IBM X-Force 2011 Trend and Risk Report

March 2012

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 contribution to the publication of this report.

Contributor	Title
Bryan Casey	Market Manager, IBM Security Systems
Carsten Hagemann	X-Force Software Engineer, Content Security
Colin Bell	Security Solution Architect, Lab Services & Support, IBM Security Systems
Clay Blankenship	Senior Incident Response Analyst
Cynthia Schneider	Information Developer
David McMillen	Security Intelligence Analyst, IBM Security Services
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
Gina Stefanelli	X-Force Marketing Manager
Jason Kravitz	Techline Specialist for IBM Security Systems and E-Config
John C. Pierce	Threat Intelligence Analyst, AI, MSS
John Kuhn	Security Intelligence Analyst, IBM Security Services
Kimberly Madia	Data Security Strategy, InfoSphere Guardium & Optim
Leslie Horacek	X-Force Threat Response Manager
Marc Noske	Database Administration, Content Security

Contributor	Title
Mark E. Wallis	Senior Information Developer, IBM Security Systems
Marne Gordan	Regulatory Analyst, IBM Security Systems
Michael Applebaum	Director of Product Marketing, Q1 Labs
Michael Montecillo	Managed Security Services Threat Research and Intelligence Principal
Michelle Alvarez	Manager, MSS Global Operations
Paul Sabanal	X-Force Advanced Research
Phil Neray	Q1 Labs Marketing Leader, IBM Security Systems
Ralf Iffert	Manager X-Force Content Security
Randy Burton	Senior Incident Response Analyst
Robert Lelewski	Senior Incident Response Analyst
Ron Black	Senior Incident Response Analyst
Ryan Berg	Cloud Security Strategy Lead
Scott Moore	X-Force Software Developer and X-Force Database Team Lead
Shane Garrett	Team Lead, X-Force Advanced Research
Tom Cross	Manager, X-Force Strategy and Threat Intelligence
Veronica Shelley	Segment Marketing Manager IBM Security Systems

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.

DEDICATION

The IBM X-Force 2011 Trend and Risk Report is dedicated in memory of our friend and colleague ***Marne Gordon*** *who passed away during the production of this report. A regulatory analyst with the IBM Security Division's security strategy team, Marne's knowledge and focus on cloud and social media security are featured in this report. She was a frequent speaker at industry events and published numerous articles on the topics of security and compliance. Marne, and the contributions she made to security, compliance, and to IBM will be greatly missed.*

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 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.
- Professional Security Services (PSS) delivers enterprise-wide security assessment, design, and deployment services to help build effective information security solutions.
- The IBM X-Force 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 AppScan® OnDemand Premium Service. This service combines application security assessment results obtained from IBM AppScan with manual security testing and verification.
- IBM Security Services supports the cloud in two ways: Security Services for the Cloud that help clients begin their journey to the cloud by providing security expertise and Security from a cloud-based model that helps reduce costs and complexity, improve security posture and meet compliance requirements.
- IBM Identity and access management solutions enable organizations to efficiently centralize and automate the management of identity profiles and access privileges for authorized users. These solutions can further strengthen security with strong authentication, single sign-on, and audit/reporting tools for monitoring user access activity.
- IBM data and information security solutions deliver capabilities to help protect data and access management that helps address information lifecycle security across the enterprise.
- IBM InfoSphere® Guardium® provides a scalable enterprise solution for database security and compliance that can be rapidly deployed and managed with minimal resources.
- The QRadar Security Intelligence Platform from Q1 Labs, an IBM Company, offers an integrated solution for SIEM, log management, configuration management and anomaly detection. It provides a unified dashboard and real-time insight into security and compliance risks across people, data, applications and infrastructure.

Contents
Section I

Contents
Section II, III and IV

Section I
Threats

In this section we explore threat-related topics and describe many of the enterprise attacks that security specialists face. We discuss 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

2011 was a remarkable year for IT security. By mid-year, in the midst of frequent reports of data leaks, DoS attacks, and social hacktivism, IBM X-Force declared 2011 "Year of the security breach." By the end of the year, the frequency and scope of these incidents have persisted, and continues to bring awareness to the basic tenets of operating a business and for protecting its assets in an increasingly connected world. The sheer number of high profile and highly public incidents throughout 2011 has been a catalyst for executives and business leaders to reevaluate the effectiveness of existing structures, policy, and technology in the enterprise.

With any great challenge, there comes a great opportunity to learn and improve. While companies have been forthcoming in disclosing when a breach has occurred, and what impact it might have for their customers, little is said about how it happened and what could have been done to prevent it. One difficulty we face in the security industry is how to responsibly disclose a breach so that the technical details may help to ensure that other businesses are not affected similarly. In this report we reflect on what we might discern from these unfortunate incidents, and how we might take an affirmative step in communicating breach information that could contribute to a culture of beneficial disclosure for the future.

Through the disclosure of breaches that have occurred, we still see SQL injection as a choice point of entry for attackers. Automated SQL injection attacks like LizaMoon are successfully scanning the Internet and exploiting vulnerable hosts. These SQL injection attacks have been common for a long time. Recently, we have also started seeing an increase in attacks targeting Shell command injection vulnerabilities. By the end of 2011 X-Force saw two to three times more Shell command injection attack activity than we saw earlier in the year. We have also noticed large spikes in SSH password cracking activity near the end of 2011.

We have seen unprecedented new attacks such as the compromise of several certificate authorities. This type of attack breaks a basic trust for users—that visiting an encrypted SSL page means we are communicating securely. Old methods of attack such as traditional phishing and spam are being replaced with new methods of deploying malware. Social media attacks are increasing and a prime target area for attackers who are successfully encroaching on their target's circle of trust by infiltrating their friends and followers.

Despite these difficulties, throughout the report, we have also observed some positive trends and improvements. The total number of reported web application vulnerabilities is lower than we've seen since 2005 and X-Force is seeing a significant decline in the number of true exploits that have been publicly released. When exploit code is released on the Internet it can provide an easy means for attackers to target vulnerabilities. In the past few years, exploit code was released for about 15 percent of the vulnerabilities that were publicly disclosed. This year that number has dropped to 11 percent. The frequency of exploit code releases targeting web browsers as well as document readers and editors was down to levels not seen in over four years. Publicly disclosed vulnerabilities were also more likely to have patches than ever before. The percentage of unpatched vulnerabilities was down to 36 percent from 43 percent last year.

In web application vulnerability testing, the IBM AppScan team has seen significant improvements in both the areas of Cross-Site Request Forgery (CSRF) and Cross-Site Scripting (XSS).

As we immerse ourselves and our business in an ever more connected, open, mobile, and social online presence, so too do opportunistic individuals devise new ways of exploiting the system with versatility and ease. Using a lowest common denominator approach, they not only are targeting the technology, but the individual directly, taking advantage of basic human nature, and preying on trust. Social media and mobile devices continue to blur the lines between the boundaries of the enterprise and the outside world.

Along these lines, in this report, we will continue to explore how companies are keeping up with the complexities of mobile devices and cloud. The mass adoption of mobile devices brings the discussion of "bring your own device" (BYOD) programs to the forefront, and how to mitigate the risks associated with these policies, along with the greatest threats affecting this platform.

Cloud adoption faces similar discussions. The question is not whether the cloud is more or less secure, but on what specific controls, and business processes, do we need to be focused to address risk and help ensure security in a cloud environment. It is important for any organization when planning to more widely adopt cloud-based infrastructures to have an understanding of the role of the organization versus the role of the cloud service provider when it relates to security and risk mitigation.

Throughout 2011 security teams were repeatedly challenged to do better. Many were challenged to improve processes, technology, to educate employees and customers on safe practices, and to raise security intelligence by increasing visibility into the security posture of the business. IBM believes the way to help clients get ahead of security threats is to connect our analytics and intelligence capabilities across an organization for better prediction and detection. IBM made a big move by acquiring Q1 Labs in October 2011 and creating the new Security Systems division. Continued news on how we're advancing our security intelligence platform shows how seriously we're addressing the market. With awareness comes action and change. It is our hope to make change.

2011 Highlights

Threats:

Malware and the malicious web

- An explosion of data breaches opened 2011 and continued throughout the year with IBM X-Force declaring "2011—Year of the security breach." (page 12)
- SQL injection continued to be a major exploited weakness in targeted companies. SQL injection has been around for some time but continues to be a successful means of entry. (page 17)
- Another milestone occurred in 2011 after attackers compromised several certificate authorities; the most publicized being the Dutch firm DigiNotar. Attackers were able to generate unauthorized certificates which they could later intercept using a man-in-the-middle type of attack as a way to listen on an encrypted connection. This type of attack breaks a basic trust for users—that visiting an encrypted SSL page means we are communicating securely. (page 33)
- A proof-of-concept tool to perform a denial-of-service (DoS) attack against servers communicating via SSL/TLS was released in 2011. This tool showed the potential for an everyday laptop on an average connection to take down an enterprise web server. (page 33)
- Top high-volume signatures from IBM Managed Security Services (MSS) group demonstrate favorite attacker methods to be SQL injection, increases in SSH brute forcing and Shell command injection activity, and proxy bouncing continue to rank at the top of MSS sensor traffic. (page 16)
- More than in any previous year so far, 2011 has seen the most activity in the Mac malware world. This applies not only in volume compared to previous years, but also in functionality. In 2011, we started seeing Mac malware with functionalities that we've only seen before in Windows® malware. (page 39)

Web content trends, spam and phishing

- In the first term of 2011, anonymous proxies have steadily increased, more than quadrupling in number as compared to three years ago. However, in the second half of the year for the first time since the beginning of 2009, we did not see another increase of this volume. Anonymous proxies are a critical type of website to track, because they allow people to hide potentially malicious intent. (page 44)
- In 2011, spam volumes continued to decline into the end of the year where a shift to spam delivering malware with zip attachments became a method of choice. (page 49)
- Top countries for distributing spam in the year—India continued to dominate the top of the list by sending out roughly 14 percent of all spam registered today. The USA which was at the top of the list one year ago is down to less than 2 percent of spam sent overall. Behind India follows Vietnam, Indonesia, Russia, Brazil, and for the first time Australia, rounds out the top six lists responsible for 5.6 percent of all spam distributed by the end of 2011. (page 55)
- Nearing the end of 2011, we began seeing the emergence of phishing-like emails that link to websites which do not necessarily perform a phishing attack. These emails use the good name of a well-known brand to get users to click on a malware link or in some cases a link to an otherwise innocuous site such as a retail site. One possible explanation for the latter type of emails might be click-fraud, wherein spammers drive traffic to these sites in exchange for advertising fees. Regardless of the explanation, this nuisance contributed to a large increase in phishing-like emails seen in the later months of the year. (page 56)

Operating secure infrastructure:

Vulnerabilities and exploitation

- 2011 reported just over 7000 new security vulnerabilities. While this is a significant decline from 2010, when we saw more vulnerabilities than ever before, there has been a two year, high-low cycle in vulnerability disclosures since 2006, and the levels of each high point and each low point keep climbing. (page 74)
- For the past few years about half of the disclosed security vulnerabilities were web application vulnerabilities. However, this year that number was down to 41 percent, a percentage that hasn't been seen since 2005. (page 75)
- One category of web application that is subject to both public vulnerability disclosure and a lot of attack activity is web-based content management systems (CMS). We took a look at four popular web-based content management systems, and our data shows that the most important weaknesses in these systems come from the ecosystem of third-party plug-ins that they support. (page 77)
- In 2011 X-Force has seen a significant decline in the number of true exploits that have been publicly released. It was the lowest number we've seen since 2006. This number is lower on a percentage basis as well as a real basis. For the past few years the percentage of vulnerabilities with public exploits has hovered around 15 percent, but this year it was 11 percent. (page 78)
- High and critical browser vulnerabilities continue to rise and we have also observed an increase in drive-by-download attacks that have moved into targeting third-party browser plug-ins rather than the browser itself. Document readers are one such third-party component that has been a favorite of attackers as malicious document files can be used in drive-by-download scenarios as well as attached to emails. (page 78)
- We continue to see increases in the number of vulnerabilities being disclosed in multimedia players and we saw just as many public exploits for multimedia vulnerabilities in 2011 as we saw in 2010. This continues to be an area of focus for attackers. (page 81)
- The largest enterprise software vendors have represented a constantly increasing percentage of the total number of vulnerabilities disclosed, from 19 percent in 2008 to 31 percent in 2011. We don't believe that this is merely a measure of software industry consolidation. Secure development practices have become an increasingly important part of the software development lifecycle, and responsible vendors have taken steps over the past few years to improve their ability to identify and eliminate vulnerabilities in their code. (page 84)
- In the last seven years, social networking has gone from a fringe pastime to become the number one online activity in the world, eclipsing even use of search engines. Naturally, such concentrated activity represents a fertile environment for the criminal element. Frauds and scams that were successful years ago via email found new life on the social media forums as well as a fresh group of potential targets. (page 89)

Software development security practices

Web application vulnerabilities

- Many issues from the 2010 OWASP Top Ten showed up frequently in software submitted to the IBM AppScan OnDemand Application Vulnerability Testing Service. Broken Authentication and related issues with session control were found in nearly eight tests out of 10. Many applications tested failed to restrict session tampering and were exposed to session fixation style attacks. In addition issues relating to session termination and session reuse also attributed to this high statistic. (page 113)
- Cross-Site Request Forgery (CSRF) in 2011 was found in 28 percent of tests undertaken, but this number was reduced from 2010 where the percentage was 59 percent. Some of this reduction appears to be in the greater awareness in this type of vulnerability and also improvements in methods used to include CRSF tokens. (page 116)
- The fact that Cross-Site Scripting (XSS) is still found in over 40 percent of applications tested highlights that there are likely still many applications that do not adhere fully to secure coding practices. There is no doubt that things are improving, but that is no reason to be complacent. The likelihood of 40 percent for XSS vulnerabilities is still high, especially for something that is so easily understood, so easily demonstrated, and so easily fixed. Web application vulnerabilities remain the key to many data breaches, and data breaches continued to rise in the first half of 2011. So much so that X-Force declared 2011 to be the "Year of the security breach".(page 114)
- Another important data point that we capture is "the average number of a given finding per security test". What we are seeing is a reduction in instances of XSS when this vulnerability is found. In 2009 the average number was over 40 while in 2011 it was just over 3. It is now much less likely to find an application with absolutely no input control in place. (page 114)
- In 2011 Financial applications were again the best performing segment. Government applications were the worst performers in all three of these categories. It is not clear why this is the case, but reputational damage could be a factor. Breaches in Government applications are less likely to drive an investment in security mitigation than they would for financial applications. (page 116)

Emerging trends in security:

Mobile

- Mobile devices are another area that is gaining in importance. There are many mobile operating system vulnerabilities being disclosed, and there are a number of exploits being publicly released for these vulnerabilities. The desire to jail break or root mobile devices is one motivating factor that leads people to post mobile exploit code online. Of course, once that code is available, it can be used for malicious purposes against phones that are not jail broken. (page 82)
- Large botnets of infected mobile devices have started to appear on the scene and this is only the beginning. (page 83)
- Mobile devices (because they usually have GPS hardware, along with voice, messaging, and data services) have detected presence of spy applications that monitor multiple aspects of their users behavior—including recording location, messages, email, and voice calls to their attacker for review. This is particularly disconcerting when we compare it to the kinds of attacks we see on personal computers. Because mobile devices really have become "your office in your pocket," they can provide an opportunity for a spy attack. (page 122)

- One of the more recent developments this year has been the increased interest in providing the ability to separate enterprise applications and data from the employee's personal applications and data. Obviously, a primary driver for this development has been the pervasive nature and interest in BYOD programs. (page 125)

Cloud security

- The question is not whether the cloud is more or less secure, but what specific controls, and business processes, should we focus upon to address risk and help ensure security in a cloud environment. It is important for any organization looking to more widely adopt cloud-based infrastructures that they have an understanding of the role of the organization versus the role of the cloud service provider when it relates to security and risk mitigation. (page 126)
- Success in secure cloud computing can be more than a question of simple contract management, but it can be critical to the success of the cloud deployment. Creating a flexible Service Level Agreement (SLA) that takes into account lifecycle management and exit strategy could be beneficial. (page 128)
- SLAs should be true agreements, specific in both terms and scope, changeable only with appropriate notice, and cognizant of the specific business and information security requirements of the organization. (page 131)

2011—Year of the security breach

From mid-year to the new year—the breach plays on

At the mid-year, IBM X-Force declared 2011 "Year of the Security Breach" which was marked with a litany of significant, widely reported external network security breaches and other incidents, notable not only for their frequency, but for the presumed operational competence of many of the victims.

The second half of 2011 continued to demonstrate common reports of weekly wide-scale network security breaches, leaving a wake of leaked customer data, inaccessible web services, and billions of dollars of damages. IT security is now a board room discussion affecting business results, brand image, supply chain, legal exposure, and audit risk. In the IBM X-Force 2011 Mid-year Trend and Risk Report, we looked at the underlying motivations, attack methods, and basic security practices which were circumvented to set 2011 apart as the year of the security breach.

These incidents did not discriminate against any industry or sector. Law Enforcement, governments, social network communities, retail, entertainment, banks, non-profits, Fortune 500, and even security companies, were attacked. No single geography was the focus, but clearly these attacks occurred on a global scale. As the year came to a close, the trend showed no signs of slowing. December marked some of the largest impact-by-cost breaches that affected several massive social and entertainment sites in China with billions of dollars of potential losses.

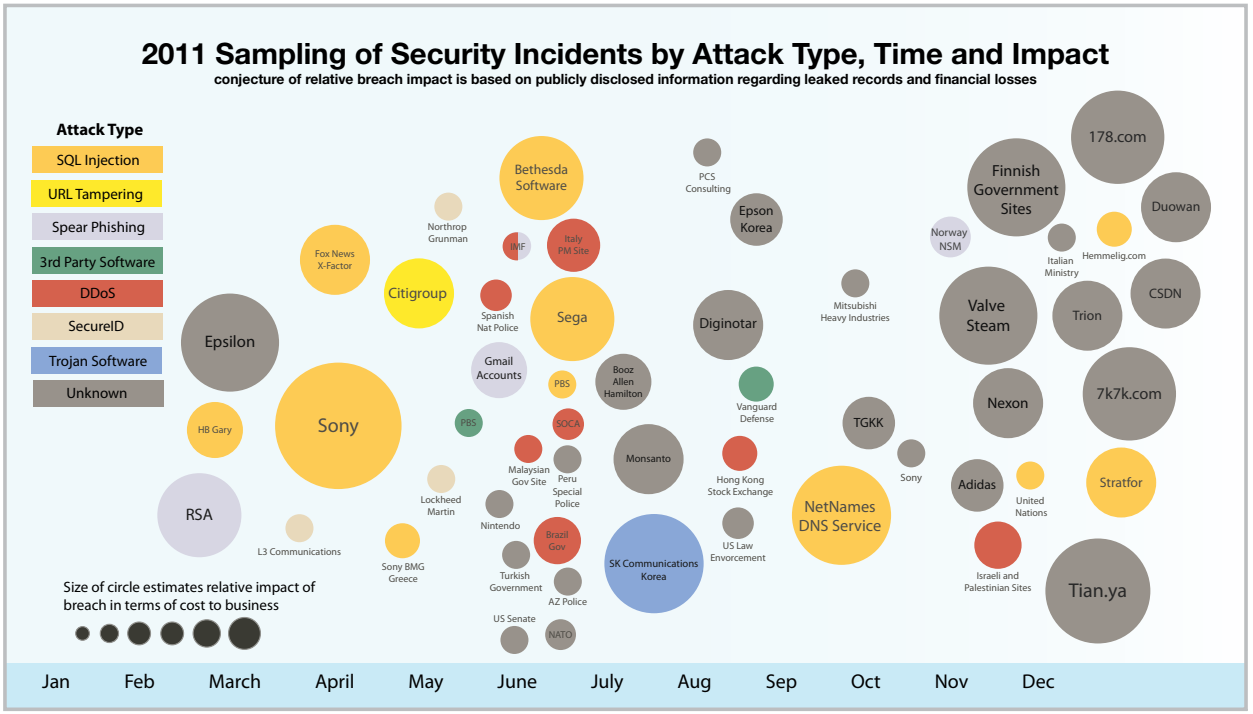


Figure 1: 2011 Sampling of Security Incidents by Attack Type, Time and Impact

Game changers in the second half of 2011

As illustrated in Figure 1, **SQL injection** continued to be a major exploited weakness in targeted companies. SQL injection has been around for some time but continues to be a successful means of entry. Later, we discuss the complexities of SQL injection, and why it is so difficult to discern and protect networks from it.

Adding an additional dimension of sophistication to breaches in 2011, we saw several examples of compromised core technology leading to a wide-scale exploitation of other targets. Early in the year, an attack at RSA resulted in the theft of sensitive code and data associated with the company's SecureID authentication product. Later it was disclosed[1] that the compromised technology was used to gain entry into at least three other corporations. This represents an increase in complexity where attackers are not only exploiting a specific end target, but also gaining a foothold into the underlying technologies used by a wider base of potential victims.

Another emerging trend which has continued into 2012 is attackers targeting DNS servers as a means of redirecting unsuspecting users to malicious variants of well-known sites. Every time a user enters a web domain in a browser such as http://www.somecompany.com, the name has to be translated to the IP address of the server that hosts the site.

A SQL injection on a NetNames DNS name server allowed attackers to update the DNS records for several high profile sites such as The Register, The Daily Telegraph, and UPS.[2]

By compromising the DNS name server itself, attackers reroute requests to a server of their choice, often creating a similar looking variant of a well-known site that contains downloadable malware or forms set up for phishing sensitive information. This type of attack breaks a basic principle of trust, that typing in a website name will take us to the correct server.

Another milestone occurred after attackers compromised several certificate authorities,[3] most publicized being the Dutch firm DigiNotar. Certificate authorities distribute security certificates, which provide the secure function of the HTTPS protocol used to encrypt traffic from users to online services. Attackers were able to generate unauthorized certificates which they could later intercept using a man-in-the-middle type attack as a way to listen in on an encrypted connection. This again breaks a basic trust for users—that visiting an encrypted **SSL** page means we are communicating securely. Later in this report, we discuss in more detail the risks associated with the current SSL trust model. In each of these cases, we see attackers using a multi-tiered strategy, comprising some core technology and then using it to cast a wide net of potential targets.

1 http://www.nytimes.com/2011/06/04/technology/04security.html
http://www.infosecisland.com/blogview/14142-RSA-SecurID-Breach-Spreads-to-L3-and-Northrop.html
2. http://www.theregister.co.uk/2011/09/05/dns_hijack_service_updated/
3. http://www.theregister.co.uk/2011/10/27/ssl_certificate_authorities_hacked/

Lessons learned

As you can see from the chart of security breaches in 2011, there were many cases in the latter half of the year where a breach was publicly reported but we do not have any information about how the breach occurred. There are a number of different motivations that drive the public disclosure of security breaches, but a desire to inform the public about the technical vulnerability that was exploited by the attacker is not usually one of those motivations. Disclosure is often motivated by a desire to inform customers whose personal information or corporate data may have been exposed, or that a technology the victim produces has been compromised. Recently, financial analysts have begun to take an interest in using information about computer security risks when evaluating investment decisions. However, it's relatively rare that companies disclose security breaches specifically because they want to bring attention to a computer security problem that other firms might face. We believe this is unfortunate, as security professionals could stand to benefit from the hard lessons learned by others.

Many boating and flying magazines print monthly columns describing real world situations that were either very dangerous or resulted in an accident. Through reading these reports on a monthly basis, pilots and skippers get the benefit of regularly analyzing each other's actions in hindsight. Through this process they can learn how to handle difficult situations and develop a confidence that can be valuable in a crisis. Similarly, people who are responsible for protecting computer networks from attack should regularly immerse themselves in information about security failures, so that they can develop good instincts about what pitfalls to avoid. Knowing the exact technical as well as process failures that led to a breach can shed light on the gaps in one's own posture.

Often, computer security is perceived to be a cost of doing business, and businesses seek to avoid investing resources into fixing security flaws that may never be exploited. There is a desire to find the "sweet spot" where enough money is being spent on the right security investments to protect the company, but not a dollar more. This means that merely identifying a technical or procedural gap is usually insufficient to convince a business to invest in closing that gap—there must be a demonstrable real world risk that this gap will be exploited if it is not fixed. When the victims of security breaches expose particular technical and procedural gaps that led to the breaches that they have experienced, this information helps to provide the business justification that others require in order to obtain the investment needed to close similar gaps. When situations arise that result in technically similar breaches across multiple firms, disclosure of the specific technical vulnerability involved can drive discussion toward addressing that sort of vulnerability throughout the marketplace.

Victims of computer crimes should consider the value of talking with the public about the technical details of "what went wrong" when they otherwise publicly disclose a security breach. There are going to be cases where the risks involved in disclosing that sort of information might outweigh the benefits. Obviously, providing too much technical detail could create a roadmap for future attacks. However, it's important to understand what the benefits of disclosure might be. Helping other people learn from your misfortune is an affirmative step that might be taken to limit the future success of the sort of criminals who have compromised your network.

The way forward

In our X-Force 2011 Mid-year Trend and Risk Report we identified ten steps that X-Force would suggest taking to mitigate some of the attacks that have happened this year. None of the steps we suggested is a ground breaking revelation for IT security pros. The challenge is not knowing what to do, but executing consistently across a complex, decentralized organization. In order for a security program to be successful it must have the resources, political support, and institutional respect needed to ensure compliance with best practices throughout the organization. Achieving that level of effectiveness is the true challenge of IT security leadership.

If IBM X-Force were running the IT department

1. Perform regular third party external and internal security audits
2. Control your endpoints
3. Segment sensitive systems and information
4. Protect your network
5. Audit your web applications
6. Train end users about phishing and spear phishing
7. Search for bad passwords
8. Integrate security into every project plan
9. Examine the policies of business partners
10. Have a solid incident response plan

For more detailed information on any of the above points, please download and read the IBM X-Force 2011 Mid-year Trend and Risk Report.

IBM Managed Security Services—A global threat landscape

IBM Managed Security Services (MSS) monitors tens of billions events per day in more than 130 countries, 24 hours a day, 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. Threat trend identification is vital to establishing future security strategy, and understanding the significance of the threats.

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 2011 as compared to year end 2010. Four of the top ten signatures from 2010 have retained a spot on the 2011 year end list. SQL_Injection and SQL_SSRP_Slammer_Worm have managed to remain high on our list for two years now, although Slammer activity has been trending slightly downward. The downward trend of SQL_Injection was reversed in 2011. SSH_Brute_Force continues to hold a position in the top ten but has fallen to ninth place. HTTP_Unix_Passwords persists in the top ten from the 2011 report as well, although it has dropped from sixth place to tenth, despite its continued growth upward.

Event Name	2011 Rank	Trend	2010 Rank	Trend
SQL_Injection	1	Up	2	Down
HTTP_Suspicious_Unknown_Content	2	Down		
SQL_SSRP_Slammer_Worm	3	Slightly Down	1	Down
SNMP_Crack	4	Down		
HTTP_GET_DotDot_Data	5	Up		
Cross_Site_Scripting	6	Slightly Up		
SSH_Brute_Force	7	Slightly Up	4	Slightly Down
HTTP_Unix_Passwords	8	Up	6	Slightly Up
Shell_Command_Injection	9	Up		
Proxy_Bounce_Deep	10	Up		

Table 1: Top MSS high volume signatures and trend line—Year End 2011 vs Year End 2010

SQL injection

Our heuristic SQL signature, which ranked second in 2010, climbed to first place and has been trending upward. 2011 was a banner year for exploiting SQL weaknesses and several high profile and newsworthy episodes of successful SQL injection attacks were made public. The hacktivist groups Anonymous and Lulzsec were major players in SQL injection tactics and continue to hone their skills with new injection attack vectors. Additionally, there are automated SQL injection attacks like LizaMoon that scan the Internet for vulnerable hosts and that is the origin of most of the activity we are seeing. IBM MSS has added multiple additional attack vector coverage to its Security Information and Event Management (SIEM) rule sets and continues to monitor, and analyze for any new vectors every day. The following section titled "The Continuing Threat of SQL injection" discusses the nature of this threat in depth as well as explains actions organizations can take to help protect their web application code, servers, and networks from SQL injection.

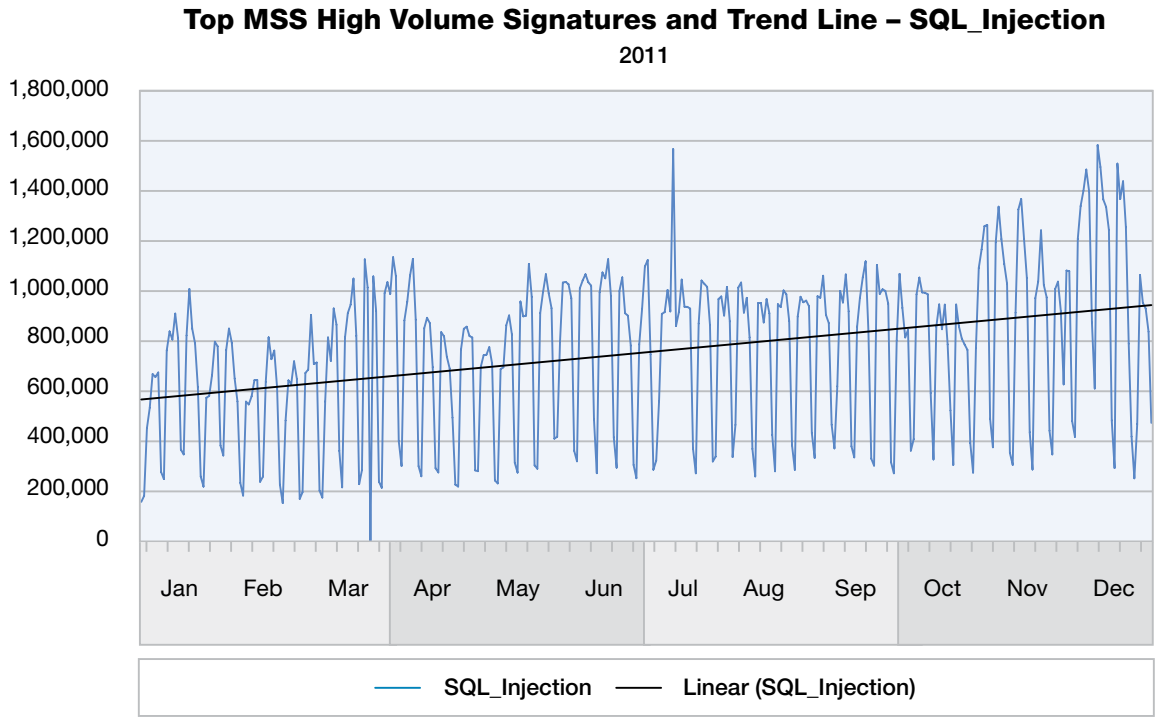


Figure 2: Top MSS high volume signatures and trend line—SQL_Injection 2011

Zeus in our midst?

The HTTP activity which triggers our second place signature, HTTP_Suspicious_Unknown_Content, might be normal. However, it could indicate that a BotNet, such as Zeus, is active on your network. Zeus is a widely known banking Trojan that was first identified in July 2007 and became widespread by mid-2009. The primary infection vectors are drive-by-downloads and phishing. There are many different individuals and groups that have Zeus botnets set up. The goal of Zeus botnets is usually to steal personal information. Commonly this is online banking data that can be used to access bank accounts to transfer money.

The FBI has been aggressively pursuing various groups that create botnets using Zeus. However, despite the successful take down of many of the original Zeus command and control servers in 2010, MSS tracked a large number of Zeus infections. Because Zeus is very difficult to defend against and antivirus products are at best temporary stop gaps for Zeus propagation, user education has become the primary focus of combat. Training employees to not click hostile or suspicious links in emails or on the web while also keeping up with antivirus updates has become the primary defensive strategy.

SQL Slammer's continuous decline

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). Slammer continued to be a pervasive threat over the years, with Slammer infection packets still accounting for a sizable portion of UDP traffic on the Internet. In fact, the top signature in 2010 was SQL_SSRP_Slammer_Worm. However, this signature had fallen to second place when we checked mid-year and dropped to third place when we assessed the year-end data. "The day that SQL Slammer disappeared" section in our X-Force 2011 Mid-year Trend and Risk Report discusses the dramatic fall in SQL Slammer activity in March 2011 which contributed to the lower placement of this signature on our list.

There were a few times throughout 2011 where the activity staged a moderate comeback only to drop off again, as shown in Figure 3.

We noted a larger than normal volume for Slammer in December, but it does not appear to be a resumption of the pre-March patterns. We are monitoring the situation and will note any new trends that emerge going forward.

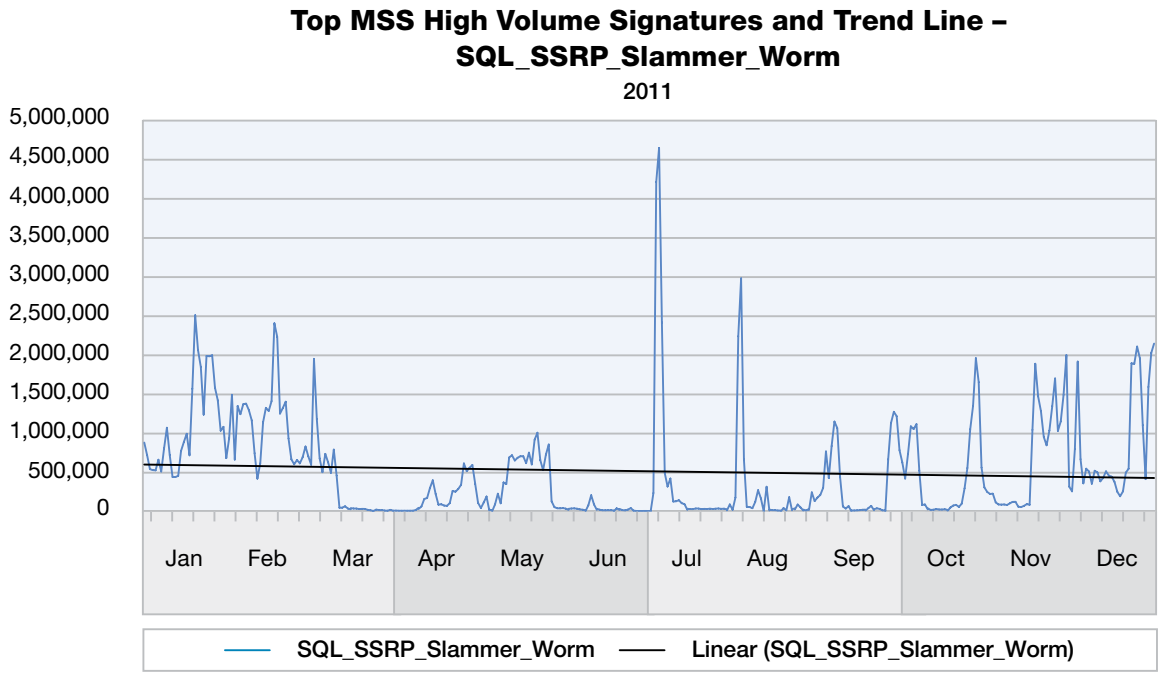


Figure 3: Top MSS high volume signatures and trend line – SQL_SSRP_Slammer_ Worm 2011

SNMP vulnerabilities

Our SNMP_Crack signature indicates an attempt to brute force SNMP Community Strings. SNMP is a service that makes it easy for network administrators to monitor the status of network devices, and sometimes to control their configuration. Operating systems, hubs, switches, and routers all utilize SNMP. SNMP uses Community Strings like passwords to protect access to sensitive information and controls. Often SNMP services are configured with default community strings and attackers will search for that first. Otherwise, attackers may attempt to guess Community Strings through brute force. We recommend that organizations assess the need to have SNMP active on their devices and disable it if it isn't necessary.

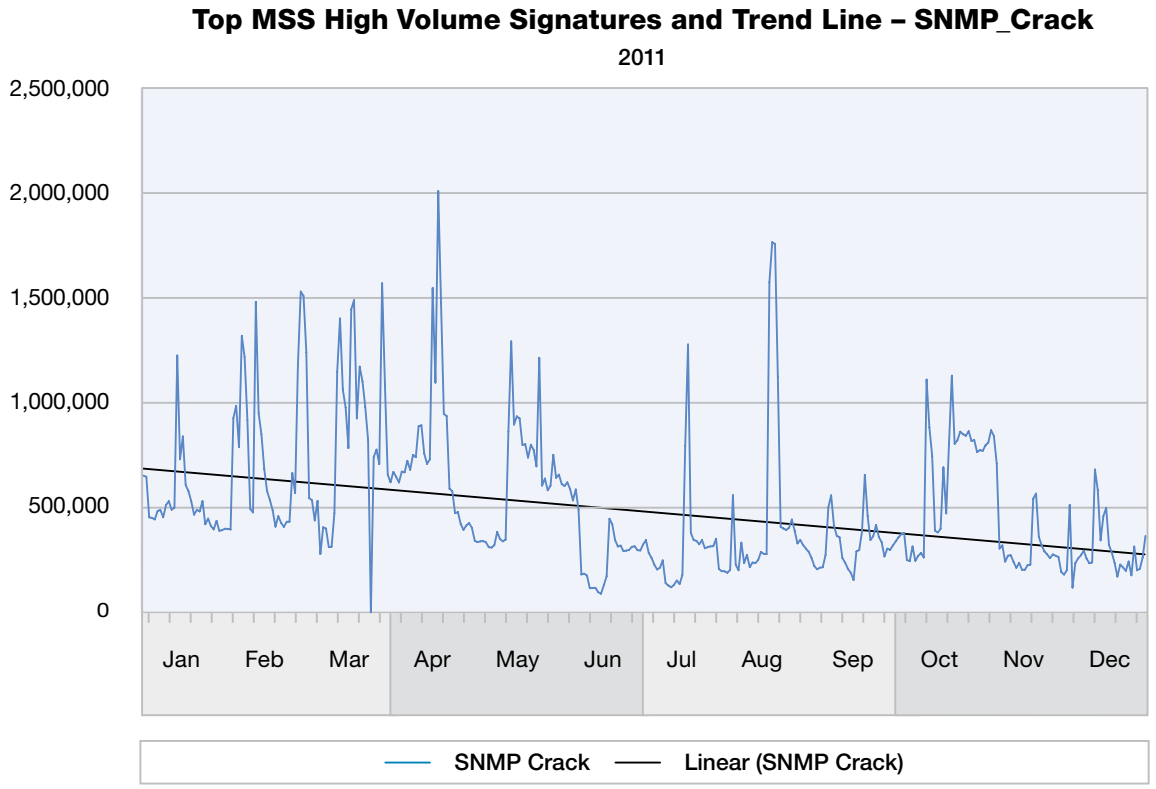


Figure 4: Top MSS high volume signatures and trend line – SNMP_Crack 2011

Traversing directories

The HTTP_GET_DotDot_Data signature detects an attacker's attempt to bypass the normal security imposed by the web server 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 worldreadable 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.

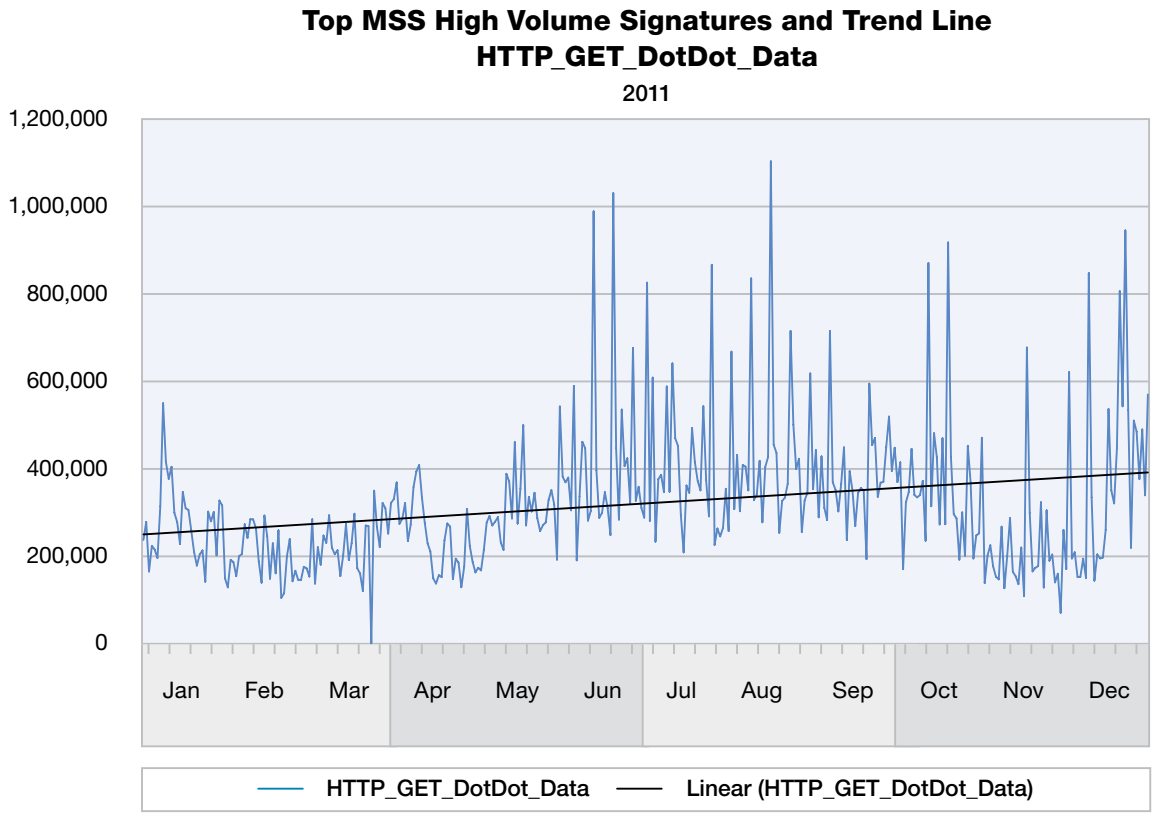


Figure 5: Top MSS high volume signatures and trend line – HTTP_GET_DotDot_Data 2011

Cross_Site_Scripting

Usually found in web applications, a cross-site scripting attack enables attackers to inject client-side script into web pages viewed by other users. This attack can also be used by attackers to bypass access controls. This attack has an extremely high popularity and is a significant security risk. Cross site scripting has been popular since the 1990's and it is the most common type of web Application vulnerability. Our Cross_Site_Scripting signature falls into position number eight in our list of the top ten signatures tracked by volume. Reducing the threat largely entails several tactics including validating HTML input, cookie security and disabling client side scripts. More recently emerging technologies are currently available such as Mozilla's Content Security Policy, Javascript Sandbox tools, and Auto-escaping templates that although still evolving, help to reduce the threat.

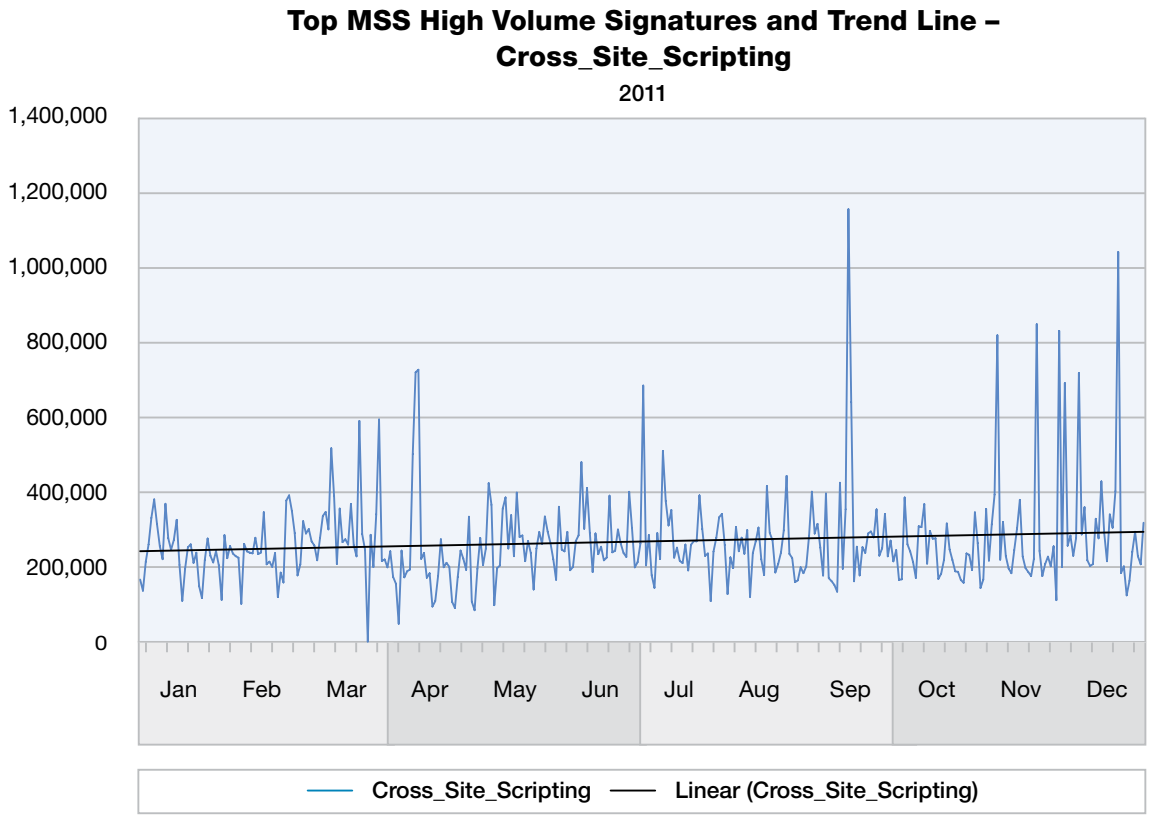


Figure 6: Top MSS high volume signatures and trend line – Cross_Site_Scripting 2011

Brute force attacks

SSH_Brute_Force holds seventh place, down three slots from its fourth place rank in 2010. A brute force attack involves an attacker attempting 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. In 2011, we observed constant activity scanning the Internet for insecure SSH servers with weak passwords. Organizations should mitigate against brute-force attacks by disabling direct access to root accounts and using strong usernames and passwords.

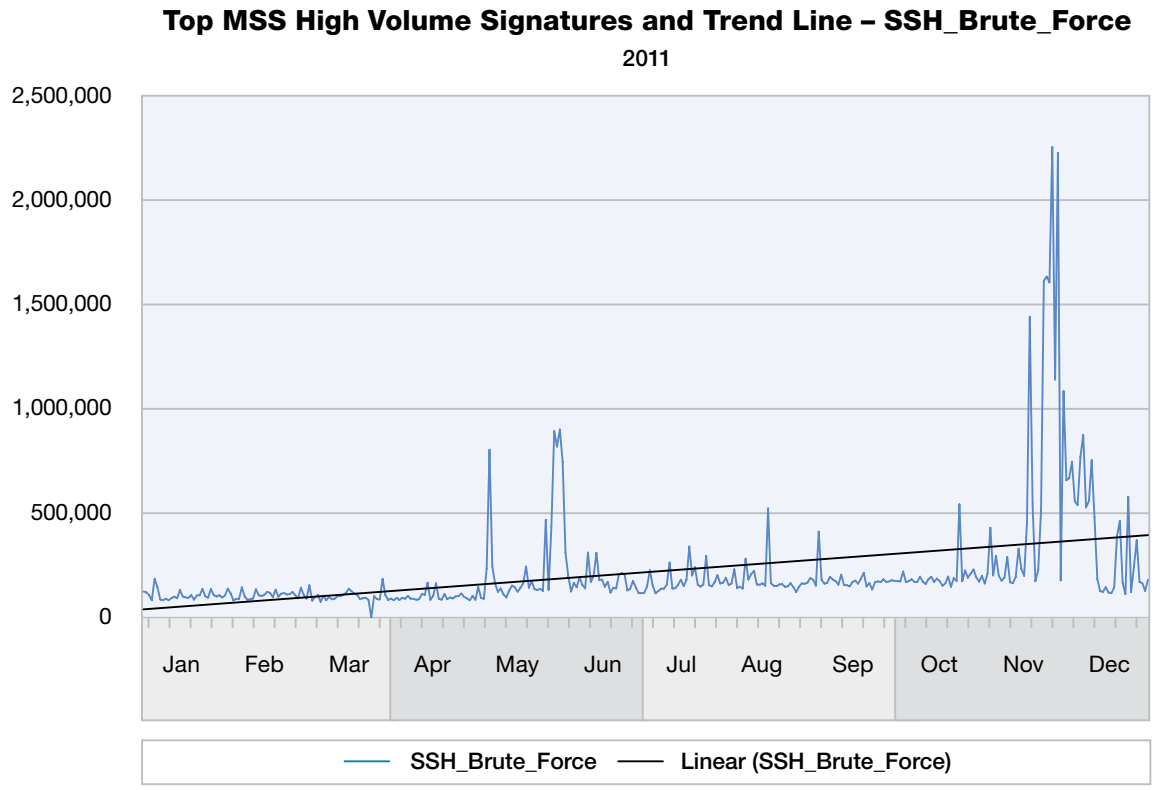


Figure 7: Top MSS high volume signatures and trend line – SSH_Brute_Force 2011

Attacks against UNIX

While the signature HTTP_Unix_Passwords remains in the top high-volume list and continues to see an upward trend, it drops from sixth place in 2010 to 10th place in 2011. This signature detects attempts to access the /etc/passwd file on UNIX systems via a web (HTTP) server. While this activity may be authorized, it can sometimes be suspicious. This is a very old attack, but is still successful today.

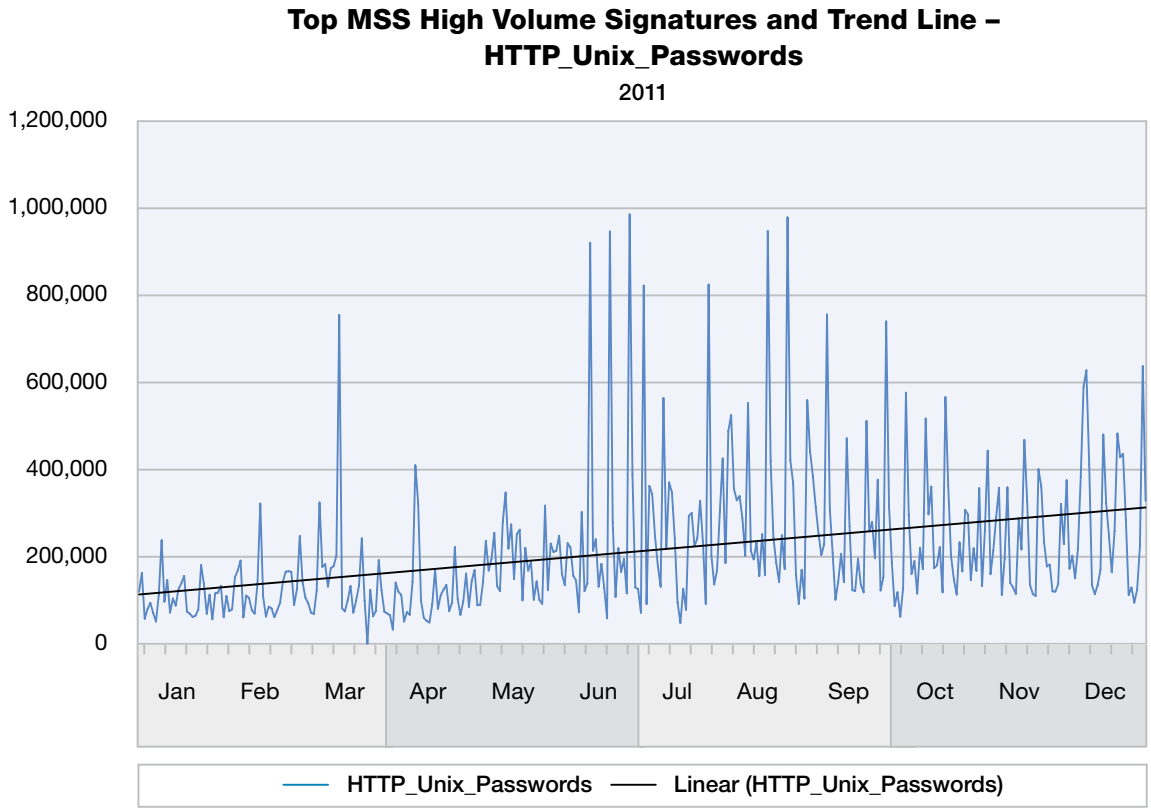


Figure 8: Top MSS high volume signatures and trend line – HTTP_Unix_Passwords 2011

Remote command injection

MSS has been tracking Remote command injection attacks globally. These vulnerabilities exist when user input is not properly sanitized and then it is used with functions that execute system shell commands, such as PHP functions like exec() and system(). This allows attackers to execute commands on the web server. This is a very basic, yet often successful attack for the same reason as SQL injection, proper security at the application level has not taken place.

Many of the payloads we have witnessed, consist of getting the web server to download a remote script via wget, store it in a tmp directory, and finally executing it. The script is designed to maintain remote access to the system, gather intel, and establish command and control back to the attacker's server. The server is then used to scan and attack other servers it finds, locally, and remotely via Google. This is a very quick and effective means for attackers to gain control of hundreds of vulnerable websites. In 2012 we can only expect to see a steady increase of the activity as some botnets grow and other attackers start to use the vulnerabilities for their own use.

Protection can be as simple as sanitizing any inputs from your website to exclude many popular shell commands such as passwd, wget, dir, and so on.

Also removing the command wget from the server can certainly hinder what an attacker can do without further digging.

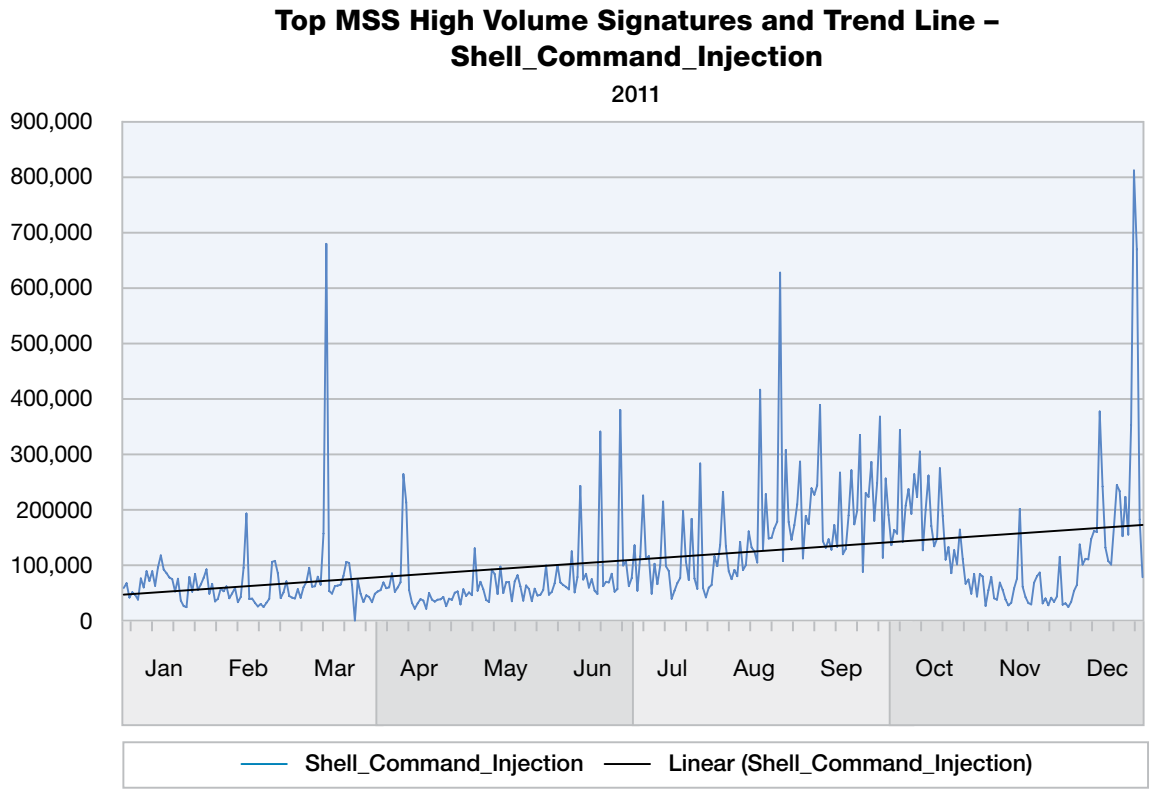


Figure 9: Top MSS high volume signatures and trend line – Shell_Command_Injection 2011

Nested anonymous proxies

The X-Force Proxy_Bounce_Deep signature detects situations where clients are attempting to access websites through a chain of HTTP proxies. We have seen large batches of this activity pop up on the networks of different clients. This could represent extremely paranoid but ultimately legitimate web surfing, but attackers sometimes do this to obfuscate the source address from which they are launching attacks against web servers. We have seen significant increases in the number of anonymous proxies on the Internet over the past few years that can be used for this purpose. You can learn more details on this topic of **anonymous proxies in the web content section** of this report.

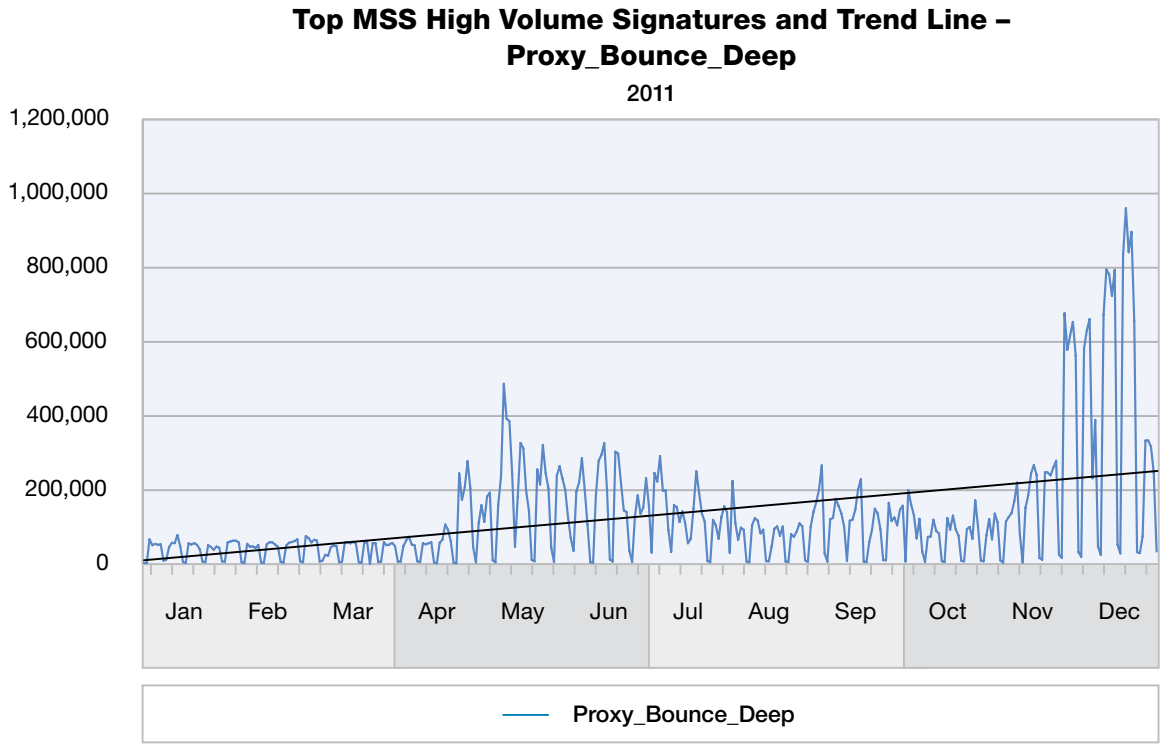


Figure 10: Top MSS high volume signatures and trend line – Proxy_Bounce_Deep 2011

The continuing threat of SQL injection

SQL injection

The Structured Query Language (SQL), originally conceived of in the 1970's, is a powerful language used to manage data in relational databases. While invented for use in large-scale data farms, the relational database has taken on new roles when paired with the interactivity of the web. Search forms, account management, order tracking, and collaboration tools are all possible through the union of these two technologies. This combination has resulted in innovation, but it has also created the risk of data leakage and provided an effective avenue of attack.

For years, attackers have used specially formatted strings in web forms and against web application programming interfaces. These strings are designed to manipulate the underlying database by injecting SQL statements into the web application's code. This process, known as SQL injection, can be used to bypass authentication, access the unpublished contents of the database, or even to compromise the operating system that hosts the database.

Initially, SQL injection was a targeted attack since the database schema and web application code is different for every site. Once an attacker found a vulnerable web application, the attacker would use crafted queries to map out the database. Armed with table and field names, the attacker could access information and explore permission issues. The attacks were slow, targeted, and the process was fairly manual. This type of targeted attack still exists. When HBGary Federal CEO Aaron Barr stated that he had been able to identify high-ranking members of Anonymous, the group used a SQL injection attack against HBGary's website. That attack ultimately led to a root compromise of their network, the disclosure of their sensitive data, and the resignation of Aaron Barr. When Sony announced that they had secured their network after the largest breach of customer data in history, LulzSec responded by posting over a hundred and fifty thousand customer details that they were able to get with SQL injection.

Starting in 2008, a new type of SQL injection attack emerged that no longer required knowledge of the underlying database structures or web application code. Instead of trying to access the data stored in the database, the attacker would inject a script and get the database to execute it. Since the only reconnaissance needed for this type of attack was to find a vulnerable server, it was very easy to automate. The first mass SQL injection attacks were born. Rather than go after the contents of the databases, these attacks generally seek to gain root access or use the web server to attack the users accessing the site. This can be accomplished by inserting a Cross-Site Scripting (XSS) vulnerability or

other malicious content into the web application or its cache. A hallmark of these attacks is SQL injection attempts that include DECLARE statements to insert the script and EXEC statements to execute the script. IBM Managed Services saw a large increase in the number of these types of attacks in 2011 as shown in Figure 11, especially in the latter half with jjghui and other variants attacking ASP.NET sites. jjghui is a mass SQL injection attack which refers back to the website it redirects its traffic towards.

A new mass injection technique emerged in 2011 that combines a scripted payload with some knowledge of the underlying database structure. This was first seen in March with the LizaMoon attacks. These attacks use UPDATE and REPLACE commands against a valid table instead of a blind DECLARE and EXEC. This requires a bit more work, but is more difficult to detect with simple pattern matching—especially when the URL is obfuscated.

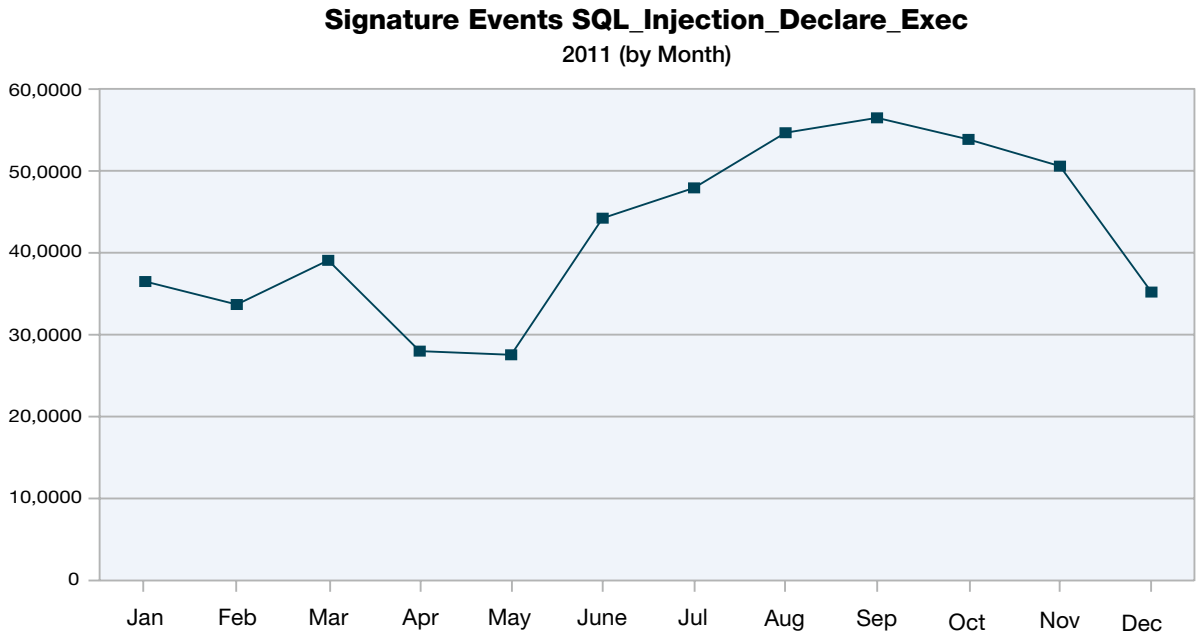


Figure 11: Signature Events – SQL_Injection_Declare_Exec 2011 (by month)

The nature of the threat

SQL injection attacks have been around for a long time, but they are still the most common type of attack on the internet. They are often successful, but generally can be prevented by sanitizing all user input and securing the database. From a security perspective, there are two types of systems that are vulnerable to this attack. There are the web connected databases that you know about, and the ones that you don't. Your network may contain web pages with login accounts, employee services, a storefront, or any number of public facing sites. These are the sites that you know about and they are likely to contain sensitive information like user accounts, credit card numbers, or customer contact information. If databases containing this type of information interact with one of your web servers, you have likely taken steps to secure the data. But, are they enough?

You may have a secure coding policy in place and may have done a thorough security review when the first version of your website was deployed. But over time, there are many opportunities for vulnerabilities to crop up. As new features are deployed, are they code reviewed? As new scripts or software applications are added, are they being researched and tested for vulnerabilities? As new tables and fields are added to the database, are the permissions being set properly?

As new developers are hired, are they given training on secure web programming? The loss of confidential data can have serious repercussions. Not only are there direct financial costs, but it can also create trust issues with your customers.

In addition to the servers that you know about, there may be some on your network that you don't know about. With the advent of open source databases and web tools, integrating database and web servers has become fairly simple. With an Apache web server, MySQL or Postgres database, and community-supported web code—anyone with an idea for a new web application can build one if they are willing to put in the research time. Knowledge bases, collaboration tools, ticket tracking, and testing tools are common examples of these internal applications. While this path can lead to great innovation, the builders of these applications may not have received training in secure web development. There are many resources available for learning best practices, but part time web developers are generally more concerned with function than security. Without proper training, they may not be aware that SQL injection is a possibility. Novice developers are also more likely to download canned modules or copy example code—two things that can greatly increase the chances of being a victim of a mass injection attack.

While these types of systems are less likely to have information like credit card account numbers, they may still contain sensitive data. Even if the data stored in the database is not sensitive, the database user names and passwords might be. If the database permissions are too permissive, the attacker may gain root access to the machine running the database. With a foothold in your network, the attacker can proceed to attacking higher value targets. They can also install bots and use your network to attack others.

Helping to protect your code

Like any other vulnerability, the key to helping stop SQL injection is a layered defense. The web application code is your first line of defense. This is the entry point of a SQL injection attack. To help protect the database from this code:

- Remove all SQL escape and unneeded reserved characters from any user-supplied data. We recommend using the peer reviewed libraries provided by your chosen programming language rather than attempting this yourself. There are many ways to encode dangerous characters and you may not be aware of them all.
- Validate encoding and data types returned by users—if you expect an integer, verify that you get an integer.
- Never allow user-supplied data to interact directly with the database. Even if you have sanitized the user-supplied data, you should never construct SQL statements with that data. Instead, use prepared statements, parameterized statements, or stored procedures to separate your SQL code from the data the user is supplying.
- Never return debug information to the user—log it locally instead.
- Periodically check to see if your programming language, server framework, or any third-party software you use has any known vulnerabilities.

If you sanitize all user-supplied data, then you deny the attacker a way to get to the database. You cannot rely on this alone, however, since it only takes one unchecked field to allow an attacker the opportunity to invade. You should ensure that everyone that changes code in your web application has had training on how to program securely. Consider making it a requirement to gain access to the code and periodically raise awareness of the importance of secure code.

Even the best developers can make mistakes if they are in a hurry or believe they are making a small change. The best way to catch this is through a peer code review. A second set of eyes helps reduce the chance of simple mistakes. With the deployment of a new technology, large feature addition, or any significant changes to a system with highly sensitive data, consider an outside code review or penetration test prior to making the application public.

Helping to protect your server

Your second line of defense is the connection with the database. You should:

- Never allow your web application to use a root or superuser account.
- Use the most restrictive permissions possible for the account you use to access the database server. Only grant permissions to the fields that the database must access and only allow write access to required fields.
- Remove default accounts, example code, and test applications that may have been installed with your database server. If you didn't write it and you don't use it, there is no reason to keep it.
- Use strong passwords and never store passwords in plain text.
- Routinely audit your database and web application logs for strange or repeating errors.
- Consider using database or log monitoring software to either prevent or notify you of a compromise.

Having a properly configured database server can be the difference between losing some data and a root system compromise. You should ensure that database security is a priority when interacting with a web server even when the data is not considered sensitive. You should periodically audit any such database for proper permissions and unneeded accounts. It is easy for these to get corrupted as new fields and tables are added.

If an attacker does manage to execute a SQL injection attack and gains sufficient permissions, the security of your operating system will be your last line of defense. Some steps you can take to help secure your system include:

- Secure the accounts and file system permissions for your database and web servers.
- Use host-based intrusion detection or protection that watches for intrusion attempts.
- Use antivirus and malware detection to look for bot infections.
- Monitor web application, web server, and database logs for suspicious behavior.

Helping to protect your network

Following the steps in the previous section will help keep the servers you are protecting safe from SQL injection. Locking down these servers may not be enough to protect your network from SQL injection though. If you have unprotected or unknown servers on your network, they could provide fertile ground for an attack. Proper use of firewalls and network-based intrusion protection or detection can help fill this gap. Blocking inbound web requests to addresses other than to authorized servers can help shield internal applications from outside attack. You should consider using web application firewalls or proxy based defenses for the web traffic that you do allow into your network. In addition, all major network-based intrusion detection vendors provide some level of SQL injection detection. The detection methods can vary by vendor and range from simple regular expression matches on known attack strings to complex scoring algorithms. Consider the following when protecting your network from SQL injection:

- Read the description of any SQL injection signatures that your vendor provides. Some signatures are very specific and only fire in limited circumstances while others are broader and prone to false positives. It is important to know what criteria are used for each alert.
- A lot of things look like SQL to an Intrusion Detection System but are not. Search results, the Yahoo Query Language (YQL), the Facebook Query Language, and twitter feeds are common false positives. Outbound SQL injection events can be a concern if there are a lot coming from the same address, but there are likely to be false positives in many SQL injection signatures due to user triggered events. Of more interest are inbound SQL injection attempts as these are the ones that could compromise your network.
- Periodically scan your network for unknown web servers. If one is found, track down the owner and ensure that steps have been taken to mitigate a SQL injection attack. Also consider outsourced penetration testing or purchasing software that specifically looks for sites vulnerable to SQL injection.
- Provide a security policy or guidelines that identify potential security problems and offer solutions.

Network protection alone is not adequate enough protection but, if you can identify and address a security breach early, you can mitigate the damage the attacker is able to cause. This is especially important for systems that don't contain sensitive data as they are the least likely to be secured. Even if the system itself is not critical, a root-level compromise on your network is a dangerous thing.

Conclusion

If proper precautions are taken, the risk of a SQL injection attack succeeding is low. It is important to remain vigilant though, as new business needs tend to push new features and technology. Every time a database connected web application is deployed or undergoes a code revision, you introduce risk. All code changes should be reviewed, and the education of your web developers should be an ongoing process. As long as attackers can successfully exploit unchecked user input, they will continue to try SQL injection attacks. And with the push toward bot-driven mass injection attacks, you can almost guarantee that someone will attempt to attack your sites. Will you be ready?

For more information on securing your servers against SQL injection, please visit the following links:

Securing Java:
http://today.java.net/pub/a/today/2005/09/08/handling-java-web-app-input.html

Securing ASP.NET:
http://msdn.microsoft.com/en-us/library/ff648339.aspx

Securing PHP:
http://php.net/manual/en/security.database.sql-injection.php

Database Security Tips:
http://en.wikipedia.org/wiki/Database_security

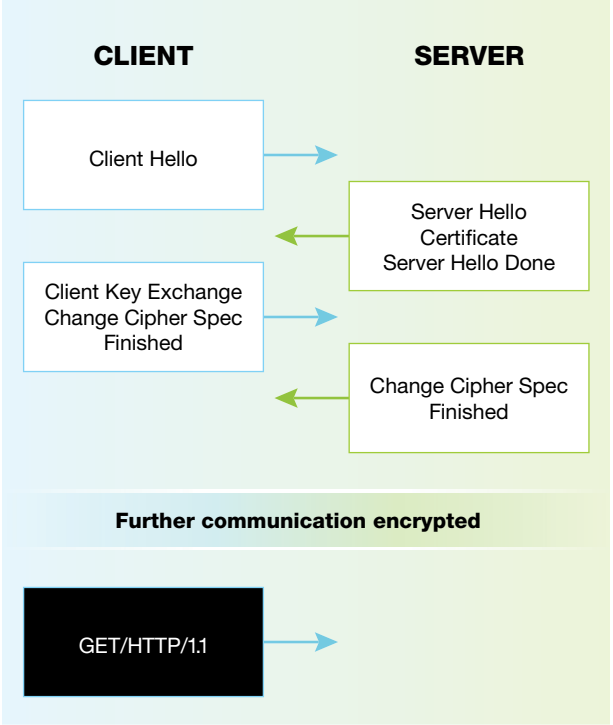
Challenges to SSL security

The release of the Firesheep plug-in in 2010 demonstrated the inherent insecurity of HTTP and how widespread its use was on popular websites that host sensitive personal information. In response, sites like Facebook and Twitter correspondingly embraced HTTPS to bring greater security and privacy to their users. In light of that progress, it was somewhat disheartening to see 2011 host a number of high-profile issues with the SSL and TLS protocols that underpin HTTPS. This article presents a close look at three of the year's prominent incidents affecting SSL/TLS and their impact on the threat landscape.

THC-SSL-DOS

In February (and again with more publicity in late October), a security group, The Hacker's Choice, provided a proof-of-concept tool to perform a denial of service (DoS) against servers communicating via SSL/TLS. The tool showed the potential for an everyday laptop on an average connection to take down an enterprise web server. It worked by exploiting understood asymmetries in the computational resources required to set up cryptography during the TLS handshake.

The TLS Handshake



During a typical TLS handshake, a number of things happen. The client initiates the handshake with a "Client Hello" message, listing a series of cipher suites it can perform. The server then responds with a "Server Hello" message, indicating the cipher suite it has selected for encrypting communication. In the same packet, the server also includes a "Certificate" message that contains the site's certificate which establishes its identity and provides a public key for cryptography.

If the provided information checks out with the client, it then responds with a "Client Key Exchange" message (followed by "Change Cipher Spec" and "Finished" messages). The "Client Key exchange" contains a pre-master secret that is encrypted with the server certificate's public key. Upon receipt of the "Client Key Exchange", the server decrypts the pre-master secret with its private key (and responds with its own "Change Cipher Spec" and "Finished" messages). At this point, both the client and the server can generate their master key and set up their encryption which they use for the rest of the session.

In the handshake (and later symmetric encryption of the traffic), the greatest hit in computational cost is found in the encryption and decryption of the pre-master secret in the "Client Key Exchange" message. Although both client and server use the RSA algorithm, the server has a greater (and, depending on things like the RSA key length, an even greater) computational cost. The specifics of this behavior are interesting but beyond the scope of this article.

The tool could be particularly effective in bringing down a server due to its use of client-initiated cipher suite renegotiation. Built into the TLS protocol is the ability for either end of the encrypted channel to renegotiate the cipher suite used. The renegotiation essentially causes another handshake to occur and with it, the same computational cost. Client-initiated renegotiation allows a single client to cause the server to perform TLS handshakes as fast as the client can request them. Using renegotiation allows for an attack to use a small number of machines and therefore be under the radar for typical distributed denial of service (DDoS) connection thresholds.

Mitigation

The impact of exploitation can be mitigated by a number of things but none of them is a silver bullet. The simplest way is to disable client-initiated renegotiation. 99 percent of sites don't need to support this feature, so it should be disabled when it isn't needed. Due to a previous TLS man-in-the-middle (MITM) vulnerability (CVE-2009-3555), many web servers disable this feature by default.

Not all cipher suites incur the server-side computation cost that RSA does. Unfortunately, not all browsers support these cipher suites and they would be poor choices on low-performance clients like mobile devices. Not supporting RSA is not a viable option, especially considering that it's mandatory for TLS 1.1 and 1.2.

If client-initiated renegotiation is disabled, the computational cost of 10,000 handshakes over a single connection can be achieved with the initial handshake over 10,000 connections. This attack then becomes a traditional DDoS attack. The damage from distributed attacks can be mitigated with the old standbys of IDS and/or throwing in more hardware.

It is worth noting that the computational asymmetry in certain cipher suites will not be fixed in TLS/SSL. The computational asymmetry can be pushed onto the client side via a "client puzzle" that requires the client to do more work but this isn't a generally acceptable solution. As Eric Rescorla pointed out on a blog post about this issue: "…DoS attackers generally use botnets (i.e., other people's compromised computers) to mount their attacks and therefore they have a very large amount of CPU available to them. This makes it very hard to create a puzzle which creates enough of a challenge to attackers to reduce the attack threat without severely impacting people with low computational resources such as those on mobile devices."[4]

The BEAST

On September 23rd, security researchers Juliano Rizzo and Thai Duong demonstrated an attack, for Ekoparty security conference attendees, that decrypted session cookies of a client's HTTPS connection to paypal.com. The tool was called BEAST (Browser Exploit Against SSL/TLS). The attack exploits the long-known weakness in SSL 3.0 and TLS 1.0's use of an implicit Initialization Vector (IV) when using cipher block chaining (CBC) mode. While this vulnerability was understood for a long time, it was mostly considered hypothetical until these two researchers demonstrated how feasible it is.

The previous discussion of cipher suites in the TLS handshake focused on the key exchange algorithm. The cipher suite also defines the bulk encryption algorithm that is used to encrypt the data over the established TLS connection. TLS supports two families of bulk encryption algorithms, stream ciphers and block ciphers. It is the block ciphers that operate in CBC mode.

Block ciphers work by first breaking the plain text into discrete blocks of a fixed size and then encrypting them. In CBC mode, the plain text of a block is first XORed with the ciphertext of the previous block and then encrypted. The first block to be encrypted, however, has no preceding ciphertext and has to substitute the IV. The vulnerability lies in how affected versions of SSL/TLS use an implicit IV of the last block's ciphertext when encrypting a new record.

The attacker has to meet a few requirements to pull off the attack. First, he needs to be able to monitor the client's encrypted HTTPS data. Second, he needs to be able to control parts of the plain text sent from the client over the HTTPS channel, such as a URL path, and also a way to control the plain text of the last encrypted block. The first requirement, while harder to achieve than a drive-by-download, is hardly insurmountable. The TLS/SSL protocol exists because people don't trust the intermediaries on the Internet; also you could just

4 http://www.educatedguesswork.org/2011/10/ssltls_and_computational_dos.html

hop onto an unsecured wireless network. The researchers indicated a number of common technologies like Java and Silverlight that satisfy the second requirement's ability to craft traffic (and include a cookie). In the demo, exploitation of the second requirement was facilitated by leveraging a (now addressed) vulnerability in the Java plug-in's same-origin policy (SOP) checks that allowed an applet in one origin, say http://www.attacker.com, to send requests to another, https://www.paypal.com.

With these, the attacker can effectively decrypt unknown parts of the plain text one byte at a time. Generally speaking, the attacker gets the client to make a request where, for a given block, the attacker already knows all but a single byte of the plain text. The encrypted block on the wire is then intercepted and remembered. He then makes a guess as to what the unknown byte was and places that at the end of a record so that it will be used as the IV for the next record. On average, it will take 126 guesses before the attacker sees an encrypted block that matches what he's looking for. Now, the attacker knows what that byte was and adjusts the next request to decipher the next byte. This continues until the attacker deciphers the session cookie (or whatever else he was looking for).

Mitigation

The problem in using an implicit Initialization Vector has been known about for years and was fixed in TLS 1.1. Unfortunately, few browsers actually support TLS 1.1, so switching servers to using only TLS 1.1 might not be an option. Stream ciphers, like RC4, do not have this problem, so prioritizing a server to use stream ciphers for communication was the only reasonable server side fix for this. A solution for this issue needs to be found on the client side. Unfortunately, there are some SSL/TLS implementations that do not work with the back-ported TLS 1.1 fix. Vendors have been addressing this problem, though the problems with interoperability complicate the issue. Microsoft, for example, released a patch to address this vulnerability in the January 2012 monthly update.

DigiNotar and Comodo compromises

In March, a registration authority (RA) associated with the Comodo certificate authority (CA) was hacked and nine fraudulent certificates for common domains like "*.google.com" and "*.yahoo.com" were issued from a trusted root certificate for UTN-USERFirst-Hardware. This was a security disaster. In order to continue secure communications to these domains, browser vendors like Microsoft, Google, Mozilla and Apple had to immediately release updates to their products to revoke these certificates. Fortunately, according to the Comodo reports, only one of the fake certificates (for Yahoo) was seen live on the Internet.

In mid-July, there was another compromise, this time of the DigiNotar CA. If Comodo was a disaster, this was a security catastrophe. Over 500 fraudulent certificates were issued for domains like "*.google.com" as well as the incredibly broad "*.*.com". The official Fox-IT report on the breach indicated that over 300,000 unique IPs had accessed a fraudulent certificate for Google. The response to this, like the former breach, required browser vendors to scramble to release updates to their products to revoke these certificates as quickly as possible.

Product updates may seem like a dramatic measure to revoke certificates but due to the current mechanisms they are understandable.

Certificate revocation

The necessity for revoking certificates due to fraud or an information update was understood and solutions were created for it. Two methods are commonly used for checking revocation status: certificate revocation lists (CRLs) and the online certificate status protocol (OCSP). Unfortunately, neither of these solutions is really effective.

Using the first method, CRL information can be included in a certificate. When the client is authenticating a certificate, it can download the indicated CRL, download a list of revoked certificate serial numbers, and check to determine if any certificates encountered are revoked. OCSP, on the other hand, is a protocol made so that a client can issue a request for an individual certificate, as opposed to downloading an entire list, to check for its revocation status.

Neither approach works well because implementations default to failing open. If the client doesn't receive a revocation notification, then it assumes the server was down and that the certificate was valid. The obvious problem is that if a MITM can intercept traffic and present an invalid certificate, then he can also probably block any revocation response as well.

More importantly than not having a good method for revoking fraudulent certificates is that these compromises indicate much larger problems in the SSL trust model itself.

SSL trust model

Two primary SSL and TLS goals are to establish authenticity and secrecy of communication. To provide authenticity and prevent a MITM from impersonating a server, SSL was designed with the notion of certificates and certificate authorities (CAs). If a site wants to provide HTTPS, it requires a certificate which it requests from a certificate authority. The CAs are trusted entities, represented by companies such as Verisign, Thawte, Comodo, or DIgiNotar, whose job is to verify that the site is who they say they are, and then issue a certificate to represent this fact.

Web browsers can then come pre-installed with certificates for these trusted authorities so that when an arbitrary certificate is presented, the browsers can check the certificate's validity against their trusted certificates. However, going back to the MITM case, an attacker can create a bogus certificate for a site and then present it to a client trying to connect to a legitimate site. Since the attacker's certificate was not signed by a trusted certificate authority, the client's browser presents a warning and the client then knows it has not reached the authentic site.

Problems with the SSL trust model

This model has a few problems. In the system, all CAs are treated as equals. A certificate issued by one CA is just as valid as one issued by any other. As an example, a certificate issued for "*.google.com" from a random CA is just as valid to a browser as one issued from Google's actual registrar.

According to the Electronic Frontier Foundation's (EFF) SSL observatory project, there are over 600 entities that can issue certificates. With that many companies, it's not surprising to find out that there are differing levels in quality in site security as well as levels of verification of a requested certificate. CAs don't have to be hacked for a certificate to be issued to the wrong party. It has happened in the past and will undoubtedly continue in the future.

Another consequence of the system is that once a CA is trusted, it stays trusted. CAs can issue certificates for millions of sites. If you decide you no longer trust a CA, and you remove its certificate from your store, then all of the sites signed by that CA are no longer available over HTTPS.

Revising SSL trust

Solutions have been proposed to address these problems. There are proposals for using DNS to handle the trust, such as DNS-based Authentication of Named Entities (DANE) and Certificate Authority Authorization (CAA). Both of these proposals allow for information about authorized CAs to be embedded into a domain's DNS record. A CA can use this information to check who is supposed to be issuing certificates for a given domain as a means to help prevent inadvertent issuing to incorrect parties. Clients can use this information to verify that a certificate is issued by an appropriate CA.

DANE also allows for an alternate to CA trust. A site can embed certificate information into its domain record. When a client connects to a site, it can compare the domain certificate to the provided certificate. If the certificates match, the site can be assumed to be authenticated. The shortcoming with these approaches is that both CAA and DANE require Domain Name System Security Extensions (DNSSEC) in order to be secure and DNSSEC is still not widely utilized.

The case for an alternative to the current trust mechanism and the use of DNS was proposed by Moxie Marlinspike[5] in a blog article and later presented at BlackHat USA. He pointed out that the use of DNS doesn't make a certificate more trustworthy. If a client is concerned that certificates issued by a government's CA is being abused to eavesdrop on traffic, how does the situation improve when relying on certificate data from that same nation's DNS servers?

Marlinspike called for the necessity of "trust agility" in SSL. The two core tenets are that a client can revoke the trust of an agency at any time and can choose where to anchor its trust. Consequently, Marlinspike developed a plug-in for Firefox called "Convergence" that implements these requirements by allowing a user to choose multiple "notaries" which validate a site's certificate, resulting in a flexible system where notaries are free to impose their own security requirements on validation and users are free to choose which notaries they trust.

5 http://blog.thoughtcrime.org/ssl-and-the-future-of-authenticity
http://www.youtube.com/watch?v=Z7Wl2FW2TcA

In the current CA model, once a CA is trusted it's essentially trusted forever. CAs can issue millions of certificates used on websites throughout the Internet. As an example, if at some point you decide not to trust Verisign and you remove their CA certificate from your store, then any site with a certificate issued by them, even one issued in the past when you still trusted them, won't be accessible over HTTPS. Contrast this to trust agility using Convergence where you may stop trusting a given notary, but other, trusted notaries could still vouch for the security of the site's certificate. The notary system seems like a great idea, particularly for users, however it's uncertain how organizations would be motivated into starting and sustaining themselves without a clear financial incentive.

Another solution that also provides trust agility is to extend TLS/SSL to support multiple certificates. If a site could provide multiple certificates signed from different CAs, such as one from DigiNotar and one from Verisign, and then you decided not to trust DigiNotar, you would still trust Verisign and therefore establish a secure connection. A large impediment to this solution is that it requires a change to the TLS protocol and vendor adoption of new protocol versions is slow.

What does the future hold?

SSL was initially developed in the early 1990s to secure communications to a handful of sites. It is now at version TLS 1.2 and protects over two million websites. TLS 1.1 has yet to see widespread adoption but as problems previously considered to be theoretical have proven to be a reality, there will likely be a more rapid acceptance of future versions. Hopefully, there will be more widespread deployment of newer cipher suites like ECDHE_RSA on clients and servers which can provide forward secrecy so that a leaked certificate private key can't be retroactively used to decrypt previously recorded traffic. It is almost inevitable that the current trust model for SSL will have to be changed but such a change to an entrenched and flawed aspect of the protocol is likely to be too long in coming.

The emergence of Mac malware

Introduction

More than in any previous year, 2011 has seen the most activity in the Mac malware world.[6] This applies not only to volume, but also in functionality. In 2011, we started seeing Mac malware with functionalities that we've only seen before in Windows malware. This may indicate that cyber criminals are now becoming aware of how profitable targeting OS X might be.

Let's take a look at some of the more noteworthy Mac malware discovered in 2011.

MacDefender

MacDefender was first discovered in May 2011, with subsequent variants (named MacSecurity, MacProtector, MacGuard, and MacShield) discovered in the months that followed. What makes MacDefender interesting is that it is the type of malware with a spreading mechanism that has been rampant in the Windows world in the last couple of years. MacDefender

6 http://blog.intego.com/dl/Intego-year-in-Mac-security-2011.pdf

belongs to the category of malware called “Rogue Antivirus,” which disguise themselves as legitimate antivirus programs. Once installed, it pretends to scan your system, flagging random files as malicious to make it look like your system is heavily infected.

The user interface is professional looking and well made to make it more believable to the user that it is a legitimate app. The user interface contains a Register button that will take the user to a website where they can supposedly purchase a license for MacDefender using a credit card. MacDefender displays a message that says to remove the detected malware, you should pay for the licensed version, so a user may feel forced to register. The user’s credit card will then be charged for the amount and on top of that, his credit card number may be used for other purposes as well.

Figure 12: MacDefender malware screenshot 2011

MacDefender and its variants are spread by targeting users through SEO poisoning attacks, wherein malware authors manipulate search engine results to make their links, which hosts the malware, to appear near the top of search results. When a user clicks on one of these links, Javascript downloads the MacDefender installer into their system. If the browser setting was set to automatically open safe files after downloading, the installer opens automatically.

Rogue antivirus is a highly profitable scam, so X-Force believes that we are going to see more of this type of malware in the future. Users should be wary about clicking links in search results. Before you click on a link, check if the domain name of that link is related to what you’re looking for. Also, do not install software unless you are sure that it is from a reputable source.

Flashback

Flashback is a Trojan that was discovered in September 2011. Variants of this malware appeared in the months that followed, each with various improvements on the original. Flashback disguises itself as a Flash Player installer that can be downloaded when visiting malicious websites, showing a download or install Flash player icon.

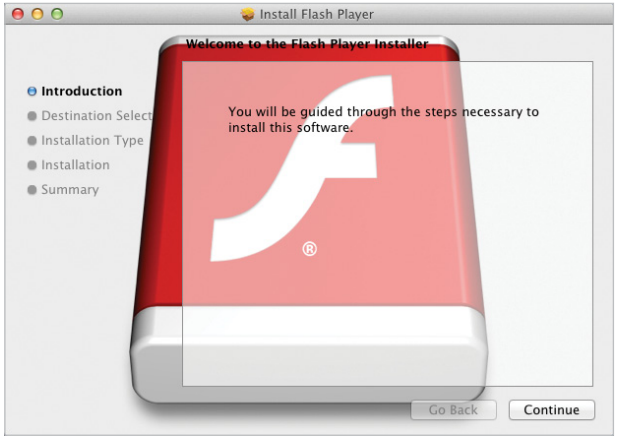


Figure 13: Flashback trojan screenshot 2011

When installed, Flashback drops a dynamic shared library file and uses the DYLD_INSERT_LIBRARIES environment variable to inject code into the application launched by the user. Later variants target specific applications, such as Safari and Firefox, to inject into. The injected code is responsible for contacting a remote server to download updates or to send data from the infected machine. This code injection technique is similar to what some notorious Windows malware like Zeus does when injecting code into web browsers. Zeus intercepts web pages passed from the server to the web browser and modifies them on the fly before showing them to the user. The modified web page typically shows a fake login page, allowing the malware to steal sensitive information. Fortunately, no web injection functionality has been observed in any of the variants of Flashback thus far.

Flashback also tries to prevent future updates to XProtect by overwriting some relevant files. XProtect is Apple's built-in basic malware protection system that uses string matching to detect malware. Apple updates XProtect whenever a high-profile Mac malware is discovered.

Flashback also tries to thwart analysis by researchers by detecting if it is running on a VMWare virtual machine. Using this detection evasion mechanism is common in Windows malware but this is the first Mac malware we've seen that employs this technique. This demonstrates that Mac malware technology is catching up to Windows malware technology.

DevilRobber

DevilRobber is the latest OS X malware to make news in 2011. It was discovered in October 2011 with variants released in the months of November and December. DevilRobber was discovered inside Mac applications that were illegally shared in BitTorrent, such as GraphicConverter, Flux, CorelPainter, and Pixelmator.

DevilRobber is the most sophisticated Mac malware we've seen so far and contains several components. It is primarily a backdoor that opens a port in the infected machine to receive commands from a remote attacker but one interesting functionality it has is BitCoin mining, where it installs the BitCoin mining application DiabloMiner to use the computing power of the CPU and GPU (for users with high performance graphics cards) of the infected machine to mine for Bitcoins. It also attempts to steal the Bitcoin wallet if found. DevilRobber also steals the Keychain of the user along with other information from the infected machine and uploads them to a remote FTP server.

DevilRobber also has the ability to detect if the infected machine is behind a gateway device, and then enable port-forwarding via UPnP. This enables the attacker to remotely access the infected machine using the port opened by DevilRobber, even if the infected machine is behind a gateway device.

Conclusion

As you may have noticed, none of the malware just mentioned use any software vulnerability exploits to spread. We speculate that this is due to the lack of exploits for OS X that are publicly available for reuse. Most Windows malware that use exploits often reuse publicly available exploits, such as those found in exploitation frameworks like Metasploit, with minor modifications. However, there are fewer publicly available exploits for OS X. This may be due to lack of interest in developing exploits for a platform with a

relatively low market share, or because of the lack of available technical information to do so. The barrier of entry is higher now with the recent security improvements in the latest version of OS X. OS X Lion implements full ASLR, 64-bit processes by default, and a sandboxing framework. Starting on June 2012, Apple will also require all applications submitted to the Mac App Store to have sandboxing enabled, thus mitigating exploitation attempts through third-party applications.

This is not to say that Mac users should be complacent. As shown in the examples above, malware authors will find alternative means of delivery. Also, these improvements focus on exploit prevention and mitigation and don't really address the types of malware we mentioned above, so we expect to see more Mac malware in 2012.

On the other hand, Apple is certainly taking steps to further increase the cost of developing malware for OS X. In the recently announced next version of OS X, Mountain Lion, they have added a new feature called Gatekeeper. Gatekeeper allows the user to choose which applications can be installed and run on their system based on where the application came from. Users can choose to allow applications from the App Store only, or from both the App Store and from identified developers (applications with an associated Apple Developer ID). They can also disable this feature if they choose to. By default, only applications from the App Store or from identified developers can be installed or run. We believe this will go a long way in preventing large-scale and long-term malware outbreaks.

As attackers take note of OS X, so do security vendors. As such, X-Force predicts that the next wave of Mac malware will employ ways of evading detection and analysis. Surprisingly, most of the Mac malware we've seen so far did not bother with any evasion mechanism. We predict that techniques common in the Windows world such as packing, anti-debugging, and virtual machine detection will be used more. We also expect to see more advanced techniques that work well in Windows malware being adapted by Mac malware, such as Zeus-style web injection and stealth technology such as root kits. New malware will eventually have to deal with the aforementioned Gatekeeper as well, so we may see malware that will attempt to circumvent it in some way.

The number still pales in comparison with what we see from Windows malware, but it is clear that the attackers are starting to notice that Macs are becoming viable targets. Mac users should be aware that malware previously seen only on Windows is also possible in OS X.

Web content trends

The IBM Content security team constantly reviews and analyzes new web content data and 150 million new web pages and images each month. We have analyzed 16 billion web pages and images since 1999.

The IBM web filter database has 68 filter categories and 70 million entries with 150,000 new or updated entries added each day.

This section provides a review of:

- Analysis methodology
- IPv6 deployment for websites
- Increase in the amount 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 Systems 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.

IPv6 deployment for websites

As IPv4 is running out of space, we expect that more and more Internet sites are switching to IPv6. However, when looking at the last five months, this expectation was not met. To measure the IPv6 deployment for websites, we have done DNS requests (check for an AAAA record in DNS) for millions of hosts every month.

The percentage of domains having at least one host supporting IPv6 remained relatively flat and ranged between 2.2 and 2.6 percent.

It will be interesting to see whether there is a significant increase of IPv6 support on the next World IPv6 Day[7] on June 6, 2012, when many companies and organizations plan to implement permanent IPv6 deployment.

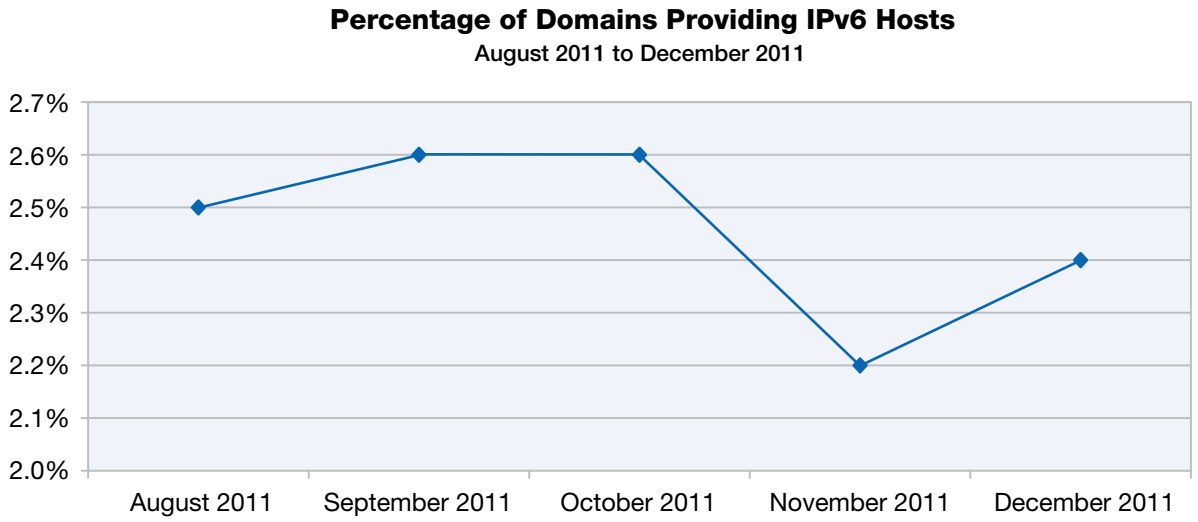


Figure 14: Percentage of Domains Providing IPv6 Hosts – August 2011 to December 2011

7 http://en.wikipedia.org/wiki/World_ipv6_day

Increase of anonymous proxies

As the Internet becomes a more integrated part of our lives at home, at work, and at school, organizations responsible for maintaining acceptable environments in these public settings increasingly find the need to control where people can browse.

One such control is a content filtering system that prevents 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 would have normally 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 four times as many anonymous proxies registered compared to three years ago. In the second half of 2011, there were still more than three times as many anonymous proxies registered compared to three years ago. However, this is the first time since the beginning of 2009, that we did not see another increase of this volume. Perhaps internet activities are more focused on social networks. In many cases, these sites are not blocked at work or in schools so people no longer need to circumvent the content filtering system.

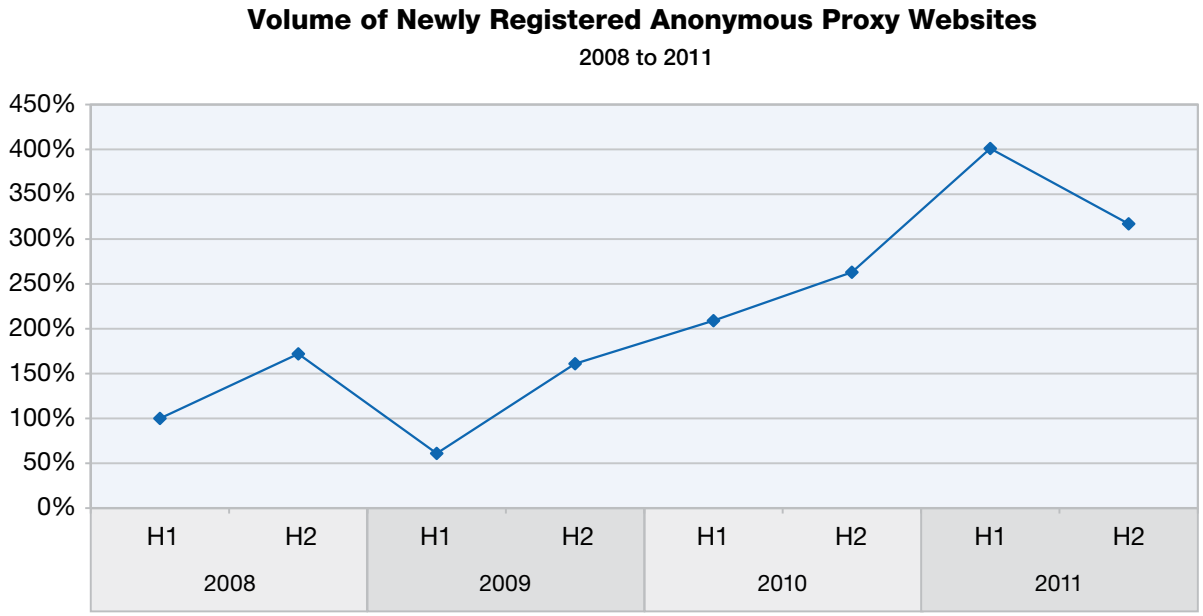


Figure 15: Volume of Newly Registered Anonymous Proxy Websites – 2008 to 2011

However, the use of social networking platforms issues new challenges, particularly for companies that need to control what information is shared with other users and to prevent the sharing of confidential information. Thus, many companies are starting to use web application control systems, often as a part of next generation firewalls.

Anonymous proxies remain a critical type of website to track, because of the ease with which proxies allow people to hide potentially malicious intent.

When looking at the top-level domains of newly registered anonymous proxies, the trend of the first half of 2011—as reported in detail in the IBM X-Force 2011 Mid-year Trend and Risk Report—has continued. The .tk and .com domains continue to prevail, representing more than 70 percent of all new anonymous proxies.

Malicious websites

This section discusses the countries responsible for hosting malicious links along with the types of websites that most often link to these malicious websites. The Vulnerability disclosures in 2011 section contains more information on malicious websites in the exploit context.

Geographical location of malicious web links

The United States continues to reign as the top hosts for malicious links. More than one-third of all malware links are hosted in the U.S. The runner-up is Romania, hosting 8.5 percent. China has been in the top two for the last three years. China is now tied with France for third place, claiming 5.7 percent as shown in figure 16.

The second-tier countries have also shifted, but these changes are less than one percent between 2010 and 2011 figures.

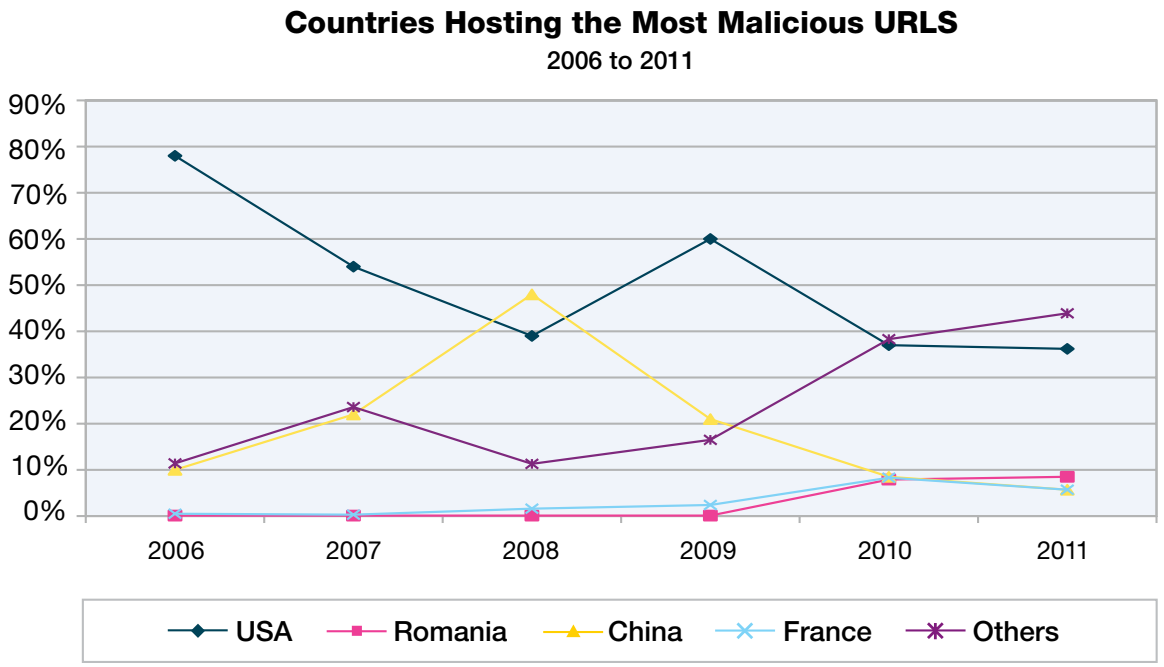


Figure 16: Countries Hosting the Most Malicious URLs – 2006 to 2011

Good websites with bad links

As reported in previous IBM X-Force Trend and Risk Reports, attackers are focusing more and more on using the good name of trusted websites to lower the guard of end users and hide their attempts with 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 content.

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

Search engines, blogs, bulletin boards, and personal websites fall into this second-tier category. Most of these websites allow users to upload content or design their own website. In other words, it is unlikely that these types of websites are intentionally hosting malicious links.

The following chart shows the history of the distribution of malware links.

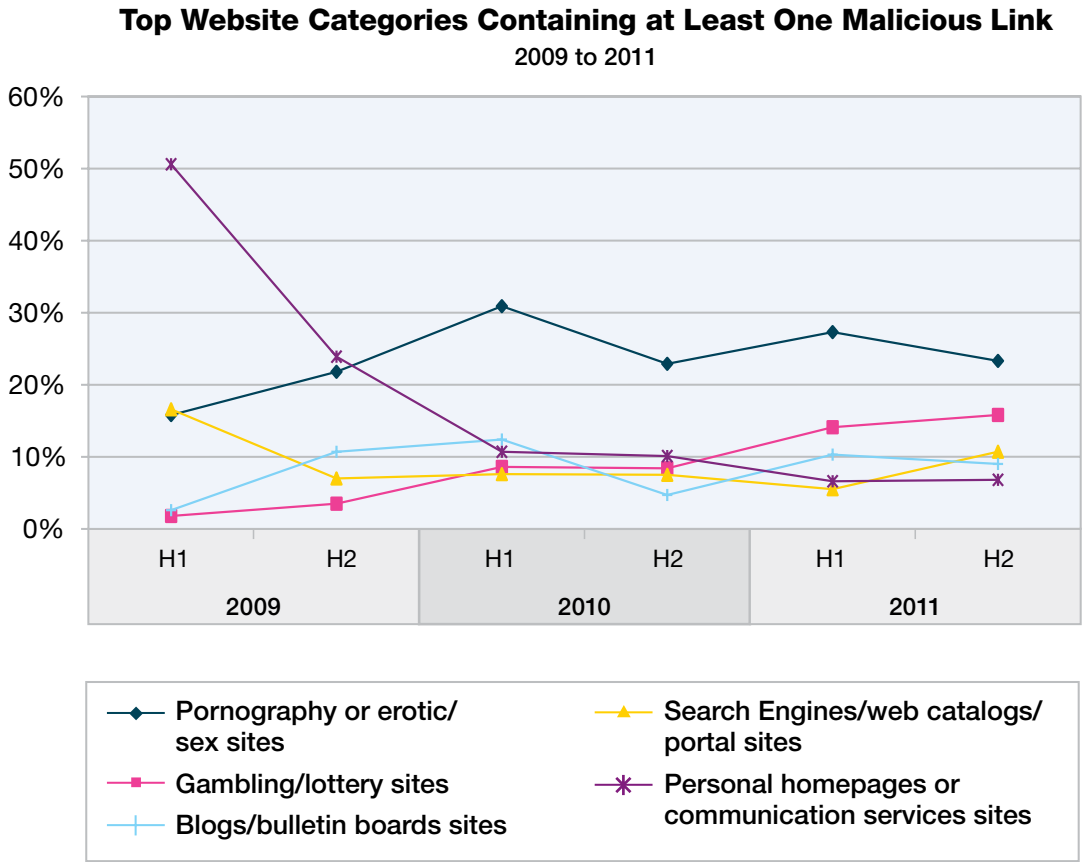


Figure 17: Top Website Categories Containing at Least One Malicious Link – 2009 to 2011

When looking back at the last three years, interesting trends appear.

- The professional "bad" websites like pornography and gambling now clearly dominate the scene and systematically distribute malware.
- Pornography is on the top, stable at about 23 percent.
- Gambling is the only category with a significant term over term increase. Against the background of 0.6 percent of the adult population having problems with gambling issues[8] gambling sites are a popular target for malware distributors.
- Blogs/bulletin boards have decreased to nine percent within the last six months.
- Personal homepages– 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.
- Search engines, web catalogs, and portals sites recovered and reached more than 10 percent for the first time in two and a half years.

8 http://en.wikipedia.org/wiki/Gambling_addiction#Prevalence

Spam and phishing

The IBM spam and URL filter database provides a world-encompassing 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 continues to decline
- Major spam trends in 2011
- Common top-level domains in URL spam
- Spam—country[9] of origin trends
- Email scam and phishing
- Flashback and future prospects on spam

Spam volume continues to decline

In the last Trend and Risk Report, we provided in-depth details on how spam continues to decline in the past months and even years. We believe this is due to several botnet take downs, as discussed in previous reports. You can see in the following chart how those overall numbers continue to decline.

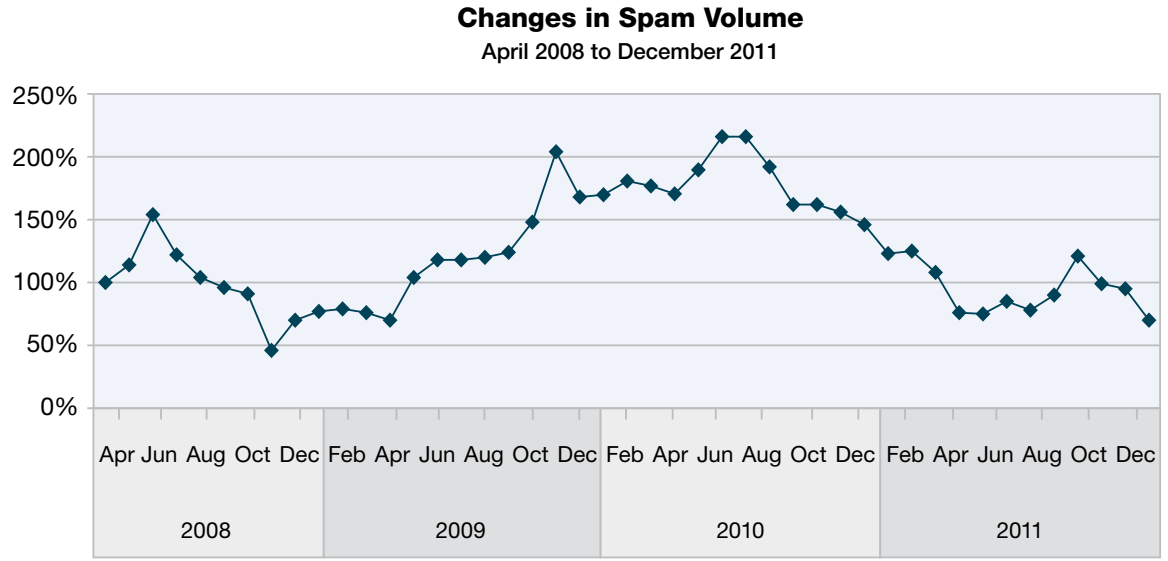


Figure 18: Changes in Spam Volume – April 2008 to December 2011

9 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.

Major spam trends 2011

The following chart summarizes the major trends in spam we observed in 2011.

One can derive several changes concerning aspects of the spam sent out in 2011. We have defined several phases to highlight these changes:

- **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:** May 19, 2011 until August 22, 2011
- **Phase 5—Second recovery of spam volume:** August 23, 2011 until November 29, 2011
- **Phase 6—Year-end decline of spam volume:** Since November 30, 2011

We have discussed the phases zero to four in detail in the IBM X-Force 2011 Mid-year Trend and Risk Report. The new phases five and six were dominated by ZIP or RAR Malware spam (again) and image-based spam, as discussed in the next two sections.

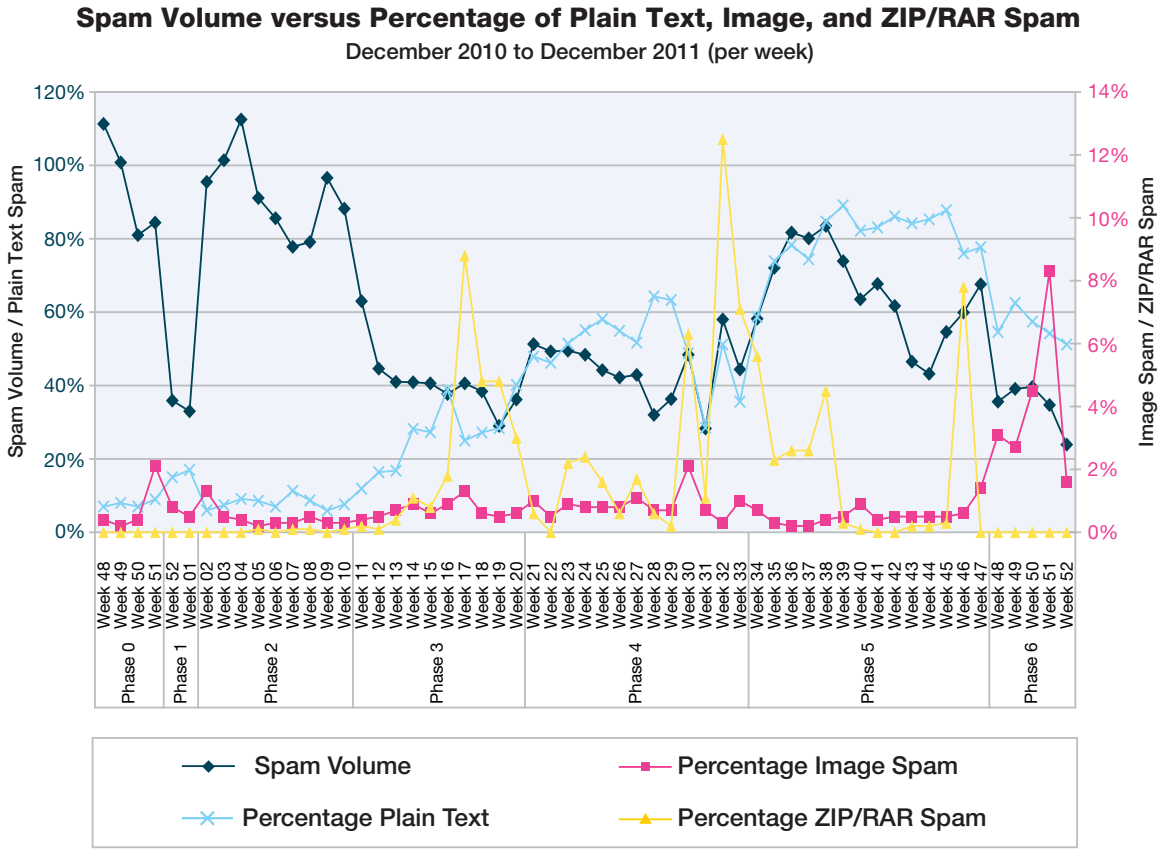


Figure 19: Spam Volume versus Percentage of Plain Text, Image, and ZIP/RAR Spam – December 2010 to December 2011 (per week)

When looking at the whole time frame, the nearly continuous increase of plain text spam is particularly significant. In previous years we have seen between five and 30 percent of spam written in simple plain text. This is the first time that we observed these high values—sometimes more than 80 percent in phase five—over a longer period of time. During phase six, it declined to around 55 percent.

Spam in plain text makes it even harder for content-based spam detection because there is no fixed feature like a special kind of attachment or suspicious html code sequences that can be used to build patterns. However, the trend in legitimate email is reversed. There are only a few remaining types of status messages or newsletters that do not use html. Sooner or later, simple plain text spam as an email characteristic becomes more and more suspicious. Someday it might even be used as a blocking criterion.

2011's Malware ZIP spam

The ZIP attachments of spam in phase three were discussed in detail in the IBM 2011 Mid-year Trend and Risk Report.

In the second half of 2011, we saw three spikes of emails with ZIP attachments between 18 and 43 percent, each measured on a daily basis. Trojans are the favorite type of malware attachment. More than 50 percent of ZIP attachments during the peak at the end of July contained the Trojan:Win32/Fivfrom.gen!B. To encourage users to open those attachments and click on the malware binary, spammers used several variants similar to those used in phase three. One of the major ones was a 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.

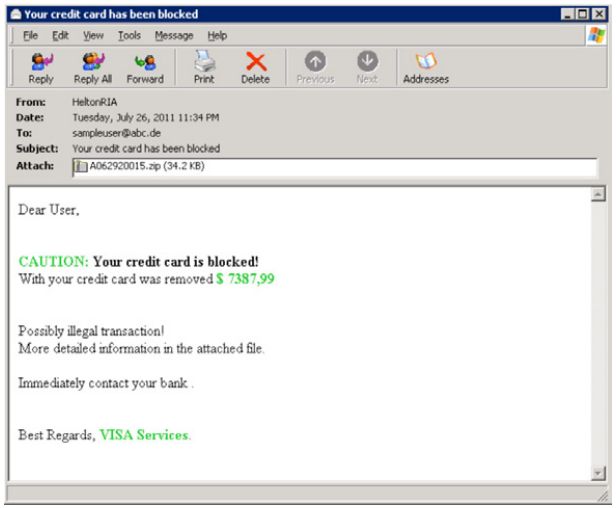


Figure 20: Faked message about charged credit card – July, 2011

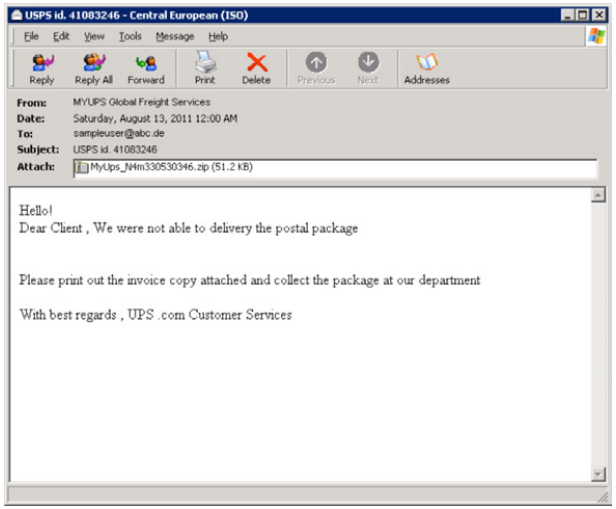


Figure 21: Faked delivery notification – August, 2011

We could see a similar picture during both of the other spikes: the dominant type of malware in mid-August was TrojanDownloader:Win32/Cbeplay.M. Two weeks after this peak, the percentage of ZIP attachments was about 10 percent per day. Typical of this type were faked delivery notices from a well-known parcel service to try to convince the user to open the attachment and to click on the binary.

The third peak was around the 20th of September. The dominant type of malware was the TrojanDownloader:Win32/Chepvil.N.

Image spam in 2011

The rebirth of image spam was a bit surprising. During the last two years we did not see major amounts of this type of spam. On most days the percentage of this type of spam was below one percent. However, since the end of November we have seen major spikes in these statistics.

Previous image spam used the image to transport the actual spam message, e.g. showing some pills or displaying the URL and requesting the user to type that into his browser. There still exist a few of these old-styled image spams, but the majority of the latest image spams have been logos of legitimate organizations or companies. The text of the email states something similar to:

- Your transaction failed, please click on the link to see the details.
- We have received a complaint about your business, please click here.

The actual purpose of using these logos is to make users click on the provided link—a malware link that infects the user's machine. This type of email looks like phishing. Please see the "Email scam and phishing" section for further details on this type of spam.

It will be interesting to see the other approaches that spammers might use in 2012 to get users to click on malicious links.

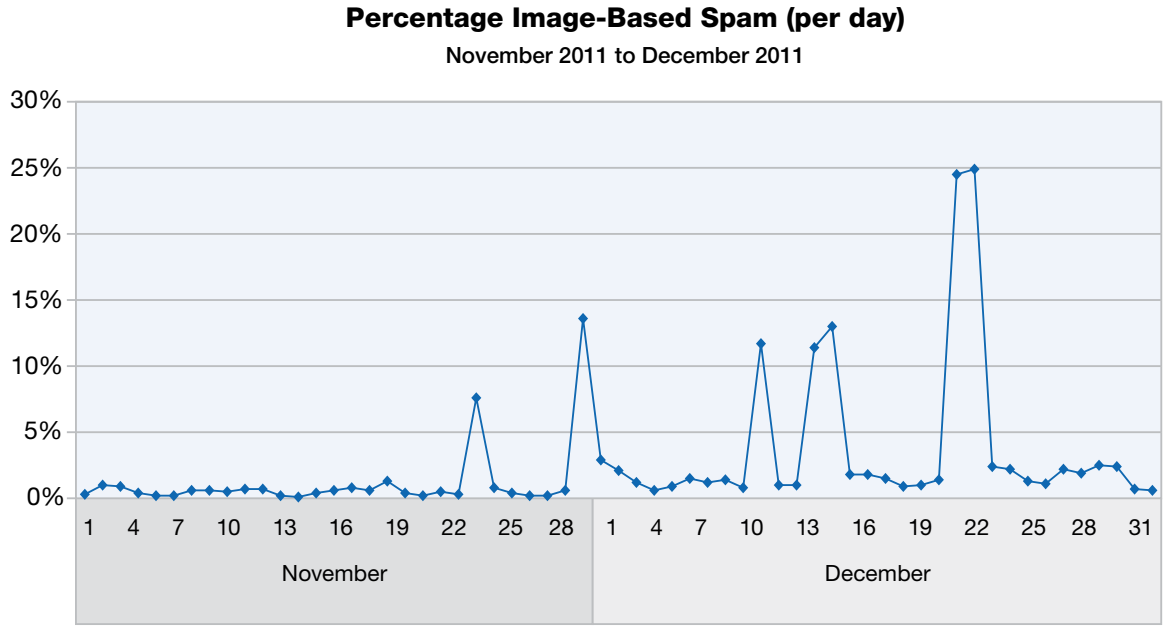


Figure 22: Percentage Image-Based Spam – November 2011 to December 2011 (per day)

Common top-level domains in URL spam including long-term trends

Top-level domain use by spammers in 2011 was similar to that in 2010. The one exception was .ua, the top-level domain of the Ukraine. This domain was used to get new content placed on the Internet. Spam and phishing always intend to get the user to click on the provided link. It is worth looking at the long term trends of the top-level domains used by the bad guys. The last four years have brought major changes.

- The most used top-level domain from 2008 through 2011 is .com, always staying in either the first or second position.
- The other generic top-level domains .net, .info, and .org remained popular for spammers over the years. However, they significantly declined in 2011.
- Since the beginning of 2010, .cn (China) significantly declined and never returned to the top 15.
- .cn was replaced by .ru (Russia) that entered the top 15 in 2008 and since 2010 alternates with .com as the top position.
- Newcomer in 2011 is .ua (Ukraine), staying in third place since spring of 2011.

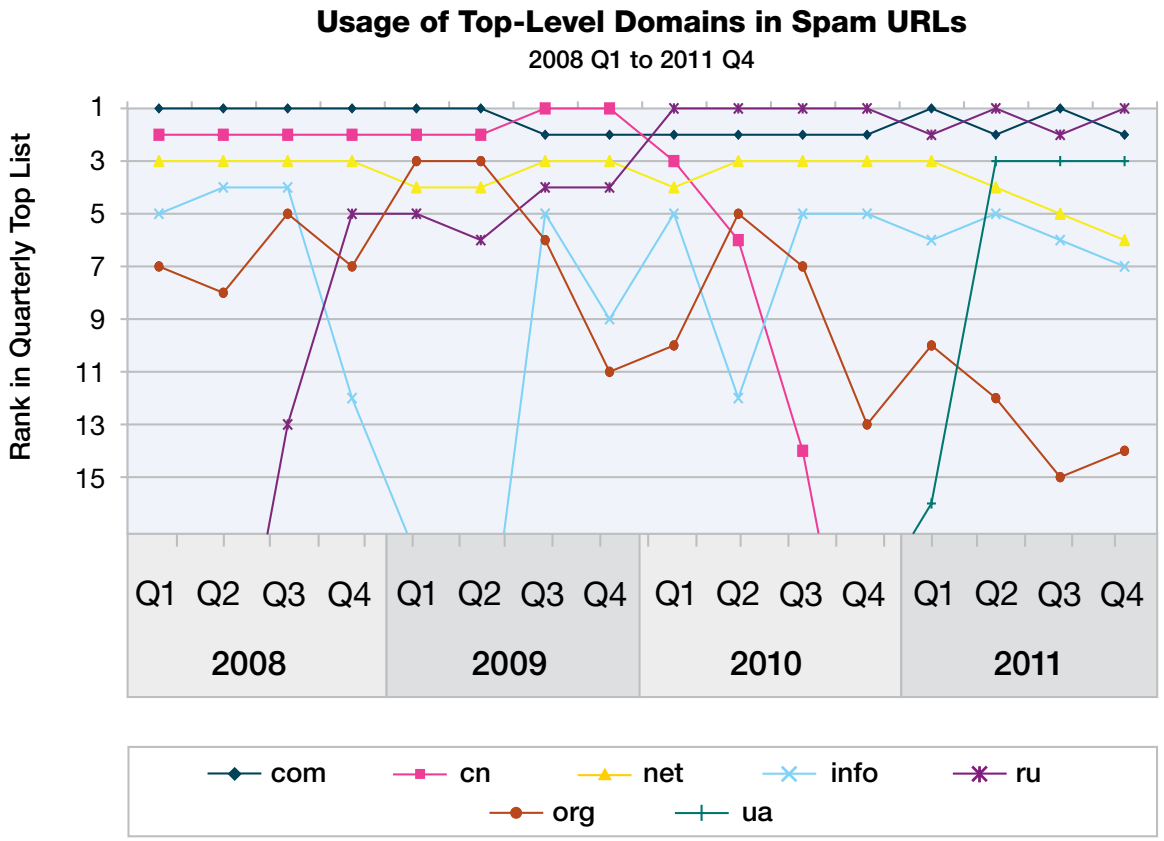


Figure 23: Usage of Top-Level Domains in Spam URLs – 2008 Q1 to 2011 Q4

Some interesting questions occur based on these long term statistics:

- **Why is .com so popular for spammers?** The .com domain is by far the most used top-level domain in the internet. A .com domain is cheap and easy to register. Furthermore, a .com URL in an email is totally unsuspicious.
- **What happened to .cn (China) top-level domain?** In 2008 and 2009 Chinese domains were the favorite domains of spammers. However, since China has tightened the rules on registering a .cn domain[10] as of mid-December 2009, this appears to have deterred spammers.
- **Why doesn't Russia (.ru) do the same as China?** They tried. On April 1st, 2010, the Russian NIC also tightened their rules to register new domains.[11] Eighteen months later, they tightened the rules again.[12] However, spammers continue to choose .ru domains to provide their offers. Currently, .ru is still the topmost used country code top-level domain used for spam.
- **As there are only a few top-level domains widely used for spam, wouldn't this be a point of action to fight spam?** Yes and No. If there would be a concerted action by the registrars to apply the same rules as China does, this could help. However, this is an unrealistic expectation. Registration is a legal issue that each country handles differently. It is likely that there will always be a loose registrar out there that provides open doors for spammers. Also, registering domains is only one way to get spam content hosted. Another way is to use other content hosts without the need to register domains.

10 http://www.cnnic.net.cn/html/Dir/2009/12/12/5750.htm
11 http://news.softpedia.com/news/Enhanced-Security-Measures-for-RU-Domain-Registrations-138234.shtml
12 http://www.abuse.ch/?p=3581

Spam—country of origin trends

When looking at the countries that sent out the most spam over the last three years, some interesting long-term trends become visible.

- Three years ago Brazil and the U.S. dominated the marketplace.
- India has shown nearly continuous growth and now dominates the scene by a large margin, sending out more than 14 percent of all spam.
- The USA owned the top position one year ago and now sends only two percent of all spam.
- Vietnam which was rising in 2009, has significantly declined in the first quarter of 2011, but recovered considerably in the second half of 2011, sending more than 10 percent of all spam.
- Brazil has halved its percentage within the last 18 months.
- Indonesia is the newcomer. It has shown a continuous growth for three years and now generates 10 percent of all spam.
- Australia is another newcomer, responsible for 5.6 percent of all spam by the end of 2011.

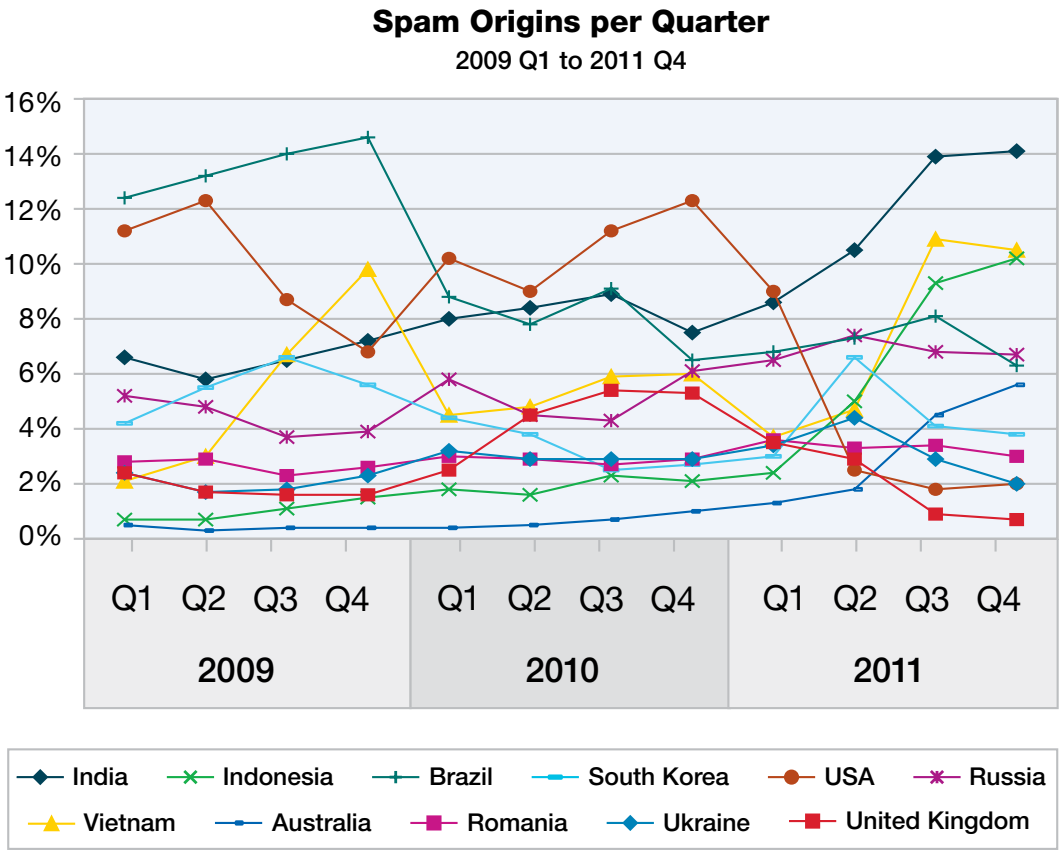


Figure 24: Spam Origins per Quarter – 2009 Q1 to 2011 Q4

Email scam and phishing

Methodology of the provided scam and phishing statistics

As reported in earlier Trend and Risk Reports, we have seen declines of traditional email phishing in 2010 and in the first half of 2011.

However, this industry is not dead. Traditional email phishing has been replaced by some new approaches the bad guys are using, but the differences are not obvious. We still see a lot of spam that looks like normal phishing, such as:

- Emails that look like they are sent from banks that ask users to click on the provided link to update the account, confirm their data, and so on.
- Emails that look like they are sent from social networks regarding a new friend request that can be confirmed by clicking on the provided link.

In the context of traditional email phishing we notice a major change after some time. Many of the phishing pages contained in the phishing emails are no longer placed on a newly registered domain. The advantage of these domains is that phishers are able to choose a domain name that is similar to the phishing victim, e.g. hxxp://www.<banknamewithsmallspellingmistake>.com. Phishers took advantage, sometimes in combination with the new internationalized domain names.[13] One counteraction was to shut down such domains quickly. New domain shut down services were created around these phishing sites.

Ever clever and trying, the phishers have found other ways to circumvent this shutdown issue. Today many phishing pages are placed as sub-pages of legitimate websites, such as hxxp://www.<legitimatesite>.com/<anyword>.html. The advantage for phishers is that the domains of these sub-pages cannot be shut down as they belong to legitimate websites, in some cases even business-related websites. In order to put this sub-page in place, the legitimate website is attacked. Once in, the phishers can simply place the additional page—that consists only of a few kilobytes—on the web server. Third, they send out their phishing emails containing a link to this new sub-page. Phishers collect the credentials entered by the users on this sub-page, that—as usual—looks like the expected banking login site.

Phishers could even make it more difficult for the security guys to detect these pages using this approach by presenting the page only to a certain percentage of users clicking on this link.

However, the maybe more surprising thing is that not of all these phishy looking emails provide a link to such a phishing site. Instead, there is often:

a) An online shop for medical products, fashion accessories, or software identical to the links provided by normal spam.
b) Malware that can infect your computer when clicking on the link.

So why do phishers change their approach to something that looks illogical, especially in case (a). Some reasons might be:

- It is a well-proven approach to get users to click on a link when the email looks like it is coming from a legitimate organization such as banks or social networks. Thus, it is simply click fraud.[14] It's possible that these websites are paying these phishers to advertise their sites and they are unaware of how that advertising is working.
- It is too much work to set up fake banking sites that might be blocked by security products within a few minutes. It is much cheaper and more convenient to install a Trojan at the user's computer, because the Trojan can capture banking credentials independently from the user's main bank.

13 http://en.wikipedia.org/wiki/IDN_homograph_attack

14 Click fraud is a kind of Internet crime in the context of pay per click (see http://en.wikipedia.org/wiki/Pay_per_click) online advertisement. The fraud is done by imitating or provoking clicks on advertisements. Each click generates a charge. In contrast to a user who has a real interest in the target of the ad's link these clicks do not have any interest, thus, the charge is paid without reward. For more details please see http://en.wikipedia.org/wiki/Click_fraud

- Selling counterfeit medical products, software, and fashion accessories is still a profitable business and some users might not think about why such an online shop appears when clicking on a link provided in an email from a bank or a social network. Thus, it is only one method among many to get users on their online shopping sites.

There is another mathematical and statistical consequence from these recent phishing trends that we have seen, particularly in 2011. Much of the spam that looks like phishing emails are normal medical or malware spam. However, in many statistics they are counted as phishing emails. This is not necessarily a mistake because, in the case of provided malware links, the data-stealing malware might phish for credentials and therefore, one can still consider it correct to call this spam phishing.

The boundaries between normal spam, phishing, and malware spam become more and more blurred. Other facets can have a major impact on phishing statistics including:

- This section only considers phishing coming from normal email. It does not include phishing messages from within social networks.
- The provided statistics count the absolute number of received phishing emails. Compared to 2008, these numbers declined until mid-2011. On the other hand, there are many reports about an increase of phishing. This is not a conflict because this represents the number of attacks. There might be many more attacks but each attack consists of fewer emails. In the case of spear phishing, (see sidebar) there might be only a single email.
- The provided statistics do not count spam with malware attachments or malware links where its text does not relate to the targeted brand, even if the malware is targeting your banking credentials.

Thus, there are many aspects that might result in different phishing statistics.

The following statistics include this “phishing-like” spam because of the aforementioned aspects. It is interesting to measure and analyze which kinds of brands the bad guys are abusing to get users to click on their bad links. A generic term for these fraudulent emails is “scam.”

Spear phishing
Spear phishing is phishing that is personalized. At first, phishers gather many kinds of personal data by applying social engineering. In the second step, 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.

Latest trends in email scam and phishing

When we take the aforementioned methodology into account, we see a significant decline of traditional email phishing, particularly in 2010. However, in the second half of 2011, the trend of using the names of trusted brands to make users click on the provided link resulted in a significant increase of these phishing-like emails or scams respectively.

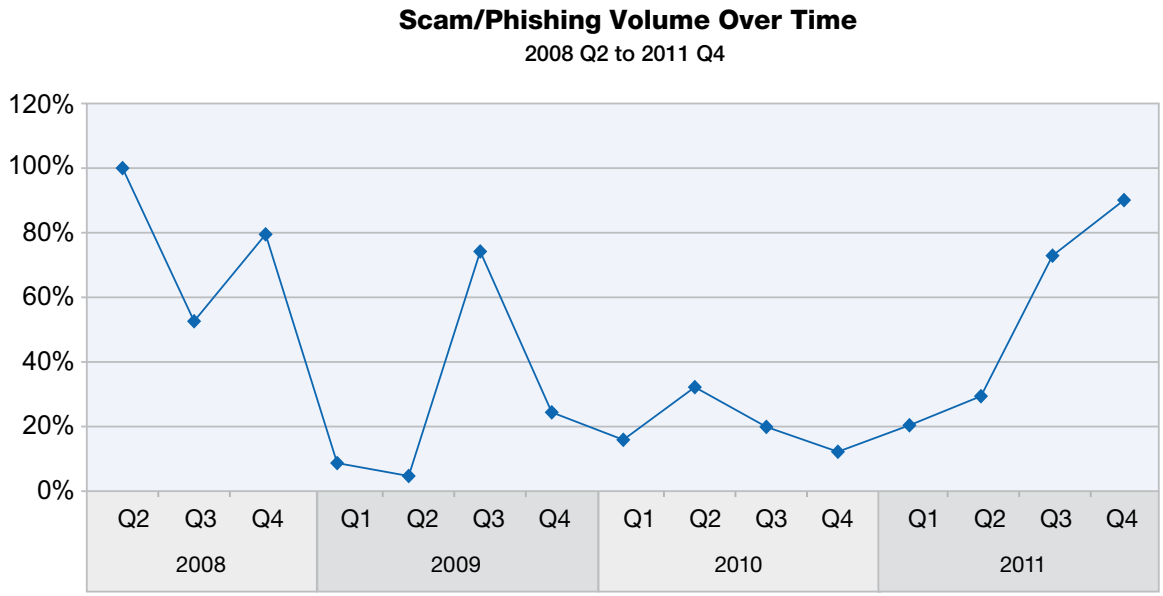


Figure 25: Scam/phishing Volume Over Time – 2008 Q2 to 2011 Q4

The following map shows from which countries the phishing-like emails are sent.[15]

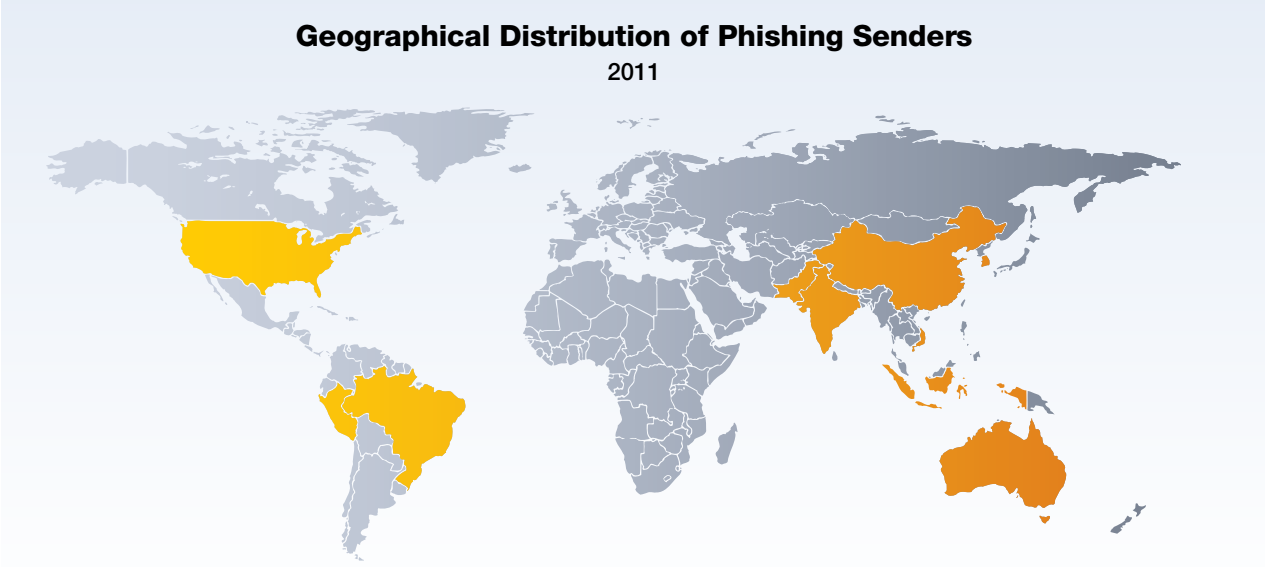


Figure 26: Geographical Distribution of phishing Senders – 2011

Country	% of phishing
Indonesia	15.1%
India	10.7%
China	6.9%
Brazil	5.9%
Vietnam	5.8%

Country	% of phishing
Australia	5.0%
South Korea	4.5%
USA	4.4%
Peru	3.8%
Pakistan	2.6%

Table 2: Top 10 Countries of Scam/phishing Origins—2011

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

The changes in email-like phishing/scam described at the beginning of this section are also reflected in the targeted industries.[16]

- Until 2009, traditional email phishing that targeted financial institutions dominated the scene, representing more than 50 percent of all phishing emails. They lost ground in 2010 and until autumn, 2011, but recovered to about 15 percent by the end of 2011.
- Online shops were the most favored targets of mid-2010 but did not play a role in 2011.
- Parcel services were used widely to dupe users during the second term of 2010 when they reached about 20 percent of all scam/phishing-like emails. In the second quarter of 2011, more than 50 percent of this spam used the good name of parcel services. This type nearly disappeared by the end of 2011.
- Since the beginning of 2010—when we started to monitor this class of emails—social networks have dominated the statistics by always staying in the top two. At the beginning of 2011, more than 80 percent of the good name of legitimate brands using emails bet on social networks, stabilizing at 43 percent during the second term of 2011.

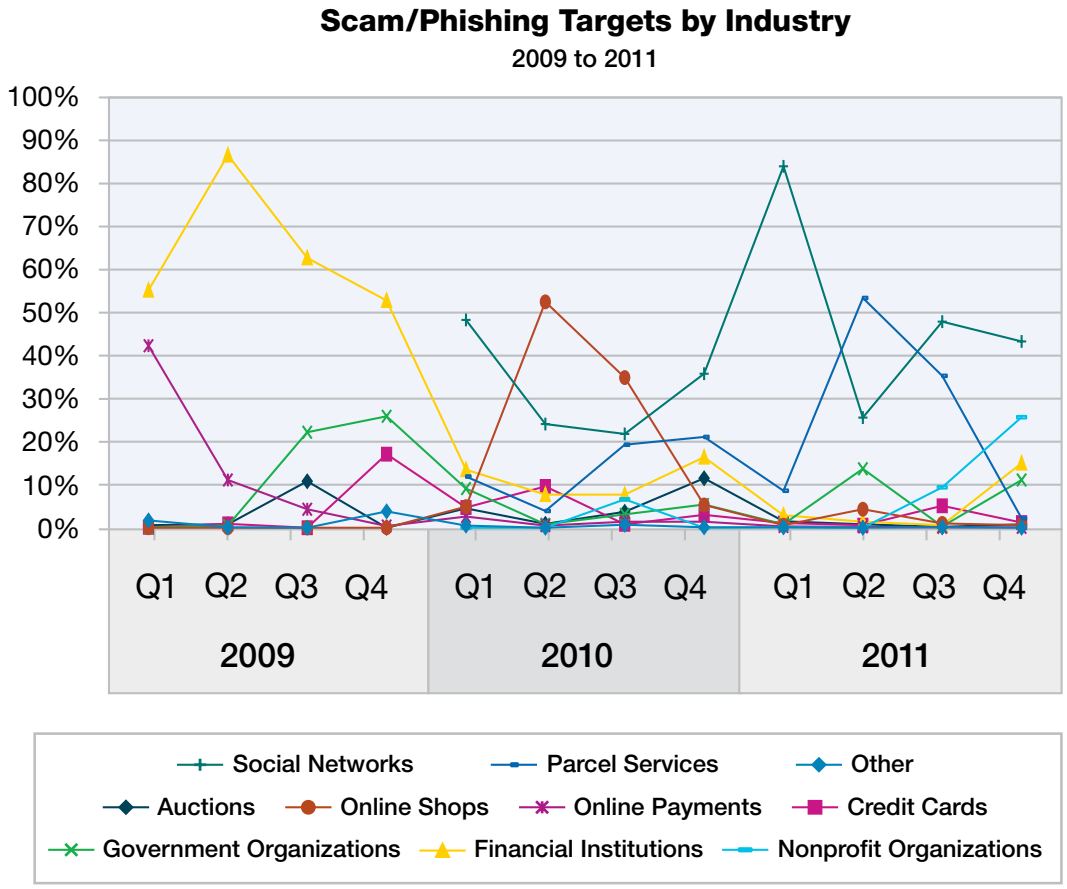


Figure 27: Scam/phishing Targets by Industry – 2009 to 2011[17]

16 In previous Trend and Risk Reports, the numbers are significantly different because they did not incorporate social networks, parcel services, and nonprofit organizations. Furthermore, the emails that "only" misused the name of the brand without doing real and traditional phishing were not counted.

17 The numbers concerning social networks, parcel services, and nonprofit organizations were not recorded before the beginning of 2010.

Evolution of spam

Over the years, we have seen many trends and types of spam that were discussed in previous IBM X-Force Trend and Risk Reports. We thought it would be interesting to take a look back at the ways spam has changed over time.

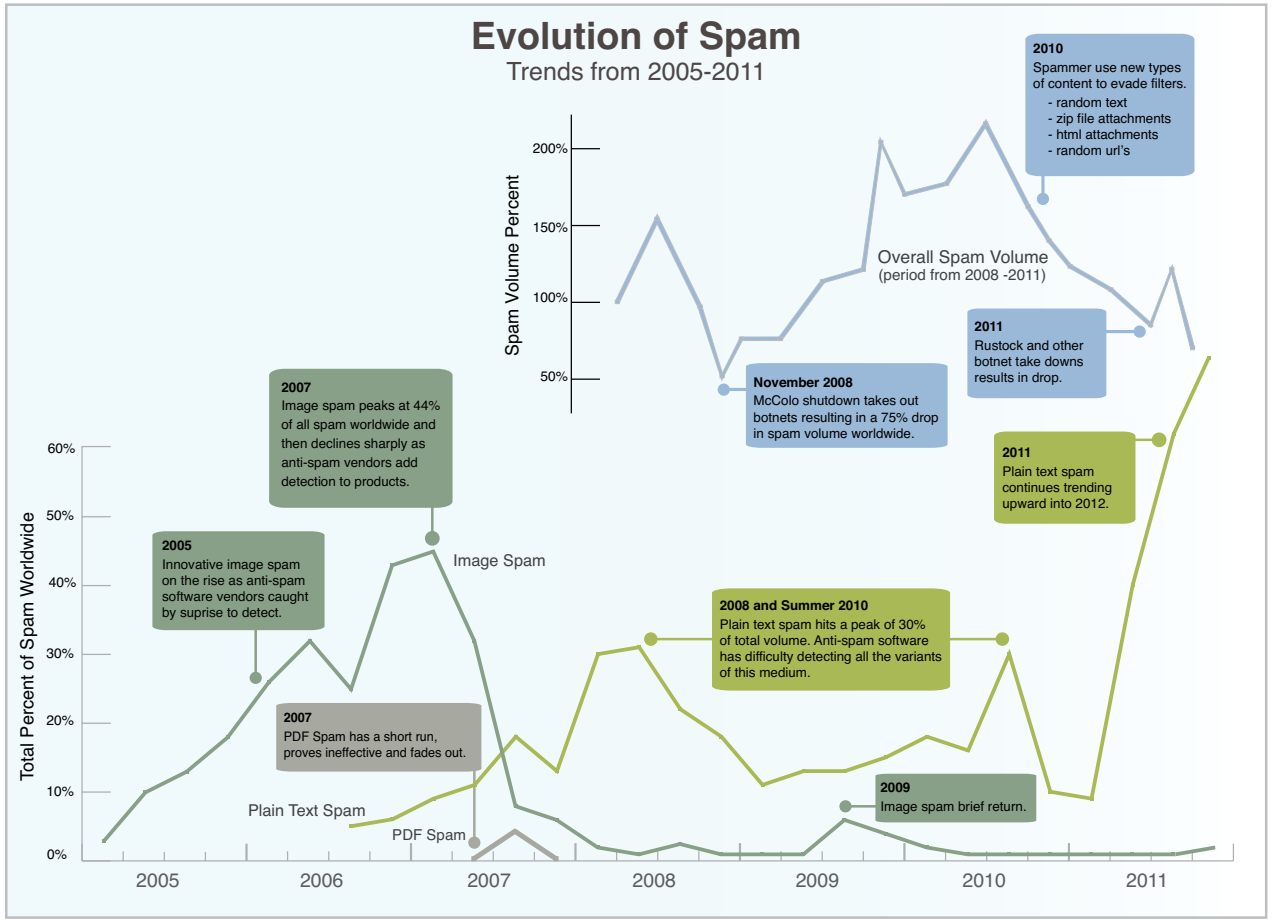


Figure 28: Evolution of Spam – Trends from 2005-2011

2005 to 2006—Image spam

In 2005 spammers started the massive use of image-based spam. By the end of 2005 nearly 20 percent of all spam was image-based and had reached an all-time high of more than 44 percent by the beginning of 2007. Afterwards it significantly declined. But why?

In the beginning spammers had good results, because at the time most anti-spam vendors did not expect this kind of spam and may even have considered attached images to be an indication of legitimate email. But after two years of image-based spam even the last anti-spam vendor adjusted its detection methods to include this kind of spam. Since this type of spam predefines many characteristics of the spam email, it was rather easy to check for suspicious patterns, so by early 2007 almost all of this type of spam was being reliably blocked.

2007—PDF spam

After the dramatic decline in image-based spam in the spring and early summer of 2007, spam that uses PDF attachments started to take its place. In August, 2007, large amounts of PDF spam representing nearly 20 percent of all spam (on some days) were seen. You can read more details about these PDF spam threads in the Frequency-X blog.

PDF spam was short lived. Maybe spammers tried to repeat the initial "success" they had with image-based spam and hoped anti-spam vendors were not prepared for this kind of attachment. This was not the case and spammers gave up on this approach quickly.

MP3 spam had an even shorter shelf life than PDF spam. We saw this technique emerge in October, and it lasted for only a few days. The volume was much lower than the PDF spam activity over the summer. Interestingly, the MP3 spam source code was very similar to the PDF spam. Details on MP3 spam can be found at the Frequency-X blog.

2008—First major decline of spam caused by McColo take down

In the first term of 2008, the percentage of plain text spam (without HTML code) significantly increased. Spam written in plain text reached a volume of 30 percent for the first time. After a similar peak in summer 2010, this type of spam reached its all-time high—generating more than 70 percent of all spam—at the end of 2011. Spam in plain text is even harder to detect, because there is no fixed feature, such as an abnormal HTML code fragment or a special kind of attachment around which patterns can be built. However, the trend in legitimate email is just the opposite. Today there are fewer types of messages or newsletters that do not use HTML. Simple plain text spam as an email characteristic has become more suspicious, and some day it might be used as a blocking criterion.

But the biggest blow to spam evolution in 2008 was the McColo shutdown on November 11th. On that day the worldwide spam volume decreased by a

whopping 75 percent! More interesting, perhaps, is the marked change we noticed in the origins of spam (the country location of the spam bot, generally). While McColo was operated out of the United States, the sudden and extreme volume and country distribution changes observed after the shutdown point to McColo as the base operator of spam bots all around the world. The United States has, for years, maintained a top spot in the spam origin list. Six days before the take down, it was in the number one spot.

Six days after the take down, spam production in the U.S. was reduced to a mere 14 percent of its original capacity. So, it was not a terrible surprise that the U.S. finally lost its top spot on the list.

2009—First climax of spam volume

In March, spammers once again started launching several threats using image-based spam. Technically, there were no new techniques in their approach, so most anti-spam filters had no trouble recognizing and blocking them. But there were differences in the content of the images attached in this new round of spam. In 2007 most imaged-based spam focused on stock trading. With the financial crisis that was happening, the focus took a more lucrative turn toward drugs. More information about the reborn image spam can be found in the Frequency-X blog.

So, why would the spammers return to an old technique? Especially when its success depends on the user to actually type the URL (that he only sees in the image and is not able to click) into the browser themselves. An answer might be that during the course of 2009, spammers increased the overall spam volume significantly. In that sense, image spam might be one part of the strategy to fire from all guns.

By November 2009, one year after the McColo shutdown, a first peak of the worldwide spam volume was reached.

Top 5 Spam Sending Countries before McColo Take Down	
USA	14.2%
Russia	11.0%
Turkey	7.4%
Spain	5.9%
Brazil	4.8%

Top 5 Spam Sending Countries after McColo Take Down	
China	12.7%
Russia	11.4%
USA	8.0%
South Korea	6.2%
Brazil	5.8%

Top 5 Spam Sending Countries at End of 2008	
Brazil	11.7%
USA	8.1%
China	6.6%
Turkey	5.7%
Russia	5.7%

Table 3: Top Spam Sending Countries Before and After the McColo Take Down

2010—First long-term decline of spam, but major and fast content variations of spam including HTML attachments

In contrast to all previous years, this was the first year where we did not see a significant increase of the spam levels. Instead, there were more content variations than were seen in all previous years. Examples witnessed in 2010 were:

- Spam with random text combined with random URLs, significantly increased the average byte size of spam.
- At the beginning of August 2010 , spammers began sending spam threats with ZIP attachments. X-Force looked into these messages, and each ZIP file contained a single malicious EXE file. More details on these spam threats with ZIP attachments can be found at the Frequency-X blog.
- The diversity in spam content seen in only a one year time span might suggest that spammers placed more emphasis on "quality" rather than quantity. Volume was no longer the solution to making it through spam filters.

2011—Another decline of spam, mainly caused by the Rustock take down

On March 16, more excitement came when spam volume was cut in half by the take down of the Rustock botnet. We discussed the details of that take down in the previous IBM X-Force Mid-year Trend and Risk Report. In contrast to the McColo take down of November, 2008, we did not see a fast recovery of the spam levels. However, spammers did not get tired of changing their approaches to get the spam through the filters by sending out new threats of:

- Malware ZIP spam in summer and autumn
- Image spam in December

All of which was discussed in detail in the previous sections.

Long term spam trends—country of origin

- India is the only country with a continuous growth
- Brazil was the biggest profiteer of the McColo shutdown in 2009 but is declining since then
- Russia was hit by the McColo shutdown significantly but is increasing since 2009
- Indonesia is the newcomer of 2011, the biggest profiteer of the Rustock take down in March, 2011
- The U.S. fell below 4 percent for the first time ever, mainly caused by the Rustock take down
- South Korea stabilizes at 4 percent
- France, Spain, and Turkey lost their dominating role of previous years.

Long term spam trends—that did not change

Beside all the movement described previously, some fundamentals did not change:

- We continue to see spam leveraging classic topics like replica watches, medical products, and software. This appears to be a well proven approach to earning illegitimate money.
- Spam, and particularly phishing, exists to get users to click the provided link. But spammers uncouple the content of the text provided in the spam from what happens when users click the link. This results in:
  - Perfect phishing-like spam that tries to sell products like the ones mentioned above.
  - Spam that takes advantage of topical news or other hot topics by promising more details when you click the link—and then infects the user's machine.
  - Masses of spam containing no text and only one link.

- Increase in speed. Spammers quickly adjust their approaches to try to stay ahead of every best effort to block it. The major use of image-based spam lasted more than two years (2005 to 2007) whereas the different spam phases seen in 2011 lasted 10 to 14 weeks. That said, shifts in the countries sending the spam happens much slower. Botnets are also growing slowly by way of comparison. It will be interesting to see whether the spammers will be able to speed-up their botnet acquisition in the future.

- Since 2008 the average byte size keeps returning to three kilobytes. We can consider this a standard spam size.
- Spammers always try new approaches. Please see the next section for some viewpoints on what might happen.

Future prospects on spam

In the first half of 2011 we have seen significant drops in spam volume without the quick recovery that has characterized it in the past. The business environment for traditional email spam has changed.

- Organizations or companies succeeded in taking down botnets or the needed infrastructure to send out spam, as seen in McColo or Rustock take down. (We discussed these take downs in good detail midyear.)
- Spam filters are continually improving.
- Other approaches come up that paralyze the spammer's business, such as "Click Trajectories: End-to-End Analysis of the Spam Value Chain."[18] The study showed that 95 percent of the payments of spamvertized products are handled by only three banks. The banks of the spam victim could block payments to these three banks.

This might bring the bad guys to focus on other areas such as spamming within social networks or performing distributed denial-of-service (DDoS) attacks. There are even experienced spammers who consider the spam business no longer as attractive.[19] On the other hand, there could also be aspects that might mislead old and new attackers to send out more spam.

- The number of Internet users is still growing. 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. Furthermore, there is a new type of machine to infect: the smart phone. And these hand-held computers have another advantage from the spammer's perspective: They are always online contrary to desktop PCs that are turned off when not in use. Today we still have bandwidth limits in the smart phone context because most users do not have a mobile internet flat rate. This is likely to change in the future. See the "Mobile malware perspective" section for details.
- Concerning the type of spam content, there are still some approaches spammers have not used, such as using Open Office documents as spam attachments.
- There are many well-known brand names that spammers might use as faked senders of their spam to make users click on the provided links.
- IPv6 may also provide new approaches for spammers to bother users and to annoy anti-spam vendors, particularly when spammers focus exclusively on IP blocking.

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

Section II
Operational Security Practices

In this section of the Trend Report we explore those topics surrounding 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.

Introducing Security Intelligence: An integrated approach to real-time security

In the past few years, increases in attacks, expansion of computing models (and hence, attack surfaces), and explosion of data have created significant challenges for security practioners. Organizations are defending against more, and more varied, threats than ever before.

Even determining that a breach has taken place can be challenging, leaving many firms unaware of serious compromises for months. They often have the raw data but lack the visibility and analytics to detect the breach. The 2011 Verizon Data Breach Investigations Report concluded that in 69% of breaches there was good evidence of the breach in the organization's log files, but such evidence is rarely found due to data overload.

Threat detection today therefore hinges on two elements: *identifying* suspicious activity among billions of data points, and *refining* a large set of suspicious incidents down to those that matter. For both tasks, organizations need approaches that can 1) analyze *all* the relevant data, 2) intelligently identify the signal in the noise, and 3) deliver that intelligence in a practical way.

This has led to the development of a new class of solutions called Security Intelligence, which provides unified visibility and real-time analytics across the spectrum of security operations.

In recognition of the new reality, IBM has made a bold move to drive the future of security intelligence and analytics. Through a commitment to unite the various disciplines of information security via a single Security Systems division, and the acquisition of Q1 Labs, a leader in SIEM (Security Information and Event Management) and security intelligence, IBM is tackling this problem head-on.

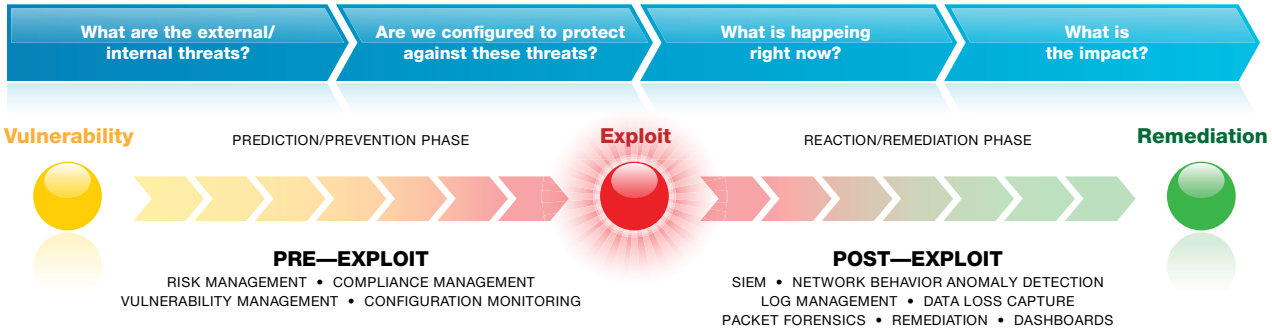
Defining Security Intelligence

Let's begin by considering a definition of Security Intelligence:

Security Intelligence (SI) *is the real-time collection, normalization, and analysis of the data generated by users, applications, and infrastructure that impacts the IT security and risk posture of an enterprise. The goal of Security Intelligence is to provide actionable and comprehensive insight that reduces risk and operational effort for any size organization.*

Data collected and warehoused by Security Intelligence solutions includes logs, events, network flows, user identities and activity, asset profiles and locations, vulnerabilities, asset configurations, and external threat data. Security Intelligence provides analytics to answer fundamental questions that cover the before/during/after timeline of risk and threat management.

Security Intelligence Timeline



Security Intelligence provides a unified view of the security and risk posture of an organization, spanning the four primary risk domains: People, Data, Applications, and Infrastructure.

Those familiar with SIEM and log management products might view Security Intelligence as the next logical step in the journey of these technologies. Adding pre-exploit capabilities, broader data capture, and deeper intelligence, Security Intelligence extends SIEM and log management. It can enable better prevention, detection, and prioritization of threats (both external and internal) and it automates compliance monitoring and reporting.

Security Intelligence also can provide for broad visibility into security incidents. For example, by analyzing network flows via deep packet inspection, and monitoring user activity for anomalies, Security Intelligence can help identify when an employee's actions look suspicious, suggesting possible insider data theft or account compromise by external parties. Furthermore, by marrying IPS alerts with vulnerability scan results and knowledge of network topology, Security Intelligence can help identify which intrusion attempts are attacking vulnerable assets and which can be ignored.

The analogy to Business Intelligence

It's instructive to look at the parallels between Business Intelligence (BI) and Security Intelligence. BI synthesizes large volumes of business information to glean actionable business insights:

Which products are selling well, and with which customer segments?

Which geographies responded most strongly to a recent promotion?

Why is my profitability increasing with one product line, but falling with another?

Similarly, Security Intelligence (SI) synthesizes large volumes of security information to obtain actionable security insights of relevance to both IT and the line of business:

Which types of attacks are we likely most vulnerable to? (How should we adjust our security practices and controls?)

Which business partners and vendors may be creating the greatest security risks for us? (Should we adjust their access or require stronger controls on their end?)

Are we seeing any new security or compliance risks from mobile computing? (If so, which risks should we focus on?)

One difference between SI and BI is that Security Intelligence provides real-time insight and monitoring, while Business Intelligence typically reflects point-in-time information. Both can be invaluable management tools, but in the world of security and compliance, up-to-the-minute information is critical.

Business Intelligence has become a standard tool for business planning and executive visibility. Likewise, Security Intelligence is becoming a standard tool for security planning and executive visibility. Moreover, it can serve as the fact basis for security conversations between IT and the line of business, to help evaluate risk/reward considerations about business practices and offerings.

The tenets of Security Intelligence

The three tenets of Security Intelligence—**Intelligence, Integration, and Automation**—help make it easier for users to get productive quickly.

Here are some examples of what this looks like in practice:

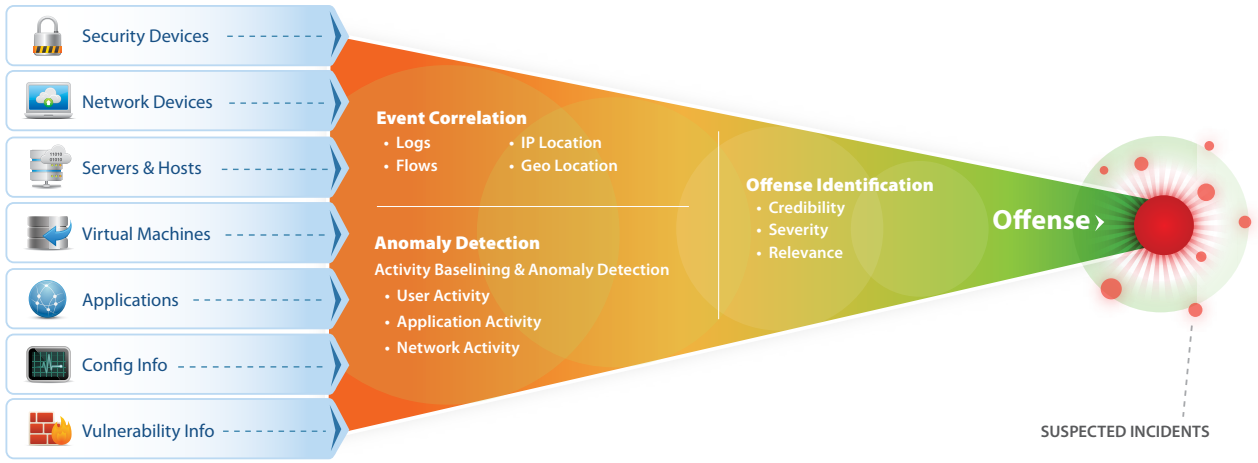
1. **Intelligence:** The ability to make sense of large amounts of security- and compliance-relevant data. This means storing, correlating, reporting on, and querying a wide variety of information at Big Data scale (security information "is" Big Data)—in order to deliver actionable insight.

2. **Integration:** The foundation of intelligence, enabling consistent, normalized analysis of disparate data. By gathering and combining security-relevant data—in type and volume—you can expand a limited, two-dimensional view of a security event into a rich, three-dimensional view supported by context.

– **Example:** The integration capabilities delivered out-of-the-box by Security Intelligence solutions make a huge impact on a security analyst's productivity. Normalization of the data from hundreds of sources helps prevent customers (and consultants) from having to become experts in each vendor's data schema. For example, a compliance mandate might require documenting authentication events (failed logins, successful logins, successful logins followed by a privilege escalation, etc.). With SI, organizations may no longer track that manually across dozens of assets, each with its own data schema.

3. **Automation:** The element that brings Security Intelligence into the modern era by helping drive out unneeded complexity and reduce the total cost of ownership (TCO). This includes tasks automated by the use of broader data (such as network flows) and intellectual property packaged for easy application.

Applying Advanced Analytics to the Broadest Set of Data



How does Security Intelligence differ from SIEM?

Security Intelligence surpasses first-generation SIEM technologies in several meaningful ways:

Network activity monitoring and flow analytics. In the past, logs from devices, applications, servers, and infrastructure services gave you a rough idea of what was happening. Today, logs are just a starting point. Network flow collection, deep packet inspection, and packet (content) capture are required for three-dimensional context and visibility. Security Intelligence uses flow analytics to help deliver real-time insight into user behavior, social media usage, mobile activity, cloud activity, and more:

Is that conversation using port 80 web traffic or a hidden botnet IRC communication?

Are intruders using a compromised employee account to exfiltrate sensitive data?

Are employees accessing sensitive intellectual property inappropriately?

Packet-level visibility, which comes from the integration of network activity monitoring (content capture) and SIEM, can provide such insight.

Predictive analytics and pre-exploit awareness. Security Intelligence integrates pre-exploit configuration and vulnerability management capabilities. This allows an organization to identify, prioritize, and systematically address the risks created by misconfigured devices (such as firewalls) and unpatched vulnerabilities.

Anomaly detection. Many traditional security solutions focus on protecting the organization from known threats such as publicly disclosed vulnerabilities and common malware. In today's security environment there is an increased desire to detect sophisticated, targeted attacks that may employ entirely new attack methodologies. Additionally, insider threats can often only be detected through the analysis of authorized behaviors. An anomaly-centric approach can shed light on these kinds of activities.

Easier to deploy and staff. When the first SIEM products were released, early adopters were willing to spend considerable time and money to bring them into production. Connectors and rules needed to be written, users needed to be trained, and so on. Once in production, their staffing requirements could also be significant, due to a high rate of "false positive" alerts requiring investigation. Security Intelligence solutions now use a broader set of information (event, flow, asset profile, network topology, vulnerability, etc.) and greater automation to help achieve significant data reduction and reduce staffing requirements.

What are the main benefits?

Let's examine the benefits organizations are gaining from their SI deployments:

Improved compliance

Security Intelligence aids compliance activities by logging and proactively monitoring diverse information across the enterprise—which users are accessing high-value systems (appropriately and otherwise); is any sensitive data being transmitted unencrypted over open networks; are firewalls configured properly; and so on. SI also can improve operational efficiency—in some cases, saving thousands of hours of manual effort—through automated reporting and easy searching of logs, and flows.

Faster detection and remediation of threats

In the post-perimeter world, focusing solely on either prevention or detection/remediation is a losing proposition. Organizations need to perform both. Boundaries can be porous due to mobile computing, social media, and cloud computing, leading to what Forrester Research calls a **"zero-trust" environment**. Security Intelligence addresses this by helping businesses detect and remediate breaches faster, in addition to helping prevent them in the first place (see "Pre-Exploit Risk Reduction" below). By correlating massive data volumes in real-time, SI can help find the needle in the haystack—analyzing events from network and security devices, servers, applications, directory servers; network activity flows (with packet capture); asset information; configuration data; and vulnerability information. Security Intelligence can also accelerate remediation by helping identify which assets and users were potentially affected by a compromise, and by leveraging content capture for forensic research.

For example, when the Conficker worm began to spread in late 2008, this caused a dramatic increase in TCP port 445 traffic on the Internet. Security intelligence systems highlighted this traffic increase as suspicious even before Conficker had been given a name by security researchers. This sort of preemptive detection can help protect computer networks against advanced and zero day threats, for which there might not be a signature or a patch.

Reduction of insider fraud, theft, and data leakage

External attacks garner most of the headlines, but insider threats can be even more damaging—compromising invaluable intellectual property and even jeopardizing national security. Security Intelligence enables organizations to identify and mitigate these types of threats by helping detect:

- Unauthorized application access or usage
- Data loss such as data being transmitted to unauthorized or unfamiliar destinations
- User access issues such as privileged access exceptions
- Application performance issues such as loss of service or over-usage

Pre-exploit risk reduction

Security Intelligence builds on foundational prevention tools like firewalls and IPS devices with new correlations that help the organization prevent attacks:

- Automatic monitoring of device configurations (such as firewalls) and alerting on security gaps and policy violations
- Prioritization of vulnerabilities seen by VA (vulnerability assessment) scanners, based on network topology and asset value
- Predictive threat modeling and simulation of network changes

Security Intelligence solutions can apply greater intelligence to a broader set of inputs than previously possible. Network activity flows based on content capture, for example, can provide a more reliable

view of the effectiveness of security device rules than configuration data by itself. As a recent blog post observed, "[Configuration data alone can] miss situations where a configuration is thought to be adequate but for some reason still allows potentially risky network traffic to propagate." Similarly, knowledge of network topologies can "minimize false positives common among vulnerability scanners and … [prioritize vulnerabilities] that can be easily exposed because of the way the network is configured."

Simplified operations and reduction of effort
SI solutions are applying intelligent automation to simplify security operations and reduce the burden on security and network professionals. This can result in significant cost reductions. These benefits stem from greater efficiencies and elimination of tedious manual tasks.

Best practices for Security Intelligence

When building Security Intelligence competency, there are both organizational approaches and technical capabilities that increase the chances of success. Here are several that can be prioritized:

Definition of incident escalation policy. The security intelligence solution can be viewed as an internal cloud service, serving groups such as firewall management, systems management, and network management. Just as with a public cloud service, the provider of the SI solution (typically the security or risk management group) should define a contract with the consumers of the solution that governs how security incidents are handled and escalated. Immediately reporting issues to executive management may not be optimal, and can also damage the relationship with the consumers and cause them to withhold data in the future.

Definition of key use cases and reports for initial deployment. The organization should decide on which topics to initially focus its monitoring and reporting efforts. Common categories include generic external threats (such as botnets and traffic from darknets), industry-specific risks, insider threats, policy violations, and privileged user activity.

Intelligent anomaly detection. To detect unusual behavior, the solution should generate activity baselines across dimensions of interest (users, applications, and networks) based on observed behavior, and then identify anomalies that fall outside the norm. Dynamic baselining that automatically learns baseline changes can reduce subsequent manual work.

Flow analytics based on deep packet inspection. As described earlier, flow analytics with packet capture can provide deep visibility into security and compliance risks. It can enhance prevention via identification of erroneous network configurations, detection via packet-level insight, and forensic investigation by showing what data was accessed by whom across a range of use cases.

Predictive analytics. Organizations seeking a more proactive security posture should also prioritize capabilities such as device configuration monitoring, compliance policy monitoring, and vulnerability prioritization.

Conclusion

In summary, Security Intelligence is a powerful enabler of enterprise security and can aid with compliance delivery of actionable information through real-time insight and deep forensics. It can provide significant benefits to both IT and the line of business through deeper intelligence, integration, and automation—areas that have historically been an Achilles heel of security solutions. Security Intelligence solutions are reasonable to implement and manage for both small and large organizations, and can yield a practical solution for real-world needs.

References:

1. http://www.verizonbusiness.com/resources/reports/rp_data-breach-investigations-report-2011_en_xg.pdf
2. http://blog.q1labs.com/2011/07/28/defining-security-intelligence/
3. http://blog.q1labs.com/2010/08/26/do-we-need-a-security-analog-for-business-intelligence-absolutely-we-do/
4. http://q1labs.com/resource-center/white-papers/details.aspx?id=113
5. http://blog.q1labs.com/2011/10/20/three-ways-to-embrace-the-zero-trust-environment/
6. http://q1labs.com/resource-center/case-studies/details.aspx?id=114
7. http://blog.q1labs.com/2011/06/16/latest-gartner-report-shines-bright-light-on-qradar-risk-manager/
8. http://q1labs.com/resource-center/white-papers/details.aspx?id=113
9. http://blog.q1labs.com/2010/09/17/siem-is-a-security-intelligence-cloud/
10. http://q1labs.com/resource-center/brochures/details.aspx?id=129

Vulnerability disclosures in 2011

Since 1997, X-Force has tracked public disclosures of security vulnerabilities in software products. Our analysts follow public mailing lists and websites where vulnerabilities, remedy information, and exploits are disclosed and we record what has been publicly reported.

In 2011 we reported just over 7000 new security vulnerabilities. While this is a significant decline from 2010, when we saw more vulnerabilities than ever before, there has been a two year, high-low cycle in vulnerability disclosures since 2006, and the levels of each high point and each low point keep climbing.

The first time we saw a decline in the total number of vulnerabilities it was 2007, and this generated a great deal of speculation as to why the vulnerability landscape was changing. However, it is clear in retrospect that this was just an aberration in the data and that the totals were going up from there. If the cycle of the past six years holds true again this year, 2012 will be another record year for vulnerability disclosure.

Web application

The category of security vulnerability that has seen the starkest decrease in 2011 is web application vulnerabilities. For the past few years about half of the disclosed security vulnerabilities were web application vulnerabilities. However, this year that number was down to 41 percent, a percentage that hasn't been seen since 2005. This is illustrated in figure 30 that shows web application vulnerabilities from 2010. Looking at the types of web application vulnerabilities disclosed, SQL injection stands out as an important category that has seen significant decline.

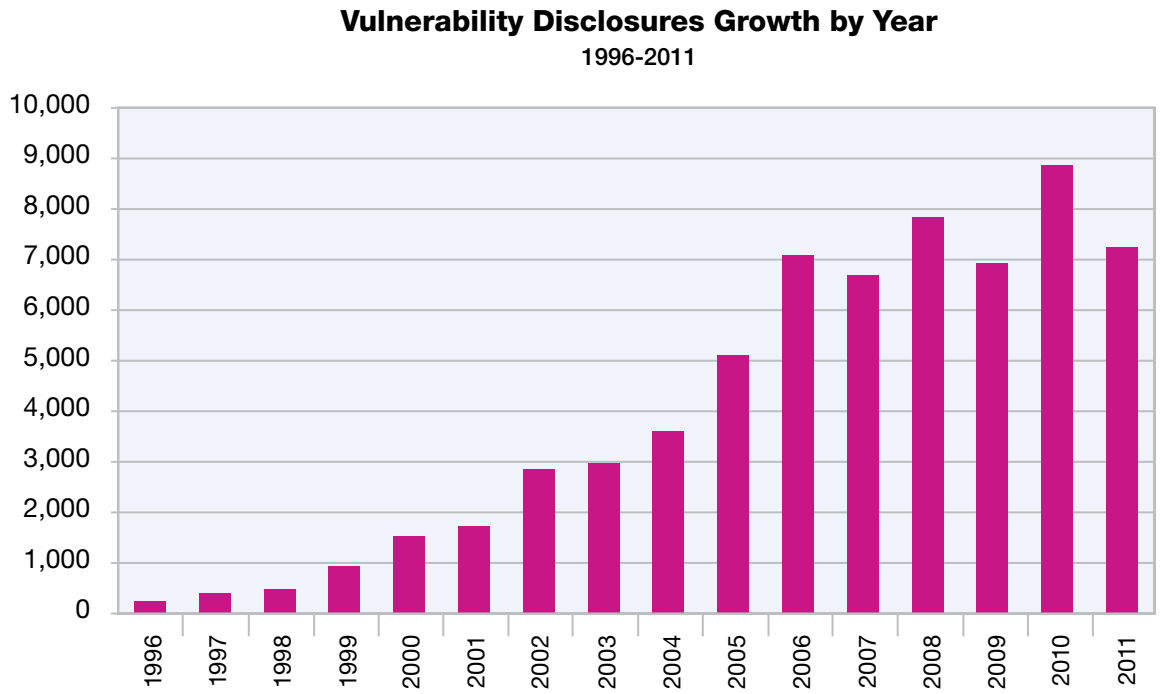


Figure 29: Vulnerability Disclosures Growth by Year – 1996-2011

SQL injection vulnerabilities are particularly important because they are the most common type of attack that IBM sees on the thousands of networks that we monitor and help protect around the world. Automated SQL injection attacks launched by financially motivated botnet builders canvas the web looking for vulnerable sites. These sites can be infected with Javascript redirectors that drive their visitors to malicious exploits. SQL injection is favored by unsophisticated attackers searching the web for easy targets to deface. SQL injection attacks have also featured prominently in several high profile breaches this year by more sophisticated attackers.

If you are running an Internet-facing web application with a SQL injection vulnerability in it—it likely will be targeted sooner or later. Therefore, it is important to get these vulnerabilities fixed. The decline in the number that we are seeing might mean that developers of web applications are getting smarter and writing less vulnerable applications. If so, this is a positive sign. However, a lot of work remains to be done.

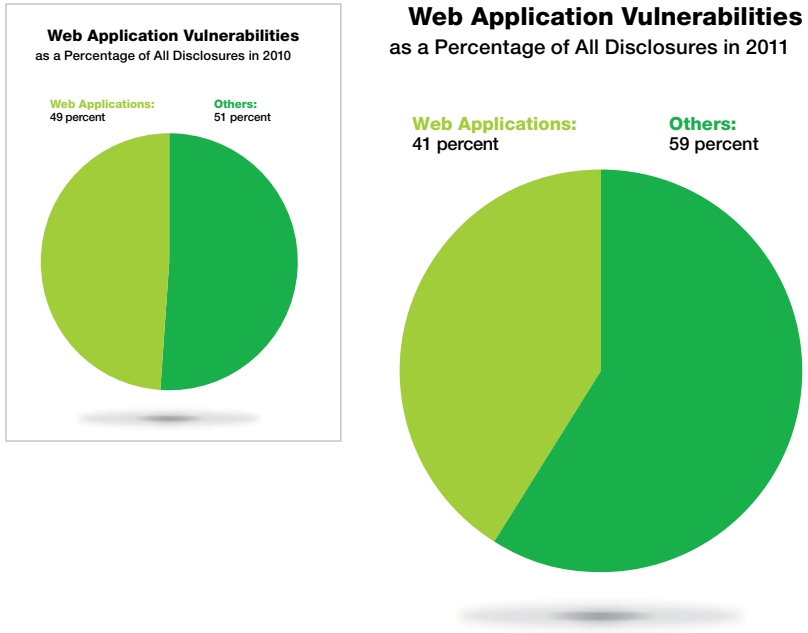


Figure 30: Web Application Vulnerabilities as a Percentage of All Disclosures in 2011

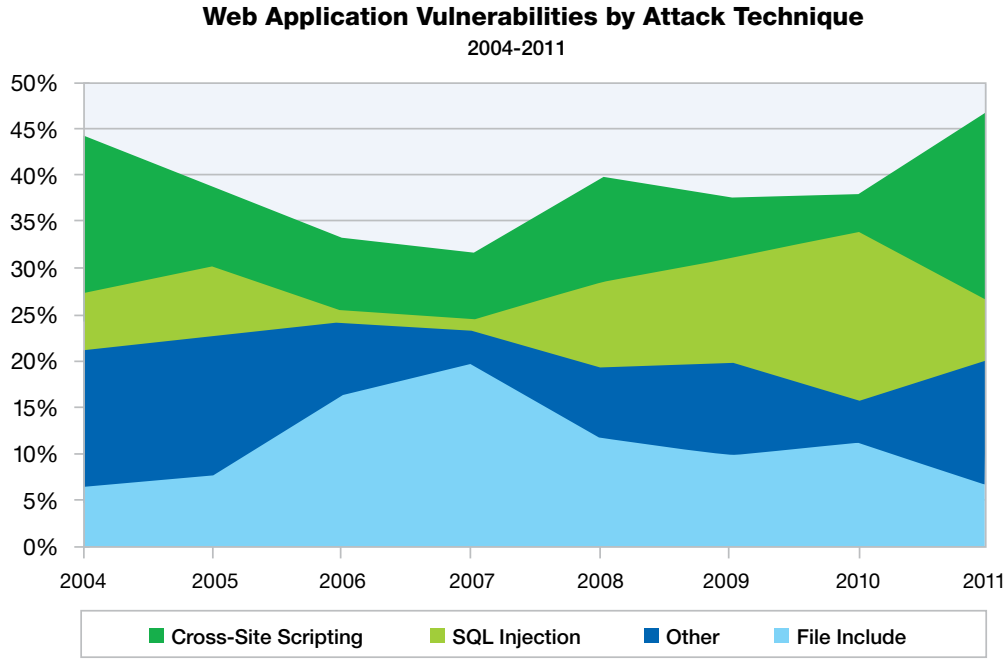


Figure 31: Web Application Vulnerabilities by Attack Technique – 2004-2011

We still saw nearly three thousand web application vulnerabilities disclosed in 2011, and the total number of web application vulnerabilities that X-Force sees may only be the tip of the iceberg of what exists on the open Internet. The reason is that X-Force tracks only publicly disclosed vulnerabilities. Web applications that are maintained by a company or an open source project for use by third parties are subject to these public vulnerability disclosures. However, most web applications are custom software developed in house or by private firms for exclusive use on a particular website. These custom web applications aren't subject to public vulnerability disclosures—they don't have third-party users so there is no need to inform the public about vulnerabilities in them.

Our data from the IBM AppScan OnDemand users provides some insight into the state of custom web applications, and it has shown some level of improvement as well. However, this sample is probably self-selecting—the developers who are smart enough to work with IBM to improve the security of their code are probably better than average at avoiding security problems in the first place. Therefore, it remains likely that the reality of web application security on the Internet is somewhat worse than our data indicates. The amount of attack activity that we're seeing certainly supports that conclusion.

One category of web application that is subject to both public vulnerability disclosure and a lot of attack activity is web-based content management systems (CMS). We took a look at four web-based content management systems, and our data shows that the most important weaknesses in these systems come from the ecosystem of third-party

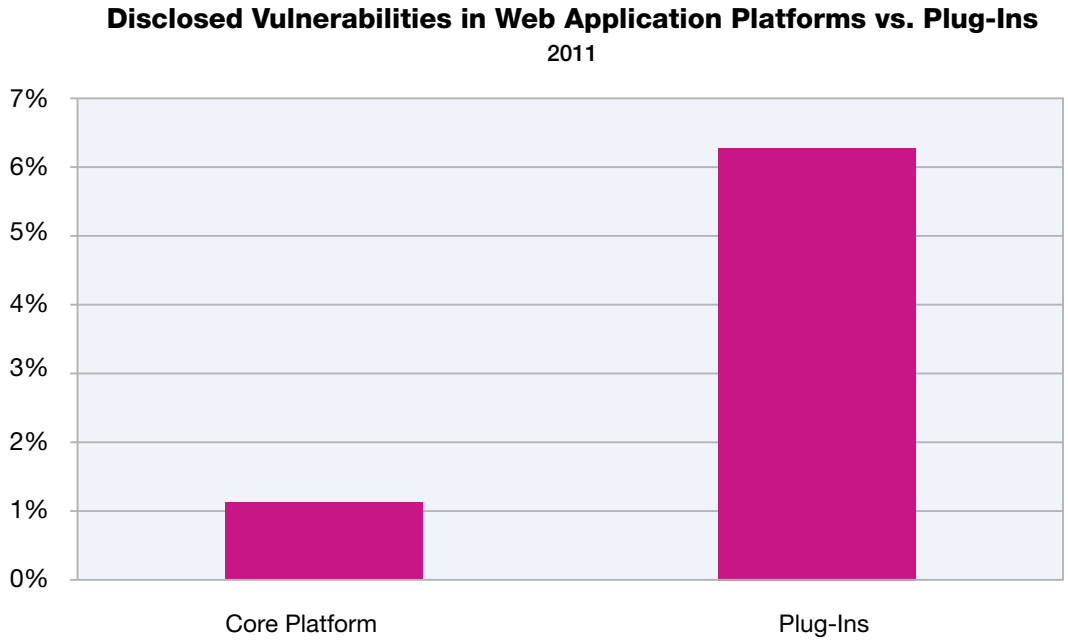


Figure 32: Disclosed Vulnerabilities in Web Application Platforms vs. Plug-Ins – 2011

plug-ins that they support. There are far less vulnerabilities disclosed in core CMS platforms than in their plug-ins, and the core platform vulnerabilities are much more likely to have patches available. Part of the reason for this is that there is wide variation in the level of support and attention to security issues offered by various plug-in developers.

Web CMS vulnerabilities are favorite targets of attackers because they are publicly disclosed and impact a large number of websites on the Internet. zero day vulnerabilities in these systems have factored into a number of breaches this year. Users of web CMS software should take care to evaluate the security practices of the maintainers of any plug-in they use. They should closely monitor security vulnerability disclosures for both core software and plug-ins and keep them patched as a top priority. They should also consider further protecting their websites with application layer firewalling or intrusion prevention.

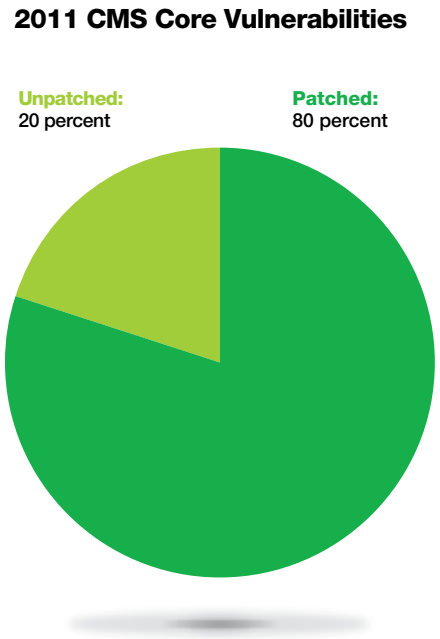


Figure 33: Disclosed Vulnerabilities in core content management systems – unpatched vs. patched – 2011

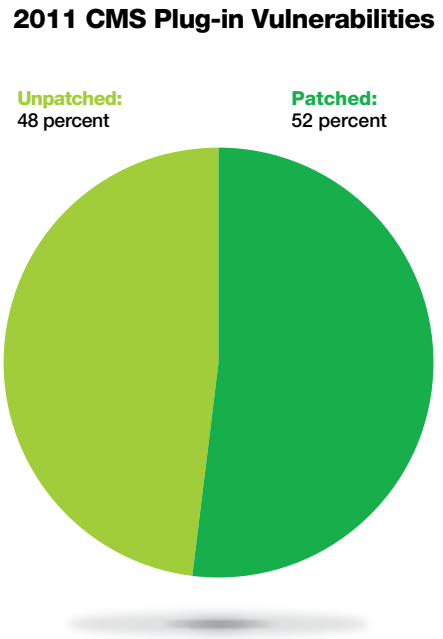


Figure 34: Disclosed vulnerabilities in plug-in content management systems – unpatched vs. patched – 2011

Declines in exploitation

Besides the improvement in web application security, there is another reason for optimism. In 2011 X-Force saw a significant decline in the number of exploits that have been publicly released, the lowest number we've seen since 2006. This number is lower on a percentage basis as well as a real basis. For the past few years the percentage of vulnerabilities with public exploits has hovered around 15 percent, but in 2011 it was 11 percent.

These decreases reflect specific areas that have been the target of a great deal of attack activity in the past few years. For years, web browsers were the primary target of drive-by-download attacks. Although the number of high and critical browser vulnerabilities was up year over year, the number of exploits released for browser vulnerabilities is lower than any year since 2006. Drive-by-download attacks have moved into targeting third-party browser plug-ins more often than the browser itself.

Document readers are one such third-party component that has been a favorite of attackers as malicious document files can be used in drive-by-download scenarios as well as attached to emails. Although document format vulnerabilities and exploits peaked last year, 2011 has seen fewer vulnerability disclosures, and exploit releases are down to a level not seen since 2007. This represents significant progress.

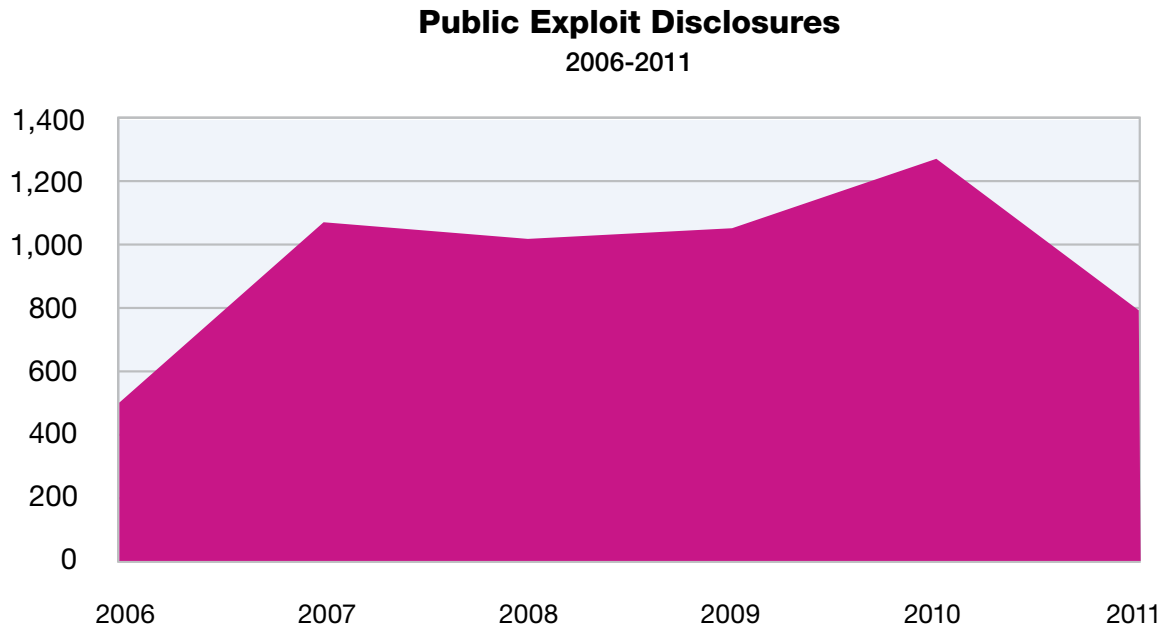


Figure 35: Public Exploit Disclosures – 2006-2011

	2006	2007	2008	2009	2010	2011
Public Exploits	504	1078	1025	1059	1280	778
Percentage of Total	7.3%	16.5%	13.3%	15.6%	14.7%	11.0%

Table 4: Public exploit disclosures – 2006-2011

X-Force believes that this progress is a result of architectural changes that have been made to software over the past few years that make exploitation more challenging. Operating system memory managers now contain a variety of features that detect memory corruption and safely stop execution. Many browsers and document readers now come with execution sandboxes that limit what successful exploits are able to do. The result is that vulnerabilities which in the past would have quickly resulted in widespread exploitation can now go for months without being successfully exploited in the wild.

To be sure, exploitation of vulnerabilities is not impossible today, in spite of these various security features. X-Force Research has published a number of papers describing the process of obtaining code execution in challenging situations. At Blackhat USA 2012, X-Force Researchers Mark Yason and Paul Sabanal presented **Playing in the Reader X Sandbox**, which discussed ways that malicious code might operate in a sandboxed application environment. In 2011, Chris Valasek presented **Understanding the Low Fragmentation Heap** at Blackhat USA, which discussed approaches for obtaining code execution in the heavily defended Windows Heap.

However, the techniques described in these papers require a great deal of time, effort, and skill to successfully apply. We have seen a growing number of situations this year where critical vulnerabilities that have been exploited in laboratory environments have not been targeted in the field. We've rarely been able to say that before, and it may mean that we are at the cusp of a new era in computer security.

Public Exploit Disclosures for Browser
2005-2011

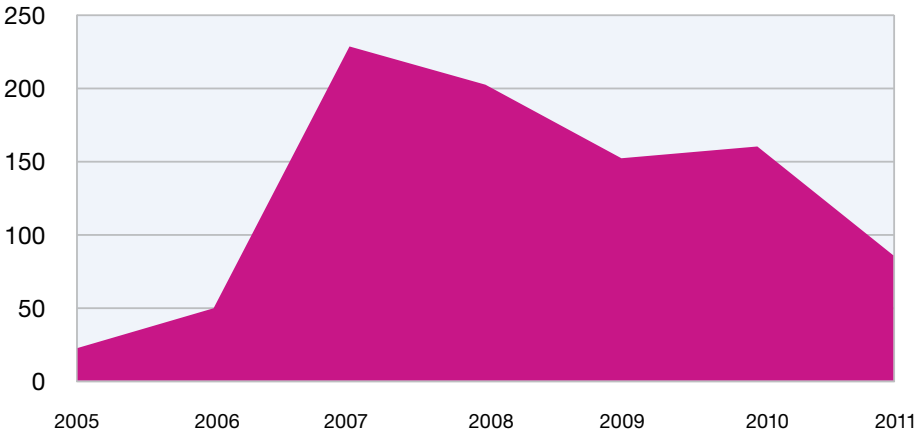


Figure 36: Public Exploit Disclosures for Browser – 2005-2011

Web Browser Vulnerabilities Critical and High
2005-2011

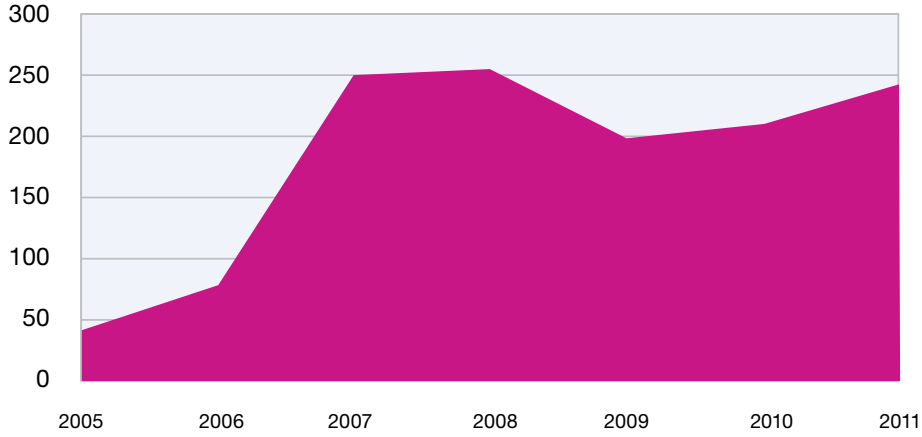


Figure 37: Web Browser Vulnerabilities, Critical and High – 2005-2011

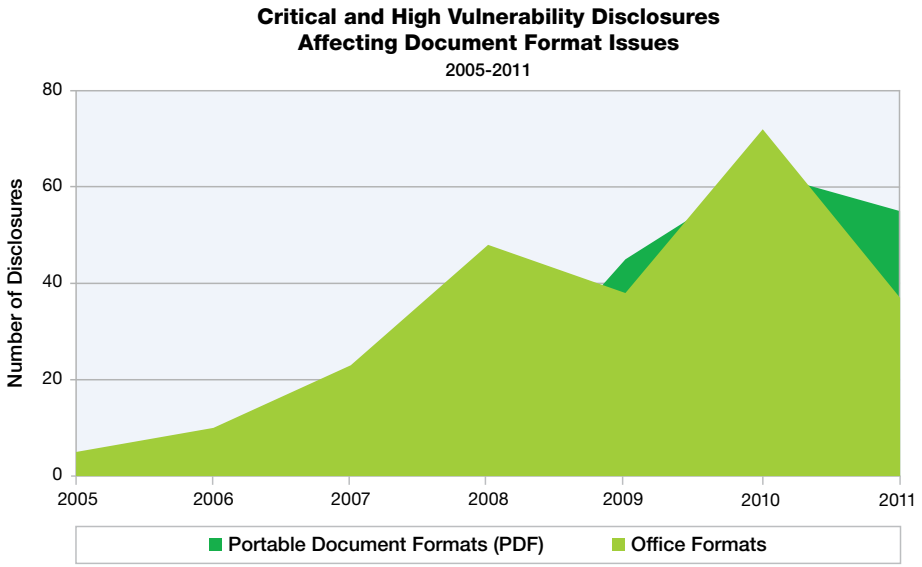


Figure 38: Critical and High Vulnerability Disclosures Affecting Document Format Issues – 2005-2011

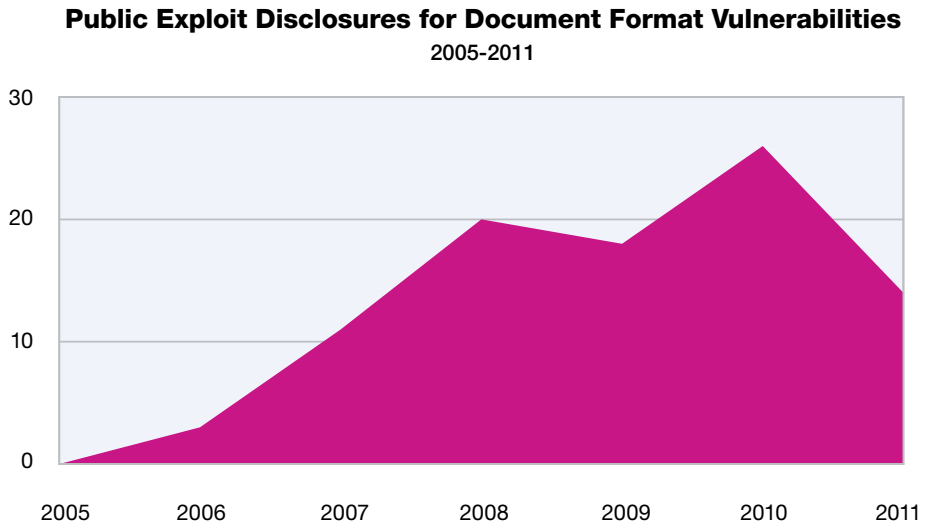


Figure 39: Public Exploit Disclosures for Document Format Vulnerabilities – 2005-2011

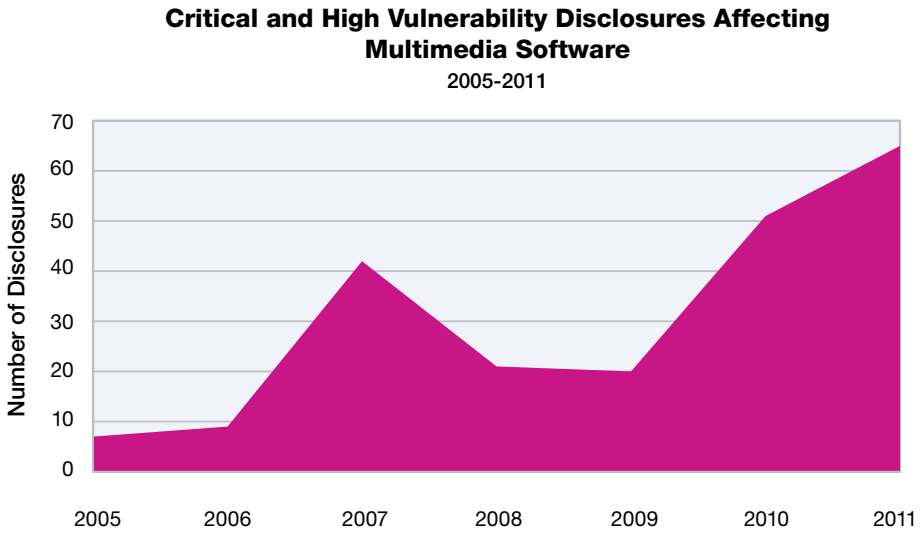


Figure 40: Critical and High Vulnerability Disclosures Affecting Multimedia Software –2005-2011

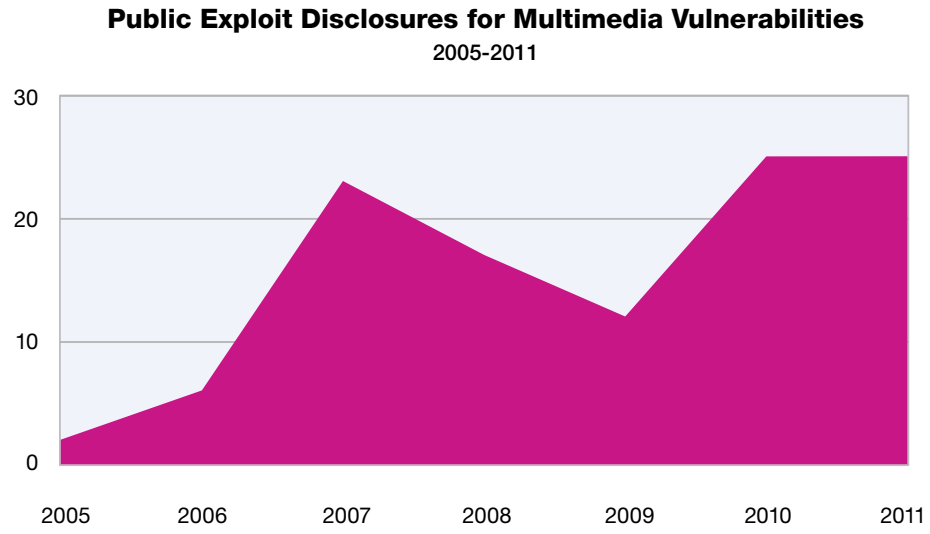


Figure 41: Public Exploit Disclosures for Multimedia Vulnerabilities – 2005-2011

Attackers shifting attention to new areas of focus

Of course, there are important gaps that remain to be closed. We continue to see increases in the number of vulnerabilities being disclosed in multimedia players and we saw just as many exploits publicly disclosed for multimedia vulnerabilities in 2011 as we saw in 2010. This continues to be an area of focus for attackers.

As of this writing, several critical multimedia vulnerabilities that were disclosed publicly early this year continue to be used in sophisticated, targeted attacks associated with Advanced Persistent Threat. These malicious files can be attached to emails, which are sent to targets along with carefully crafted email text that is tailored for the intended victim. It is critically important that multimedia players be meticulously patched or completely disabled in high-security environments.

The domain of mobile devices is another area that is gaining in importance. There are many mobile operating system vulnerabilities being disclosed, and there are a number of exploits being publicly released for these vulnerabilities. The desire to jail break or root mobile devices is one motivating factor that leads people to post mobile exploit code online. Of course, once that code is available, it can be used for malicious purposes against phones that are not jail broken.

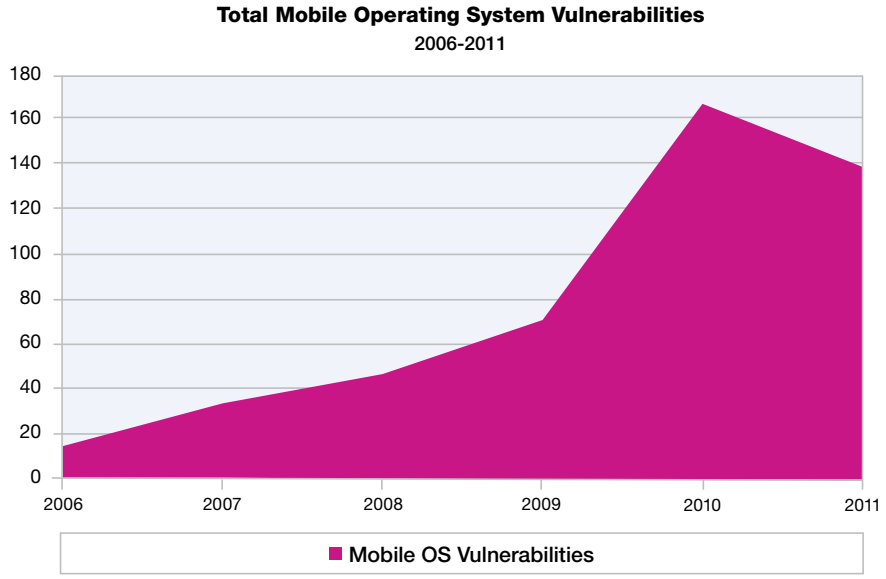


Figure 42: Total Mobile Operating System Vulnerabilities – 2006-2011

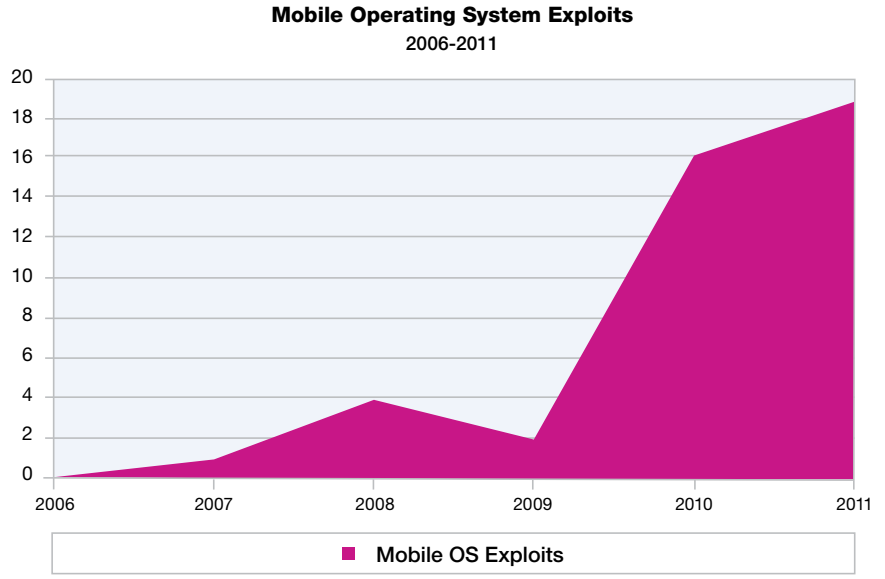


Figure 43: Mobile Operating System Exploits – 2006-2011

In 2011, we've seen an uptake in malicious activity targeting mobile devices. Some malicious applications have used publicly available jail-breaking exploits to obtain elevated privileges on phones once they've been installed. Because of the two-tiered relationship between phone end users, telecommunications companies, and mobile operating system vendors, disclosed mobile vulnerabilities can remain unpatched on phones for an extended period of time, providing a large window of opportunity to attackers. This situation is exacerbated by the proliferation of different hardware platforms as well as regulatory requirements. The amount of actual attack activity today is very small compared to the volume of activity targeting traditional workstations, but we expect attacker interest in mobile devices to grow linearly into the future. Large botnets of infected mobile devices have started to appear on the scene and this is only the beginning.

2011 has seen a 70 percent increase in the number of critical vulnerabilities disclosed this year versus last year. Critical vulnerabilities are vulnerabilities that have a Common Vulnerability Scoring System (CVSS) score of 10 out of 10. While this increase appears to be alarming, it is the opinion of X-Force that the increase represents a data aberration and we expect the volume of these kinds of vulnerabilities to smooth out in 2012.

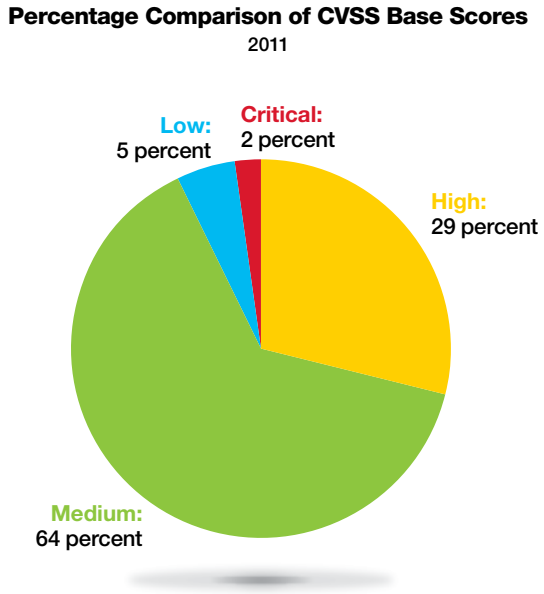


Figure 44: Percentage Comparison of CVSS Base Scores – 2011

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

Table 5: CVSS Score and Corresponding Severity Level

"Jail-breaking" is a process that allows you to install unapproved third-party applications on your device. Jail-breaking often involves the use of a privilege escalation exploit to obtain root access to phones based on unix style operating systems, and is therefore sometimes referred to as "rooting" the device. Once root access is obtained, security controls that prevent the installation of unapproved software can be subverted.

Vulnerabilities in enterprise software

An important long-term trend is the increase in the percentage of vulnerabilities being disclosed by large software vendors. The top 10 software vendors who disclosed the largest number of security vulnerabilities are also big software vendors who make the widest variety of enterprise software. A true top 10 list would also include vendors of web-based content management systems, but we have excluded those products from this analysis, in order to focus on the impact of vulnerabilities in popular enterprise software products.

These top ten vendors have represented a constantly increasing percentage of the total number of vulnerabilities disclosed, from 19 percent in 2008 to 31 percent in 2011. We don't believe that this is merely a measure of software industry consolidation.

Secure development practices have become an increasingly important part of the software development lifecycle, and responsible vendors have taken steps over the past few years to improve their ability to identify and eliminate vulnerabilities in their code. These efforts are producing surges in public disclosures from these vendors, as they fix shipped code and make patches available.

Top Ten Software Vendors with the Largest Number of Vulnerability Disclosures
2008–2011

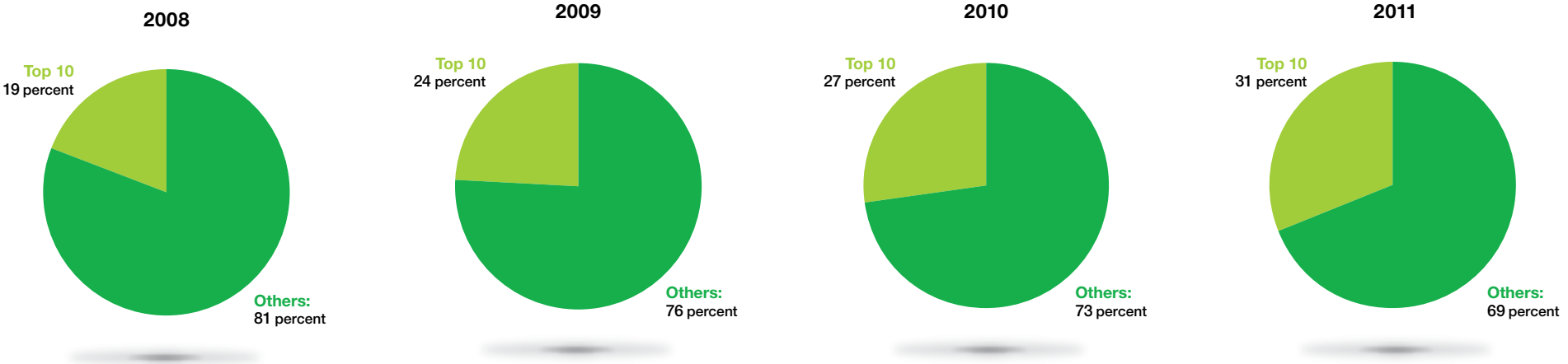


Figure 45: Top Ten Software Vendors with the Largest Number of Vulnerability Disclosures – 2008-2011

Ultimately this is a process that is helping to contribute to the declines in public exploit releases that we have seen this year. However, in the short term, the increase in the number of vulnerabilities impacting popular enterprise software, as well as the increase in critical vulnerabilities, means that IT staff who are responsible for patching and protecting production computer networks have a great deal more work to do keeping up with these disclosures than they did a few years ago. The real number of vulnerabilities from the top ten vendors has increased by 50 percent since 2008. This fact should be taken into account when planning staff capacity for vulnerability remediation.

The steps that IT staff should take to help protect the network against publicly disclosed vulnerabilities depend upon whether or not a fix is available and how quickly that fix becomes available. Fortunately, we're seeing improvements in the availability of patches. This year only 36 percent of the vulnerabilities that were disclosed have no publicly reported remedy. This is a significant improvement from previous years, when the number has hovered around 45 percent.

About 91 percent of the vulnerabilities that are patched, are patched the same day that they are publicly disclosed, which is the ideal situation. What about that other 9 percent? Most are patched within a few weeks, but the worst case scenarios can stretch out for a very long time—hundreds of days can sometimes pass between public vulnerability disclosure and patch release. This remains true even when we limit ourselves to vendors of popular enterprise software, or to vulnerabilities with public exploits. X-Force counted only 29 cases during 2011 where it took more than a week for a major enterprise software vendor to fix a publicly disclosed vulnerability with a public exploit, but it only takes one such vulnerability for an attacker to wreak havoc on a computer network.

	2006	2007	2008	2009	2010	2011
Unpatched %	46.6%	44.6%	51.9%	45.1%	43.3%	36.0%

Table 6: Percentage of publicly reported patches – 2006-2011

Patch Timeline	All	Major Vendor	Major Vendor & Public Exploit
Same Day	4054	2263	138
Week 1 (1 to 7)	132	19	4
Week 2 (8 to 14)	55	15	5
Week 3 (15 to 21)	26	3	2
Week 4 (22 to 28)	27	10	2
Week 5 (29 to 35)	27	8	2
Week 6 (36 to 42)	33	7	1
Week 7 (43 to 49)	14	6	2
Week 8 (50 to 56)	9	2	1

Table 7: Patch release timing of all software vendors vs. major software vendors – 2011 H1

These gaps are not necessarily the consequence of vendor negligence. It takes time to properly fix, package, and test an update for a commercial software application. In some cases complex interoperability concerns can have a cascading effect on different software components, requiring extensive changes in order to address a single security issue. Therefore, shaking our collective fists at software vendors may not be the best way to address this problem. There will inevitably be situations where gaps exist between disclosure and patch, and network managers need strategies to protect their networks during these gaps.

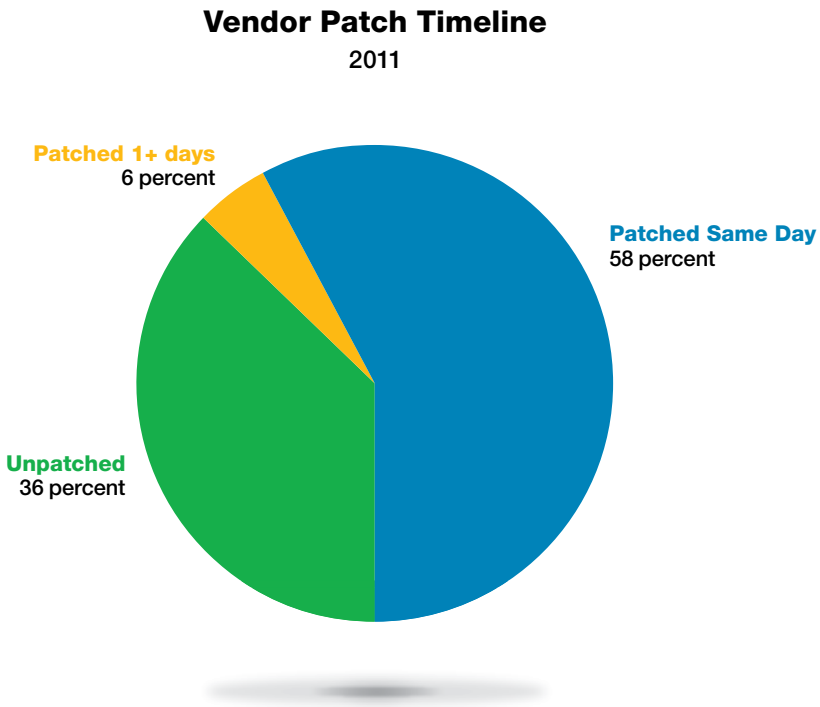


Figure 46: Vendor Patch Timeline – 2011

When the most serious security vulnerabilities are publicly disclosed, X-Force issues alerts and advisories. As a regular feature of our Trend and Risk Reports, we chart those alerts and advisories on a two dimensional graph, based on how difficult they may be to exploit as well as how valuable they may be to an attacker. These factors help us understand which vulnerabilities are likely to see widespread exploitation on the Internet.

X-Force issued thirty-four alerts and advisories during 2011. Sixteen of these vulnerabilities fit into the critical category, easy to exploit and extremely value, which is a sweet spot for malicious activity. Almost all of these vulnerabilities represent client software remote code execution issues that are exploitable through drive-by-downloads or email attachments. Most are currently being exploited in the wild.

Twelve of these vulnerabilities are categorized as valuable but more difficult to exploit—as new operating system features have made it harder to successfully obtain remote code execution from vulnerabilities, X-Force has found a growing number of serious vulnerabilities fit into this category.

Although we remain concerned that sophisticated attackers may have exploits for some of these vulnerabilities, we do not expect to see widespread exploitation on the Internet. The growth of vulnerabilities in this quadrant as opposed to the critical quadrant represents some progress in the fight against computer crime.

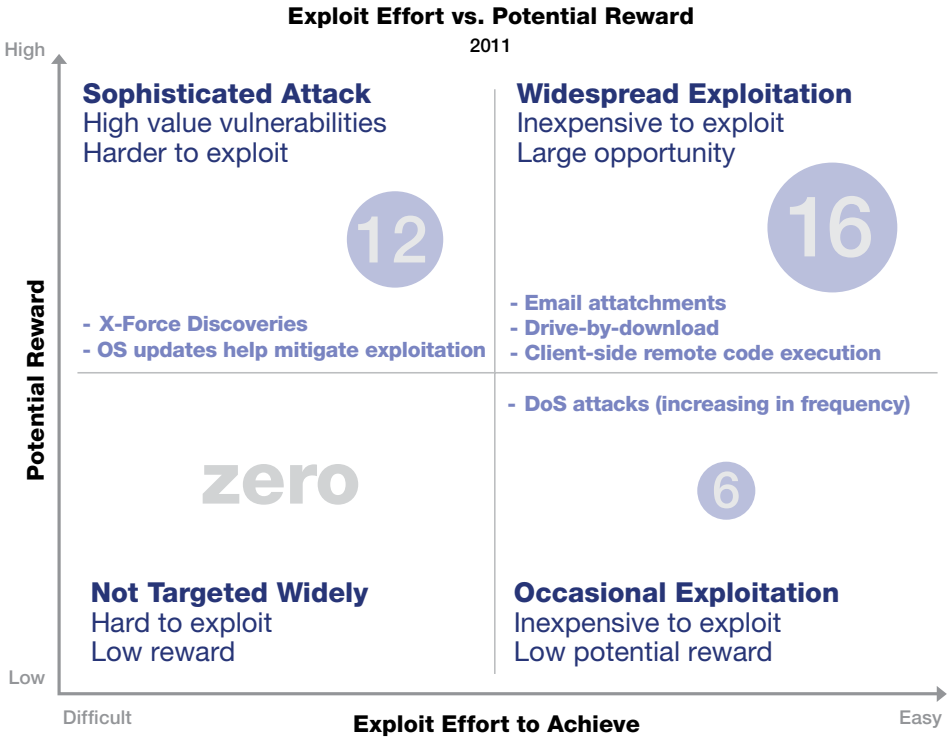


Figure 47: Exploit Effort vs. Potential Reward – 2011

Six of the vulnerabilities that X-Force issued alerts about in 2011 are denial of service issues. While denial of service vulnerabilities are less valuable than remote code execution issues, we've seen a widening interest in these vulnerabilities in past six months. Politically motivated hacktivist groups such as Anonymous have been launching denial-of-service attacks against corporate and government entities throughout the world in order to make various political statements. Most of this activity involves distributed floods of legitimate looking traffic, which can be very difficult to filter, as opposed to attacks that trigger specific vulnerabilities. However, we've begun to see some interest from these attackers in vulnerabilities that can make their attacks more effective.

The tools and techniques that these hacktivists have developed have also found their way into the hands of financially motivated attackers who appear to be using denial-of-service attacks in competitive business contexts with increasing frequency right now. With a political election in the United States this year, along with global controversies regarding intellectual property laws, we expect to see more prominent distributed denial-of-service attacks throughout 2012.

Social engineering social media: How the attackers do it

Overview

Since the widespread adoption of the Internet there have been few innovations that have had the impact of social media. Social media is shifting the way that society connects, interrelates, and shares information. The byproduct of this shift is a flood of previously difficult to gather personal and private information into a central, archivable location—namely the Internet. This treasure trove of information is particularly useful to the malicious minds of computer intrusion.

In the last seven years, social networking has gone from a fringe pastime to become the number one online activity in the world, eclipsing even use of search engines. By year-end 2011, approximately 80 percent of the global online user population (over one billion people) was using social media.[20] Naturally, such concentrated activity represents a fertile environment for an attacker. Frauds and scams that were successful years ago via email found new life on the social media forums as well as a fresh group of potential targets.

The vast amount of private information that users are pouring into social networks has shifted the paradigm of intelligence collection. Intelligence gathered from these networks has already begun to play a role in pre-attack research for the infiltration of public and private sector computing networks. As a direct result, some of the highest profile hacking attacks in 2011 began with simple Open-Source Intelligence (OSINT) collection and/or social engineering exploits executed via social media.

20 Source: comscore It's a Social World Report, December 2011

These attacks exploit a grey area in the organizational perimeter, targeting an individual and the information they volunteer, typically in a non-workplace context. Individuals associated with a targeted organization may inadvertently (or purposefully) volunteer valuable information, or introduce malware into corporate systems that result in theft or destruction of corporate data assets.

Though structuring a successful exploit leveraging social media can be challenging, the success rate of the attacks and associated payoffs have thus far proven to be worth the effort. This section explores the impact social media has had on security, paying particularly close attention to shifts in intelligence gathering and the anatomy of social engineering attacks leveraging social media platforms. The purpose of this section is to inform readers of emerging attack methodologies and their potential impact on public and private sector entities.

Intelligence gathering

It is generally accepted that intelligence gathering follows a relatively simple cycle involving the development of requirements, planning and direction, actual collection, processing, analysis, and dissemination, though the actual number of steps within the cycle may vary. A few common types of intelligence gathered within this process include Human Intelligence (HUMINT), Open-Source Intelligence (OSINT), Signal Intelligence (SIGINT), Measurement and Signature Intelligence (MASINT), and Imagery Intelligence (IMINT).

Prior to social media, methods for gathering each of these intelligence types was relatively straightforward and often required specialized focus on each. The emergence of social media has shifted collection of the intelligence sources away from individual areas towards that of simply OSINT.

HUMINT no longer requires physical contact for "interpersonal contact" and is far more public than before. SIGINT no longer requires signal interception as the media is largely publicly shared by entities, and imagery intelligence is enhanced by the world's largest repositories of pictures (Fotki, Webshots, Facebook, etc).

Social media now offers intelligence collectors a repository of information largely unparalleled in human history. Consider that a person who readily adopts social media may volunteer not only intelligence artifacts but also provide the context to those specific artifacts. By giving a public voice to the masses, social media inherently invites either the accidental or purposeful dissemination of secret information. This is evidenced by several instances where U.S. officials mistakenly posted information on classified trips or as one congressman once posted, "[b]ack in Washington. Receiving top secret intelligence briefing on Iran." Beyond the realm of the blatant indiscretion however, users often post seemingly benign information on social media such as personal email addresses, current city of residence, and educational background.

Open source intelligence gathering

The massive amount of now public or open-source intelligence (OSINT) that is available for gathering has opened a new realm in information security and attacks. The trend for conducting open-source intelligence searches grew rapidly in 2011 and is likely to continue at an increasing pace in 2012.

This massive growth has given rise to an entire realm of search tools as well as techniques. These tools include utilities that are not only focused on the actual search, but also on the mapping of found data. Commonly utilized tools such as Maltego, offer assistance in finding information and represent it in an easily consumable manner. Meanwhile, tools like Foca assist in finding information and use that information to gather more intelligence.

It has been widely publicized that law enforcement organizations are not only leveraging existing tools to mine public data on social networks but are also searching for new tools that are more powerful and granular. These efforts are interesting as they show that OSINT gathering is not only a growing trend for attackers, but also for security professionals. Indeed, the wealth of information is useful to determine who may be attacking.

In the context of computer intrusion, information like this is pure gold for social engineering attacks and authentication logic attacks such as password resets that request personal information. Attackers have been the most active in exploiting these weaknesses in social media to secure entry points into target organizations. Given the success of several high-profile attacks executed in 2011, the social engineering attack through social media is the emerging trend to watch in Advanced Persistent Threats.

How it works—not rocket science

For example, this particular exploit is a three-level attack that combines social engineering, spear-phishing, and zero-day execution to complete the agenda. Just as cheaters cruise Las Vegas casinos

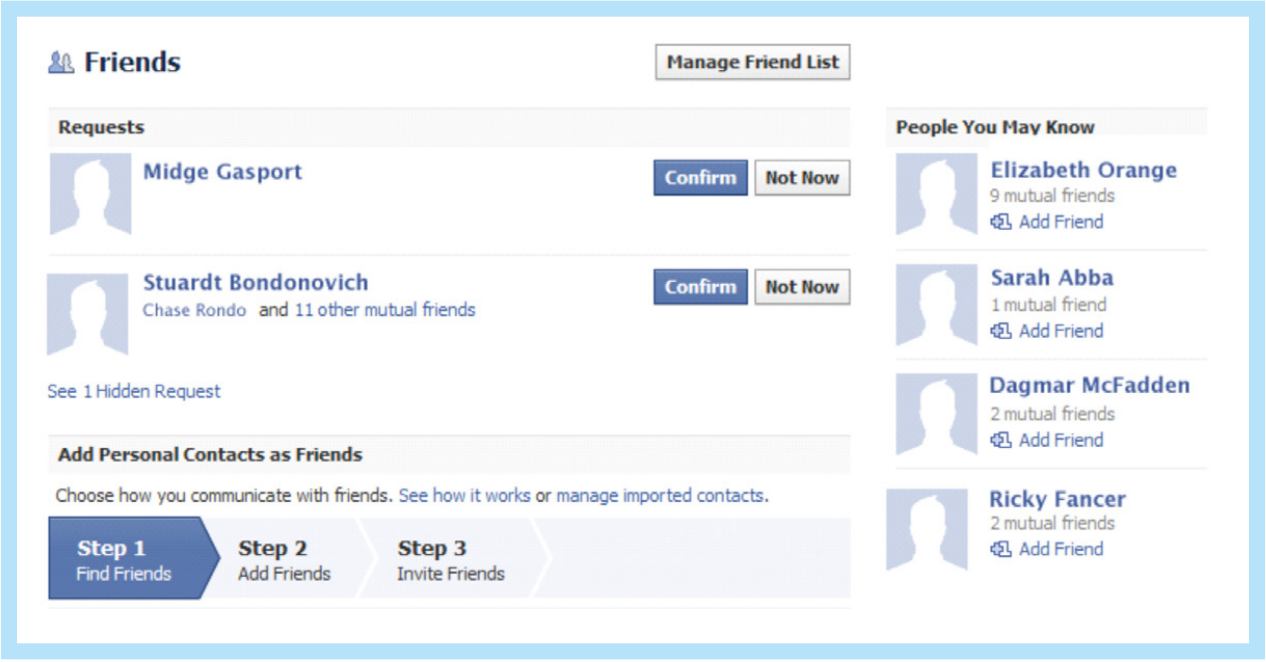


Figure 48: Example list of possible contacts to spear-phish 2011

looking for weak dealers, and cheetahs roam the Serengeti trying to spot the limping zebra, attackers troll the social networks looking for end users with large and active friend lists.

First, the attacker selects a target organization. Then they create an account on a social media forum, such as LinkedIn, and set up an alias and profile that suggests an affiliation with the organization, such as a former employee. In a down economy, in an industry experiencing a high level of merger and acquisition activity, posing as a former employee of the target can make the alias appear plausible. With the account firmly established, forums like Facebook and LinkedIn serve up to the attackers lists of potential connections from the target organization.

Once the attacker knows who to approach, the social engineering phase begins. The attacker attempts to make connections with current employees of the target. The approaches are simple but varied—getting back in touch after a few years, changed jobs and looking to broaden their professional network, recently unemployed and looking to return to the target, or wanting to connect after meeting at an industry event. A carefully worded, sometimes low-key approach has a good chance of success if the attacker approaches a large number of individuals. Getting the first connection is often the hardest. Sometimes the attacker will create another alias account from the target organization and link the two in order to create credibility. There is no mechanism for vetting false claims and representations made on social media forums, so the majority of user accounts are taken at face value and treated as legitimate.

Once the attacker has made one legitimate connection within the target, it becomes easier to gather others. LinkedIn, for example, facilitates secondary and tertiary introductions by members, as does Facebook through friends of friends. Added to that, the ability to link accounts between the major forums facilitates the attacker in establishing additional contacts from a variety of sources via the relationship with one or two legitimate individual contacts.

Then the attacker begins to analyze the profiles of each of the legitimate contacts, by collecting personal information, organization-related information, and even gauging areas of interest to determine the best approach to each individual. Establishing a baseline level of trust with these new contacts is easily accomplished—asking for simple

information, or forwarding some information that might be of interest. This allows the attacker to determine which end users are the most active, and which would be most likely to "assist" in gaining access to the target.

Finally, once the attacker has sufficiently primed the individuals, the spear-phishing phase of the attack can begin. This attack is most successful when the attacker has access to the corporate email accounts of the end users. Even one or two corporate emails allow the attacker to understand the naming conventions and guess at additional email accounts. A well crafted email—a job vacancy announcement for a disgruntled employee, a professional survey for a job seeker, a link to an educational video for a career-transition individual—anything that seems legitimate and may be loosely associated with work has the potential to attract the attention and acceptance of at least one of the end users within the target's corporate computing environment. These emails usually contain a malicious payload, a link, a download, or an *.exe file, and it is the end user that sets the final stage of the attack in motion.

Figure 49: Example of spear-phishing email 2011[21]

Then the attacker is "inside" and the zero-day attack can be executed. The one good thing—the difference between a failed attack and a successful one—is that the end user has to take an action so that the exploit is activated.

Steps organizations can take to mitigate social media risks

A September 2011 Ponemon Institute study[22] indicated that only 35 percent of respondents had a written social media policy. Of those organizations, only 35 percent actively enforce it. The same study indicated that virus and malware attacks on corporate computing systems increased by over 50 percent since their employees started using social media. Unfortunately, there is no software or suite of end-point products that can be easily deployed to defend against social engineering. As with most threats aimed at human beings, the best way to manage such risks is through policy and education.

21 Source: http://contagiodump.blogspot.com/2011/10/cve-2011-0611-pdf-2011-10-24-northkorea.html.
22 Source: http://www.ponemon.org/ Global Survey on Social Media Risks September 2011. The study surveyed 4,640 IT and IT security practitioners in the United States, Canada, United Kingdom, France, Germany, Italy, Australia, Singapore, Hong Kong, India, Brazil and Mexico with an average of 10 years experience in the field.

These efforts can be broken down into two specific focus areas: actions for business environments and actions for users. As social media is mostly a personal experience, primarily found outside of the workplace, users are largely responsible for their own privacy and security. However, it is imperative for businesses to create policies and procedures to assist in guiding employees as well as protecting company brand and assets. These efforts resemble "Security Awareness" programs of the past, but should uniquely contain guidance to end-user responsibilities such as:

Enable security and privacy settings. The major social media forums have basic privacy settings available to users. It is important that end users understand what security and privacy controls are available to them in the forums that they use regularly, even if they do not consider themselves to be active users. In order to decrease exposure to spam, scams, and opportunistic attackers, security and privacy controls should be set to maximum levels. End users must also understand that any security and privacy action they take will be minimized to the lowest levels at work within their social circle. If one friend, for example, uses only minimum security and privacy settings, it creates an avenue of exposure to all connections within their circle, regardless of the higher security postures adopted by those connections. In other words, if Facebook Friend1 limits his posts and allowable contacts to only his circle of friends, but Friend2 allows posts and contacts to be available to everyone, then anything posted to Friend2's wall can be viewed by everyone in Facebook. Depending upon Friend2's privacy settings, the posts may even be internet searchable.

Encouraging end users to adopt a "default deny" stance in their social media presence seems somewhat antithetical to participating in social media, but it is that level of security awareness that can ultimately protect them from socially engineered attacks.

Friend only friends. Social engineering attacks would not be so successful if they were not clever in some respects. Just as with real-world con artists, social media attackers begin their attacks by attempting to gain a certain level of trust from their targets. Pretending to be an old classmate, a former colleague, or a friend-of-a-friend or relative is not at all uncommon. Faking a tangential work-related relationship via LinkedIn, for example, lends almost instant credibility to the attacker given LinkedIn's status among social media as a business-oriented, drama-free forum. Yet, making connections via LinkedIn using false claims of previous working relationships, or having met the target at an industry conference or event, is plausible enough to convince the target to accept the connection. End users looking to increase their online status or sphere of influence may routinely accept random requests simply to increase their numbers.

Despite the various incentives and rewards of having large and diverse followings and friends lists, it is important to remember that this is exactly the kind of environment in which the attacker seeks to hide. End users should consider friendship requests carefully, accepting those based on prior real world relationships, or some level of trust within the social media forum, such as forum-sponsored clubs, common interests, and so on. Random requests for friendship based on secondary or tertiary mutual connections should be carefully screened. Private communications and requests for detailed personal information from new friends, known only to the end user via social media, should always be viewed with caution, particularly when the request involves real world contact information.

Use caution with links and downloads. Links and downloads have been a favorite vehicle for attackers to deliver malware to their targets since email became ubiquitous in the late 1990s. The trend has consistently evolved into social media forums. End users must exercise extreme caution, and carefully consider the source before they click on any links, or download anything (particularly executables), from unknown or untrusted sources. Many new "friends" may try to deliver a malicious payload via personal messages directing end users to hilarious YouTube videos, fun screensavers, bogus fan forums, or awesome free gameware. Random attackers often try to deliver malicious payloads via spam. Facebook and other forums routinely post warnings when they are alerted of widespread attacks. End users should subscribe to any alert services offered by the respective forums.

Be wary of contests, gifts, prizes, and special offers. "You may already be a winner." Prizes and other special offer scams also date back to the early days of email, but continue to perform strongly in social media forums. Scammers typically use this type of offer, for example 'free high value gift cards available to forum members,' to direct end users to a dead end website that will load cookies or even spyware, or more often, to fake websites that mimic legitimate businesses or brands, and require the end user to fill out complex applications or surveys to qualify for the bogus contest. Either way, the scammer is collecting personal information from their targets. Facebook hosts a user community called Facecrooks that alerts members to scams, providing details and remedy information when available. End users should subscribe to any scam alert services offered by their respective social media forums.

Consider limiting work-related information. End users should always consult their employers' appropriate use policies for social media when communicating information about the organization, colleagues, clients, products, services, and projects in which they are currently involved. Aside from specific information, end users may want to consider referencing their industry or employer only in general terms to avoid inadvertent disclosures. Careful screening of work-related information is becoming increasing important as more end users seek employment or networking opportunities via social media, and more employers scan social media to evaluate current or prospective employees.[23]

In the absence of written corporate policy, common sense may be your best guide in terms of posting work-related information.

Even a careless reference may sometimes reveal more than the end user originally intended. The best rule of thumb concerning any posts to social media forums is that, despite security and privacy settings and despite good intentions and even accidents, social networking is designed to share information globally via the internet. All posts should be considered carefully, as they go public instantly, and are essentially irretrievable.

Future trends

Social media attacks will continue to grow in influence and range in the future. This expansion includes the venture into seemingly unrelated technology. For example automobiles are already sporting interfaces to the Internet and interconnection through social media. With this expansion, social media will continue to evolve and represent an easily exploitable arena for attackers.

Enterprise organizations need to develop and enforce policies and users must become capable and knowledgeable in protecting themselves.

23 Despite negative reaction in 2011 to this trend, many employers openly use social media as part of the employment process, including background and credit checks.

Top 10 common CSIRP mistakes

Computer Security Incident Response Plans (CSIRP), a cornerstone in any environment with anything more advanced than an expensive calculator, are absolutely crucial when formulating a response to security incidents involving networks, computers, or electronic data. During an incident, a CSIRP is the map that guides your response.

This article describes several of the most common mistakes involving CSIRPs. IBM's Emergency Response Services (ERS) team is intimately involved in CSIRP plans because we frequently respond to customer emergencies and develop custom CSIRP plans for our customers. ERS is fortunate to observe what works and what doesn't. We'll describe several of the most commonly observed shortcomings of CSIRP plans in this article.

#1 Making a CSIRP too complex

When designing your CSIRP, it is best to keep in mind that the audience will be reading the document during a crisis, not while relaxing at a coffee shop with a latte and freshly baked pastry in hand, slowly absorbing the material while listening to classical music. While we may dream of an incident involving warm pastries and unlimited time to digest a plan, typically it is not going to happen. There may be stress. Individuals may be panicked and worried about their jobs. Executives who may or may not understand the fine technical points of what is happening may be upset because the local news media is asking questions. Crying and assuming the fetal position will soon follow... you get the picture.

In the situation described above, do you have time to consult a large, detailed CSIRP plan? Clearly, you may not. CSIRPs must be crisp, clear, and concise. If an employee who is unfamiliar with the document cannot quickly examine the processes described within the CSIRP, understand the chain of command, and perform the necessary actions, your CSIRP may be too complex. Of course, making a CSIRP too simple is also a potential pitfall; striking the right balance between brevity and actionable direction is essential to a successful CSIRP.

#2 Overloading key personnel

Every organization has a Joe. Joe knows everybody and every system, router, cable, and the top three coffee machines in the building. Joe is the person to whom we all look during an incident. Joe, undoubtedly, is the best person around for minor incidents and can handle them from beginning to end. When we develop CSIRPs for our customers, we quickly find the Joe of the organization during our standard questioning: Who is in charge of anti-virus? Joe. Who communicates with executives? Joe. Who schedules the company holiday party? Joe.

Joe is fantastic at what he does from eight to five. However, when an incident stretches on for days, Joe can't be your go-to guy for 72 hours straight. Separating duties during an incident and having previously designated backups in place is necessary if an organization does not want sleep-deprived, overloaded employees scheduling the holiday party in June.

#3 Treating an incident as a serial process
During a large-scale incident, multitasking is essential. Incident Managers who only look at an incident as a serial process will be unable to resolve an incident in a timely manner. While each incident is unique, they all comprise a number of short term goals. Pushing out new anti-virus signatures, patching systems, leading investigative efforts, informing employees and customers of your current status, fetching additional supplies of caffeinated beverages, and other important tasks are all unique processes and should be treated as such. One common failure is when a company focuses on only one of these tasks at a time and neglects other important tasks that may be completed in parallel.

#4 Failing to establish proper lines of communication
When responding to an incident, numerous individuals and vendors may be asked to assist. The incident manager—the individual responsible for managing the 'boots on the ground'—should be a master communicator. Communication must be orderly, efficient, and follow the proper channels. Imagine a war room with 25 different people where each of these people take orders from 15 others and no proper line of communication exists. Progress is stifled and an incident that should have been solved 24 hours ago drags on. Communication skills can be just as important as technical skills when confronting an incident. Without one voice, one vision, and one coach, the rest of the team is often doomed to fail. A CSIRP should address and codify lines of communication to ensure that all information is in the hands of the people who need it, not stifled in compartmentalized fiefdoms.

#5 Focusing on what's easy, not what needs to be done
During every incident, the urge arises to focus on the easy tasks versus what needs to be done. This is akin to filling up the window washer fluid on a car when the engine won't start. Sure, the window washer fluid does need to be eventually filled, but without a working engine, your car is useless. The same is true for an incident. There are hard tasks and easy tasks, but regardless of difficultly, some tasks just need to be completed. Failing to focus your energy on the essential problems, whether easy or hard, can cause prolonged headaches and prolonged incidents.

#6 Focusing on what's stimulating, not what needs to be done
During some incidents, the responder will discover some bits of interesting information and become focused on a chase down an unrelated rabbit hole. The newly discovered item may be extremely captivating but it does not play a material role in resolving the incident. Endless hours can be spent in the rabbit hole but the rabbit is out of the country on vacation. Remember, you are hunting rabbits and not observing the architectural variations of the rabbit hole.

#7 Ditching the CSIRP
The urge will occasionally arise to throw out the CSIRP because it doesn't address the specific situation at hand. There is a reason why the document does not address the latest email virus. The CSIRP is not meant to be an all-inclusive guide on how to confront every specific incident. Rather, the document is a blueprint for lines of communication, roles, required notifications, and steps to be taken to respond to the incident. While each incident is unique, the document should allow for a response to be formulated by quickly understanding the identities of the key players who should be included, their roles, and communication protocols. With this structure in place, the necessary steps may then be taken to address the incident at hand.

#8 Making a policy, not a plan

Always remember that the "P" in CSIRP stands for "Plan" and does not stand for "Policy." Occasionally, ERS reviews a CSIRP that reads more like a policy versus a plan. What is the difference? A plan contains actionable steps and roles while a policy states overarching guidelines to be applied within the organization. When an incident occurs, do you really want to be reading company policy to formulate a plan? Of course not. You would like a well thought out plan that tells you what to do.

#9 Failing to assign an owner

Your CSIRP may have a lot in common with your cat, Mr. Fluffy. Both develop over time, require maintenance and attention, and should have owners responsible for their well-being. Occasionally, when an incident has taken place, CSIRPs are pulled from the depths of the network only to find that the document was last updated when Vista was cool. One by one, the phone numbers of key personnel are found to be disconnected. Even the conference room that was originally designed as a war room has been re-purposed as the company daycare center. No owner was assigned to the document, and, without a caretaker, the document became outdated and its value diminished.

When establishing a CSIRP, assign an owner to the document. This owner is responsible for updating the document, ensuring that the procedures it contains are still relevant, and coordinating annual testing. Without a specific owner, the document may languish, become stagnant, and cause an increased response time to incidents.

#10 Neglecting the after-action review

The most valuable lessons from any incident can be learned from the after-action review. Even if it seems like everything went as planned during an incident, it is likely that an after-action review can bring potential improvements to light. There is no shame in pointing out mistakes or issues that need to be improved; any of these only make the CSIRP stronger and more capable to address your needs during future incidents.

At the conclusion of an incident, the major players should meet and discuss how well the CSIRP performed. Unfortunately, in the haste to forget the headaches from weeks past, the after-action review is often a neglected valuable step in the CSIRP process.

Incident response—preparing your infrastructure for response at scale

Incident response (IR) is not something most security personnel think about in our daily jobs. Rather, we think about defensive and offensive positions, identity management, code review, and other day-to-day operations. But what happens when these mechanisms fail for real? How does an organization recover from an intrusion, a virus outbreak, or a sensitive data leak? Incident response should be a planned process, laid out well in advance of its necessity to avoid quick decisions with poor consideration of repercussions. In its simplest form, IR planning mostly involves identifying expert troubleshooters within your organization that would be best at identifying and eradicating serious security issues. These individuals need not be dedicated incident responders, but would be available for immediate engagement. In this type of scenario, incident response is not typically systematic. Practitioners tend to play whack-a-mole, knocking out individual infections with local scans and solving more problems with sneaker-net and a boot CD rather than exercising pervasive monitoring and mass cleanup procedures.

All great incident response really requires is the ability to store everything and make coherent sense out of it at will.

For small organizations this can suffice. It isn't a bad approach, but it doesn't scale past a small handful of machines. Steps beyond that usually require an actual investment in infrastructure, setting up the incident response team with tools to capture and analyze data across the enterprise. With all of the logging and analysis platforms and appliances available today, it is easy to imagine that scalable Incident response is just another appliance away.

Incident response is not easy and requires the ability to store everything and make coherent sense out of it at will. Unfortunately, even this approach can fail to scale beyond a double handful of machines. Once an incident encompasses more than a few dozen machines, the more simplistic models of incident response require an inordinate amount of force to be workable. Although the truism "if brute force isn't working, you aren't using enough" can apply, most processes borne of it are correspondingly expensive and unwieldy. Say, for example, your incident response plan dictates that a system infected with an information-stealing virus must be shut down and imaged, how well (and quickly) does that work for 50 machines? For 1500? How do you even determine which machines are infected when the virus is hours or days from being detected by your antivirus solution? This article will attempt to discuss a few of the basic steps we find most helpful in preparing to deal with these types of scenarios in a manner that scales both financially and temporally.

Preparation: The solid foundation of all incident response

Although the specific acronym varies, the outlines of traditional incident response doctrine always begin with "P" standing for "preparation." Incident response at scale involves a good deal more preparation than smaller environments, but when properly prepared, the effort required in later steps can be substantially similar. Fortunately, many (if not all) of the steps involved in preparing for good incident response are simply good infrastructure practices in general and as such are already necessary components of a well-managed environment. Indeed, much system administration could be considered "low-grade" incident response. Centralized authentication, patch management, inventory management, logging, access control, and automation are all basic components of running a successful computing infrastructure and each of these has specific implications to incident response. Discussing all of these is beyond the scope of this article, but two of them in particular are incredibly important to scaling incident response: logging and automation. These are two key success factors that we see customers miss time after time.

Not logging will hurt you more than it hurts me

One of the first things a seasoned incident responder will ask is "Where are your logs?" When that responder encounters a situation where the answer to that is less than desirable, they will do everything they can to help, but with the knowledge that their chances of successfully identifying the issue at hand and rooting out the cause are rapidly diminishing. They have learned that success in incident response is not unattainable perfection but sufficient closure. It can hurt the customer more that they cannot identify the individuals involved in a data breach by anything more than having been "in the system" during a broad swath of time, elevating exposure from dozens to millions of records.

Logging provides both the incident responder and the system administrator critical knowledge traction to determine what happened in the infrastructure at a given time, past or present. Unfortunately, much like the rest of a well-run security environment, pervasive logging also tends to be one of the first areas cut in contemporary infrastructure, as it consumes precious system, network, and financial resources for something that is only occasionally needed. The primary keys to successful logging are filtration and centralization. It is an atypically well-planned environment that can support full logging of every single operation. Far more often, incident responders must work with system administrators to determine the minimum set of logging necessary to provide reasonable response while avoiding consuming excessive resources. The key is to strike a balance between retention and cost/performance. As an example, it is rarely necessary (but entirely possible) to log every object access in a Windows systems, but failing to log privilege use can have critical implications for both response and administration. Imagine that in a well-configured domain, a domain administrator (properly using low-privilege personal credentials) briefly elevates their privilege to change a DNS setting, accidentally breaking it. In this situation, administrators can quickly see what happened, who did it, and what to fix. Now imagine that the low-privilege credential was actually not that user but an attacker that compromised the credentials.

Once logs are collected, they must be stored, and there are few worse places to store them than on the system that is generating them. Logs consume valuable disk space, can be lost in a system failure, or even modified by an attacker in the event of an intrusion. Centralized storage can help mitigate or at least offload these issues to a separate system. Transferring logs to a central system can take multiple forms depending on the organization's needs and a risk calculation of what constitutes an acceptable loss of data. From the perspective of the incident responder, an ideal arrangement would often be real-time delivery with end-to-end guarantees via a mechanism like the Reliable Event Logging Protocol (RELP), effectively eliminating an intruder's window to alter the logs. RELP can provide reliable event logging over the network. The key again is balance and to not allow the best to be the enemy of the good. It is more valuable to have a sub-optimal log collection system polling logs in batches from systems, than to have nothing at all.

A rough rule of thumb in determining how often to poll logs is to decide how long of a window would be acceptable for an attacker to be able to modify logs, and then divide by two. Some organizations avoid central log storage because it appears expensive, quoting fast SAN disk costs and application server hardware prices. Central logging need not be so costly. Other than ideally being standalone and administered separately from the rest of the environment (no trust or shared credentials), and unless log analysis is being performed on-system, the hardware and availability requirements should not be any greater than those of a typical file server.

Automation is your second, third, and Nth best friend

Automation in system administration and incident response is the difference between the major league baseball and Saturday pick-up softball. It is what allows some organizations to operate at a server-to-administrator ratio of over 1000:1 and incident responders to surgically deal with thousands of compromised machines at once. Fortunately for incident response, automation is usually at least partly built into environments comprising more than a handful of machines as an administrator generally is not going to choose to singly install patches on more than two systems without some sort of patch management tool to control the process. In addition, many environments have "agents" installed to provide endpoint security, asset management, antivirus management, and a plethora of other necessary day-to-day administrative functions. It is often the case that the greatest difficulty an incident responder can face with these tools is knowing which of them are available and how to use them. Any of these automation tools may be used to provide the incident responder with valuable custom queries of various system states, but should be carefully selected according to their refractory period and how much they modify a system.

For example, assume that your incident response team has identified a new virus that is yet undetected by the antivirus solution but known to create files matching a certain regular expression in a certain set of directories. The infrastructure teams have patch management, asset management, and antivirus tools running on all the potentially-infected systems. One solution could be to deliver a batch file that searches for the indicator files on the systems via the patch management tool and reports back by uploading a results file to a central server. For some situations, this may be the only approach, but it modifies the potentially affected systems and tends to leave the response process more open to interference from a malicious party. If, however, the asset management tool can report on files matching a specific pattern, it could be preferable as it may not modify the end system and could appear to be normal activity to a malicious party. Cleaning up the infection could well be done by either the patch management or antivirus systems, depending on the specific situation. The key is to first be aware of what tools/capabilities are already available for the environment (or work with infrastructure teams to implement mutually-beneficial tools) and then choose the right tool for the job.

Last and certainly not least in the way of automation is scripting. Although automation tools often have their own scripting languages that can be put to good use, little is more effective and efficient in the incident response process than incident responders that are capable of writing scripts in a generic language like Python or Perl to control the automation tools and fill the gaps the tools might miss. Having a seasoned system administration programmer on the incident response team or available on short notice can be an invaluable asset for speed and completeness of response.

Last and foremost: Authentication

A third and somewhat forgotten critical key for incident response is authentication. One may note that much of the above is (or should be) predicated on having strong, centralized authentication. Without central authentication the administrator and responder typically have no way to efficiently query systems, apply fixes, or otherwise deal with systems without resorting to brutish tactics like gathering and storing per-machine passwords. Some notably large environments do get along without centralized authentication, instead relying on remote agents that periodically execute scripts under administrative privileges, but the response time of this type of setup can be prohibitive to good administration and incident response and should be avoided with all diligence.

Work smarter and make good friends

Good infrastructure practices should translate directly to good incident response. The tools and procedures you need to leverage to build a more fault-tolerant, repeatable, and scalable computing infrastructure are the same tools and procedures you should leverage to respond to incidents smoothly and quickly. Cultivate good relationships with system administration teams and learn what tools they already have in place to be sure you understand and know how to operate those tools. That way, the incident response team can support them when necessary.

Data security and privacy, understanding the differences to help achieve compliance

Companies rely on data to support daily business operations, so it is essential to ensure privacy and protect data no matter where it resides. According to the Verizon Data Breach Investigation Report, database servers are the primary source of breached data, representing 92 percent of compromised records. Unfortunately there is a large disparity in the time required for attackers to penetrate databases compared to the time required to recognize a break in and to remediate the breach. It takes attackers days to penetrate defenses but often weeks or months for organizations to figure how, where, and when they were compromised and then often weeks or months again to remediate the issue.

Data security and privacy are made even more complex because different types of information have different protection and privacy requirements; therefore, organizations must take a holistic approach to protecting and securing their information. This includes:

- **Data discovery and classification**—Organizations need to understand where data exists across the enterprise and how it is related. This allows them to classify sensitive data properly so it gets proper treatment throughout its lifecycle.
- **Data redaction**—Sensitive data also resides in documents, forms, and scanned images. Protecting this unstructured data requires privacy policies to redact (remove) sensitive information while still allowing needed business data to be shared. These unstructured documents could be attachments in the database.
- **Data encryption**—Encrypting databases may be required by many regulatory mandates. Organizations need a single solution that scales to help protect heterogeneous data types. This can be a nice complement to database activity monitoring because organizations can build a defense in depth approach.
- **Static data masking**—Much focus is given to production environments, but the security of non-production environments shouldn't be overlooked. De-identifying sensitive data in non-production databases yet maintaining usability for application development, testing, training processes, and QA work not only helps facilitate business processes, but also helps ensure the principle of least privileges. Those without a valid business need to know should not have access to sensitive data.
- **Monitoring**—Securing and continuously monitoring access to databases, warehouses, and fileshares gives insight into the who, what, when, and how of transactions to help organizations validate the integrity of data.
- **Vulnerability assessments**—Harden databases to help mitigate risks such as mis-configurations or default settings.

Making sense of the buzz: Why the growing focus on data protection?

According to Forrester Research's February 2011 independent report, Forrsights: The Evolution Of IT Security, 2010 To 2011, IT security remains a hotbed of activity and growth as firms struggle with a more menacing, capable threat landscape; respond to a growing body of regulation and third-party requirements; and adapt to an unprecedented level of IT upheaval. Much of this focus is specifically positioned around a few key themes: new cyber security threats such as Stuxnet and Aurora; changing IT architectures such a virtualization in the data center; and growing pressures around third-party mandates.

During the past several years, according to the Forrester report, "security has steadily risen in visibility achieving board-level attention and support." For example, Forrester's research indicates 54 percent of enterprise Chief Information Security Officers (CISOs) report to a C-level executive and 42 percent of them report outside of the IT department. These percentages reflect the increasing business relevance security has in organizations of all types, across diverse industries. The number of organizations that view security as a high or critical priority is now at its highest level in recent years.

Let's delve into the details about the many factors are fueling this increased focus on data security and privacy.

Changes in IT environments and evolving business initiatives

Security policies and corresponding technologies should evolve as organizations embrace new business initiatives such as outsourcing, virtualization, cloud, mobility, Web 2.0, and social networking. This evolution means organizations should think more broadly about where sensitive data resides and how it is accessed. Organizations should also consider a broad array of sensitive data, including customer information, trade secrets, development plans, and competitive differentiators.

Smarter, more sophisticated attackers

Many organizations are struggling with the widening gap between attacker capabilities and security defenses. The changing nature, complexity, and larger scale of outside attacks are cause for concern for organizations. According to the same Forrester report mentioned, security attacks now have a far more damaging business impact compared to ten years ago. Previously, the most critical concern was virus outbreaks or short denial-of-service attacks, which would create a temporary pause in business operations. Today, the theft of customer data or corporate data, such as trade secrets, could result in billions of dollars of lost business, fines and lawsuits, and irreparable damage to an organization's reputation.

Compliance mandates

The number and variety of compliance mandates are numerous, and they affect organizations around the globe.

Along with the rising number of compliance mandates is the increased pressure to show immediate compliance. Enterprises are under tremendous time pressure and need to show immediate progress to the business and shareholders, or face reputation damage and stiff financial penalties.

Information explosion

The explosion in electronic information is mind boggling. IDC estimates that 45 gigabytes of data currently exists for each person on the planet, or an astonishing 281 billion gigabytes in total. While a mere five percent of that data will end up on enterprise data servers, it is forecast to grow at a staggering 60 percent per year, resulting in 14 exabytes of corporate data as of 2011. The information explosion has made access to public and private information a part of everyday life. Critical business applications typically collect this information for legitimate purposes. However, sensitive data is subject to theft and misuse given the interconnected nature of the Internet and information systems, as well as enterprise ERP, CRM, and custom business applications.

Inside threats

A high percentage of data breaches actually result from internal weaknesses. Examples range from employees, who may misuse payment card numbers and other sensitive information, to those who save confidential data on laptops that are subsequently stolen. Organizations are accountable for protecting data no matter where the data resides—including with business partners, vendors, or other third parties.

In summary, organizations are focusing more heavily on data security and privacy concerns. They are looking beyond developing point solutions for specific pains, and towards building security policies, privacy policies, and procedures into the enterprise.

Understanding the difference between security and privacy

Security and privacy are related, but they are distinct concepts. Security is the infrastructure-level lockdown that prevents or grants access to certain areas or data based on authorization. In contrast, privacy restrictions control access for users who are authorized to access a particular set of data. Data privacy addresses limitations or restrictions on those who have a legitimate business purpose to see data. That business purpose is usually defined by job function, which may in turn be defined by compliance.

Some examples of data security solutions include database activity monitoring and database vulnerability assessments. Some examples of data privacy solutions include data redaction and data masking. In a recent case illustrating this distinction, physicians at UCLA Medical Center were caught going through celebrity Britney Spears' medical records. The hospital's security policies were honored since physicians require access to medical records, but privacy concerns arose since the physicians were accessing the file out of curiosity and not for a valid medical purpose.

The stakes are high: Risks associated with insufficient data security and privacy

According to 2010 Ponemon research, for the fifth year in a row, data breach costs have continued to rise. The average organizational cost of a data breach in 2010 increased to $7.2 million, up 7 percent from $6.8 million in 2009. Total breach costs have grown every year since 2006. Data breaches in 2010 cost their companies an average of $214 per compromised record, up $10 (5 percent) from 2009.

The most expensive breach studied by Ponemon in 2010 took $35.3 million to resolve, up $4.8 million (15 percent) from 2009. The least expensive data breach was $780,000, up $30,000 (4 percent) from 2009. As in prior years, data breach cost appears to be directly proportional to the number of records compromised.

Other potential negative impacts include fines or criminal responsibility, erosion in share price caused by investor concern, and negative publicity resulting from a data breach. Irreparable brand damage results when a company is identified as one that cannot be trusted.

Some common sources of risk include:

- **Excessive privileges and privileged user abuse.** When users (or applications) are granted database privileges that exceed the requirements of their job function, these privileges may be used to gain access to confidential information.
- **Unauthorized privilege elevation.** Attackers may take advantage of vulnerabilities in database management software to convert low-level access privileges to high-level access privileges.
- **SQL injection.** SQL injection attacks involve a user who takes advantage of vulnerabilities in front-end web applications and stored procedures to send unauthorized database queries, often with elevated privileges. Using SQL injection, attackers could even gain unrestricted access to an entire database.
- **Denial of service.** Denial of service (DoS) may be invoked through many techniques. Common DoS techniques include buffer overflows, data corruption, network flooding, and resource consumption. The latter is unique to the database environment and is frequently overlooked.
- **Exposure of backup data.** Some recent high-profile attacks have involved theft of database backup tapes and hard disks which were not encrypted.

Leveraging a holistic data security and privacy approach

Organizations should have a holistic approach to data protection. This approach should protect diverse data types across different locations throughout the enterprise, including the protection of structured and unstructured data in both production and non-production (development, test, and training) environments. Such an approach can help focus limited resources without added processes or increased complexity. A holistic approach also helps organizations demonstrate compliance without interrupting critical business processes or daily operations.

To get started, organizations should consider four key questions. These questions are designed to help focus attention on the most critical data vulnerabilities:

1. Where does sensitive data reside across the enterprise?
2. How can access to your enterprise databases be protected, monitored, and audited? How can data be protected from both authorized and unauthorized access?
3. Can confidential data in documents be safeguarded while still enabling necessary business data to be shared?
4. Can data in your non-production environments be protected, yet still be usable for training, application development, and testing?

The answers to these questions provide the foundation for a holistic approach to data protection. They help organizations focus in on key areas they may be neglecting with current approaches.

1. Organizations can't protect data if they don't know it exists. Sensitive data resides in structured and unstructured formats in production and non-production environments. Organizations need to document and define all data assets and relationships no matter the source. It is important to classify enterprise data, understand data relationships, and define service levels. The data discovery process analyzes data values and data patterns to identify the relationships that link disparate data elements into logical units of information, or "business objects" such as customer, patient, or invoice.

2. Database Activity Monitoring provides privileged and non-privileged user and application access monitoring that is independent of native database logging and audit functions. It can function as a compensating control for privileged user separation-of-duties issues by monitoring administrator activity. The technology also can improve database security by detecting unusual database read and update activity from the application layer. Database event aggregation, correlation, and reporting provide a database audit capability without the need to enable native database audit functions, which are also a part of database activity monitoring. Database activity monitoring solutions should be able to detect malicious activity or inappropriate or unapproved database administrator (DBA) access.

3. Data redaction can remove sensitive data from forms and documents based on job role or business purpose. For example, physicians need to see sensitive information such as symptoms and prognosis data whereas a billing clerk needs the patient's insurance number and billing address. The challenge is to provide the appropriate protection, while meeting business needs and managing data on a "need-to-know" basis. Data redaction solutions should protect sensitive information in unstructured documents, forms, and graphics.

4. De-identifying data in non-production environments is the process of systematically removing, masking, or transforming data elements that could be used to identify an individual. Data de-identification enables developers, testers, and trainers to use realistic data and produce valid results, while still complying with privacy protection rules. Data that has been scrubbed or cleansed in such a manner is generally considered acceptable to use in non-production environments and helps ensure that even if the data is stolen, exposed, or lost, it will be of no use to anyone.

A three-tiered approach to ensure holistic data protection

Understand and define

Organizations should discover where sensitive data resides, classify and define data types, and determine metrics and policies to ensure protection over time. Data can be distributed over multiple applications, databases, and platforms with little documentation. Many organizations rely too heavily on system and application experts for this information. Sometimes, this information is built into application logic and hidden relationships might be enforced behind the scenes.

Finding sensitive data and discovering data relationships requires careful analysis. Data sources and relationships should be clearly understood and documented so that no sensitive data is left vulnerable. Only after understanding the complete landscape can organizations define proper enterprise data security and privacy policies.

Secure and Protect

Data security and privacy solutions should span a heterogeneous enterprise and protect both structured and unstructured data across production and non-production environments. They should help secure sensitive data values in databases, in ERP/CRM applications, and in unstructured environments such as forms and documents. Key technologies include database activity monitoring, data masking, data redaction, and data encryption. A holistic data protection approach helps ensure a lockdown of all organizational data.

Structured data: This data is based on a data model and is available in structured formats like databases or XML.

Unstructured data: This data is in forms or documents which may be handwritten or typed, such as word processing documents, email messages, pictures, digital audio, and video.

Online data: This is data used daily to support the business, including metadata, configuration data, or log files.

Offline data: This is data in backup tapes or on storage devices.

Monitor and audit

After data has been located and locked down, organizations may need to prove compliance, be prepared to respond to new internal and external risks, and monitor systems on an ongoing basis. Monitoring of user activity, object creation, database configuration, and entitlements help IT professionals and auditors trace users between applications and databases. These teams can set fine-grained policies for appropriate behavior and receive alerts if these policies are violated. Organizations should quickly show compliance and empower auditors to verify compliance status. Audit reporting and sign-offs should help facilitate the compliance process while keeping costs low and minimizing technical and business disruptions. In summary, organizations should create continuous, fine-grained audit trails of all database activities, including the “who, what, when, where, and how” of each transaction.

Conclusion

Protecting data security and privacy is a detailed, continuous responsibility which should be part of every best practice. Organizations should consider data security and their privacy approach delivered through the three-tiered strategy of Understand and Define, Secure and Protect, and Monitor and Audit.

Section III
Software Development Security Practices

In the Software Development Security Practices section of this report, we present processes and techniques for addressing security during software development. 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 IBM AppScan group in all stages of application development and share insights on what was discovered.

Conclusions from real-world web application assessments

Methodology

The IBM AppScan OnDemand service is a cloud-based offering that helps customers identify and remediate web application vulnerabilities without the need to purchase and maintain software or employ highly skilled and specialized application security staff. IBM Application Security Analysts use IBM AppScan Enterprise Edition software to analyze applications for security vulnerabilities that, if left unresolved, could result in breaches and the potential loss of data such as customer and employee records or corporate intellectual property. The IBM AppScan Enterprise Edition software tests for common web application vulnerabilities including cross-site scripting, buffer overflow, and flash/flex application and Web 2.0 exposure scans. Additionally the offering includes the ability to scan and detect embedded malware in web properties providing further protection against cyber-attacks.

IBM collated real-world vulnerability data from 237 Security tests conducted in 2011 while performing Security Assessments using IBM AppScan. These assessments combine the application security assessment results obtained from IBM AppScan with manual security testing and verification. In all cases, false positives were removed from the results and the vulnerabilities were mapped to the OWASP Top 10 categories (Open Web Application Security Project):

1. Injection
2. Cross-Site Scripting (XSS)
3. Broken Authentication and Session Management
4. Insecure Direct Object References
5. Cross Site Request Forgery (CSRF)
6. Security Misconfiguration
7. Insecure Cryptographic Storage
8. Failure to Restrict URL Access
9. Insufficient Transport Layer Protection
10. Unvalidated Redirects and Forwards

For each of these categories, two core metrics were calculated:

1. The percent chance of finding at least one of these vulnerabilities in that category
2. The average number of vulnerabilities that are likely to be found in that category

Having gathered similar data since 2007, the team was also able to trend results over the past 5 years. This historical data was also mapped to the 2010 OWASP Top Ten to track this trending.

Metric points

The team also looked at additional metrics to assist in gaining deeper analysis of the data. This included:

Business Segments to attribute test data to one of the following:

- Financials
- Industrials
- Information Technology
- Logistics
- Government
- Other

Application Security Test Cycle depicting the type of test the application was involved in:

- **One Time Assessment**—Applications tested for the first time
- **Quarterly Assessment**—Applications tested on a regular ongoing basis
- **Retest**—Follow-up test to confirm the closing of findings typically from the one-time assessment

Note: Information was only categorized into these metric groups where the sample size allowed for suitable data. Where the sample size was deemed too small, the metric values were ignored. Thus, not all business segments or technologies are represented.

2011 application vulnerability trends

The following chart outlines the percentage chance of finding a vulnerability matching each of the OWASP Top 10 categories in an application security test.

The OWASP Top Ten mapping was chosen as it allows for a more focused assessment and comparisons to industry best practices. Where findings did not map directly to OWASP, the findings were captured against the security misconfiguration category and thus, the numbers in this category are naturally higher.

It is worth noting that these assessments are with organizations that appear determined to mitigate issues in their applications. They may already have security programs in place or may have had breaches in the past. As such, this data does not represent the state of web applications in general or applications that have never been examined. There is a notable downward trend in the values for some of the vulnerabilities and this likely highlights the return on their investment as much as anything else.

Broken Authentication and related issues with session control is found in nearly 8 tests out of 10. Many applications tested failed to restrict session tampering and were exposed to session fixation style attacks. Issues relating to session termination and session reuse also contributed to this high statistic.

Cross-Site Request Forgery (CSRF) in 2011 was found in 28 percent of tests undertaken, but this number was reduced from 2010 where the percentage was 59 percent. Some of this reduction appears to be in the greater awareness of this type of vulnerability and also improvements in methods used to include CRSF tokens.

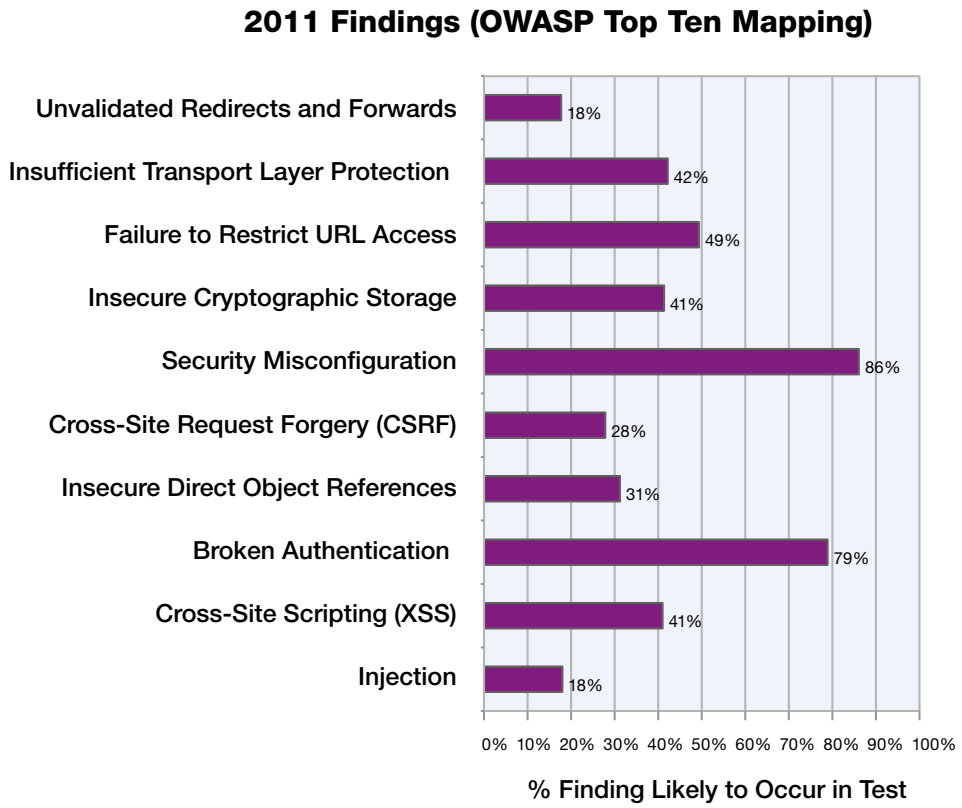


Figure 50: 2011 Findings (OWASP Top Ten Mapping)

Annual trends (2007—2011)

Since we started recording application security statistics in 2007, we have seen a steady decline in the instances of input control related vulnerabilities such as cross-site scripting (XSS) and SQL injection. In 2011, our statistics suggest that the likelihood of encountering XSS in a given test continues to decrease but shows signs of leveling out at approximately a 40 percent chance of occurring. Injection vulnerabilities and specifically SQL injection appears to have leveled out at around a 20 percent chance of occurring in a given test.

Although not clear from the statistics, our testing found that applications that use best practices and secure coding practices to filter invalid input had little to no instances of input-related issues such as XSS. The fact that XSS is still found in over 40 percent of applications tested highlights that there are still many applications that do not adhere to secure coding practices. There is no doubt that things are improving, but that is no reason to be complacent. The likelihood of 40 percent for XSS vulnerabilities is still high, especially for something that is so easily understood, so easily demonstrated, and so easily fixed. Web application vulnerabilities remain key to many data breaches, and data breaches continued to rise in the first half of 2011. So much so that X-Force declared 2011 to be the "Year of the Security Breach."

Another important data point that we capture is "the average number of a given finding per security test." What we are seeing is a reduction in instances of XSS when this vulnerability is found. In 2009 the average number was over 40 while in 2011 it is just over three. It is now much less likely to find an application with absolutely no input control in place. Most applications where XSS was found now appear to have some form of input control, but there were specialized attack vectors that were able to get around these filters/controls.

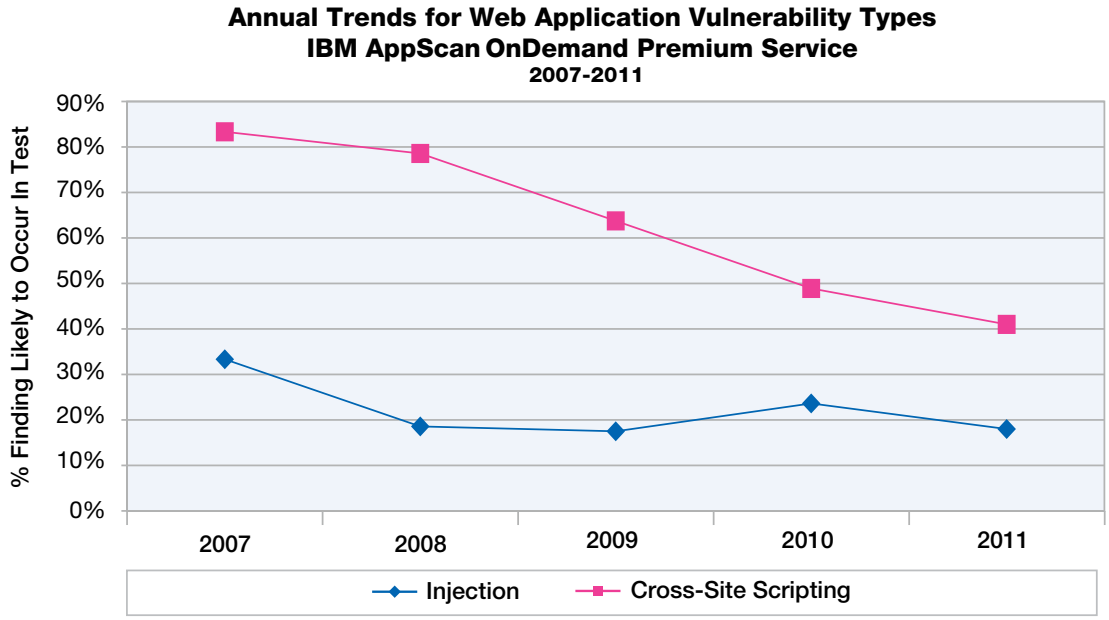


Figure 51: Annual Trends for Web Application Vulnerability Types
IBM AppScan OnDemand Premium Service – 2007-2011

ANNUAL TRENDS										
Vulnerability Type	**2007**		**2008**		**2009**		**2010**		**2011**	
	Avg. vulnerability per test	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**
Injection	1.3	33%	5.3	19%	1.7	18%	2.3	24%	2.1	18%
Cross-Site Scripting (XSS)	12.7	83%	17.9	79%	40.8	64%	5.8	49%	3.3	41%
Broken Authentication	11.2	83%	4.8	84%	3.2	65%	2.5	53%	9.7	79%
Insecure Direct Object References	2.6	50%	3.2	54%	3.0	51%	1.9	33%	1.6	31%
Cross-Site Request Forgery (CSRF)	1.9	22%	1.8	20%	7.9	59%	3.8	53%	2.0	28%
Security Misconfiguration	46.9	83%	22.6	74%	23.5	68%	15.3	56%	10.7	86%
Insecure Cryptographic Storage	21.7	38%	17.9	56%	29.1	38%	19.8	45%	11.9	41%
Failure to Restrict URL Access	7.2	13%	6.0	19%	9.7	13%	6.6	15%	5.0	49%
Insufficient Transport Layer	7.3	28%	2.4	17%	2.5	35%	1.6	22%	9.8	42%
Unvalidated Redirects and Forwards	1.7	7%	0.5	5%	0.1	3%	0.4	4%	0.3	18%

Table 8: Annual Trends for Web Application Vulnerability Types, 2007-2011, IBM Rational IBM AppScan OnDemand Premium Service

Business segments

As in 2010, we split out our 2011 statistics by business segments. We were able to split out data for five segments where the number of data points would allow.

In 2011, financial applications were again the best performing segment. The following chart shows how each of the five segments compared in relation to XSS, injection, and CSRF vulnerabilities. Government applications were the worst performers in all three of these categories. It is not clear why this is the case, but reputational damage could be a factor. Breaches in Government applications are less likely to drive an investment in security mitigation than they would for financial applications.

CSRF is significantly lower for financial applications than for any of the other sectors. It is likely that this form of attack is taken far more seriously in this sector because of the perceived consequences. The main object of this type of attack is to defraud the victim and it is probable that banking applications and applications that use financial transactions are the main targets.

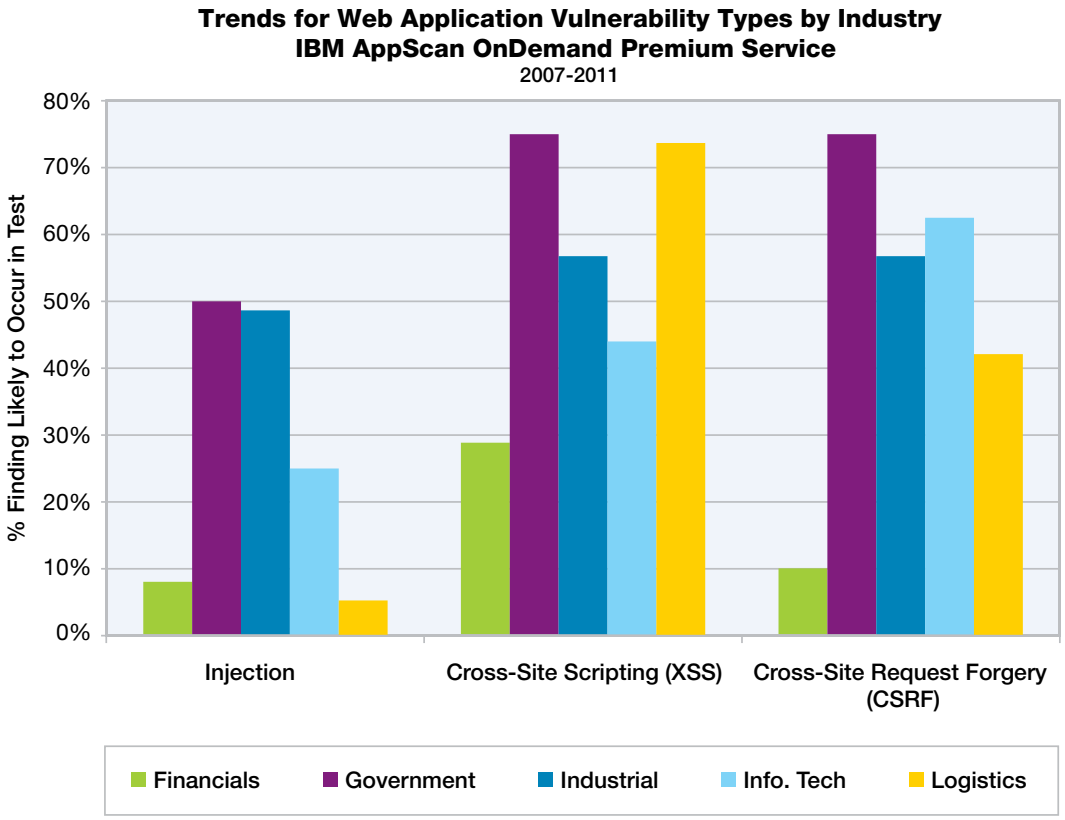


Figure 52: Trends for Web Application Vulnerability Types by Industry IBM AppScan OnDemand Premium Service – 2007-2011

BUSINESS SEGMENT										
Vulnerability Type	**Financial Services**		**Government**		**Industrial**		**Info. Tech**		**Logistics**	
	Avg. vuln per test	**% one vuln likely to occur**	**Avg. vuln per test**	**% one vuln likely to occur**	**Avg. vuln per test**	**% one vuln likely to occur**	**Avg. vuln per test**	**% one vuln likely to occur**	**Avg. vuln per test**	**% one vuln likely to occur**
Injection	0.1	8%	3.5	50%	10.9	49%	0.6	25%	0.3	5%
Cross-Site Scripting (XSS)	0.4	29%	5.8	75%	13.2	57%	6.1	44%	2.5	74%
Broken Authentication	5.1	73%	12.7	94%	4.8	84%	26.5	100%	38.9	84%
Insecure Direct Object References	0.3	18%	5.6	94%	2.1	35%	4.8	63%	4.5	47%
Cross-Site Request Forgery (CSRF)	1.1	10%	3.9	75%	3.0	57%	2.3	63%	5.7	42%
Security Misconfiguration	2.9	82%	18.9	100%	25.9	97%	39.7	100%	10.5	74%
Insecure Cryptographic Storage	4.8	22%	19.4	100%	12.3	51%	39.9	94%	37.1	79%
Failure to Restrict URL Access	1.0	44%	14.9	100%	0.9	19%	29.4	81%	15.2	79%
Insufficient Transport Layer	3.8	25%	1.4	75%	13.6	59%	36.3	88%	34.3	79%
Unvalidated Redirects and Forwards	0.2	14%	0.2	19%	1.1	46%	0.1	6%	0.0	0%

Table 9: Most Prevalent Web Application Vulnerabilities by Industry, IBM AppScan OnDemand Premium Service

Application security test cycle

In most cases, the IBM AppScan service where this data is collected offers a retest option for any tested application. Typically this retest occurs within 60 days of the initial test and it is not always possible to close all the issues in that timeframe.

It is certainly expected that the results returned from an application retest would be less than those from an application being tested for the first time. By looking at the average number of a given finding for a test, this difference is however highly significant. For each of the OWASP Top Ten categories, the difference is more than double.

The chart below highlights the difference in the average findings found per test between a one-time assessment and the later retest.

In general, our customers should retest results to validate that things are fixed. If the act of initial application testing was enough, then our quarterly results would yield similar results to these retests. This is clearly not the case. We believe that knowing that the application will be retested immediately must act as a motivator for the development team; otherwise the quarterly results would look at lot like the "retest" results. Another factor here is that customers who undertake regular quarterly testing may be motivated by compliance factors and not the pressing need to mitigate vulnerabilities. This would suggest that a best practice is to always retest to confirm that items are fixed. To achieve that cost-effectively, customers should consider using in-house tools and expertise.

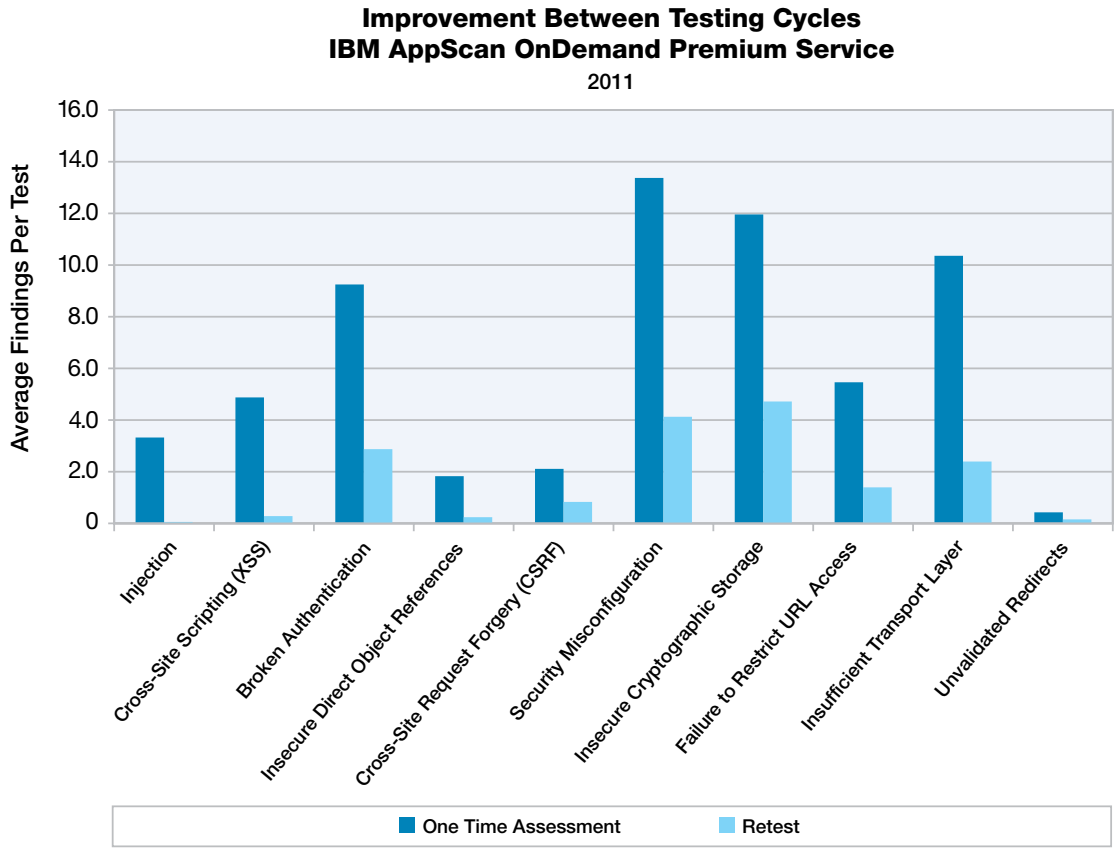


Figure 53: Improvement Between Testing Cycles
IBM AppScan OnDemand Premium Service – 2011

SECURITY TEST CYCLE						
Vulnerability Type	**One Time Assessment**		**Quarterly Assessment**		**Retest**	
	Avg. vulnerability per test	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**	**Avg. vulnerability per test**	**% one vulnerability likely to occur**
Injection	3.3	27%	0.2	5%	0.1	4%
Cross-Site Scripting (XSS)	4.9	46%	3.0	76%	0.3	21%
Broken Authentication and Session Management	9.2	82%	36.3	86%	2.9	70%
Insecure Direct Object References	1.8	37%	4.1	43%	0.2	15%
Cross-Site Request Forgery (CSRF)	2.1	34%	5.5	43%	0.8	10%
Security Misconfiguration	13.4	91%	14.1	76%	4.1	79%
Insecure Cryptographic Storage	12.0	50%	35.7	76%	4.7	14%
Failure to Restrict URL Access	5.5	51%	13.9	76%	1.4	38%
Insufficient Transport Layer Protection	10.4	46%	31.0	71%	2.4	27%
Unvalidated Redirects and Forwards	0.4	23%	0.0	0%	0.2	13%

Table 10: Security Test Cycles by Vulnerability Type, IBM AppScan OnDemand Premium Service 2011

Section IV
Emerging Trends in Security

The Emerging Trends in Security section looks at fast-developing technology that presses upon enterprises considering whether it is time to make investments in these future areas. We explain where threats and exploits are being used in these early technology adoptions and how enterprises can stay focused.

Mobile security and the enterprise—a year in review

Mobile enablement and related security was a primary focus item for nearly every enterprise. They are being challenged to adopt increasing levels of mobility as technology innovation has helped drive capabilities that allow increases in efficiency and allow nearly any business to increase its pace through the constantly-connected workplace that mobility is driving. Best practices in helping secure mobile devices is in its infancy, although progress is being made in this area.

Lack of clarity around best practices in helping secure mobile devices is also compounded for many enterprises that are embracing or at least facilitating Bring-Your-Own-Device (BYOD) programs that have previously never allowed or supported such models. Due to increased prevalence of employee ownership of such devices, both senior executives and employees alike are interested in making this program work, with often the enterprise's CISO being the primary hurdle in moving forward due to security concerns. While many CISOs have continued to say "No" rather than "How," indications are that this approach may result in multiple projects to detect and prevent employees from finding ways around existing infrastructure to enable themselves. Clearly this is not a favorable position for enterprises to be in and the approach being taken is to enable and control limited uses of mobile devices with a focus on the classification of data elements.

Sound analysis of existing security requirements associated with the data elements under discussion for enablement adds some clarity for enterprises struggling with what controls are required. This data-focused approach leverages existing security standards and will eventually result in best practices for mobile security. In many ways this is just a common sense approach to securing data on any computing device—and certainly we would all acknowledge that today's smartphones and tablets are just computing devices.

For some industries, this may mean BYOD approaches that result in some data elements being present on the personally owned mobile device may not be appropriate. It really is a matter of focusing on the data under consideration for enablement and then applying the associated, required controls.

Mobile malware visibility has certainly increased in the past year. It is important to look at this within the context of the overall threat landscape with which enterprises deal. There have been many mainstream IT press articles highlighting mobile-specific malware attacks that would lead one to believe that it surpassed the traditional Windows XP threat landscape. Of course, this couldn't be farther from the truth, but it does provide a good data point in that mobile malware is on the increase and a sound security program must plan for this challenge.

Malware related to mobile devices is not the only thing increasing over the past year. Certainly new mobility management solutions (commonly called MDM or Mobile Device Management solutions) seem to be popping up on a weekly basis. This is to be expected and just as a lot of technology innovation is focused on the mobile space, this should lead to increased need and opportunity for such solutions. Choice and competition is always good for the customer and increases the likelihood that all the security control requirements that enterprises need will be covered in a choice of solutions at a competitive price. Lately there have also been increases in the numbers of secure isolation or separation solutions. These are sometimes referred to as Data Leakage solutions, though in the mobile context, are much different than traditional DLP solutions that exist for workstations. While these solutions provide the eventual promise of being able to better address enterprise data and applications residing on employee-owned devices in a BYOD program, most are relatively limited and immature at present.

Mobile malware perspective

As we look at changes in the mobile malware threat landscape over the past year, it has gotten more visibility as an area of concern. In some ways, this has been beneficial by making IT executives aware of the real possibilities of what to expect and to allow enterprises to plan for proper controls. Readers of previous X-Force Trend and Risk Reports may notice that IBM has been watching and anticipating such increases.

It is worth mentioning the nature of the mobile malware threats that have been exposed over the past year. In nearly every case, they existed and were delivered to devices within what is considered legitimate application stores associated with the mobile platform. It is also worth noting that this has occurred across all major mobile platforms and stores and has not been confined to one uniquely. This is important for multiple reasons. As the selection of applications has exploded in nearly all application stores, the effectiveness of reviewing

submissions has not increased (with the exception of the Google App marketplace highlighted below) and we're starting to see the results of this. The other important aspect of this is that most device owners (and enterprise employees) would expect that confining application downloads to only legitimate application stores would protect them from malicious applications. This is a fallacy.

In fairness, curators of popular application stores certainly respond reactively at the presence of a malicious application and remove it but often, this is well after it has been downloaded by many users and is purely reactive. It is also best practice guidance to avoid most third-party applications stores that cannot revoke existing applications. It makes sense that the probability of encountering a malicious application increases as oversight decreases. There is not a valid model to feed awareness to the curator from security researchers because it does not provide a method to monetize their research. Unfortunately, the user and enterprise are the real losers here because of the assumption of trust implied by the app store model.

To assist enterprises with this problem, there is a growing selection of malware prevention approaches available from security vendors. While many enterprises initially overlooked this need, more and more have embraced the reality that mobile malware will continue to increase and provide the criminal enterprise that drives most malware an increased opportunity (and threat for the enterprise). These solutions are available for most platforms and platform coverage becomes easier and easier as the market determines which platforms will survive.

Without detection being present and required, some malicious applications could go undetected by the device's user. We should emphasize some because others exist solely to perform fraudulent transactions that should be detected by the user upon review of their monthly bill. As with personal computer malware, monetary attacks remain a primary focus and mobile devices that support SMS provide a very attractive target.

Another practical example we've observed that is somewhat unique due to the nature of mobile devices (because they usually have GPS hardware, along with voice, messaging, and data services) is the detected presence of spy applications that monitor multiple aspects of their users behavior—including recording location, messages, email, and voice calls to their attacker for review. This is particularly disconcerting when we compare it to the kinds of attacks we see on personal computers. Because mobile devices really have become "your office in your pocket," they provide a broad opportunity for a spy attack.

Recently, Google disclosed the implementation of an application review capability that begins the process of security oversight in what is accepted and maintained in their App Marketplace. This is particularly noteworthy because it's a proactive step toward improving the security of apps within their store and an example for other app store curators to follow. While it should be expected that this will not be perfect and will likely be a cat-and-mouse exercise between Google and those who seek to submit malicious applications, it is clearly a statement of action and the need to protect users from malicious applications. Time will tell if other application store owners/curators will follow suit.

With respect to mobile malware, an area of risk that should be of concern is mobile operating system currency. While we have not seen broadly pervasive malware attacks that self-replicated across a mobile operating system that occurred as a result of an underlying platform vulnerability, it likely is only be a matter of time before this occurs and certainly some platforms are in a better position to address this than others. From a purely enterprise perspective, nearly all MDM solutions available allow for the ability to control synchronization of enterprise information based on operating system version (hence allowing the enterprise to discontinue support of vulnerable versions of unpatched operating systems). This will likely lead to the frustration of enterprise employees who can be caught in the middle of this support issue with their carrier (particularly in BYOD programs where models and carriers can often not be completely managed as in corporate-provided programs with contractual controls in this area). We could see a time in the not too distant future where employees and device owners are stuck with vulnerable devices, unsupported within their enterprise, and their only option will be device upgrade before the completion of their current

contract where subsidized contract models are popular. Many suspect that hardware OEMs have intentionally left devices behind in order to drive owners to upgrade more frequently. This particular problem may prove a challenge in terms of consumer acceptance at some point because it is quite different than the accepted model for other consumer computing devices such as laptops.

BYOD and secure isolation

As highlighted earlier, one of the more recent developments this year has been the increased interest in providing the ability to separate enterprise applications and data from the employee's personal applications and data. Obviously, the primary driver for this development has specifically been the pervasive nature and interest in BYOD programs. While some solutions existed prior to this year, selection was sparse and most solutions were limited in function and usability. Over the past year, solutions in this space are popping up like flowers in the spring. We should expect that most are largely works-in-progress and perhaps each will have their own limitations, usability quirks, and hurdles to implement but they are clearly a sign that the recognition of this problem and enterprise need is being heard and recognized across the industry. This is significant progress from last year when these solutions were a niche used by specific industries because related enterprises held very specific, regulated data that could find its way to employee's mobile devices.

As this market segment matures and improves, we should expect to see solutions fall into a couple of different categories. There are activities and collective work underway in the Android space toward approaches that use hardware-based virtualization approaches. Progress in this approach is limited to extensive adoption of this chip-level capability and then corresponding adoption and support across large numbers of different devices running on an array of carriers globally for it to be an effective approach for large multi-national enterprises. While this may take 24 to 36 months to happen that broadly, the "movement" is clearly in progress as chip makers, hardware OEMs, and carriers recognize this enterprise requirement and corresponding market opportunity. We'll see whether this approach becomes pervasive enough to work in BYOD approaches within large enterprises.

In the meantime, a number of solutions that allow separation—whether via a container, a virtual container, or asset management approaches—are filling a gap for early adopters. These approaches provide a level of separation and additional control that enterprises can express on employee devices without continuing to control the whole device but active work is still needed to determine what controls are required at a device level in order to trust it as a host for the separation solution. While this same concern applies in the virtualization approaches mentioned above, an application container or separation approach within the same instance of the mobile operating system is at more risk to the existence of malware or malicious applications. This is another case where best practice is yet to be defined but as enterprises adopt such solutions and needed security technical testing is conducted, accepted practices will likely emerge.

Importance of device management convergence in role-based enterprises

As mobile device usage continues to explode within the enterprise—whether purely corporate-owned, employee-owned, or a mixture of both—the need to manage them within the context of enterprise risk management will increase in importance. This will be particularly true as adoption rivals use of other computing devices like laptops. It will likely not be uncommon for the user-to-device ratio to become two or three devices per employee among laptops, tablets, and smartphones. This will mean that the distribution of enterprise data likely will continue to increase and challenge the use of role-based security profiles and enterprise risk management.

As enterprises seek to address role-based user security profiles that are tailored to the role and types of data specific user roles are associated with, this approach will become increasingly difficult as device management

is spread across multiple device management solutions. In fact, as enterprises migrate away from one-size-fits-all programs that exist under purely corporate provided computing programs to reliance on managing a wider array of operating platforms common in BYOD programs, this ability to drive device management into a single platform will likely become the primary factor that enables BYOD within reasonable cost. For smaller homogeneous enterprises, this may be avoided because of the absence of significant numbers of different roles or use of the same data classifications across the majority of their population. They may be able to live with a couple of solutions (perhaps one for standard computing assets and one for mobile assets like smartphones and tablets) but for larger enterprises, this is likely to be a severe limitation that ends up being an unsatisfactory compromise. Imagine a large enterprise having to secure all assets, mobile or otherwise, to meet the highest level of security needed for particularly sensitive contracts, customers, or projects because they cannot effectively implement multiple roles across the different kinds of devices used by their employees. In the end, the loss of efficiency and inability to support the best device form factor for differing roles really drives the necessity to seek a uniform platform to manage all endpoint devices.

The second and equally important reason to converge the management of all endpoint devices is the desire for collective visibility and enterprise risk management. While it is certainly possible to try to tie together disparate management systems into a single enterprise risk console, it is far easier and more likely to succeed if this can be supported by a single framework technology. It's also far more likely to be able to integrate this single platform into advanced persistent threat (APT) analysis and response.

Fundamentally, for most enterprises concerned with advanced persistent threats, tying together operational analysis and analytics so that it includes endpoint status, information, and the ability to interact with endpoint systems in real time, becomes fundamental in the ability to provide a closed detection/response ecosystem. Managing all endpoints consistently and programmatically with well-defined and controlled security policies should be easily done with the selection of the correct security management technologies, along with providing efficiency and oversight to improve the whole enterprise security landscape by focusing across the endpoint population.

A retrospective look at the state of security in the cloud

Much has been spoken about the state of security in cloud environments and organizations have been left looking for answers when trying to understand how to adopt cloud solutions and ensure their security. As more and more organizations look to embrace the cloud, security remains the top priority. Many organizations remain hesitant about moving business-critical applications to public clouds and have in many cases chosen to leverage private clouds. This thinking is similar to when the Internet was in its infancy when many organizations were hesitant to move business-critical applications to this "new" network and instead relied on private networks (often based on leased lines). Just as the economies of scale eventually pushed some of the most critical business applications onto the Internet, the same transformation is occurring in cloud computing. The question is not whether the cloud is more or less secure, but what specific controls and business processes should be used to help reduce risk and help ensure security in a cloud environment. It is important for any organization looking to more widely adopt cloud-based infrastructures that they have an understanding of the role of the organization versus the role of the cloud service provider when it relates to security and risk mitigation.

As with any business-critical application or service, the business organization should ensure alignment between the risks specific to the organization and the policies and procedures provided by the service provider. Security best practices should be adhered to when adopting any new Internet technology and cloud computing is no different. When considering any cloud deployment it is beneficial to think about security across all phases of a cloud deployment.

Design

Security by Design
Focus on building security into the fabric of the cloud.

Deploy

Workload Driven
Secure cloud resources based on the security requirements of each workload

Consume

Service Enables
Govern the cloud through ongoing security operations and workflows

Adopting security for the cloud

One question that many organizations have is whether cloud-based applications and services are more secure than traditional Internet and intranet applications. Although no single deployment scenario offers more inherent security, one common observation is that security is a greater focus area when considering cloud-based deployments. Often, more trust is placed in an organization's deployment of applications and services when those services are considered to be inside the perimeter of the organization's trust boundary. It is obvious that just a conversation around security does not in and of itself create more security, but because security is front and center when considering cloud deployments, it is much more common to see strict controls, processes, and procedures around security in many cloud application and service engagements.

Design considerations

Corporate security development practices should be in place and adhered to. When considering third-party cloud application providers, it is important to make sure that their secure development standards and practices meet or exceed your own.

Appropriate network and endpoint security safeguards should be in place. In a multi-tenant environment it is important to make sure that sensitive and critical applications are not sharing the same hypervisor without appropriate security zones and data segregation processes in place.

Understand the data security requirements. Many applications that leverage sensitive and private information have strict security requirements mandated by organizations, governments, and applicable standards and regulations. You must ensure that the cloud service provider can adequately address these.

Deployment considerations

Manage virtual endpoints the same way that you manage non-virtual endpoints. It is important that virtual libraries and catalogs do not suffer from "security drift" when it comes to patch and configuration management.

Enforce security controls consistently across both cloud and non-cloud environments. Make sure that applications deployed to virtual environments receive the same security scrutiny as public internet applications—especially development and test environments that often lack basic security controls.

Scan all cloud applications on a regular basis. Leverage source code and dynamic application services to limit security exposure of any application deployed to the cloud.

Consume considerations

Appropriate identity and access management. Enforce identity and access rights accordingly, taking into consideration federation of identity when it comes to third-party cloud SaaS services.

Log and security event management. Have effective log and security event management of virtual devices.

Data Forensics. If considering a third party, understand how data forensics is managed in the event of a security incident.

Following a secure-by-design approach is the best way to help achieve greater security and help reduce the risks in moving to a cloud-based infrastructure. Moving to the cloud has re-engaged many IT organizations and, because of the lack of control, greater emphasis is being brought to the forefront when it comes to security. In many cases, this heightened attention paid to addressing the challenges involved in securing an environment you don't control 100 percent, results in greater security being achieved. Even though the details of the infrastructure are less transparent and, quite honestly, cloudy, greater security can be the result.

Improving cloud security through SLAs

Introduction

2011 was a big year for data breaches in the cloud. Many large and high-profile organizations were exploited, and millions of consumer records put at risk. The breach of a single large-scale cloud entity in 1Q11 started a chain reaction, affecting retailers' and financial institutions' customer databases, whose consumer financial records were subsequently exposed. 2011 was dubbed by IBM X-Force as the Year of the Security Breach, and led many organizations to question whether cloud computing could reasonably be secured.

Success in secure cloud computing is more than a question of simple contract management, but it can be critical to the success of the cloud deployment. Standard contracts and Terms of Service (TOS) are typically written for the benefit of the cloud provider, in order to define basic services and limit exposure and liability. It is extremely unusual for the cloud vendor to alter its standard contract in order to accommodate the needs of client organizations. The Service Level Agreement (SLA) is the more flexible document that allows the client organization to define requirements unique to its business model, legal and regulatory requirements, or other considerations. Unfortunately it is the nature of cloud computing—its flexibility, scalability, and rapid deployment capability—that can make it very difficult to structure and maintain a meaningful SLA.

Issues to consider

Because of the limited impact that the organization can realistically exercise over the cloud computing environment, the most effective means for managing information security may be through the SLA. Therefore, it is important for the organization to be proactive in its approach, and take as long term a view as reasonably possible over each of its cloud computing projects. Far too many early adopters took a short-term view, concerning themselves primarily with vendor selection and service launch, not taking lifecycle management and exit strategy into account.

Resiliency is at the core of most cloud SLAs, and what cloud vendors often make the focus of their standard statements of services. Resiliency includes guarantees of uptime, performance and response times, error correction time, and so on. Some may go so far as to include issues such as segmentation and isolation in multi-tenancy situations or change management policies and procedures. More often than not, standard SLAs include only general representations regarding information security. The organization must look carefully at the policies, procedures, and control measures that are offered as standard service, and then create custom requirements for each specific workload, as driven by the data that each will process, transmit, or store.

For effective information security management over the long term, the organization should consider the following when crafting SLAs:

- **Ownership.** The organization should scan the cloud provider's standard contracts, TOS, SLAs, and others for any provision related to the joint or outright ownership of applications, functionality, data sets, or related work product resulting from the cloud engagement, prior to putting any sensitive or critical data, processes, or intellectual property in the hands of the cloud provider. The organization should ensure in writing that it retains ownership of data or assets that it exposes to the cloud provider, in order to facilitate the transfer of that asset to another service provider or to bring that asset back in house at will. This is particularly important for organizations using cloud-based Anything-as-a-Service (XaaS). Software and processes proprietary to one cloud vendor may not be easily replicated should the organization need to bring the project in house or to transfer the project to another provider. Finding out after the fact that the organization has given up partial or full rights to its assets as a condition of service can further complicate a difficult situation.
- **Access management.** Just as the organization sets limits on authorized users of sensitive or critical data in house, it should oversee the access management policies and mechanisms in place in a cloud environment. Specific requirements for access management by the cloud provider's staff to the organization's data should be driven by the unique demands of the workload. But, in general, the organization should have a good understanding of how the principle of least privilege is applied in the cloud provider's live production environment(s). This is absolutely critical in a multi-tenancy public cloud environment. Just as each tenant's deployment should be isolated within the shared hosting environment, access should be restricted (to the extent reasonably possible) to a set of technical employees designated to provide services to the client organization. This is dependent on the cloud provider's business model, but the organization must understand precisely how the cloud provider manages physical, logical, remote, and emergency access to the tenant environment and data. The organization should assess the legal and regulatory requirements for the data in the workload, and make certain that the cloud provider understands and can meet those requirements, and provide demonstrable evidence of a good faith effort so to do. We discuss Access Management in the cloud a bit more in the following section.
- **Governance.** How the cloud provider makes representations regarding its information security posture and capabilities should be a key factor for the organization when determining the type of cloud and provider appropriate to the workload. The organization should examine any documents that the cloud provider makes public regarding its information security capabilities, including redacted audit reports or summaries (such as an SSAE 16 SOC 2 report or SOC 3 seal), certifications (such as an ISO 27001 registration for the production environment), or other documents of conformity with compliance standards such as the BITS Shared Assessments AUP or COBIT. The organization should state its need to have access to such documentation from the provider, in order to satisfy any legal and regulatory requirements. The organization should negotiate in its SLA with the cloud provider:
  - Verification of security training and awareness for technical staff.
  - Access to logging and monitoring information directly related to tenant environments.
  - Documented security responsibility and liability in the event of a data breach. This is particularly critical when compound SLAs exist.

  - Access to forensic information related to data breaches for purposes of consumer notification and investigation by law enforcement.
  - Documentation outlining how the cloud provider will respond to law enforcement requests for information, investigation, subpoenas, and so on.

- **Termination.** Most cloud providers' standard contracts and TOS statements contain provisions related to termination for cause on the part of the provider (such as non-payment), or on the part of the customer (such as failure to meet uptime guarantees). In addition to this, the organization should inspect these standard documents for any other conditions of breach of contract, and should clearly define an exit strategy in anticipation of material changes to the services offered by or abilities of the provider, changes to its own business model, or simply due to failure of the cloud project. The organization should retain the reasonable right to terminate its contract with the provider without the imposition of unreasonable penalties, including:
  - Changes in business model of the cloud provider, such as the introduction of compound SLAs after the commencement of the engagement without sufficient notice or the opportunity for due diligence by the organization.
  - Changes in ownership of the cloud provider, such as merger or acquisition.
  - Substantial changes in fees without sufficient notice.
  - Cancellation or significant changes to services without sufficient notice.

Ideally, the organization should plan for termination of its cloud service with sufficient time to implement a transition plan. This assumes, of course, that the organization has a written transition plan in place. Reasons for this can vary depending upon workload, type of cloud, and provider performance, but cloud deployments can fail for many reasons—the anticipated costs savings never realistically materialized, the project was too difficult to manage in an outsourced situation, the product or service itself failed, and so on. Whatever the organization's reason, it should have an exit strategy that plans for the need to move the project to another outsourced provider or return the functions in house. A documented transition plan should include:

- Reasons for termination documented for the benefit of the cloud provider.
- Sufficient transition timing in cases where the function or service was not originally designed to be turnkey.
- Transition assistance, including data format and transfer from the cloud provider back to the organization.
- Return of all data and assets belonging to the organization, including backups.
- Secure disposal and/or destruction of residual data in the cloud environment, including backups.
- Contingency for complications created by data encryption.

Clearly, this is not a comprehensive set of issues that the organization should consider. The specific requirements of the workload allow the organization to choose the most appropriate type of cloud (public, private, hybrid, or managed) and the most suitable vendor to provide cloud services. These considerations vary depending on the type of cloud model in which the organization deploys. For example, isolation in a multi-tenancy environment is not applicable when the organization deploys to a private managed cloud. In general, however, these are some often overlooked yet critical issues that the organization should plan for and document to successfully manage its cloud deployments.

Conclusion

Cloud computing is moving rapidly from emerging to mainstream technology, and rapid growth is anticipated through year-end 2013. There are valuable lessons to be learned from early adopters of cloud technology, particularly where information security is concerned. Taking a long-term view of any proposed cloud computing project, and carefully reviewing the service and security requirements as dictated by workload, allows the organization to select an appropriate cloud model and provider.

Negotiating strong and favorable SLAs can be critical to the success of the mission and have benefits for all parties involved. This exercise requires careful planning, and avoiding take-it-or-leave-it agreements with standard, non-negotiable terms. If the cloud provider is not willing to negotiate the SLA, then they may not be the right provider for the deployment. SLAs should be specific in both terms and scope, changeable only with appropriate notice, and cognizant of the specific business and information security requirements of the organization. SLAs may seem to be a passive tool, but they may be the most effective means to manage and maintain an effective security posture in an outsourced environment.

Identity and access management in the cloud

Security challenges in cloud environments

With its flexibility, cost efficiencies, and scalable "on demand" model, cloud computing has become ever more popular. The ability to share services and information with various departments, partners, and customers is a major advantage of cloud computing. As an added benefit, cloud computing can enhance the user experience without adding to its complexity. Users do not need to know anything about the underlying technology or implementations.

Although the benefits of cloud computing are clear, so is the need to develop proper security in cloud implementations. As more and more organizations embrace or consider cloud computing, they also worry about the associated security risks. **A Global Risk Survey conducted by IBM's Institute for Business Value** found that cloud computing raised serious concerns about the access to and use and control of data: 77 percent of respondents believe that adopting cloud computing makes protecting privacy more difficult; 50 percent are concerned about a data breach or loss; and 23 percent worry about a weakening of corporate network security.

Data and applications are often hosted on public domains, so access management becomes a concern. Will cloud computing be as secure as the data center? What happens when business units begin using public cloud services in combination with the data center or a private cloud? How can you be sure only authorized people are accessing your sensitive data and applications? Is your cloud provider able to provide audit reports to demonstrate your compliance with industry and government regulations? Addressing the issues raised by these questions is critical to successful cloud security.

Organizations should balance protection, privacy, governance, and accessibility to key resources—whether in the traditional data center, the private cloud, or the public cloud. Cloud computing requires a delicate balance between the requirement to share resources and the need to protect those resources from unauthorized access, data leakage, and other exposures. It's obvious that you won't want inappropriate individuals to have access to your organization's private data and applications. To help ensure that your company's IT resources are safe wherever they're located and whenever they're needed, identity and access management must be built into the fabric of your cloud.

The need for security in the cloud should not be overlooked or "bolted on" later during the transition, but built into the overall cloud implementation plans. These plans may need to include updates to business processes and policies, as cloud security requires more than just technology. Just as in traditional security environments, organizations should agree to document and execute security mandates for the cloud environment to meet their business and regulatory objectives. These mandates may include service level agreements with the cloud provider, separation of duty requirements for various cloud user groups, and the creation of "trust zones" to isolate your data from other cloud customers that share the same physical hardware.

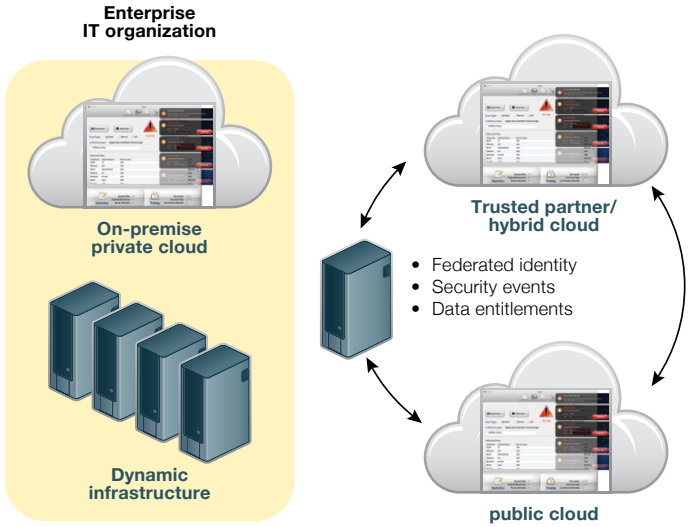
Identity and Access Management (IAM) solutions for the cloud

Whichever applications or information you decide to move to the cloud, a solid identity and access management (IAM) solution can lead the way. It can encompass both cloud and traditional computing environments so you do not have to manage two sets of credentials. The primary goal is helping ensure authorized users have access to the applications, data, and tools they need, when they need them, while blocking unauthorized access. With their ability to limit access to only authorized and appropriate users, IAM solutions are an invaluable component of any cloud security plan.

With an IAM solution, you can set and enforce policies for who can access what information, when, from what locations, and how much they can access in a set time period. You can use the solution to reconfirm entitlements over time and promptly revoke them as necessary. Tools should also be available to monitor, report, and proactively prevent policy violations.

As in traditional IT environments, an IAM solution for the cloud should incorporate the following capabilities: user provisioning (including separation of duty, roles-based access controls, and fine-grained entitlements), password management, web and federated single sign-on, logging, and audit reporting capabilities. Finally, privileged identity management is especially critical because of the catastrophic damage that insiders can cause, intentionally or inadvertently.

Securing access to cloud-based applications and services



With Identity and Access Management (IAM) solutions, the organization can centrally control access for large numbers of users to its cloud-based services hosted by external providers such as salesforce.com.

The cloud extends services, applications, and resources to a large and diverse community of users that may include employees, customers, and business partners coming from trusted and untrusted external locations. Organizations should tie cloud-based applications together with internal applications and enable users to access them easily with single sign-on. Identity federation and capabilities for rapid onboarding must be available to coordinate authentication and authorization with the enterprise's back-end or third-party systems. Federated identity management provides an approach to managing identities and access in a cloud, and within traditional computing infrastructures. It also can simplify the provisioning in the self-service environment of the cloud. A standards-based, single sign-on capability simplifies end-user logins for both internally hosted applications and the cloud, allowing end users to easily and quickly leverage cloud services.

In a typical scenario, authentication of the user takes place outside the cloud. The user's identity is then federated into the cloud. The entire process is transparent to the user. Single sign-on capabilities enable the user to go directly to cloud-based applications and information without having to manage identities within the cloud.

When it comes to compliance, organizations should have enterprise-wide capabilities to help ensure that both internal and external access are governed by effective authentication, to monitor authorization and network traffic, and to support the system with comprehensive audit and reporting capabilities. Regardless of the type of user, the solution should enhance security by helping to fill gaps in security measures. It should mitigate the risk of threats such as fraud, theft of intellectual property, or loss of customer data. It should help reduce costs by streamlining business and IT processes that grant users access to resources.

In summary, Identity and access management offers tangible, operational benefits of improved user productivity while reducing the risk of security breaches. An automated identity and access management (IAM) solution can address cloud security challenges and encompass both cloud and traditional computing environments.

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

Produced in the United States of America
March 2012

Please Recycle

WGL03012-USEN-00