Challenges of Automated Web Application Scanning

"Why automated scanning only solves half the problem."

Blackhat Windows 2004
Seattle, WA

Jeremiah Grossman (CEO)
WhiteHat Security, Inc.
Speaker Bio

Jeremiah Grossman (Chief Executive Officer)

- Founded WhiteHat Security in 2001
- Former Yahoo! Information Security Officer
- Primary developer of WhiteHat Arsenal, Web Server Fingerprinter, CIS Apache Benchmark Tool, and platform for Sentinel
- Performed over 300 web application security assessments
Topics of conversation

- Web application security landscape
- Top 5 myths of web site security
- Tools only solve the half the problem
- 6 ways to improve web application security

Mission: Develop a process for securing web sites
Web Security is about Layer 7

Layer 1-6 security solutions are ineffective for web security

User requests a web page.
Firewalls only allows port 80 and 443.

Web Server queries the Database records.
Internal Firewall only allows traffic from the Web Server to the Database.

Internal Database only accessible by the Web Server.
No external network access.
Web Security Hacks in the News

_guess

200,000 credit card numbers compromised

SQL Injection permitted a properly-crafted URL to have direct access to the customer database. FTC Settlement prohibits the company from misrepresenting the security of personal information collected from or about consumers.

Howard Beales - FTC's Bureau of Consumer Protection
"Consumers have every right to expect that a business that says it's keeping personal information secure is doing exactly that"
Victoria’s Secret
Fined $50,000 by the FTC

Customer order information accessible by changing a number in the URL. Must implement security reforms outline by the Federal Trade Commission.

Eliot Spitzer - New York State Attorney General
"A business that obtains consumers' personal information has a legal duty to ensure that the use and handling of that data complies in all respects"

Also hacked: Travelocity, RIAA, FTD.com, Creditcards.com, Tower Records, CD Universe
Attractive Targets

Delivering data that’s valuable

- Credit card numbers
- Bank account information
- Personal and private information
- Medical history

The Gartner Group
“97% of the over 300 Web sites audited were found vulnerable to web application attack.”

“75% of the cyber attacks today are at the application level.”

© 2004 by WhiteHat Security, Inc.
The Fed’s are getting restless

- More business applications are moving to the web.
- FTC Privacy Issues
- Sarbanes-Oxley
- Graham-Leach-Bliley
- HIPAA
Who’s in charge?

Network  Security  Developers  QA  Management
Top 5 myths of web site security

I am secure because...

- We use 128-Bit SSL
- Firewalls protect the web site
- ISS/Eeye/Nessus shows no issues
- My application scanner found no issues
- We have annual security assessments
SSL only protects traffic passing between the web site and the client.
Opening 80 and 443

Hackers
Firewall
Port 443 Open
SSL
IDS
A scanner for every day of the week
How web applications scanners scan

Network Security Scanning:
“Identifying known vulnerabilities in known code.”

Web Application Security:
“Identifying known classes of vulnerabilities in unknown code.”
Today’s best practices fall short

- One time assessments are the norm
- Web sites change frequently
- Cost prohibitive to perform more than once
- Scanning tools only solve half the problem
- Scanners only handle technical vulnerabilities
- Current products are not robust

Insecure

<table>
<thead>
<tr>
<th>1 Week</th>
<th>51 Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weekly Code Pushes</td>
<td></td>
</tr>
</tbody>
</table>
Tools, the incomplete solution
WhiteHat’s Team

Performed security assessments on over 1,000 web sites. Spent the last two years developing Sentinel, our web application scanning technology. Sentinel required an incredible R&D effort to overcome unforeseen challenges.

Challenges of Automated Scanning

Remote Scanning, Automated Login, Infinite Web Sites, Rate-of-Change, Strange URL Structure, Client-Side Scripting, Anti-Automation, Multi-Page Sequence, Authentication System Auditing, and Non-Standard Errors to name but a few.
Two kinds of issues

Scanners are unable to indentify flaws in business logic.

The most devastating attacks are found by people.
Technical vulnerability

- String of code or repeatable pattern that a computer can be programmed to recognize.
- If I put a single quote there and get an ODBC error then there is a SQL Injection vulnerability.
Logical flaws

“At step 3 of the wire transfer process, change the account parameter to point to the account you wish to transfer funds from. Continue changing the parameter on the next 2 steps of the transfer process.”

A scanner is unable to determine context of good or bad
Humans vs. Scanners

Humans, as well as automated scanners, are best suited for identifying different types of security issues.

Scanners can be expected to be very thorough in the testing process, but are only able to identify “technical” vulnerabilities.

These automated scanners will not uncover multi-page sequence problems that often occur in complex web application.

A human possess the ability to analyze a large set of circumstances and determine if a weakness in a process exists.
Automating vulnerability discovery

Halting Problem

The halting problem is a decision problem which can be informally stated as follows:

“Given a description of an algorithm and a description of its initial arguments, determine whether the algorithm, when executed with these arguments, ever halts.”

Undecidable Problem

“Not all problems can be solved. An undecidable problem is one that cannot be solved by any algorithm, even given unbounded time and memory.”
State-of-the-Art

Technical vs. Logical

50/50
Remote black box scanning

- Starting from zero
  - No access to source code
  - No access to binaries
  - No preexisting knowledge about the software distribution or architecture

- There is nothing known!
Automated Login

The web application scanner must be able to generically login to a web application on demand.

Most authentication systems are set-up differently.

If a valid session is not maintained, the scan is invalid because full application functionality cannot be exercised.
Infinite web sites

Many web sites are enormous and crawling the entire site in a reasonable amount of time is impossible. Must compile an accurate structural map.

@ 2 HTTP Request Per Second = 2.9 days to crawl

Dynamic Web Sites: Rate of addition

Rate of decay
Database of items 500,000+ links
Dynamic URL creation
Model the web site structure

Condense the amount of links we need to crawl and create a complete application structural map of the web site.

Locate all web applications and all unique parameter name instances
Detecting Logout

A scanner will at some point become logged out. How does the scanner know when that happens?

Logout can occur by:

Clicking logout links, Timing out, Application errors, Session expiration, etc, etc, etc, …
Detecting and Invalid Session

We designed system that performs preliminary tests on the web application to learn the login/logout nuances.
Multi-Page Sequence

Web sites will commonly have multi-page business processes utilizing HTML forms. The user must properly perform certain steps through the process to move on.

Application flow is VERY difficult to traverse and mapped automatically by a web application scanner. The scanner has no context to decide what to put into a form field or if answer they receive is either good or bad.
Web applications are constantly changing. Normally there is one or more new revisions a year with incremental updates.

Every new line of code potentially introduces new security issues.

Rate of change negatively affects the ability to complete a scan and maintain login state.
Map the URL structure

- The normal web application URL structure has a “?” delimiting the file name from the parameters. However, developers have realized that many web spiders will not index dynamic data so they have opted for some non-standard URL trickery.

- The goal is to identify:
  - Web application filename
  - Web application parameter names and values

- Even if:
  - There is no question mark
  - No “&” and uses strange delimiters.
  - Strange file extension (like .html)
Normal URL Structure

- Normal:
  - /articles/03/08/19/1748206.shtml?tid=109&tid=111&tid=126
  - /news?hl=en&edition=us&q=a&btnG=Search+News
  - /shopping/category.asp?categoryID=11
  - /weeknight_survival.asp?wday=3&ww=this

- Inject into the name value pairs
Strange URL Structure

Strange, where is the injection point?

- /gp/browse.html/10217298046144934?node=1036592
- /exec/obidos/ASIN/B00009J5VW/ref=e_hp_cb_3_1/12-1729804-6144934
- /srs7/sid=030803095821064050032/g=home/search/detail/base_pid/271134/
- /catindex/computers.html?ssPageName=MOPS5:HEC03
- /exec/obidos/subst/home/home.html/102-17298046144934
- /shop/enter.asp?category=2378467~2378483
Client Side Scripting

Dynamic Link Menus
Sometimes web sites will have menus and style-sheets which create hyperlinks on the fly. In these cases, web crawlers have a extremely difficult time traversing the site since the links are not yet built or parse-able.

HTML Encryption
Some web sites will have their HTML encrypted and then decrypted by javascript when read into the browser.
Many web application authentication systems are inherently weak. Many are susceptible to session hi-jacking, session replay, etc.

- Cookie: T=user=100 or Cookie: T=user=101
- Or
  - Cookie: S=UID=ae5fad5ad6a8asd6as9

Even if the scanner does twiddle the bits, how does scanner know when something works or does not work, or what's good or what's bad?

How does a scanner know when it accesses another users bank account?

“Scanner is not able to generically determine context of good or bad”
Non-Standard Not-Found

Not Found does not always mean, “Not Found”.

Not everyone is RFC compliant
Universal Error Catching
Error strings are different
Standard Responses

Web application scanning will become harder in the future as more systems are configured by default to suppress error messages.

Server Error in '/' Application.

The resource cannot be found.

Description: HTTP 404. The resource you are looking for (or one of its dependencies) could have been removed, had its name changed, or is temporarily unavailable. Please review the following URL and make sure that it is spelled correctly.

Requested URL: /dashbo'ard.aspx

Version Information: Microsoft .NET Framework Version:1.0.3705.288; ASP.NET Version:1.0.3705.288
Error Response Messages

Application Errors:

- SQL Injections
- Cross-Site Scripting
- Command Injection

Removing response messages helps prevent against exploitation. However, prevents scanners from finding the vulnerabilities. Lots of false positives.
Anti-Automation

If a computer is not supposed to automate the process, then how can a scanner?
Lessons learned?

- We all have security issues in our web sites
- One-time security assessments are not enough
- Continuous security review by qualified personnel
- Scanners help, but not a complete solution
Process: Combined Approach

Technology
- Thorough scan
- Continuous scanning
- Augments people

People
- Evaluate results
- Assess business logic
6 ways to improve web security

- Experienced web security staff must perform assessments at least once a quarter.
- Automated vulnerability scans should be performed at least twice a month or as application change demands.
- The security scanner should cover all technical security issues, able to support large web sites, maintain a logged-in state, yield a low volume of false positives, and remain current.
- Application developers must consider security from the beginning.
- Involve security staff early in the development process.
- Stay patched and configure properly.
Thank You - Questions?

Jeremiah Grossman
jeremiah@whitehatsec.com
http://www.whitehatsec.com