than security solutions can respond. That's why SPIM is likely to become a very popular means of spamming in the near future. Indeed, based on a recent US survey, 30% of online adults who use IM have already been subjected to SPIM [11, 20].

SPAM OVER INTERNET TELEPHONY (SPIT)

Telephony over the Internet, or Voice over IP (VoIP), means cheaper phone calls, which is a great prospect for consumers and businesses alike. On the other hand, it also means that audio advertising messages can be sent out for very little cost, much like spam in e-mail communication. Spam over Internet Telephony (SPIT) is potentially one of the most imminent threats associated with VoIP (once the technology becomes widely used), and we can compare it with telemarketing campaigns. How does it work? Spammers discover the IP addresses used for voice, upload an audio file (the spam) to a server, and send it to a large number of voice mail in-boxes. If you are already nervous about spam through e-mail, you have good reason to get mad about voice spam, because you will have to listen to it!

Some people may argue that this scenario is unlikely. The victims of e-mail and IM spam are just one click away from a spammer's destination Web site, but what about SPIT? How exactly are SPIT senders going to earn money? Unluckily, there are two potential ways to achieve that: (1) they can persuade the listener to call them back (we should not forget that even in the case of untrustworthy e-mail, where the victim has time to think, there are still people who respond); or (2) trick her to press a number and let her speak with a real salesperson, who knows well how to handle the rest of the successful contact [6, 7]. In any case, having costs near zero, SPIT senders have nothing to lose!

SPAM GOES FROM ANNOYING TO THREATENING

With the rise of phishing, spam has shifted from annoying to threatening. Phishing refers to the acquisition of Internet users' credit card numbers, bank account numbers, and other sensitive personal information by means of bogus messages purporting to be from legitimate businesses. Because the referred Web sites look practically identical to the companies' real sites, users are misled into divulging critical information, such as usernames and passwords.

According to a recent study conducted by the Anti-Phishing Working Group (APWG), the most targeted industry is the financial sector, though there was no lack of phishing activity targeting customers of ISPs and retail e-shops and even donors to the Red Cross after Hurricane Katrina! The report further states that the average time online for a scam Web site is five to six days, and the US, China, and South Korea together share more than 50% of the hosting phishing Web sites worldwide. Europe and Canada were reported to be the new future targets of phishing spammers [2]. With the number of unique phishing Web sites constantly rising and methodologies varying from confirming personal information, to keystroke logging, to legitimate URLs redirecting users to deceit Web sites (a phenomenon called "pharming" [22]), phishing attacks are expected to impact consumers' confidence in e-banking and e-commerce [2, 10].

If you are already nervous about spam through e-mail, you have good reason to get mad about voice spam, because you will have to listen to it!

POSSIBLE COUNTERMEASURES FOR THE NEWEST SPAM

When attempting to tackle the mobile spam problem, we should first underline that there is much to learn from our experience with e-mail spam, although there are a number of differences in each situation. For instance, mobile spam messages, in contrast to e-mail spam messages, are hard to filter at the client end. Thus, efforts should focus on prevention and blocking messages before they reach customers' phones. In addition to compliance with high-security standards, mobile carriers have to prove their sensitivity to privacy concerns in order to promote m-commerce and other value-added mobile services. Due to the limited number of mobile operators (versus ISPs), the possible countermeasures listed below can be adopted to combat mobile spam. Some of them have already proven very effective in Japan, where operators have dealt with vast amounts of mobile spam. These countermeasures include:

- Blocking messages sent to a large number of recipients

- Blocking messages if a number of recipients are invalid or a dictionary attack is discovered
- Blocking messages from specified sources (i.e., domains), similar to DNSBLs
- Blocking access to premium-rate services included in the spam messages (which is something similar to URLs identified in SURBLs)
- Limiting the amount of messages sent daily from a single mobile phone
- Providing the users (free of charge) with the ability to block specific phone numbers or domains of e-mail senders, or even all e-mails from the Internet
- Bringing spammers and sponsors of the spamming action immediately to court and being committed to taking legal action

Regarding SPIM, some basic protection rules for IM users are [21, 25]:

- Activating alerts when other users add you into their contact lists
- Blocking strangers from seeing when you are online
- Allowing only people you know to send you messages
- Avoiding opening attachments, even from people you know if you haven't confirmed that they sent you something
- Finding out how to activate the scanning of all files with your installed antivirus program

Finally, in order to be protected against phishing and pharming [10], some good practices are:

- Staying up to date with antivirus software and firewalls on every computer you use
- Installing an antispam tool on your computer, even if your e-mail provider offers one
- Setting up a public e-mail address to be used when entering chat rooms, forums, online dating sites, job search sites, auctions, and generally surfing the Internet
- Checking out the privacy policy before giving away any personal information, even to trustworthy companies
- Not unsubscribing to suspicious lists, as you are just confirming that your e-mail address is valid
- Not replying to any e-mail that asks you to verify your personal data
- Never giving out sensitive information such as passwords and credit card numbers, Social Security number, and bank account numbers to forms inside e-mail messages or to Web sites without a valid security certificate

Ultimately, none of the above guidelines and practices will totally solve the spam problem, but each one is a step in the right direction and can save us from further troubles. It's essential for IT directors, managers, and technical staff to be aware of the risks posed to security and productivity by spam. Lifetime training programs for employees along with informational leaflets and published tips on the corporate intranet are just some ideas for educating users on ways they can reduce both personal and enterprise risks from spam.

CONCLUSION

The new mobile "cousins" of spam presented in this article seem to be just in their infancy, but that doesn't guarantee they won't be as troublesome as e-mail spamming in a very few years. Spam over Internet and mobile telephony, for instance, would be significant problems inside a corporate voice network, especially as they can raise common deficiencies such as latency, dropped calls, or distortion, which can lead to the disruption of the phone service, a mission-critical application for all enterprises.

We have witnessed how difficult the regulation of e-mail spamming has proved to be, since this is generally opposed to businesses' aim of increased sales. In contrast to what many researchers and antispamming companies initially thought, the contribution of users in combating the spam phenomenon is critical. We strongly believe that *spam tends to be a social and ethical issue rather than a technological issue only*, and regulation efforts should be focused on that aspect, too. Let us not forget that all problems on the Internet start as a small snowball but quickly turn into a huge snowstorm. Isn't it time to rethink our strategy?

REFERENCES

1. Accenture Institute for Strategic Change. *The Future of Wireless: Different than You Think, Bolder than You Imagine*. Accenture, 2001.

2. Anti-Phishing Working Group (APWG). *Phishing Activity Trends Report*. APWG, September 2005 (http://antiphishing.org/apwg_phishing_activity_report_sept_05.pdf).

3. The Apache SpamAssassin Project. "Tests Performed: v3.0.x" (http://spamassassin.apache.org/tests_3_0_x.html).

4. Brodt, Torsten, and János Heé. *Insights into Mobile Spam: World's First Collaborative Empirical Study*. University of St. Gallen, Switzerland, 2005.

5. CAN-SPAM Act of 2003, Public Law 108-187, 108th Cong., 1st sess., 16 December 2003 (www.spamlaws.com/pdf/pl108-187.pdf).

6. Cyber Security Industry Alliance (CSIA). *Cyber Security for IP Telephony*. CSIA, May 2005.

7. De Guzman, Mari-Len. "Spam May Be a Future Threat to VoIP." *ITWorld Canada*, 7 September 2005 (www.computerworld.com/networkingtopics/networking/story/0,10801,104442,00.html).

8. EuroCAUCE. "Final Amendments to E-Privacy Directive" (www.euro.cauce.org/en/amendments1a.html).

9. F-SECURE. "Barrier of 100 Mobile Viruses Reached." Press release, 10 November 2005 (www.f-secure.com/wireless/news/items/news_2005111100.shtml).

10. Hinde, Stephen. "Spam: The Evolution of a Nuisance." *Computers & Security*, Vol. 22, No. 6, September 2003, pp. 474-478.

11. Keizer, Gregg. "Expect Bigger IM Attacks After Microsoft, Yahoo Merger." *TechWeb News*, 12 October 2005 (www.techweb.com/wire/showArticle.jhtml?articleID=172300558).

12. Koh, Eung Lyeol. "Anti-Spam Toolkit: Existing and Emerging Technical Measures against Spam." Slide presentation. Spam Response Team, Korea Information Security Agency, 8 September 2004.

13. Levitt, Mark, Robert P. Mahowald, Brian E. Burke, and Christian A. Christiansen. *What You Can and Should Do About the Rising Cost of Spam*. IDC, March 2004.

14. Mertz, David. *Spam Filtering Techniques — Six Approaches to Eliminating Unwanted E-Mail*. IBM, 1 September 2002 (www.ibm.com/developerworks/linux/library/l-spamf.html).

15. MessageLabs. "Spam Intercepts Timeline: September 2005" (www.messagelabs.com/Threat_Watch/Threat_Statistics/Spam_Intercepts).

16. Microsoft Corporation. "Disabling Messenger Service in Windows XP." 9 January 2004 (www.microsoft.com/windowsxp/using/security/learnmore/stopspam.mspx).

17. Microsoft Corporation. "Messenger Service Window That Contains an Internet Advertisement Appears," revised 15 March 2005 (www.microsoft.com/windows2000/techinfo/administration/communications/msgrspam.asp).

18. Morris, John. "Best Practices for End Users." Presentation to *Internet Engineering Task Force (IETF) 56*, San Francisco, California, USA, March 2003 (www.ietf.org/proceedings/03mar/slides/asrg-3/index.html).

19. MX Logic. "Spam Classification Techniques." MX Logic, October 2004 (http://searchsecurity.techtarget.com/0,293857,sid14_gci1010908,00.html).

20. "Nearly a Third of IM Users Encounter Spim." *TechWeb News*, 23 February 2005 (www.techweb.com/wire/security/60403038).

21. ninemsn Pty Ltd. "How You Can Help Reduce Your Instant Message Spam." 2005 (http://microsoft.ninemsn.com.au/protectfromspam.aspx#messenger).

22. Phan, Matthew. "'Pharming' Is Latest Internet Security Threat." *Inc.com*, 20 June 2005 (www.inc.com/criticalnews/articles/200506/pharming.html).

23. Scharl, Arno, Astrid Dickinger, and Jamie Murphy. "Diffusion and Success Factors of Mobile Marketing." *Electronic Commerce Research and Applications*, Vol. 4, No. 2, Summer 2005, pp. 159-173.

24. The Spamhaus Project. "Effective Spam Filtering" (www.spamhaus.org/effective_filtering.html).

25. wiredsafety.org. "Instant Messaging Safety — Basic Safety Tips" (www.wiredsafety.org/safety/chat_safety/im/index.html).

Charalampos Z. Patrikakis received his Dipl.-Ing. degree and his Ph.D from the Department of Electrical and Computer Engineering of the National Technical University of Athens. He is currently working in the field of computer networks research in national, European, and international projects. Dr. Patrikakis's main research interests are in the area of IP service design and implementation, multicasting in IP networks, IP transport protocols and peer-to-peer networking, and streaming media distribution algorithms and protocols. He is a member of IEEE, a member of the Greek Computer Society, a certified trainer by the National Accreditation Centre of Vocational Training Structures and Accompanying Support Services, and a member of the Technical Chamber of Greece.

Dr. Patrikakis can be reached at the Telecommunications Laboratory of the National Technical University of Athens, Heroon Politechniou 9, Zographou, Greece 15773; Tel: +30 210 7721513; Fax: +30 210 7722534; E-mail: bpatr@telecom.ntua.gr.

Anastasios A. Pallas is an IT professor in secondary education in Greece. He is currently finishing his M.Sc. in IT at the University of Paisley, Scotland, in collaboration with the Technological Educational Institution of Pireaus, Greece. Mr. Pallas's research focuses on investigating the spam phenomenon in Greece and evaluating the effectiveness of comparatively unexploited techniques such as spam traps. He has experience consulting on and deploying Internet/intranet-based application projects, and his broad interests include information security, Web services, e-commerce, e-learning, and the social impact of the Internet.

Mr. Pallas can be reached at the School of Computing, University of Paisley, PA1 2BE, Scotland; E-mail: tpallas@sch.gr.