Web vulnerability in 2021: large scale inspection, findings, analysis and remedies

Prof. Borka Jerman Blažič
International postgraduate school Jožef Stefan and Institute Jožef Stefan
Ljubljana, Slovenia
The web site presence and content design

• Since its appearance in the early 1990s, the World Wide Web has evolved from a platform to access text and other media to a framework for running complex web applications.

• These applications appear in many forms, from small home made to large scale commercial services (Google Docs, Twitter, Facebook)

• The system that enable and manage the web site content is known as Web Content Management System

• WordPress is a free open-source content management system based on PHP scripting language and MySQL database.

• Word Press is easy to be learnt and used. It allows users to create websites, even without knowing the HTML or PHP language, which means that also suitable for the beginners.
Word Press is the most popular WCMS system on the Internet

1. WordPress (WP) appeared on the web in 2003, when it was introduced to the public as a simple blogging platform where users could write text, post images and link to other websites.

2. WP is due to its ease of use is one of the most widespread WCMS.

3. The popularity of WP lays in the easiness in setting web pages for particular website, the low cost of use and maintenance.

4. The abundance of plug-ins for developing very different type of services and scenarios, like blogs, social network applications, webmail service, banking, e-commerce, educational services contribute to its popularity.

5. WordPress is applied for the creation of websites of companies like NBC, CNN, TED, New York Times, Forbes, eBay, Best Buy, Sonny, UPS, CBS Radio, TechCrunch and others
Some differences regarding Possibilities of an attack among the applications provided on the client side (the browser) and the server side of the Web service

The attacks and the vulnerabilities are happening mainly on the server side.

The server, the WCMS and associated plug-ins are the focus of the web attacks

<table>
<thead>
<tr>
<th>Basis of Comparison</th>
<th>Client-side scripting</th>
<th>Server-side scripting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Use</strong></td>
<td>Works on the frontend and is visible among users</td>
<td>It works in the backend, which cannot be seen at the end of the client</td>
</tr>
<tr>
<td><strong>Processing</strong></td>
<td>Does not need interaction with the server</td>
<td>Requires server interaction</td>
</tr>
<tr>
<td><strong>Supported Languages</strong></td>
<td>HTML, CSS, JavaScript, etc.</td>
<td>PHP, Python, etc.</td>
</tr>
<tr>
<td><strong>Running</strong></td>
<td>It runs on user’s computer</td>
<td>It runs on web server</td>
</tr>
<tr>
<td><strong>Execution</strong></td>
<td>The scripting process for the server side is done on remote computer and hence the response is comparatively slower than the client side one</td>
<td>The scripting process of client server is executed on a local computer and thus the response is comparatively quicker when compared to server-side scripting</td>
</tr>
<tr>
<td><strong>Database Connection</strong></td>
<td>Does not connect to the databases</td>
<td>Connects to the databases that are already present in the server</td>
</tr>
<tr>
<td><strong>Access to Files</strong></td>
<td>No access to all the files on the server</td>
<td>Access to all the files on the server</td>
</tr>
<tr>
<td><strong>Source Code</strong></td>
<td>Source code is visible to user</td>
<td>Source code is not visible to user</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Less secure as the scripts are usually not hidden from the client end</td>
<td>Relatively secure, but more secure than client-side scripting as the server-side scripts are usually hidden from the client end</td>
</tr>
</tbody>
</table>

Table 1: Client-side and server-side comparison.
More Than 162,000 WordPress Sites Used for Distributed Denial of Service Attack
Crawl:
Crawl entire web, optimize crawler performance and stability.

Collect Data:
Collect as much data as possible.

Awareness:
Prediction model based on collected results.

Secure:
Notify national Certs about potential security risks.
Web site security can be inspected on two different ways: Open source and Close source.

Other types of inspection are dynamic (going with crawlers) and static where source code is required to be provided by the owner.

Top vulnerabilities of a web site according to the Open Web Application Security Project:

<table>
<thead>
<tr>
<th>ID</th>
<th>Vulnerability</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Injection</td>
</tr>
<tr>
<td>A2</td>
<td>Broken Authentication and Session Management</td>
</tr>
<tr>
<td>A3</td>
<td>Sensitive Data Exposure</td>
</tr>
<tr>
<td>A4</td>
<td>XML External Entities (XXE)</td>
</tr>
<tr>
<td>A5</td>
<td>Broken Access Control</td>
</tr>
<tr>
<td>A6</td>
<td>Security Misconfiguration</td>
</tr>
<tr>
<td>A7</td>
<td>Cross-Site Scripting XSS</td>
</tr>
<tr>
<td>A8</td>
<td>Insecure Deserialization</td>
</tr>
<tr>
<td>A9</td>
<td>Using Components with Known Vulnerabilities</td>
</tr>
<tr>
<td>A10</td>
<td>Insufficient Logging &amp; Monitoring</td>
</tr>
</tbody>
</table>
Crawlers on the Internet

- Use of crawling tools is very popular, but they are not always welcomed
- They are accepted but as well denied to access to web site by the owners
Some owners is hiding the version of the used SW release, with an idea that this protect the web site from an attack

Services for crawling the internet 24/7 in order to provide the latest Internet Intelligence, are located all over the world.

Services for crawling are used around the world by researchers, security professionals, large enterprises, CERTs and everybody in between.
They are used mainly in inspecting servers instead of the content they like WMCS and the applied data
How hacked web sites are compromised

Parameter Manipulation
Malicious File Upload
Authentication Bypass
SQL Injection
Elevation of Privilege
Spam
Session Hijacking

How Hacked WordPress Sites Were Compromised

- plugin
- brute force
- core
- theme
- hosting
- file permissions
- old files
- password theft
- workstation
- phishing
- insider
- server
- ftp

0.0% 10.0% 20.0% 30.0% 40.0% 50.0% 60.0%
Security scanning tools and their functionality

- Shodan
- Censys
- IVRE
- Tools that support large-scale scan
  - All use Port Scanners
- Port Scanners
  - ZMap
  - Masscan
- Custom Solutions
- Tools that support detailed scan
  - Nmap
  - Nmap extension supports vulnerability assessment
- Nmap
- Nessus
- Skipfish
- WPScan
Comparison of the most known vulnerability scanning tools with crawlers

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>General Characteristics</td>
<td>OpenSource</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes (A)</td>
</tr>
<tr>
<td>URL or IP</td>
<td>IP</td>
<td>IP</td>
<td>URL</td>
<td>IP</td>
<td>URL</td>
<td>URL</td>
<td>URL</td>
<td>Both</td>
</tr>
<tr>
<td>Results are Freely Accessible on the Internet</td>
<td>Yes (P)</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (E)</td>
</tr>
<tr>
<td>Internet-connected Devices</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Automatic Scanning</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Ethical</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes (P)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Web UI and Command Line (CL)</td>
<td>Both</td>
<td>Both</td>
<td>CL</td>
<td>CL</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Both</td>
</tr>
<tr>
<td>Real-time Visualization while Scanning</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Free API</td>
<td>Yes (P)</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>More than 1 million scanned IPs or Websites</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

WP Specific

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CMS Scan</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CVE Exposer</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Plugins Scan</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes (L)</td>
</tr>
<tr>
<td>Scoring</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

P = Partly, E = Extension, A = Attended, L = Limited.
The system and component scheme of the new developed tool at IJS for large scanning WP web sites over the Internet – VULNET enables fast and ethical scanning at large
WEB site presence on the internet in Numbers found by VULNET in the first large scan of the Internet

- **126 Million+**: There were over 126 millions scanned websites (194 countries).
- **16 MILLION**: WordPress Websites
- **5 MILLION**: More than 5 millionion web sites had vulnerability, with a score of 5 or more
Owners and the hacked compromised WP web sites

From the 1,032 survey respondents who answered to the submitted questions about their own web site vulnerability awareness after the successful attack, 61.5% didn’t know how the attacker compromised their website!

- Plugins: Plugins are the biggest risk. There are cca 50,000 plugins available.
- Theme
- Brute Force Attacks: A brute force attack is a password guessing attack.
- File permissions
- Core
The measured vulnerability
of one affected domain
with VULNET tool
## Comparison of known scanning tools and VULNET according to properties

### The table legend:
- ‘•••’ is used to denote a strong support,
- ‘••’ to denote a moderate support,
- and ‘•’ for a weak support (i.e., unavailable) of a specific feature.

CPE, Common Platform Enumeration, CVE, Common Vulnerability and Exposure.
What was studied and what was found

• The aim of the analysis was to find out the percentage of insecure WP websites among the WP websites found.

• The level of Digital skills to be compared with the appearance of web insecurity

• The cost of fixed access normalized with GNI was the second parameter studied as impact factor that affects the appearance of insecure websites.

• The subset of studied countries was composed from: Germany (DE), Netherlands (NL), France (FR), Great Britain (GB), Italy (IT), Denmark (DK), Poland (PL), Spain (ES), Sweden (SE), Switzerland (CH), Czech Republic (CZ), Ireland (IE), Finland (FI), Austria (AT), Romania (RO), Belgium (BE), Hungary (HU), Bulgaria (BG), Norway (NO), Slovakia (SK), Estonia (EE), Slovenia (SI), Portugal (PT), Croatia (HR), Lithuania (LV), Luxembourg (LU), Greece (GR), Iceland (IS), Latvia (LT), Cyprus (CY), and Malta (MT)
Security state in the EU WP web space

The web space with domains representing 28 members states and three other (Switzerland, Norway) was inspected for discovering the presence of insecure and secure web sites with WP core and added plug-ins. Web sites that did not provided information of WP core version were classified as unknown, the others with score higher than 5 from a scale of 10 were classified as critically insecure and the other with identified vulnerability but having lower insecure score were classified as insecure.

The top five ranked countries with highest numbers of vulnerability and out-dated plug-ins were found in the general internet at large scan to be: USA, France, Germany, Netherlands and Japan (this is related also to the highest numbers of present web sites.)
• On average, we could not determine the vulnerability status in the first scan for a third of the sites in a domain, due to unknown (hidden) core or plug-in versions. Then the second scan with 31 countries gave more significant result.

• The percentages of secure sites were between 21% and 40%, with an average of 28%. The percentage of insecure sites is in the range of 30% and 47%, with a mean of 38%, a median of 40%, and a standard deviation of 4.4%. There were no univariate outliers in either variable.

Table 1. Summary statistics of all the WP sites found, the vulnerability percentages, and the digital-skills (DS) index.

<table>
<thead>
<tr>
<th></th>
<th>total WP</th>
<th>unknown [%]</th>
<th>secure [%]</th>
<th>insecure [%]</th>
<th>score [%]</th>
<th>DS</th>
</tr>
</thead>
<tbody>
<tr>
<td>count</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
<td>31</td>
</tr>
<tr>
<td>Mean</td>
<td>120712.19</td>
<td>33.56</td>
<td>28</td>
<td>38.43</td>
<td>31.27</td>
<td>59.06</td>
</tr>
<tr>
<td>Std</td>
<td>182007.02</td>
<td>4.3</td>
<td>5.08</td>
<td>4.44</td>
<td>4.76</td>
<td>14.89</td>
</tr>
<tr>
<td>Min</td>
<td>10</td>
<td>20</td>
<td>21.22</td>
<td>30.31</td>
<td>22.13</td>
<td>29</td>
</tr>
<tr>
<td>25%</td>
<td>14715</td>
<td>31</td>
<td>24.28</td>
<td>35.5</td>
<td>27.27</td>
<td>49</td>
</tr>
<tr>
<td>50%</td>
<td>49345</td>
<td>33.57</td>
<td>26.12</td>
<td>39.65</td>
<td>32.24</td>
<td>57</td>
</tr>
<tr>
<td>75%</td>
<td>128049</td>
<td>36.86</td>
<td>31.4</td>
<td>41.59</td>
<td>34.9</td>
<td>71</td>
</tr>
<tr>
<td>Max</td>
<td>805361</td>
<td>40.15</td>
<td>40</td>
<td>46.96</td>
<td>40.83</td>
<td>85</td>
</tr>
</tbody>
</table>
The correlation between insecure and secure sites in a country shows that the countries with well developed digital skills (high percentage of DS) among the population have low percentage of insecure sites. The group is consisted from SE, CH, NL, DE, NO, DK, FI.

The insecurity among plug-ins was found to be the factor that has most influence on the site to be insecure.
The study about appearance of insecure sites and the Digital skills among the country population has shown that those countries that have high percentage of population with digital skills have the lowest percentage of Insecure web sites. This group is consisted from the countries SE, CH, NL, DE, NO, DK, FI.

The study as well has shown that countries with low cost of fixed internet access have also low percentage of insecure web sites.
VulNet service offered to the public a service for checking the website vulnerability in safe and privacy protected manner.
CONCLUSION

• Vulnerability presence is still a risk issue for the security of the web space
• Improvement were noticed with the third large VULNET scan that happened 6 months later due to the replacement of old WP versions with the new releases of the web core system
• The best remedy was found to be continuous checking of the insecurity presence in the web sites by the maintainers and owners with a public service that guarantee the owner privacy