IBM X-Force® 2011
Mid-year Trend and Risk Report

September 2011

IBM

Contributors

Producing the IBM X-Force Trend and Risk Report is a dedication in collaboration across all of IBM. We would like to thank the following individuals for their attention and dedication to the publication of this report.

Contributor	Title
Bryan Casey	Market Manager—IBM Security Solutions
Carsten Hagemann	X-Force Software Engineer, Content Security
David Merrill	STSM, IBM Chief Information Security Office, CISA
Dr. Jens Thamm	Database Management Content Security
Dr. Ashok Kallarakkal	Sr. Manager—Product Management and Beta Ops
Jason Kravitz	Techline Specialist for IBM Security Systems and E-Config
John Kuhn	Senior Threat Analyst, MSS
John C. Pierce	Threat Intelligence Analyst—AI, MSS
Jon Larimer	X-Force Advanced Research, Malware
Leslie Horacek	X-Force Threat Response Manager
Marc Noske	Database Administration, Content Security
Marc van Zadelhoff	Director of Strategy, IBM Security Solutions
Mark E. Wallis	Senior Information Developer for IBM Tivoli Security
Michelle Alvarez	Team Lead, MSS Intelligence Center(aka Eagle Eyes)
Mike Warfield	Senior Wizard, X-Force
Ory Segal	Security Products Architect, AppScan Product Manager
Patrick Vandenberg	Manager, Rational Security & Compliance Marketing
Pete Allor	Senior Cyber Security Strategist
Phil Neray	Data Security Strategy, InfoSphere Guardium & Optim
Ralf Iffert	Manager X-Force Content Security
Randy Stone	Senior Incident Response Analyst
Ryan McNulty	IBM Managed Security Services & SQL Querier Extraordinaire
Scott Moore	X-Force Software Developer and X-Force Database Team Lead
Scott Van Valkenburgh	Market Segment Manager, IBM Security Solutions
Tom Cross	Manager—X-Force Strategy and Threat Intelligence
Vidhi Desai	Product Marketing—IBM Security solutions

About X-Force

The IBM X-Force® research and development teams study and monitor the latest threat trends, including vulnerabilities, exploits and active attacks, viruses and other malware, spam, phishing, and malicious web content. In addition to advising customers and the general public about emerging and critical threats, X-Force also delivers security content to help protect IBM customers from these threats.

Contents
Section I

Contents
Section II, III and IV

Navigating the Report

Welcome. This year we have made some usability improvements to the format and content of the Trend Report. These improvements are designed to enable readers to draw practical applications from the findings. We understand that computer and network security is about focusing on awareness of threats and helping to protect the systems and networks from these threats. But then what? As an organization matures in its stance on computer security and known threats, how can they begin to develop a deeper focus towards improvement?

We asked ourselves that question and determined the answer is to provide to our readers a deeper understanding of what we experience and have learned from the breadth of capabilities that is IBM Security Solutions.

For this report we divided the content into four sections.

- Threats
- Operating Secure Infrastructure
- Developing Secure Software
- Emerging Trends in Security

We start by talking about the **Threats** that our systems and networks are facing, because we have to begin by understanding the problem we are all working to solve. Once a threat is understood, we can work towards realistic technology controls and educational awareness to help secure our enterprise and systems. In both the **Operating a Secure Infrastructure** and **Developing Secure Software** sections we discuss threats and provide logical advice on how to help improve or detect those threats in your environment. In the **Emerging Trends in Security** section, we explore and examine the emerging technologies that are pressing into discussions as future business concerns.

X-Force believes this new layout better organizes the material we want to present, and helps you focus on what is most important to your organization.

Section I
Threats

In this section we explore threat-related topics and describe the enterprise attacks that security specialists face. We address malicious activity observed across the spectrum by IBM and how we help protect networks from those threats. We also update you on the latest attack trends that IBM has identified.

Executive overview

Sometimes, to find the way forward, we must look to recent history so that we can correlate, understand, and assimilate the lessons and trends we encounter.

In the trend report released at the end of 2010, we began discussing what we at IBM call the Smarter Planet.

"A world that is more interconnected, intelligent, and instrumented. As much as these innovations can increase our efficiency and ability to instantly connect on a global scale, so too can the risks and dangers of a connected world become more sophisticated and difficult to contain."

Little did we know that 2011 would provide such an acute, first hand demonstration of just how interconnected we are, and how this confronts us in our day-to-day world. Enterprises and governments are being shown on a near-daily basis how the decisions we make in the cyber-world can affect our physical world.

An unprecedented number of high-profile security breaches reported throughout the first half of this year demonstrate the potential weaknesses within technology and the impact a breach can have on enterprises. Each new breach reinforces the awareness that basic network security is not just a technical problem, but rather a complex business challenge where risk exposure, communication, end-user education, and technology must be considered in a delicate balance.

The attackers of networks and enterprises are also adjusting and evolving from fairly indiscriminate groups who want to break into as many networks as possible using off-the-shelf tools, to highly targeted and sophisticated attackers who study their targets, biding their time for entry into high-value networks and data. Political hactivism that we reported on in 2010 continues to evolve. The actions witnessed by security breaches in the first half of the year, are blurring the lines between political values and moral standards, to simply attacking companies based on apparent personal bias.

As major botnet operators are taken down and off-line by law enforcement officials, we see a trend in the decline of spam and traditional phishing tactics. We discuss the continued success of law agencies addressing these botnet take downs and how these actions are changing the manner in which criminals make money. Are these declining methods now forcing malicious operators to consider more lucrative choices such as spear-phishing specific targets?

The mobile and smartphone journey continues its integration into the enterprise with a few key topics. First, we report that many enterprises have moved past initial discussions around basic enablement decisions and are now dealing with the new generation of security-related topics we discussed at the end of last year. The maturity of how large enterprises approach this enablement journey becomes more important as role-based mobile policies for the enterprise are considered and implemented. Secondly, mobile vulnerabilities, exploits and malware continue to grow rapidly as the rate of user adoption soars.

The security discussion evolves into a deeper dive towards understanding risk exposure, natural disaster mitigation (such as in Japan) and how a high-level security breach can effect even the most common of businesses. It's not a matter of asking "Why would they attack us?" but rather how each company must take the self-responsibility to say "Are we prepared when this happens to us?" The presumption of "Could it happen?" turns to the reality of "When it happens, how will we respond?"

We believe this mid-year report will help organizations better prepare for the changes we face.

2011 highlights

Threats

Malware and the malicious web

- An explosion of security breaches has opened 2011, and near daily reports continue to mark this year as the "Year of the security breach." Page 10
- SQL injection continues to be a favorite attack vector among malicious groups, as demonstrated by the numerous mass SQL injection attacks occurring over the past several years. Page 23
- Top high-volume signatures from IBM Managed Security Services (MSS) demonstrate that favorite attacker methods are SQL injection, and the brute forcing of passwords, databases, and Windows shares that continue to rank at the top of MSS sensor traffic. People are scanning the Internet for open services and attempting to break into them. Page 25
- SQL Slammer worm, once at the top of MSS sensor traffic, has fallen down the list after a dramatic "disappearance" that occurred in March 2011. Page 27

Web content, spam, and phishing

- In the first half of 2011, anonymous proxies have steadily increased, more than quadrupling in number in comparison to three years earlier. Anonymous proxies are a critical type of website to track, because they allow people to hide potentially malicious intent. Page 34
- In 2011, spam volumes continue to decline with the critical take down of the Rustock botnet. Page 40
- Top countries for spam origination have changed this year. India now dominates the top of the list by sending out roughly 10 percent of all spam registered today. Behind India follows Russia, Brazil, and South Korea, with Indonesia rounding out the top five. The USA, which was at the top of the list in 2010, is in the number ten position with less than three percent of spam sent. Page 48
- In the first half of 2011, spammers said adieu to traditional email phishing. When looking at the percentage of spam that is phishing on a weekly basis, we have measured less than 0.01 percent each week. Page 49
- The top phishing email-originating country in 2011 is the USA at 41.5 percent, followed by the United Kingdom at 6.8 percent. Page 50
- In the first half of 2011, financial institutions continue to be the number one target for phishing attempts, representing 69 percent of the targeted industries up from the 2010 year-end report when it was 50 percent. Page 51
- In 2011, as reported previously, North America is still the number one region for email phishers. However, in the second quarter, Europe increased significantly, reaching nearly 30 percent. Page 52

Operating a secure infrastructure

Vulnerabilities and exploitation

- In the first half of 2011, we saw fewer total security vulnerability disclosures than we saw last year at this time. The volume of security vulnerability disclosures appears to follow a two-year alternating cycle. Page 58
- 2010 saw the largest number of vulnerability disclosures on record, over 8500. This year it appears on track for just over 7,000 disclosures, a significant decrease from last year but approximately the same amount that was seen in 2006. Page 58
- For the past few years approximately half of the security vulnerabilities that were disclosed were web application vulnerabilities. The number is down to 37 percent this year, with a significant drop in the volume of SQL injection vulnerabilities in particular. Page 59
- So far only about 12 percent of the vulnerabilities that have been disclosed have seen exploit releases, whereas in previous years the number was closer to 15 percent. Page 61
- Security vulnerabilities with a Common Vulnerability Scoring System (CVSS) score of 10 out of 10 are up to three percent for the year and have already exceeded the total for 2010. Almost every one of these critical vulnerabilities is a serious remote code execution issue that impacts an important enterprise-class software product. Page 65
- Two areas that have seen significant increases are vulnerabilities in document readers and multimedia players. As the browser market has become more competitive, attackers zeroed in on software that consumers are running regardless of what browser they prefer—allowing attackers to net the highest number of victims with a particular exploit. Page 67

Developing secure software

Web application vulnerabilities

- The IBM Rational® Application Security Group research tested 678 sites (Fortune 500 plus 178 popular websites on the Internet.) Out of the websites tested, 40 percent (271 sites) contained client-side JavaScript vulnerabilities. Page 75
- Out of the vulnerable applications, 90 percent included one or more vulnerabilities that were introduced through third-party JavaScript code, such as marketing campaigns, code that embeds Flash animation, and AJAX libraries. Page 77
- DOM-based cross-site scripting (3214 issues out of 3683) is still the single most common security issue type. Page 78
- A new type of vulnerability was detected for the first time: DOM-based email Attribute Spoofing. This vulnerability occurs when a web application uses JavaScript code to automatically craft an email for the user to fill in and send, using user-controlled data. In such scenarios, an attacker could potentially manipulate the content, subject, or CC and BCC fields of the email, resulting in a leak of private information. Page 78

Emerging trends in security

Mobile

- The first half of 2011 saw an increased level of malware activity targeting the latest generation of smart devices, and the increased number of vulnerability disclosures and exploit releases targeting mobile platforms seen in 2010 continues into 2011, showing no signs of slowing down. Page 68
- Mobile devices are quickly becoming a malware platform of choice. This malware increase is based on premium SMS services that can charge users, a rapidly increasing rate of user adoption, and unpatched vulnerabilities on the devices. Page 79
- Two popular methods of malware distribution models are to create infected versions of existing market software and to publish software that claims to be a crack, patch, or cheat for some other software. Page 79
- Besides sending SMS messages, Android malware has been observed collecting personal data from the phone and sending it back to a central server. This information could be used in phishing attacks or for identity theft. We have also seen Android malware that has the ability to be remotely controlled by a remote command and control server—just like a bot that infects a Windows desktop machine. Page 80
- Enterprise security management of mobile endpoint devices will struggle to handle massive expansion. One solution may be the convergence of endpoint security configuration management to incorporate all these new devices. Page 81

Database security

- The old adage still holds true—"Why do you rob banks? Because that's where the money is..." A company's data must be continuously connected to its customers, partners, and employees; however, that exposes sensitive data to more automated and targeted attacks than ever before." For example, we're now seeing numerous attacks that easily bypass traditional perimeter defenses by exploiting web application vulnerabilities such as SQL injection, or by leveraging stolen administrative credentials, to compromise back-end databases. Page 84
- Databases have become an important target for attackers. Critical data used to run our organizations—including financial/ERP, customer, employee, and intellectual property information such as new product designs—is stored in relational databases. Page 85

IBM security collaboration

IBM Security represents several brands that provide a broad spectrum of security competency.

- While the X-Force® research and development teams are busy at work analyzing the latest trends and methods used by attackers, other groups within IBM work to use that rich data to develop protection techniques for our customers.
- The IBM X-Force research and development team discovers, analyzes, monitors, and records a broad range of computer security threats and vulnerabilities.
- IBM Managed Security Services (MSS) is responsible for monitoring exploits related to endpoints, servers (including Web servers), and general network infrastructure. MSS tracks exploits delivered over the Web as well as other vectors such as email and instant messaging.
- IBM Professional Security Services (PSS) delivers comprehensive, enterprise-wide security assessment, design, and deployment services to help build effective information security solutions.
- Our content security team independently scours and categorizes the Web through crawling, independent discoveries, and through the feeds provided by MSS.
- IBM has collated real-world vulnerability data from security tests conducted over the past several years from the IBM Rational Services team. This data is a combination of application security assessment results obtained from IBM Rational AppScan® with manual security testing and verification. From requirements, through design, code, and production, IBM Rational AppScan provides comprehensive application vulnerability management across the application lifecycle.
- IBM Cloud Security Services allows clients to consume security software features through a hosted subscription model that helps reduce costs, improve service delivery, and improve security.
- Identity and access management solutions provide identity management, access management, and user compliance auditing. These solutions centralize and automate the management of users, authentication, access, audit policy, and the provisioning of user services.
- IBM data and information security solutions deliver capabilities for data protection and access management that can be integrated to help address information lifecycle security across the enterprise.
- IBM endpoint management solutions combine endpoint and security management into a single offering that enables customers to see and manage physical and virtual endpoints—servers, desktops, roaming laptops, and specialized equipment such as point-of-sale devices, ATMs and self-service kiosks.
- IBM InfoSphere® Guardium® provides a scalable enterprise solution for database security and compliance that can be rapidly deployed and managed with minimal resources.

2011—Year of the security breach

The first half of 2011 has been marked by a litany of significant, widely reported external network security breaches, which are notable not only for their frequency, but for the presumed operational competence of many of the victims. The questions that executives in every industry are asking their security teams are "What is going on?" and "Can this happen to us?" To answer those questions one must start with an understanding of the different groups responsible for these attacks and their motivations and capabilities.

Who is attacking our networks?

External threats can be categorized according to the focus of their attacks as well as their level of operational sophistication. Some network attackers are fairly indiscriminate; they want to break into as many computer systems as possible regardless of where they exist. Others are targeted; they have an interest in penetrating specific victim networks. Some botnet operators lack sophisticated technical skills and mostly know how to use a tool chest of exploit and malware kits they have purchased. Others work in well-organized, state-sponsored teams that discover new vulnerabilities and develop totally unprecedented attack techniques.

For several years now, the most common external network threat is financially motivated malware and botnet builders—attackers who infect millions of computers with malicious software that steals credit card numbers, sends spam, and launches denial of service attacks. These attacks are broadly targeted and they tend to employ known, off-the-shelf attack techniques. In the past we have seen new zero-day vulnerabilities being used in conjunction with botnet construction, but it appears that black market sellers of these exploits are currently being outbid by buyers with different kinds of aspirations.

There can be a great deal of money to be made in broadly targeted botnet activity and that has attracted a large number of players to the game from all around the world. The statistics published over the years in the X-Force Trend Report show the signs. Broadly targeted botnet attack activity is so commonplace on the Internet that it shows up prominently in the statistics we publish. We have also published charts showing malicious exploits from toolkits, botnet command and control activity, as well as the effects of these botnets, such as the growing volume of spam and distributed denial of service attacks. Keeping all of this activity out of an enterprise network—by keeping up with vulnerability patches and detecting attacks at the perimeter—can be a significant challenge, and the threat landscape appears to be getting only more complicated.

Attacker Types and Techniques 2011 H1

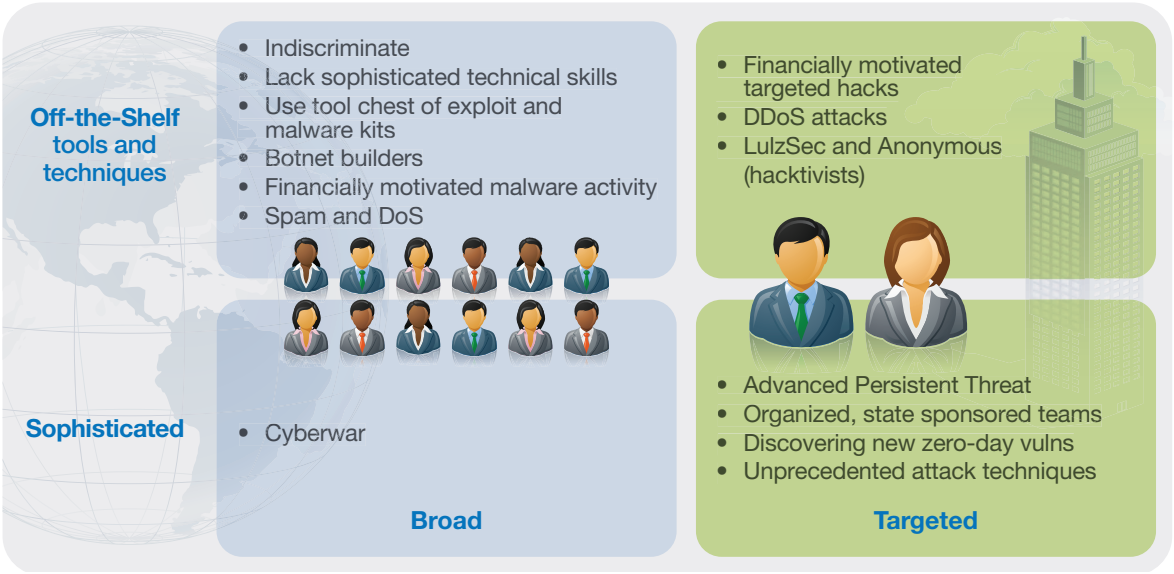


Figure 1: Attacker Types and Techniques – 2011 H1

Advanced persistent threat

Although government organizations have always worried about the threat of state sponsored computer intruders, it is apparent that both large and small private enterprises also face this type of threat. A number of prominent publicly reported breaches in 2010 and early 2011 appear to drive this point home. Sophisticated, targeted, state sponsored attacks (which are sometimes referred to using the moniker Advanced Persistent Threat) are at the opposite end of the spectrum from the financially motivated botnet operators. These attacks usually are specifically targeted. They often show evidence of extensive pre-operation intelligence collection and careful, patient, long-term planning. They often also involve never-before-seen vulnerabilities and obfuscation techniques. The standard prescription of keeping up with patches and running commercial security products typically does not defend a network from this kind of adversary.

Previous IBM X-Force Trend and Risk Reports as well as publications on the **X-Force blog** have delved into the topic of defense from sophisticated attackers. The approach is a paradigm shift from the usual "audit and patch" approach to network security, one which involves developing operational assurances that the network is invulnerable to known attack techniques. Sophisticated attackers may at times, employ unknown attack techniques that you are not prepared to prevent, so the focus must shift to detection and analysis.

Your organization should be willing to embrace approaches to detection that may not be 100 percent effective. For example, many organizations shy away from investing heavily in educating their end users about spear phishing because they know that end-user education may not be completely effective. However, it is somewhat effective, and when you are dealing with a stealthy, sophisticated attacker, those detections that you get from a partially effective detection process are detections you wouldn't otherwise have. **In the next section of this report, we dive deeper into the differences among Phishing, spear phishing, advanced persistent threat and targeted network attacks.**

Once you've detected something, the next step is forensic analysis. The normal inclination when you discover a breach is to clean it up so you can go along with business as usual. When you are dealing with a sophisticated attacker, it might be better to suppress that instinct. You are running counter intelligence. The attacks that are being launched against you are custom built to target you, and you need to learn as much about them as possible, even if you've managed to detect them before they were successful. You might want to let them continue to unfold, with the knowledge that you are watching what is happening. In this way, forensic analysis becomes an everyday part of your operational approach to security.

In other words, when dealing with APT, prevention is eventually going to fail you—deep analysis is going to be necessary, so you should plan ahead and have capacity to handle these types of events. You should do so with the understanding that the information gained through deeper analysis can provide successful prevention of future attacks and at the same time creates a feedback loop for improvement. The information that you gain from analysis can lead to the detection of other compromises that you don't yet know about. It's the feedback between detection and analysis that helps get you ahead of your adversary.

For a deeper understanding of the APT topic we recommend you **download this discussion on APT** from the IBM 2011 Pulse event.

Undermining common security practices

Many of the most prominent breaches of 2011 were committed by attackers who don't fit into either of the previous two categories. Hacktivist groups, such as Lulzsec and Anonymous, lack the technical and operational sophistication of state sponsored attackers and lack the financial motivation of the botnet operators, but they have been very successful at breaching networks and damaging reputations. These groups fit into a category of attackers who commit targeted attacks using known, off-the-shelf techniques. Of course, we have also seen financially motivated attacks that fall into this category, such as the litany of breaches attributed to Albert Gonzales.

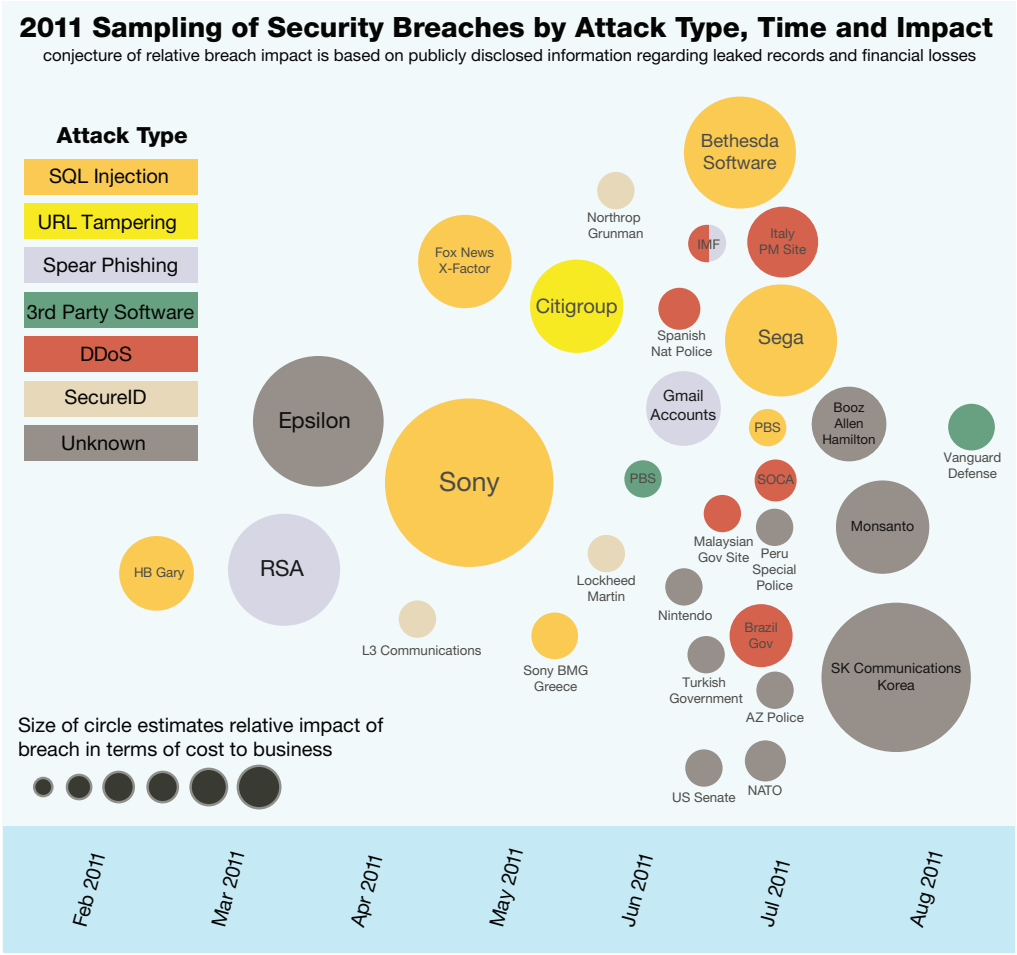


Figure 2: 2011 Sampling of Data Breaches by Attack Type, Time and Impact – 2011 H1

As illustrated in figure 2, many of the 2011 published data breaches[1] were successful in spite of the fact that well-known techniques, such as SQL injection, were often a factor. Some of these attacks might not have been successful had victim organizations consistently followed best practices with regard to computer security.

Common points of entry

There are two common IT points of entry into every corporation. The first and most obvious is the public website and data servers. Every page and script on every public facing website, as well as every other Internet facing service, is an opportunity for a motivated individual to find a hole. The second point of entry is employee workstations or endpoints. Every employee with access on a corporate network is a potential target for an attacker.

The challenge of locking down an organization's public facing Internet presence is complexity; large corporate websites may contain thousands of scripts, written by many different departments. A simple temporary promotional page may contain a flaw that can lead to a full-scale exploitation of a company's internal network.

1 2011 Cyber Attacks (and Cyber Costs) Timeline (Updated):
http://paulsparrows.wordpress.com/2011/06/28/2011-cyber-attacks-and-cyber-costs-timeline-updated/
http://blog.thomsonreuters.com/index.php/cyber-attacks-timeline-graphic-of-the-day/
http://www.ponemon.org/blog/post/cost-of-a-data-breach-climbs-higher

A good security policy for what gets put on a public server is a great first start. Scripts should be audited using web code scanning software, and every single input form should be routinely scanned for common injection and cross-site scripting vulnerabilities.

Third-party scripts and applications such as blogging software or forums can be particularly vulnerable to security issues since more popular applications are used by many businesses and attackers can use a single exploit on many different sites. Most of these holes can be scanned automatically and attackers can then hone in on vulnerable servers. These servers in turn can be used to exploit the end business, or to inject malware onto the legitimate pages to infect all visitors to those sites.

In many cases, one small weakness can lead to further opportunities for attackers to find ways into a network. For example, in one recent breach, a common file-include vulnerability was reportedly used to access the file system on a web server. The attackers claim to have then discovered SSH keys on the server, allowing authentication to other servers. This paved the way for a full-scale takeover of the company's network[2].

In several breaches, poor passwords or reused passwords resulted in wider scale exploitation. When a user reuses the same password in multiple places, their Intranet, their personal email, social networking sites, the corporate VPN, and the company website, it becomes much easier to gain a larger foothold into an enterprise which may have otherwise been secure. Strong password enforcement is essential with a policy against password reuse and periodic expiration enforcement.

The second common IT entry point is the endpoint. By using email-based social engineering attacks, it is possible to persuade people within an organization to click on malicious links, allowing back doors or other malware to be installed on their endpoint. This can provide a further launching point into the infrastructure. The endpoint can be targeted by attackers of all classes, from financially motivated attackers looking to build botnets to sophisticated, state sponsored attackers wielding carefully crafted social engineering attacks. Clearly, keeping endpoints patched and up to date is a critical security task.

Not a technical problem, but a business challenge

None of the above recommendations should surprise seasoned Internet security professionals. Regular external penetration testing, web application vulnerability testing, web application firewalling, effective password policies, end-user education, network policy enforcement, encryption, and intrusion prevention should have identified and closed many of the holes that were instrumental to newsworthy breaches that have occurred this year. So why did they happen?

Because, this sort of basic network security is not just a technical problem, it is a business challenge. Large organizations have complicated network operations and these networks are constantly changing. A significant effort is often required to inventory, identify, and close every vulnerability, and keep them closed on an ongoing basis as the network grows and changes. These efforts meet with resistance, not just financial but operational. The business wants to get business done and not waste time dotting every "i" and crossing every "t" in the security auditor's checklist. How much of an investment in all of this is enough?

The answer to that question has evolved over time as prominent breaches teach us that we still haven't found the right investment level. Compliance programs have created a useful lowest common denominator for security programs but experience has demonstrated that compliant networks, systems, and infrastructures can still have real security gaps. What would happen if you coupled the sophisticated technical and operational capabilities we associate with Advanced Persistent Threat, with the broad focus of financially motivated botnet operators?

2 LulzSec versus Bethesda & Senate.gov: http://pastebin.com/i5M0LB58

This brings us to the cyber-war scenario that has been the focus of a lot of hand wringing in policy circles over the past few years. We define cyber-war as sophisticated attackers using broad based computer attacks to achieve tactical and strategic advantages by damaging the infrastructure and capabilities of their enemies. Fortunately, we have not seen a lot of that kind of activity. However, the fact remains if we can't repel a cyber-army of hacktivists using off-the-shelf techniques, then we are not prepared to handle a more formidable threat.

Clearly, there is more work to be done in external network security. Getting that work done boils down to making sure you have an up-to-date understanding of the gaps in your external network security profile. With that understanding, you can more effectively communicate the risks associated with those gaps to executive decision makers within your organization. You can also develop a plan to address those gaps and obtain the resources that you need to execute on that plan. Ultimately, effective network security programs work because they have the right level of political support within an organization. Considering the right approach to developing that support should be an ongoing dialog within the security community.

Key lessons learned

One key lesson that IBM has learned in working with clients on this problem is that it makes sense to prioritize efforts that create the most visible and demonstrable results with minimal impact on existing business processes. You should build a marketing campaign for your security efforts, wherein you focus on achieving meaningful milestones, and communicating to the business when each milestone is achieved, focusing on the benefits accrued. Through repeatedly demonstrating success, you can raise awareness about and support for your efforts. This support is important when the time comes for more significant investments, or when some disruptive change to an ongoing process is required.

None of this can happen quickly and the clock is ticking. The longer gaps exist within defenses, the greater the probability of compromise. It is probable that most of the organizations that have been breached over the past few months had real and vital network security programs in place at the time they were breached. And the race isn't over for any of them—or any of us; the scale and complexity of the threats faced by our networks grows with each passing year.

IF IBM X-FORCE® WAS RUNNING THE IT DEPARTMENT

Many readers have asked, if IBM X-Force were running the IT department and saw what happened this year, what would you do? Well, here are ten actions beyond the basics that X-Force would do if we ran the IT department.

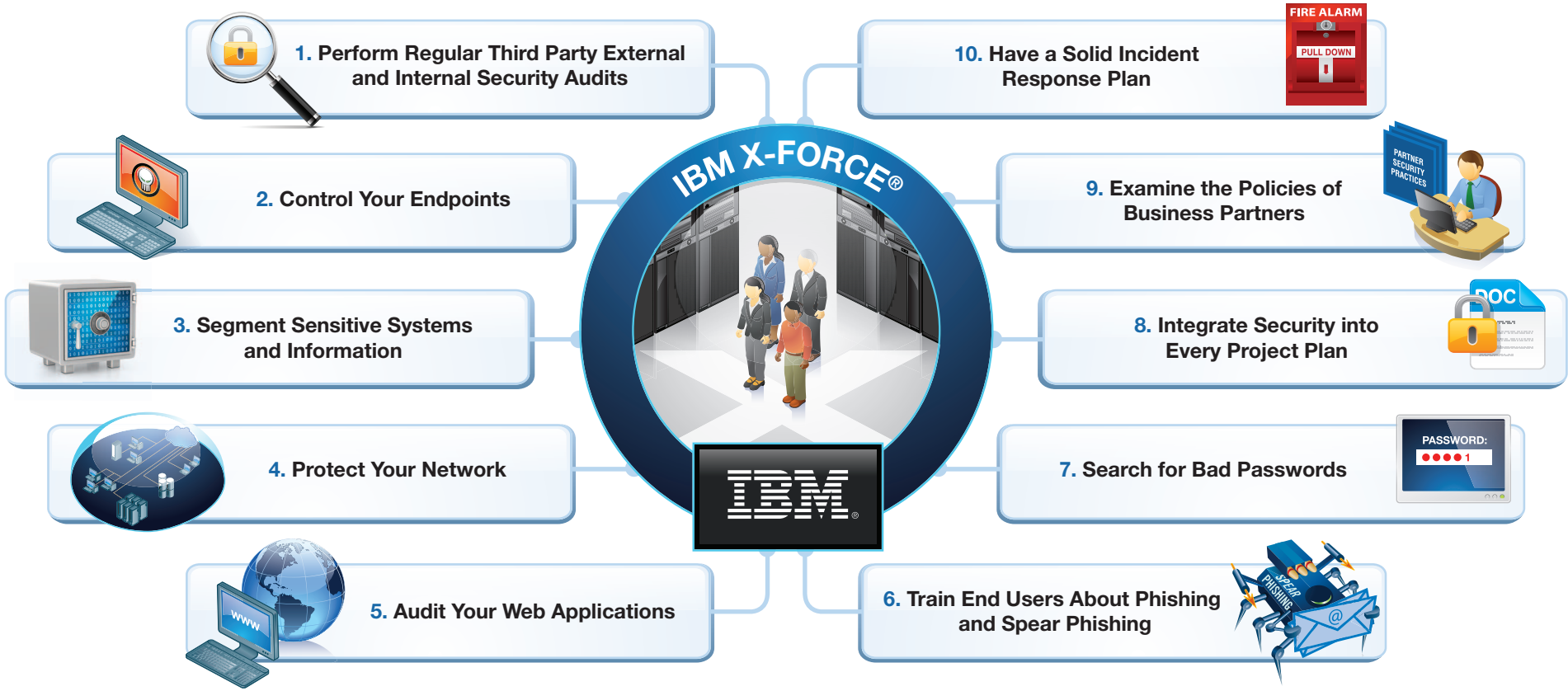


Figure 3: If IBM X-Force Was Running the IT Department

1. Perform regular third party external and internal security audits

Your network is constantly changing. When new security problems are introduced, you need to find them before the bad guys do. Regular third-party security audits coupled with constant vulnerability assessment and scanning are the best ways to ensure that you understand the complete landscape of your network and where the weaknesses are located.

2. Control your endpoints

Do you know what systems you have in your network, what software is running on them, and what patch levels and configurations you have? To what depth? The closer you can get to total endpoint awareness and control, the more secure your infrastructure should become. Do you have a dynamic IT environment that allows you to keep up with security fixes or do you struggle to patch systems due to lack of resources, legacy code, or custom code that is incompatible with the latest technologies? Legacy systems and long patch deployment cycles can become a security liability.

3. Segment sensitive systems and information

In environments where people work with particularly sensitive information, such as classified data centers, employees are typically given separate desktop systems for web surfing and doing email versus the real work. You may not be working with classified information in your office, but it still makes sense to eliminate unnecessary interconnectivity between sensitive data and insecure networks, particularly if your organization is targeted by sophisticated attacks. It's important to keep in mind that interconnectivity takes many forms, such as USB tokens.

4. Protect your network

You need to understand what resides in your network, and you also need to understand who has access. Breaches often happen in areas where intrusion prevention systems were not deployed or were not carefully monitored. When breaches occur, successful investigations depend upon having access to rich log information. The more you are monitoring your network and the more you know about what has occurred in the past on your network, the better prepared you are for breaches.

5. Audit your web applications

Web application vulnerabilities continue to be a common gap that is targeted by attackers of every motivation and skill level. Whether a web application was developed in-house, purchased from a software vendor, or downloaded from the Internet, if it is running on your network, you need to check it for vulnerabilities. If you don't, someone else will do it for you.

6. Train end users about phishing and spear phishing

Many sophisticated attacks involve social engineering or a spear phishing element. Attacks may target personal as well as business accounts and systems. Savvy users may suspect that something is out of the ordinary. If your organization knows that it could potentially be targeted, employees are more likely to report something suspicious rather than ignore it.

7. Search for bad passwords

Even after decades of experience, bad passwords remain a common security weakness. Security audits may make cursory attempts to find bad passwords but constant, pro-active efforts to crack bad employee passwords are much more comprehensive, particularly when coupled with effective policies and end user education. For an example of intelligent thinking regarding password policies, see this recent cartoon: http://xkcd.com/936/

8. Integrate security into every project plan

The security team must not operate on a footing in which they are constantly chasing down projects that have just “gone into production” by introducing massive security gaps into the network that happen to show on a vulnerability assessment report. Security must be applied into new infrastructure from the beginning. Achieving this requires political finesse—the security organization should be enabled and not a bureaucratic barrier. The security team must constantly demonstrate its value to the rest of the business at all levels.

9. Examine the policies of business partners

In this world of cloud computing and complex outsourcing relationships many of the systems you are responsible for may be operated by other companies. Many “insider” attacks come from employees who work for business partners of the targeted firm. Has your security team audited the practices of your partners? Are their practices consistent with yours? How confident are you in their execution?

10. Have a solid incident response plan

Eventually, prevention fails. Managing sophisticated, targeted attacks is an ongoing process that involves not just being able to identify that a breach has occurred, but being able to respond and investigate, learn and adapt. If you are an important strategic target and you are not aware of any breaches, it may mean you are not looking carefully enough.

Phishing, spear phishing, advanced persistent threats, and targeted network attacks

Introduction

Last year a major cyber intrusion incident involving Google occurred and was termed "Aurora." Before the Aurora incident, the term "Advanced Persistent Threat" or APT received only light passing mention in cyber security circles and the news in general. The term itself originated with the military years earlier and was occasionally mentioned and discussed privately and in professional circles, but after Aurora, it seemed to be all over the press and it has been in the news since, tied to a number of intrusions.

In contrast, the term phishing has been bantered about in cyber security news for many years. Spear phishing has also been a common term for years, though not in use nearly as long as the more common phishing.

Spear phishing has been associated closely with recent APT incidents. The Google Aurora incident became a well-publicized APT incident which turned out to be initiated through a spear phishing attack.

These terms have been subject to mixed usage in the press recently, and even in professional discussions, such that the meaning of each tends to blur. Phishing sounds very similar to spear phishing. Misuse of these terms seems to be on the rise of late. But they are quite different from one another and we should pay careful attention to properly using, understanding, and recognizing them.

Recently some observers have introduced the term Targeted Network Attacks as an alternative to Advanced Persistent Threat in an attempt to reduce some of the confusion regarding that definition.

Phishing

Phishing derives its name from the analogy of fishing in a large lake. You cast your line into that lake and you do not care if there are 10,000 fish in there that do not find your bait tasty. You also do not care if you catch the biggest fish in the lake. All you care about is the half a dozen or so reasonably sized fish who will take the hook and make your dinner. Your dinner is in numbers, not size.

Phishing in general is the method of attacking users by pretending to be a legitimate, trusted site like a bank or email service or some store with which the user might be doing business[3]. Some phishing can lead to banking fraud where the attackers rifle a victim's account while others can lead to malware sites attempting to further compromise the victim.

Phishing relies on mass mailings with relatively little personalization beyond, possibly, a customized address in the "to" field and maybe a name in the subject and message body. They are often sent in bulk from botnets and mass mailers. Phishing may look rather unprofessional with a number of spelling or grammatical errors. It may appear to come from an institution with which the recipient may or may not have a business relationship. Many of these types of things may serve as red letter warnings that this email is not legitimate. Often, phishing attacks lead the users to malware, like fake Anti-Virus and Trojans, or malicious URLs run by attackers trying to hijack connections. Because they are mass mailings, the various security organizations and services rapidly pick up on the malicious sites, URLs, and software and it is quickly detected. The people behind phishing really do not spend a lot of time preparing these messages and their sites may only be up for a few hours before they are taken down by security organizations. Senders of phishing emails really do not care if 99.99 percent of the people receiving the email trash them.

While phishing originally referred to email activity, it has since grown to include instant messaging, social networking pages, and other delivery mechanisms by which a user can be duped into visiting a malicious site masquerading as something they should trust.

3 Phishing: http://en.wikipedia.org/wiki/Phishing

Spear phishing

Spear phishing is a form of phishing[4] but, outside of sharing part of a name and a core paradigm, bears little resemblance to common phishing. In contrast to common phishing, it is highly targeted.

Spear phishing is highly directed and targeted at relatively few and very specific individuals within an organization. Spear phishing is deeply customized and personalized to make it appear as though it has come from a legitimate friend or business colleague. The attackers know the targets well and may spend considerable time and effort in studying these targets and crafting the attacks. It probably will not appear to be a generic message from a large institution. Rather, it may seemingly come from an individual friend or colleague with whom there has been frequent past messages. It may even relate to recent events or activities both individuals would know about. The message itself may not even be "spoofed" but may actually come from the other individual's compromised account. The malware or malicious site the email leads to will not have been mass mailed so the security and anti-virus organizations may be unaware of the sites and software. In short, these people have picked their targets and tools carefully. The attackers have decided they want you and have put a great deal of effort into getting you. Consequently, it must be worth the while of the attackers. The yield percentage must be much higher, 0.01 percent success is just not going to cut it. The attackers require a significant percentage of the targets to fall for the trap. The value of the target also must be much higher, to provide a better "return on investment".

To extend the fishing analogy, this is the fisherman standing, spear in hand, in the water, on a dock, or in a boat, watching a really big fish he wants. He waits patiently while watching the fish's movements and learning those movements. When the time is right, the spear phisher acts quickly and decisively either getting the fish or not. If not, he picks another fish or another spear. A lot of time and effort goes into this type of fishing. His dinner is more about the size of the fish rather than the number of fish.

Like phishing, spear phishing can take place over a number of delivery channels and not just email or Instant Messaging.

Whaling, in regards to spear phishing, is a term that is used occasionally to describe spear phishing which specifically targets the highest level officers within a company. They are targeting the biggest fish in the organization. This may not necessarily be the C-level officers with valuable financial information, but could be well positioned people of authority or people with high levels of access to the data. The term whaling is not as popular as spear phishing and most often you will hear spear phishing referring to any such targeted attack whether it is a high level officer or some common employee.

Over 2000 years ago, Sun Tsu wrote, in The Art of War "Attack where there is no defense. Defend where there is no attack." The first part is easy to understand and is the attacker's view. Attack where you see a weakness and a weakness is where the defender is not protected. The second part is not as easy to understand and has been subject to interpretations through the centuries. One interpretation is that he is merely saying the same thing from the view of the defender only telling the defender to protect those weak areas which may be perceived by an attacker and come under an attack.

We can take this advice to defend against spear phishing which may target any employee, high or low within the organization. It is not just the high-level, high-value executives who may be targeted but anyone an attacker perceives as giving a foothold into the enterprise and behind your security perimeters.

Advanced persistent threats

While phishing and spear phishing may seem very similar and use a few similar techniques, they target a very different audience, have a very different attack profile, and come from a very different class of attacker. One area where a spear phishing attack may be utilized is in the opening activities of an Advanced Persistent Threat, APT[5]. While precise numbers are impossible

4 Spear Phishing: http://www.webopedia.com/TERM/S/spear_phishing.html
5 Advanced Persistent Threats: http://en.wikipedia.org/wiki/Advanced_Persistent_Threat

to come by, it is believed to be large organizations behind APTs that devote significant specialized teams of attackers and spend months studying a number of employees at all levels of an organization. They study the target employee's family and friends and professional acquaintances looking for weaknesses. They may only find a few but a few is enough. They may compromise a friend's account and study their past correspondence with the chosen target to mimic their style and behavior or maybe even take part in follow up to a real conversation if the opportunity arises. The real target gets a very legitimate looking email from a friend with an update to a file he recently sent or some pictures of the kids or some video they are both interested in viewing. Nothing that would match a similar email the guy at the next desk just received from a business colleague from a conference he attended two weeks before. And it may not be email. It could be in a chat with a friend on IM. The attacker knows the targets well enough to mimic friends. No real red flags here. The resulting compromise starts off very small and very quiet, low and slow, but expands, as opportunities arise, to other systems and other people within the organization. The malware may involve one or more zero-day vulnerabilities, unknown to the greater security community or even the general underground community. The malware goes to significant lengths to not disrupt anything or disturb things or draw attention to it. Finally, after months of preparation and months of spreading into vital areas of the target organization, it can begin to exfiltrate data and can sustain itself from eradication efforts making it a very persistent threat by the time it is discovered. The essence of this working is that the early attacks, even the ones that failed, do not get detected or reported. The victims do not know they have been attacked.

With APTs, early detection is the best defense, the earlier the better. Spotting a spear phishing attack is challenging even for those who are prepared. Even the very differences between phishing and spear phishing can lead people into a false sense of confidence that they know how to spot a sophisticated spear phishing attack. People should verify things received unexpectedly, even from a friend or family member and should not blindly trust URLs. A simple reply back to someone saying "hey thanks for that video, yeah that was great" may be followed by the remark "uh, what video?" Those sorts of replies should be done "out of band" since you don't know if it would be the attacker reading and deleting that reply. People should be trained to spot and report any suspicious activity like this, even peculiar instant messaging chats with people they know. If they suspect something is wrong, encourage them to report this activity. Never penalize someone for reporting that they may have just fallen for a trap. Your entire organization is your eyes and ears for these sorts of attacks. You will not be able to rely on your anti-virus to spot the malware or your anti-spam software to block the email.

After the compromise, it may be some of the subtlest of clues that lead you to a compromise, such as unusual DNS activities which we illustrated in a focused series on DNS in the **Managed Security Services daily assessments** last year. But, you have to be looking for those clues to see them or they will easily be lost in the noise of everyday business activities. Be paranoid. Even the paranoid have enemies.

Targeted network attacks

The term Advanced Persistent Threat has spawned some controversy in the security world. One of the questions is whether APT is a proper name for a specific group of attackers or a description of a type of attack methodology. Those in the "proper name" camp face the challenge of how to refer to attacks that have similar characteristics but have been launched by other groups of attackers with different motives. Often correct attribution of Internet attacks can be its own challenge. Complicated situations can arise wherein financially motivated botnet operators sell access to a compromised network to more sophisticated players. Those in the "methodology" camp face the challenge of describing situations where sophisticated attackers use different types of attack techniques, which they certainly have the capability to do. Furthermore, the word "Advanced" has caused some disagreement in the "methodology" camp as security professionals often view that term from a technical perspective rather than in terms of the operational sophistication of the

attacker, and whether or not a particular technique is "Advanced" can be a matter of opinion.

The term Targeted Network Attack describes a pervasive attempt to maintain control of the computer network of a targeted organization, regardless of who is doing the targeting or why. It therefore provides a way to characterize the situation faced by an organization irrespective of who is launching the attack, or how "Advanced" their tools and techniques appear to be. Whether adoption of this term as a more neutral alternative to "Advanced Persistent Threat" will take off within security circles, remains to be seen.

"There is no patch for ..."

When the subject of phishing and spear phishing comes up, invariably someone will ask "how could anyone be so stupid?" That question may be understandable for common phishing. It is not quite so applicable to spear phishing and APTs, however. Spear phishing and APTs are highly sophisticated. They are not so easy to identify.

We have many common derogatory terms used in cases where someone makes a mistake and falls for a trap such as "operator headspace," "the nut that holds the keyboard," "PEBKAC" (Problem Exists Between Keyboard And Chair, or "PICNIC" (Problem In Chair, Not In Computer). These terms are summed up in a comment we see in a lot of presentation slides when it comes to human error—"There is no patch for stupid." But, these terms may disregard the sophistication of a number of these attacks and doing an injustice to some of the individuals ensnared. They may even be making the problem worse.

By categorizing these problems as such we may be giving people a false sense of confidence that they would never fall for something like that. They won't be stupid. But the attackers are not stupid either and they are picking their targets carefully and crafting their attacks. The person who falls for these may not have been stupid but merely unprepared and they may have been unprepared because of excessive references to these being stupid.

By categorizing these problems as such, we may put victims on the defensive. They have heard the snide remarks and here they are or they suspect (but are not sure) that something bad might have happened to them. Do they dare tell anyone and risk ridicule for falling for a trap? They should be encouraged to report anything out of the ordinary. We should be cautious about terminology and emphasize that some of these attackers are good and getting better.

Examples from the news

Google and Aurora [6]

When Google first announced it had uncovered a major intrusion, the term APT had not become nearly so common as it is today. Over the course of weeks and months, information about the extent of the intrusion and how long it had been present along with details of the malware, were disclosed. It became more apparent that this attack was out of the ordinary. Finally, in June of 2010 at the Annual General Meeting of the Forum of Incident Response and Security Teams (FIRST), Heather Adkins of Google delivered an outstanding presentation detailing the attack, its discovery, and the following incident response. She described how it was only detected by data mining the extensive DNS logs and that the malware, while basically mundane, had been recompiled to evade detection by their anti-virus defenses. She also described how they used forensic analysis and further data mining to track back to the original "patient zero" compromise and further followed the attack forward as newer versions of the malware were reintroduced into their systems.

The original attack was a focused Instant Messaging (IM) attack. The attackers had done a lot of research and compromised a friend's account. The attack methodology was advanced in its research, even if the malware itself was not. The attackers were very persistent. This fit the entire criterion for an APT.

6 For Google, DNS log analysis essential in Aurora attack investigation: http://searchsecurity.techtarget.com/news/1514965/For-Google-DNS-log-analysis-essential-in-Aurora-attack-investigation

Stuxnet[7]

Stuxnet is considered by a number of security professionals to be another example of an APT but there is some controversy about that designation. Stuxnet does not appear to have been triggered through spear phishing. The attack methodology was advanced and the attackers had done a great deal of research into their targets. But the triggering attack (apparently infected USB keys and possibly other mechanisms) was broad and not focused on particular individuals even though the grand target was very specific. Once the attack was underway, however, it had the "low and slow" infiltration approach and communication of keeping quiet and doing no harm while it worked its way into the network and looked for data to exfiltrate and finally its ultimate mission of doing damage to certain industrial control systems. The malware was updated several times while exhibiting a tenacious persistence. This would appear to fit the behavior of an APT but did not involve spear phishing.

RSA[8]

The attack against RSA certainly appeared to be similar to that against Google. Initially, RSA said very little about the compromise or to its extent of damage / penetration. After some time went by, they announced that they had been hit by an APT attack that had been caught early. It has also been reported that the attackers used a phishing attack with a malicious attachment targeting a zero-day vulnerability, but the emails that have been disclosed publicly do not appear to have been carefully targeted. It has also been reported that the attackers have used data gleaned from RSA in attacks against other targets including some large Department of Defense contractors. This was a serious compromise that did not merely impact RSA, but a wide array of other organizations that depend on the technology RSA produces to protect their own infrastructures.

Conclusion

Phishing, spear phishing, APTs and targeted network attacks seem here to stay for the foreseeable future. As explained, phishing and spear phishing are quite different in scope and execution. Spear phishing does not always indicate an APT, and APT attacks may not always be synonymous with targeted network attacks. Understanding the intricate differences between these techniques helps provide better understanding, education, and remediation.

What is most worrisome about these attack types is that we are witnessing a paradigm shift and an unprecedented assault on the fabric of trust. These series of attacks against the physical supply chain are now targeting the intellectual, e-commerce as well as software and firmware that sustain our networks.

Everything that we know or do regarding the Internet is impacted as the human element represents the strength in seeing what can be made, as well as the weakest link and easiest point to overcome.

Organizations must realize that some, not all, of our data has to be protected. Not only in defensive technologies, but also in monitoring and reviewing at all stages within the data's life cycle. Further, we need input from others to know what is happening across the Internet so we can cue in on the quick changes in techniques using these processes. This awareness will allow us to better monitor, forensically chase, and remove the plethora of arrayed threats and attacks on the organization at its many endpoints.

7 How digital detectives deciphered Stuxnet, the most menacing malware in history: http://arstechnica.com/tech-policy/news/2011/07/how-digital-detectives-deciphered-stuxnet-the-most-menacing-malware-in-history.ars
8 RSA: Anatomy of an attack: http://blogs.rsa.com/rivner/anatomy-of-an-attack/

IBM Managed Security Services—A global threat landscape

IBM Managed Security Services (MSS) monitors several billion events per day in more than 130 countries, 24 hours a day, and 365 days a year. The global presence of IBM MSS provides a first-hand view of current threats. IBM analysts use this wealth of data to deliver a unique understanding of the cyber threat landscape. This section focuses on SQL injection, JavaScript activity, brute force attacks, and scans among other threats that are discussed throughout this report. The trend of these threats is vital to determining what directions threats are taking and to understanding the significance of the threats to our networks.

SQL injection attack activity in 2011 H1

A SQL injection vulnerability occurs when user input is improperly filtered allowing an attacker to execute SQL commands on a target server. Specifically, if escape characters are not filtered, an attacker can alter a query to produce unintended results. SQL injections often lead to information disclosure or the ability to alter information stored in the database.

SQL injection has witnessed a slight upward trend in 2011 (see Figure 4). It is too early to tell what the overall trend will be for 2011. It continues to be a favorite attack vector among malicious groups as demonstrated by the numerous mass SQL injection attacks we have seen over the past several years.

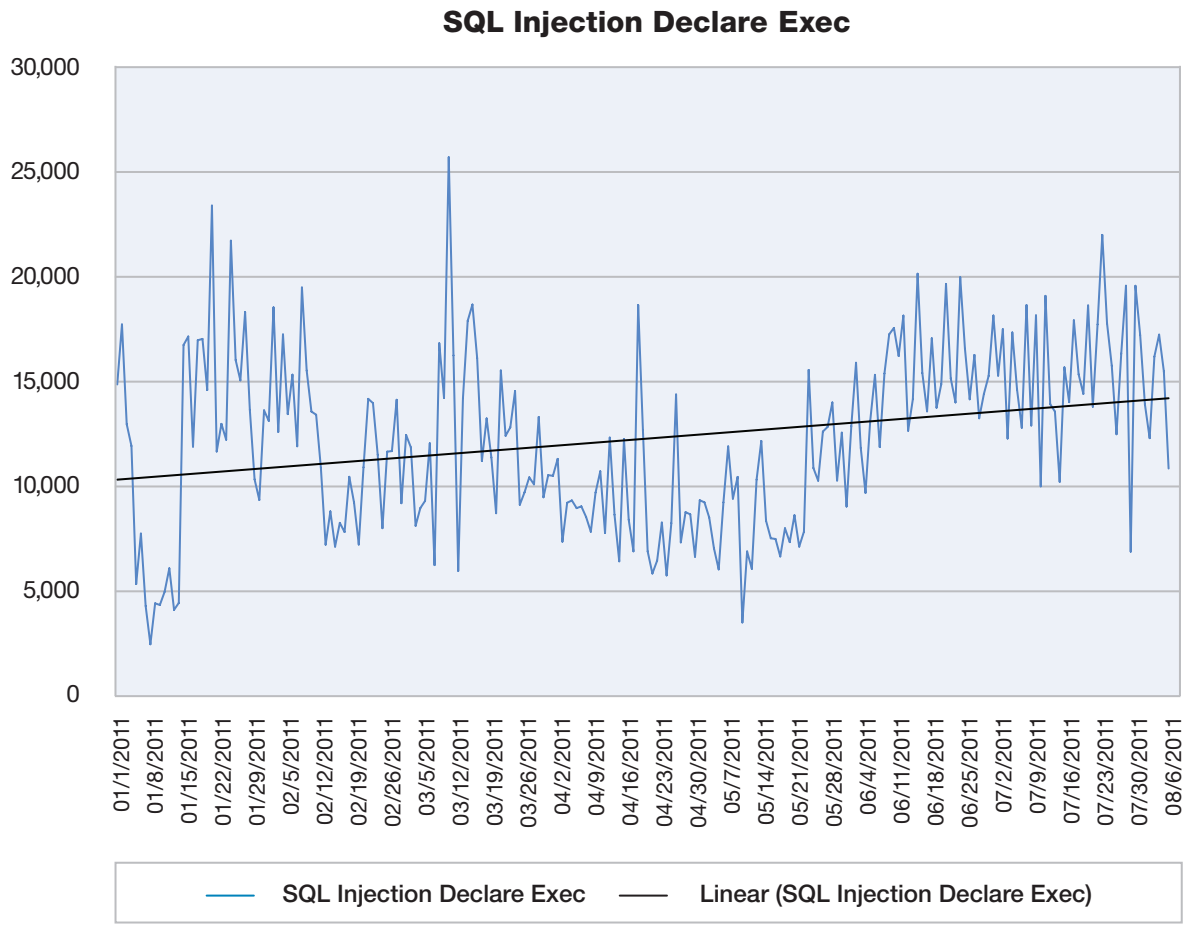


Figure 4: SQL_Injection_Declare_Exec Activity January 2011 – June 2011

However, it is possible that the activity may slow or decrease resulting in a flat or downward trend for the full year.

Historically, targeted SQL injection attacks have focused primarily on data—obtaining unauthorized access or altering it. However, beginning in 2008, we began seeing multiple coordinated and simultaneous mass compromises of tens of thousands of websites using SQL injection as the attack vector. Rather than pilfering data, the attackers injected custom HTML content containing invisible iframes that redirected users to sites hosting drive-by exploits and other malicious content. **The IBM X-Force 2010 Trend and Risk Report** highlights the various mass attacks observed over the past three years.

This trend continues in 2011. In March, IBM Managed Security Services began tracking a mass SQL injection attack named **LizaMoon** because of a URL that was ultimately injected into the SQL tables of the target site. While this attack reached noteworthy levels, we did not see near the volume of activity that we saw with the "Asprox" (2009) and "dnf666" (2010) attacks. This is because the attacks seemed to source from only a few specific IP addresses which corresponded back to the site being injected into the victim's database. Contrast this with the Asprox SQL injection attack which used a botnet to do the mass injection, giving attackers far more reach and bandwidth.

One of the main problems for organizations trying to address this issue is that attackers are not just exploiting vulnerabilities in the actual Web server software, such as IIS and Apache. It is not enough for Web server administrators to stay up to date on vendor patches. Attackers are also analyzing Web application packages (written in .ASP, PHP, etc.) running on the Web server in order to find SQL injection vulnerabilities they can exploit. In some cases, once a vulnerable web application has been identified, attackers use search engines to automate the process of finding target sites that use the vulnerable applications.

Unfortunately, attackers have many available options for remaining undetected. There are multiple ways of disguising and obfuscating their attack to prevent discovery. In order to prevent SQL injection, Web applications should be scanned or audited for places where user input is allowed to pass unfiltered to the database. This includes web forms, URL parameters, and cookie values. Where possible, parameterized SQL statements should be used so that the underlying database driver can escape the harmful characters. Also, web logs should be monitored for intrusion attempts such as hex-encoded strings or SQL keywords such as DECLARE, CAST, CONVERT, UNION, INSERT, or UPDATE. Organizations that suspect they have been a victim of the recent SQL injection attacks should automate the database cleanup.

In addition to web application auditing, administrators should review their remote access policies and verify that reusable passwords are prohibited in favor of strong authentication mechanisms such as SSH "authorized_keys." Remote access to administrative accounts should be disabled entirely with the possible exception of tightly controlled applications and keys. To help minimize the risk of becoming infected when visiting compromised sites, client systems should ensure that they have applied the latest security patches for browsers and plugins (Flash, Realplayer™, etc.). Additionally, "ghost" accounts (expired accounts or accounts where individual owners are no longer present), should be removed.

Public defacement, confidential data leakage, and database server compromise can result from these attacks. Complete compromise of vulnerable client systems is also possible. It is imperative that organizations treat SQL injection vulnerabilities as a serious threat and address them accordingly.

MSS—2011 top high-volume signatures

Top high-volume signatures

Table 1, shows the placement of the top Managed Security Services high volume signatures and their trend line for mid-year 2011 as compared to year end 2010. What is interesting is that the top two signatures have flip-flopped placement on our chart. SQL_injection is now number one and seeing an upward trend and SQL_SSRP_Slammer_Worm is in second place and continues to observe a downward trend. Six of the top ten signatures from 2010 have claimed a spot on the 2011 mid-year list. Also, with the exception of SMB_Empty_Password_Failed, the top half is made up of the same signatures; their placement has simply changed.

Event Name	Trend	Mid-Year 2011 Rank	Year End 2010 Rank
SQL_injection	Up	1	2
SQL_SSRP_Slammer_Worm	**Down**	**2**	1
SMB_Empty_Password_Failed	Slightly Up	3	
SSH_Brute_Force	Up	4	4
HTTP_Unix_Passwords	Up	5	6
PsExec_Service_Accessed	**Slightly Down**	**6**	3
HTTP_DotDot	Up	7	
Shell_Command_Injection	Slightly Up	8	
MSRPC_RemoteActivate_Bo	Up	9	
SMB_Mass_Login	Up	10	7

Table 1: Top MSS high volume signatures and trend line – Mid-year 2011 vs Year End 2010

SQL injection—upward trend and higher volume

SQL injection continues to be a favorite attack vector amongst attackers. Our heuristic SQL signature has climbed to first place. Contributing to this higher volume of activity are the mass SQL injection campaigns that continue to plague users, as discussed in the **SQL injection** section. The overall trend since 2010 through mid-year 2011 has been upward.

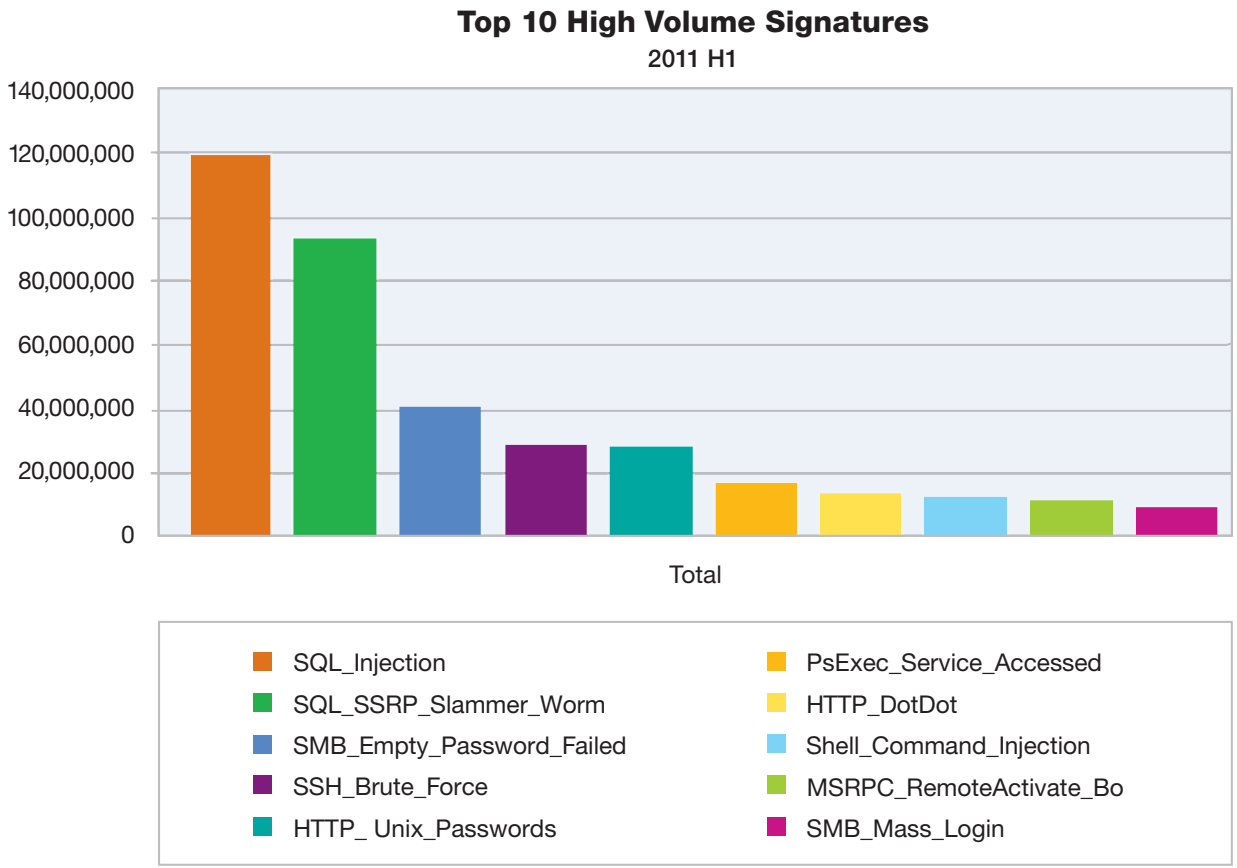


Figure 5: Top 10 High Volume Signatures – 2011 H1

SQL Slammer—no longer top dog

The top high-volume signature in 2010 was SQL_SSRP_Slammer_Worm. This signature has fallen to second place in our mid-year 2011 table. SQL Slammer targets a buffer overflow vulnerability in the Resolution Service in Microsoft SQL Server 2000 or Microsoft Desktop Engine (MSDE) 2000 installations. **"The day that SQL Slammer disappeared"** section, highlights a dramatic fall in SQL Slammer activity in March which contributed to the lower placement of this signature on our list for mid-year 2011.

Targeting Server Message Block (SMB) servers

Two of the top decodes once again (two) identify attacks against threats targeting server message block (SMB) servers—SMB_Empty_Password_Failed (third place) and SMB_Mass_Login (tenth place). While SMB_Mass_Login was in our year-end 2010 list (seventh place), SMB_Empty_Password_Failed is new to the list.

SMB_Empty_Password_Failed detects when an unsuccessful login attempt with no password is made to an SMB server. If attackers are attempting to connect to SMB servers with no password, this signifies that this method of attack continues to be fruitful for attackers. The SMB_Mass_Login signature detects an excessive number of granted NETBIOS sessions originating from the same IP address. This may indicate a stolen account being used in a scripted attack.

The existence of these signatures in the list highlights a possible lack of basic security with SMB shares. Recent threats, such as the Conficker and Stuxnet malware, use SMB shares to spread across networks.

Brute force attacks and scans

SSH_Brute_Force is another interesting signature in this list, holding onto fourth place. A brute force attack involves an attacker trying to gain unauthorized access to a system by trying a large number of password possibilities. This signature detects an excessive number of SSH Server Identifications from an SSH server within a specified time frame. Through this type of attack, a malicious individual may be able to view, copy, or delete important files on the accessed server or execute malicious code.

The drop in placement of this signature on the list may indicate a move towards mitigating brute-force attacks by disabling direct access to root accounts and using strong usernames and passwords. On the other hand, this activity is trending upward, and we may see this signature climb its way back up the list by year-end 2011.

Do not forget about UNIX . . . Several of the aforementioned signatures detect attacks utilizing Microsoft vulnerabilities; however, UNIX systems are not immune to threats. The signature HTTP_Unix_Passwords remains in the top high-volume list, but climbs up one spot from sixth to fifth place. This signature detects attempts to access the /etc/passwd file on UNIX systems via a Web (HTTP) server. While this activity is sometimes authorized, it can sometimes be suspicious. This is a very old attack, but is still successful today.

PsExec—a remote administration tool

Third on our list for 2010, PsExec_Service_Accessed, has dropped to sixth place for the mid-year 2011. PsExec is a command-line based remote administration tool and is used for legitimate purposes. However, worms and advanced threats also take advantage of PsExec. The "Here you have" worm, for instance, includes a PsExec tool that allows it to copy itself onto other computers over the network. If this application is used in your organization, you should ensure that best security practices are employed.

Traversing directories

The HTTP_DotDot signature detects an attacker's attempt to bypass the normal security imposed by the Web server and in order to access normally restricted files. An attacker can traverse directories on vulnerable Web servers by using "dot dot" (/../) sequences in URLs, allowing the attacker to read any file on the target HTTP server that is world-readable or readable by the ID of the HTTP process. For example, a URL of the form (http://www.domain.com/..\..) allows anyone to browse and download files outside of the Web server content root directory. URLs such as (http://www.domain.com/scripts..\..\) script-name could allow an attacker to execute the target script. An attacker can use a listing of this directory as additional information for planning a structured attack, or could download files elsewhere in the file system.

Shell commands

Our eighth signature on the list, Shell_Command_Injection, comes as no surprise. This signature detects a Shell Command injection attempt by scoring various combinations of commands and symbols used when executing shell commands. We have seen an increase of a very rudimentary attack utilizing shell commands that are injected into unsanitized inputs. Why would an attacker do this? Because it works! Rather than attempt to use SQL injection the attacker runs his code command via the web.

Targeting Microsoft

The signature MSRPC_RemoteActivate_Bo looks for a specially-crafted MSRPC Remote activation request that is used to conduct a buffer overflow. Microsoft Windows is vulnerable to a buffer overflow in the Distributed Component Object Model (DCOM) interface of the RPC (Remote Procedure Call) service. By sending a malformed message to the RPC service, a remote attacker can overflow a buffer and execute arbitrary code on the system with Local System privileges.

The day that SQL Slammer disappeared

Origin of the SQL Slammer worm

On January 25, 2003, an aggressive worm exploiting a buffer overflow in the Microsoft Resolution Service began a mass infection of Internet-connected servers. While the worm did not use a SQL vulnerability to propagate, the vast majority of infections occurred on servers running the Microsoft SQL Server Desktop Engine (MSDE). The worm exists only to propagate itself and immediately seeks to infect as many machines as possible by attacking random IP addresses. The worm is very small at only 376 bytes, so it is able to send a copy of itself to a vulnerable machine in a single UDP packet.

Although Microsoft released a patch for the vulnerability six months before the first appearance of Slammer, there were enough exploitable servers for the growth to become exponential. According to analysis done by the **Cooperative Association for Internet Data Analysis (CAIDA)**, 90 percent of all vulnerable systems were infected within the first 10 minutes of the worm's release. Since the worm targeted random IP addresses, even networks with no vulnerable servers were brought to their knees by the sheer volume of the infection attempts. Because SQL servers were the main victims of the worm and infected boxes were generating enough attacks to overwhelm Internet infrastructure, it was dubbed the SQL Slammer worm by then Internet Security Systems CTO Chris Rouland.

The worm's speed of propagation was due in large part to its small size, but the size limitation meant that it had no facility for becoming persistent. Removing the worm was as easy as rebooting the affected server, but it would quickly become reinfected if the proper patch had not been applied.

Given its effects and frightening propagation speed, it became a large story even outside of the security and IT communities. Microsoft launched an education campaign to get server administrators to apply the patch and the main media outlets devoted an unusually high amount of coverage for a malware infection. This raised awareness had the desired effect; the pool of vulnerable servers rapidly declined. Combined with filtering by security devices and Access Control List (ACL) changes to routers, the traffic eased up enough to stop the Internet-wide denial of service.

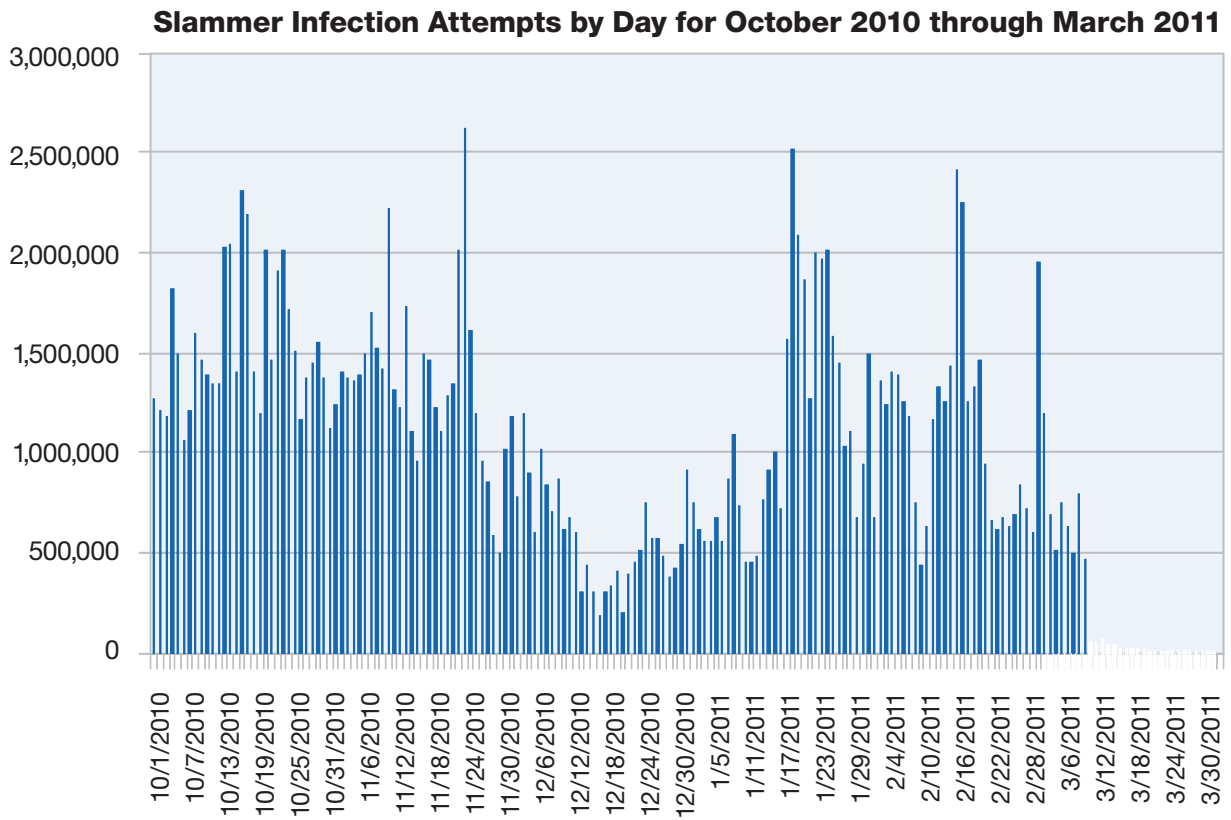


Figure 6: Slammer Infection Attempts by Day for October 2010 through March 2011

Slammer, however, did not go away. As late as the beginning of March 2011, Slammer infection packets still accounted for a sizable portion of UDP traffic on the Internet. For IBM customers, this translated into infection attempts measured in the hundreds of thousands per day. That all changed March 10th 2011 through March 11th 2011. Within a 24-hour time period, that rate dropped to below 2,000 a day, as shown in Figure 6. We first reported on this topic in the **April 2011 Frequency-X blog**.

Analysis of the drop in activity

The daily numbers alone show that this disappearance was coordinated and likely triggered by a single cause. This dramatic drop-off in such a short amount of time could not be a naturally occurring phenomenon, nor could the shutdown of a few servers account for such a steep decline of infection attempts. The IBM Managed Security Services team set out to discover the mechanism by which Slammer was almost universally disabled. The first step was to determine what the affected time window looked like hour by hour.

Rather than an abrupt drop occurring over two to three hours (signaling the use of a command and control style kill switch,) we see a phased draw down occurring over a 20-hour period (Figure 7). This data pointed to a clock-based shut off issued by the same coded trigger across all of the affected servers. To test for this, we used geographic location on the source IP addresses for these attacks. Geo location works by using the whois records for registered addresses to identify the geographic location of those addresses. The system is not perfect—it cannot tell you the location of private address space (10.x.x.x, 172.16.x.x, 192.168.x.x), not all real addresses have sufficient registration to identify physical locations, and you cannot be sure that the server has the correct timezone information or time set. But even with these caveats, our data turned up some interesting results.

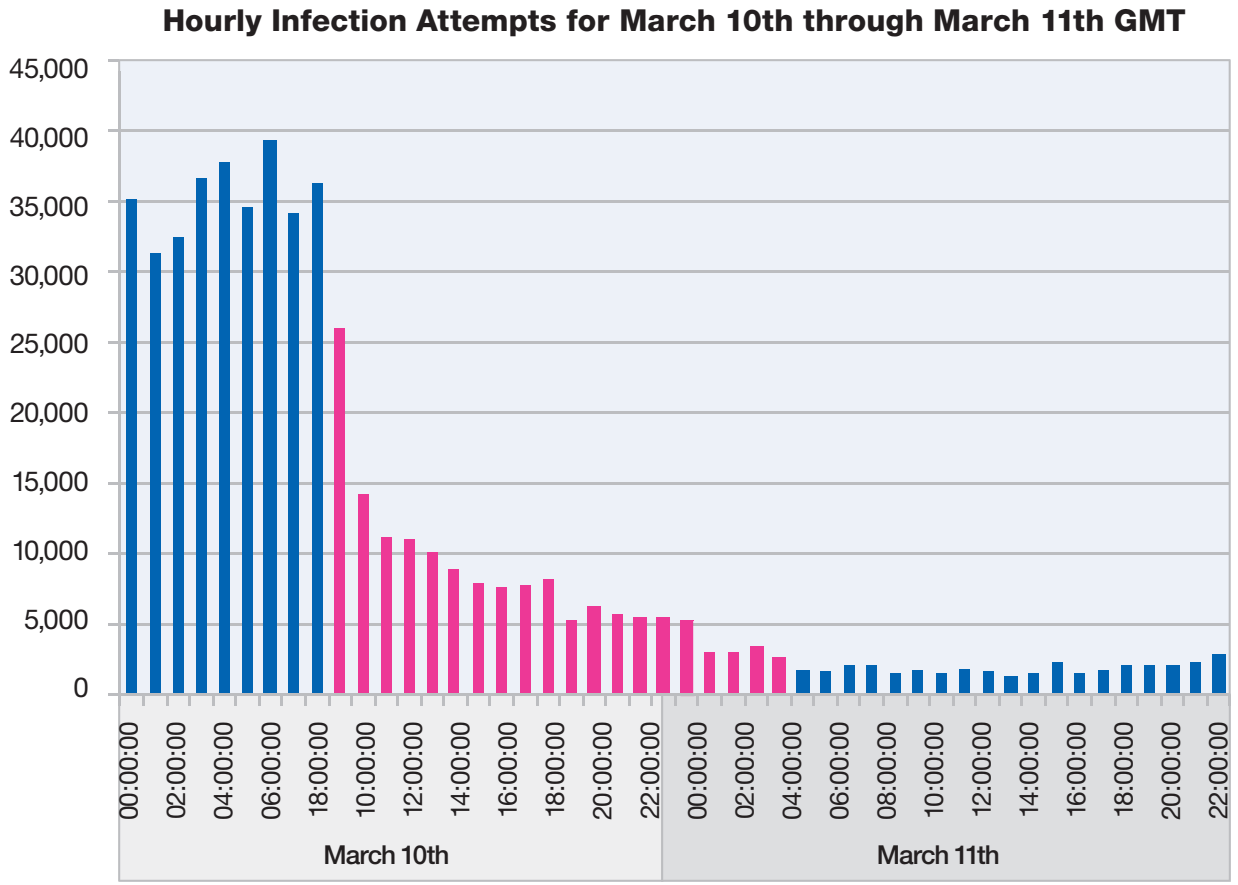


Figure 7: Hourly Infection Attempts for March 10th through March 11th GMT

We took all the source IP addresses for which we could get good geo-location data and used their probable offsets to look at the infection attempts based on the attacker's estimated local time. What we see is an almost 50 percent drop between 11:00 AM and 12:00 PM (attacker's local time) on March 10th 2011 (Figure 8). We also see a sustained drop after 12:00 PM for these IP addresses. While we do not see a complete drop off in a one-hour period, the fact that the adjusted data falls off quicker than the non-adjusted data is significant. The difference between a gradual draw-down over 20 hours as opposed to a sharp curve occurring in eight hours is a pretty compelling argument for a trigger set to occur between 11:00 AM and 12:00 PM based on the Slammer server's clock. If there were no time-based trigger, the curve should have been broken rather than clarified by the time-based adjustments. We would expect to see a more erratic graph that no longer created a discernible pattern.

Conclusion

While geo-location data is not a perfect methodology, the correlation between estimated attacker local time and the drop off of events is strong enough to point to an automated shut down triggered by an individual server's local clock. While this data supports a very plausible answer as to "how", it does not really answer "why." There are two likely alternative theories as to why. The first, is that the attack packets generated by

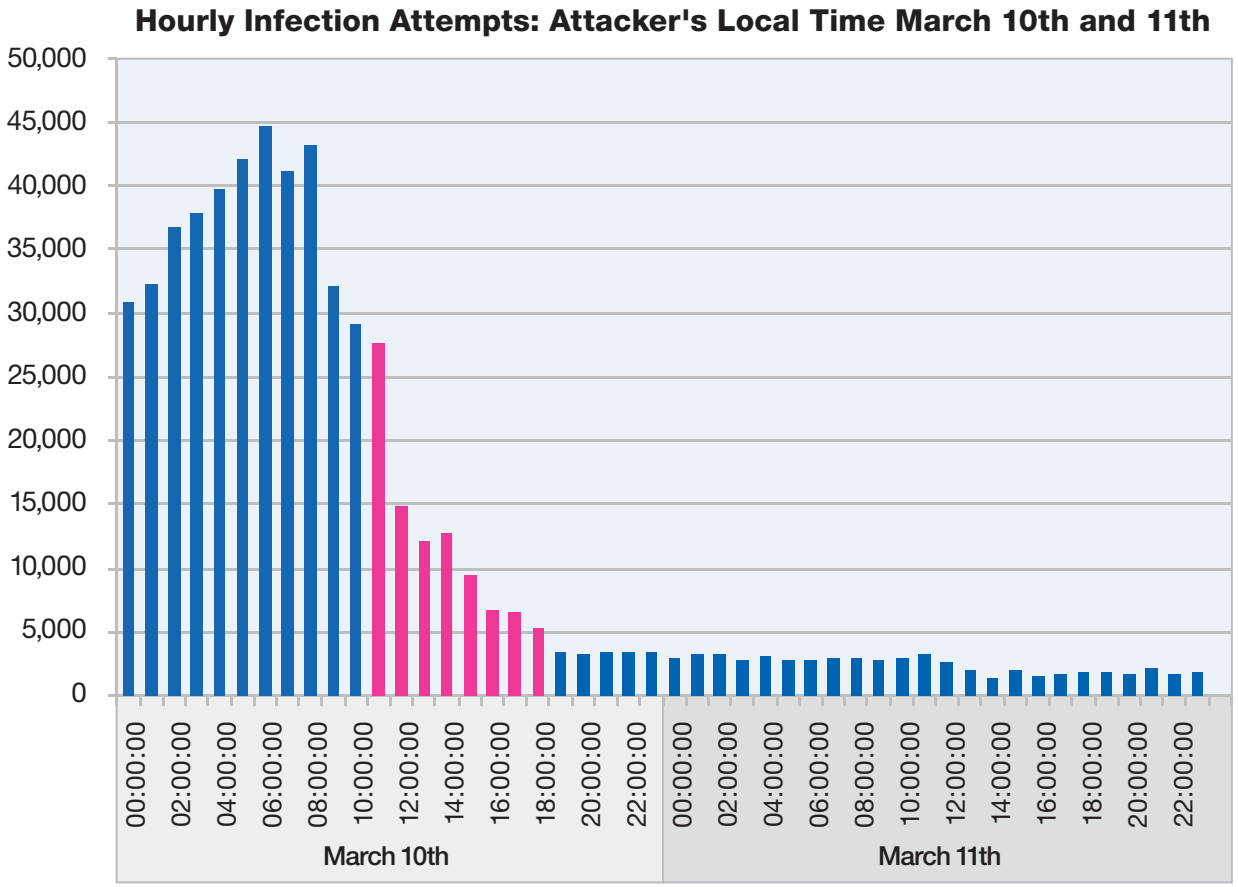


Figure 8: Hourly Infection Attempts: Attacker's Local Time March 10th and 11th

the infected servers gave a map to easily exploitable machines. An attacker compromised all of those servers manually or through an automated process and installed botnet code on them. Once the attacker had achieved control of the servers, the attacker set an automated patch and reboot so that they would stop giving themselves away. The second is that a White Knight security professional decided to rid the world of Slammer and did the exact same thing (hopefully without the persistent botnet code.) Those theories ascribe two very different motives, but the implementation is basically the same.

The most interesting thing about this operation is the use of the time-based trigger. If the servers were patched as they were compromised, we should have seen either a more gradual draw-down over a much larger time period or a steep draw-down that happened within one to two hours of the start of the patching. The fact that they used a time-based delay points to one of three possible motivations. If it was a malicious botnet, the attacker may have wanted enough time for fingerprints to leave the router and firewall logs so that the attacker could not be traced. If it was a White Knight, he or she may have wanted to use a trigger to know how effective the methodology had been. The third possibility is that, whatever other motives the attackers had for doing this, the person responsible wanted to get the security community's attention and, if that was the motivation, than the goal was certainly accomplished.

We have seen infection attempts rise and fall since March, but they have come nowhere near the volume we saw before March 10th (Figure 9.) Given that the vulnerability has been patched for nine years, it seems unlikely that these were new systems that were infected naturally. The new events form a wave pattern and began appearing in April 2011, soon after the slammer disappearance had gained some media traction. It is possible that many of these were intentional infections by members of the security community trying to answer "why."

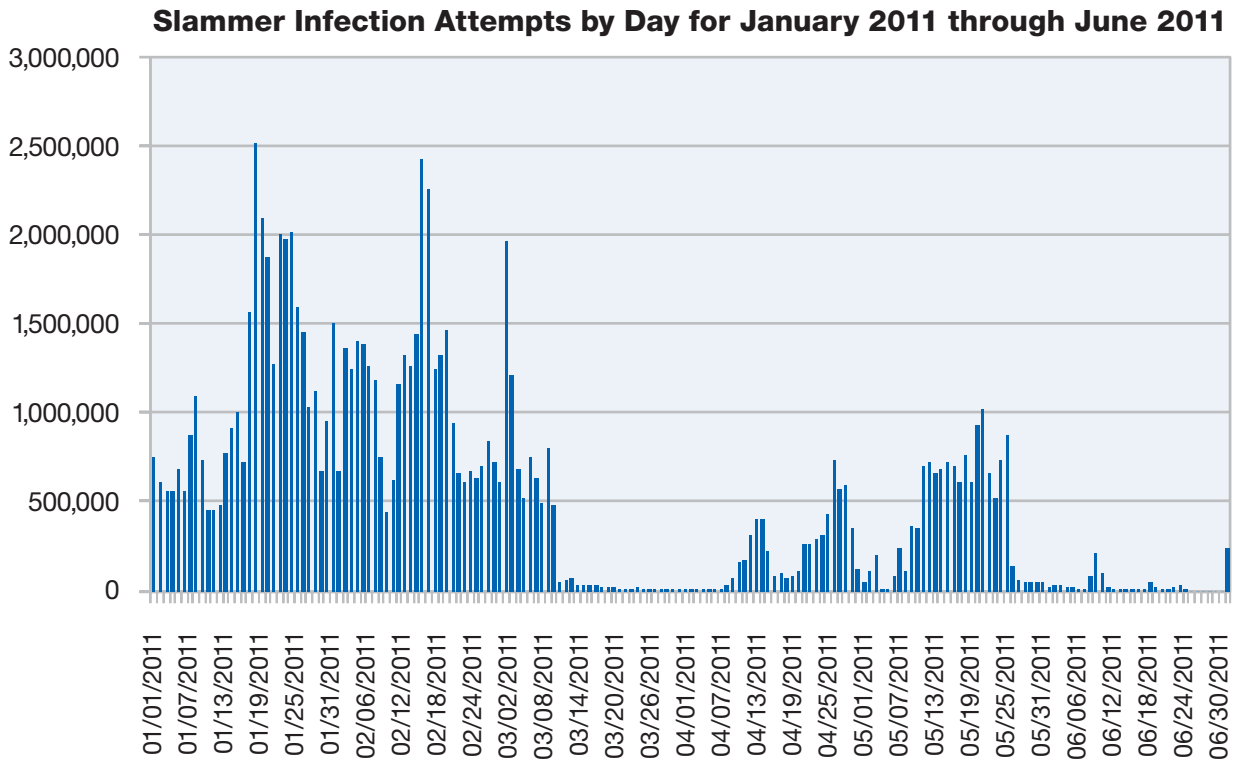


Figure 9: Slammer Infection Attempts by Day for January 2011 through June 2011

Web Content Trends, Spam and Phishing

Web content trends

The IBM Content data center constantly reviews and analyzes new web content data and analyzes 150 million new web pages and images each month and has analyzed 15 billion web pages and images since 1999.

The IBM web filter database has 68 filter categories and 69 million entries with 150,000 new or updated entries added each day.

This section provides analysis for:

- Analysis methodology
- Internationalized top-level domains
- Increase in the amount of anonymous proxies
- Top-level domains of anonymous proxies
- Malicious websites

Analysis methodology

X-Force captures information about the distribution of content on the Internet by counting the hosts categorized in the IBM Security Solutions web filter database. Counting hosts is an accepted method for determining content distribution and provides a realistic assessment. When using other methodologies—such as counting web pages and subpages—results may differ.

Internationalized top-level domains

Since the beginning of 2010 it is possible to register internationalized country code top-level domains[9]. Therefore URLs can be displayed without using any ASCII letters. The first domains were registered in the Arabic and Cyrillic alphabet. However, the usage amounts on the Internet differ widely for different languages. While there are only a few Arabic websites using these new domains, there is a significant increase of these domains in Russia.

In April, 2011, nearly five percent of the newly online Russian domains were .рф[10] domains. This seemed to be some kind of spring-time promotion as in May and June it decreased significantly to one percent.

In the **Trend reversal of spam volume** section we provide more information about the usage of this new top-level domain.

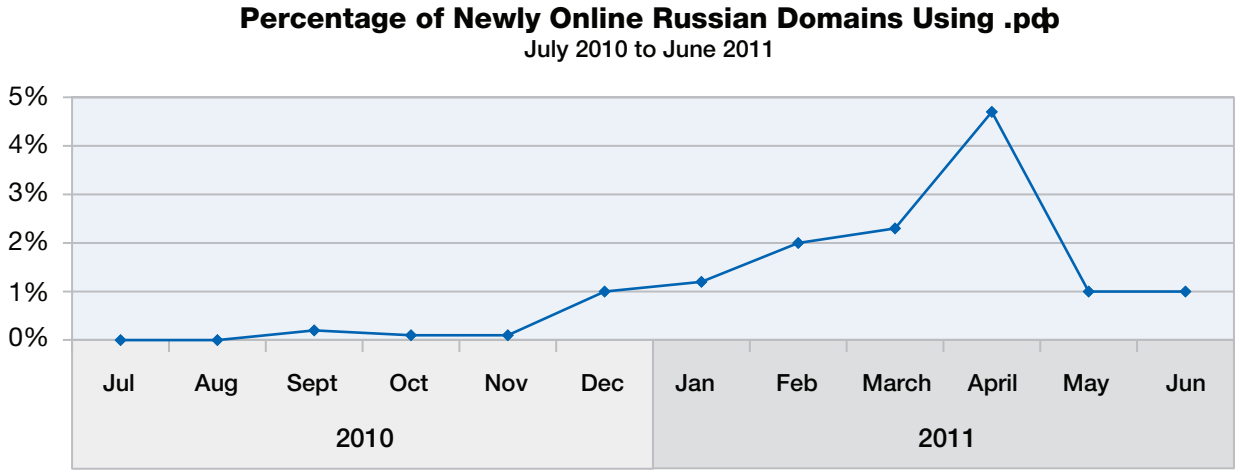


Figure 10: Percentage of Newly Online Russian Domains Using .рф – July 2010 to June 2011

9 See also: http://en.wikipedia.org/wiki/Internationalized_country_code_top-level_domain
http://en.wikipedia.org/wiki/List_of_Internet_top-level_domains
http://en.wikipedia.org/wiki/Internationalized_domain_name

10 "рф" are the letters rf in the Cyrillic language and mean "Russian Federation".

Increase in the amount of anonymous proxies

As the Internet becomes a more integrated part of our lives, not only at home, but also at work and at school, organizations responsible for maintaining acceptable environments increasingly find the need to control where people can browse in public settings.

One such control is a content filtering system that helps prevent access to unacceptable or inappropriate websites. Some individuals attempt to use anonymous proxies (also known as web proxies) to circumvent web filtering technologies.

Web proxies allow users to enter an URL on a web form instead of directly visiting the target website. Using the proxy hides the target URL from a web filter. If the web filter is not set up to monitor or block anonymous proxies, then this activity (which normally would have been stopped) bypasses the filter and allows the user to reach the disallowed website.

The growth in newly registered anonymous proxy websites reflects this trend.

In the first half of 2011, there were about four times as many anonymous proxies registered as there were three years ago. Anonymous proxies are a critical type of website to track because of the ease that proxies provide in allowing people to hide potentially malicious intent.

We have chosen in this report to make a slight methodology change from previous reports. In the past, we have always counted the total number of anonymous proxy websites. The disadvantage to this type of calculation is that it does not demonstrate the dynamic nature of this activity.

Many anonymous proxies go offline and the same number of proxies might come online with no change shown in previous charting method. Therefore, we decided to choose a new way of presenting the data by counting the newly registered anonymous proxies from term to term.

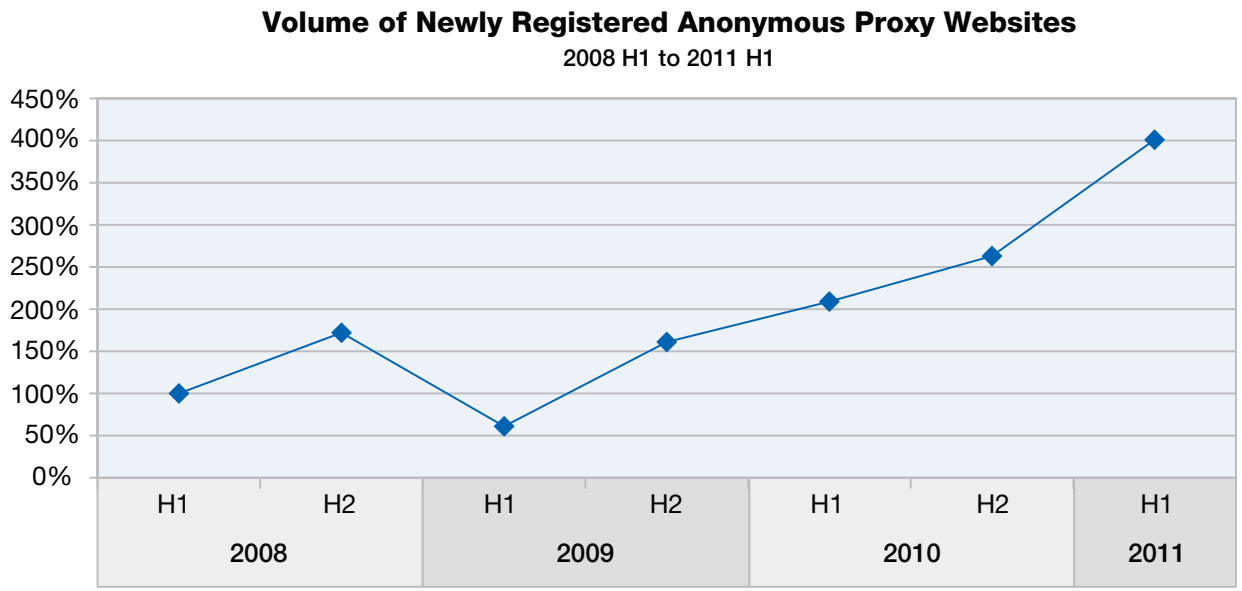


Figure 11: Volume of Newly Registered Anonymous Proxy websites – 2008 H1 to 2011 H1

Top-level domains of anonymous proxies

Figure 12 illustrates the top level domains (TLDs) of the newly registered anonymous proxies.

In 2006, more than 60 percent of all newly registered anonymous proxies were .com domains, but since the middle of 2007, .info has been at the top until the beginning of 2010 (while .com was runner-up for most of the time).

But why is .info no longer in the prime position? It seemed to be a proven top-level domain (TLD) for anonymous proxies for years. A reason could be that .info, similar to .com, is running out of names. So the question arises why are anonymous proxies now provided on .cc and .tk top level domains. These are the domains of Cocos (Keeling) Islands (.cc), an Australian territory, and Tokelau (.tk), a territory of New Zealand. The domain .cc is administered by VeriSign. Nearly all .cc anonymous proxies websites are registered on the domain co.cc. There is no charge to register a domain anything.co.cc[11]. The same is true for .tk[12]. Thus, it is both cheap and attractive to install new anonymous proxies on .co.cc or .tk.

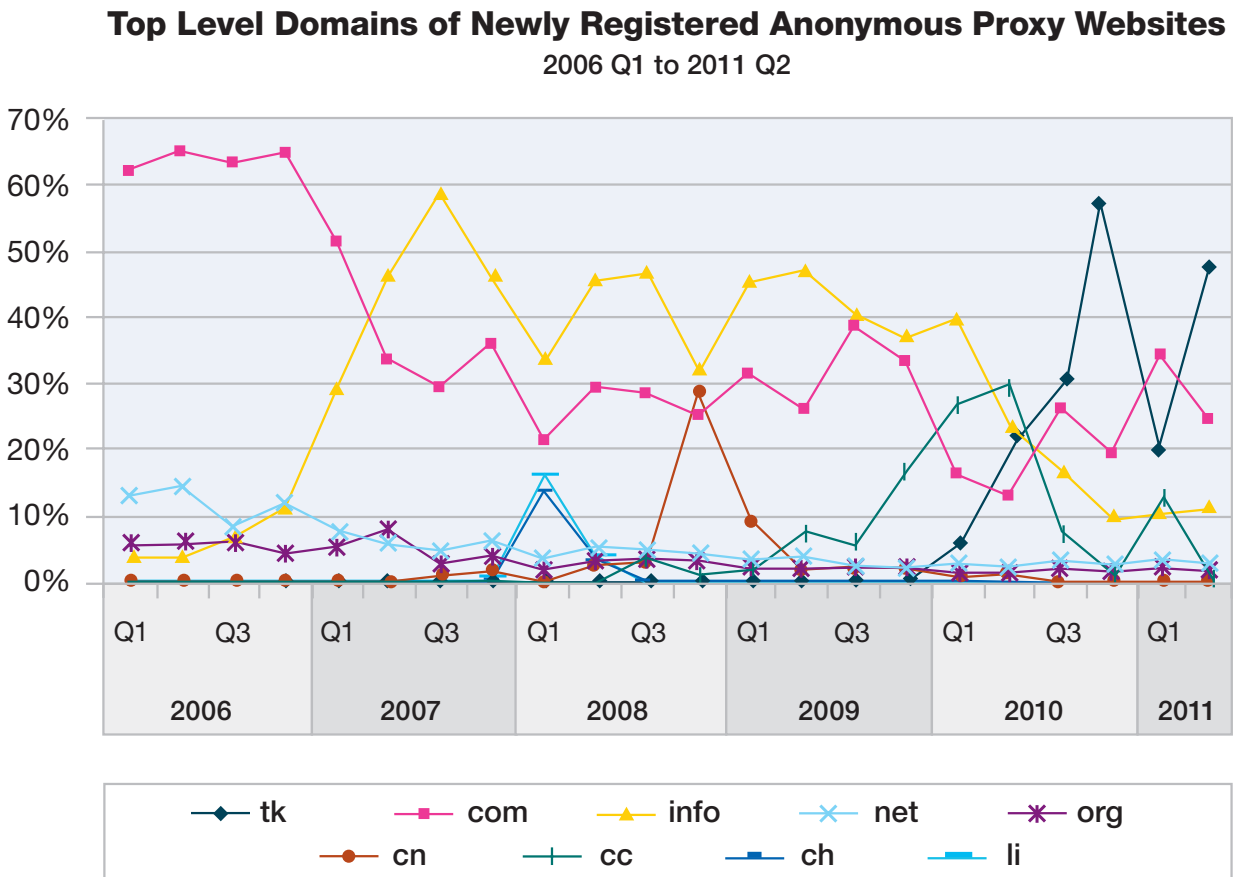


Figure 12: Top Level Domains of Newly Registered Anonymous Proxy websites – 2006 Q1 to 2011 Q2

11 http://www.co.cc/?lang=en
12 http://www.dot.tk/

Additional trends:

- At the beginning of 2008, the top-level domains of neighboring countries Switzerland (.ch) and Liechtenstein (.li) together represented about 30 percent of the newly registered anonymous proxies.
- In the fourth quarter of 2008, the top-level domain of China (.cn) reached nearly 30 percent of the newly registered anonymous proxies.
- At the end of 2009 .cc (Cocos (Keeling) Islands) started to increase significantly and even reached the number one position in the second quarter of 2010. Nevertheless .cc went out of vogue at the end of 2010 nearly completely, then had a short comeback in the first quarter of 2011.
- In the second quarter of 2010, another new star in proxy heaven, .tk (Tokelau), reached about 23 percent of new anonymous proxies. It dominated the rest of the year with nearly 30 percent in the third quarter and more than 56 percent in the fourth quarter of 2010. In the second quarter of 2011 it is frontrunner again with 46.5 percent.
- During that same time period, .info decreased dramatically and fell below 10 percent for the first time at the end of 2010, recovering only slightly to 12 percent in the second quarter of 2011.
- In the first quarter of 2010, even .com fell significantly below 20 percent for the first time, recovering to 26 percent in the third quarter and 19 percent in the fourth quarter of 2010. In the first quarter of 2011, it topped the list for the first time in four years and ended in second place in the second quarter of 2011.
- When looking at the last 12 months, .tk and .com are clearly dominating the scene.

It will be interesting to see whether .tk has a similar destiny as .co.cc, being the star of anonymous proxies for one and a half years and then sinking into obscurity.

Concerning .co.cc there was another interesting action: At the beginning of July, 2011, Google announced that they would remove .co.cc sites from its search index[13] in order to modify its malware detection system to identify subdomain services which allow attackers to register thousands of domains. One would think that such a sanction could help stem anonymous proxies. Unfortunately, new anonymous proxies are published in many other ways including mailing lists and Twitter feeds. Therefore, they do not need to be findable via search engines. Even if there were harder actions against some domains or top-level domains—perhaps comparable to the McColo or Rustock take downs (see the **Trend reversal of spam volume section**)—it would help only temporarily. It seems likely that there will always be some loose registrar out there who provides open doors for anonymous proxies because domain registration is an issue that each country handles differently.

13 See http://www.h-online.com/security/news/item/No-more-Googling-for-co-cc-domains-1274332.html.

Malicious websites

This section discusses the countries responsible for hosting malicious links along with the types of websites that most often link back to these malicious websites. More information on malicious websites in the exploit context can also be found in the section Exploit effort versus potential reward matrix.

Geographical location of malicious Web links

The United States continues to reign as the top host for malicious links. More than one third of all malware links are hosted in the US. New second place is now Romania, hosting 7.8 percent. China was on top two and a half years ago, now it is in third place, claiming 7.2 percent. This is 1.4 percent more than France as shown in Figure 13.

The second-tier countries have also shifted, but these shifts are below one percent.

Good websites with bad links

As described later in this report, page 58, we discuss how the total number of vulnerabilities are declining. Regardless, attackers still focus on using the good name of trusted websites to lower the guard of end users and attempt to obfuscate their attempts from protection technologies. The use of malicious web content is no different. The following analysis provides a glimpse into the types of websites that most frequently contain links to known, malicious websites.

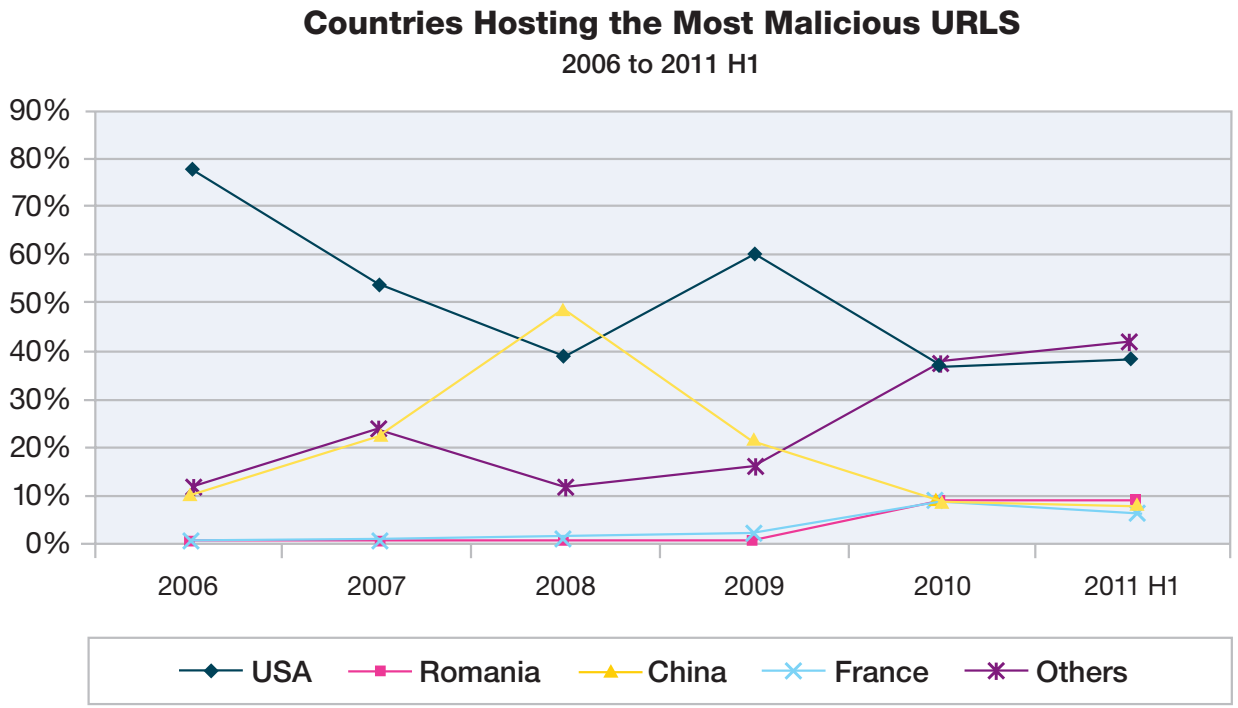


Figure 13: Countries Hosting the Most Malicious URLS – 2006 to 2011 H1

Some of the top categories might not be surprising. For example, one might expect pornography and gambling to top the list. Together they host more than 40 percent of all malicious links. However, the second-tier candidates fall into the more "trusted" category.

Blogs, bulletin boards, personal websites, search engines, education, shopping sites, online magazines, and news sites fall into this second-tier category. Most of these websites allow users to upload content or design their own website, such as personal content on a university website or comments about a purchase on a shopping website. In other words, it is unlikely that these types of websites are intentionally hosting malicious links. The distribution is probably more representative of the types of websites that attackers like to frequent in hopes of finding a loop hole (like a vulnerability or an area that allows user-supplied content) in which they can incorporate these malicious links in hopes of compromising an unsuspecting victim.

Figure 14 lists the most common types of websites that host at least one link that points back to a known malicious website.

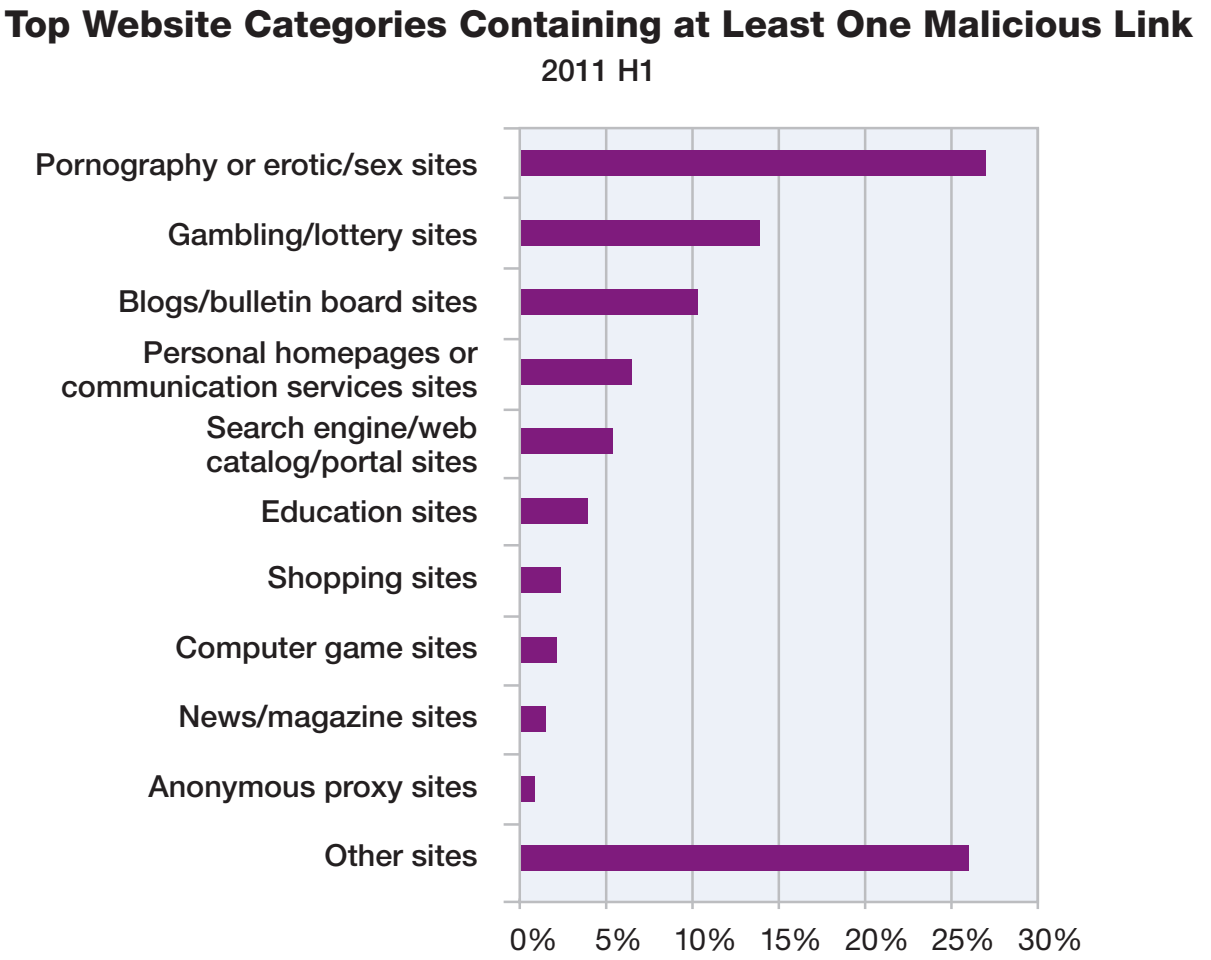


Figure 14: Top website Categories Containing at Least One Malicious Link – 2011 H1

Figure 15 shows the history of the top players.

When looking back the last two and a half years, some interesting trends appear.

- The professional "bad" websites like pornography and gambling now dominate the scene to systematically distribute malware
- Pornography is at the top and has stabilized at about 25 percent
- Gambling is the only category with a significant year-over-year increase. Against the background of 0.6 percent of the adult population having problem gambling issues[14], gambling sites are a popular target for malware distributors.
- Blogs and bulletin boards are at roughly the same level from a year ago at about 10 percent
- Personal homepages and search engines, web catalogs, and portal sites—the classical Web 1.0 websites—significantly lost ground. One reason may be that personal homepages are more out of style in favor of Web 2.0 applications like profiles in social or business networks.

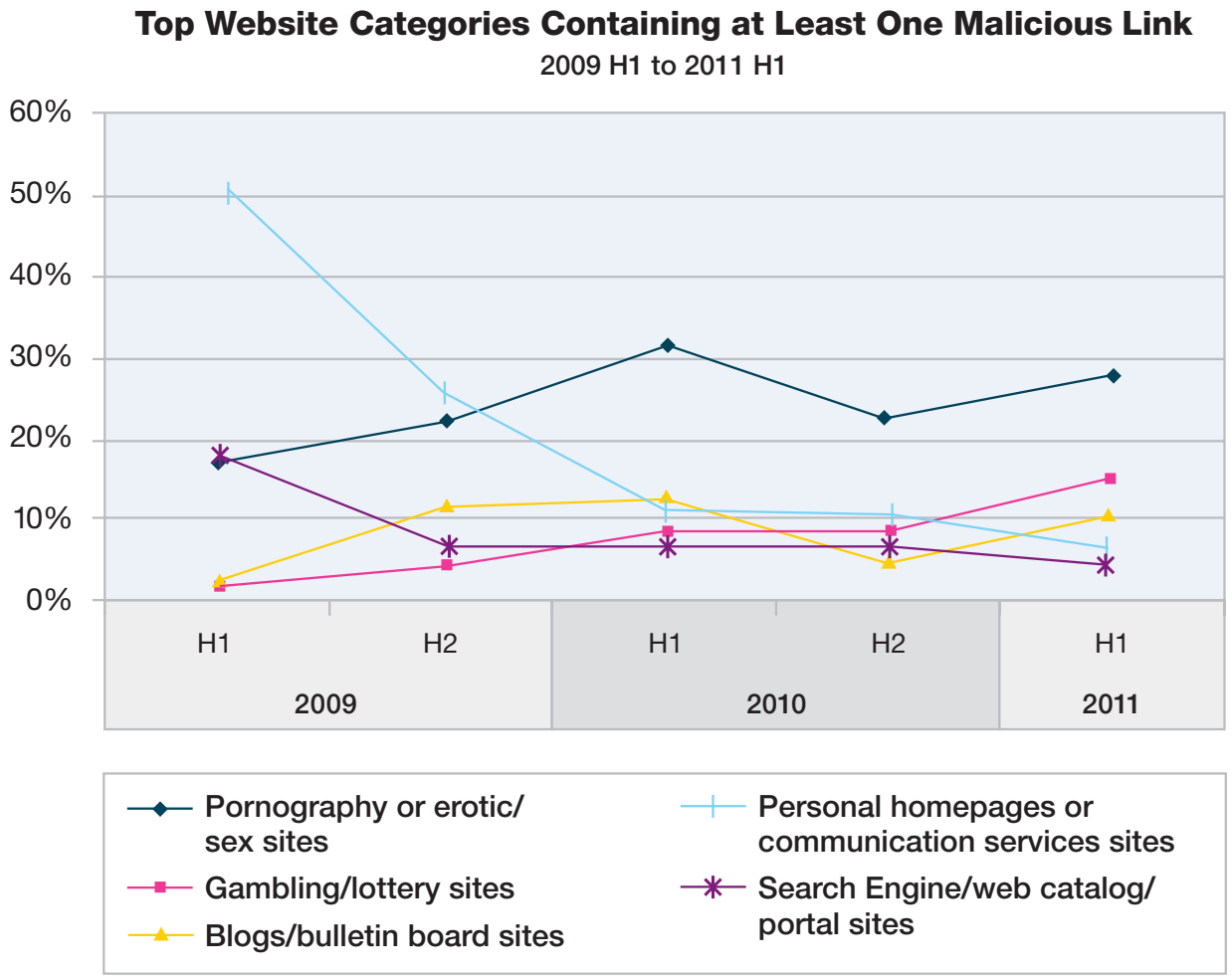


Figure 15: Top website Categories Containing at Least One Malicious Link – 2009 H1 to 2011 H1

14 See http://en.wikipedia.org/wiki/Gambling_addiction#Prevalence

Trend reversal of spam volume

The IBM spam and URL filter database provides a global view of spam and phishing attacks. With millions of email addresses being actively monitored, the content team has identified numerous advances in the spam and phishing technologies that attackers use.

Currently, the spam filter database contains more than 40 million relevant spam signatures. Each piece of spam is broken into several logical parts (sentences, paragraphs, etc.). A unique 128-bit signature is computed for each part and for millions of spam URLs. Each day there are approximately one million new, updated, or deleted signatures for the spam filter database.

This section addresses the following topics:

- Spam volume and botnet take downs
- Common top-level domains in URL spam
- Spam—country[15] of origin trends
- Email phishing
- Future prospects on spam

Spam volume and botnet take downs

After years of significant growth until the middle of 2010 with only one major fallback at the end of 2008, we have seen a decline in spam volumes within the last 12 months.

One interesting story discussed in spam circles since the end of December 2010 is the "lull in activity" at the end of the year. In a January post of the Frequency-X blog we speculated on why these volumes suddenly dropped. Did the spammers go on holiday? Was the business drying up? Was this the first take down of the Rustock botnet? We had more questions than answers in the month of January but since then some interesting news has played out.

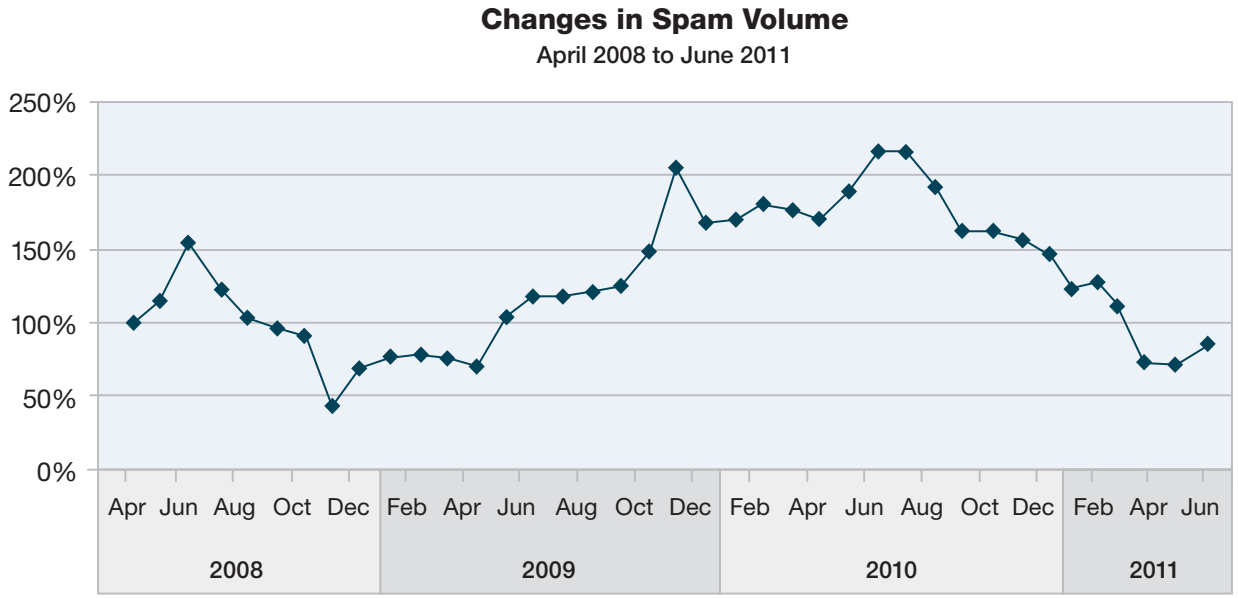


Figure 16: Changes in Spam Volume – April 2008 to June 2011

15 The statistics in this report for spam, phishing, and URLs use the IP-to-Country Database provided by WebHosting.Info (http://www.webhosting.info), available from http://ip-to-country.webhosting.info. The geographical distribution was determined by requesting the IP addresses of the hosts (in the case of the content distribution) or of the sending mail server (in the case of spam and phishing) to the IP-to-Country Database.

Let's look closer at the last six months that were dominated by two Rustock take downs in December 2010 and March 2011.

- **First Rustock take down** December 25, 2010 until January 9, 2011: In early January, several news agencies, including a **New York Times article** began reporting on the decline in spam as the first Rustock was taken down and also reported that key business in Russia was drying up.

- **Second Rustock take down** since March 16, 2011: By March it was clear when Microsoft and US Marshals were able to take down the command and control capabilities of this botnet, as reported on **Microsoft's blog site**. The drop in spam volumes was picked up by the IBM spam traps and can be seen in Figure 17.

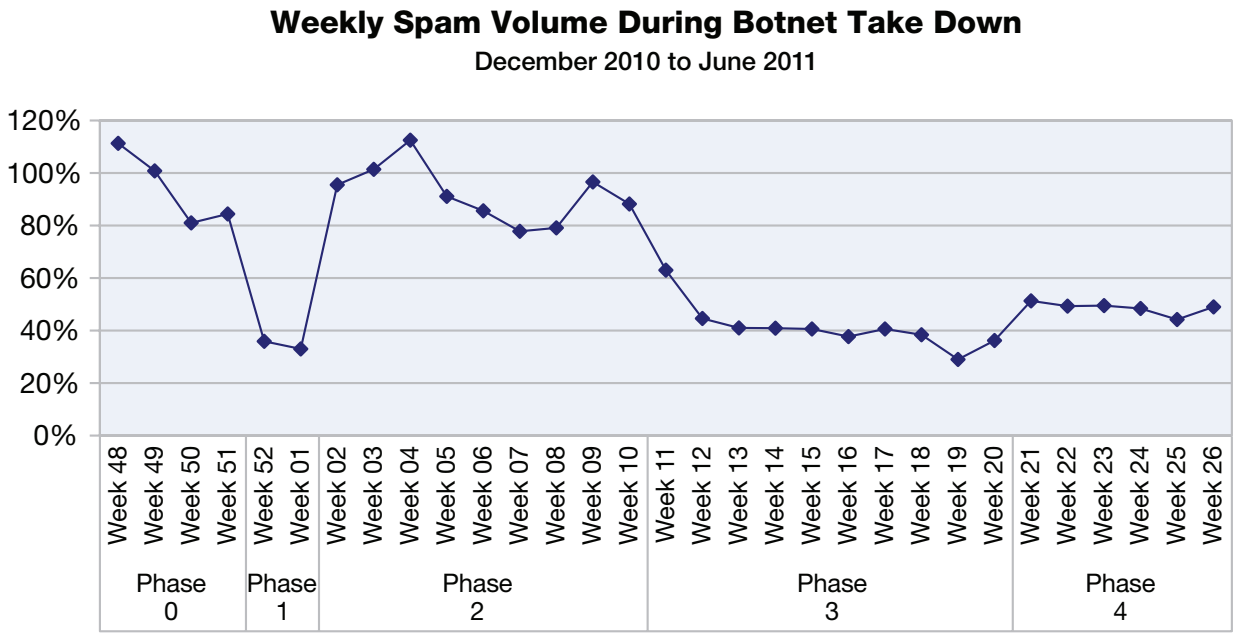


Figure 17: Weekly Spam Volume During Botnet Take Down – December 2010 to June 2011

Based upon this, one can derive several changes concerning other aspects of the spam sent out until mid-2011. To highlight these changes we have defined several phases:

- **Phase 0—Initial situation:**
  Beginning of December, 2010
- **Phase 1—First Rustock take down:**
  December 25, 2010 until January 9, 2011
- **Phase 2—Between the Rustock take downs:**
  January 10, 2011 until March 15, 2011
- **Phase 3—After the second Rustock take down:**
  March 16, 2011 until May 18, 2011
- **Phase 4—First recovery of spam volume:**
  Since May 19, 2011

We first look at some top players regarding the country of spam origins. India, Indonesia, and the US were the countries that have shown the most significant changes from phase to phase.

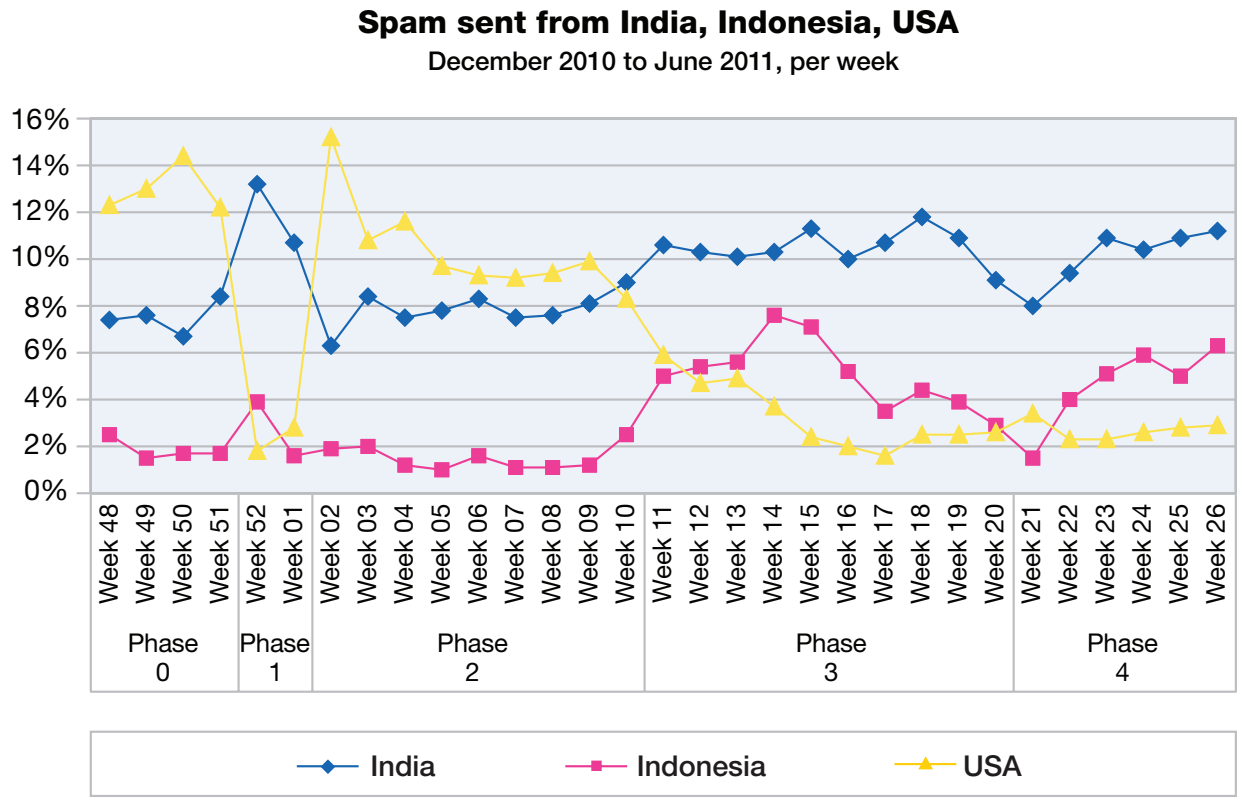


Figure 18: Spam sent from India, Indonesia, USA – December 2010 to June 2011, per week

When Rustock was taken down, the spam volume sent from the US was down, too, in most cases below three percent. At the same time, the percentage of spam sent from India increased to more than 10 percent and spam sent from Indonesia increased to four percent. When Rustock was active (phase zero and two), significantly more spam—more than nine percent in most cases—was sent from the United States while India went down to eight percent and Indonesia reduced to two percent. Thus, Rustock was widely in use on US computers but much less in India and Indonesia. In the fourth phase, when spam volume increased again, the levels of spam sent from US-based computers did not recover as they did before. It seems that new botnet clients are recruited more outside the United States than in earlier years. But why do botnet infections avoid US-based computers? Possible answer: It is much easier to infect computers in other countries because:

- Non-Windows 7 installations in other countries are more susceptible[16].
- The last two major take downs (McColo in November, 2008 and Rustock in March, 2011) were driven by US-based organizations or companies. Perhaps spammers are avoiding this area and focusing on the rest of the world.

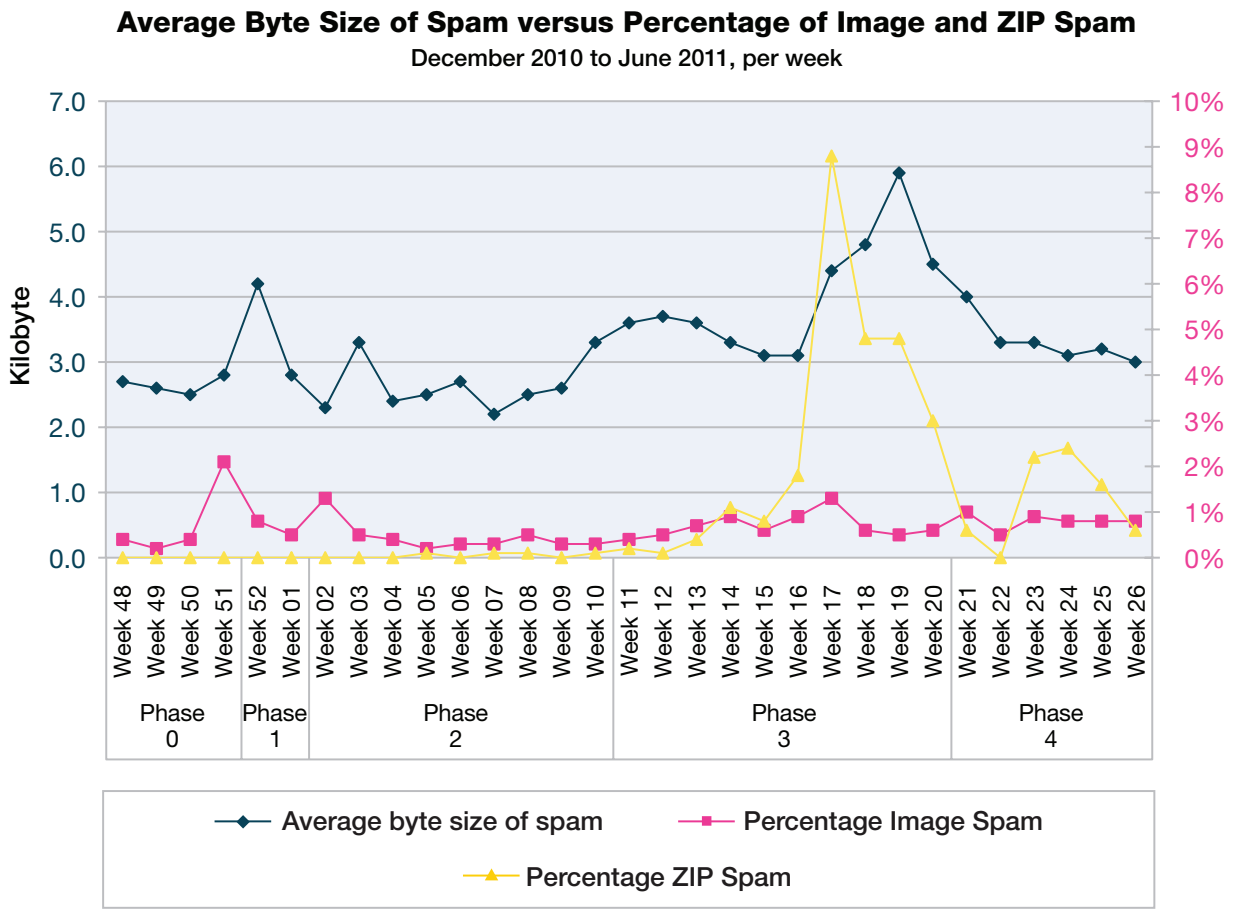


Figure 19: Average Byte Size of Spam versus Percentage of Image and ZIP Spam – December 2010 to June 2011, per week

16 According to StatCounter (http://gs.statcounter.com/) in June, 2011 about 35 percent of all computers in the US were running on Windows 7 but only 29 percent on Windows XP. In India we see only 28 percent Windows 7 but 64 percent Windows XP. This is even more significant in Indonesia where we have only 21 percent Windows 7 but 75 percent Windows XP.

The content characteristics of the spam are another way to analyze changes.

- **Average byte size of spam:** When spam volume is up and Rustock is active (phases zero and two) the average byte size of spam is down, mostly below three kilobytes. When the spam volume is down (phase one and three) the average byte size of spam is high, above three kilobytes in most cases. The spam sent by Rustock were small in size. When the ZIP attachment spam threats started in April, the spam size increased significantly as expected.
- **Image based spam:** As in previous years, this type of spam does not play a significant role. In most cases its volume is below one percent.
- **ZIP spam:** In the first quarter, spam with ZIP attachments were scarcely seen in the wild. But since mid-April, we have seen several threats accounting for two to eight percent of the spam volume (measured on a weekly basis).

When looking at the ZIP attachments of spam during early May of this year, more than 90 percent of that spam contained the **TrojanDownloader:Win32/Chepvil.K**. As a Trojan downloader, it downloads malware, rather than having intrinsic malicious capabilities. And, it may download not just one piece of malware but multiple malware applications with different intentions.

To convince users to open the ZIP attachment, some typical variants are used.

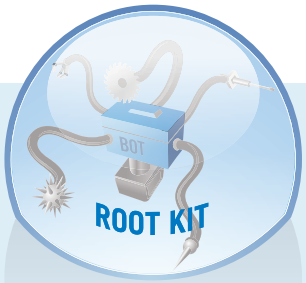
- Faked order confirmation including the message that the user's credit card will be charged for an amount over one-hundred USD and that the user can find the details in the attached file.
- An email stating that the user's IP address was logged on to several illegal websites. The "fake sender," the FBI, requests that the user answers the attached questions.

In another threat during May, emails contained the **TrojanDownloader:Win32/Ufraie.A**. In this case users were convinced to open the ZIP attachment by announcing that it contained a naked picture.

The Rustock take down paralyzed some major "sales channels" of the spammers. Figure 19 suggests that these ZIP attachment spam threats are an answer to that because, shortly after sending out these ZIP spam threats, the spam volume has begun to increase (phase four), perhaps with the involvement of new botnet clients infected some days before by the ZIP attachments. However, the levels are still about 50 percent below the levels of the fourth quarter of last year.

Shortly before publication of this report, we recognized a significant increase of the spam volume. The rise was initiated by higher levels of ZIP attachment spam. In mid-September 2011, the spam volume reached 80 percent of the levels it had reached nine months earlier.

We will be providing more detail on this new spam activity in the **Frequency-X blog**.



Are botnet take downs a continued trend in 2011?

In the IBM X-Force® 2010 Trend and Risk Report the Managed Security Services group reported on an upward trend in Trojan botnet activity during 2010. This growth was significant because, despite increasing coordinated efforts to shut down botnet activity (as seen with the Mariposa and Bredolab botnets), the threat appeared to be gaining momentum at the time.

In that same report, we also discussed the efforts by Microsoft to crack down on botnets. Specifically, Microsoft's "Operation b49" which took down the Waledac botnet in late February 2010.

In 2011, this trend of taking down botnet operators continues and we ask, "Is Security making advances on Botnet take downs?" Two more examples have come to light in early 2011.

March 16, 2011 – Rustock botnet

The Rustock botnet was one of the most problematic botnets within recent years. It was dedicated to sending out spam. A machine infected by Rustock sent out an average of more than 190 spam messages. Reportedly, there were between 150,000 and 2,400,000 computers infected with Rustock. On March 16, 2011, the botnet was taken down through a coordinated effort by vendors, researchers, and law enforcement. For more details see http://en.wikipedia.org/wiki/Rustock.

April 13, 2011 – Coreflood botnet

Since the early 2000's, the CoreFlood botnet has been used to compromise millions of systems in the United States and around the world. In an effort to take down the botnet, the FBI executed a first of its kind operation which involved setting up a custom command and control server which issued a stop command to the malware running on the infected PC's. Additionally, Microsoft has added CoreFlood removal to its Malware Removal Tool and assisted in cleaning up infected systems. As a result of this crackdown, CoreFlood botnet activity has been reportedly reduced by as much as 90 percent in the United States and 75 percent worldwide.

2011 References:

1. Microsoft: http://blogs.technet.com/b/microsoft_on_the_issues/archive/2011/04/13/fbi-and-doj-take-on-the-coreflood-botnet.aspx
2. US DOJ: http://www.justice.gov/opa/pr/2011/April/11-crm-466.html
3. http://blogs.technet.com/b/microsoft_on_the_issues/archive/2011/04/07/initial-revelations-and-results-of-the-rustock-takedown.aspx
4. http://www.computerworld.com/s/article/9216199/Feds_to_remotely_uninstall_Coreflood_bot_from_some_PCs
5. http://www.darkreading.com/database-security/167901020/security/client-security/229401635/coreflood-botnet-an-attractive-target-for-takedown-for-many-reasons.html

2010 References:

- Massive Mariposa botnet shut down—http://www.net-security.org/secworld.php?id=8962
- Bredolab botnet shut down—http://nakedsecurity.sophos.com/2010/10/26/bredolab-botnet-shut/
- Cracking Down on Botnets—http://blogs.technet.com/b/microsoft_on_the_issues/archive/2010/02/24/cracking-down-on-botnets.aspx

Common top-level domains in URL spam

Internationalized country code top-level domains on the increase

Table 2 shows the five most frequently used top-level domains (TLDs) in spam by month.

Similar to 2010, the first half of 2011 is dominated by .ru, reaching rank one or two in every month. In March, the newcomer .me[17] made it into the top five, caused by the massive use of the advertising service zrink.me and some URL shortening services such as ino.me, shortn.me, and widg.me.

In April of this year, Russia even appears twice, one time with its traditional top-level domain .ru and a second time with its internationalized country code top-level domain .рф[18]. This top-level domain first appeared at the end of 2010, and within half a year, it has reached the top five. Since March of 2011, it has stayed within the top 15, as Figure 20 shows.

Rank	January 2011	February 2011	March 2011	April 2011	May 2011	June 2011
1.	ru (Russia)	com	com	ru (Russia)	ru (Russia)	ru (Russia)
2.	com	ru (Russia)	ru (Russia)	com	com	com
3.	uk (United Kingdom)	net	me (Montenegro)	ua (Ukraine)	net	net
4.	net	nl (Netherlands)	us (USA)	рф (Russia)	info	info
5.	info	info	net	net	cl (Chile)	cl (Chile)

Table 2: Most common top-level domains with real spam content, 2011 H1

17 .me is the Top Level Domain of Montenegro that was part of the former Yugoslavia.
18 "**рф**" are the letters rf in the Cyrillic language and mean "Russian Federation."

Because of the massive use of this new internationalized top-level domain by spammers, it is worth analyzing in more detail ***how*** spammers use this domain. For this purpose, Figure 21 shows how long spammers use one domain.

Most traditional .ru spam domains—nearly 43 percent—are used for less than 24 hours. Only 12 percent are used for longer than one month. In contrast, less than one third of the .рф spam domains are used for only 24 hours or less (a significant lower percentage for classical .ru domains) but also 32 percent of the domains are used for 30 days or longer (which is a significant higher percentage for classical .ru domains). This brings up some interesting conclusions:

- Spammers use this new top-level domain alternative right away.
- An internationalized domain is used for much longer than a traditional .ru domain. It may be because spammers expect that such URLs are not recognized correctly by some spam filters. Therefore, they do not need to change the URLs as frequently.

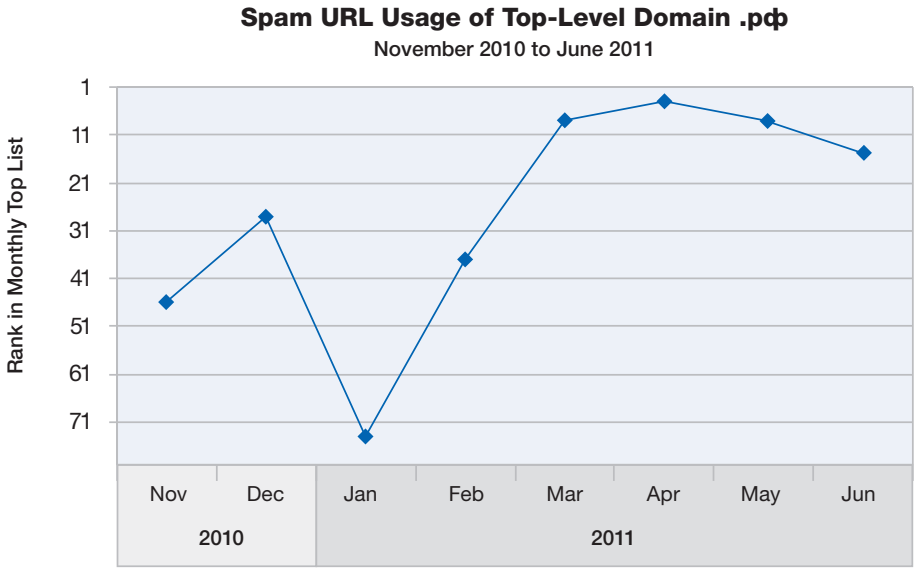


Figure 20: Spam URL Usage of top-level domain .рф – November 2010 to June 2011

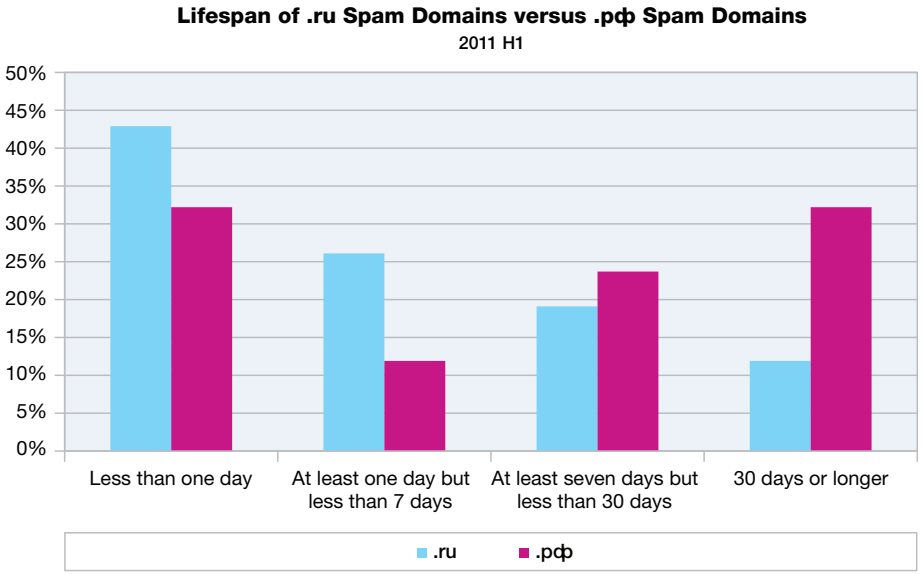


Figure 21: Lifespan of .ru Spam Domains versus .рф Spam Domains – 2011 H1

Spam—country of origin trends[19]

When looking at the countries that sent out the most spam over the last 30 months some interesting long-term trends become apparent.

- Two and a half years ago Brazil and the US dominated the market place.
- India has shown continuous growth and now dominates the scene by a large margin, sending out more than 10 percent of all spam.
- The USA owned the top position in each quarter of 2010 and now comes in last, sending out less than three percent of all spam.
- Vietnam was a major spam provider in 2009, has significantly declined in the first quarter of 2011, but has recovered a bit in the second quarter.
- Brazil has halved its percentage in the last 18 months.
- Indonesia, a relative newcomer, has shown continuous growth for 2.5 years and is now responsible for five percent of the spam.

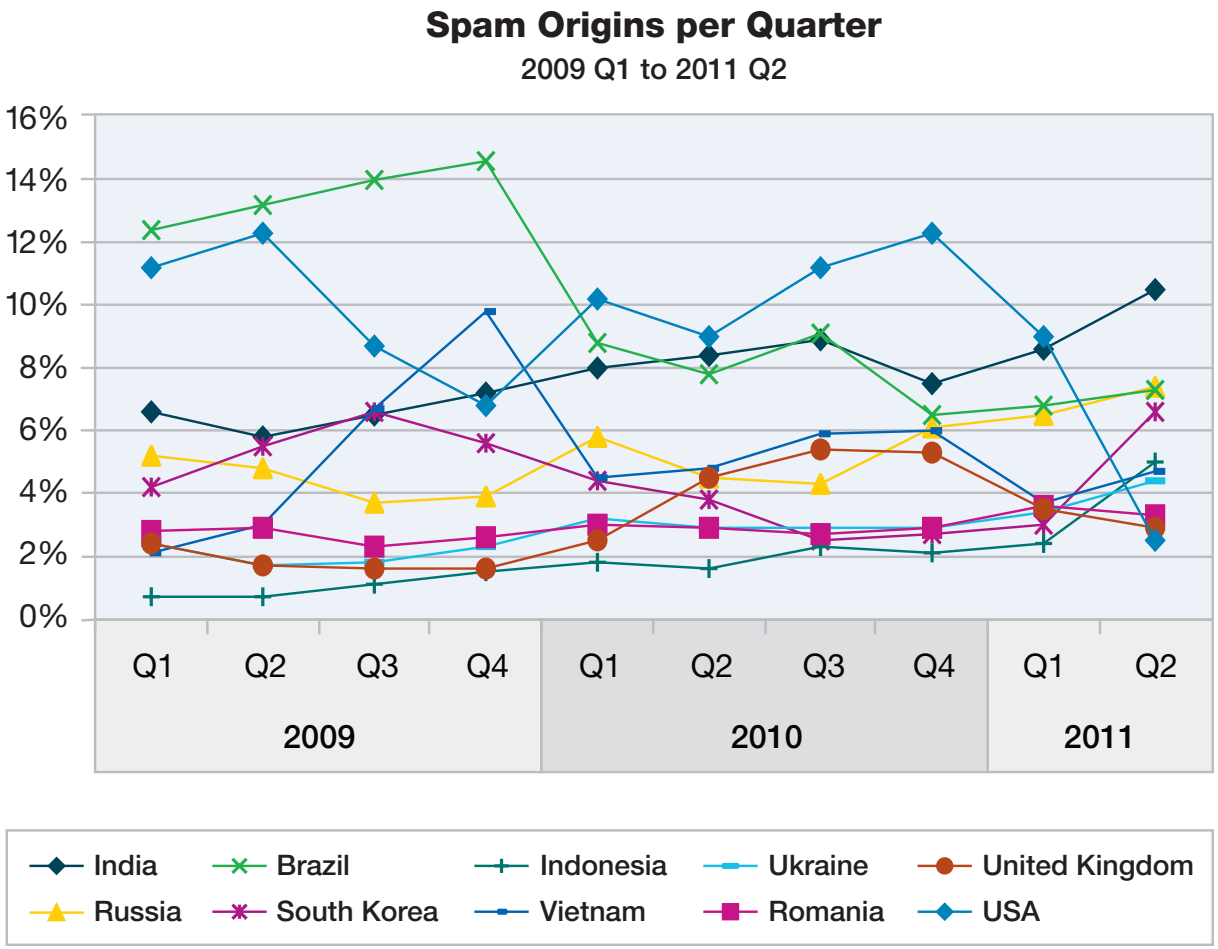


Figure 22: Spam Origins per Quarter – 2009 Q1 to 2011 Q2

19 The country of origin indicates the location of the server that sent the spam email. X-Force believes that most spam email is sent by bot networks. Since bots can be controlled from anywhere, the nationality of the actual attackers behind a spam email may not be the same as the country from which the spam originated.

Email phishing

In the first half of 2011, spammers said adieu to traditional email phishing. When looking at the percentage of spam that is phishing on a weekly basis we have measured less than 0.01 percent for every month.

Figure 23 reflects the significant decline of traditional email phishing, particularly within the last two years.

Figure 23: Phishing Volume Over Time – April 2008 to June 2011

The time when we observed huge email phishing threats luring people onto faked banking websites with a link in a more or less legitimate looking email appears to have passed. The following map shows from which countries the remaining phishing emails are sent[20].

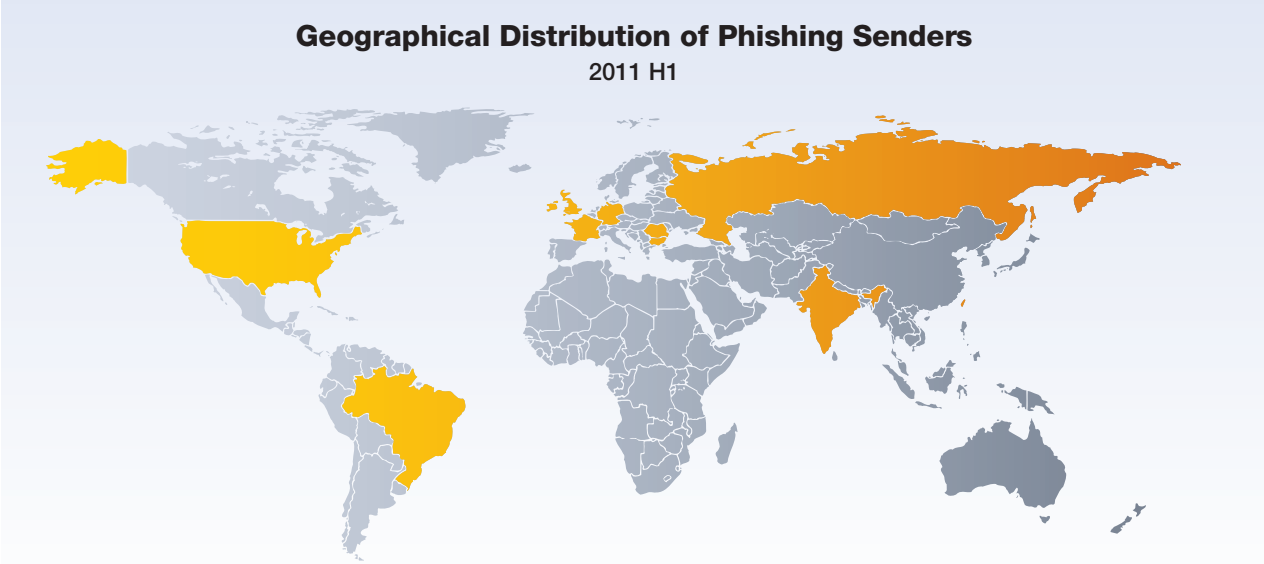


Figure 24: Geographical Distribution of Phishing Senders – 2011 H1

Country	% of Phishing	Country	% of Phishing
USA	41.5 %	India	3.0 %
United Kingdom	6.8 %	France	2.9 %
Brazil	3.5 %	Taiwan	2.7 %
Bulgaria	3.2 %	Germany	2.7 %
Romania	3.2 %	Russia	2.6 %

Table 3: Geographical Distribution of Phishing Senders – 2011 H1

20 The country of origin indicates the location of the server that sent the phishing email. X-Force believes that most phishing email is sent by bot networks. Since bots can be controlled from anywhere, the nationality of the actual attackers behind a phishing email may not be the same as the country from which the phishing email originated.

Email phishing still targets financial institutions, which represent more than 80 percent of all phishing emails in the first quarter and 31.1 percent in the second quarter. In the second quarter, online payment reached the top position for the first time at 31.7 percent. Online shops reached nearly 19 percent, and auctions 13.5 percent.

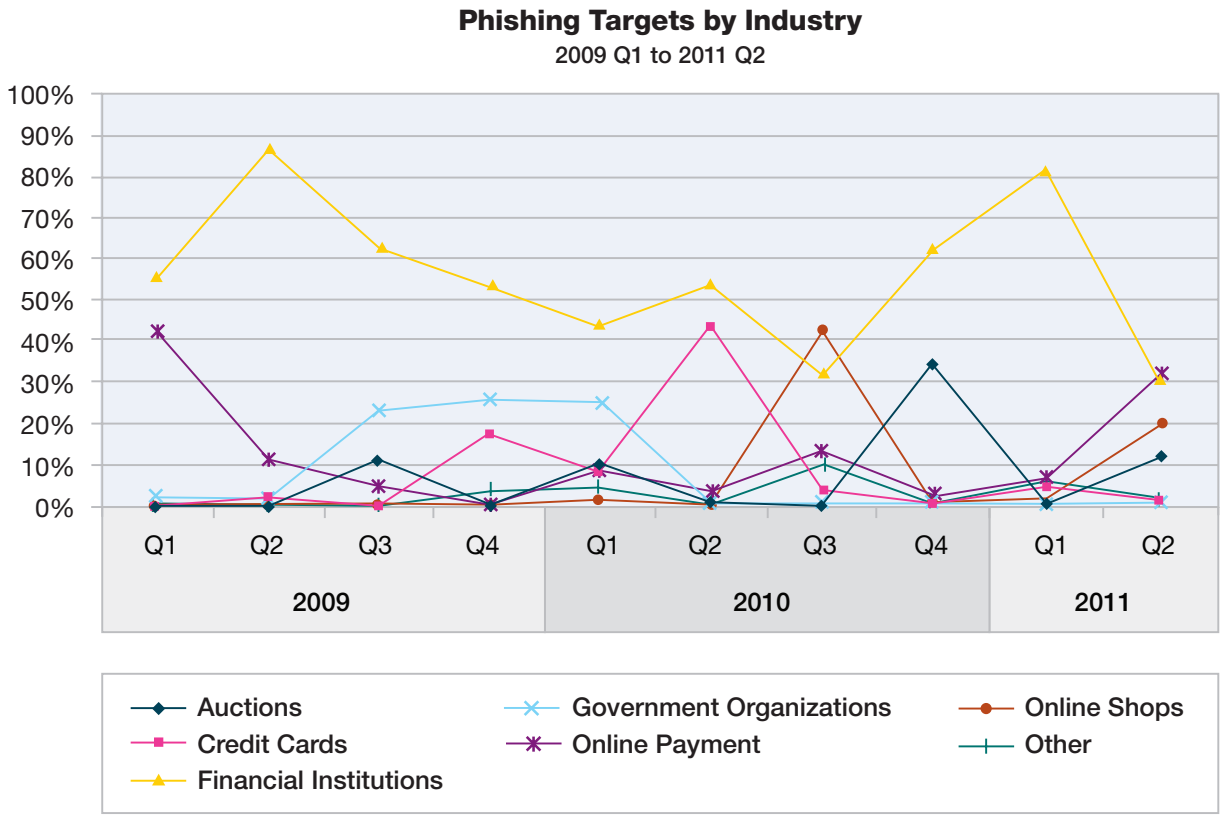


Figure 25: Phishing Targets by Industry – 2009 Q1 to 2011 Q2

Figure 26 shows the geographical distribution of financial institutions targeted by phishing emails.

As in 2010, North America is the number one region for email phishers. In the second quarter, Europe significantly increased, reaching nearly 30 percent.

Spear phishing
Spear phishing is phishing that is personalized. Phishers first gather many kinds of personal data by applying social engineering techniques. Then this data is used to compose a personal message to the victim. The personalized content assures the victim that the message is legitimate, hence, he walks right into the trap. For more information see http://en.wikipedia.org/wiki/Spear_phishing#Phishing_techniques.

ATM skimming
ATM skimmers put a device over the card slot of an ATM that reads the magnetic strip when the unsuspecting users pass their cards through it. More information about this topic can be found on http://en.wikipedia.org/wiki/Credit_card_fraud#Skimming.

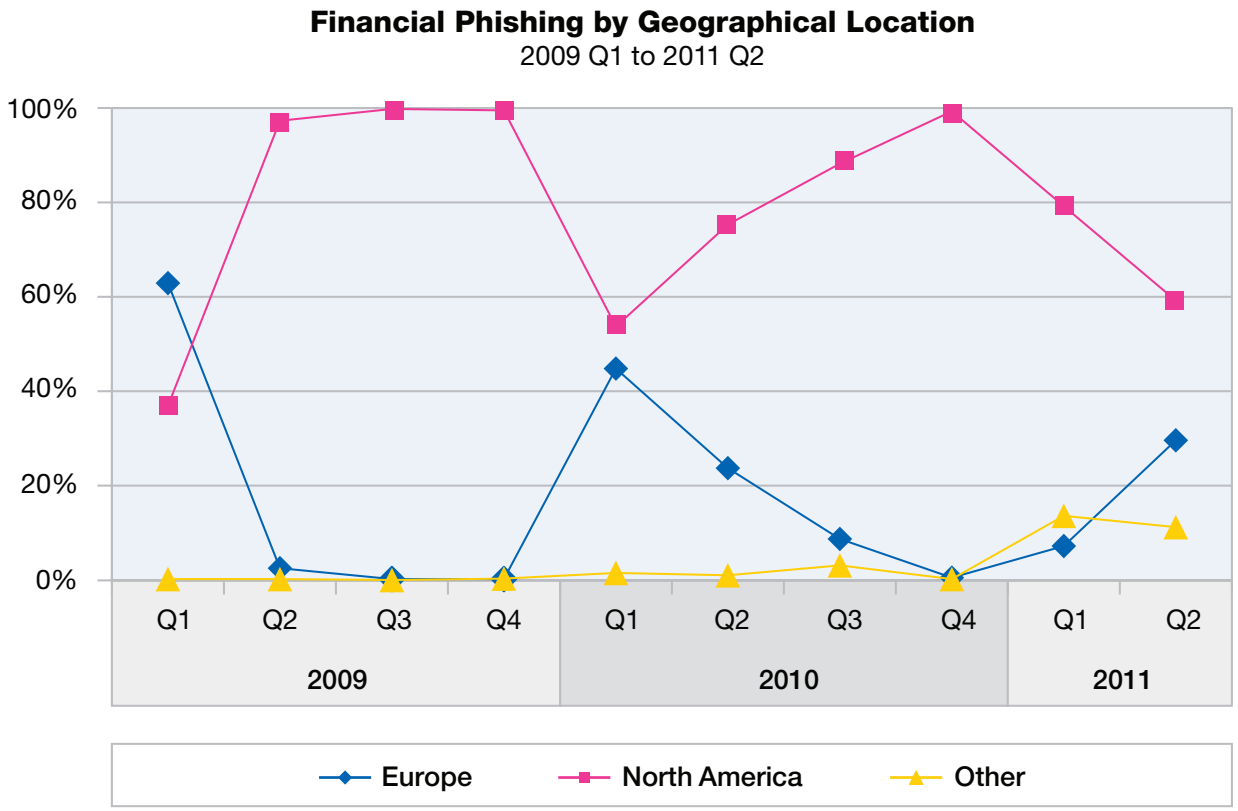


Figure 26: Financial Phishing by Geographical Location – 2009 Q1 to 2011 Q2

Note: *These statistics do not support the conclusion that phishing for passwords and credentials is dead in general. It simply suggests that phishers no longer rely on simple email phishing. It seems reasonable to conclude that they are focusing on other approaches like spear phishing or ATM skimming.*

Future prospects on spam

In the first half of 2011 we have seen significant drops in spam volume without the recovery that we have seen in the years before. The "business environment" for traditional email spam has changed.

- Organizations or companies succeeded in taking down botnets and infrastructures used to distribute spam, as seen in McColo and Rustock take downs.
- Spam filters continue to improve.
- Other approaches appear that affect the spammer's activities, as noted in "Click Trajectories: End-to-End Analysis of the Spam Value Chain"[21]. The study stated that 95 percent of the payments of spamvertized products are handled by only three banks. The banks of the spam victim could block the payment to these three banks.

This might cause the attackers to focus on other areas such as spamming within social networks or performing distributed denial of service attacks. There are even experienced spammers who consider the spam business no longer as attractive[22]. On the other hand, there are other aspects that might mislead the old and new attackers to send out more spam.

- The number of Internet users is escalating, hence, there are always new victims of spam and phishing attacks, even if only one of ten thousand spam emails reaches an inbox.
- The number of available machines is also still growing permanently. Furthermore, there is a new type of machine to infect: smart phones. And these hand-held computers have another advantage from the spammer's perspective: Contrary to desktop PCs that are turned off when not in use, smart phones are always online. Today we still have bandwidth limits in the smart phone context because most users do not have a flat rate for mobile Internet usage. This will likely change in the future. In this context, see also the section on **Mobile vulnerabilities continue to rise**.
- Regarding the type of spam content, there are some approaches spammers have not used so far such as Open Office documents as spam attachments.
- IPv6 may also provide many new approaches for spammers to bother users and torture anti-spam vendors, particularly when they rely exclusively on IP blocking.

Thus, there are many aspects that may influence the development of spam volume in the future. Assuming that the number of attackers does not decrease, we are curious whether they will still use spam to do damage or whether they will focus on other techniques.

21 See http://cseweb.ucsd.edu/~savage/papers/Oakland11.pdf.
22 See http://www.itworld.com/security/178991/internet-evolves-there-place-spam.

Section II
Operating a Secure Infrastructure

In this section of the Trend Report we explore those topics surrounding the weaknesses in process, software, and infrastructure targeted by today's threats. We discuss security compliance best practices, operating cost reduction ideas, automation, lowered cost of ownership, and the consolidation of tasks, products, and roles. We also present data tracked across IBM during the process of managing or mitigating these problems.

Preparing for a breach: incident response handling (IRH)

"An infinite number of monkeys with an infinite number of typewriters and an infinite amount of time could eventually write the works of Shakespeare "—The Infinite Monkey Theorem

Previous IBM X-Force® Trend and Risk Reports have discussed numerous mechanisms by which a network or its users can be exploited and compromised. Recent high profile hacks and attacks have demonstrated that it is likely that your organization will need to perform incident response handling (IRH) to a suspected or valid incident sometime in the future. Some organizations have in-house capabilities to perform IRH while others perform the initial first responder steps, and then seek additional resources from an outside provider.

In order to facilitate an initial response by first responders which will help provide a good analysis environment for secondary responders, IBM Emergency Response Service (ERS) has developed some suggestions to assist with the initial response.

Stop, drop and roll

Every organization should have a Computer Security Incident Response Plan (CSIRP) to implement when a computer security incident occurs. Depending on the size of the organization, its resources and the frequency at which it encounters incidents, the CSIRP can range from something as simple as a phone number to an organization to which they have outsourced support to running their own full-service incident response organization. When an incident is declared: *Plan the work. Work the plan. Don't run around like your clothes are on fire*. The CSIRP is there to make sure that all aspects of the incident are covered, that decision makers are informed so they can make correct and timely decisions, and that support can be provided while mistakes are avoided. Practice the CSIRP in advance to make sure that it is a sound plan and provides the needed structure.

"An infinite number of hackers with an infinite number of keyboards, an infinite amount of caffeine, and an infinite amount of time could eventually compromise a network." —Stone's Corollary to the Infinite Monkey Theorem

Train first responders

First responders involved in the IRH process should have enough training to fulfill the position to which they are assigned on the response team. This starts with having enough situational awareness of the threats to recognize a situation that is hazardous to your network. Responders should be trained to recognize that there is a difference in the level of concern between an incident with a single virus and an incident with multiple fraudulent user accounts created along with trojans, keyloggers, and Zbot or Zeus (which is a Trojan horse that steals banking information by keystroke logging.) The concern, remediation steps, and notifications for the two incidents would not be the same.

Move forensic efforts earlier into the process

Frequently, there is a period of time between when the incident is recognized as an event and it is designated as an incident and the CSIRP is implemented. Actions during this crucial period can determine whether a successful analysis can be accomplished. Keep in mind that activities such as running Anti-Virus and malware scans, patching, deleting logs and changing configurations can have a destructive impact on the file system in addition to the goal of solving the problem. For example, part of the process for malware incidents is to turn off and delete restore points. It is common for relevant data to be recovered from restore points and deleting the restore points makes that data unavailable for analysis. This data might include copies of the initial dropper files, malware, keystroke logs, and other files to identify the initial infection vector and content of stolen data. As the event progresses and it begins to appear that this is an actual incident, implement forensic procedures early in the process to capture volatile data and drive images. Otherwise, you may destroy data valuable to the analysis.

Know where your PI, PII, HIPAA and Secret Sauce reside

PI stands for Personal Information

PII stands for Personally Identifiable Identification

HIPAA stands for Health Insurance Portability and Accountability Act. HIPAA is a privacy act that was created in 1996 by U.S. Congress; its sole purpose is to protect individuals and their medical privacy.

Incident responders should know details about the construction of your network, what network devices may have relevant logs, and what type of sensitive data should or should not be found on the analyzed computers. During an analysis, data may be recovered that points to IP addresses of other computers in the network as either a source or target of malicious traffic. You should be able to locate the owner and physical location of an IP address quickly if it is an ongoing incident. Also, you should be prepared to provide passwords and decryption keys for encrypted data.

You should have a knowledgeable and reachable legal contact

During an incident, numerous legal questions can arise that require a decision from the legal representative for the organization. These questions can range from the legal authority requiring access an employee's personal storage devices to the legality of taking specific data types from a foreign country. Additionally, if there is an anticipation of HR or legal action as a result of the incident, the legal adviser may have specific questions they need answered. The legal adviser should have enough training in IT legal issues to provide appropriate answers. They should also be included in discussions at decision points to help ensure the company is taking legally correct action.

The client representative should have the authority to make things happen

The point of contact between the organization and the responding ERS group should have enough authority to get things done, both organizationally and on the network. This person doesn't have to be at the top of the CSIRP management hierarchy, but should have sufficient knowledge of the organization and network to obtain needed resources. Additionally, the first responder should have sufficient network access privileges to run tools that capture volatile data and forensic images.

You should invite somebody with a checkbook to the party

The incident response process often involves the expenditure of money. The bulk of this expense is identified and planned if support has to be purchased. Frequently, what is not planned are the day-to-day expenditures at the responder level. Even though a multi-thousand dollar contract is approved for the engagement, the process may be held up by the inability to obtain a $50 data storage hard drive in less than a week.

You should understand the capabilities of the malware

During an incident involving malware, incorporate the known capabilities of the malware while planning and executing the response plan. For example, malware traveling via USB autorun is a frequent trait of many of today's malware. Because of this, the use of USB drives to transfer remediation tools or volatile data captures and logs between contaminated computers and responder computers frequently can result in spreading the malware from a contaminated computer to a non-contaminated computer.

Quarantine, don't delete

Many incident response activities involve malware. The malware is frequently identified by an anti-virus scanner which can then either delete or quarantine the malware. If the malware is deleted, it is no longer available for analysis to determine its capabilities and the dates and times associated with the malware file. These dates and times can help identify the circumstances under which it was installed on the computer and other file system activity occurring at the same time, such as the creation of keystroke logging files.

Windows 7 configurations

For those networks that have transitioned to Vista and Windows 7 (or are in the process), consider turning on the last access update key in the registry. In Windows 7, one of the dates that can be tracked related to a file is the date the file was last accessed. This could be the date it was printed, scanned by anti-virus, copied, or opened. To save time, a default install of Windows 7 has the "update" of the last access date and time turned off. This last access date and time is frequently critical during the analysis of an incident to establish a timeline to determine what data may have been stolen or touched.

Logging—Yes/No

The default logging for Windows XP is not configured to assist most enterprises with post-incident analysis. Some logging is not activated, appropriate events aren't being logged, and default log sizes are small. Misconfigured logging has been seen to document as little as 52 minutes worth of activity before the small log size causes the newest entries to overwrite the oldest.

Good company cafeteria

While a good company cafeteria isn't necessary, the point is that the support functions occur behind the scenes. These functions range from on-site food support to a rotation plan for personnel involved in the response and remediation process. For example, it is not uncommon for ERS to respond to an organization that is in the midst of response and remediation and to learn the responding staff has been without sleep for over 24 hours. While sometimes necessary to achieve the goals of the operation, this raises several concerns:

- Pre-incident planning and preparation did not provide for more than one person to have the skills and ability to respond to and remediate the incident. Plan on having more than one person with the appropriate skills.
- Studies have shown that after 24 hours with no sleep, a person's cognitive and physical performance is the equivalent to being legally intoxicated. Their decision-making ability is reduced and the time it takes to make decisions is extended.
- From a safety perspective, the health of the responders can be affected by the stress and long work hours imposed by the incident management process. The military conducts operations in excess of 24 hours but they train for it and the physical conditioning of soldiers is typically better than that of IT personnel. Accidents are more likely to occur as decision-making ability decreases and response time increases.

Over the years, police tracking teams have developed the concept of the "Time-Distance Gap" which compares the difference in time it takes for a fleeing person to cover a distance compared to the time it takes for the tracker to cover the same distance. A fleeing felon can run 50 yards in seconds but the tracker may spend an hour following the minuscule signs of the person's passage over the same 50 yards. Incident response suffers from the same "Time-Distance Gap" delay in detecting the attack, accumulating and analyzing data, and implementing defenses based on the analysis. The details of the two minutes worth of activities performed by an intruder may take hours or days to obtain and analyze from log files, delaying the implementation of defenses. Whether the incident response is performed in-house or outsourced, each of the items discussed here is intended to help responders reduce the "Time-Distance Gap" in IRH to achieve a quicker detection, data analysis, and defense implementation.

Potential Sources of Data for IRH Analysis	
Network Topology Diagram	Notes on First Responder Actions (scans, updates)
Network Logs (Firewall, DNS, Proxy, IDS/IPS)	Baseline Image (used to exclude known files)
Host IDS Event Logs	Packet Captures/Port Scans
Operating System Logs	Application Logs (WWW, FTP, VPN)
Database Logs	System Backups/Forensic Images
RAM and other Volatile Data	Samples of Suspicious Files Located
User Interviews	Anti-Virus Event Logs

Table 4: Potential Sources of Data for incident response handling analysis – 2011 H1

Vulnerability research

Attacks on computer networks often leverage vulnerabilities in the software running on those networks. To operate networks safely, you should be aware of those vulnerabilities and the patches that fix them. Since 1996, the X-Force Database has kept track of every public report of a security vulnerability disclosure or remedy that we've been able to get our hands on. For years we've been reporting on this data in the biannual X-Force Trend Report. This data tells us a great deal about the nature of the security issues we're mitigating on our networks and how these issues change over time.

Please note that, in many of the charts on vulnerabilities included in this section, totals for 2011 are presented as projections. It is difficult to compare data from half of a year to previous trends based on annual totals. In order to make the comparisons easier to see and understand, we have, in some cases, doubled the number of vulnerabilities that we have seen so far in 2011 in order to create a projected total for the full year which can then be compared with previous full-year totals and identified as 'P' within the charts. Of course, the trends we've seen in the first half of 2011 may or may not hold through to the end of the year, so the final charts that we publish in our end of the year report may well look different.

Total number of vulnerabilities decline—but it's cyclical

In the first half of 2011 we saw fewer total security vulnerability disclosures than we saw last year at this time. This might not be surprising to those who have been following vulnerability disclosure for many years. The volume of security vulnerability disclosures seems to follow a two-year alternating cycle. In 2007 there were fewer vulnerability disclosures than in 2006, the number was back up again in 2008, and then back down in 2009, but over time the totals seem to be creeping higher. Last year, 2010, saw the largest number of vulnerability disclosures on record, over 8500. This year we're on track for just over 7,000 disclosures, a significant decrease from last year, but about the same amount that we saw in 2006.

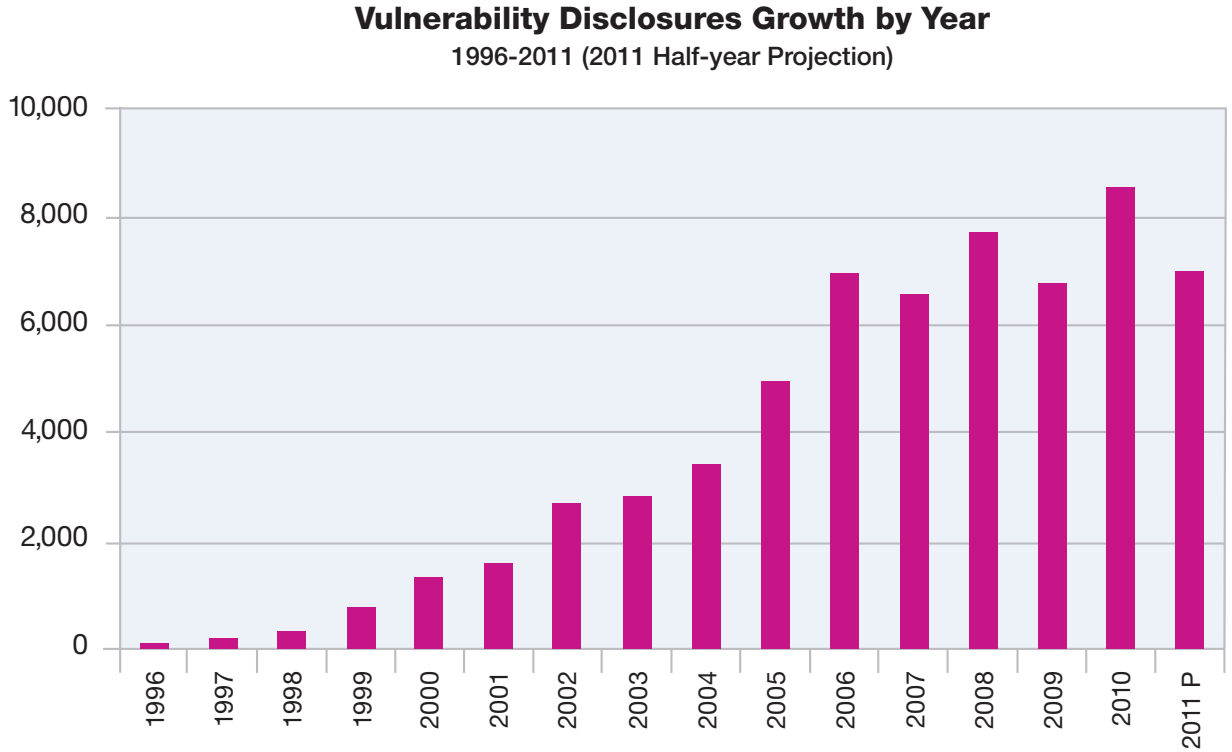


Figure 27: Vulnerability Disclosures Growth by Year – 1996-2011 (2011 Half-year Projection)

Where is the decrease? It is primarily in web application vulnerabilities. For the past few years approximately half of the security vulnerabilities that were disclosed were web application vulnerabilities. This year the number is down to 37 percent, with a significant drop in the volume of SQL injection vulnerabilities in particular. Does this mean we can stop worrying about web application security problems? Of course not. There are still a large number of these vulnerabilities being disclosed.

In fact, we are loathe to make any long-term predictions about this category of security issues. A decrease this year might be a sign of progress—it might mean that SQL injection vulnerabilities are getting harder to find—meaning that web application developers are writing better code that is less susceptible to them. Over time, this might mean that the web will become more secure. However, we've made this kind of prediction in the past regarding declining categories of security vulnerabilities and then been surprised when the disclosure numbers took off again. It will take a longer sustained trend before we're comfortable making predictions.

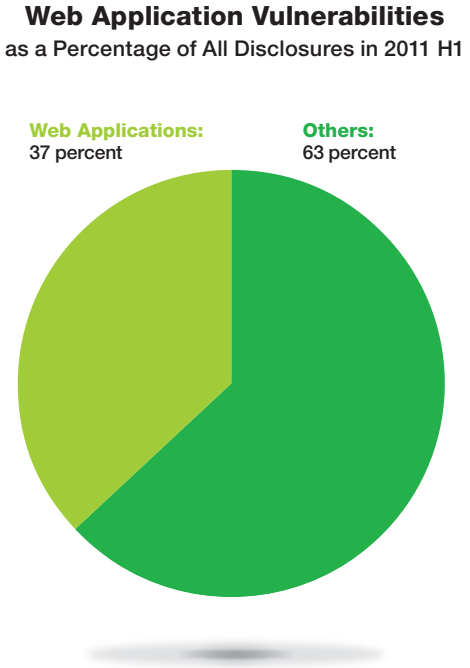


Figure 28: Web Application Vulnerabilitiesas as a Percentage of All Disclosures in 2011 H1

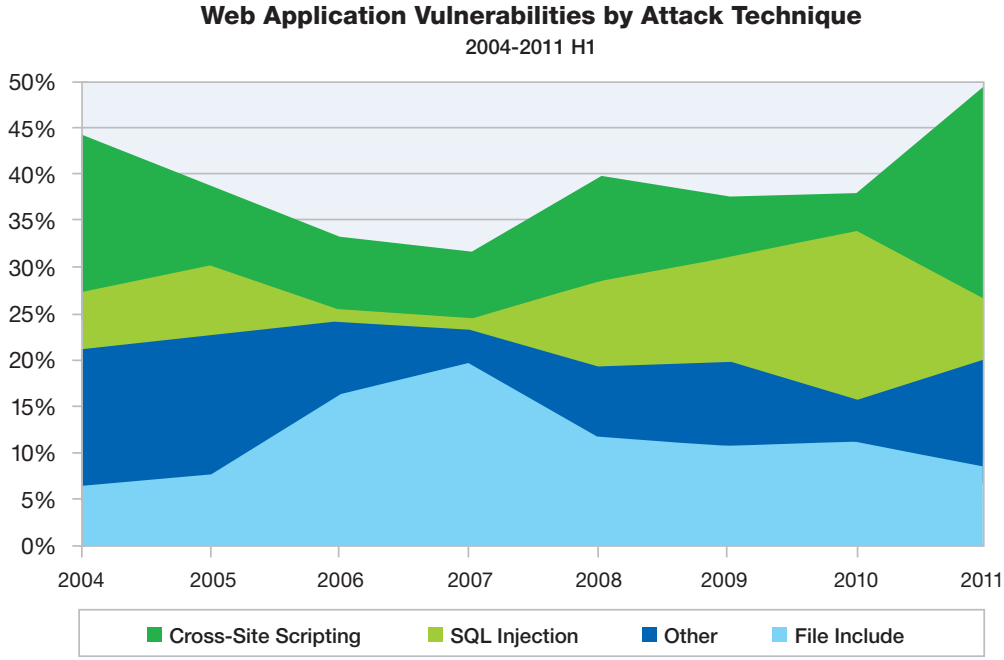


Figure 29: Web Application Vulnerabilities by Attack Technique – 2004-2011 H1

Are Web browsers safer?

For example, in 2009, we saw fewer high and critical web browser vulnerabilities disclosed than in 2008. (We consider high and critical vulnerabilities to be those that have a Common Vulnerability Scoring System (CVSS) score of 7.0 or greater.) At the time this appeared to be a victory. There is a lot of attack activity targeting web browser vulnerabilities, and so there is a corresponding amount of focus on finding, disclosing, and patching these vulnerabilities in hopes of reducing the attack surface area of the browser. When we saw the numbers start to drop, we thought, maybe we've rounded the corner on this. Maybe we're really moving toward a day when the browser is safer.

Unfortunately, the number of high and critical browser vulnerabilities was back up in 2010 and that year, the total number of browser vulnerabilities was up significantly. In the first half of 2011, that total number is still climbing, but the number of high and critical vulnerabilities has dropped to a level that the industry has not seen since 2007. When you take a look at the downward trend in high and critical vulnerabilities from 2009 through 2011 (projected), there does seem to be a steady decline. The industry seems to be getting better at making safe browser software, even as the browser market has become more competitive. Perhaps these are signs of progress.

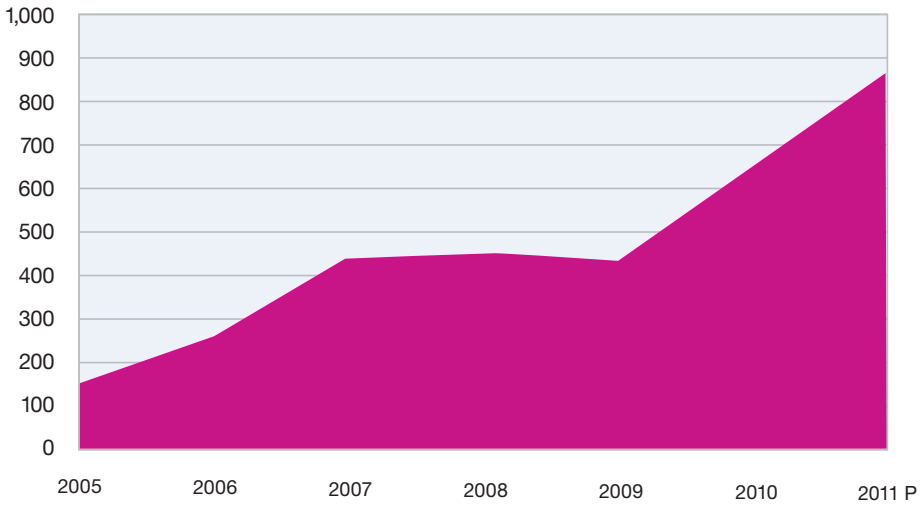


Figure 30: Web Browser Vulnerabilities – 2005 – 2011 H1 (projected)

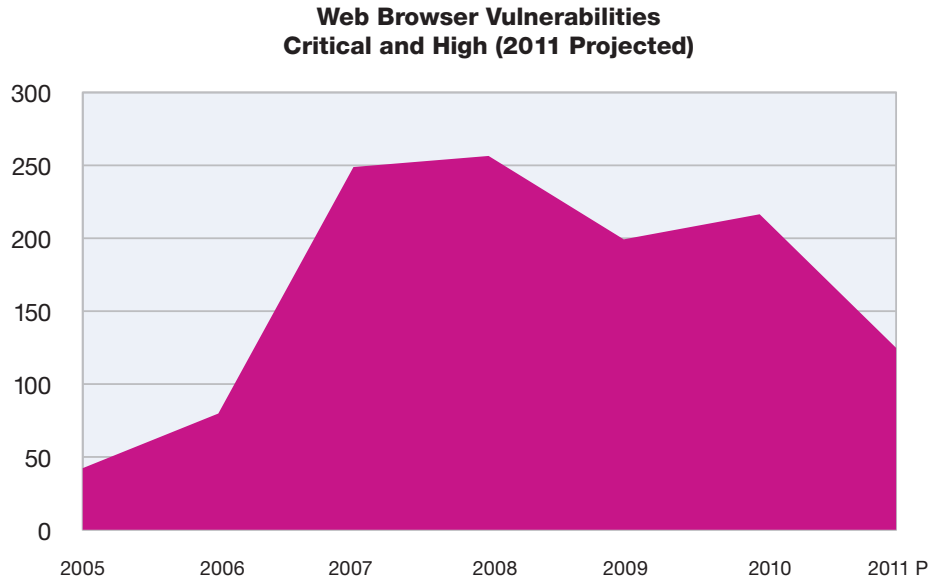


Figure 31: Web Browser Vulnerabilities, Critical and High – 2005 – 2011 H1 (projected)

Another area that saw significant decline in the first half of 2011 was the public release of exploit code targeting security vulnerabilities. We saw fewer true exploits released so far this year since 2006. Although we can only speculate as to the cause, the decline occurred on both real terms and on a percentage basis versus the total number of vulnerabilities disclosed. So far only about 12 percent of the vulnerabilities that have been disclosed have seen true exploit releases, whereas in previous years the number was closer to 15 percent.

There is a window of opportunity that an attacker has to target a security vulnerability. That window opens when the vulnerability is first discovered, and it closes when the vulnerability is finally patched on a vulnerable system. The period of time between vulnerability disclosure and patch release constitutes one part of this window, and the other part is the period of time that it takes people to install that patch on vulnerable systems in their networks.

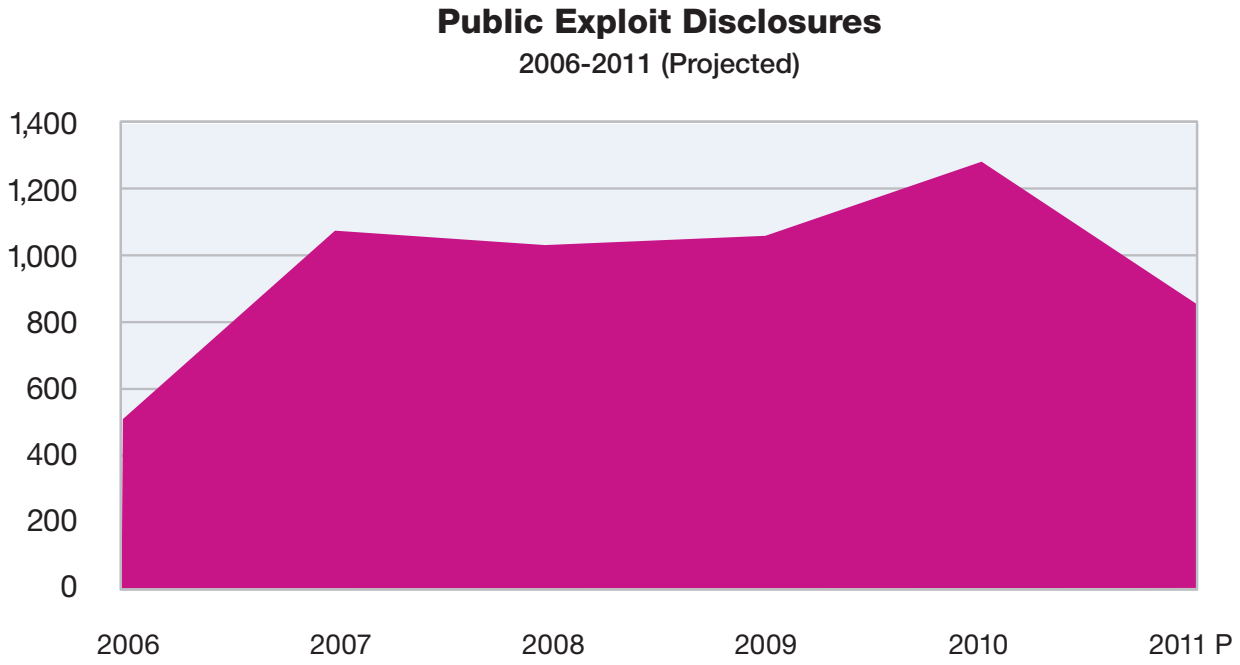


Figure 32: Public Exploit Disclosures – 2006-2011 (Projected)

True Exploits	2006	2007	2008	2009	2010	2011 Projected
Percentage of Total	7.3 %	16.5 %	13.4 %	15.7 %	14.9 %	12.0 %

Table 5: Public exploit disclosures – 2006-2011 (Projected)

About 58 percent of the vulnerabilities that were disclosed during the first half of 2011 had a remedy available on the same day that they were publicly disclosed—which is the ideal case. On the other hand, about 37 percent have no remedy available as of this writing. This is a significant improvement from previous years—the number of unpatched vulnerabilities hasn't dropped below 44 percent of the total in over 5 years. The remaining 5 percent in the middle represent cases where a patch was made available sometime after public disclosure of the vulnerability. The worst case in our data set was 171 days. Fortunately, there is only a handful of high-severity, enterprise software vulnerabilities that fit into this middle category.

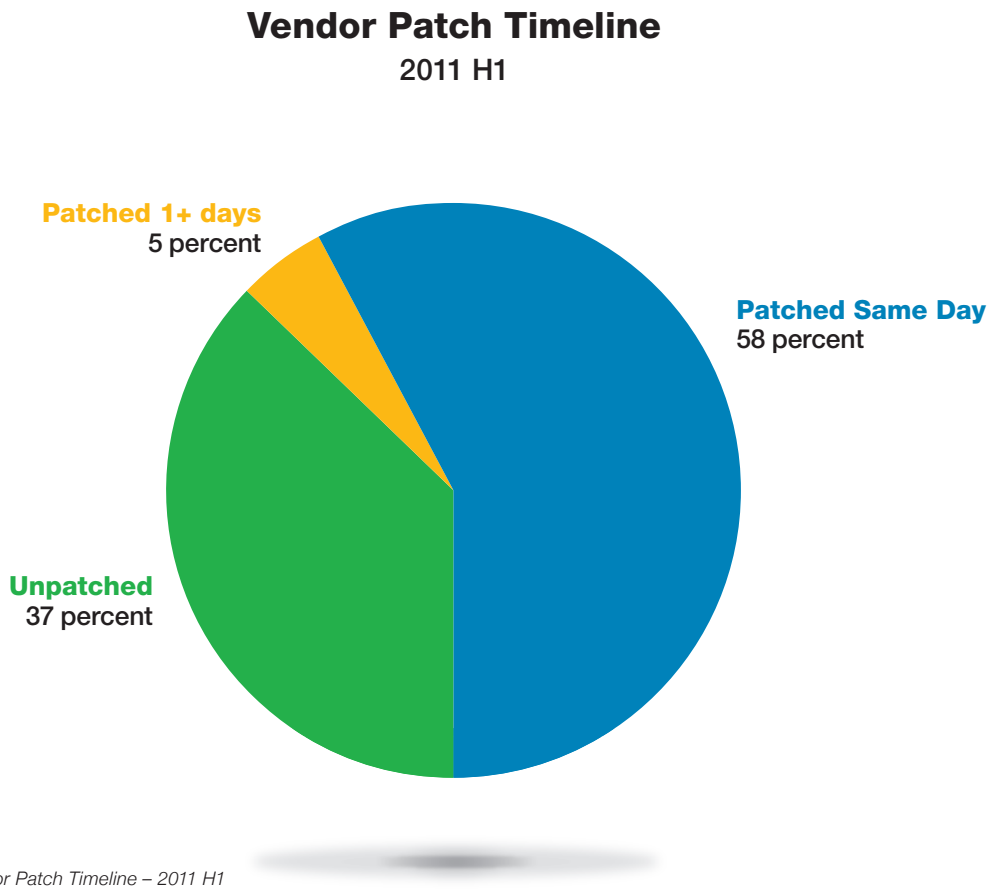


Fig. 33: Vendor Patch Timeline – 2011 H1

Once a patch becomes available, network administrators should install it in a timely manner. We don't have a direct data source for measuring the amount of time it is taking to patch networks, but there may be an indirect way of measuring this window. As we previously stated, exploit code was publicly released for about 12 percent of the vulnerabilities that were disclosed this year. The timing of these exploit releases is interesting. Often exploits are released the same day or shortly after a vulnerability is disclosed, but in some cases many weeks or months go by before the exploit code surfaces. Some of this time delay may represent situations where exploit code is being used to target vulnerable networks and only surfaces publicly once its value has waned, because the vulnerabilities have finally been patched.

Although a number of the vulnerability statistics decreased during the first half of this year, there are important areas of increase. Our security vulnerability database covers vulnerabilities in all different kinds of software products from the most critical enterprise software to minor software packages with a handful of users. A decrease in the overall number of security vulnerabilities might not have an impact on the workload experienced by enterprise IT operations unless that decrease is focused on vulnerabilities that impact enterprise software. In fact, so far this year we've seen more security vulnerabilities disclosed in major enterprise software packages than we've seen in previous years, which means that IT security professionals may have more work to do patching and remediating these vulnerabilities than they have in the past.

Patch Timeline	Software Vendors	Major Software Vendors
Same Day	2058	1229
Week 1	72	9
Week 2	30	6
Week 3	7	0
Week 4	14	5
Week 5	15	0
Week 6	11	4
Week 7	5	2
Week 8	2	1

Table 6: Patch release timing of all software vendors vs. major software vendors – 2011 H1

Exploit Timing	0 Days	1 Month	2 Months	3 Months	4 Months
0 Days	854	308	23	12	6

Table 7: Public Exploit Disclosure Timing by Weeks – 2011 H1

Companies that ship a lot of software tend to be subjected to a lot of security vulnerability disclosures. When we look at the ten software vendors with the largest number of vulnerability disclosures, excluding web content management systems, we get a list of the largest enterprise software companies that is fairly consistent year over year. In 2009 this group represented 24 percent and in 2010 this group represented 25 percent of the total number of security vulnerability disclosures. This year that number rose to 34 percent as the total number of vulnerabilities disclosures decreased.

In 2010, the number of vulnerabilities disclosed by this group of ten software vendors increased an average of 66 percent over 2009, with eight of the 10 top vendors seeing increases. This increase seems to have held in 2011 despite the overall decline in vulnerability disclosures. During the first half of 2011, we've seen another 28.6 percent average increase from this group, with half of the group increasing and the other half decreasing. The bottom line is that enterprise IT staff are spending just as much, if not more time installing patches this year as they have in the past.

Top Ten Software Vendors with the Largest Number of Vulnerability Disclosures
2009 – 2011 H1

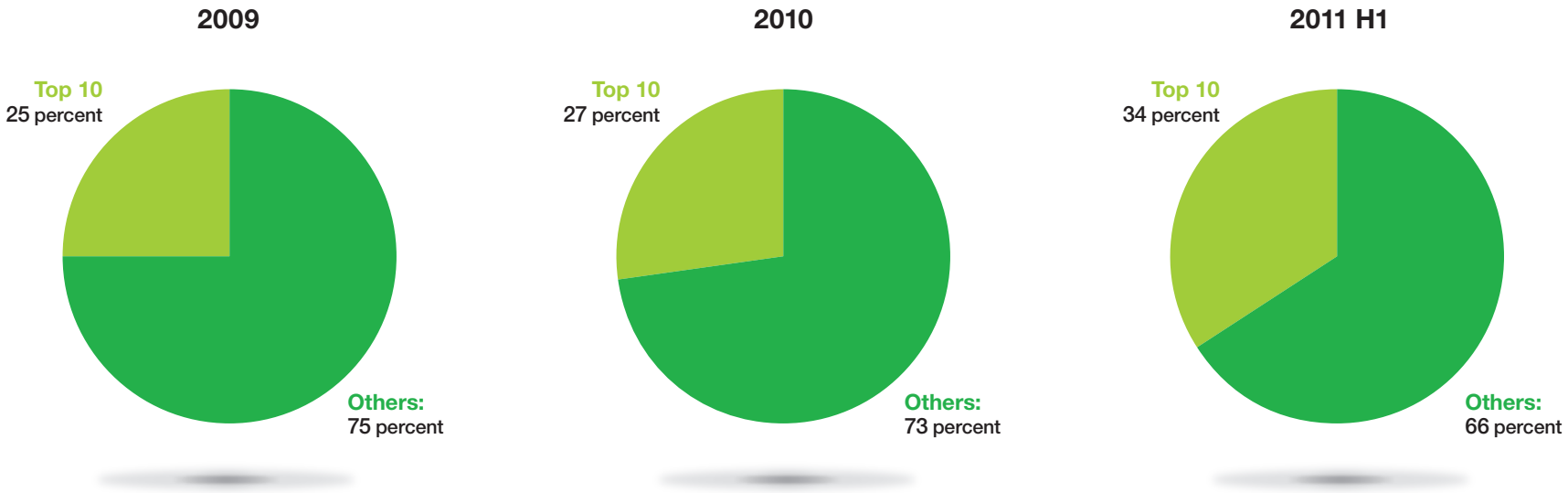


Figure 34: Top Ten Software Vendors with the Largest Number of Vulnerability Disclosures – 2009 – 2011 H1

Critical vulnerabilities are on the rise

Another variable that is up substantially is the number of critical vulnerabilities. These are security vulnerabilities with a Common Vulnerability Scoring System (CVSS) Score of 10 out of 10. For the past two years approximately 1 percent of the vulnerabilities that have been disclosed have had this score, but so far in 2011 the number is up to 3 percent, and it has already exceeded the total for 2010. Almost every one of these critical vulnerabilities is a serious remote code execution issue impacting an important enterprise class software product. This is another reason why there has been little rest for IT security professionals this year.

CVSS Score	Severity Level
10	Critical
7.0-9.9	High
4.0-6.9	Medium
0.0-3.9	Low

Table 8: CVSS Score and Corresponding Severity Level

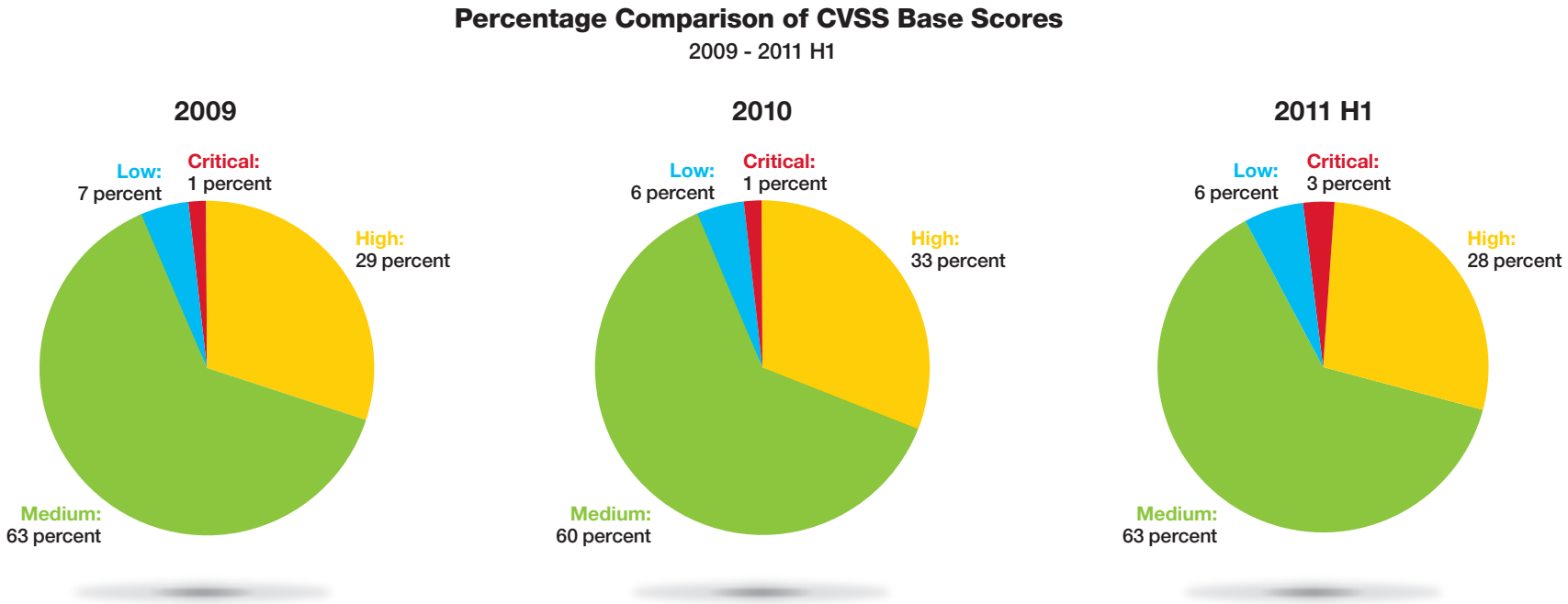


Figure 35: Percentage Comparison of CVSS Base Scores – 2009 – 2011 H1

Changes in client-side, multi-media and document readers

Back in 2005 and earlier the most common type of client-side vulnerability was a vulnerability in the desktop operating system. In fact, desktop operating system vulnerabilities were a very important attack vector during that period of time as some turned out to be exploitable by Internet worms. In the latter half of the past decade, better security practices moved the focus from the OS to the browser. However, so far in 2011, we've seen a large number of desktop OS vulnerabilities, with high and critical desktop operating system vulnerabilities exceeding high and critical vulnerabilities disclosed in browsers. These vulnerabilities fit into a number of different categories. Some of them could be exploited by worms in theory, but this has not occurred in practice. Advanced security features in modern operating systems have made exploitation of some of these vulnerabilities more challenging than they were many years ago, and so far that seems to have had a positive impact.

The major types of vulnerabilities affecting clients continue to fall into one of four main categories shown in Table 9.

Category	Description
Browser	Client Web browser software and plug-ins.
Document Reader and Editor	Software that allows users to create or view documents, spreadsheets, presentations, and other types of files that are not images, music, or movies.
Multimedia	Software that allows users to view or create music and movies.
Operating System	The base operating system, excluding applications that are in the other three categories.

Table 9: Key Vulnerability Categories Related to Client-Side Vulnerability Disclosures in 2011

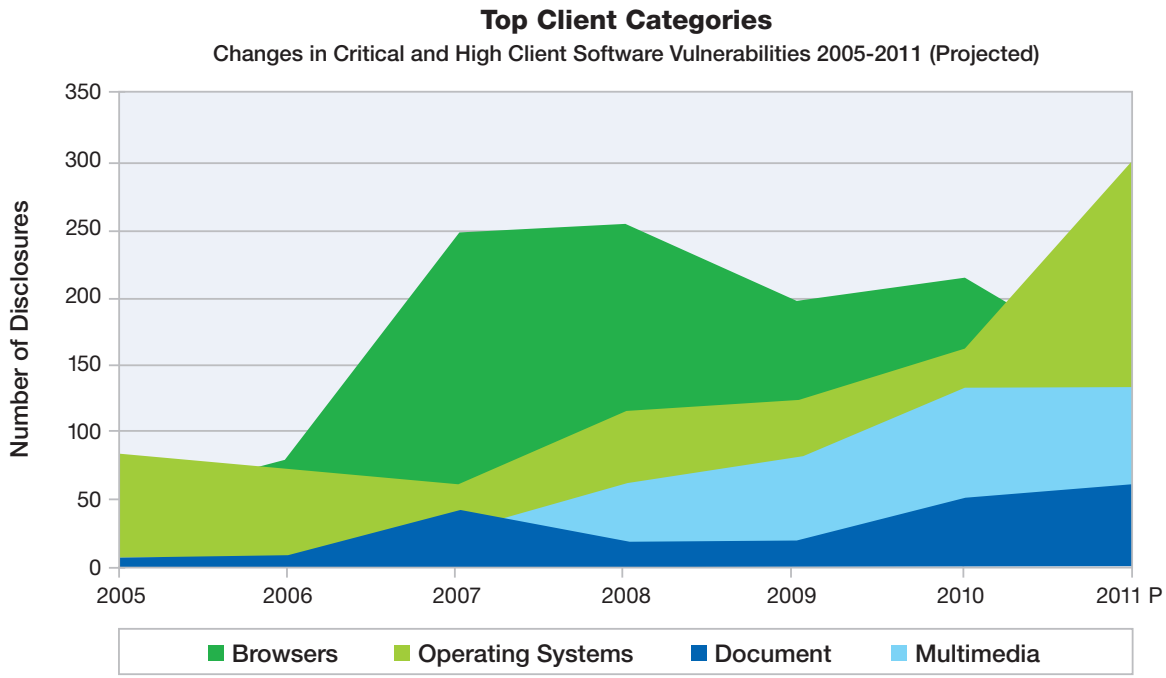


Figure 36: Top Client Categories—Changes in Critical and High Client Software Vulnerabilities 2005-2011 (Projected)

Two other areas that have seen significant increases are vulnerabilities in document readers and multimedia players. As the browser market has become more competitive, attackers have zeroed in on software like this that consumers are running regardless of what browser they prefer—allowing attackers to net the highest number of victims with a particular exploit. Recent efforts to sandbox some of these applications should force attackers to change strategies, but sandbox technology is not perfect. This topic was the subject of a presentation by X-Force Researchers Mark Yason and Paul Sabanal at Blackhat USA 2011.

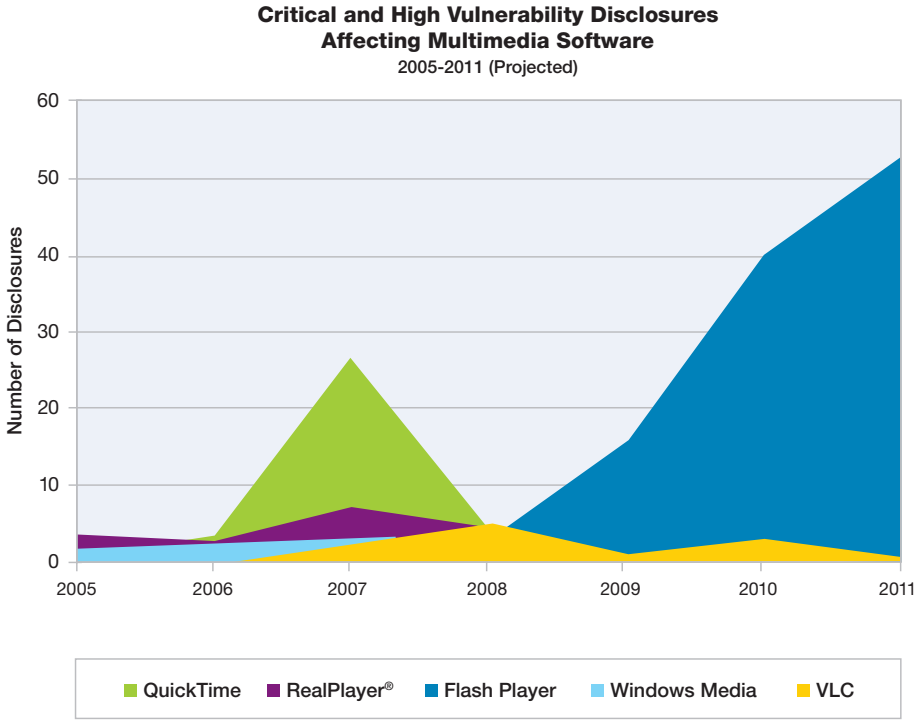


Figure 37: Critical and High Vulnerability Disclosures Affecting Multimedia Software –2005-2011 (Projected)

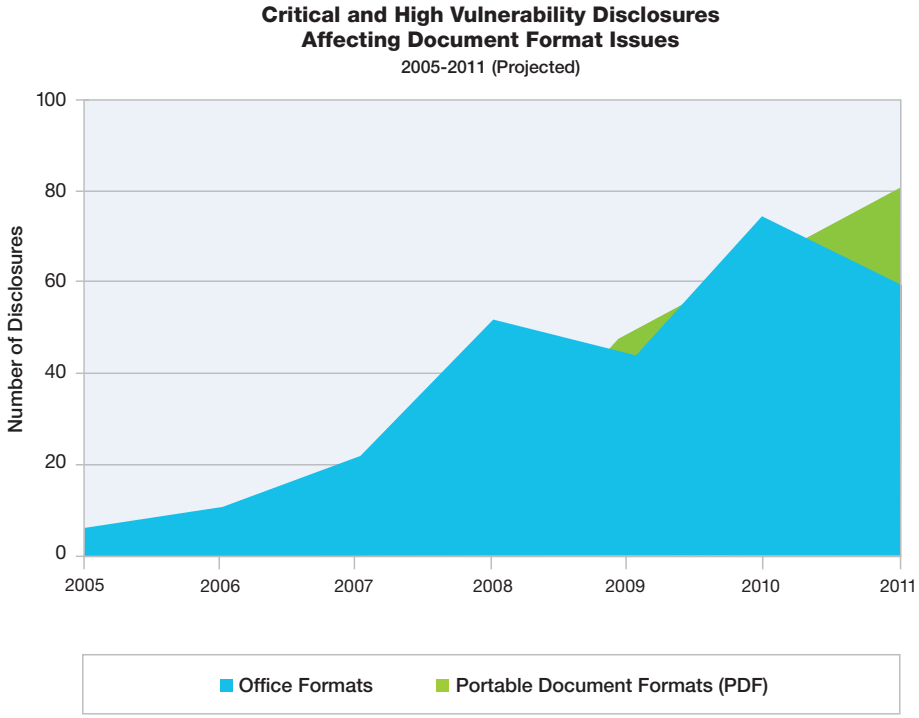


Figure 38: Critical and High Vulnerability Disclosures Affecting Document Format Issues – 2005-2011 (Projected)

Mobile vulnerabilities continue to rise

We've seen continued interest in Mobile vulnerabilities as enterprise users bring smartphones and tablets into the work place. The first half of 2011 saw an increased level of malware activity targeting the latest generation of smart devices, as attackers are finally warming to the opportunities these devices represent. The increased number of vulnerability disclosures and exploit releases targeting these platforms that we saw in 2010 has been sustained into 2011 and shows no sign of slowing down.

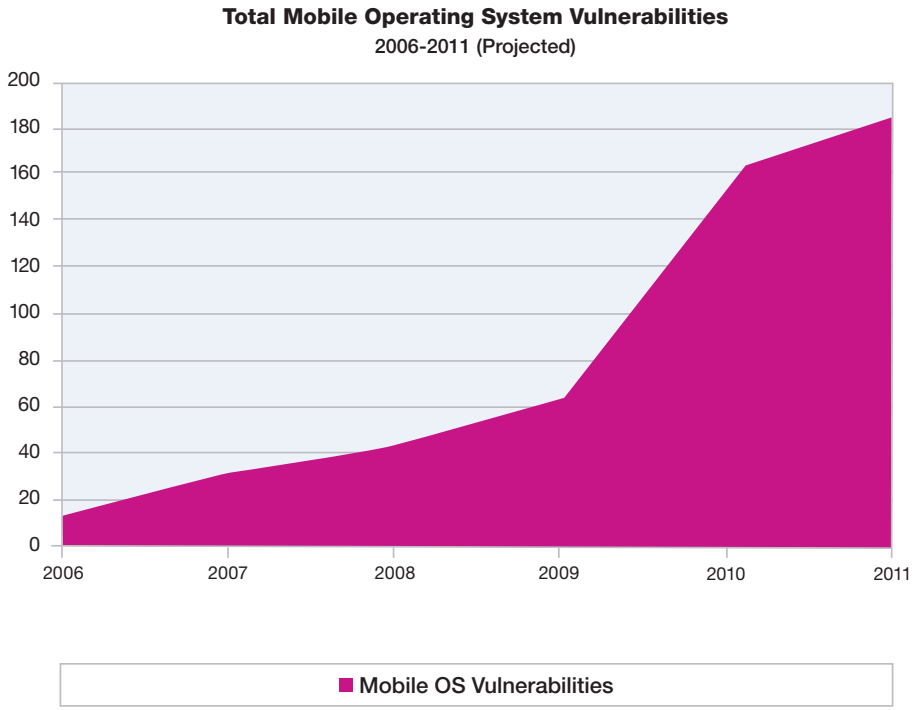


Figure 39: Total Mobile Operating System Vulnerabilities – 2006-2011 (Projected)

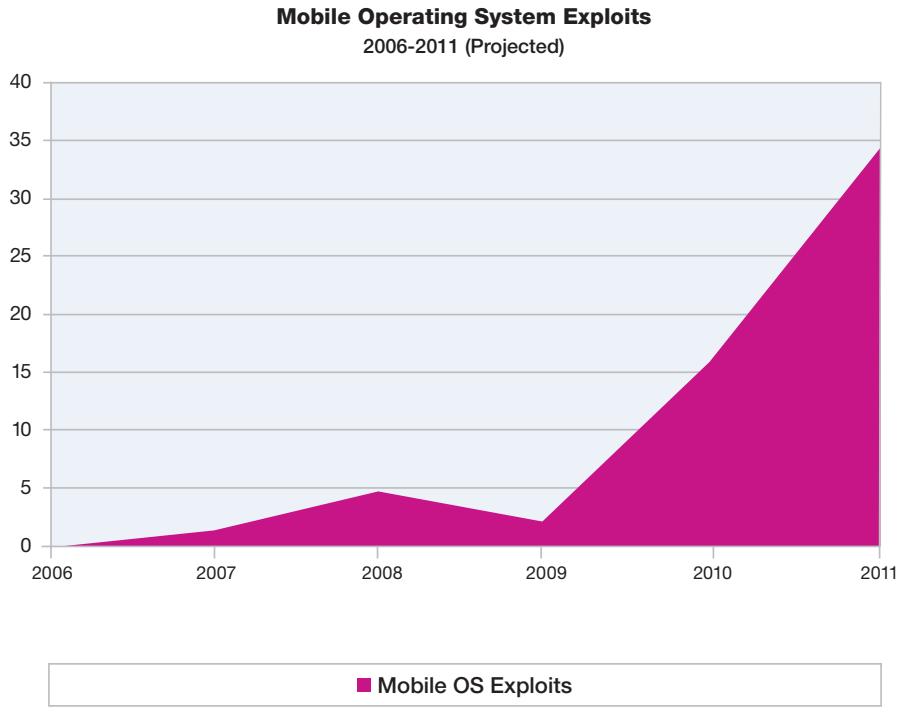


Figure 40: Mobile Operating System Exploits – 2006-2011 (Projected)

Exploit effort versus potential reward matrix

Every day, as X-Force tracks vulnerability disclosures, we make decisions regarding which vulnerabilities require deeper investigation with an eye toward product coverage and which vulnerabilities are less likely to be exploited in the wild. The exploitation probability matrix provides an abstraction of the thought process we apply in making these choices. It functions by attempting to chart the opportunity that each vulnerability represents to attackers from a financial perspective. On the X axis we chart the cost associated with exploiting a vulnerability and leveraging it to commit computer crime. Vulnerabilities that fit readily into an existing process that attackers have for breaking into computer systems score high on this dimension. Vulnerabilities that are hard to exploit or which require attackers to develop new processes around them score low. On the Y axis we chart the overall opportunity that a vulnerability represents to attackers who do exploit it—how many systems out on the Internet are vulnerable and what kind of data to they host? How much value can be extracted out of exploiting this vulnerability?

A chart of these two axes breaks out into four quadrants. The first quadrant represents vulnerabilities that are relatively inexpensive to exploit and represent a large opportunity to attackers. These are exactly the sort of vulnerabilities that are likely to see widespread exploitation in the wild. The second quadrant represents vulnerabilities that are high value but harder to exploit—cases which might be targeted by more sophisticated attackers but are less likely to see widespread exploitation. The third quadrant represents low value, high cost vulnerabilities that are unlikely to be targeted widely. The fourth quadrant represents lower value, lower cost vulnerabilities which are sometimes targeted if it is sufficiently easy for attackers to do so.

Figure 41: Exploit Effort vs. Potential Reward – 2011 H1

In the first half of 2011 X-Force published 24 vulnerability **alerts and advisories**, highlighting the most critical vulnerabilities disclosed during this timeframe from our perspective. We place 12 of these vulnerabilities in the first quadrant—high value vulnerabilities that are cheap to exploit. In fact, there are publicly available exploits for nine of these 12. Almost all of these vulnerabilities represent client software remote code execution vulnerabilities that are exploitable by malicious web servers through the browser or the browser environment. These vulnerabilities directly fit the drive-by-download approach of attracting victims to malicious websites that has been the pattern of a great deal of attack activity in the past few years. One interesting exception is a client-side cross-site scripting vulnerability in Internet Explorer (**CVE-2011-0096**) which could be used to steal cookies needed to access secure websites even if those websites themselves do not have an inherent cross-site scripting vulnerability.

We place nine vulnerabilities in the second quadrant—harder to exploit but high value. This percentage is unusually high versus previous publications of the exploitation probability matrix. Two of these vulnerabilities are Adobe Shockwave vulnerabilities (**CVE-2010-4306** and **CVE-2010-4307**)-exploitable with a malicious website, which were discovered by X-Force Researchers. Hopefully our early discovery and coordination of these vulnerabilities with the vendor has discouraged attackers from targeting them. Four of the nine are in this category because they involve setting up a malicious server that is not a web server (such as a DNS server or file server). It can be relatively easy to get victims to access various kinds of malicious servers by embedding requests in HTML content that the victims are accessing. This is a little bit more complicated than merely hosting malicious content on a web server and most attack activity seems to favor that model instead.

The other three vulnerabilities in the second quadrant are serious, server-side remote code execution vulnerabilities that could have been exploited by self-propagating Internet worms. We have not seen public exploits emerge for any of these three vulnerabilities, in spite of their extremely high value. We know of one private exploit for one of these vulnerabilities that was developed by a couple of former X-Force Researchers, but it has not been disseminated publicly, and the technical difficulty associated with achieving code execution in this case was relatively high. (See Chris Valasek's paper **Understanding the Low Fragmentation Heap** from Blackhat 2010.) Over the past few years, developers of memory management systems have created a wide array of exploitation prevention features that protect serious vulnerabilities from being exploited. Although some of the world's best vulnerability researchers have found ways to cut through these protection mechanisms, the fact that extremely high value vulnerabilities like these can now sit for months without exploits emerging publicly likely means that these memory management security features are having a measurable impact on the security of the Internet.

We place three vulnerabilities in the fourth quadrant, vulnerabilities that are easier to exploit but lower value. All of these are denial-of-service vulnerabilities in Internet services. Denial of service attacks can have a serious impact on network operations, so it is appropriate to publish alerts about these issues. However, they are obviously less valuable to attackers than remote code execution vulnerabilities, and so we rate them appropriately. In fact, we find that most denial of service activity on the Internet involves distributed traffic flooding rather than the exploitation of specific vulnerabilities.

Endpoint management: continuous patch compliance and visibility

Malware attacks continue to rapidly exploit vulnerable computer systems before the patches are published by software vendors and applied by customers. These attacks can cause loss of organizational productivity, risk of sensitive data loss, and potential litigation and fines, costing the U.S. economy an estimated $266 billion annually, according to the Cyber Secure Institute, a Washington, D.C.-based advocacy group.

Although vendors typically are diligent in providing patches and are issuing more and more patches to address vulnerabilities, most organizations take weeks or even months to deploy them throughout the environment. According to some estimates, it can take organizations as long as four months to achieve a 90 to 95 percent patch compliance rate.

Despite the risks, some organizations are slow to patch due to massive complexity and the barriers they face when implementing effective patch management practices. Time and labor involved, lack of visibility, potential business impact, network bandwidth limitations, lack of manageability, long remediation times, scalability issues, heterogeneous environments, and roaming endpoints are just some of the hurdles they may encounter.

With software and the threats against that software constantly evolving, organizations should have an effective way to assess, deploy, and manage a constant flow of patches for the myriad operating systems and applications in their heterogeneous environments.

Changing the patch management paradigm

While there is no single, official patch management best practice, the general approach involves a closed-loop process with six basic steps: research, assess, remediate, confirm, enforce, and report. Historically, many of these steps were implemented via separate, non-integrated technologies, making it virtually impossible to create a closed-loop, real-time patch management process.

Step 1: Research

The first step in the patch management process involves discovering which patches are available and then evaluating and testing them for compatibility within the organizational environment. If not automated, this process can consume a significant amount of time and resources. Accepting automated vendor updates without testing them can put organizations at huge risk, since there is no enterprise control over timing or reporting. Relying on users to apply updates can be risky and unreliable.

Patch management solutions that automate acquiring, testing, and distributing patches from operating system, anti-malware and common third-party application vendors directly to customers can remove considerable patch management research overhead. When new patches are received, they should be analyzed and deployed according to highly granular policies which contain information such as patch dependencies, applicable systems, and severity level. Deployment should be targeted to specific machine profiles, so that specific patches are sent only to the endpoints that need them.

Step 2: Assess

For each identified patch, the IT organization should determine the applicability and criticality of the update. Since many patches are time critical, and the process of risk assessment and patch prioritization

should take place as quickly as possible, it is important for organizations to have access to the complete, current asset and configuration data set to quantify the scope and impact of patches across the organization. There are tools that can help acquire the data but may take days or weeks to collect the data after scanning every endpoint.

Continuous monitoring of the endpoints and reporting on their status, configuration, and compliance state with the policies defined such as mandatory patch levels and standard configurations is highly recommended. This information is especially critical during emergency patch scenarios when a vendor releases high-priority patches delivered outside regular release schedules and organizations are required to respond rapidly.

Once the total number of patches is mapped to the endpoints that need them, and the business criticality is defined, the IT organization can proceed to the remediation step.

Step 3: Remediate
Remediation is often difficult to accomplish quickly on an organizational scale due to many reasons. Some of these reasons include determining patch prerequisites, ensuring that the patch is safe, network capacity, and inadvertently skipping certain endpoints, such as those not currently connected to the corporate network. All of these factors can result in low first-pass patch rates.

Additionally, many tools do not provide the fine-grained, policy-based controls to effectively deploy patches to all affected endpoints. Controls such as patch installation time windows, whether or not a user must be present, reboot options, method of distribution, system type, and user notification options should be available inputs into the automated update process.

A comprehensive report can help organizations determine which endpoints need the update. Operators can then determine when the patch should go out, what notification to display to end users (if any), whether or not to allow users to delay a patch implementation and for how long, and whether to force (or delay) reboots.

Once it is determined that the patch is applicable to a particular endpoint, it can be downloaded and applied while reporting back success or failure. With endpoint management solutions that can easily reach Internet connected devices, network load can be significantly reduced and first-pass success rates can be improved to 95+ percent.

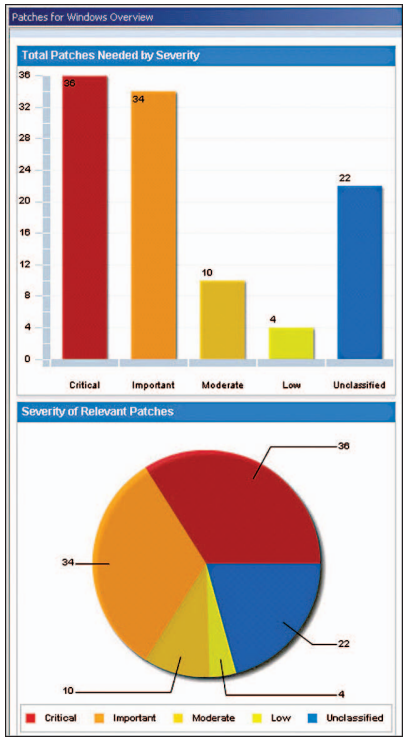
To help ensure greater security, you should employ only cryptographic identities. This helps ensure that only authorized administrators can create and distribute patches.

Step 4: Confirm
After patches are applied, successful installation on all endpoints should be confirmed so that IT knows when the patch cycle is complete. This data should be communicated back to a central reporting system that updates personnel on the process, including exceptions, in real time. This step is critical in supporting compliance requirements, which require definitive proof of patch installation and for closing the loop on patch management.

Step 5: Enforce
If a patch is uninstalled for any reason (users intentionally or accidentally uninstall patches, new applications or patches corrupting existing updates, malware deliberately removing patches), the policy can specify that the patch should be automatically reapplied to the endpoint as needed. In the event of problems with a patch, administrators should quickly and easily issue a rollback to endpoints. If the endpoint compliance status is reported in real time, IT administrators can easily control and monitor the state of all managed endpoints.

Step 6: Report

Reporting is a critical component of the patch management process. Compliance and corporate policies often require highly detailed, up-to-date dashboards and reports that indicate the organization's risk position and patch management status for a variety of consumers, including compliance auditors, executives, management and even end users.



Reports showing patch management progress in real time

IBM CIO deployment of patch management solution

IBM needs to protect its internal infrastructure, which covers over 425,000 employees, just like any other organization today. At the same time, IBM's evolving business models have increased the challenge of maintaining IBM endpoints and infrastructure, and the number of nonstandard IBM endpoints has increased significantly.

IBM has begun a worldwide internal deployment of Tivoli® Endpoint Manager. At the time that this document went to publication, IBM had deployed Tivoli Endpoint Manager to over 550,000 endpoints within six months, out of a total of 750,000 Windows, Mac, and Linux endpoints targeted for deployment by the end of 2011. IBM has estimated it can reduce workstation security issues by 50 percent within the first year, an estimated US$10 million in cost savings.

Summary

Traditional patch management approaches that use manual processes and cumbersome scan- and poll-based mechanisms generally are no longer fast or cost-effective enough to meet business and regulatory requirements, leaving organizations with unacceptably high risk and costs. When vendors introduce regular patch release cycles, attackers get the opportunity to exploit un-patched endpoints without having to work to uncover new vulnerabilities.

Effective solutions that automate patch management tasks and support a closed-loop process can help organizations drastically increase patch success rates, improve regulatory compliance, and reduce expenditures.

User access and the insider threat

As more and more business applications embrace the Web as the preferred platform and as the users on the web grow by 35-45 percent year over year, security risks exposed by web applications can assume hitherto unforeseen proportions in today's world. While high risk threats in the form of SQL injection and cross-site scripting lead Web Application vulnerabilities, some studies associate vulnerabilities common to insider threats just as risky. Insider based vulnerabilities include 'Broken Authentication and Session Management' related vulnerabilities, especially since interactions and computer sharing between insiders (employees) are likely to be much more prevalent than external users. These are cases where users or anonymous parties leverage flaws in the user authentication or session management capabilities of target applications to impersonate genuine users and steal or modify business critical data.

Primary cause

Many organizations tend to have web applications with custom authentication and session management capabilities built as internal functions by application developers themselves. These application developers are often not specialized in security solutions and hence the security and access modules in these applications could contain flaws/ vulnerabilities in critical areas such as password management, security policy management, timeouts, and session management. Since frameworks, methods, and technologies used in security implementations often differ from application to application, testing and code reviews frequently fail to catch many vulnerabilities associated with them.

Typical attack scenario

Consider an application that does not have appropriate session timeout capability. If a user uses a shared computer to access the application and leaves it without proper logout, subsequent users may be able to use the session to perform additional transactions. This gets compounded if the session remains open for long periods in the absence of an appropriate session-inactivity timeout.

Consider an online application which supports URL re-writing thereby including session ID in the URL. If a user, after a specific transaction, shares the link, the recipient could misuse the session ID for carrying out additional transactions.

Typical solutions adopted by enterprises

Considering the seriousness of the Broken Authentication and Session Management related vulnerabilities and the enormity of associated business losses, today's organizations assign great importance to enabling sound security and access control mechanisms for their web applications.

Despite this, web application vulnerability category climbed four positions in the Open Web Application Security Project (OWASP) 2010 top ten vulnerability report compared to its previous release.

A commonly used approach to combat this is to externalize access and session management functions from web applications and use a specialized solution to handle these functions for all web applications used by the organization. These specialized web access solutions would be built by security experts or acquired from competent vendors. Additional capabilities such as centralized gate-keeping and Single Sign-On help to reduce Insider threats.

Section III
Developing Secure Software

In the Developing Secure Software section of the report, data is presented on processes and techniques for creating secure software. We discuss how enterprises can find existing vulnerabilities and help prevent new ones from being introduced. If you use networked or web applications to collect or exchange sensitive data, your job as a security professional is harder now than ever before. We take a look at both the static and dynamic security testing done by the Rational® AppScan® group in all stages of application development and share insights on what was discovered.

Further details on hybrid analysis of client-side JavaScript code

In the first half of 2011, IBM's Rational Application Security Group continued to enhance, evolve, and focus its research into the prevalence of client-side JavaScript vulnerabilities in Web 2.0 applications. This research is based on a unique technology called **JavaScript Security Analyzer (JSA)**, which is available as a part of **IBM Rational AppScan Standard Edition**. JSA performs hybrid analysis of client-side code, by applying static taint analysis on JavaScript and HTML code collected from web pages and extracted by an automated deep web crawl process (dynamic analysis).

Similar to previous research, we used a sample set of 678 websites, including the Fortune 500 websites, and a list of 178 most popular websites such as social networks and media sites. The research used a newer version of the JSA technology, which includes superior analysis algorithms and enhancement to reduce susceptibility to noisy results. This can result in a lower false positive rate. The outcome of the research showed a dramatic increase in the amount of vulnerabilities that could be detected.

Performing manual code review to modern JavaScript is not a simple task!

Since this research was conducted on the same website data as the previous research, the difference in results can be attributed to higher accuracy in the enhanced JSA algorithms. Table 10 shows that our current research found that out of the 678 websites, 40 percent (271 sites) contain client-side JavaScript vulnerabilities.

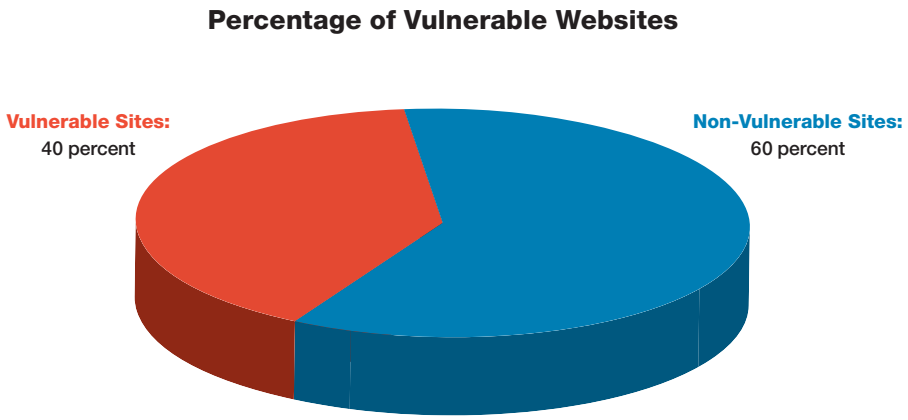


Figure 42: Percentage of Vulnerable websites

Total Sites Scanned	678
Total Issues Found	3683
Vulnerable sites (number)	**271**
Vulnerable sites	**40.0 percent**
Non-vulnerable sites	**60.0 percent**
Non-vulnerable sites (number)	**407**
Applications with issues in 3rd party code	90 percent
Applications with issues only in in-house code	10 percent
Sites vulnerable to DOM-based XSS	252
Total DOM-based XSS issues found	3214
Sites vulnerable to Open Redirect	226
Total Open Redirect issues found	266
Sites vulnerable to DOM-based email Attribute spoofing	5
Total DOM-based email Attribute Spoofing issues found	203

Table 10: Breakdown of total sites scanned

Issue Types	Sites Vulnerable	Total Issues
DOM-based XSS	252	3214
Open Redirect	226	266
DOM-based email Attribute Spoofing	5	203
Total Issues found		**3683**

Table 11: Overview of total issues discovered

In figure 43, we see that, out of the vulnerable applications, 90 percent* included one or more vulnerabilities that were introduced through third-party JavaScript code, such as marketing campaigns, code that embeds Flash animation, and AJAX libraries.

*** Note:** The statistics in figure 43 (Percentage of sites that include a vulnerability in third-party client-side code) are different than the statistics presented in our earlier research results, which counted the amount of third-party issues out of the total issue types.

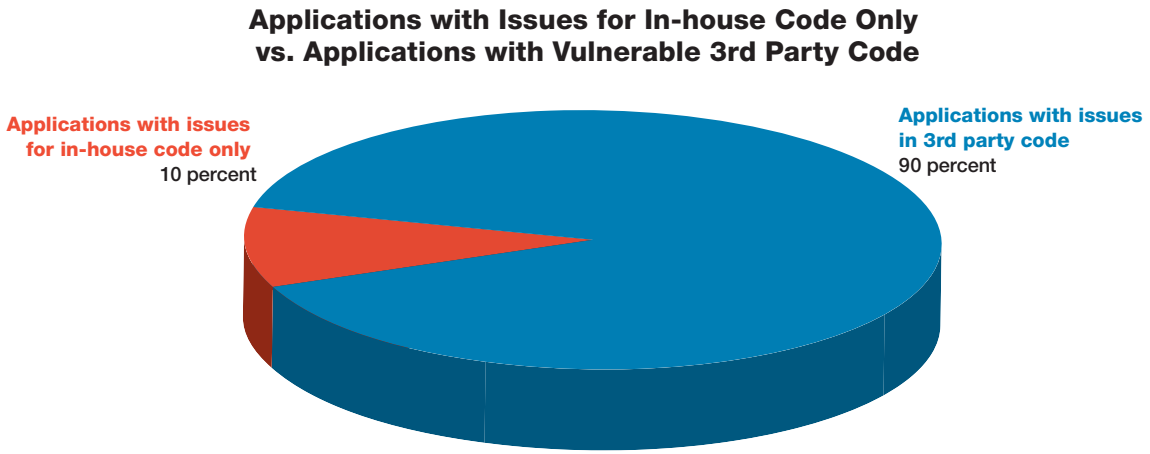


Figure 43: Applications with Issues for In-house Code Only vs. Applications with Vulnerable 3rd Party Code

Figure 44 shows that DOM-based cross-site scripting (3,214 issues out of 3,683) is still the #1 most common security issue type. It appears that the number of sites vulnerable to DOM-based cross-site (252) scripting and sites vulnerable to client-side open redirect (226) are pretty similar.

In addition, a new type of vulnerability was detected for the first time: DOM-based email Attribute Spoofing. This vulnerability occurs when a web application uses JavaScript code to automatically craft an email for the user to fill and send, using user-controlled data. In such scenarios, an attacker could potentially manipulate the content, subject, or CC and BCC fields of the email, resulting in a leak of private information.

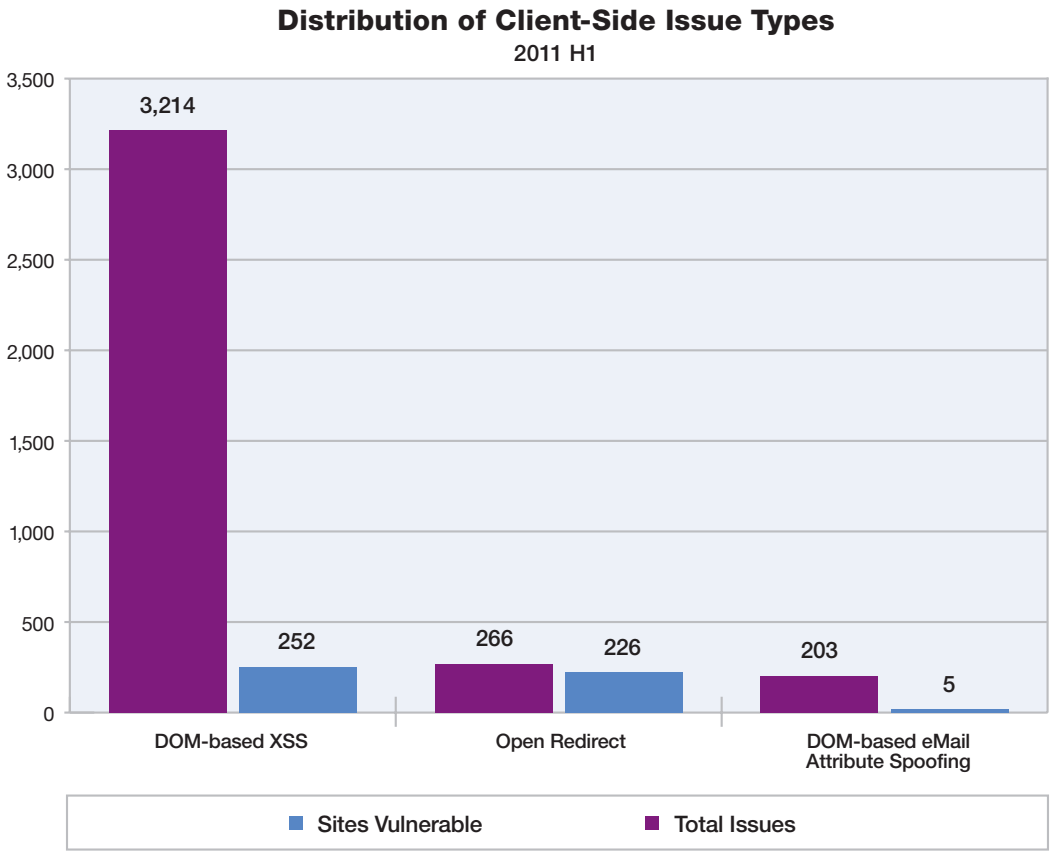


Figure 44: Distribution of Client-Side Issue Types – 2011 H1

Section IV
Emerging Trends in Security

The Emerging Trends in Security section takes a look at fast developing technology that presses upon enterprises considering whether or not it is time to make investments in these future areas. We explain where threats and exploits are being utilized in these early technology adoptions and how enterprises can stay focused.

Mobile malware

Mobile operating systems, such as Google's Android, are becoming a popular target for malware authors. However, mobile malware is not a new phenomenon. The first known mobile phone malware is believed to be Cabir, which was discovered in 2004 and affected phones running SymbianOS. The Android OS made its debut in 2008, and the first known Android malware was discovered in 2010. That malware, dubbed FakePlayer, caused infected phones to send SMS messages that would charge the user money. In 2011, the DroidDream malware was the first wide-spread infection that was hosted on Google's own application market. As smartphones become more ubiquitous, we expect the mobile malware threat to increase.

Mobile devices as a malware platform

Mobile phones are an attractive platform for malware developers for a number of reasons. First, it's easy to monetize a mobile phone infection. Malware distributors can set up premium SMS services that charge users that send an SMS message to a specific number. Most of the malware we see on Android and other mobile platforms takes advantage of this and sends SMS messages from the infected phone. Second, mobile phones often have unpatched vulnerabilities. While security is a major focus for Android, several privilege escalation vulnerabilities have been discovered that can grant root access to a malicious application. Even when a vulnerability is found and patched, there are still many unpatched devices in the wild. Many mobile phone vendors don't push out security updates for their devices. Third, mobile phones are an attractive target because of the sheer size of the user base. In late June of 2011, Google claimed that there were 500,000 new Android device activations per day.

Android malware distribution model

One of the most popular and effective ways to distribute Android malware is through application markets. Besides Google's own official market, there are many unofficial third-party markets. There are a couple of different techniques malware authors use to convince people to download their applications. The first method, used by the DroidDream malware, is to create infected versions of existing market software. These infected versions are then uploaded with a very similar name to the original software, and users unwittingly download and install the infected version. Another trick to lure victims is to publish software that claims to be a crack, patch, or cheat for some other software. A malware family dubbed Plankton used this method; it was disguised as a cheat for the Angry Birds game.

Application markets aren't the only way to distribute Android malware though. We've seen infected applications on peer-to-peer networks, hosted on websites, and even on Usenet. These off-market applications are usually targeted at people looking for pirated versions of commercial Android applications.

Android malware capabilities

Once malware is installed on a mobile device, there is a substantial amount of damage it can do. A root exploit could be used to increase privileges and give the malware full access to the phone. Even without root access, the malware has the ability to do anything within the limits of the permissions that the user grants it.

When an Android application is installed by an end-user, the required permissions are displayed so the user can verify what the application does. For example, if an application needs to send SMS messages or read accounts stored on the phone, the user can decide to allow that before installation. If the user doesn't want to grant those permissions, the application is not installed. If a user is not careful about checking permissions, they could install a rogue application that requires more permissions than the original application needed—such as the ability to send SMS messages. The GoldDream malware, discovered by researchers at NC State University in July 2011, was distributed as trojanized versions of existing games on unofficial Chinese application markets. One particular example was

supposed to be a game called "Blood vs Zombie" that was actually a copy of an existing game, "Draw Slasher," but included a number of extra permissions that allowed it to steal user information. For a comparison of the permissions required by the legitimate and malware-infected game, see figures 45 and 46 below.

Besides sending SMS messages, Android malware has been observed collecting personal data from the phone and sending it back to a central server. This information could be used in phishing attacks or for identity theft. We have also seen Android malware that has the ability to be remotely controlled by a remote command and control server just like a bot that infects a Windows desktop machine.

As mobile platforms become more powerful, they begin to gain the features of a desktop PC. These features help make mobile platforms an even more attractive target for malware authors.

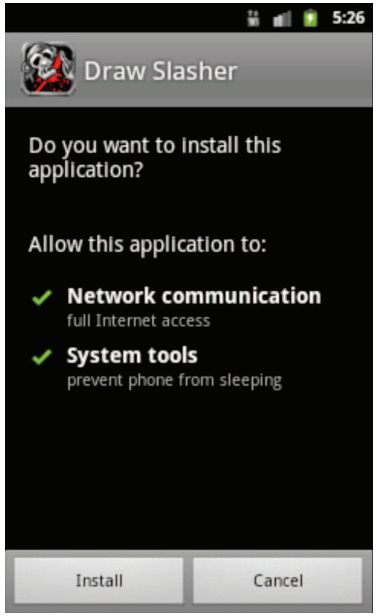


Figure 45: Draw Slasher, a legitimate game that requires minimal permission

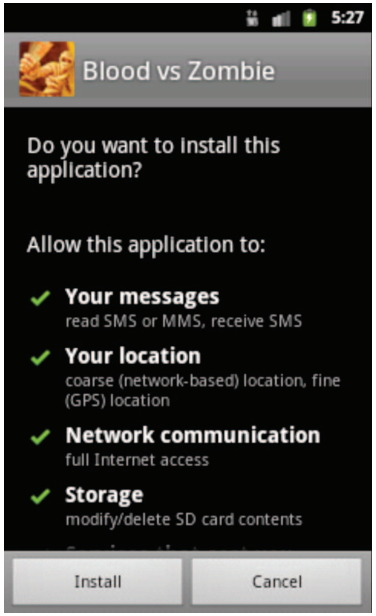


Figure 46: Blood vs Zombie, a malicious copy of Draw Slasher that contains more permissions than a game should need—including GPS and SMS access.

Protecting yourself from Android malware

It's possible to avoid infection with Android malware by using common sense when installing applications. First, stick with a reputable application market, such as the official Google Market or Amazon's Android application market. It's still possible for malware to sneak into the official market, so be wary of any applications that you install. Double check the permissions and ensure that you are comfortable with the level of access you are giving applications. A game should not require GPS or SMS access. Also, be careful of the type of software you install. Trying to get a free copy of a paid application is a sure way to get infected. Another tip is to only install applications that have a large number of installs (100,000 or more) with a high review rating.

Be sure to run security software on your mobile phone, regardless of what operating system it runs. Most major anti-virus software vendors have mobile versions of their products that can protect you from many types of malware.

Transformation in the enterprise with mobile endpoint devices

Over the past year or two, many enterprises have proceeded from traditional Blackberry-only, primarily or solely enterprise-liable smartphone programs to those where at least limited numbers of smartphones and tablets are running non-Blackberry operating systems with varying levels of enterprise access. In many cases, this progression has occurred in response to a combination of executive and employee needs, combined with realization that the trend is increasingly expensive to prevent rather than embrace. At least some of the mainstream consumer smartphone platforms have matured to include at least the minimal security and management functionality required to allow this progression into the enterprise.

This transformation trend varies significantly across various industry sectors. Enterprises with application requirements combined with application data requirements that extend to regulated data are still challenged by the security controls in some mobile operating systems. However, for enterprises with limited amounts of regulated data, the transformation is clearly underway.

With this trend, it has become increasingly important to strategically plan for increased expansion because the "normal" computing model is blurring the distinction among smartphones, tablets, and personal computers. While they may seem distinct and separate now, enterprise security management will struggle to handle the expansion with the convergence of endpoint security configuration management.

Today, in most enterprises, the security configuration management of personal computers (typically laptops and desktops with a sprinkling of Windows-based tablets) are completely separate and distinct from smartphones. This is primarily because it is based on the legacy of Blackberry Enterprise Servers as the required management platform for the existing mobile enterprise. While this worked in a simple endpoint model that included only PCs and Blackberries, this model likely will become unwieldy as more smartphone platforms are supported. Although many of the smartphone MDM solutions today are cross platform, there are still gaps as tablets enter use along with Blackberries. Additionally, most non-Blackberry smartphone programs are relatively limited and small in comparison to personal computer programs. If current trends continue, it will be common for most employees to also have a smartphone and tablet used whenever they are away from the desk. In this enterprise computing model, employees will have multiple devices with a mixture of funding liability across the spectrum.

This transformative trend should drive the importance of convergence as it applies to the management of endpoint computing devices. This likely will vary from enterprise to enterprise, depending on risk appetite, operating budget, and business sector. But for nearly all, this diversity and sprawl of management technology presents a challenge.

Endpoint security management convergence

Nearly all enterprises will be affected by some form of endpoint device differentiation, which will continue to drive the need for endpoint security management convergence.

Regardless of risk appetite, enterprises will increasingly pursue roles-based security management. The need for this within the enterprise varies depending on the diversity of roles within the company as well as the variance of data classifications from role to role. In typical larger enterprises, the most sensitive data is always limited to only those who require its access (least privilege) and this limitation itself can lead to varying security across roles. Ideally, this makes sense– why "over secure" some devices when not needed. Roles-based security is a primary driver toward the convergence of the security management of endpoints.

Derived from a set of roles, each with security requirements that are based on the data classifications associated with the role, this approach is much more easily applied to a myriad of devices in a converged environment. Without it, the replication of roles management is required across each of the corresponding smartphone, tablet, and personal computer management infrastructures. A proper roles-based management system could guide the boarding of employees to devices with corresponding security configuration and any needed security agents. This could help ensure that employees are not boarded to devices that do not support the minimum data protection requirements associated with their role. This becomes even more important in a personally-owned device scenario because some devices owned by employees may not be suited to support the required controls for the employee's role. In this case, employee and device should not be boarded to their enterprise data.

Cost and device management complexity are other driving forces toward the convergence of endpoint security management. While enterprises were typically willing to invest in different technologies to provide the secure use of smartphones, as this expands to include the majority of employees across the enterprise using increased numbers of platforms, the associated technology sprawl can become unaffordable. If not affordable, it certainly can become more costly than the selection of strategic, converged, cross-platform management technologies. Ideally, the enterprise can select the technologies that best integrate into their roles-based directory services, leveraging a roles-based approach.

Cost aside, the convergence of endpoint security management enables uniform enterprise risk management as well as increased audit readiness. The fewer tools, consoles, and reports required to be properly managed, the more likely it could result in efficiency. This is just common sense: technology sprawl and complexity make it harder to ensure everything is as it should be. Operationally, convergence simplifies the amount of effort and expertise required to keep it all running. The selection of a single converged tool should improve the enterprise opportunity to maintain deep, expert-level skills that lead to a best-of-class, audit-ready implementation.

A side benefit to all of this, is ease of off-boarding. When all devices from smartphones to tablets to personal computers are managed by a single solution, both boarding and off-boarding processes are simplified and are more likely to occur as intended. This extends to ease of integration for automation if the enterprise wants an automated off-boarding process.

The real differences among these computing devices is expected to continue to blur until they become a continuum of functionality from your pocket to your desktop.

Isolation/separation of enterprise and employee applications and data

Today, many enterprises allow use of employee-owned or financially liable devices (where the employee is financially liable for the device versus the enterprise.) This does not extend to all enterprises, particularly those in financial and healthcare sectors. The willingness to embrace employee devices contrasts with the organization's risk appetite but, rather than spend money at preventing incorrect usage, enterprises have moved forward to manage employee devices in a way that meets enterprise security requirements.

Typically, this has meant that enterprise security requirements have been applied to employee devices. Examples include the use of an enterprise-compliant password and wipe on failed access attempts. While initially few employees pushed back on this, there are use cases where enterprise management of personally owned devices in today's model can be overbearing. Let's face it, entering an alphanumeric password to pause music, or access a pedometer application presses the limits of what employees may be willing to live with in terms of enterprise security on their personal smartphone. While push back today comes from a minority of employees (as the consumerization model expands to include most employees) it will become a strategic necessity to clearly separate access from enterprise applications, associated data, and security control requirements from the employee's personal data and applications.

Today many financial and healthcare enterprises have avoided the support of employee-owned devices for multiple reasons. These reasons extend from the inability to achieve required controls on the devices to concerns related to data lifecycle management and to incident management. In some cases, enterprises have led the way by looking for virtualized mobile solutions so that enterprise data never ends up on devices. Remote virtualization solutions often have their own challenges such as the need for continued, robust connectivity.

Regardless of industry sector (and associated risk appetite), the area in which nearly all enterprises have the same goal is with solutions that allow enterprises to secure the enterprise applications and data, while allowing the employee to determine the security of personal applications and data. Approaches will likely vary from the use of secure encrypted containers in which enterprise data and applications reside to a variety of local and remote virtualization approaches. Virtualization may include the ability to run a virtual machine dedicated to the enterprise or the extension of today's virtual desktop or application streaming approaches to tablets and smartphones.

Because this is a wide-open solution space for smartphones and tablets, we should expect to see a variety of approaches. From the employee's use perspective, the underlying technology solution is immaterial; they simply want to determine the security controls for their personal mail, photos, music, video, and other content. While enterprises may look at this requirement to be more about employee satisfaction that helps enable consumerization (and enterprise cost savings), these approaches should also greatly improve the off-boarding process. These approaches should allow removal of enterprise data and applications without destroying employee files. The enterprise should not have to worry about deleting "once in a lifetime" photos that may reside on an employee device. This can be a win for both employee and enterprise.

Solutions that allow isolation and/or separation are still relatively limited and immature but maturation should occur rapidly, much as it did with mobile MDM solutions over the last three years. Enterprises should put thoughtful effort into developing the strategy of how they seek to achieve this separation. This can facilitate efficient identification of potential solutions to pilot and test.

Beating the breach: trends in database security and compliance

Life for security professionals used to be simpler. You could stop outsiders from accessing your data by establishing perimeter defenses such as firewalls and anti-virus systems, and by restricting physical access to the machines that process it. You could have on-site security guards and identity checks at the entrance to your corporate data center.

In today's interconnected world, that's no longer the case, because the boundaries of our business infrastructure are constantly being extended by the emergence of cloud, mobility, social business, big data, and more.

To be useful, a company's data should be continuously connected to its customers, business partners, and employees. That can expose sensitive data to more automated and targeted attacks than ever before. For example, we're now seeing numerous attacks that easily bypass traditional perimeter defenses by exploiting web application vulnerabilities such as SQL injection, or by leveraging stolen administrative credentials to compromise back-end databases. Despite more attention being paid to secure coding practices, SQL injection continues to be a favorite

attack vector amongst malicious groups as demonstrated by the numerous mass SQL injection attacks we have seen over the past several years.

Perimeter defenses are also ineffective against insiders such as disgruntled or rogue employees, because they are already "behind the firewall." In most current IT environments, privileged users such as DBAs, developers, and outsourced personnel have unfettered access to sensitive data, with little or no monitoring controls around their activities.

According to a major breach study, 92 percent of all compromised records are stolen from database servers[23]—far surpassing other sources of sensitive data leaks such as stolen laptops or data theft via email or USB drives.

It's no surprise that databases have become an important target for attackers. Critical data used to run our organizations—including financial/ERP, customer, employee, and intellectual property information such as new product designs—is stored in relational databases. Key enterprise applications such as SAP, PeopleSoft, Siebel, and Cognos all have commercial DBMS systems at their core.

The data security landscape

According to Forrester Research, over 75 percent of firms do not have a database security plan in place. Forrester also estimates that DBAs currently spend less than 5 percent of their time on database security.

Faced with these realities, many organizations are now seeing heightened C-level focus on tightening controls around application and database infrastructures. In our conversations with clients, we see five key drivers behind these initiatives:

Attackers are highly motivated to compromise databases with weak defenses, with crime syndicates willing to pay hard cash for personal information stolen from customer databases.

Cyber-espionage targets intellectual property (IP) such as new product designs, algorithms, strategic plans, and information about strategic resources such as oil, energy, and infrastructure.

Hacktivism is a phenomenon in which sites are attacked for political reasons rather than financial gain. State-sponsored cyber-attacks can also be used to support political goals such as gathering personal information to suppress internal dissent.

Insider threats are often considered the biggest threat because employees can easily exploit legitimate access to commit fraud, download large amounts of sensitive or proprietary data, or commit acts of vandalism such as inserting logic bombs in critical databases. The risk is especially high for "superusers" such as administrators.

Compliance requirements are constantly evolving and increasing in complexity, especially for global organizations. As a result, lowering costs by streamlining compliance processes is an important financial driver for implementing new database security technologies. In particular, many organizations are now looking to replace their current ad hoc collections of manual compliance processes with a single set of centralized, standardized, and automated controls for all of their key applications and compliance mandates.

23 2010 Data Breach Investigations Report by Verizon Business.

Ten best practices for database security and compliance

Based on our engagements with Global 1000 organizations, the following best practices have emerged for strengthening database security and compliance in enterprise environments:

1. Discover. Data can't be secured if you don't know that it exists in the first place. It's important to discover the locations of sensitive data and have a good mapping of sensitive assets. This includes rogue database instances, sensitive data inside databases, and relationships between data elements that can make them more sensitive (such as an association between "Last Name" and national insurance number.)

Also, don't forget about non-regulated data such as corporate intellectual property (IP) including strategic plans, product designs, algorithms, and M&A analyses that may be of interest to attackers. Finally, automate the discovery process and execute it on a regular basis because the location of sensitive data is constantly changing.

2. Assess vulnerabilities and configurations. It's important to assess database configurations to help ensure that they don't have security holes. There are several best practices checklists for accomplishing this, such as the CIS Database Server Benchmarks and the Security Technical Implementation Guides (STIGs) for databases developed by the U.S. Defense Information Services Agency (DISA).

This process includes verifying both the way the database is installed on the operating system (for example, checking file privileges for database configuration files and executables) and configuring options within the database itself (such as how many failed logins result in a locked account or checking permissions for various roles in the database itself). Note that these database-specific assessments are typically not performed by traditional network vulnerability assessment solutions. Also, verify that outmoded database versions with known vulnerabilities are not being run.

3. Harden the database. The result of a vulnerability assessment is often a set of specific configuration recommendations to take as next steps. Other elements of hardening involve removing all database functions and options that you do not use.

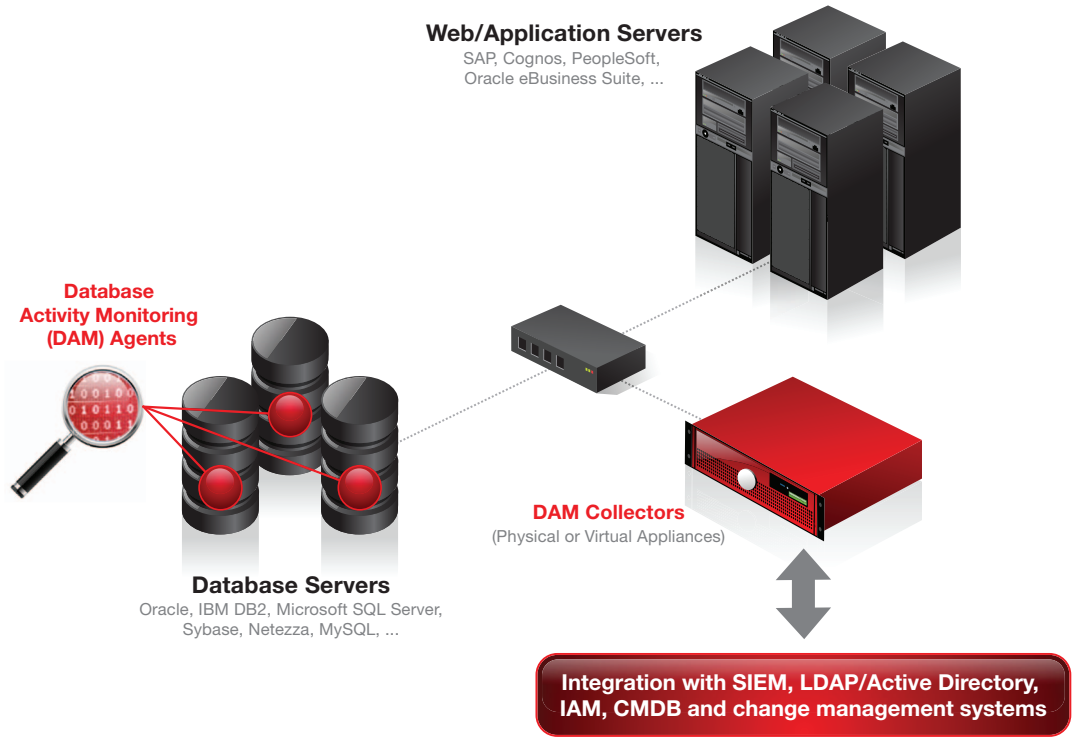
4. Audit configuration changes. Once the hardened configuration is established, it's important to continually track it to help ensure the "gold" configuration hasn't changed. This can be done with change auditing tools that compare snapshots of the configurations (at both the operating system level and at the database level) and then immediately alert whenever a configuration change is made that could affect your security posture.

5. Deploy Database Activity Monitoring (DAM) and Database Auditing. Continuous, real-time monitoring (see diagram below) is crucial for rapidly detecting suspicious or unauthorized activity—such as a customer service rep downloading hundreds of sensitive data records in a single day—and limiting exposure to attacks and misuse.

This is important because, according to a recent survey of database professionals, 75 percent of organizations can't prevent privileged users from reading or tampering with data in their databases, and close to half said an end-user with common desktop or ad hoc tools either could gain unauthorized direct access to sensitive information (or they weren't sure about it).

Monitoring privileged users is also important for detecting intrusions from outside attackers, since cyber attacks such as SQL injection frequently result in the attacker gaining control of privileged accounts. DAM is also an essential element of vulnerability assessment, because it goes beyond traditional static assessments to include dynamic or "behavioral vulnerabilities" such as users sharing generic service accounts and other privileged credentials.

Database auditing allows organizations to generate a secure, non-repudiable audit trail for all database activities that impact security posture (such as creation of new accounts), data integrity (such as



Database Activity Monitoring (DAM) technologies continuously monitor and audit all database traffic in order to rapidly identify suspicious or unauthorized activities at the database tier.

DAM Agents reside on database servers and are used to capture all database traffic, including activities by privileged users such as DBAs, developers and outsourced personnel.

DAM Collectors are used to evaluate database security policies in real-time, store a secure audit trail of captured traffic, perform forensics on audit data, and generate compliance reports, security exception reports and real-time alerts.

DAM solutions can also provide related capabilities such as blocking, database vulnerability assessment (e.g., identifying unpatched systems, misconfigured privileges and default accounts), sensitive data discovery, configuration auditing, entitlement reporting, and application layer monitoring to identify end-user fraud in multi-tier enterprise applications such as SAP and PeopleSoft.

Figure 47: Deploy Database Activity Monitoring (DAM) and Database Auditing.

changing financial data values or schemas), or data privacy and confidentiality (such as viewing of Personally Identifiable Information or PII). In addition to being a key compliance requirement, granular audit trails are important for forensic investigations.

6. Authenticate, control access and manage entitlements. Authenticating, controlling access, and managing entitlements is essential to helping ensure full accountability and managing privileges to limit access to data. These privileges should be enforced, even for the most privileged database users. It's also recommended that you periodically review entitlement reports (also called User Right Attestation reports) as part of a formal audit process.

7. Monitor the application layer. Well-designed DAM solutions can associate specific database transactions performed by the application with specific end-user accounts, in order to deterministically identify individuals that are violating corporate policies.

In addition, combining database auditing information with traditional logs from other applications and systems (such as Windows, UNIX/Linux, and firewalls) via a Security Information and Event Management (SIEM) system to see everything that a user has done, can also provide critical information for forensic investigations.

8. Encrypt. Use encryption to render sensitive data unreadable, so an attacker cannot gain unauthorized access to data from outside the database. This is most easily accomplished by encrypting data at the file level, via the operating system, in order to avoid costly and time-consuming application changes required for field-level encryption at the database layer. File-level encryption, when combined with granular real-time monitoring and access control at the database layer, is typically accepted as a practical alternative to column-level encryption and a compensating control for Requirement 3.3 of PCI-DSS.

9. Mask test data. According to a recent industry survey, close to two out of five of organizations ship live production data to development teams and outside parties. Masking is a key database security technology that de-identifies production data, replacing it with realistic but fictional data that can then be used for testing, training, and development purposes, because it is contextually appropriate to the production data it has replaced.

10. Automate and standardize compliance processes. Laws and regulations can require implementation of data security measures and provisions to help reduce risks and vulnerabilities to a reasonable and appropriate level. Achieving compliance is not only important because no one likes to fail an audit, but it also provides third-party validation that your organization has implemented the proper controls to help ensure the confidentiality, integrity, and availability of your data. Automating and standardizing compliance processes is essential for helping to reduce compliance costs, minimize last-minute audit fire drills in your organization, and address ever-changing regulations.

Why existing security technologies are insufficient

Traditional security technologies are essential building blocks for a layered defense, but unlike database-specific technologies like DAM, they weren't designed with embedded knowledge about database protocols, structures, activity patterns and context that would allow them to easily identify unauthorized or suspicious database activities. Specific examples follow.

Firewalls, network IDS/IPS and Web Application Firewalls (WAFs) have limited understanding of database constructs and SQL commands. Even WAFs only require understanding of 10 or so HTTP constructs, while analyzing database traffic requires a detailed understanding of more than 350 SQL commands as well as that of a full programming language (PL-SQL).For example, these technologies weren't designed to identify a rogue DBA that is using DDL (Data Definition Language) commands to modify database schemas, which is an important requirement for SOX compliance. Similarly, they were not designed to identify an attacker that has compromised the application server to gain access to the database server, or is using stolen credentials in order to read the entire contents of a sensitive database.

In addition, traditional network security systems typically are not designed to handle the massive amounts of database audit data generated by enterprise applications such as Oracle EBS, PeopleSoft, and SAP. This requires a scalable architecture for efficient collection, storage, and analysis of database audit data, including automated, enterprise-wide aggregation of data across multiple servers and locations. Traditional network IDS devices—which are optimized for network packet monitoring rather than audit logging—can be ineffective for continuously monitoring and auditing database environments in real-time, because this requires intelligent storage algorithms and advanced relational database tools to extract the critical information required for auditors and forensic analysis.

Native DBMS audit logging facilities, triggers, and other DBMS-resident approaches, combined with home-grown scripts to analyze audit data, are often the first avenue for organizations looking to monitor database activities, but they can suffer from several significant disadvantages, primarily because native DBMS logging was originally developed for performance tuning and recovery purposes rather than for security and compliance.

The principal drawback is that native audit logging facilities are controlled by the same DBA teams that auditors are looking to monitor, thereby creating an important separation of duties (SoD) issue. As a result, DBAs and other privileged users can disable logging or tamper with audit logs in order to "cover their tracks." Similarly, attackers that compromise databases via SQL injection or stolen credentials typically gain super user privileges allowing them to disable logging (known as anti-forensics).

A second drawback is that native audit logging facilities impose a high level of performance overhead on database systems, particularly when used to capture a high volume of activities, such as capturing all access to sensitive data (as required by PCI-DSS requirement 10, for example).

Third, multi-tier enterprise applications such as SAP and PeopleSoft use generic service accounts to access the database layer, thereby concealing the identity of application end-users that initiate transactions at the application layer. As a result, native database audit logging technologies may not be sufficient to detect end-user fraud and other suspicious actions performed by authorized end-users, because they associate all database transactions with the generic service account rather than with specific application IDs.

Fourth, homegrown script-based approaches that rely on DBMS-resident audit functions are difficult to develop and maintain in distributed heterogeneous DBMS environments, because each DBMS platform performs audit logging in a different way. This can lead to unique, siloed tools and processes for each DBMS environment, with inconsistent audit policies and reporting. It can also make it much more difficult to create enterprise-wide views of audit information for database compliance, analytics, and forensics.

Finally, native audit logging facilities are "after-the-fact" detective controls that don't provide proactive, real-time preventive controls such as alerts or blocking.

Security Information and Event Management (SIEM) relies on collecting native DBMS audit logs and therefore can suffer from the same disadvantages described above (lack of separation of duties, overhead, etc.). The same can be said for other solutions that rely on collecting native audit logs, such as audit data vault or audit repository solutions. Finally, SIEM systems typically don't provide any real-time protection, and lack database-focused analytics and reporting.

Data Leak Prevention (DLP) technologies are an important element of a defense-in-depth strategy, but they aren't typically used to protect sensitive data at the source—that is, in the data center—which is the focus of most attacks. Instead, these technologies are designed to catch sensitive data as it leaves the network perimeter via email, or as it exits endpoints via USB drives—after it's been extracted from sensitive databases. For example, because they don't monitor database access activity, DLP technologies would not be used to identify an analyst who just launched a SQL query to access 1,000 records from the company's document database server, CRM system, payment card processing system, or CAD system.

Database encryption is an important technology for protecting database files and media (such as backup tapes) from theft or snooping, but it doesn't provide monitoring capabilities to identify or prevent unauthorized activities by authorized users. In addition, it can't protect against attackers that hijack application servers to gain encrypted access to back-end databases, nor can it defend against administrators and developers with access to encryption keys. Database encryption also isn't effective as a granular access control mechanism because it can take years to modify existing application architectures to support field-level encryption at the database tier (for example, to address the performance impact of encrypted indexed fields).

Overview of database security technologies

Over the last few years, the security industry has responded with new technologies designed specifically to address database security and compliance challenges. These new technologies address the limitations of existing security solutions described above, by providing the following features.

Database Activity Monitoring (DAM) Continuous, real-time monitoring and auditing of all database activities, including creating a granular audit trail of all activities by privileged users or all access to sensitive tables, with minimal impact on performance. Policy-based rules are used to rapidly identify unauthorized or suspicious activities, across heterogeneous DBMS environments (Oracle, DB2®, and SQL Server), with real-time alerts and exception reports. Best-of-breed solutions are based on scalable, multi-tier architectures with centralized policy management as well as centralized aggregation of audit data for enterprise-wide compliance reporting, analytics, and forensics.

Configuration and Vulnerability Assessment (VA) Libraries of automated tests to find database-specific vulnerabilities such as default vendor passwords, misconfigured privileges and roles, unprotected database configuration files, and missing patches.

Change and Configuration Auditing Identify critical changes to databases such as schema changes, as well as configuration changes that can impact security posture such as changes to database configuration files and permissions, registry variables, environment variables, and scripts.

Discovery Automated database discovery to identify new or rogue databases, combined with data discovery and classification technology to locate sensitive data in their databases such as credit card numbers and social security numbers.

Blocking and Fine-Grained Access Control Policy-based blocking of unauthorized database activities, typically used to block transactions by privileged users such as outsourced DBAs. Rules can be based on incoming queries (users, activity

performed, database objects, time of day, location, or source application), as well as outgoing result sets such as an abnormally high number of sensitive records being returned to the client.

Compliance Workflow Automation Auditors want to know that organizations aren't simply generating database access reports, but that they also have a formal oversight process for addressing exceptions and violations of corporate policies. Best-of-breed DAM solutions automate the compliance oversight process, including report distribution, electronic sign-offs, comments, and escalations. This is usually combined with libraries of best practices compliance reports and policies.

Masking De-identifies sensitive data for use in test and development environments.

Encryption Encrypts database files and media such as backup tapes to prevent theft or snooping of sensitive data.

Integration with Existing IT Infrastructures Organizations are looking for broad heterogeneous DBMS support (Oracle, SQL Server, DB2, Sybase, Informix, MySQL, Teradata, Netezza, and PostgreSQL) on all key OS platforms (Linux/UNIX, Windows, and z/OS) as well as integration with key infrastructure components such as LDAP/Active Directory, SIEMs, CMDBs, change ticketing systems, and so on.

Data security, virtualization and the cloud

Despite the agility, scaling, and cost benefits of moving to the cloud, many organizations are hesitant to adopt cloud computing services, often citing data security as a concern. However, certain cloud deployment models such as Private Clouds, as well as cloud service models such as Infrastructure as a Service (IaaS), allow organizations to ensure higher levels of data security by providing it themselves. The same is true for organizations with virtualized infrastructures.

Here are some questions to ask when considering migrating existing data security technologies to virtualized infrastructures or the cloud.

Monitoring Approach Does the solution provide software-based monitoring that automatically moves with the database instance as it migrates within the cloud or virtualized infrastructure, or does it rely on physical access to traditional network resources such as SPAN ports or TAPs? These resources are typically not available in virtual or cloud environments because communication occurs over hardware backplanes rather than over traditional networks.

Virtual appliances Can the solution be deployed as virtualized appliances, or does it rely on hardware appliances that live "outside" the cloud or virtual environment?

Web-based Management: Can the solution be managed from any device using a standard browser?

Single solution for both physical and virtual or cloud infrastructures: Can you use a single solution that spans both environments, or is it dedicated to one or the other?

IBM Corporation
Software Group
Route 100
Somers, NY 10589 U.S.A.

Produced in the United States of America
September 2011

WGL03009-USEN-00