Web site security issues
White paper
November 2009

Rational. software

IBM

Maintaining trust: protecting your Web site users from malware.

Contents

Is your Web site attacking your users?

The proliferation of malware designed to infiltrate computer systems without the owners' informed consent has become one of the most challenging security issues facing users today. Hackers are engineering ever more sophisticated viruses, worms and Trojan horses that can increasingly outsmart traditional defense mechanisms. Instances of Web-based malware alone increased by 508 percent in the first half of 2009 compared to the first half of 2008.[1] And the vast majority of it is served or linked off of legitimate, trusted Web sites—maybe even your own.

The increasing popularity of interactive Web sites—such as social networks, blogs and wikis designed to facilitate the sharing of user-created content—has made it easier for cybercriminals to distribute malicious software through the very same sharing channels, planting the malware on a visitor's computer without his or her knowledge. To date, Web site owners have fallen short in defending against this problem, leaving it up to users to defend themselves. In most cases, site owners don't even discover their sites are serving malware until long after they have been compromised. And that creates an array of risks for businesses, which can potentially include:

- *Loss of customer trust and loyalty.*
- *Steep fines for noncompliance with regulations.*
- *Legal liabilities from customer lawsuits.*

Given potential costs to the business as well as ethical considerations, Web site owners have a responsibility and a business interest in ensuring that their Web sites are not serving malware. This paper explores the problem of malware and how it is increasingly being delivered through legitimate Web sites. It also introduces new techniques from IBM that are designed to go beyond standard security measures to help organizations proactively defend against threats by scanning their Web sites for instances of embedded malware.

Highlights

Familiar culprit, new MO

By now most IT professionals are all too aware of malware and its various forms, including worms, Trojan horses, spyware and viruses. *Malware* is short for malicious software, and it is software that infiltrates a user's system and performs actions without the user's informed consent.

Today, malware is a favorite tool of organized criminals seeking to make money.

Malware was initially the territory of anarchists and "script kiddies" seeking to disrupt business or to win bragging rights. Today it's a favorite tool of organized criminals seeking to make money through the sale of stolen information or through spamming operations, for example. And it's not propagated just by amateurs. Criminal organizations have the means to hire highly skilled developers and software engineers to produce sophisticated malware. It can even be easily purchased like many pieces of legitimate commercial software and used without deep knowledge. Even governments of warring nations are in the game, developing malicious software for espionage and intelligence gathering. In fact, it is now widely reported that cyber warfare is a component of almost all major conflicts.

The latest front in the push to infect: the Web

Malicious software can be distributed in a variety of ways—and attackers generally do not limit themselves to a single channel. For a long time, e-mail was the primary delivery mechanism and is still significant. Network vulnerabilities and instant messaging have also been useful for pushing worms from one machine to another.

Web applications are now the primary delivery mechanism for malware.

Today, Web applications are the primary delivery mechanisms for malware via "drive-by downloads" or "social engineering." A drive-by download happens when a user's machine becomes compromised simply by browsing an infected Web page. The browser executes components that are maliciously crafted to exploit vulnerabilities in the browser, operating system or other plug-ins as the page renders images, in-line scripts and videos, for example.

Highlights

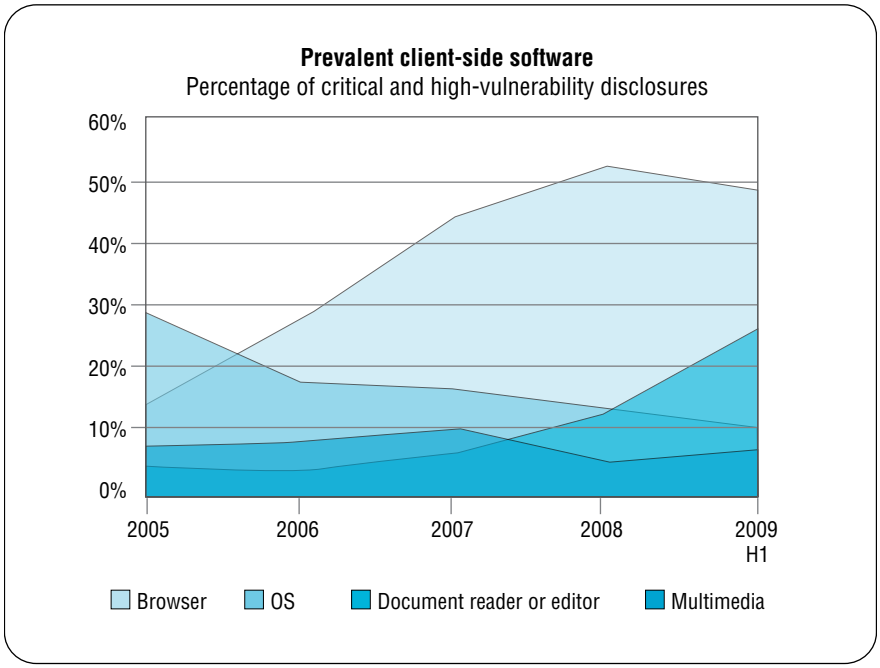

Figure 1: Critical and high-vulnerability disclosures affecting client-side applications by application category, 2005 through H1 2009.[1]

Scareware designed to trick users into downloading malware is showing up on trusted Web sites.

Social engineering is a term used to describe tricking a user into performing some action, such as downloading a file or accepting a prompt. "Scareware," such as an alarming pop-up that prompts users to perform an action, is an apt example. A pop-up designed to look like an antivirus alert may read "A virus has been detected on your system" and prompt a user to download a cleanup utility, which is actually malware (often a Trojan horse). In the fall of 2009, a major national newspaper in the United States faced a version of this tactic in the form of a scam that was designed to scare users into buying useless antivirus software.

Highlights

The most common method for delivering malware through a trusted Web page is through iframes pointing to external pages.

Malware can be delivered from many different components of a Web page. Because today's Web pages have images, have iframes pointing to external pages, and contain links to other pages or to other domains, they provide many potential vehicles for attacks. The most common method is using a malicious iframe on a page leading to malicious content. Through this approach, many compromised sites can share one malware-delivering host. Unfortunately, the end result to the user is the same whether a page links to malicious content or hosts it. And there isn't anything users can do to prevent this type of attack, so site owners need to protect them.

Serious implications

Owners of trusted Web sites rarely know if their sites have been compromised.

In recent years, occurrences of legitimate Web sites serving malware have become more widespread. Previously, cautious Web surfers who avoided questionable sites, such as adult-oriented or illegal download sites, could reasonably expect to avoid attacks. Not so today. Moreover, site owners rarely even know that the compromise has occurred. Consider the consequences. Users are no longer able to avoid exposure through good judgment alone. The malware is delivered through the sites they use and trust on a regular basis—for personal and business needs. Web gateways can no longer rely on blacklists of malicious sites without blocking legitimate sites as well. So how are users expected to protect themselves, and how can Web site owners avoid putting their users in harm's way? That question can't be addressed without understanding how legitimate sites are compromised.

Highlights

Trusted Web sites are typically attacked using vulnerability exploitations, malware uploads on user-driven sites, internal attacks or third-party content.

A look at how legitimate Web sites are compromised

In most cases, reputable Web sites are attacked using one or a combination of four primary methods.

Vulnerability exploitation

Vulnerabilities on a site are a favorite target of criminals. These could be 0-day vulnerabilities in the software running the Web site or vulnerabilities in the application-specific code. Such vulnerabilities can allow attackers to deface the site, making it link or serve malicious content. Exploiting 0-day or very recent vulnerabilities in Web infrastructure (for example, Web servers, application servers and operating systems) is the primary method of compromising Web sites today.

Uploaded malware on user-driven sites

Any user-driven Web 2.0 community sites—including blogs, wikis and social media sites—that let users create and post data likely provides another popular malware delivery source. And technical vulnerabilities aren't even necessary. If users are allowed to add content and links to the site, they may be uploading malicious items. For example, PDF document files holding malicious content or images that exploit a security hole in a graphics library can be the cause of a legitimate Web site to serve malware.

Internal attacks

Web site defenses are often not as robust when accessed from within an internal network. As a result, internal resources, such as disgruntled employees or an employee who has been blackmailed, can modify a Web application from within and make it serve or link to malware.

Highlights

Malware is commonly delivered through unvetted third-party banners and ads via Flash applications, for example.

Web site owners have significant responsibilities for protecting users from malware.

Third-party content

Including third-party content such as ads or mashup applications can multiply the risk of malware on your Web site. Third-party sources may be malicious or may have been compromised by yet another party, resulting in malware being served through your application's pages. Consider an advertising service serving Flash technology–based advertising banners. Flash applications are powerful and dynamic and have powerful scripting engines. If an advertising company is not properly vetting and analyzing each banner it posts, it may be serving malicious banners that deliver malware.

Where existing approaches fall short

How can users be expected to protect themselves from malware on legitimate Web sites? Certainly users need to take precautions by installing appropriate endpoint security solutions, such as antivirus software, firewalls and other security tools. But this will only get them so far. As a result, Web site owners have significant responsibilities in the matter. Just like organizations have a responsibility to ensure that applications are properly protected to keep users' personal data confidential, users should be able to expect a reasonable level of protection against being unknowingly served malicious code.

Currently, companies and organizations generally employ two types of approaches to protect the server side: an intrusion prevention system (IPS) or similar network protection device that monitors outgoing traffic, and server-side antivirus solutions. An IPS can examine all traffic returned from the site and block anything deemed malicious. The problem with this approach is the depth of the analysis—the IPS needs to work at a very high velocity to support huge volumes of data and thus can only afford a fraction of a second to analyze passing content.

Highlights

As a result, its analysis is mostly limited to matching known malicious patterns against the content; it cannot afford to really examine a passing piece of functionality let alone examine content in links on a returned page. This makes it easy for criminals to evade the IPS with simple obfuscation/modification of the malicious payload or by linking to the malicious Web parts on another site.

IPS devices and antivirus solutions can provide only limited protection of Web pages from malware.

A server-side antivirus solution can be used to examine files on the server and identify whether they are malicious. The problem with this approach is one of visibility. Antivirus solutions are designed to look for viruses in files but are limited in their ability to examine content residing in the databases where most applications store their dynamic content. Similarly, antivirus solutions don't see or understand Web pages, making them blind to content that is linked from the Web site but not hosted on them. Although an antivirus solution can identify malware reasonably well, it is of no use if it doesn't have access to the files that need analyzing.

What's the solution? A new approach to malware scanning

IBM has developed an alternative malware scanning approach that can more deeply analyze Web sites to find hidden sources of malware.

As mentioned above, the most common way for criminals to make legitimate Web sites serve malware is by injecting an iframe that leads to a malicious site. Unfortunately, existing solutions cannot find this very common manifestation of the problem. That's why IBM has developed an alternative approach: HTTP-based malware scanning. IBM Rational® AppScan® software uses this new approach, combining the HTTP view with antivirus-like capabilities. The solution works in three phases: discovery, content analysis and link analysis.

Highlights

By reviewing the application over HTTP, Rational AppScan software can see the application from a user's—or a potential victim's—point of view, including files and complete pages, as well as the links on them, regardless of how they were composed.

The VPS models the behavior of the discovered files, identifying and flagging all malicious or suspicious behaviors.

Discovery

In the discovery phase, Rational AppScan software completes a full discovery of virtually all Web content and links on the Web site. The software's proven deep-scanning technology reviews the site's pages, properly handling asynchronous JavaScript and XML and other rich content by executing JavaScript and Flash applications, reaching deeper into the site by filling in forms, analyzing private sections through extensive support for login mechanisms and server-side state, for example. By reviewing the application over HTTP, Rational AppScan software can see the application from a user's—or a potential victim's—point of view, seeing complete pages and files as well as the links on them, regardless of how they were composed.

Content analysis

In the content analysis phase, all discovered content is passed through a virus prevention system (VPS) engine developed by the IBM Internet Security Systems (ISS) X-Force® team.* The VPS is an advanced behavioral analysis engine, which is conceptually similar to an antivirus solution. The VPS models the behavior of the discovered files, identifying and flagging all malicious or suspicious behaviors, such as attempting to overwrite system files or modify network settings. Files that are deemed malicious or suspicious are properly flagged as security issues to the user. Thanks to the high visibility the HTTP scan offers, the scanner can optionally download and analyze the content *linked* from the site to uncover malicious iframes and compromised third-party elements. The content analysis phase is conceptually similar to running an antivirus scan on your site, but instead of analyzing files, it analyzes everything served or linked from the site.

**Applies only to AppScan Standard Edition and AppScan Enterprise Edition software.*

Highlights

As a last step, Rational AppScan software matches all the external links on a Web site against a database with a known blacklist of links.

To perform the link analysis, Rational AppScan software uses a Web filter repository that is a constantly updated database of link categorizations, used by many Web gateways to help protect users.

Link analysis

In the last phase, link analysis, Rational AppScan software matches all the external links seen on the Web site against a database of known link categorizations—effectively a blacklist of links. Any link that appears on the site that is known to be malicious, such as malware, phishing or spam sites, is flagged as a security issue. In addition, suspicious links—such as links to illegal domains or short-lived domains—are flagged with their own issue types.

To perform the link analysis, Rational AppScan software uses the IBM Proventia® Web filter database. This repository is a constantly updated database of link categorizations used by many Web gateways to help protect and improve the productivity of corporate users. Every day, millions of Internet pages are scanned by the X-Force servers (similar to how search engines crawl the Web) and sorted into dozens of different categories, such as news, sports, malware and phishing. The categorization is done using a variety of algorithms, created by IBM and other organizations, and is manually refined through user feedback. When a malicious server has been discovered and published, this repository is quickly updated with that information, allowing Rational AppScan software to flag links to such known malicious locations.

Highlights

Rational AppScan HTTP-based malware scanning and detection capabilities can help you overcome the inherent problems of existing security technologies.

Why IBM?

Rational AppScan HTTP-based malware scanning and detection capabilities can help you overcome the inherent problems of existing technologies. Using HTTP-based discovery, Rational AppScan software analyzes the content a Web site user might see, including content from external domains. Because the scanning capability doesn't have to process content in real time, the content can be deeply analyzed using both content and link analysis. Moreover, Rational AppScan software performs the scan using two leading technologies—deep scanning technology from IBM Rational software and malware content and link detection from IBM ISS, helping to achieve high-quality results.

For more information

To learn more about IBM Rational AppScan software, contact your IBM representative or IBM Business Partner, or visit:

ibm.com/software/rational/offerings/websecurity/webappsecurity.html

For more information on Malware, see the following Webcast:

http://download.boulder.ibm.com/ibmdl/pub/software/rational/web/email/Jul16v1.html

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

Produced in the United States of America
November 2009

---

1 IBM, *IBM Internet Security Systems™ X-Force® 2009 Mid-Year Trend and Risk Report*, August 2009 (http://www-935.ibm.com/services/us/iss/xforce/trendreports).

RAW14158-USEN-00