Web Application Security Statistics Project 2007

Purpose

The Web Application Security Consortium (WASC) is pleased to announce the WASC Web Application Security Statistics Project 2007. This initiative is a collaborative industry wide effort to pool together sanitized website vulnerability data and to gain a better understanding about the web application vulnerability landscape. We ascertain which classes of attacks are the most prevalent regardless of the methodology used to identify them. Industry statistics such as those compiled by Mitre CVE project provide valuable insight into the types of vulnerabilities discovered in open source and commercial applications, this project tries to be the equivalent for custom web applications.

Goals

1. Identify the prevalence and probability of different vulnerability classes
2. Compare testing methodologies against what types of vulnerabilities they are likely to identify.

TABLE OF CONTENT

Methodology

The statistics was compiled from web application security assessment projects which were made by the following companies in 2007 (in alphabetic order):

Booz Allen Hamilton
BT
Cenzic with Hailstorm and ClickToSecure
dblogic.it
HP Application Security Center with WebInspect
Positive Technologies with MaxPatrol
Veracode with Veracode Security Review
WhiteHat Security with WhiteHat Sentinel

Identified vulnerabilities for assessment technologies have been aggregated using the Web Security Threat Classification (http://www.webappsec.org/projects/threat/) as a baseline.

The statistics includes 2 different data sets: automated testing results and security assessment results made using black and white box methodology.

Automated testing results contain data about the scanning of hosting provider sites without any customizing the settings (with standard profile). While analyzing this data it is recommended to consider that not every site tested uses interactive elements. Additional customized settings made by an expert would improve vulnerability detection effectiveness by automated scanners.

It is also recommended to consider automatic scanning type 1 and 2 errors: the scanner might miss the vulnerability or suggest the vulnerability which does not exist. A manual expert assessment allows to eliminate type 2 errors and to minimize type 1 errors (but not eliminate them).

Black and white box security assessment statistics contain manual and automatic analysis results. The analysis includes scanning with preliminary settings followed by manual analysis, manual search for vulnerabilities which cannot be detected by automated scanner, and source code analysis.

Consequently 3 data sets were obtained:

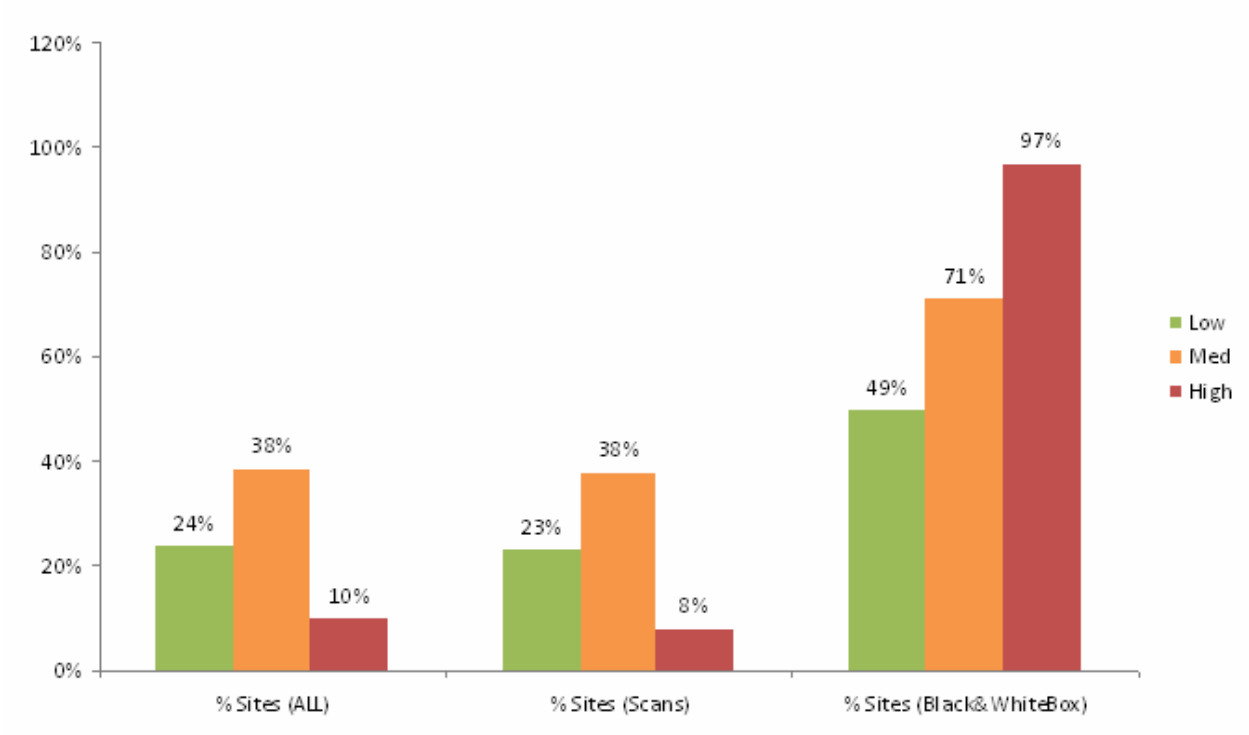
1. Overall statistics
2. Automated scanning statistics
3. Black and and white box methods security assessment statistics

The overall statistics includes analysis results of 32,717 sites and 69,476 vulnerabilities of different degrees of severity. The detailed information can be found in Statistics chapter.

Data analysis

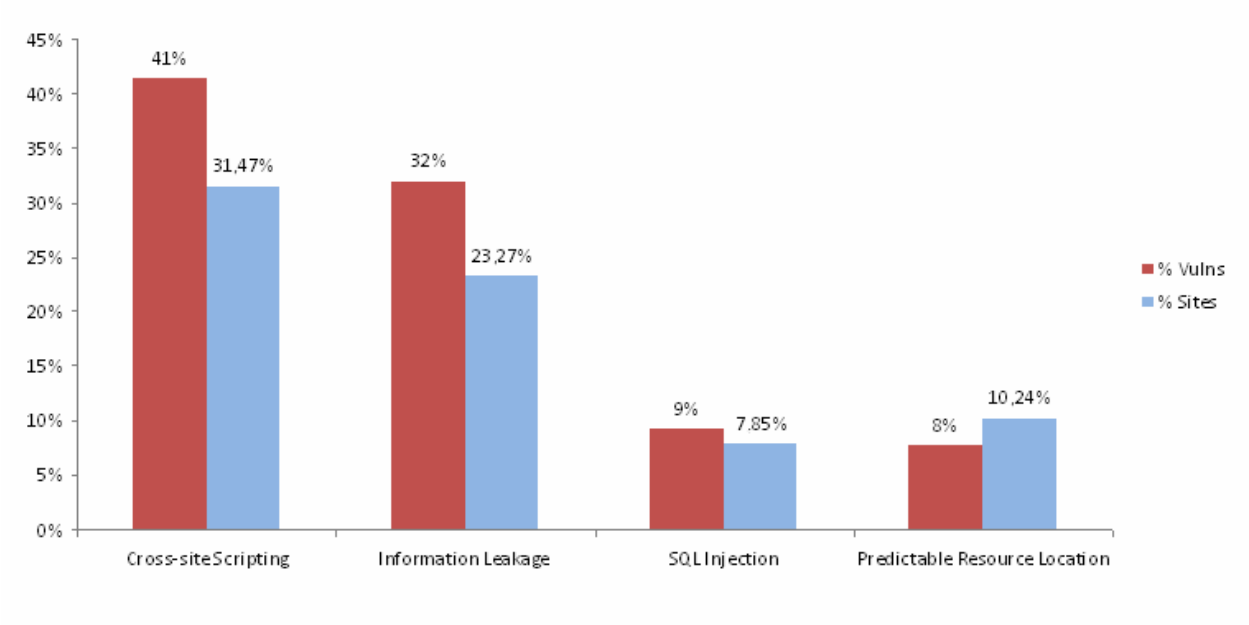
Data analysis shows that more than 7% of analyzed sites can be compromised automatically. About 7.72% applications had a high severity vulnerability detected during automated scanning (P. 1). Detailed manual and automated assessment using white and black box methods shows that probability to detect high severity vulnerability reaches 96.85%.

So automated scanning represents data for an average Internet site and black and white box methods results refer to interactive corporate web applications.

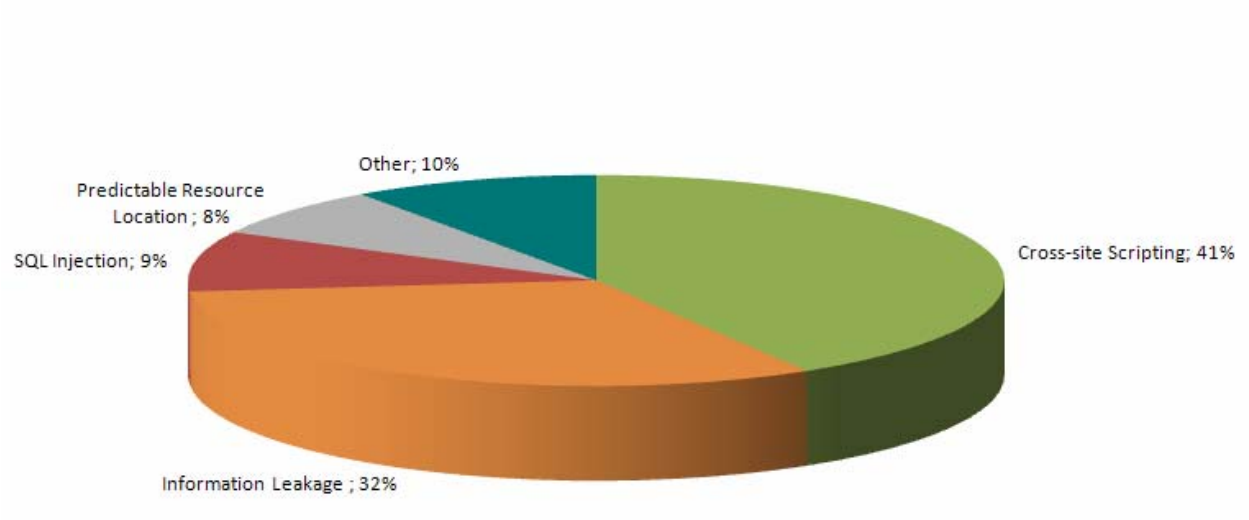


P. 1 Probability to detect vulnerabilities of different risk degree

The most prevalent vulnerabilities are Cross-Site Scripting, Information Leakage, SQL Injection and Predictable Resource Location (P. 2, P. 3). As a rule, Cross-Site Scripting and SQL Injection vulnerabilities appears due to system design errors, Information Leakage and Predictable Resource Location are often connected with improper system administration (for example, weak access control).

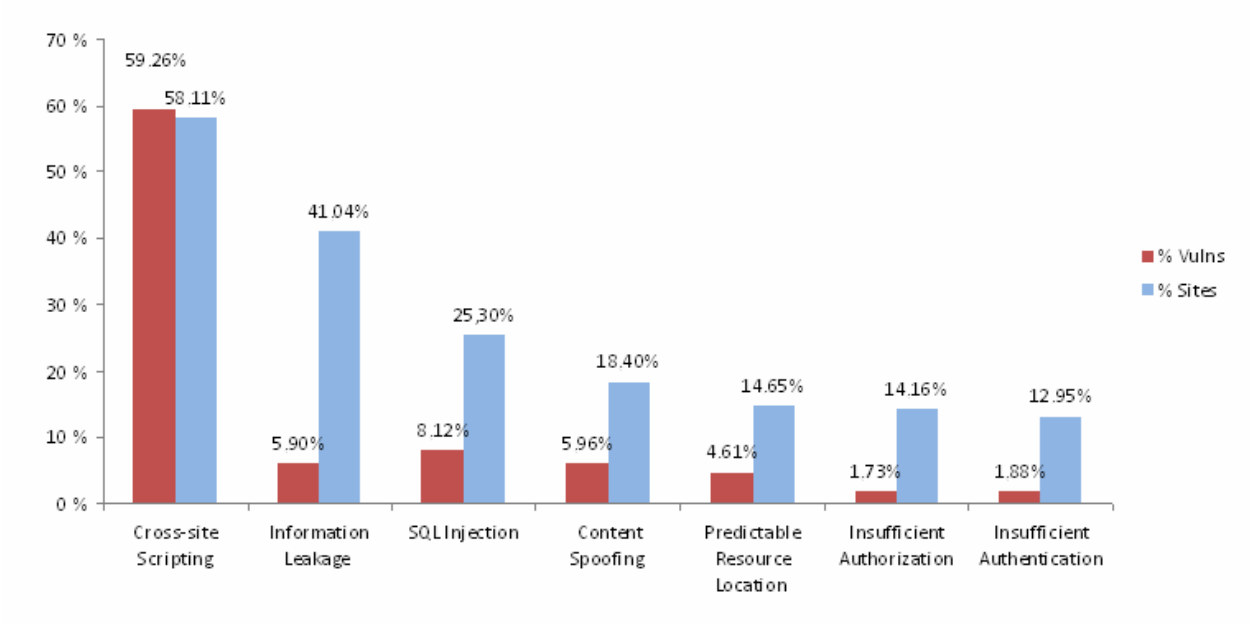


P. 2 The most prevalent vulnerabilities

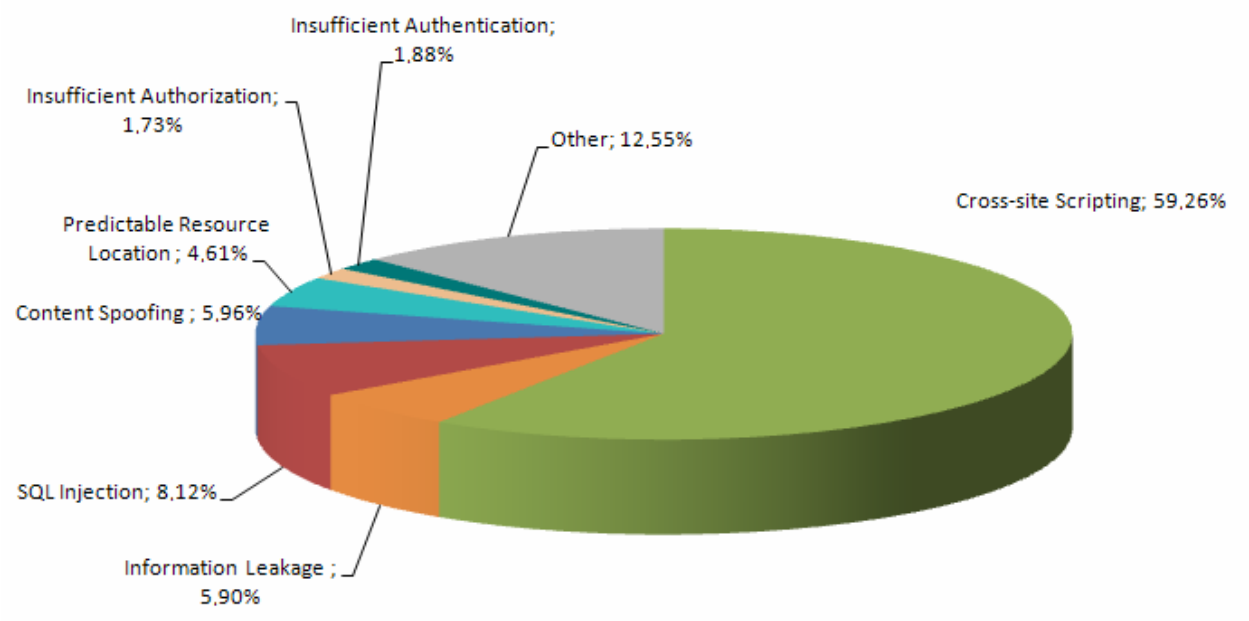


P. 3 Vulnerability frequency by types

While detailed system analysis with BlackBox and WhiteBox methods appreciable percentage of sites are vulnerable to Content Spoofing, Insufficient Authorization and Insufficient Authentication (P. 4, P. 5). With this approach to security assessment the probability to detect SQL Injection reaches 25%.



P. 4 The most prevalent vulnerabilities (BlackBox & WhiteBox)

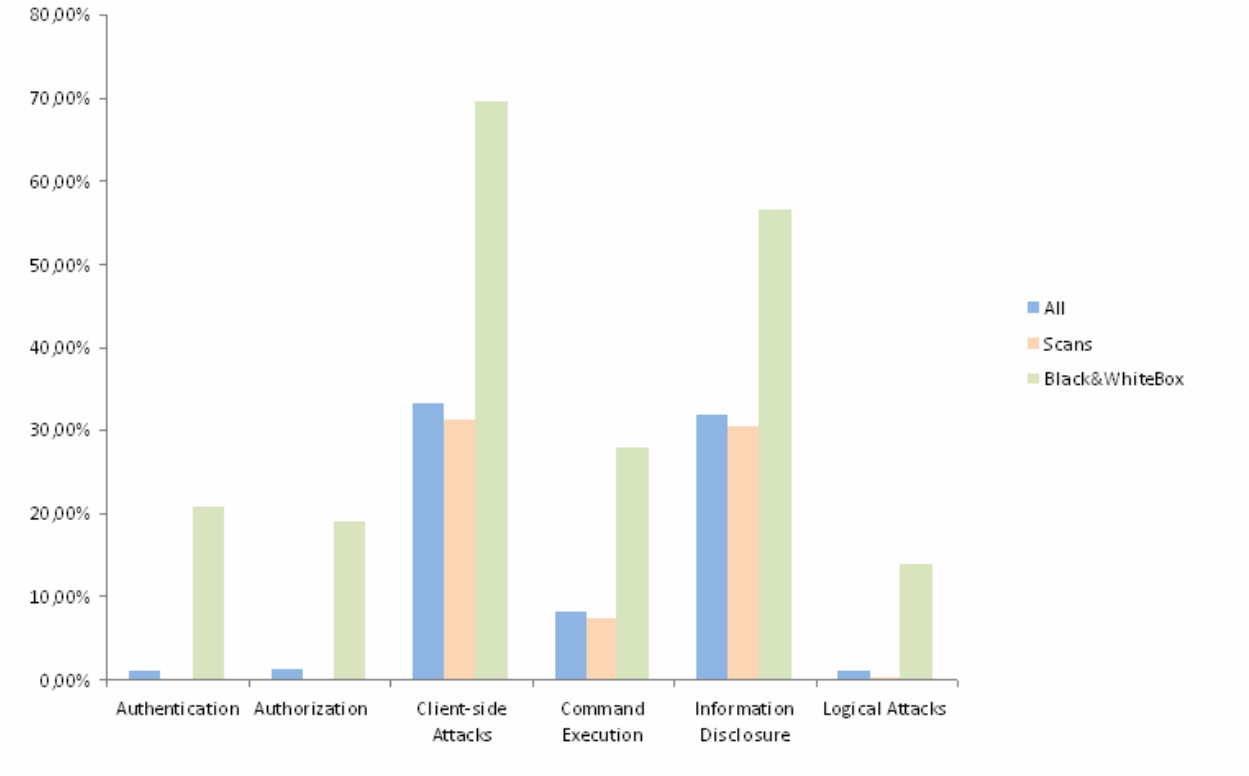


P. 5 Vulnerability frequency by types (BlackBox & WhiteBox)

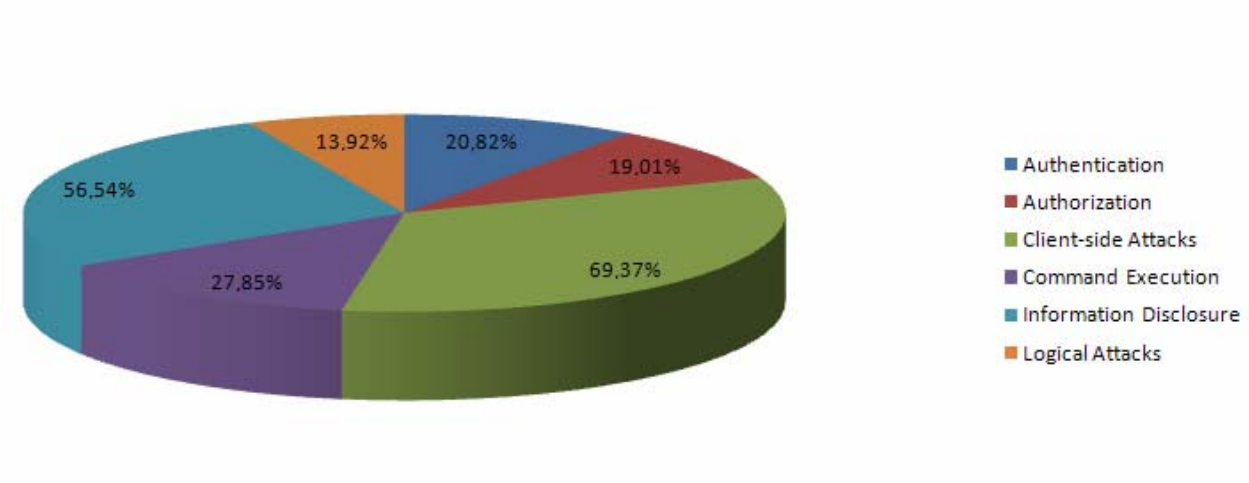
In terms of Web Application Consortium Threat Classification version 1 classes (T. 1 and P. 6) the most prevalent classes of vulnerabilities are Client-side Attacks, Information Disclosure and Command Execution. The detailed analysis shows the popularity of Authentication and Authorization classes (P. 7).

T. 1 The probability distribution of vulnerabilities detection according to WASC TCv1 classes

	% ALL	% Scans	% Black & WhiteBox
Authentication	1.17%	0.02%	20.82%
Authorization	1.28%	0.07%	19.01%
Client-side Attacks	33.13%	31.17%	69.37%
Command Execution	8.15%	7.32%	27.85%
Information Disclosure	31.78%	30.42%	56.54%
Logical Attacks	0.90%	0.20%	13.92%



P. 6 The probability distribution of vulnerabilities detection according to WASC TCv1 classes



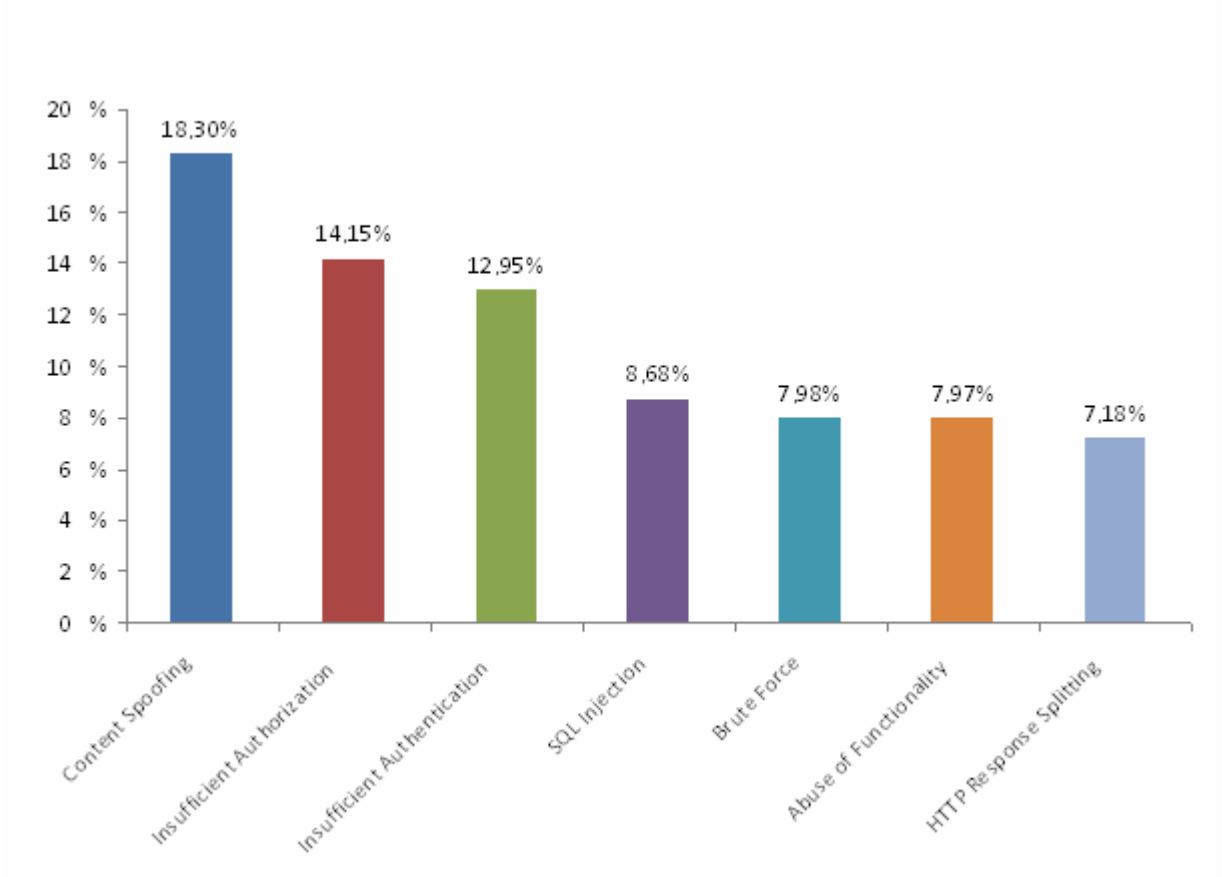
P. 7 The probability distribution of vulnerabilities detection according to WASC TCv1 classes (BlackBox & WhiteBox)

The Comparison of security assessment methods

While compared automated scanning with detailed Blackbox and Whitebox analysis methods, it is evidently clear that detailed analysis is much more effective to detect Authorization and Authentication class vulnerabilities and logic flaws (T. 2, P. 8).

T. 2 Automated scanning vs Blackbox and Whitebox analisys (% Sites)

Threat Classification	Scans vs Black & WhiteBox
Content Spoofing	18.30%
Insufficient Authorization	14.15%
Insufficient Authentication	12.95%
SQL Injection	8.68%

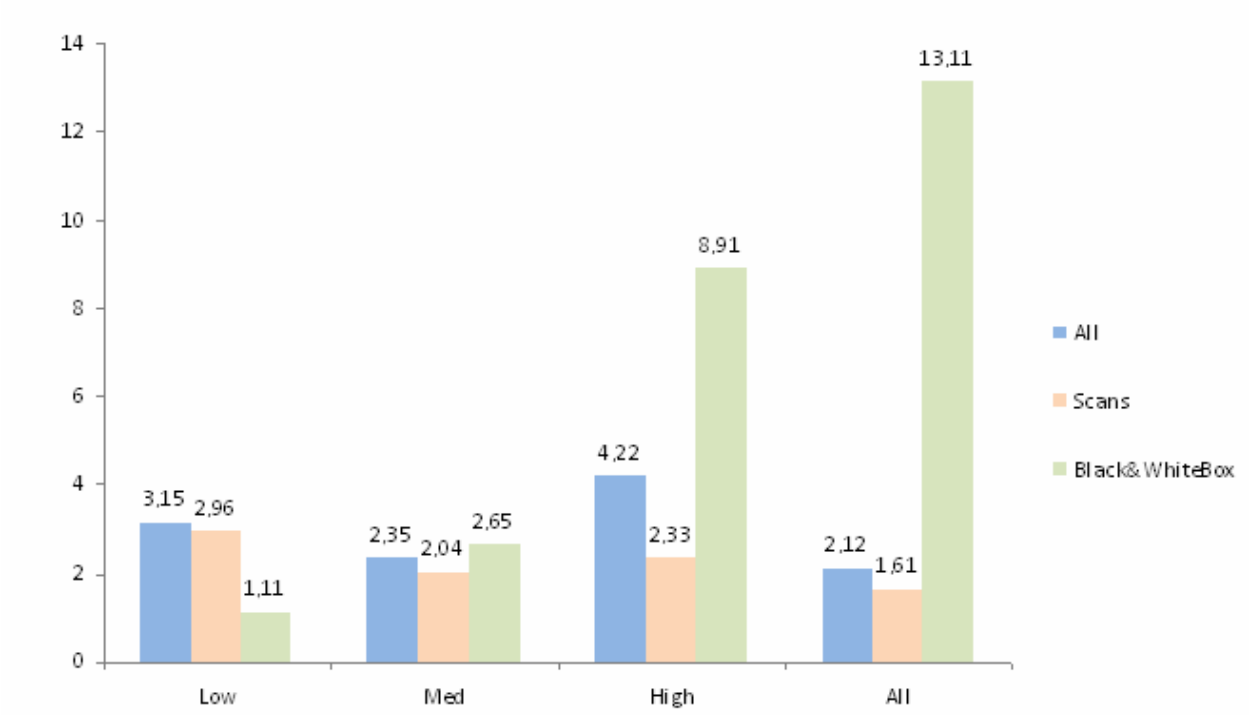


P. 8 The difference in probability of vulnerabilities detection using different methods

As mentioned above (P 1), the probability to detect high risk degree vulnerability using detailed analysis is 12.5 times higher than using automated scanning. According to the number of vulnerabilities detected for a site (T. 3 and P. 9) the detailed analysis allows to detect on average 9 high risk degree vulnerabilities per site while automated scanning allows to detect only 2.3 vulnerabilities of this rank.

T. 3 Number of vulnerabilities per site

	All	Scans	Black&WhiteBox
Low	3.15	2.96	1.11
Med	2.35	2.04	2.65
High	4.22	2.33	8.91
All	2.12	1.61	13.11



P. 9 Number of vulnerabilities per site

Additional notes

Web Application Security Consortium Threat Classification version 1 was used in this research. Therefore some types of vulnerabilities are not included into the overall results. We plan to use WASC TC version 2 in future.

The most prevalent vulnerability Cross-Site Request Forgery in this statistics is not on top because it is difficult to detect in automatically and because a lot of experts take its existence for granted.

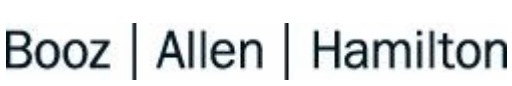
Vulnerabilities which existence depends on platform are also not included into the statistics (for example, buffer overflow in Apache).

Contributors

WASC would like to thank the following organizations and persons for making this initiative possible. Each organization is responsible for contributing sanitized data from web application security projects which was then combined to produce aggregated statistics.

*Sergey Gordeychik**
Jeremiah Grossman
Michael Sutton
Mandeep Khera
Peter Ahearn
Brian Martin
Simone Onofri
Matt Latinga
Chris Wysopal

**Project Leader*



Statistics

Overall Data

T. 4 General statistics

Threat Classification	N of Vulns	N of Sites	% Vulns	% Sites
Abuse of Functionality	169	99	0.24%	0.30%
Brute Force	291	125	0.42%	0.38%
Buffer Overflow	171	19	0.25%	0.06%
Content Spoofing	1399	213	2.01%	0.65%
Credential/Session Prediction	79	46	0.11%	0.14%
Cross-site request forgery	993	126	1.43%	0.39%
Cross-site Scripting	28769	10297	41.41%	31.47%
Denial of Service	55	44	0.08%	0.13%
Directory Indexing	281	87	0.40%	0.27%
Fingerprinting	120	60	0.17%	0.18%
Format String Attack	104	12	0.15%	0.04%
HTTP Response Splitting	749	265	1.08%	0.81%
Information Leakage	22156	7614	31.89%	23.27%
Insufficient Anti-automation	288	115	0.41%	0.35%
Insufficient Authentication	356	229	0.51%	0.70%
Insufficient Authorization	343	218	0.49%	0.67%
Insufficient Process Validation	117	38	0.17%	0.12%
Insufficient Session Expiration	200	91	0.29%	0.28%
LDAP Injection	21	11	0.03%	0.03%
OS Commanding	25	9	0.04%	0.03%
Path Traversal	178	145	0.26%	0.44%
Predictable Resource Location	5331	3349	7.67%	10.24%
Session Fixation	183	65	0.26%	0.20%
SQL Injection	6420	2567	9.24%	7.85%
SSI Injection	185	40	0.27%	0.12%
URL Redirectors	210	195	0.30%	0.60%
Weak Password Recovery Validation	164	31	0.24%	0.09%
WSDL Exposure	60	20	0.09%	0.06%
XPath Injection	59	16	0.08%	0.05%
Total	69476	32717		

T. 5 Vulnerabilities distribution by risk

Threat rank	N of Vulns	N of Sites	% Vulns	% Sites
Low	24433	7760	35.17%	23.72%
Med	29575	12596	42.57%	38.50%
High	13765	3263	19.81%	9.97%

T. 6 Vulnerabilities distribution by WASC TCv1 classes

WASC Classes	N of Vulns	N of Sites	% of Vulns	% Sites
Authentication	811	384	1.17%	1.17%
Authorization	805	418	1.16%	1.28%
Client-side Attacks	32120	10840	46.23%	33.13%
Command Execution	6985	2665	10.05%	8.15%
Information Disclosure	28126	10398	40.48%	31.78%
Logical Attacks	629	295	0.91%	0.90%

Automatic scans

T. 7 General statistics

Threat Classification	N of Vulns	N of Sites	% Vulns	% Sites
Abuse of Functionality	5	3	0.01%	0.01%
Brute Force	3	3	0.01%	0.01%
Buffer Overflow			0.00%	0.00%
Content Spoofing	33	21	0.06%	0.07%
Credential/Session Prediction	4	4	0.01%	0.01%
Cross-site request forgery	87	53	0.17%	0.17%
Cross-site Scripting	19171	9651	37.29%	30.26%
Denial of Service	30	22	0.06%	0.07%
Directory Indexing	91	12	0.18%	0.04%
Fingerprinting			0.00%	0.00%
Format String Attack			0.00%	0.00%
HTTP Response Splitting	182	161	0.35%	0.50%
Information Leakage	21157	7115	41.16%	22.31%
Insufficient Anti-automation	37	38	0.07%	0.12%
Insufficient Authentication	2	2	0.00%	0.01%
Insufficient Authorization	6	11	0.01%	0.03%
Insufficient Process Validation			0.00%	0.00%
Insufficient Session Expiration	3	3	0.01%	0.01%
LDAP Injection			0.00%	0.00%
OS Commanding	19	3	0.04%	0.01%
Path Traversal	118	116	0.23%	0.36%
Predictable Resource Location	4772	3213	9.28%	10.07%
Session Fixation	3	3	0.01%	0.01%
SQL Injection	5301	2298	10.31%	7.21%
SSI Injection	180	37	0.35%	0.12%
URL Redirectors	195	182	0.38%	0.57%
Weak Password Recovery Validation	1	1	0.00%	0.00%
WSDL Exposure			0.00%	0.00%
XPath Injection	4	1	0.01%	0.00%
Total	51404	31891		

T. 8 Vulnerabilities distribution by risk

Threat rank	N of Vulns	N of Sites	% Vulns	% Sites
Low	21736	7352	42.28%	23.05%
Med	24452	12012	47.57%	37.67%
High	5736	2463	11.16%	7.72%

T. 9 Vulnerabilities distribution by WASC TCv1 classes

WASC Classes	N of Vulns	N of Sites	% of Vulns	% Sites
Authentication	6	6	0.01%	0.02%
Authorization	16	21	0.03%	0.07%
Client-side Attacks	19668	9941	38.26%	31.17%
Command Execution	5504	2336	10.71%	7.32%
Information Disclosure	26138	9701	50.85%	30.42%
Logical Attacks	72	63	0.14%	0.20%

Black Box & White Box

T. 10 General statistics

Threat Classification	N of Vulns	N of Sites	% Vulns	% Sites
Abuse of Functionality	114	66	1.05%	7.99%
Brute Force	148	66	1.37%	7.99%
Buffer Overflow	1	1	0.01%	0.12%
Content Spoofing	646	152	5.96%	18.40%
Credential/Session Prediction	25	10	0.23%	1.21%
Cross-site request forgery	74	13	0.68%	1.57%
Cross-site Scripting	6418	480	59.26%	58.11%
Denial of Service	25	22	0.23%	2.66%
Directory Indexing	60	39	0.55%	4.72%
Fingerprinting	80	51	0.74%	6.17%
Format String Attack	4	2	0.04%	0.24%
HTTP Response Splitting	447	64	4.13%	7.75%
Information Leakage	639	339	5.90%	41.04%
Insufficient Anti-automation	16	14	0.15%	1.69%
Insufficient Authentication	204	107	1.88%	12.95%
Insufficient Authorization	187	117	1.73%	14.16%
Insufficient Process Validation	77	18	0.71%	2.18%
Insufficient Session Expiration	17	16	0.16%	1.94%
LDAP Injection	1	1	0.01%	0.12%
OS Commanding	6	6	0.06%	0.73%
Path Traversal	60	29	0.55%	3.51%
Predictable Resource Location	499	121	4.61%	14.65%
Session Fixation	120	22	1.11%	2.66%
SQL Injection	879	209	8.12%	25.30%
SSI Injection	5	3	0.05%	0.36%
URL Redirectors	11	11	0.10%	1.33%
Weak Password Recovery Validation	13	10	0.12%	1.21%
WSDL Exposure	0	0	0.00%	0.00%
XPath Injection	55	15	0.51%	1.82%
Total	10831	826		

T. 11 Vulnerabilities distribution by risk

Threat rank	N of Vulns	N of Sites	% Vulns	% Sites
Low	452	408	4.17%	49.39%
Med	1549	584	14.30%	70.70%
High	7127	800	65.80%	96.85%

T. 12 Vulnerabilities distribution by WASC TCv1 classes

WASC Classes	N of Vulns	N of Sites	% of Vulns	WASC Classes
Authentication	365	172	3.37%	20.82%
Authorization	349	157	3.22%	19.01%
Client-side Attacks	7596	573	70.13%	69.37%
Command Execution	951	230	8.78%	27.85%
Information Disclosure	1338	467	12.35%	56.54%
Logical Attacks	232	115	2.14%	13.92%

Participation

If you represent an organization that performs vulnerability assessments on websites, particular in those in custom web applications, through a manual or automated process and would like to participate please let us know. Once statistics are compiled, a report will be distributed, and all contributors will receive a logo on the project pages as well as on other deliverables in appreciate of their contribution. Please contact Sergey Gordeychik. Statistics will be collected once per year one month after December 31.

License

Terms and Conditions for Copying, Distributing, and Modifying Items other than copying, distributing, and modifying the Content with which this license was distributed (such as using, etc.) are outside the scope of this license.

1. You may copy and distribute exact replicas of the OpenContent (OC) as you receive it, in any medium, provided that you conspicuously and appropriately publish on each copy an appropriate copyright notice and disclaimer of warranty; keep intact all the notices that refer to this License and to the absence of any warranty; and give any other recipients of the OC a copy of this License along with the OC. You may at your option charge a fee for the media and/or handling involved in creating a unique copy of the OC for use offline, you may at your option offer instructional support for the OC in exchange for a fee, or you may at your option offer warranty in exchange for a fee. You may not charge a fee for the OC itself. You may not charge a fee for the sole service of providing access to and/or use of the OC via a network (e.g. the Internet), whether it be via the world wide web, FTP, or any other method.

2. You may modify your copy or copies of the OpenContent or any portion of it, thus forming works based on the Content, and distribute such modifications or work under the terms of Section 1 above, provided that you also meet all of these conditions:
a) You must cause the modified content to carry prominent notices stating that you changed it, the exact nature and content of the changes, and the date of any change.
b) You must cause any work that you distribute or publish, that in whole or in part contains or is derived from the OC or any part thereof, to be licensed as a whole at no charge to all third parties under the terms of this License, unless otherwise permitted under applicable Fair Use law.
These requirements apply to the modified work as a whole. If identifiable sections of that work are not derived from the OC, and can be reasonably considered independent and separate works in themselves, then this License, and its terms, do not apply to those sections when you distribute them as separate works. But when you distribute the same sections as part of a whole which is a work based on the OC, the distribution of the whole must be on the terms of this License, whose permissions for other licensees extend to the entire whole, and thus to each and every part regardless of who wrote it. Exceptions are made to this requirement to release modified works free of charge under this license only in compliance with Fair Use law where applicable.

3. You are not required to accept this License, since you have not signed it. However, nothing else grants you permission to copy, distribute or modify the OC. These actions are prohibited by law if you do not accept this License. Therefore, by distributing or translating the OC, or by deriving works herefrom, you indicate your acceptance of this License to do so, and all its terms and conditions for copying, distributing or translating the OC.

NO WARRANTY

4. BECAUSE THE OPENCONTENT (OC) IS LICENSED FREE OF CHARGE, THERE IS NO WARRANTY FOR THE OC, TO THE EXTENT PERMITTED BY APPLICABLE LAW. EXCEPT WHEN OTHERWISE STATED IN WRITING THE COPYRIGHT HOLDERS AND/OR OTHER PARTIES PROVIDE THE OC "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. THE ENTIRE RISK OF USE OF THE OC IS WITH YOU. SHOULD THE OC PROVE FAULTY, INACCURATE, OR OTHERWISE UNACCEPTABLE YOU ASSUME THE COST OF ALL NECESSARY REPAIR OR CORRECTION.
5. IN NO EVENT UNLESS REQUIRED BY APPLICABLE LAW OR AGREED TO IN WRITING WILL ANY COPYRIGHT HOLDER, OR ANY OTHER PARTY WHO MAY MIRROR AND/OR REDISTRIBUTE THE OC AS PERMITTED ABOVE, BE LIABLE TO YOU FOR DAMAGES, INCLUDING ANY GENERAL, SPECIAL, INCIDENTAL OR CONSEQUENTIAL DAMAGES ARISING OUT OF THE USE OR INABILITY TO USE THE OC, EVEN IF SUCH HOLDER OR OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.