WhiteHat SECURITY

WhiteHat Website Security Statistics Report

October 2007
Jeremiah Grossman
Founder and CTO, WhiteHat Security

NEW

Statistics by Industry Verticals

– see page 8 –

A WhiteHat Security Whitepaper

3003 Bunker Hill Lane, Suite 220 | Santa Clara, CA 95054 | **www.whitehatsec.com**

Introduction

The Web application layer is the number one target for malicious online attacks. Millions of websites regulate access to highly sensitive information including social security numbers, credit card numbers, names, addresses, birthdates, intellectual property, financial records, trade secrets, medical data, and more. This data must be rigorously protected from intruders. To reduce the risk of financial losses, brand damage, theft of intellectual property, legal liability and fines – enterprises need timely information about how websites are penetrated and how they can be defended. Through our flagship offering, WhiteHat Sentinel, WhiteHat Security is uniquely positioned to deliver this information.

WhiteHat Sentinel is a customer controlled and expert managed service providing website vulnerability assessments on an ongoing basis. Customers subscribe annually, and weekly hundreds of the largest and most popular public-facing and pre-production websites are analyzed for vulnerabilities using our unique methodology and three-phase process. Our proprietary technology scans technical vulnerabilities, our experts creates custom checks for each website in the platform to uncover business logic flaws, and all results are verified to remove false-positives. As the only company with access to this amount of website vulnerability data, we can accurately identify which issues are the most prevalent and then trend across major vertical markets including retail, financial, insurance, healthcare and IT industries.

The data contained within this report is also completely different than the reports distributed by Symantec, Mitre (CVE), IBM (ISS) X-Force, and others. These organizations track publicly disclosed vulnerabilities in *commercial and open source software products*, which often contain Web application flaws as well. **WhiteHat's data is different because it focuses solely on previously unknown vulnerabilities** in custom web applications, code unique to that organization, **on real-world websites** (Figure 1). Also, the websites managed under WhiteHat Sentinel are likely represent the most "important" and "secure" websites found on the Web, conducting high-volume transactions and managing sensitive information. This context is helpful when estimating the current global state of website security.

This report is based on data obtained between January 1, 2006 and July 31, 2007.

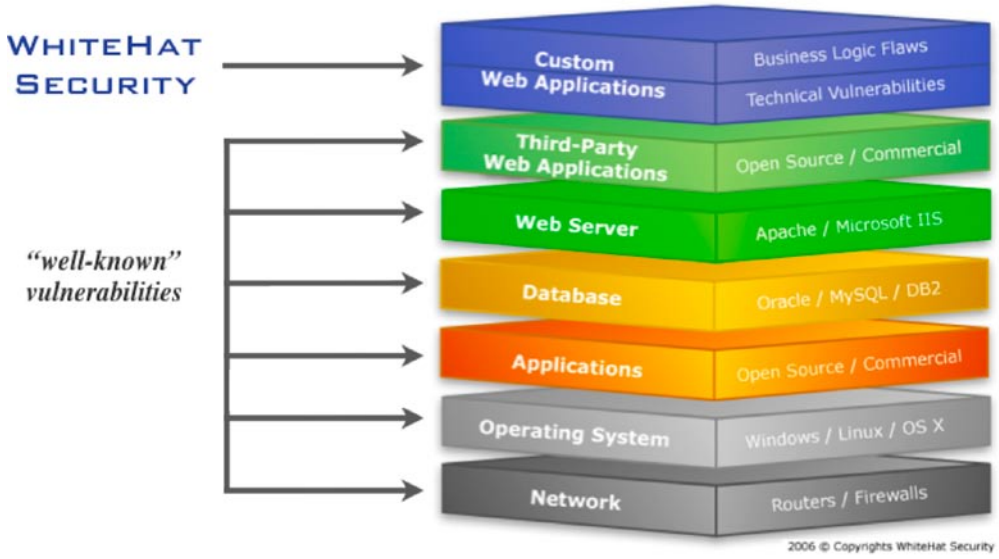

Figure 1. Software Vulnerability Stack.

Top Vulnerability Classes

The number of instances of an individual vulnerability class varies greatly across production websites. For example, one website may possess one hundred unique issues of a specific class, such as Cross-Site Scripting or SQL Injection, while another website may not contain any. As a result, "top" lists based on gross total vulnerabilities are not necessarily the most meaningful. WhiteHat's statistics also calculate the percentage *likelihood of vulnerability classes to occur* within websites (Figure 2), as well their prevalence in the overall population (Figure 3). This two-pronged approach minimizes skewing the data with findings from highly secure and extremely risk-prone website edge cases. Presenting the data in this way helps direct attention toward areas returning the most value.

It is worth noting that unless we add new vulnerability checks or significant technology improvements, our top ten lists remained largely unchanged. Also when we add large blocks of new websites to WhiteHat Sentinel, the vulnerabilities in those websites do not significantly impact the metrics. We believe this illustrates that our data set is largely representative of the current state of website security for the top-tier organizations.

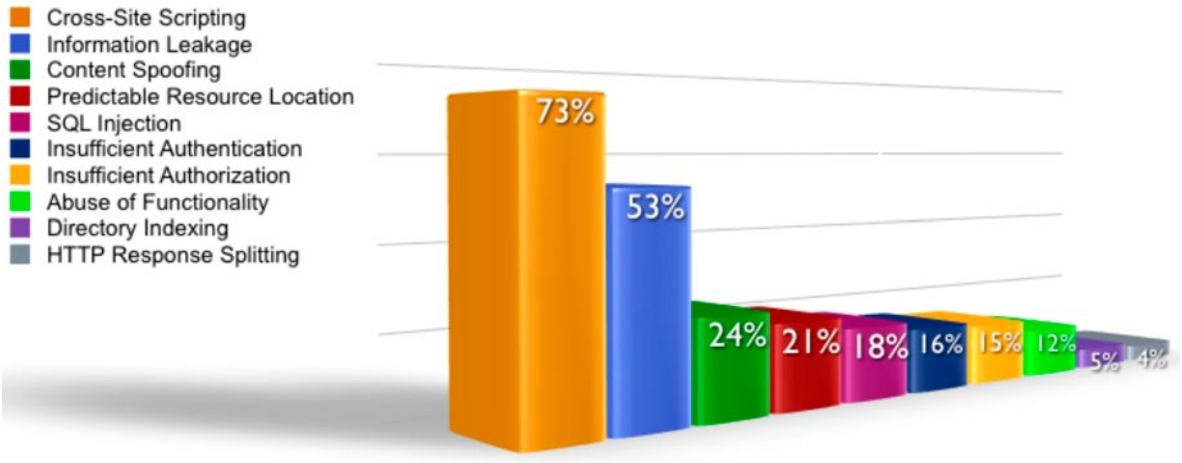

Figure 2. Top 10 vulnerability classes by percentage likelihood.

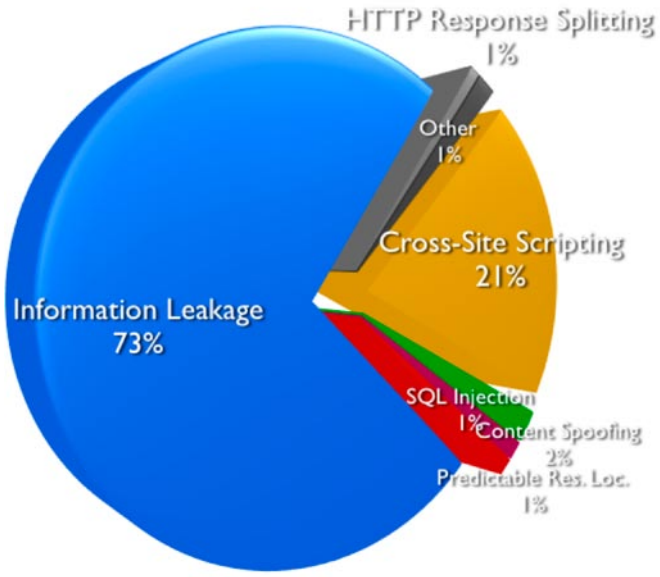

Figure 3. Top 5 vulnerability classes in the overall population.

When comparing Figure 2 against our last report there is a noticeable increase in several scanner found technical vulnerabilities classes including Cross-Site Scripting (XSS), Information Leakage, SQL Injection, and HTTP Response Splitting. This can be directly attributed to the discovery of new attack techniques and our improvement in vulnerability identification technology - not necessarily that the security of Web application software is worsening. Many customers are finding that new vulnerabilities are being regularly found in their software, even when the code is unchanged, because the web application security industry is constantly evolving. For example, a filter bypass issue was recently discovered[1] in .NET's request validator exposing what once were clean websites.

The proverbial floodgates were opened in the Information Leakage class (Figure 3). Reminiscent of the "death by a thousand cuts" metaphor, perhaps just one low/moderate severity Information Leakage issue may not lead to a serious compromise, but several used in combination certainly could[2]. In the best interest of our customer base we decided to be more expansive in our vulnerability reporting. Increasingly detailed information is provided as to what data their website is revealing that perhaps it should not. Software distribution and version numbers, local file paths, debug messages, and stack traces are extremely common types of data that are exposed.

Last, but certainly not least, HTTP Response Splitting[3] replaced XPath Injection[4] at the end of the top ten range. HTTP Response Splitting has proved to be one of the industry's most misunderstood and underestimated issues, evading most scanning technology after the attack technique was revealed some years back. Heavy second quarter R&D efforts by WhiteHat Security resulted in new checks being introduced and vetted across all websites. The results were nothing short of illuminating both in the prevalence and the potential consequences if HTTP Response Splitting exploited, such as cache poisoning and website defacement.

The Top Ten Vulnerabilities Defined

1. Cross-Site Scripting (7 out of 10 websites)

Most industry experts and researchers agree that Cross-site Scripting (XSS) is the most prevalent website vulnerability. Depending on the website, XSS can be extremely hazardous to businesses and consumers. New attack vectors employed are responsible for highly effective phishing scams and Web worms that are resistant to commonly accepted safeguards. The evolution of JavaScript malware, finding its way into more and more attackers toolboxes, has made finding and fixing this vulnerability more vital than ever.

2. Information Leakage (5 in 10 websites)

Information Leakage occurs when a website knowingly or unknowingly reveals sensitive information such as developer comments, user information, internal IP addresses, source code, software versions numbers, error messages/codes, etc., which may all aid in a targeted attack. While most of the time rated medium or low severity, several issued used in combinations could be leverage to compromise a website.

3. Content Spoofing (1 in 4 websites)

Content spoofing is often used in phishing scams as a method of forcing a legitimate website to deliver or redirect users to bogus content. For example, users often receive a suspicious link that instructs them to confirm their user name and password information. Typically, phishing websites are hosted on look-alike domain names mimicking the content of the real site. In the case of Content spoofing phishing scams, fake content is injected into the real website, making it very difficult, if not impossible, for users to detect the difference and therefore protect themselves.

4. Predictable Resource Location (PRL) (1 in 4 websites)

Over time, many pages on a website become unlinked, orphaned, and forgotten – especially on websites experiencing a high rate of content and/or code updates. These Web pages sometimes contain payment logs, software backups, post dated press releases, debug messages, source code – nothing, or everything. Normally, the only mechanism

protecting the sensitive information within is the predictability of the URL. Automated scanners have become adept at uncovering these files by generating thousands of guesses. However, although a scanner can guess at a filename, it has no contextual reasoning to tell if the data received is sensitive or how valuable it might be. Humans need to make this determination.

5. SQL Injection (1 in 5 websites)

SQL Injection has been at the center of some of the largest credit card and identity theft incidents. Today's backend website databases store highly sensitive information, making them a natural, attractive target for malicious hackers. Names, addresses, phone numbers, passwords, birth dates, intellectual property, trade secrets, encryption keys and often much more could be vulnerable to theft. With a few well-placed quotes, semi-colons and commands entire databases could fall into the wrong hands. We believe this statistic is probably higher than reported as many websites have correctly suppressed database error messages, which reduces the risk of exploitation but failed to fix the underlying issue. Genetic Blind SQL Injection techniques employed by all scanners are very immature, which results in a number of false-negatives.

6. Insufficient Authentication (1 in 6 websites)

Insufficient Authentication flaws are typically found within the business logic of an application. Successful exploitation leads to an attacker gaining unauthorized access to protected sections of a website. For example, while logged-in as a normal user, an attacker could impersonate another user on the system. These types of issues are common in financial, healthcare systems, and general content management systems where there is a high concentration of complex business logic functionality.

7. Insufficient Authorization (1 in 6 websites)

Insufficient Authorization flaws are also typically found within the business logic of an application. Successful exploitation leads to an attacker being able to escalate his or her privileges or exercise unauthorized access. For example, while logged-in as a normal user, an attacker could gain access to another user's data while still being logged-in under their current account.

8. Abuse of Functionality (1 in 7 websites)

As stated by the WASC Threat Classification[5] "Abuse of Functionality is an attack technique that uses a website's own features and functionality to consume, defraud, or circumvent access controls mechanisms. Some functionality of a website, possibly even security features, may be abused to cause unexpected behavior. When a piece of functionality is open to abuse, an attacker could potentially annoy other users or perhaps defraud the system entirely."

9. Directory Indexing (1 in 20 websites)

As a feature of most popular Web servers, Directory Indexing lists the contents of a directory if no specific file name is given and no index file is present (example: index.html). Directory listings provided in this way could reveal sensitive information that was not intended for public viewing, such as pre-released Web pages, log files, temporary files, backup files, etc.

10. HTTP Response Splitting (1 in 25 websites)

HTTP Response Splitting is an attack technique in which a single request is sent to the website in such a way that the response may appear to look like two. Depending on the network architecture of the website or the behavior of a user's Web browser, the "second" HTTP response that's under the control of the attacker can be used to poison cache servers, deface web pages, perform session fixation, etc.

What's Not on the List that Should Be

Cross-Site Request Forgery[6] (CSRF) in the last 12-18 month has captured the attention of the leading experts, malicious attackers, and the mainstream media. Most agree CSRF represents a clear and present danger to website security due to its power and pervasiveness. Attackers using CSRF can easily force a user's Web browser to send HTTP request he did not intend to make – a fraudulent wire transfer, password reset, spam relay, downloading illegal content, etc. But, it certainly does not end there as just about every important feature on every website has the potential to be abused in this way, with most being vulnerable. So this begs the question why there are so few statistics available on this vulnerability?

State-of-the art scanning technology across the industry is extremely limited at identifying CSRF and most reported issues are found by hand. The challenge with CSRF is that it's a valid request from the authenticated user. There is no "hack," so to speak, only the behavior of the website can be used to identify an issue. Once again scanners have little to no contextual reasoning to determine if a CSRF attack worked or not or, if it did, how bad it would be if exploited.

WhiteHat's Security Operations Team is leading the research to develop an efficient technology backed process to identify CSRF and make it an integral part of our complete website vulnerability assessment and management process. When this happens, expect CSRF to jump immediately into the Top 10.

Top Vulnerability Classes by Severity Rating

Using the Payment Card Industry Data Security Standard[7] (PCI-DSS) severity system (Urgent, Critical, High, medium, Low) as a baseline, WhiteHat Security ranks vulnerability severity by the potential business impact if the issue were to be exploited. The vast majority of websites (roughly 9 in 10) have had at least one urgent to high severity vulnerability in the last two years while nearly 30% have one or more critical vulnerability. This should have significant meaning for the PCI Council and credit card merchants as websites either having Urgent, Critical, or High severity issues would not pass the PCI compliance test.

Figure 4. Likelihood of websites having vulnerabilities by severity rating.

Urgent Severity Vulnerabilities

SQL Injection continues to top the list as the most common and highest severity vulnerability because it enables direct backend database access. As mentioned earlier, malicious attackers have used this method of attack to compromise millions of personally identifiable records. Insufficient Authorization vulnerabilities, typically only identified by manual assessment, are used to gain access to restricted areas of a website, yielding credit card data, addresses, or other customers' order information. Since the improvement of our testing methodology, HTTP Response Splitting quickly

surpassed Directory Traversal as the third most common Urgent Severity issue. We've seen examples in which we could poison cache servers and deface entire banks of Web servers with a single request.

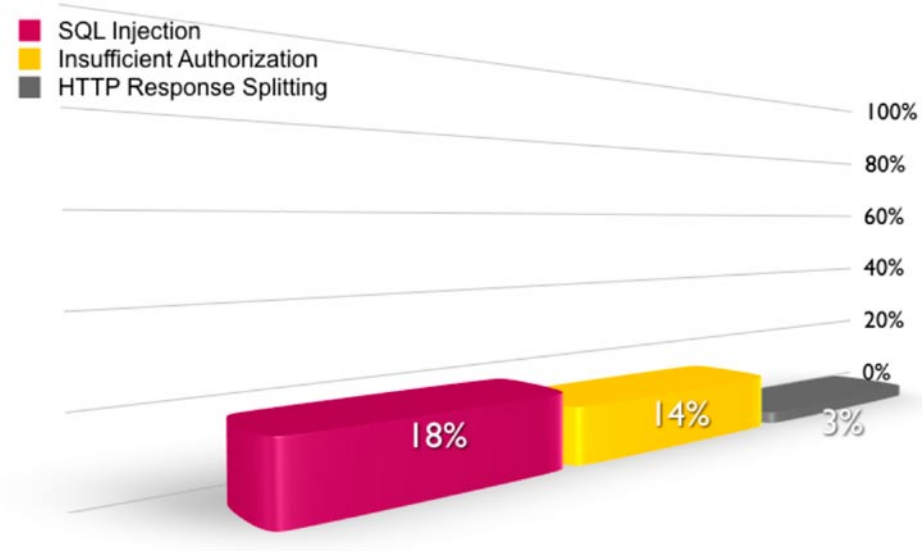

Figure 5. Top 3 Urgent Severity Vulnerability Classes.

Critical Severity Vulnerabilities

XSS is by far the most identified critical-severity vulnerability, appearing in roughly three-quarters of all websites. This is unsurprising, since most are non-persistent types and are by default assigned a critical severity rating. New attack vectors such as XSS-Phishing, Intranet Hacking and Web worms may cause enterprises to re-evaluate XSS vulnerabilities on a case-by-case basis. Insufficient Authentication is prevalent because many websites serve content or execute functionality without first authenticating a user. Typically, an attacker need only type in the proper URL. Abuse of Functionality, which is often identified in combination with other forms of attack, occupies the third spot here. An attacker uses the existing functionality of a website, such as search boxes, chat rooms, or message boards, for his own purposes

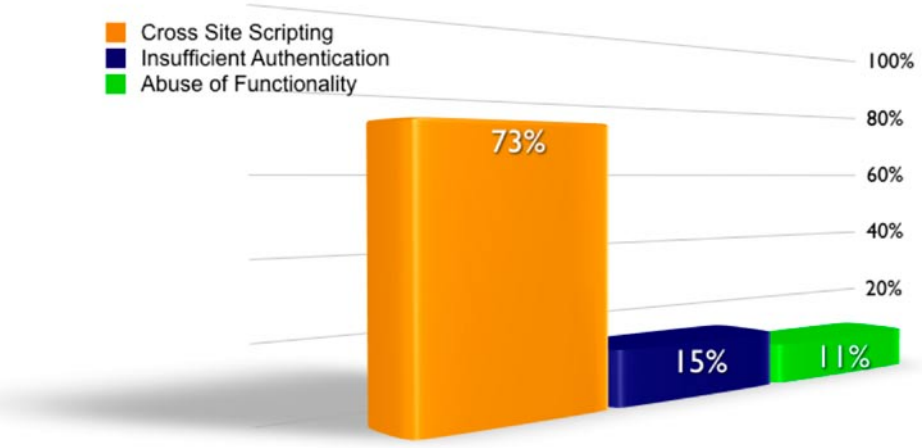

Figure 6. Top 3 Critical Severity Vulnerability Classes.

High Severity Vulnerabilities

Whether mistakenly left in Web page source code or coaxed by an attack, over half of all websites leak sensitive information including internal IP addresses, database names and passwords, software distributions and versions, etc. This type of information is extremely helpful to an attacker attempting to penetrate a system. Next, increasingly sophisticated phishing scams are starting to take advantage of Content Spoofing, found on 1 in 4 websites, as a mechanism to force the real website to host or direct users to bogus content. And, rounding out the list, 1 in 5 websites have files on their Web server that will disclose sensitive information just by requesting the information with the proper URL.

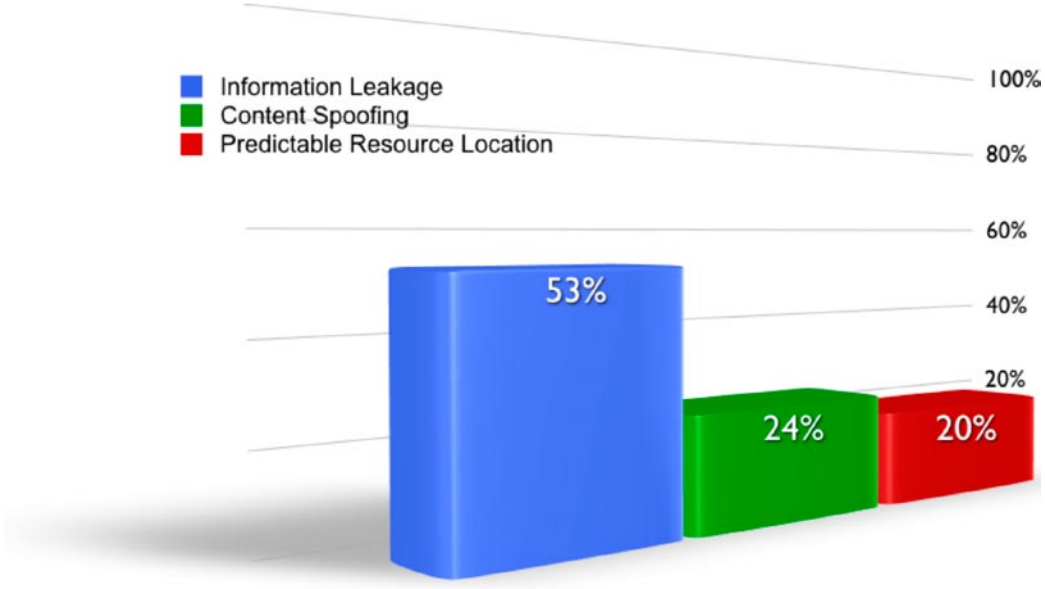

Figure 7. Top 3 High Severity Vulnerability Classes.

Comparing Industry Verticals

With the number of websites managed under WhiteHat Sentinel we feel confident that we have a solid representative sampling of several industry verticals. For the vertical to be included in the list, it had to have at least 25 websites in the category. In some verticals we have a base of hundreds of sites. Figure 8 shows the percentage of websites with at least one Urgent, Critical, or High severity issue. Quickly you'll see the majority of websites have these issues.

Beyond the generally poor state of website security, we did notice that the retail sector is performing better than other verticals. It's difficult to pinpoint exactly why this is the case, but we believe this is due to battlefield testing. The bulk of a standard retail website's functionality is accessible without the need to login. This means more external attackers are able to target these websites and spot weaknesses, which are then remedied by the organization. This is in contrast to the financial services or insurance sectors where the bulk of functionality is protected behind a login screen and an account is therefore harder to access without doing business directly with the company. So once an attacker gets an account, considerably less people have tested these areas of functionality before them.

What the numbers don't show is the average number of vulnerabilities in websites or broken down by vertical. We plan to cover this in more detail in future reports. For now some cases only contain a hand full of issues, while more often they have many dozen and sometimes hundreds. From our experience, larger enterprises are struggling with the number of websites they are responsible for and the huge number of issues that need to be resolved.

Figure 8. Industry vertical comparison.

Top 3 Vulnerability Classes by Industry Vertical

Several issues are fairly consistent across the range, such as Cross-Site Scripting and Information Leakage, while other classes tend to differ between industry verticals. For example, Predictable Resource Location occurs more on Retail websites rather than the others. We believe this is due to the high rate of change experienced on these types of websites where mistakes can occur more often, whether it is leaving on debug messages or failing to remove backup files. Financial Services sees more SQL Injection that the others, likely because these websites have many more dynamic content business processes making database access.

Retail

1. Cross Site Scripting
2. Information Leakage
3. Predictable Resource Location

Financial Services

1. Cross Site Scripting
2. Information Leakage
3. SQL Injection

Insurance

1. Information Leakage
2. Insufficient Authentication
3. Cross Site Scripting

Healthcare

1. Cross Site Scripting
2. Information Leakage
3. Content Spoofing

IT

1. Cross Site Scripting
2. Information Leakage
3. Insufficient Authentication

Conclusion

From the information above it's clear that most e-commerce websites are wide open to attack and easy victims when targeted. And while the security posture of some industries is stronger than others, the difference is insignificant when it comes to actually preventing a website compromise because intruders only needs to exploit a single vulnerability. And, it is well documented that these ever-increasing compromises lead to loss of business, system outages, incident-handling costs, brand damage, legal liability, regulatory sanctions and fines.

WhiteHat Security is dedicated to improving website security and website vulnerability management for its customers and the industry at-large. With 9 out of 10 websites vulnerable to attack, the first step toward stemming the onslaught of attacks is a thorough understanding of the nature of the problem. To make informed security decisions, enterprises require information about the vulnerabilities that exist, their impact, and how to prevent them from occurring. Through this type of industry awareness we expect to see the number and severity of vulnerabilities decrease across the board. This is especially true among enterprises that take a proactive approach to the problem. Organizations are encouraged to do the following:

- *Find and prioritize all website properties by designating their importance to the business and a party responsible for their security.*
- *Find and fix website vulnerabilities before the bad guys exploit them by assessing them for weaknesses with each code change.*
- *Timely remediate vulnerabilities based on severity.*
- *Implement a secure software development process utilizing an organizational standard development framework.*
- *Utilize a defense-in-depth website vulnerability management strategy*

Following these best practices enables organizations to conduct online business with confidence. No company can be expected to write flawless code, or have staff available around-the-clock to address all its Web application vulnerability issues. That's why WhiteHat created WhiteHat Sentinel, a website vulnerability management service that's customer controlled and expert managed. WhiteHat Sentinel is available 24/7, enabling companies to identify, prioritize and ultimately remediate the vulnerabilities that leave websites open to attack.

References

1 *Microsoft ASP.NET request filtering can be bypassed allowing XSS and HTML injection attacks: http://www.procheckup.com/Vulner_PR0703.php*

2 *Death By 1000 Cuts Case Study: http://ha.ckers.org/deathby1000cuts/*

3 *HTTP Response Splitting Revelations http://jeremiahgrossman.blogspot.com/2007/07/http-response-splitting-revelations.html*

4 *XPath Injection http://www.webappsec.org/projects/threat/classes/xpath_injection.shtml*

5 *Web Security Threat Classification: http://www.webappsec.org/projects/threat/*

6 *Cross Site Request Forgery (CSRF): https://whitehatsec.market2lead.com/go/whitehatsec/WPcsrf*

7 *PCI Data Security Standard:* https://www.pcisecuritystandards.org/tech/index.htm

The WhiteHat Sentinel Service – Complete Website Vulnerability Management

Find Vulnerabilities, Protect Your Website – The WhiteHat Sentinel Service is a unique combination of expert analysis and proprietary automated scanning technology that delivers the most comprehensive website vulnerability coverage available. Worried about the OWASP Top Ten vulnerabilities or the WASC Threat Classification? Scanners alone cannot identify all the vulnerabilities defined by these standards. WhiteHat Sentinel can. Many of the most dangerous vulnerabilities reside in the business logic of an application and are only uncovered through expert human analysis.

Continuous Improvement and Refinement – WhiteHat Sentinel stays one step ahead of the latest website attack vectors with persistent updates and refinements to its service. Updates are continuous – as often as one day to several weeks, versus up to six months or longer for traditional software tools. And, Sentinel uses its unique "Inspector" technology to apply identified vulnerabilities across every website it evaluates. Ultimately, each site benefits from the protection of others.

Virtually Eliminate False Positives – No busy security team has time to deal with false positives. That's why the WhiteHat Sentinel Security Operations Team verifies the results of all scans. Customers see only real, actionable vulnerabilities, saving time and money.

Total Control – WhiteHat Sentinel runs on the customer's schedule, not ours. Scans can be manually or automatically scheduled to run daily, weekly, and as often as websites change. Whenever required, WhiteHat Sentinel provides a comprehensive assessment, plus prioritization recommendations based on threat and severity levels, to better arm security professionals with the knowledge needed to secure them.

Unlimited Assessments, Anytime Websites Change – With WhiteHat Sentinel, customers pay a single annual fee, with unlimited assessments per year. And, the more applications under management with WhiteHat Sentinel, the lower the annual cost per application. High volume e-commerce sites may have weekly code changes, while others change monthly. WhiteHat Sentinel offers the flexibility to assess sites as frequent as necessary.

Simplified Management – There is no cumbersome software installation and configuration. Initial vulnerability assessments can often be up-and-running in a matter of hours. With WhiteHat Sentinel's Web interface, vulnerability data can be easily accessed, scans or print reports can be scheduled at any time from any location. No outlays for software, hardware or an engineer to run the scanner and interpret results. With the WhiteHat Sentinel Service, website vulnerability management is simplified and under control.

About the Author

Jeremiah Grossman is the founder and CTO of WhiteHat Security, a world-renowned expert in website vulnerability management, co-founder of the Web Application Security Consortium, and recently named to InfoWorld's Top 25 CTOs for 2007. Mr. Grossman is a frequent speaker at industry events including the BlackHat Briefings, ISACA, CSI, OWASP, Vanguard, ISSA, OWASP, Defcon, etc. He has authored of dozens of articles and white papers, credited with the discovery of many cutting-edge attack and defensive techniques and is co-author of the book *XSS Exploits*. Mr. Grossman is frequently quoted in major media publications such as InfoWorld, USA Today, PCWorld, Dark Reading, SC Magazine, SecurityFocus, CNET, SC Magazine, CSO, and InformationWeek. Prior to WhiteHat he was an information security officer at Yahoo!

About WhiteHat Security, Inc.

Headquartered in Santa Clara, California, WhiteHat Security is a leading provider of website vulnerability management services. WhiteHat delivers turnkey solutions that enable companies to secure valuable customer data, comply with industry standards and maintain brand integrity. WhiteHat Sentinel, the company's flagship service, is the only solution that incorporates expert analysis and industry-leading technology to provide unparalleled coverage to protect critical data from attacks. For more information about WhiteHat Security, please visit www.whitehatsec.com.

WhiteHat SECURITY

WhiteHat Security, Inc. | 3003 Bunker Hill Lane, Suite 220 | Santa Clara, CA 95054 | 408.343.8300 | **www.whitehatsec.com**

10.01.07