Web Tech Report 2011

Technologies in use by the worlds leading web sites

HackerTarget.com LLC
Everyone is a Target

peter <at> hackertarget.com
http://hackertarget.com

Table of Contents

Web Tech Report

Technologies in use by the worlds leading web sites

Introduction

A handful of technologies power the worlds leading websites. These technologies are the bridge between end users and information. They are also the bridge between attackers and the data.

This report aims to provide a snap shot of technology that is currently used by the worlds leading websites, based on the Alexa 1 million top sites.

Weak technology decisions, poorly managed patching and systems management deficiencies allow attackers easy access to back-end databases and systems.

Not paying attention will lead to customers, employees and business all put at risk.

About

HackerTarget.com is a leading provider of on-line security scanning services. Utilizing open source tools, advanced security testing is made available to anyone wanting to test their external facing Internet Services for security vulnerabilities or other issues.

Port Scanning, Vulnerability Testing, Web Server analysis, SQL Injection, CMS fingerprinting and open source intelligence gathering are the core automated tools.

All scan options are available for free (limited to 4 / day), additional scans are available for a minimal cost.

In depth security assessment consulting services are also available.

Methodology

Alexa (owned by Amazon) provides web metrics on the worlds top websites.

During February 2011 HackerTarget.com downloaded the list of the top 1 million sites as ranked by Alexa and proceeded to spider the root web page of each of those sites using the fingerprinting tool WhatWeb.

Redirects from sites that responded with a 302 were followed and a total of 993692 “HTTP 200 OK” pages were analyzed.

Intensity of the analysis was set to a minor level, meaning the only active scanning of the target sites was to download the HTML from the page and examine the resulting code and HTTP headers.

The data in this report is based on these sites that responded with a 200 OK HTTP Response. No attempt was made to access sub-domains or subdirectories to include additional blogs / forums and other minor parts of the site. Blogs and Forum statistics are only from sites that are based on those systems from the root page.

The nature of this data means that there is no way this report can be 100% accurate. Server administrators can hide and alter these responses for security reasons. HackerTarget.com LLC makes no guarantee on the accuracy of this report.

Generally technologies with less than 100 sites detected have not been included in the results (0.01%).

External References

http://hackertarget.com

http://en.wikipedia.org/wiki/List_of_HTTP_status_codes

http://www.morningstarsecurity.com/research/whatweb

http://www.alexa.com

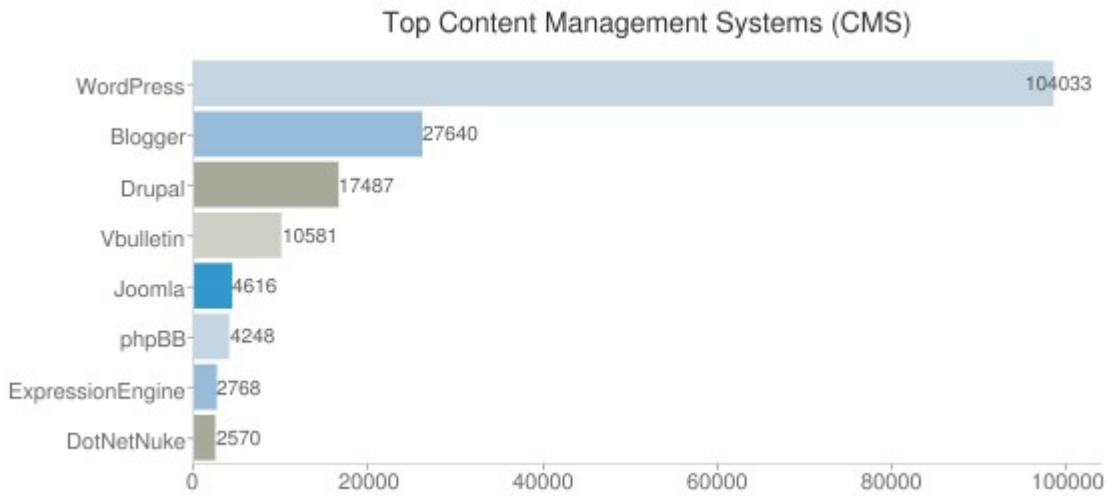
Top Content Management Systems (CMS)

Content Management System (CMS) data has been broken up into two tiers for clarity.

WordPress is by far the most popular web content framework. This was detected using the plugin in the Whatweb package, this plugin accurately determines the presence of WordPress using a number of factors. The count in the chart below does not include the sites where WhatWeb determined that WordPress is “probably in use”. If these numbers are included the total for WordPress is 152814.

We can accurately conclude that the number of systems using WordPress in the Alexa top 1 million is 10% – 15%. A significant amount and well ahead of the next competitor Google owned Blogger.

Of the Top 8 systems the only Microsoft Windows (ASP.NET) based system is DotNetNuke in 8th place.



External References

http://en.wikipedia.org/wiki/Apache_HTTP_Server

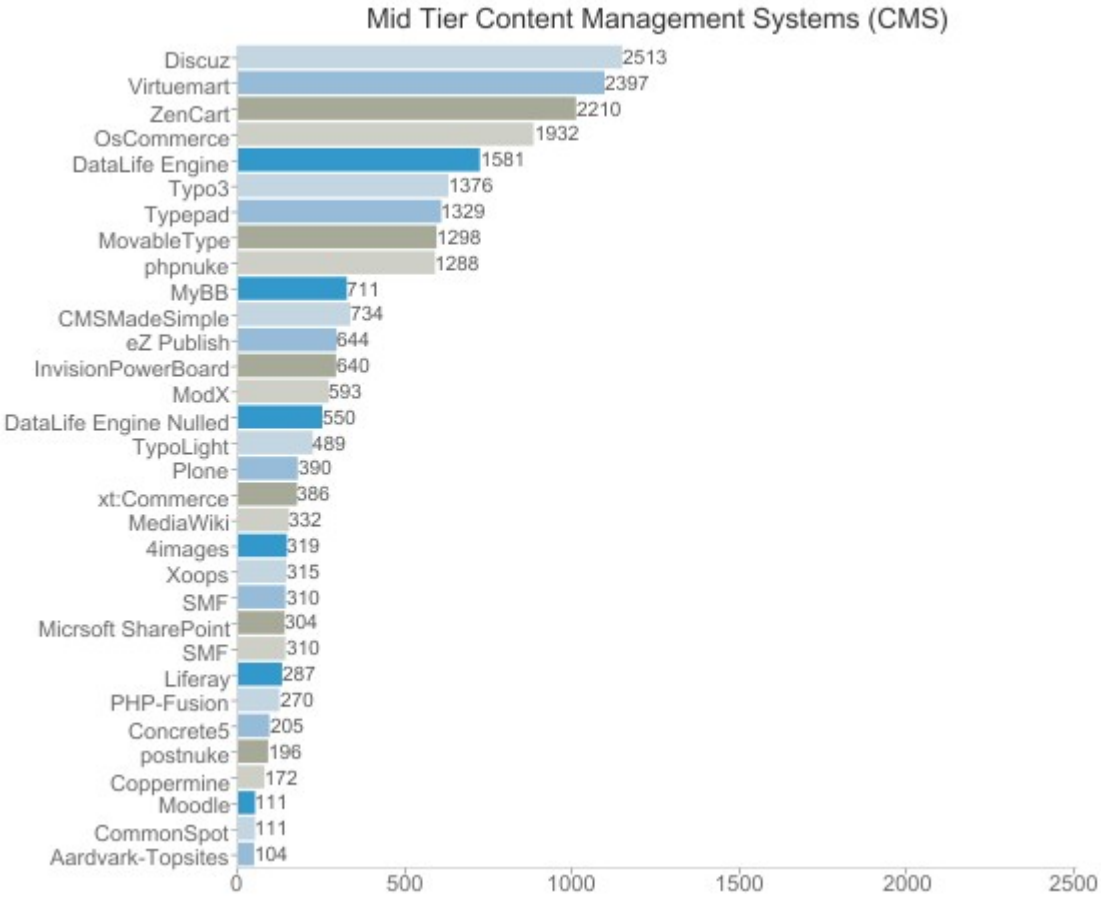
http://news.netcraft.com/archives/category/web-server-survey/

Middle Tier Content Management Systems (CMS)

This data has been collected primarily from the WhatWeb tool output, for details on the system detection requirements please see the plugins that are a part of the WhatWeb tool.

Tip: Any content management system should be updated and monitored, as exploitation is often the first step into the entire organization being compromised.

While it is a simple matter to get a CMS up and running, production sites should be “Security Hardened” using easily obtained security guides.



Note

The term “Mid Tier” is only indicative of the popularity within the Alexa Top 1 million. There is no correlation with features. The list of content management systems above range from full scale back-end systems to lightweight forums.

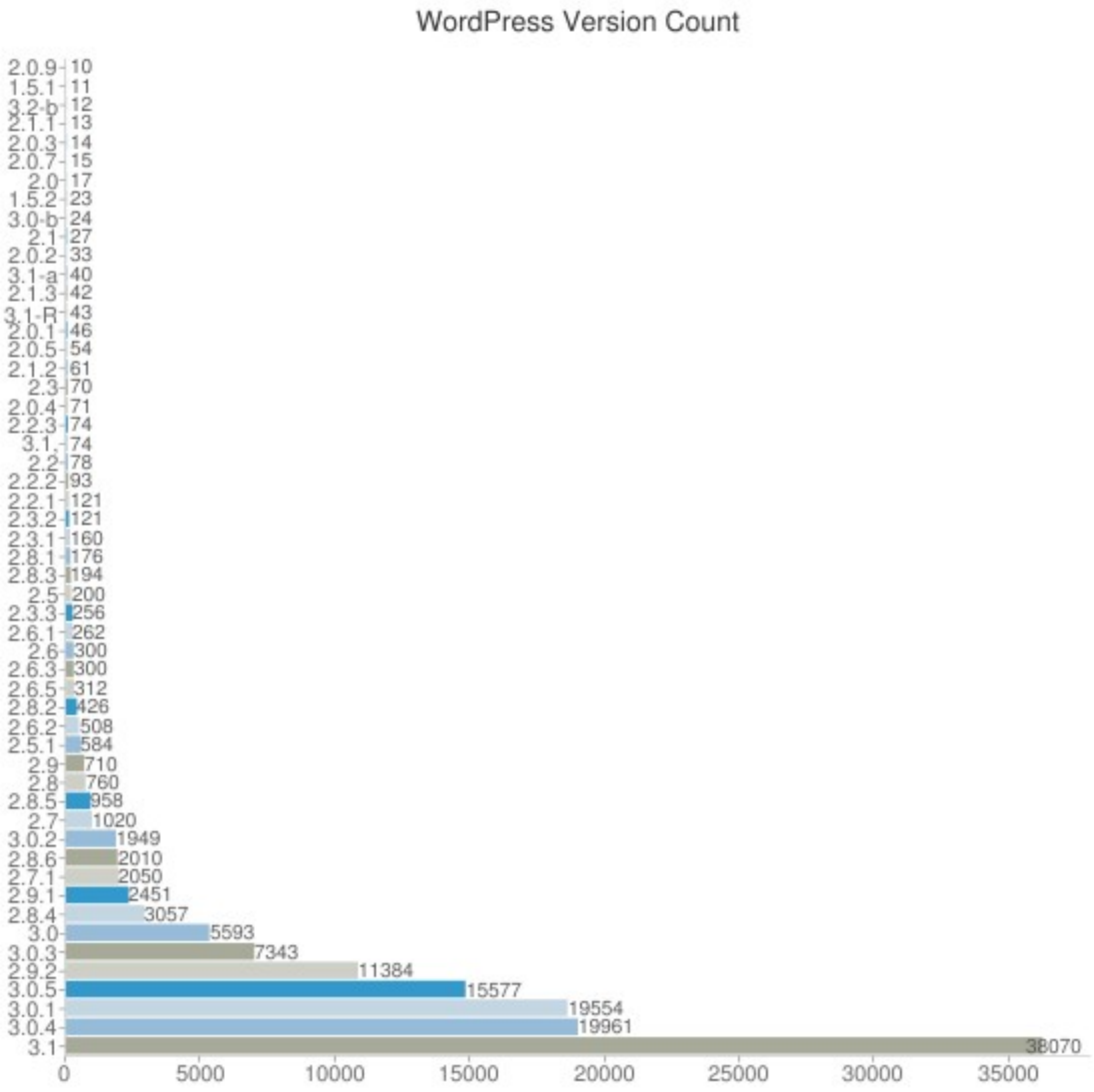
External References

http://en.wikipedia.org/wiki/List_of_content_management_systems

WordPress Versions

WordPress is the most popular CMS and as can be seen below critical security updates are not always applied when released.

WordPress security improvements have included alerting of administrators when logged into the control panel that new versions of both the core system and plugins are available. For some this has not helped in keeping the system up to date!



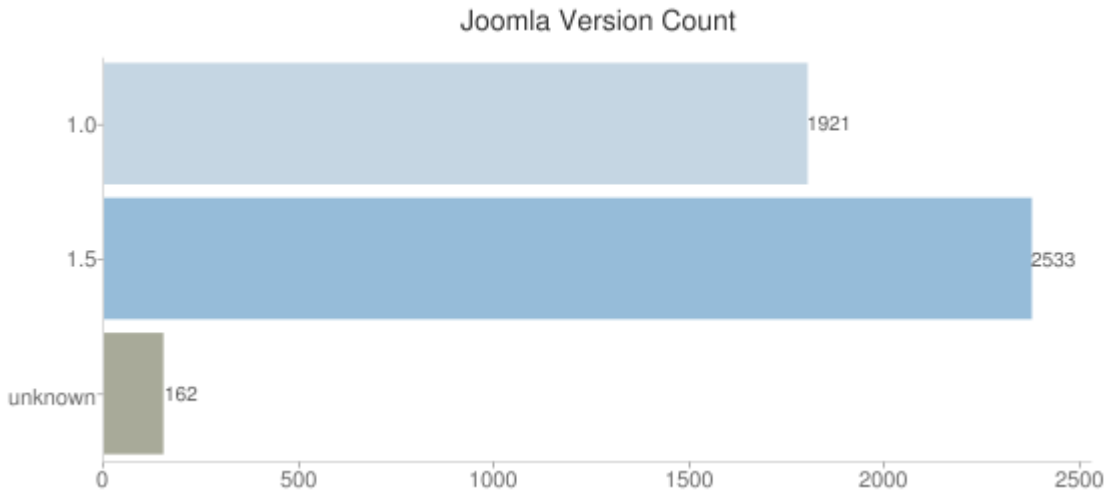
External References

Exploit DB Word Press Search

http://wordpress.org/news/category/security/

Joomla Versions

Joomla is a popular CMS with a large community and third party plugin base. This has in the past been both a blessing and a curse as many of the Joomla exploits that have been published are due to poor coding in the plugin rather than the core system.



External References

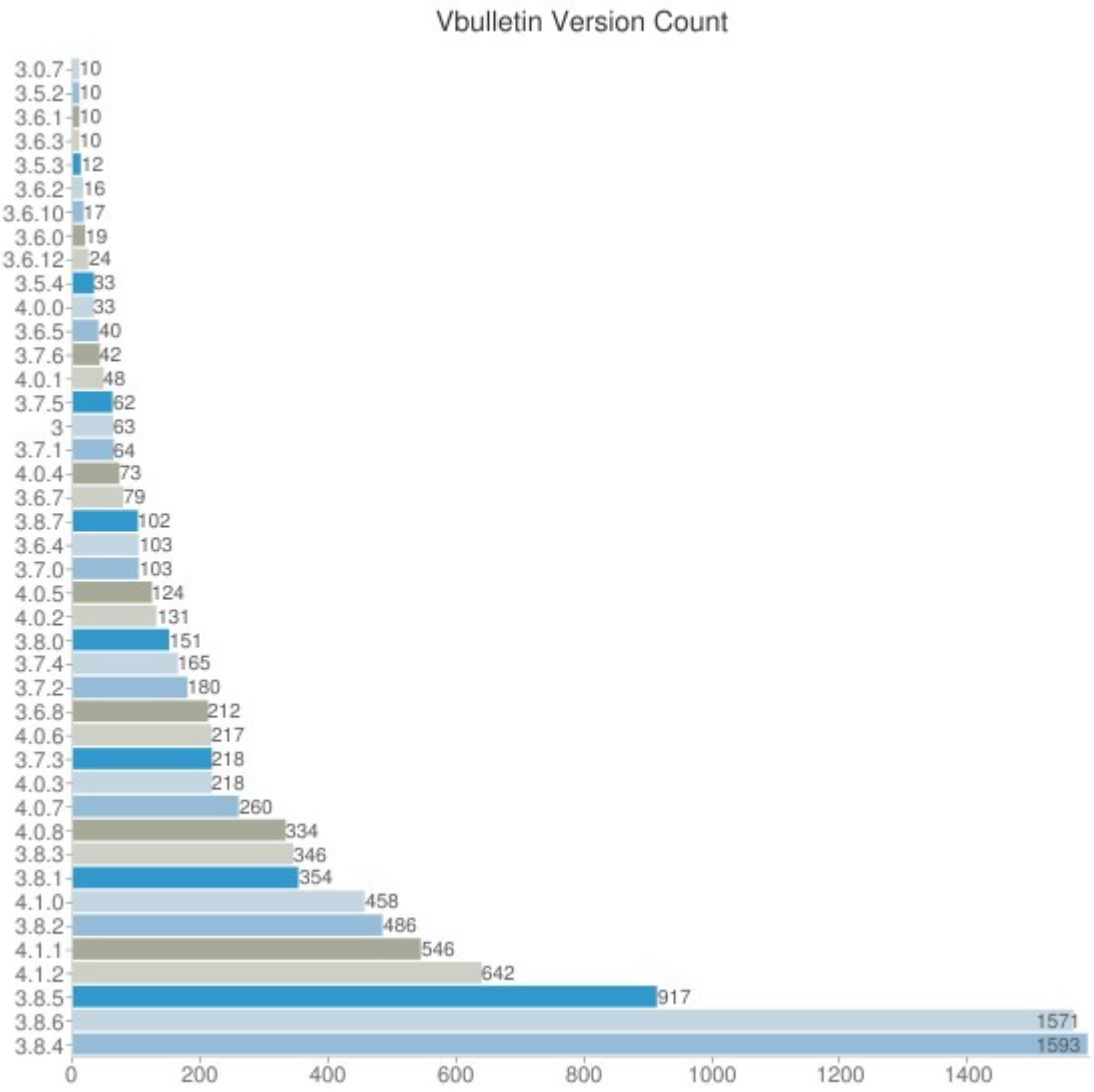
Exploit DB Joomla Search

http://docs.joomla.org/Category:Security_Checklist

Vbulletin Versions

vBulletin is a popular forum system that can also be used as a content management system. It is a commercial product.

Forums and CMS are both popular targets for attackers with many published exploits available. See the exploit-db link below for a sample of exploits. Compare that to the versions still in use and it is clear that vBulletin administrators are often as lax as their WordPress counterparts.



External References

Exploit-DB search for vBulletin

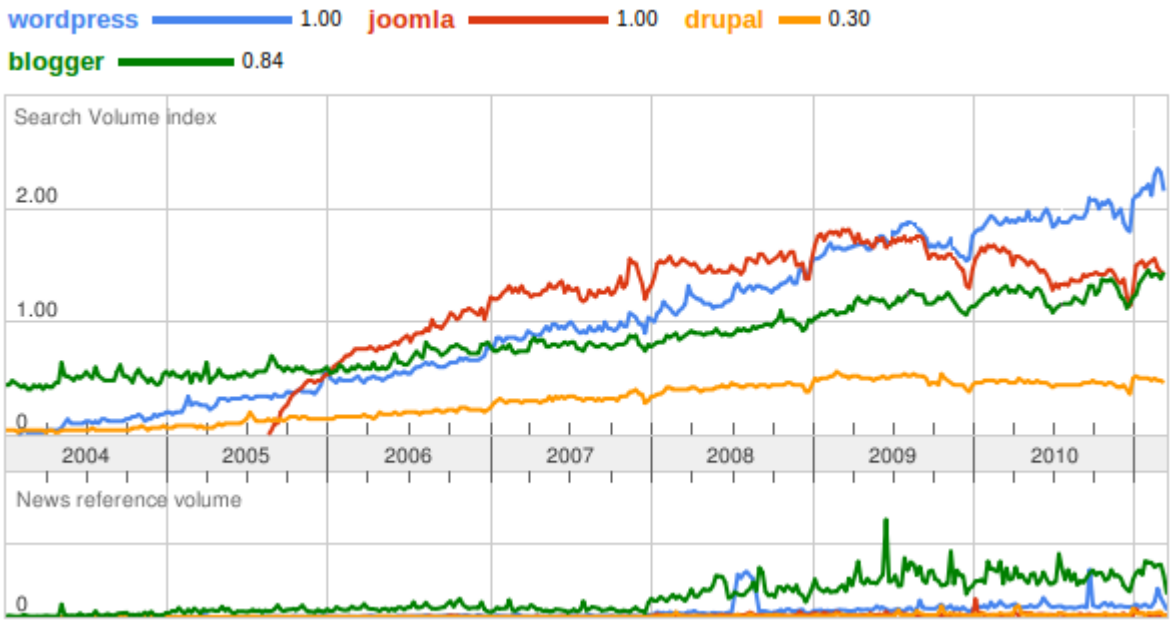
http://www.vbulletin.com/faq.php

CMS Trend Analysis

To get a view of the overall popularity of the top content management systems the following chart was created using Google Trends.

This data shows an increase to WordPress searches over the past 12 months while Joomla has steadily declined.

Drupal is more steady, however its high number in the Alexa Top 1 million (Top Content Management Systems Section) compared to search volume indicates it maybe more popular among established sites than in general use.



"Built With" Comparison

This table is comparing BuiltWith.com Top CMS results with the results from the HackerTarget.com Web Tech Report 2011.

	HackerTarget.com	BuiltWith.com
WordPress	10.50%	4.42%
Blogger	2.78%	0.1%
Drupal	1.76%	2.15%
Joomla	0.46%	0.48%

External References

http://www.google.com/trends

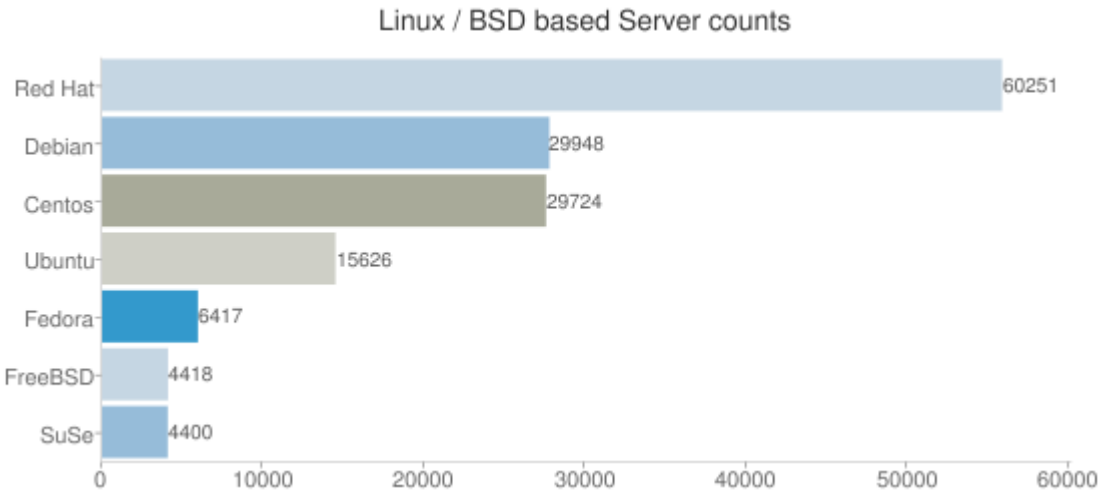
http://trends.builtwith.com/cms

Unix Based Servers

This data has been collected from the HTTPServer and X-Powered-By headers.

These should be taken as minimum counts for each distribution as systems with non-package based installs of web servers and those that have obfuscated the server strings will not be included in these counts.

Servers with Headers that indicated “Unix” in the response have also not been included as this is a generic term, that could includue Linux or other Unix based systems. .



Foot Notes

Red Hat data was counted using “Red Hat” or rhel strings.

Debian has been counted by including debian, etch, lenny, woody, sarge.

External References

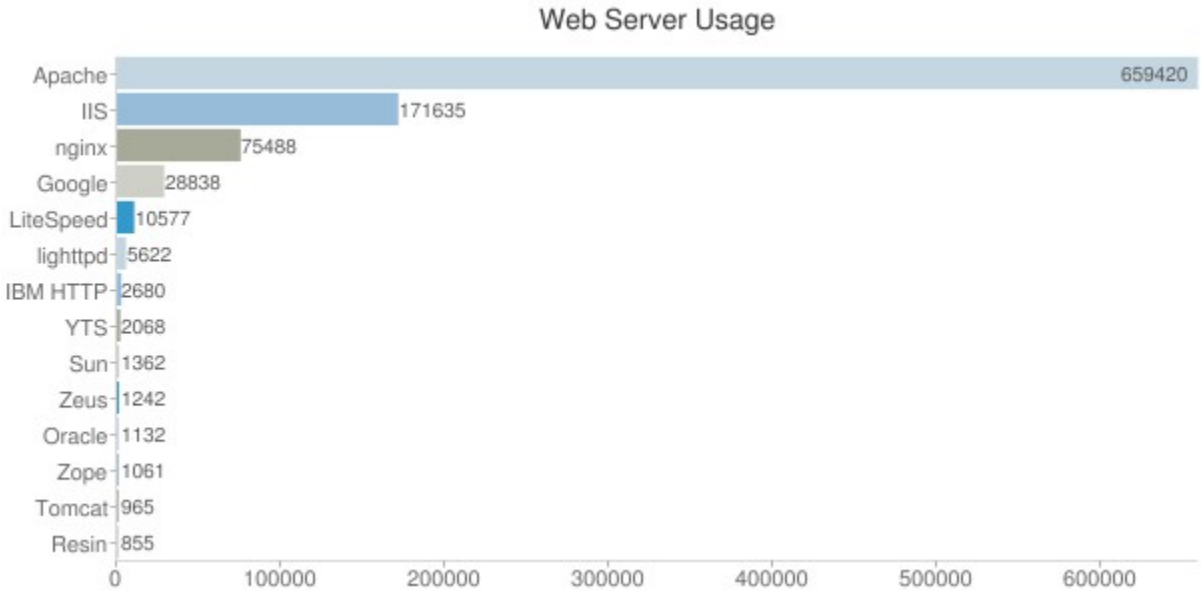
http://en.wikipedia.org/wiki/Usage_share_of_operating_systems

Web Servers

The web servers in use has been collated from HTTPServer header responses primarily however some servers such as Zope have also been counted from the X-Powered-By header.

As one would expect the Open Source Apache Server is by far the most popular server. It has held this position since April 1996 according to Wikipedia.

More information regarding web server popularity can be found at the Netcraft Web Server Survey page.



Foot Notes

The data collection tool did not receive a HTTPServer response from many "Blogger Based Sites", these have not been included in the above listing so one could surmise that the Google Web Servers number could be significantly higher (see the Top Web Content Frameworks for Blogger stats).

Google Web Servers include “Google Front End”, “GWS” and “GSE”

Sun includes; Sun One Web Server, Sun Java System Application Server.

Oracle includes mostly Oracle Application Server 10g and 11g; with a handful of other generic names.

* This chart was updated 7/4/11, error discovered in data parsing.

External References

http://en.wikipedia.org/wiki/Apache_HTTP_Server

http://news.netcraft.com/archives/category/web-server-survey/

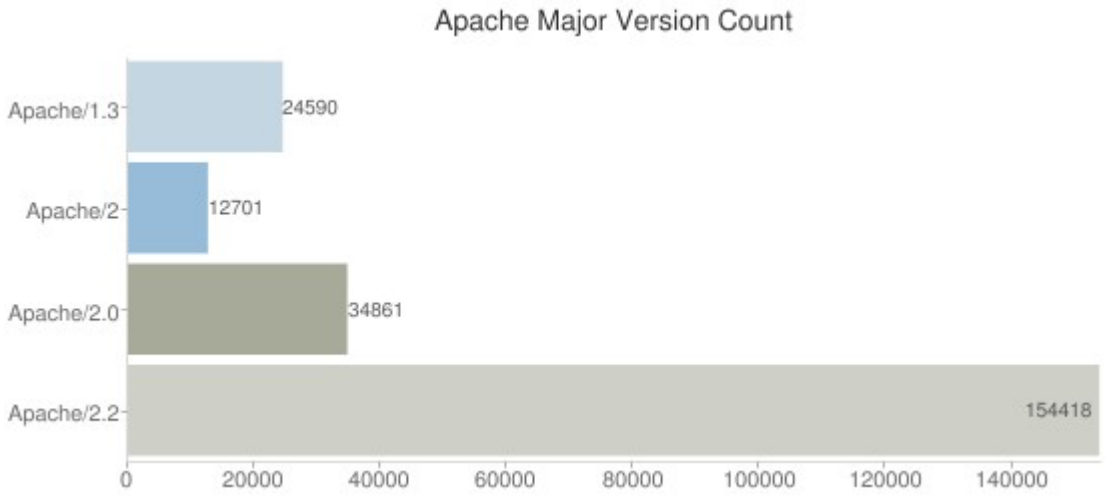
Apache Server Versions

This data has been collected from the HTTPServer Header.

While this is an indication of the Apache version distribution, the majority of Apache sites do not advertise the exact version number.

Servers running Apache in the Alexa Top 1 million number over 427000.

Apache version 1.3 is deprecated and no longer maintained.



Foot Notes

Apache has a large list of point releases for each of these versions.

External References

http://httpd.apache.org/ABOUT_APACHE.html

Microsoft IIS Server Versions

This data has been collected from the HTTPServer Header.

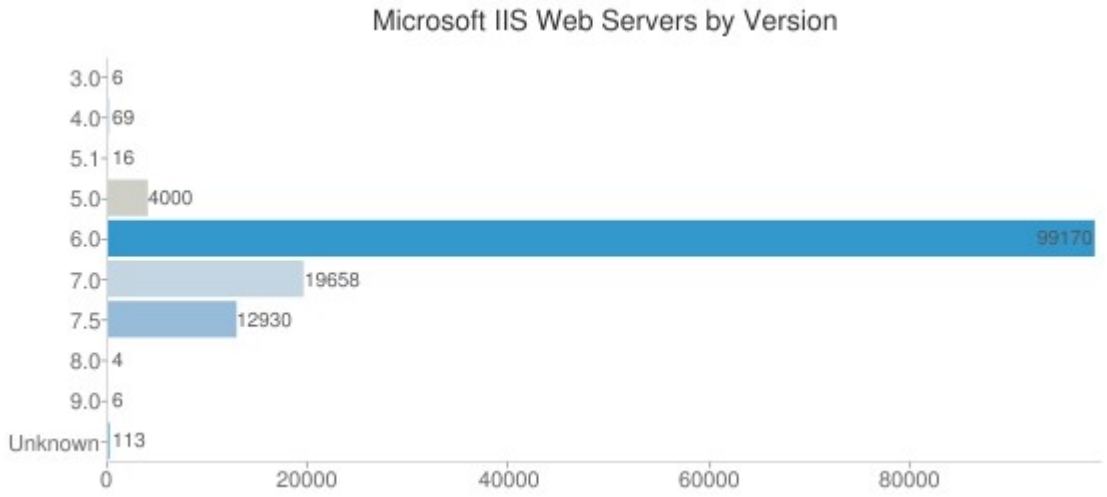
IIS 6.0 is clearly the most popular; I find it unlikely that anyone would still be running IIS 3.0, these results could have been manually changed for not only obfuscation but also amusement.

Note that if there are sites still running IIS 3.0 or 4.0 these sites will be running on Windows NT4.0.

IIS 5.0 will be running on Windows 2000 servers.

Generally IIS 6.0 was the first version released by Microsoft where Security started to become a priority.

IIS 8.0 and 9.0 have not been released so are also likely user edited server strings.



Foot Notes

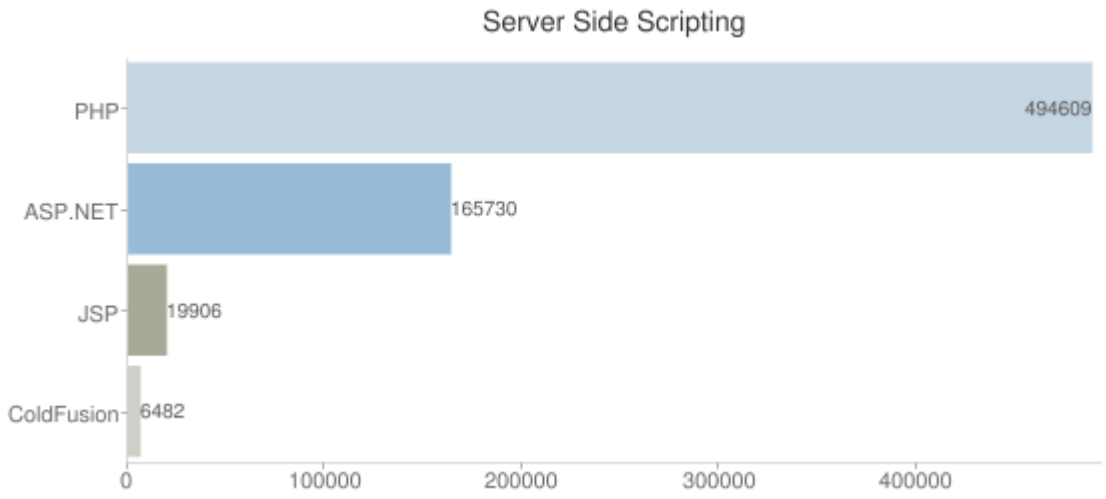
Unknown listing; a number of servers do not report the version however they do indicate they are IIS servers.

External References

http://en.wikipedia.org/wiki/Internet_Information_Services

Server Side Scripting

Numbers for PHP and ASP closely resemble the numbers for Apache and Microsoft IIS web servers.



Foot Notes

JSP (Java) pages have been detected using the detection of cookies containing JSESSIONID.

ColdFusion has been detected through the detection of CFID in the page cookie.

* This chart was updated 7/4/11, error discovered in data parsing.

External References

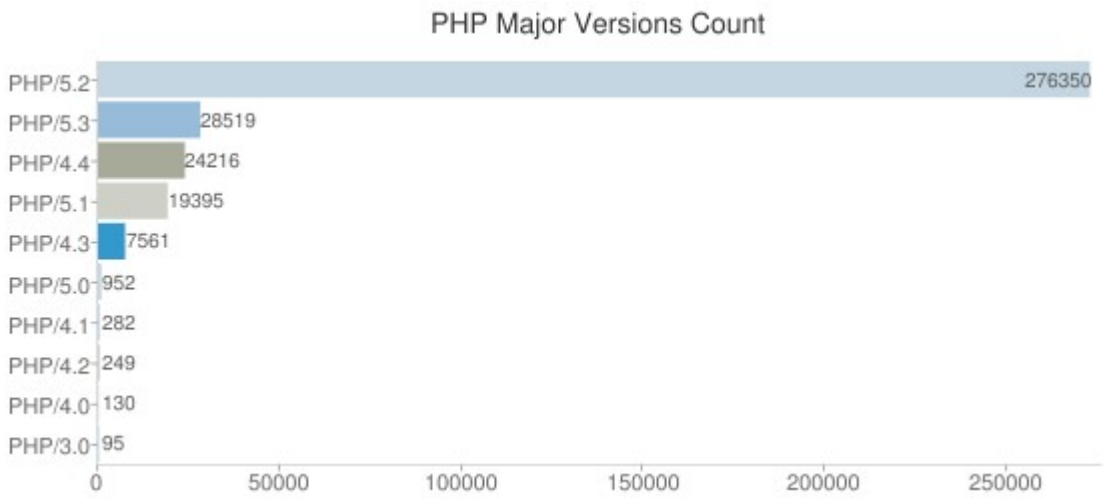
http://en.wikipedia.org/wiki/Apache_HTTP_Server

http://news.netcraft.com/archives/category/web-server-survey/

PHP Version

This data has been collected from the HTTPServer and X-Powered-By Header.

Only major versions have been counted. PHP version 3.0 was released in October 2000.

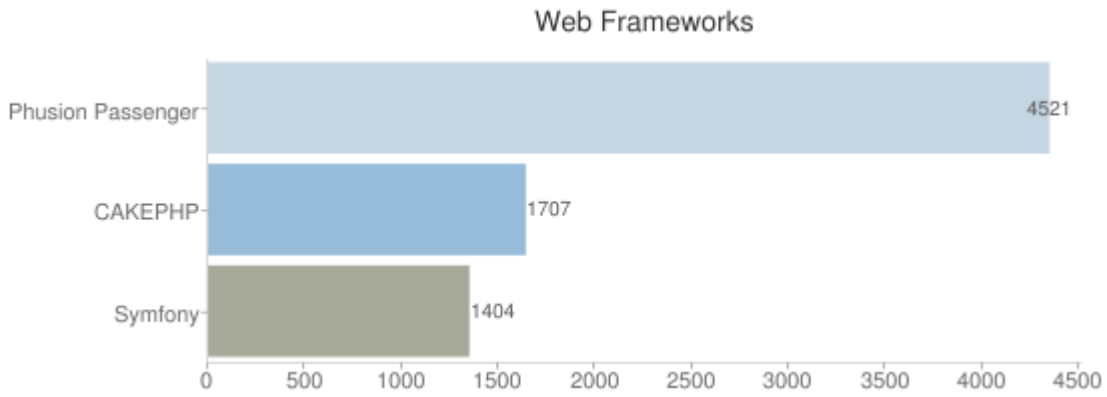


External References

http://php.net/releases/index.php

Web Frameworks

These three frameworks are easily detected hence they are the have been included in this report.



Foot Notes

Phusion Passenger (mod_rails) has been detected with the X-Powered-By header.

CakePHP has been detected using the cookie parameter (CAKEPHP)

Symfony has been detected using the cookie parameter (symfony)

External References

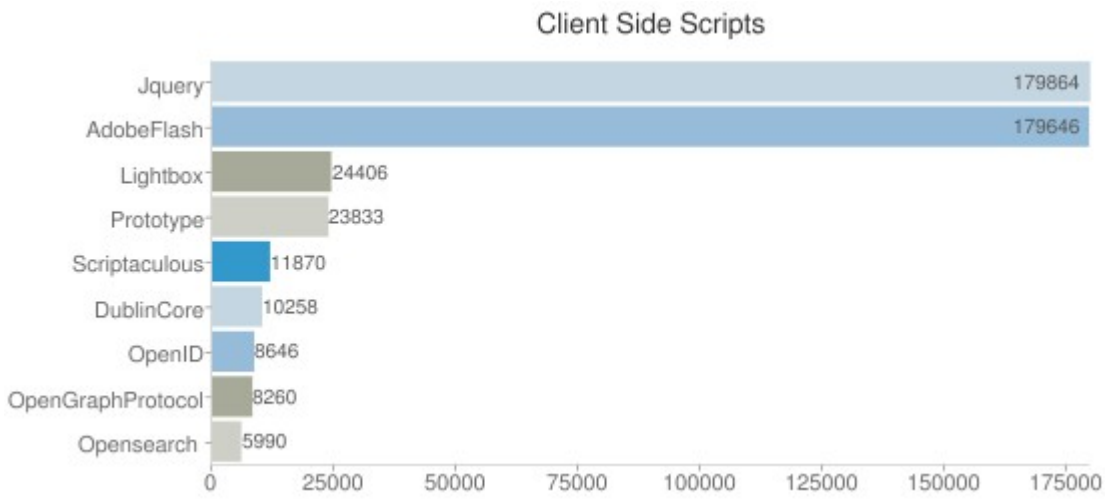
http://en.wikipedia.org/wiki/Apache_HTTP_Server

http://news.netcraft.com/archives/category/web-server-survey/

Client Side Scripting

Javascript libraries and Adobe Flash are both very popular.

HTML5 was also discovered on 22572 pages, it will be interesting to see how these numbers change once HTML5 usage becomes more widespread.



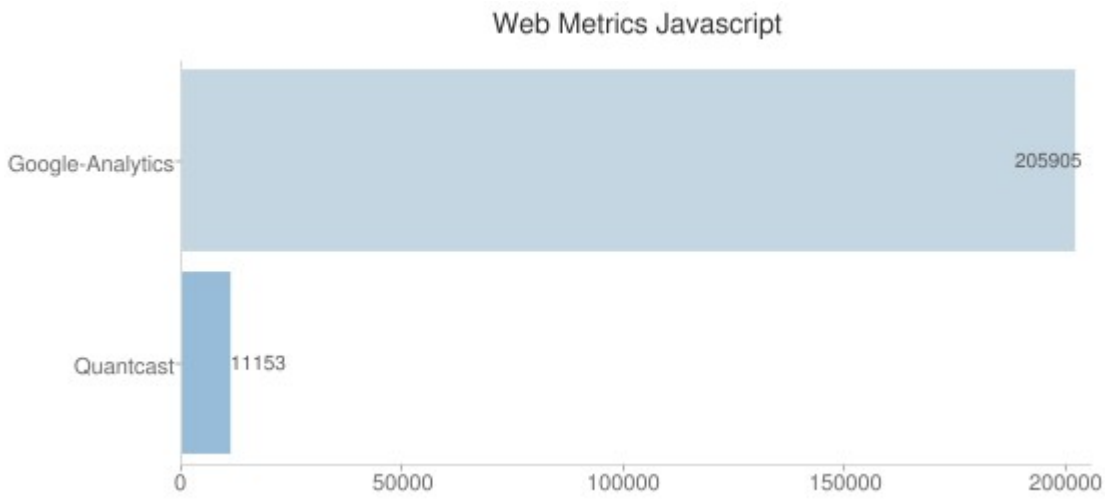
External References

http://en.wikipedia.org/wiki/JQuery

Web Metrics Javascript

Two well known web metrics services have been discovered during analysis.

Google Analytics is running on more than 1 in 5 of the worlds top websites.



External References

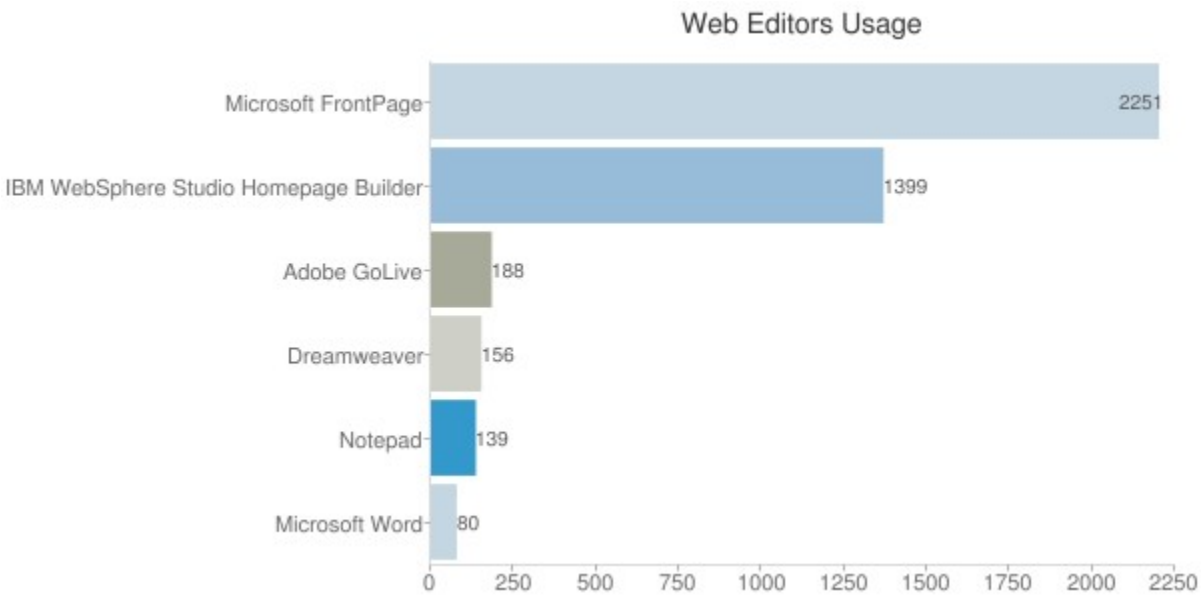
http://en.wikipedia.org/wiki/Apache_HTTP_Server

http://news.netcraft.com/archives/category/web-server-survey/

HTML Editors

Data has been collected from the MetaGenerator tag.

Dynamic websites and “Web 2.0” are all the rage, however there are still a couple of people using web editors – including the ever faithful NotePad and Microsoft Word!!



Foot Notes

The Microsoft FrontPage count is only including the MetaGenerator tags, does not include servers that have front page extensions installed.

Conclusion

When working to ensure the security of a web based environment it is essential that all aspects of the technology are addressed. Patching of operating system, web servers and content management systems are all required.

Content management systems are often the low hanging fruit, easy access and poor management practices combing to offer a juicy target to attackers.

Looking at the Content Management Application version analysis. It is not difficult to correlate published exploits to the advertised versions – allowing a potential attacker to be confident that there are thousands of sites in the Alexa Top 1 Million that are currently exploitable giving full database access or remote execution on the server.

One of many examples that can be found in the news over the past few years.

> "Here at SophosLabs we see hacked sites everyday and the majority are running PHP-driven applications such as Content Management Systems (CMS)," the blog post stated.
>
> PHP Nuke Infection Purged

This is not to say that PHP based content management systems are at fault, the ease of access, installation and management are factors that contribute to the popularity.

HackerTarget.com recommends the following:

- Ensure all software is up to date and patched including all web based applications such as content management systems (don't forget the plugins!).
- Use strong passwords on administrator accounts
- Limit password reuse between different accounts and environments
- If the application allows it move the administration panel url to an undisclosed location.
- Implement well documented web server lock down configurations.
- Perform regular security scanning as part of your overall security strategy.