

Blind Elephant:

Web Application Fingerprinting &
Vulnerability Inferencing

Patrick Thomas

Qualys

7/28/10



Outline



- Web Apps & Security
- Existing Fingerprinting Approaches
- Static File Approach
- Observations From A Net Survey
- Q & A

Well-Known Web Applications



- Every conceivable use...
 - Content Management/Blogging
 - Forums
 - Email
 - E-Commerce
 - DB Admin
 - Backup and File Storage Admin
 - Device/System/VM Admin
 - Version Control UI
 - Intranet/Collaboration

Well-Known Web Applications



Joomla!™



Special Challenges Securing Web Apps



- Remotely accessible by nature
- Lots of attack surface exposed (direct and indirect)
- Easy to set up and admin → Fly under IT radar

Special Challenges Securing Web Apps

- Fast release cycle (often open-source)
- Exploits are (often) simpler to create & comprehend
 - `wget http://example.com/wp-login.php?action=rp&key\[\]=`
 - `wget -header "Cookie: tinybrowser_lang=../../../../../../../../ZOMGSECRETS\r\n" http://example.com/plugins/editors/tinymce/jscripts/tiny_mce/plugins/tinybrowser/folders.php`
- (...and of course everything the WAF vendors are saying)

WAS Is Overkill For Well-Known Apps



- Known app + known-vulnerability list = traditional vulnerability management
- Knowing the version is good enough to infer vulnerabilities
 - It's not nearly as sexy, but it works
- Discovering the app and version → Fingerprinting

Existing Fingerprinting Approaches



- Labor intensive to add/update signatures
 - Manually locate version in files or build regexes for headers
 - If selected strings go away, human effort to notice and update
 - Decent hardening pretty much nukes them
 - Built-in options to remove identifiers (eg, meta generator)
 - Remove standard files
 - Easy to lie to
-
- Fingerprinters like this:
 - Sedusa (in nmap), Wappalyzer, BackendInfo, Plecost, etc, etc...

More Advanced Tools



- Typically improve in one area
 - Resistant to hardening
 - Less labor intensive
- Have their own downsides
 - Less specific results
 - Some request massive amounts of data (> 20 megs!)
 - Some are less generic (Plecost = Wordpress Only)
- Fingerprinters like this:
 - [Sucuri](#), WAF, WhatWeb, BackEndInfo (sortof),

Goals for a (WebApp) Fingerprinter



- Very Generic
- Fast
- Low resource usage
- Accurate (Low FP/FN)
- Resistant to hardening/banner removal
- Super easy to support new versions/apps

→ We'll come back to these later

The Blind Men and the Elephant



Collect and Eliminate Possibilities



Tree or Elephant

Spear or Elephant



Vine or Elephant

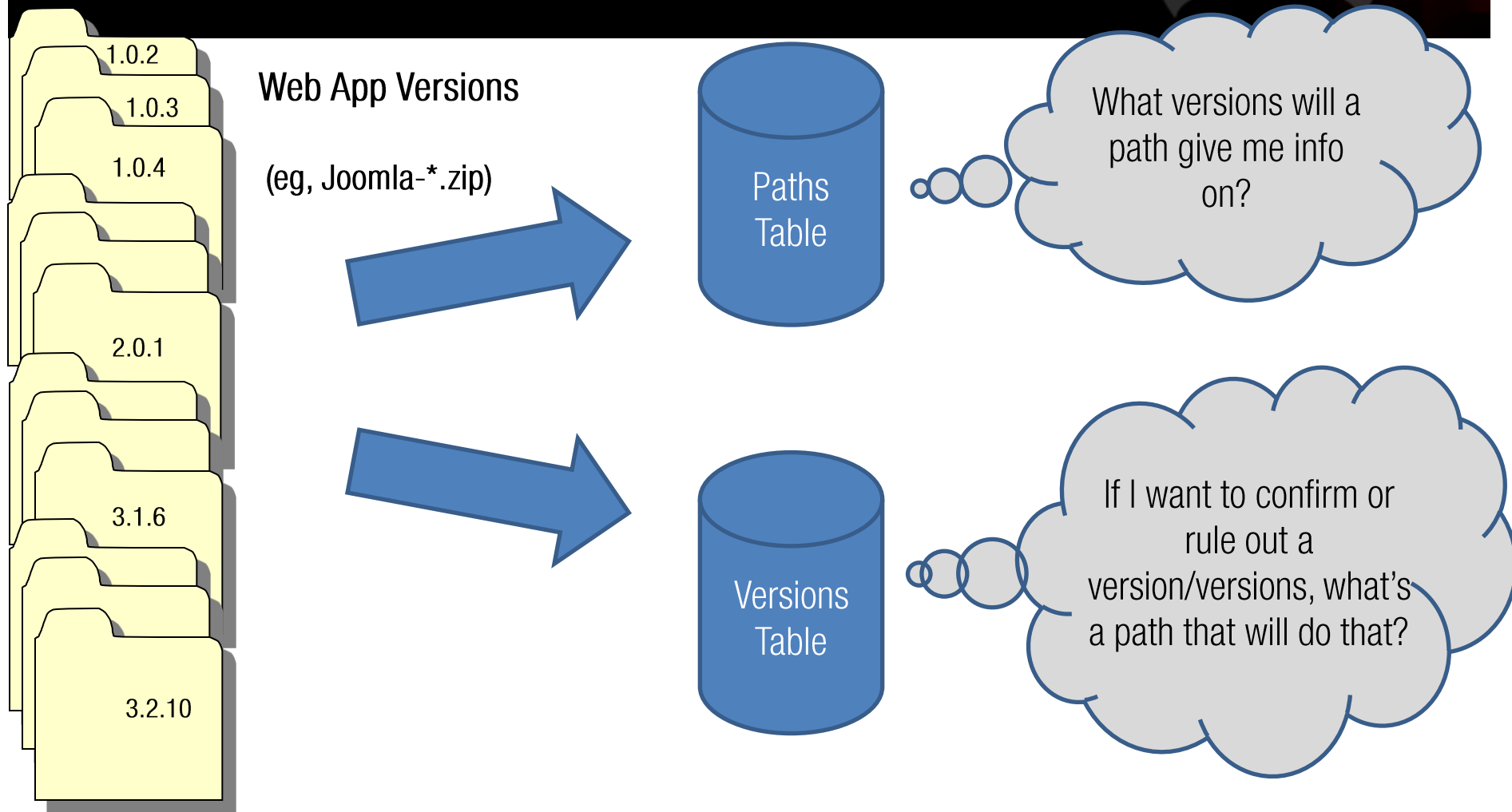
Fan or Elephant



Intersect the Possibilities and...



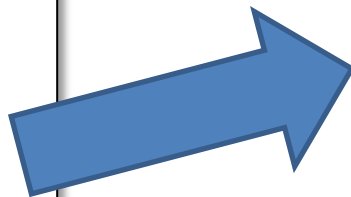
Preparing the Data



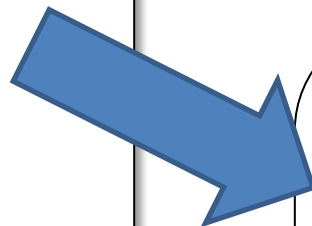
HashTable

wordpress-0.71-gold/*/*.*
wordpress-0.72-beta-1/*/*.*
wordpress-0.72-RC1/*/*.*
wordpress-1.0.1-miles/*/*.*
wordpress-1.0.1-RC1/*/*.*
wordpress-1.0.2/*/*.*
wordpress-1.0.2-blakey/*/*.*
wordpress-1.0-platinum/*/*.*
wordpress-1.0-RC1/*/*.*
wordpress-1.2.1/*/*.*
wordpress-1.2.2/*/*.*
wordpress-1.2-beta/*/*.*
wordpress-1.2-delta/*/*.*
wordpress-1.2-mingus/*/*.*
wordpress-1.2-RC1/*/*.*
wordpress-1.2-RC2/*/*.*
...
wordpress-2.9/*/*.*
wordpress-2.9.1/*/*.*
wordpress-2.9.1-beta1/*/*.*
wordpress-2.9.1-beta1-IIS/*/*.*
wordpress-2.9.1-IIS/*/*.*
wordpress-2.9.1-RC1/*/*.*
wordpress-2.9.1-RC1-IIS/*/*.*
wordpress-2.9-beta-1/*/*.*
wordpress-2.9-beta-1-IIS/*/*.*
wordpress-2.9-beta-2/*/*.*
wordpress-2.9-beta-2-IIS/*/*.*
wordpress-2.9-IIS/*/*.*
wordpress-2.9-RC1/*/*.*
wordpress-2.9-RC1-IIS/*/*.*
wordpress-1.5-strayhorn/*/*.*
wordpress-2.0.7-RC2/*/*.*
wordpress-2.2.1/*/*.*
wordpress-2.5.1/*/*.*
...

PathsTable



VersionsTable



/template/...
740
fc93
7ec
264
/install
b1f
10d
112
8db
560
ad0
590
89e
e06
ce2
efb0c11712681bdecc704ea10225594 ['3.0.5']
045634305e36af4fea75f3a95c415f49 ['3.0.6-RC4']

File

Hash → Version

Hash → Version

Hash → Version

File

Hash → Version

Hash → Version

Hash → Version

3.0.3,3.0.4,3.0.5

('styles/pros

('styles/subs

('adm/style/a

('styles/subs

('styles/pros

('styles/subs

('styles/pros

....
2.0.20,2.0.21

('language/la

('templates/s

('templates/s

('language/la

('install/schemas/mssql_schema.sql', '045c0fcfaa4f89d771b07b66a74....

('contrib/README.html', '61f46292c72f73935bcc2b74403d8b74')

Version, Version, Version

File → Hash

File → Hash

File → Hash

Version

File → Hash

File → Hash

How Many Files?

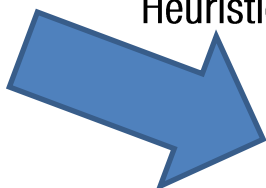


- Wordpress ~83k files in 166 versions
 - phpBB ~17k files in 32 versions
 - MediaWiki ~68k files in 68 versions
 - Joomla ~109k files in 33 versions
 - MovableType ~164k files in 95 versions
 - Drupal ~33k files in 114 versions
 - ... and many more
-
- Wordpress Plugins ~103k files in 1200 versions
 - Drupal Plugins ~76K files in 983 versions

Fingerprinting



Fitness
Heuristic



Best Candidates to Identify the Version

<code>'/htaccess.txt',</code>	14 hashes/31 versions, fitness=15.0
<code>'/language/en-GB/en-GB.ini',</code>	14 hashes/20 versions, fitness=14.64
<code>'/language/en-GB/en-GB.com_content.ini',</code>	13 hashes/20 versions, fitness=13.64
<code>'/configuration.php-dist',</code>	10 hashes/28 versions, fitness=10.90
<code>'/includes/js/joomla.javascript.js',</code>	8 hashes/28 versions, fitness=8.90
<code>'/media/system/js/validate.js',</code>	8 hashes/20 versions, fitness=8.64
<code>'/media/system/js/caption.js',</code>	8 hashes/20 versions, fitness=8.64
<code>'/language/en-GB/en-GB.mod_feed.ini',</code>	8 hashes/20 versions, fitness=8.64
<code>'/media/system/js/openid.js',</code>	8 hashes/20 versions, fitness=8.64
<code>'/language/en-GB/en-GB.com_contact.ini',</code>	8 hashes/20 versions, fitness=8.64
<code>'/language/en-GB/en-GB.mod_breadcrumbs.ini',</code>	7 hashes/20 versions, fitness=7.64
<code>'/media/system/js/combobox.js',</code>	7 hashes/20 versions, fitness=7.64
<code>'/language/en-GB/en-GB.mod_search.ini',</code>	7 hashes/20 versions, fitness=7.64
<code>'/templates/rhuk_milkyway/css/template.css',</code>	7 hashes/20 versions, fitness=7.64
<code>'/media/system/js/switcher.js',</code>	7 hashes/20 versions, fitness=7.64

Candidate Files: MovableType



- /mt-static/mt.js
- /mt-static/js/tc/client.js
- /mt-static/css/main.css
- /tools/run-periodic-tasks
- /mt-static/js/tc/tagcomplete.js
- /mt-static/js/edit.js
- /mt-static/js/tc/mixer/display.js
- /mt-static/js/archetype_editor.js
- /mt-static/js/tc/mixer.js
- /mt-static/js/tc/tableselect.js
- ...

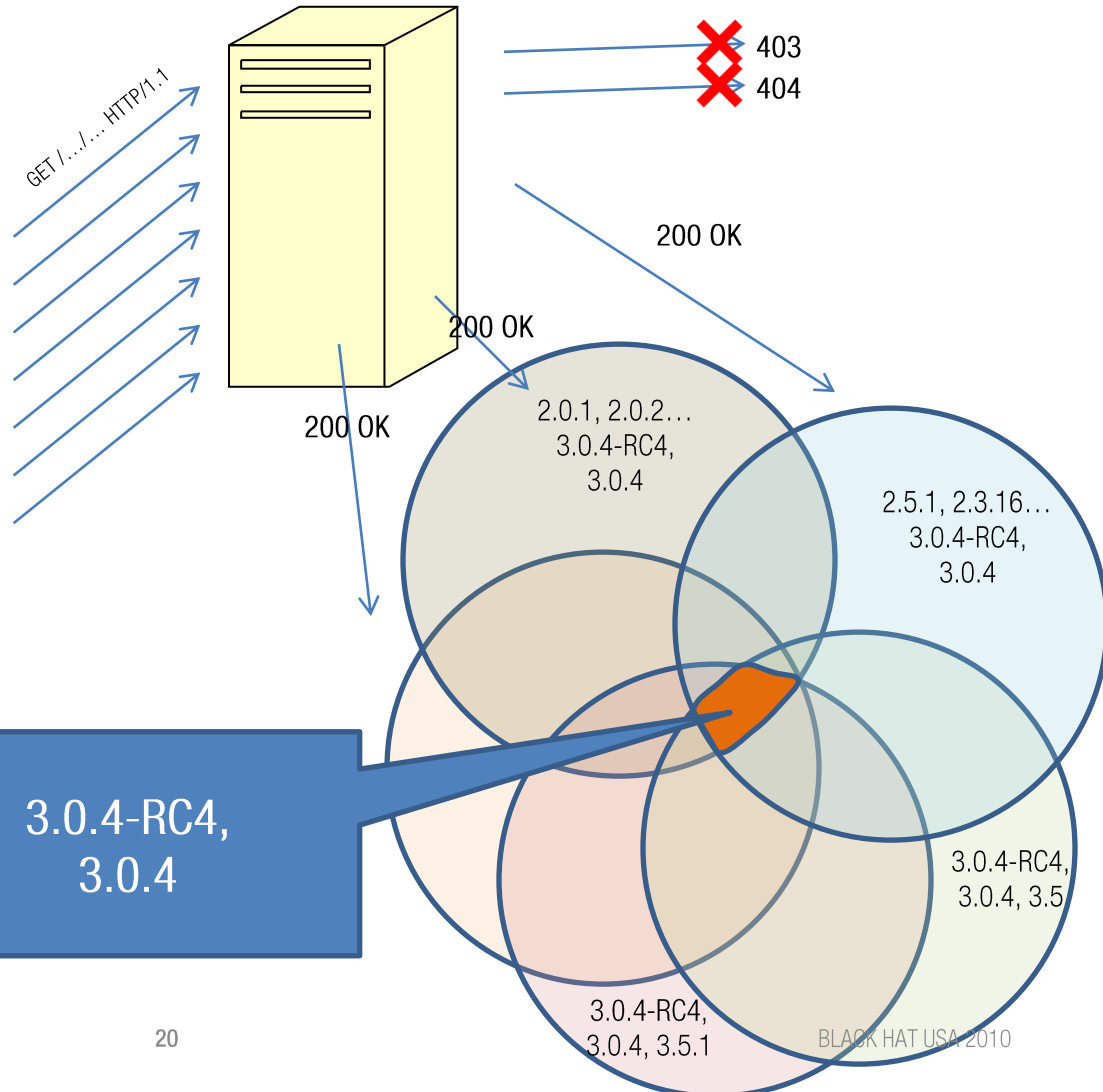
Candidate Files: Mediawiki



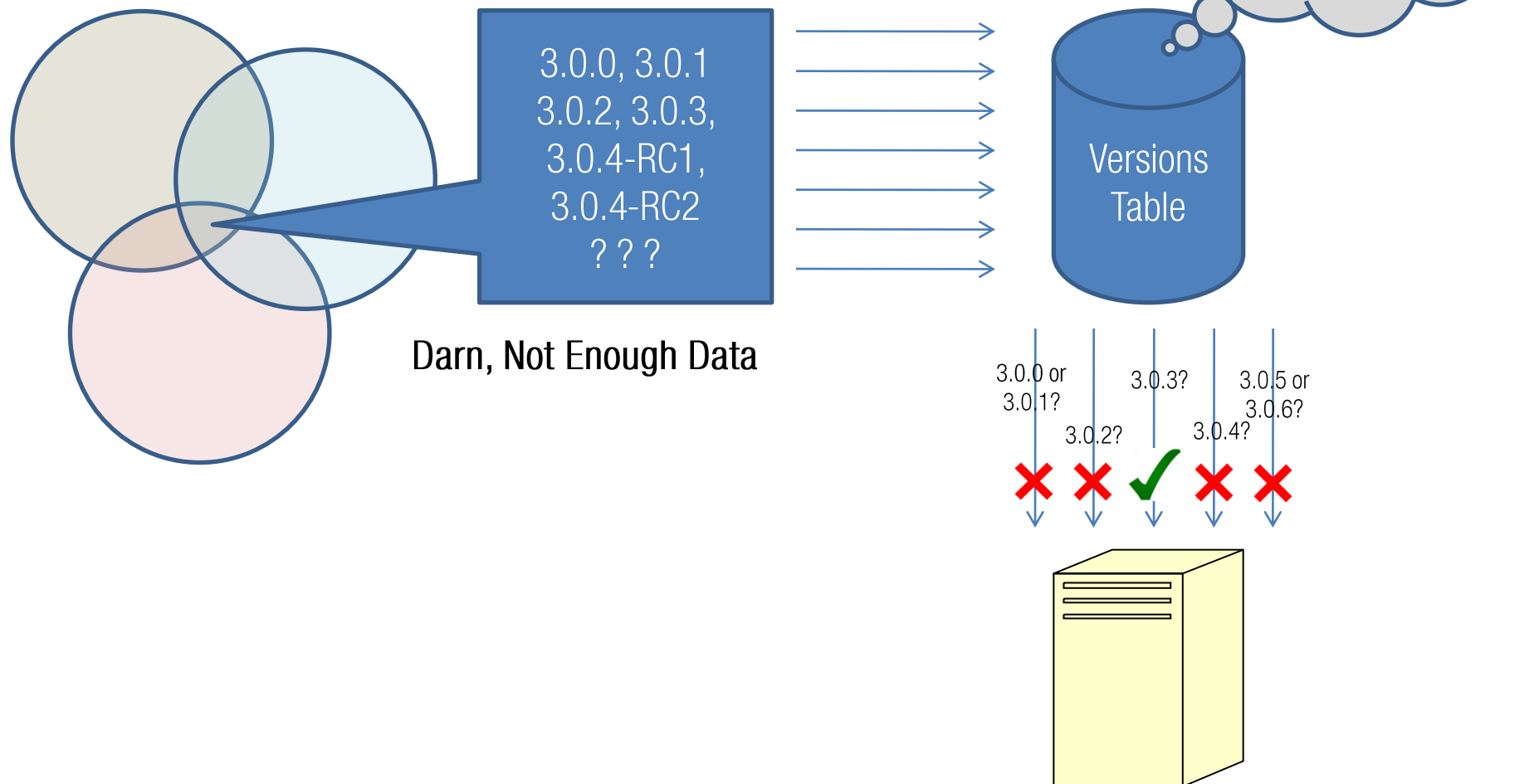
- /Documentation.html
- /ChangeLog
- /translators.html
- /README
- /scripts/create-release.sh
- /lang/sync_lang.sh
- /Documentation.txt
- /scripts/create_tables.sql
- /js/functions.js
- /lang/check_lang.sh
- ...

*Fully data-driven approach
finds useful info even in
obscure and counterintuitive
files*

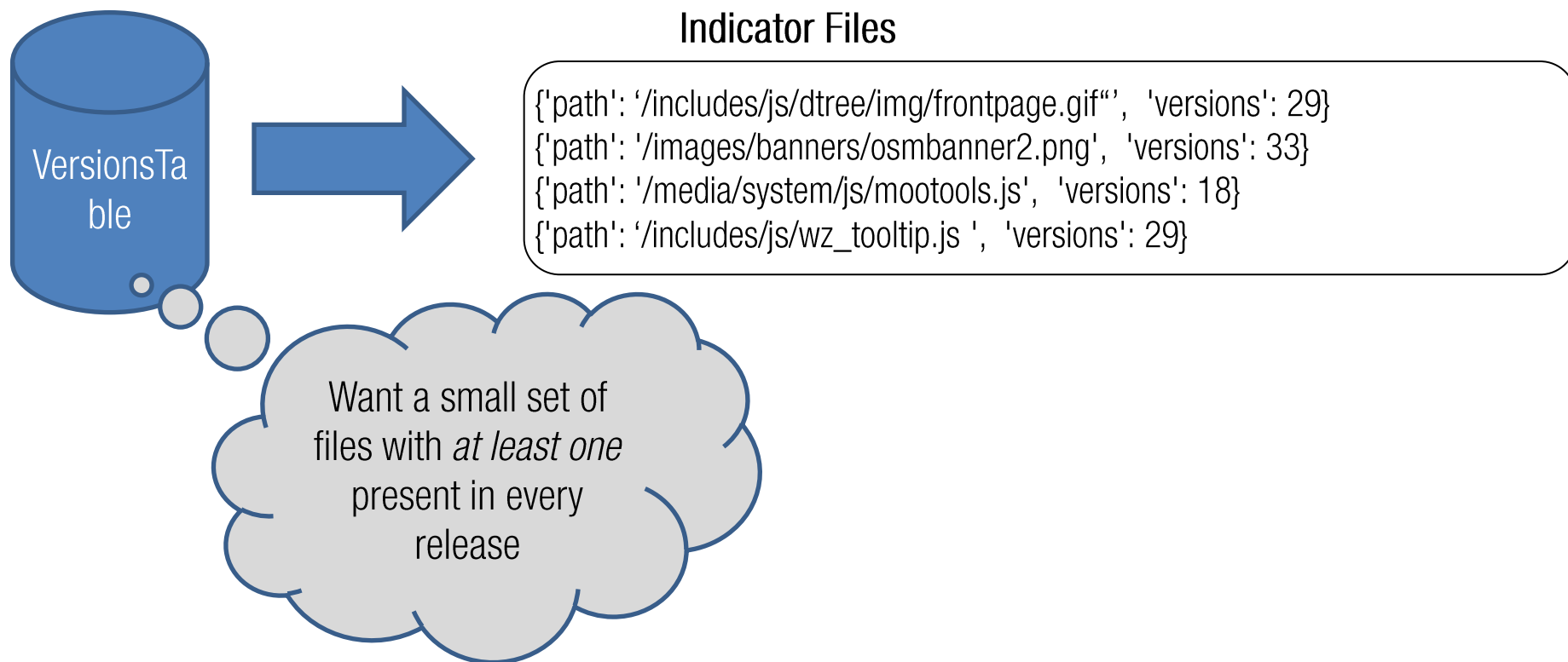
```
'/htaccess.txt'  
'/language/en-GB/en-GB.ini'  
'/language/en-GB/en-GB.com_content.ini'  
'/configuration.php-dist',  
'/includes/js/joomla.javascript.js'  
'/media/system/js/validate.js'  
'/media/system/js/caption.js'  
'/language/en-GB/en-GB.mod_feed.ini'  
'/media/system/js/openid.js'  
'/language/en-GB/en-GB.com_contact.ini'  
'/language/en-GB/en-GB.mod_breadcrumbs.ini'  
'/media/system/js/comboobox.js'  
'/language/en-GB/en-GB.mod_search.ini'  
'/templates/rhuk_milkyw/css/template.css'  
'/media/system/js/switcher.js'
```



Winnowing



App Discovery / App Guessing

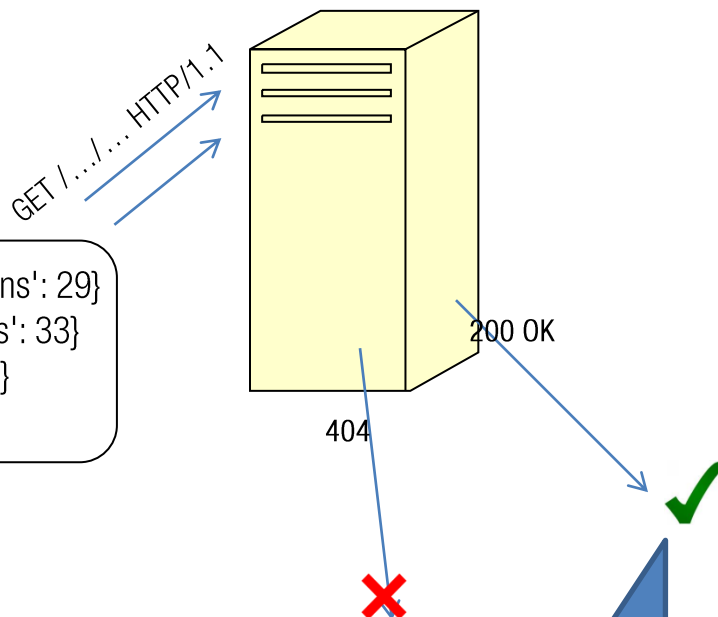


App Discovery / App Guessing



Indicator Files

```
{'path': '/includes/js/dtree/img/frontpage.gif', 'versions': 29}  
{'path': '/images/banners/osmbanner2.png', 'versions': 33}  
{'path': '/media/system/js/mootools.js', 'versions': 18}  
{'path': '/includes/js/wz_tooltip.js ', 'versions': 29}
```



It's some version of Joomla

Supporting a New App



- Gather every version you can find, dump them in a directory
- [Optional] Supply a regex to exclude directories/files from fingerprinting
 - (eg .php files, protected admin directory, .htaccess, etc)
- Use BlindElephant to build the datafiles
- Fingerprint!
- ...Profit?

Does it work?



- `./BlindElephant.py` **<http://laws.qualys.com>** movabletype
- Loaded movabletype with 96 versions, 2229 differentiating paths, and 209 version groups.
- Starting BlindElephant fingerprint for version of movabletype at <http://laws.qualys.com>
- Hit <http://laws.qualys.com/mt-static/mt.js>
- Possible versions based on result: 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc/client.js>
- Possible versions based on result: 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/css/main.css>
- Possible versions based on result: 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM
- Hit <http://laws.qualys.com/tools/run-periodic-tasks>
- File produced no match. Error: Error code: 404 (Not Found)

Does it work?



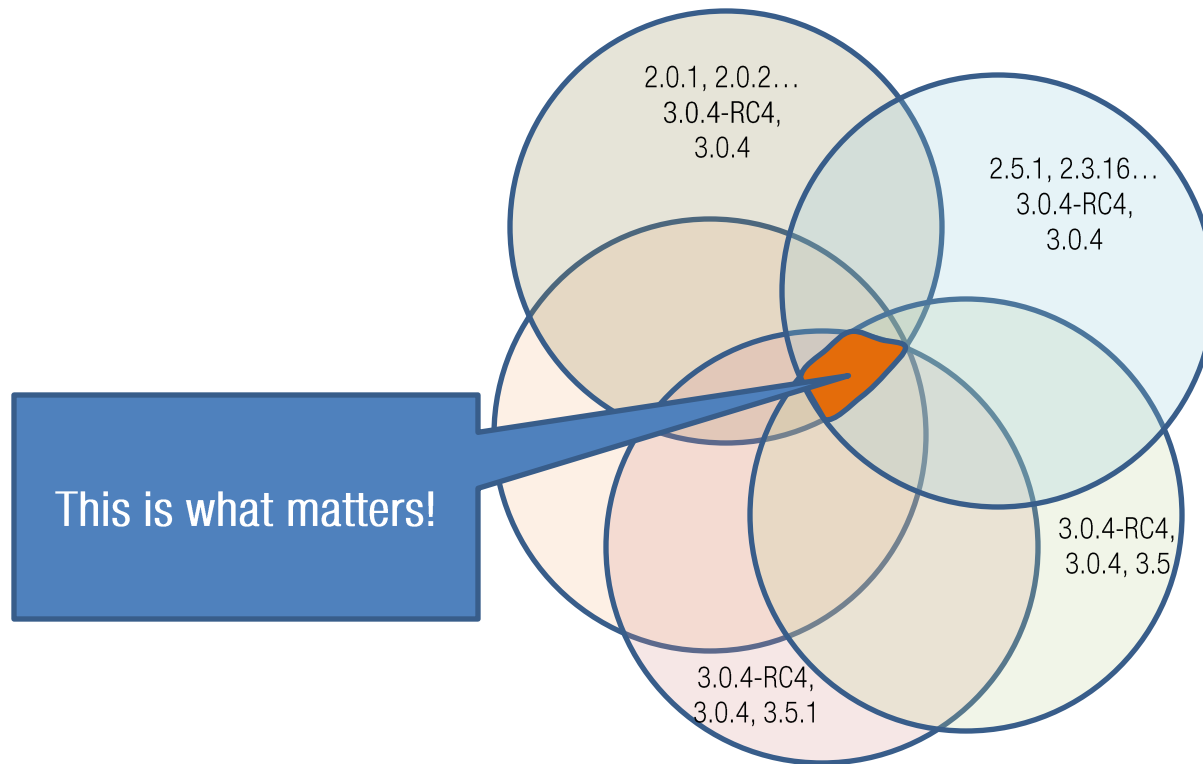
- Hit <http://laws.qualys.com/mt-static/js/tc/tagcomplete.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/edit.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc/mixer/display.js>
- Possible versions based on result: 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit http://laws.qualys.com/mt-static/js/archetype_editor.js
- Possible versions based on result: 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM

Does it work?



- Hit <http://laws.qualys.com/mt-static/js/tc/mixer.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc/tableselect.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc/focus.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM

Interlude



Does it work?



- Hit <http://laws.qualys.com/mt-static/css/simple.css>
- Possible versions based on result: 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM
- Hit http://laws.qualys.com/mt-static/mt_ja.js
- Possible versions based on result: 4.2-en, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.23-en-OS, 4.24-en, 4.24-en, 4.24-en-COM
- Hit <http://laws.qualys.com/mt-static/js/tc/gestalt.js>
- Possible versions based on result: 4.1-en, 4.1-en-CS, 4.2-en, 4.21-en, 4.21-en, 4.21-en-COM, 4.22-en, 4.22-en, 4.22-en-COM, 4.23-en, 4.23-en, 4.23-en-COM, 4.24-en, 4.24-en, 4.24-en-COM
- Fingerprinting resulted in: Best Guess: **4.23-en-COM**
- 4.22-en,
- 4.22-en-COM,
- 4.23-en,
- 4.23-en-COM

BTW: It Does Plugins Too



- `$./BlindElephant.py -s -p guess http://example.com drupal`
- Possible plugins:
- `['admin_menu', 'cck', 'date', 'google_analytics', 'imce', 'imce_swfupload', 'pathauto', 'print', 'spamicide', 'tagadelic', 'token', 'views']`

- `$./BlindElephant.py -s -p imce http://example.com drupal`
- `<snip>`
- Fingerprinting resulted in:
- 6.x-1.3

New Toy! Lets Play

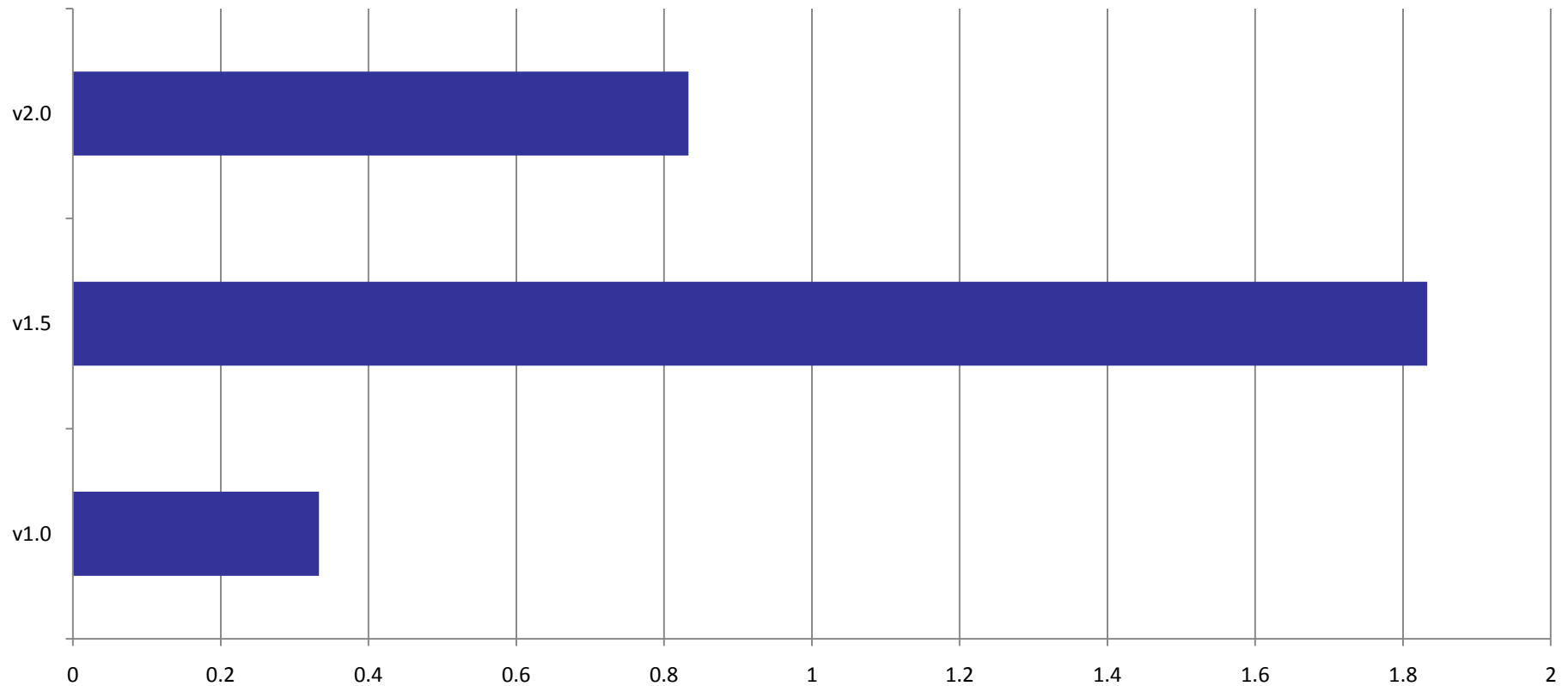


- App ID & Fingerprinting on 1,084,152 hosts
- ~34k targeted scans for bug shakeout and calibration
 - Shodan = Really, really useful (kinda expensive though)
 - Is John here? I owe him a beer.
 - Slightly biased sample (skews to default installs, s'okay though)
- ~50k and ~1M host random sample of 87M .com domains
 - Stats on accuracy and net-wide webapp population are from these

On To the Results...



Version Distribution: SomeApp



Graphing Sets of Possibilities

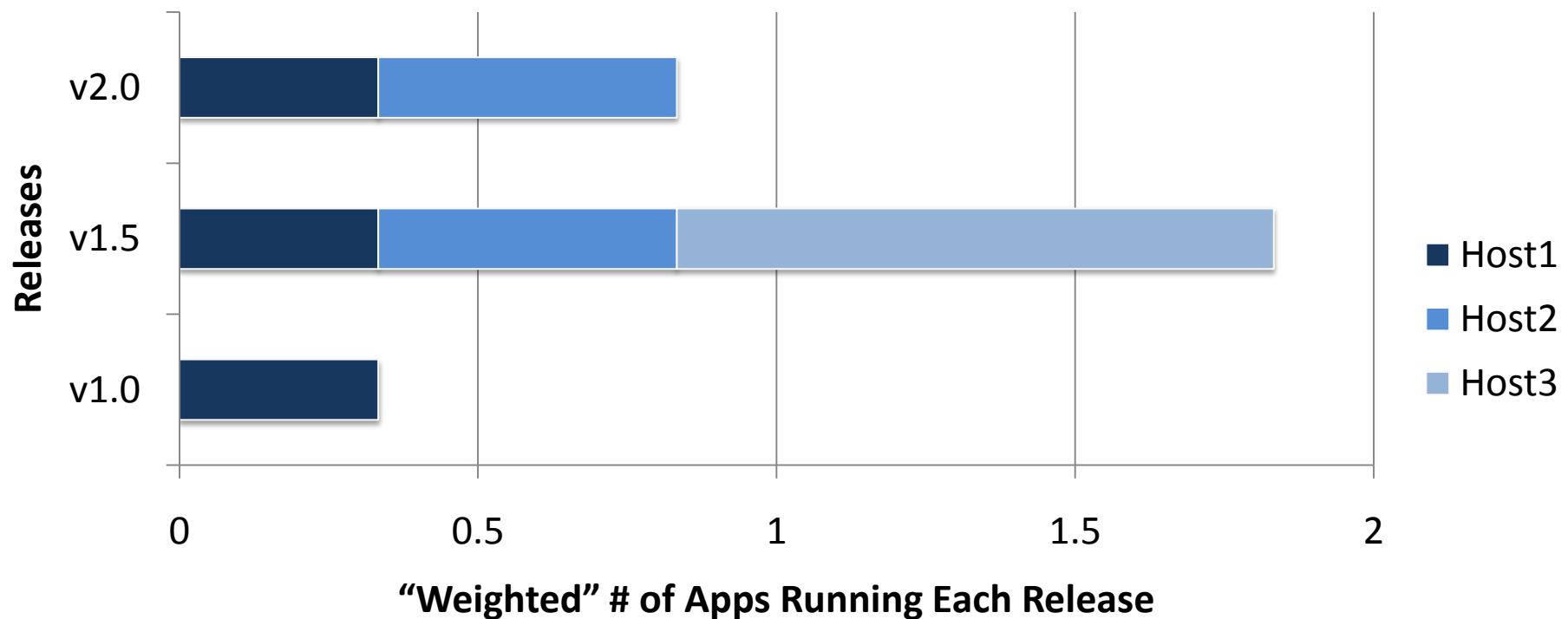


- Host1 Possible Versions: v1.0, v1.5, v2.0
 - .33 to three version columns
- Host2 Possible Versions: v1.5, v2.0
 - .5 to two version columns
- Host3 Possible Versions: v1.5
 - 1.0 to v1.5

Graphing Sets of Possibilities

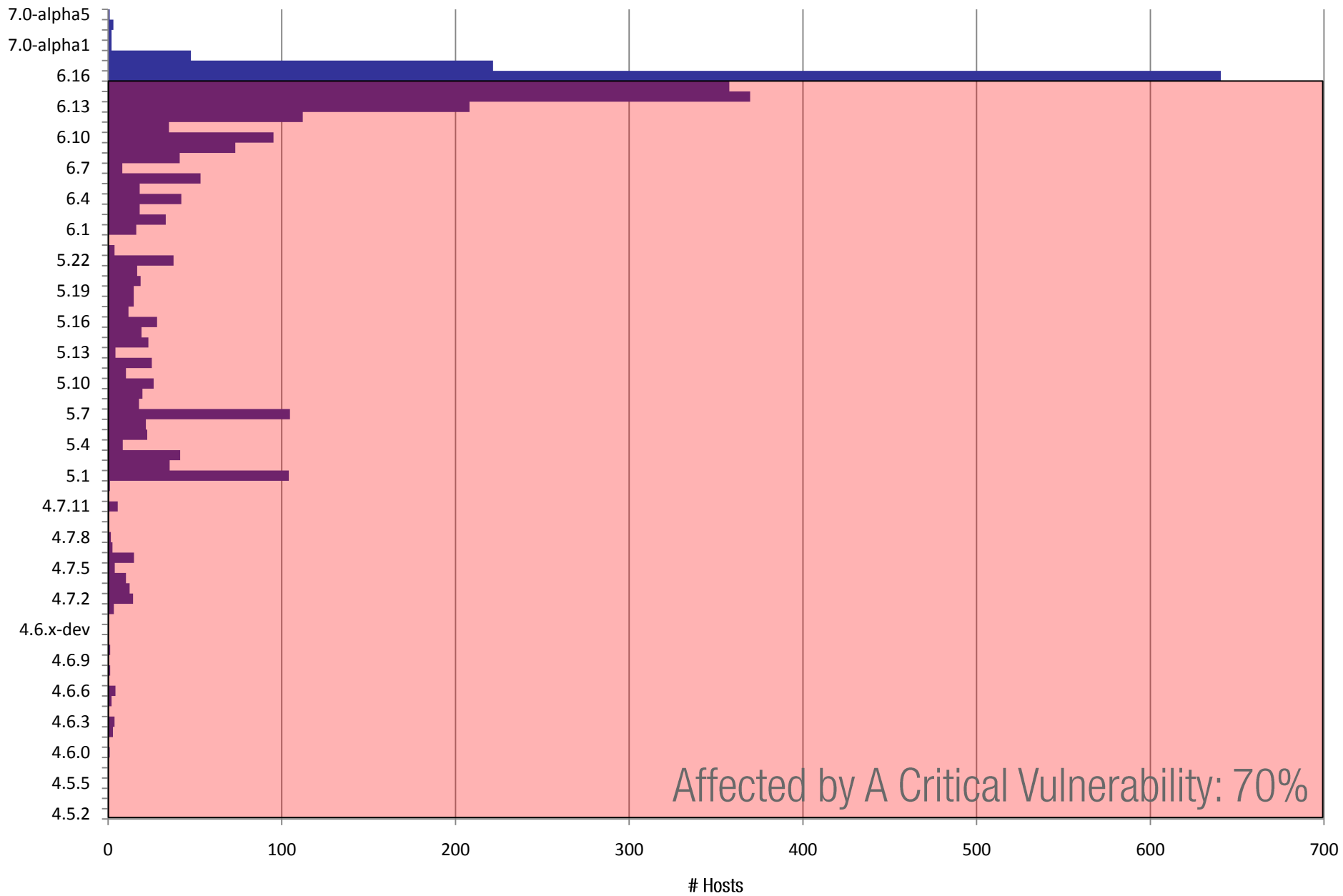


**Version Distribution: Some App
(6/18/10)**



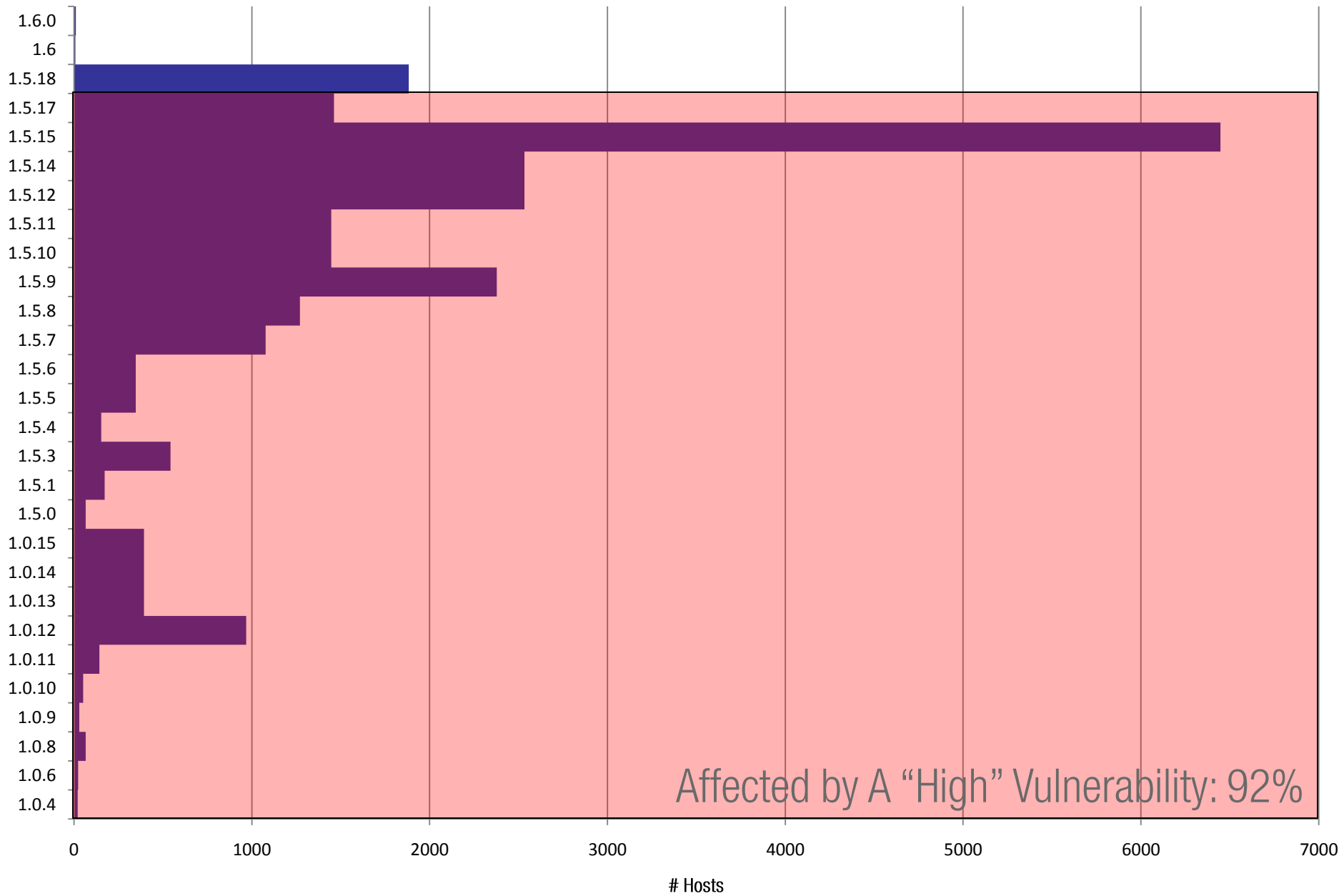
Version Distribution: Drupal

(June 18, 2010)



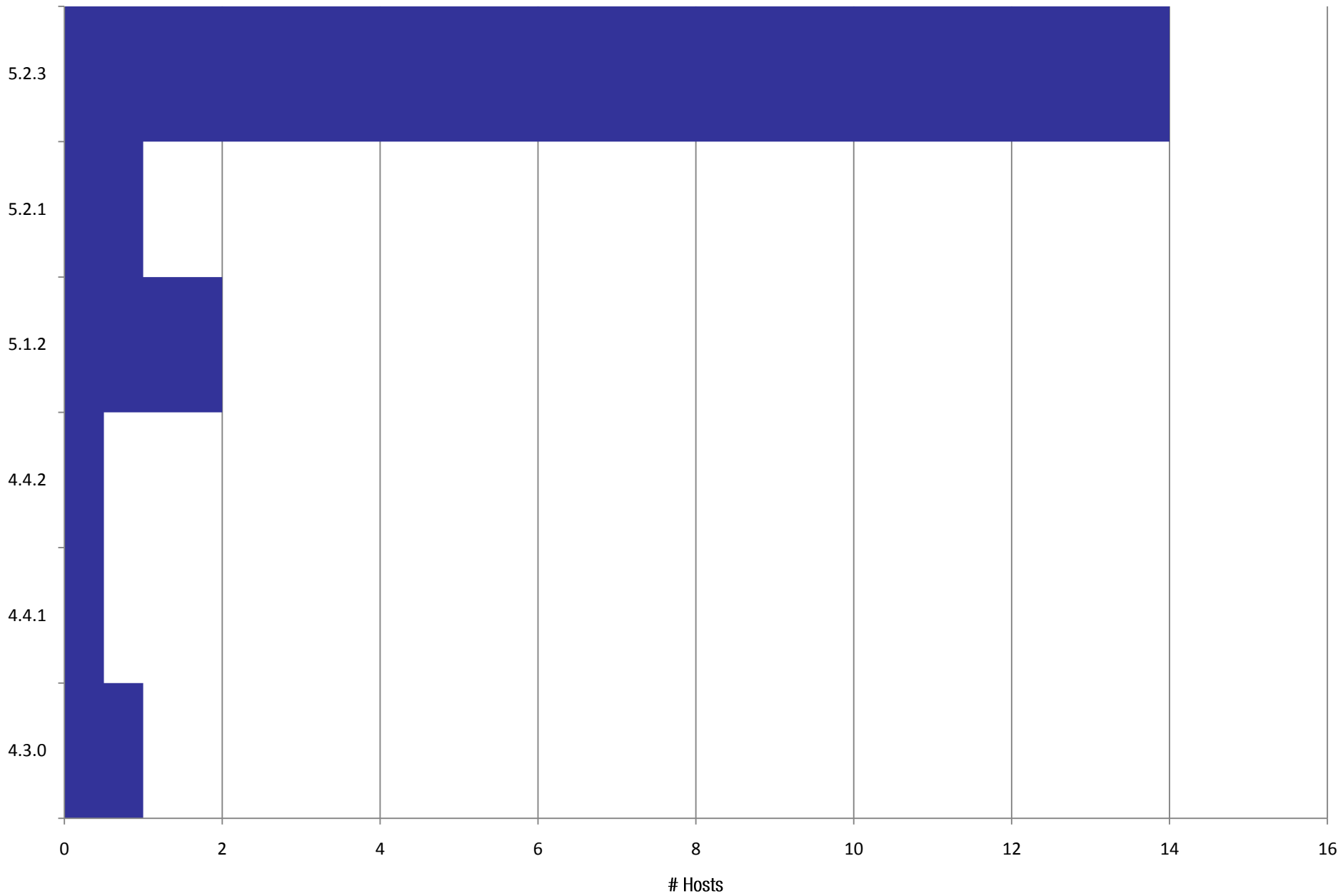
Version Distribution: Joomla

(June 18 2010)



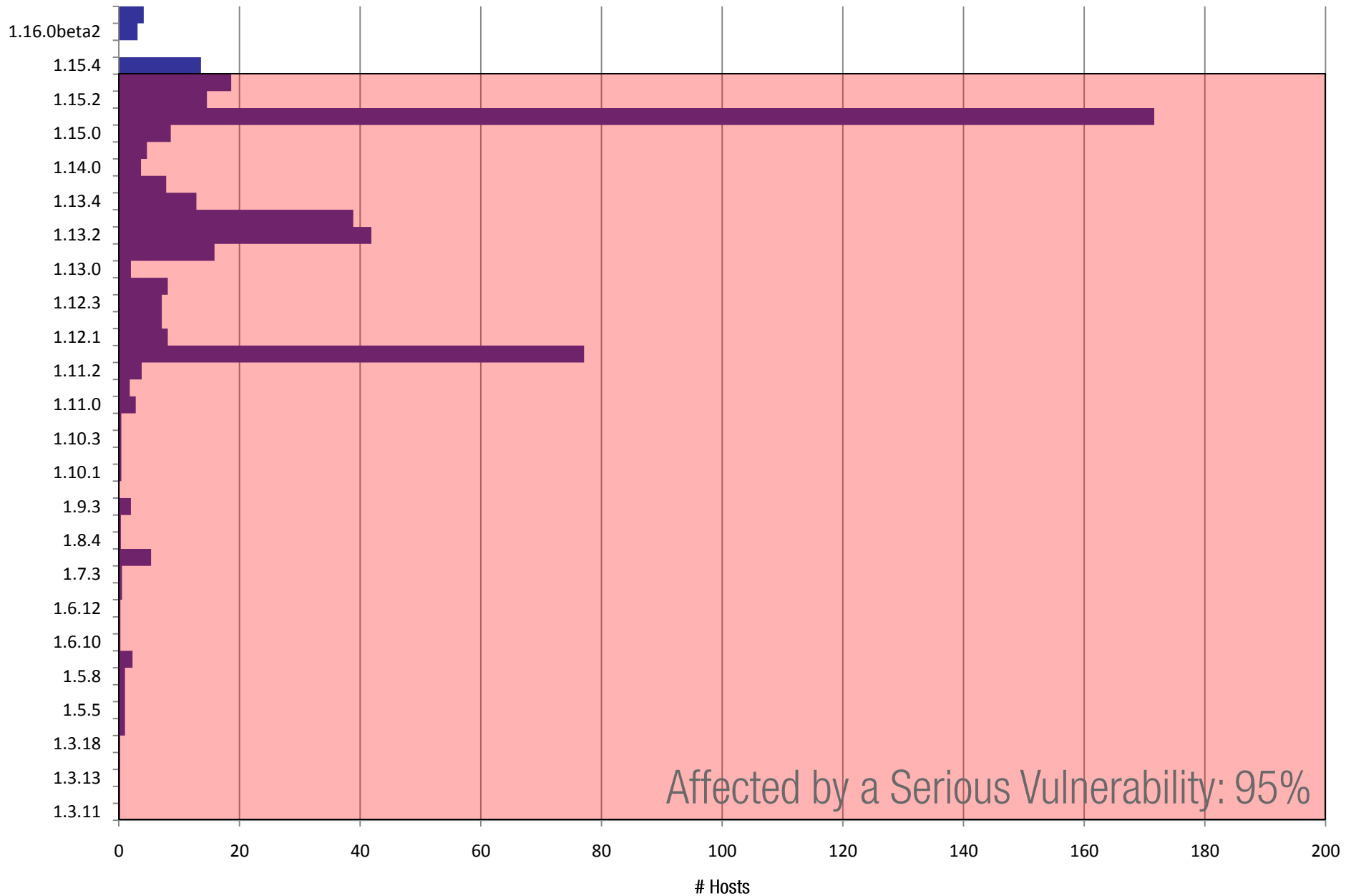
Version Distribution: Liferay

(June 18, 2010)



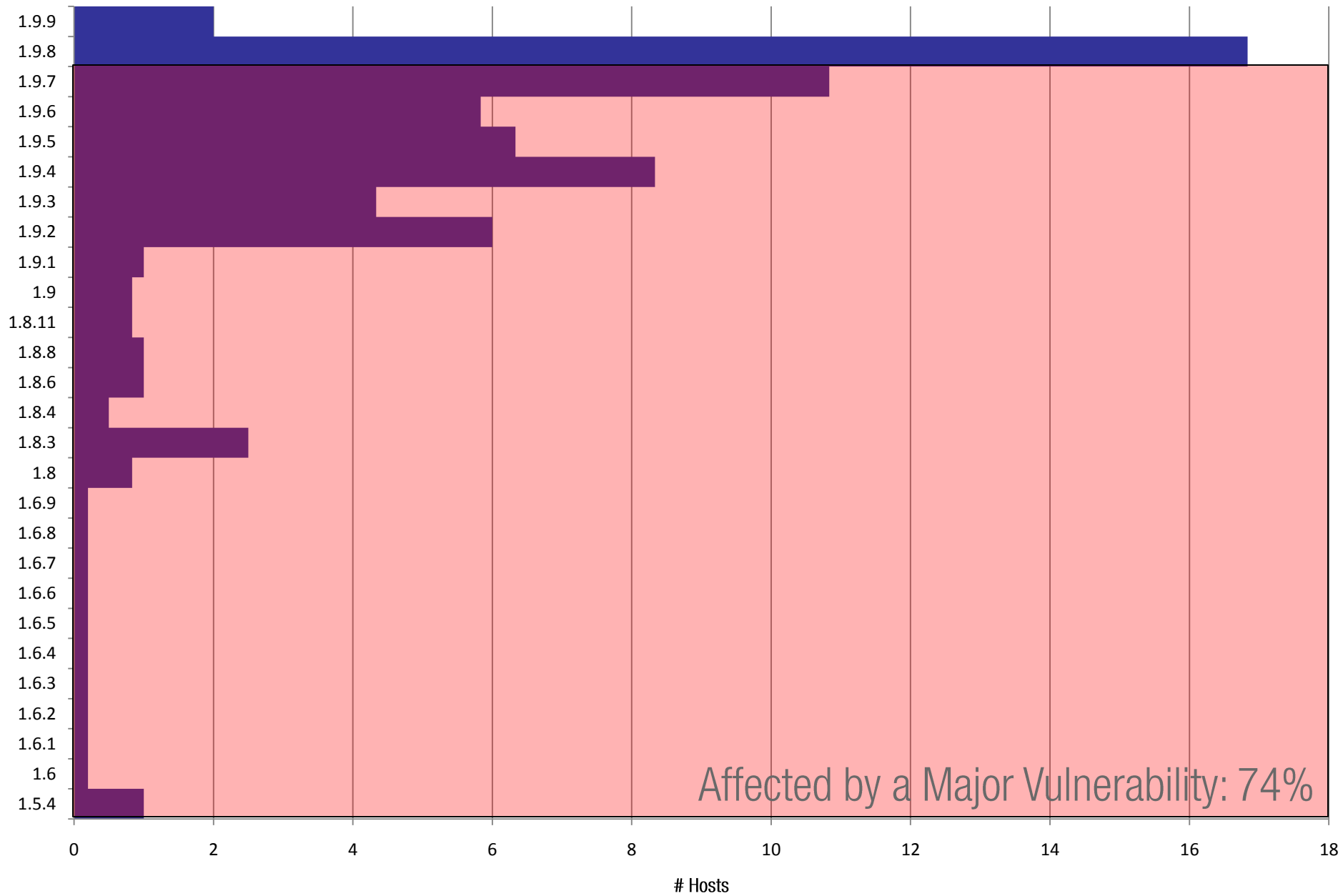
Version Distribution: Mediawiki

(June 18, 2010)



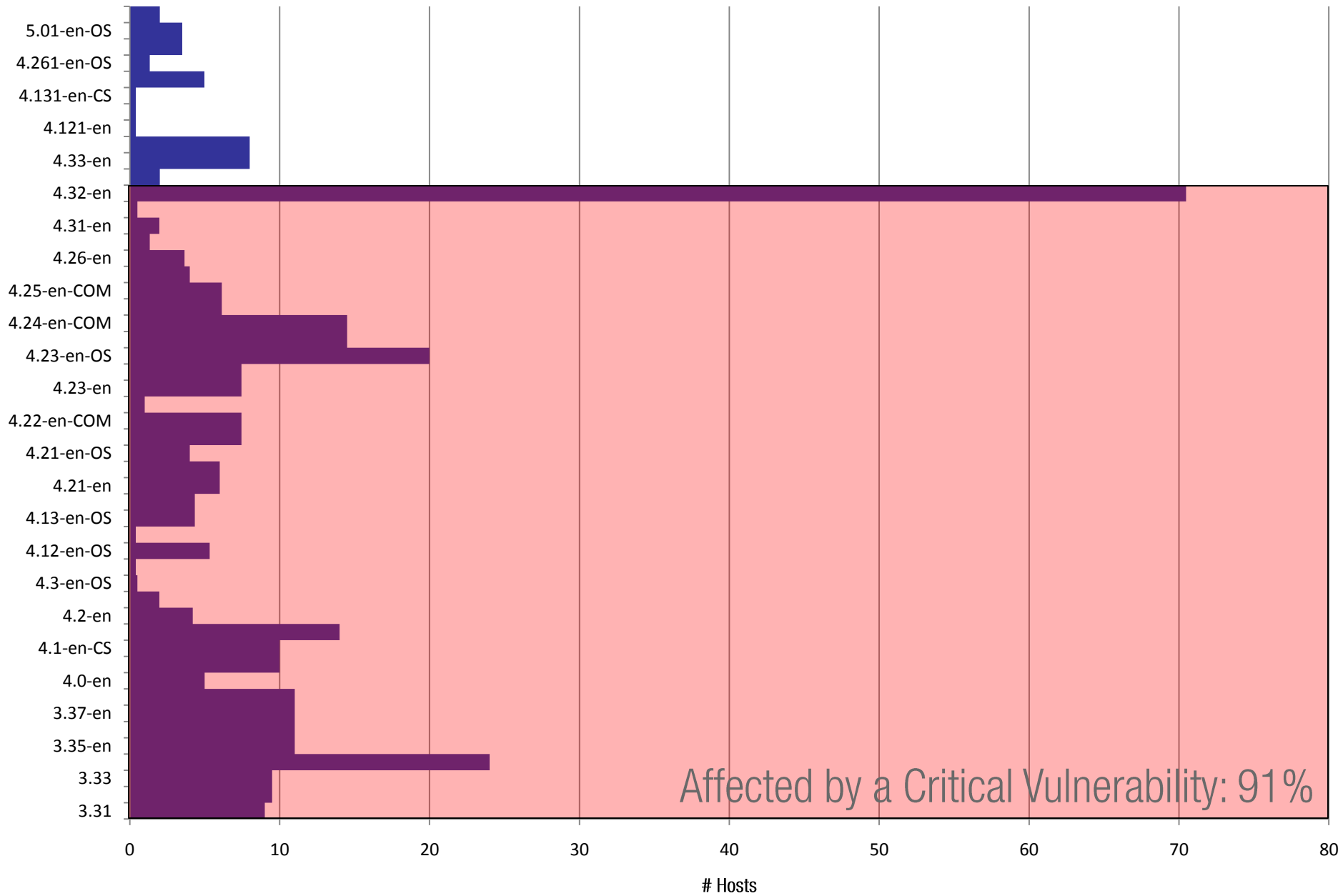
Version Distribution: Moodle

(June 18, 2010)



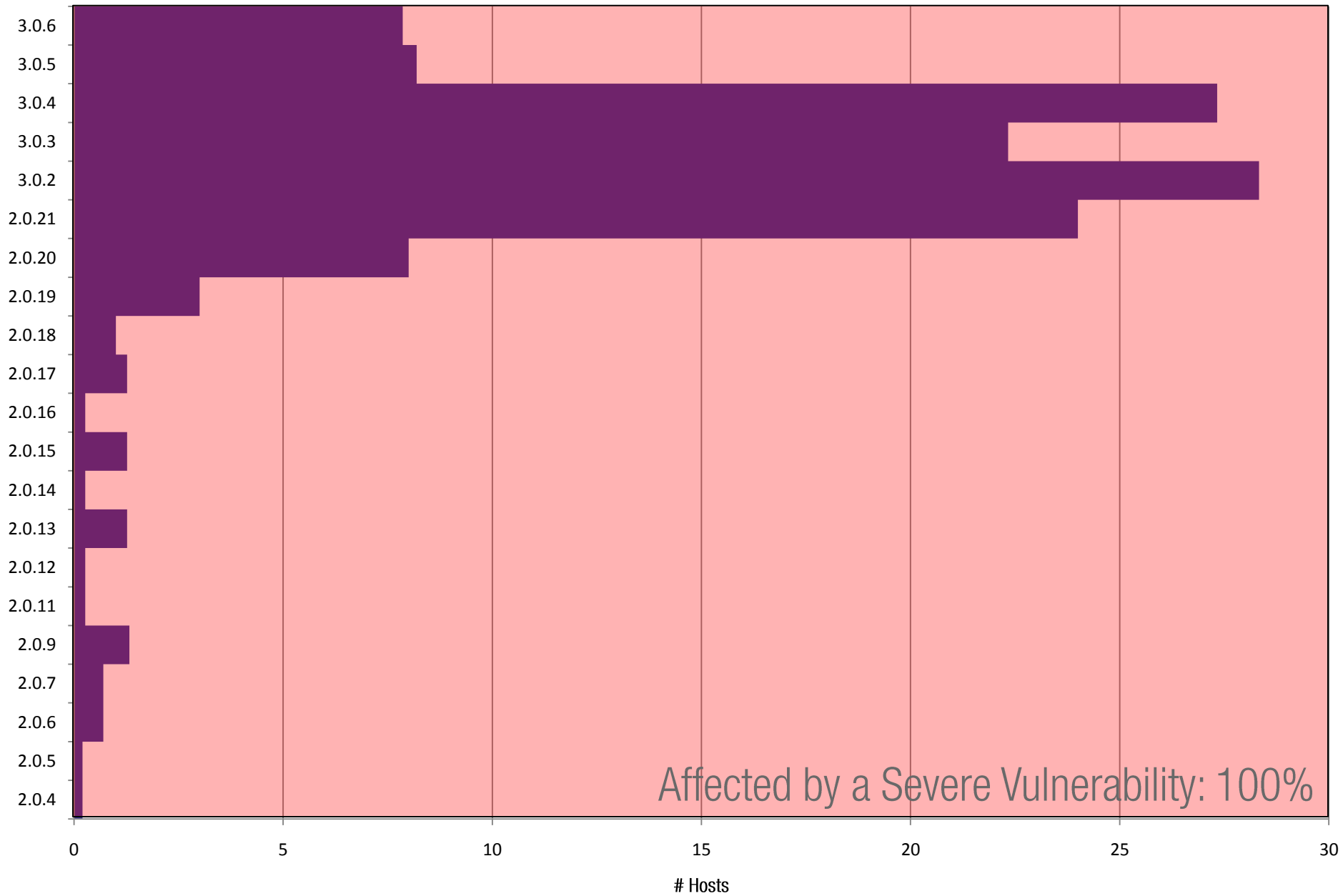
Version Distribution: MovableType

(June 18, 2010)



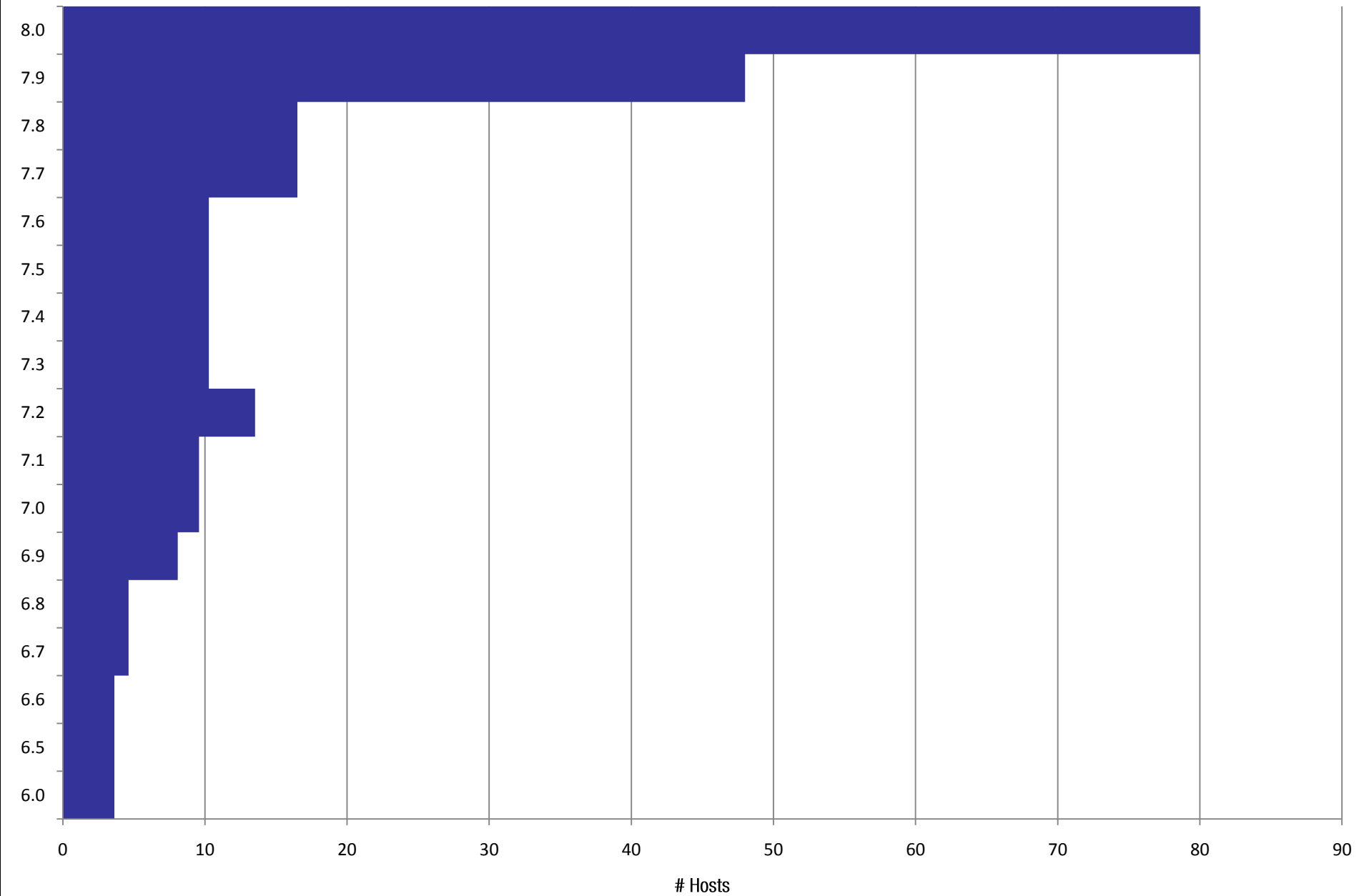
Version Distribution: phpBB

(June 18, 2010)



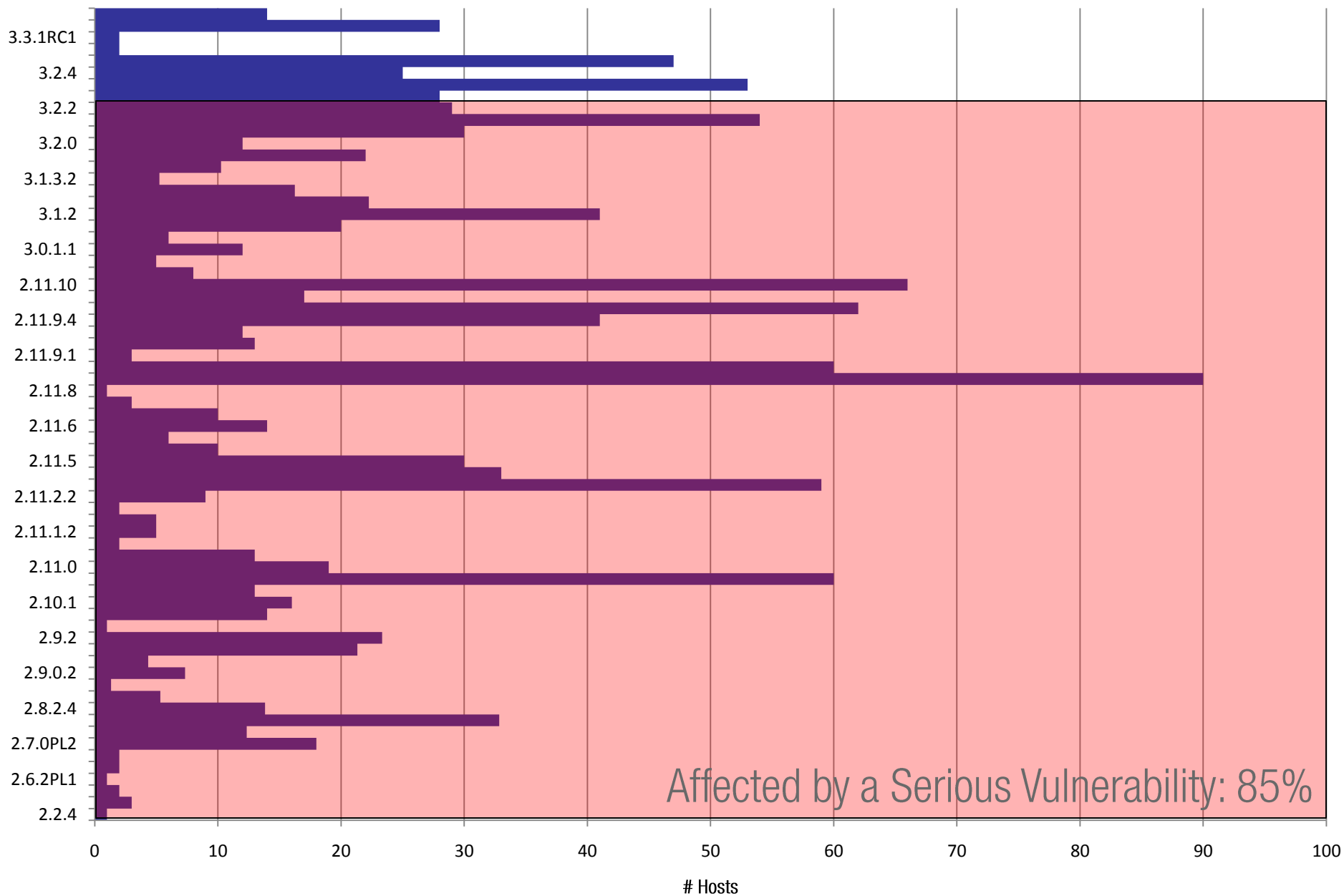
Version Distribution: PHPNuke

(June 18, 2010)



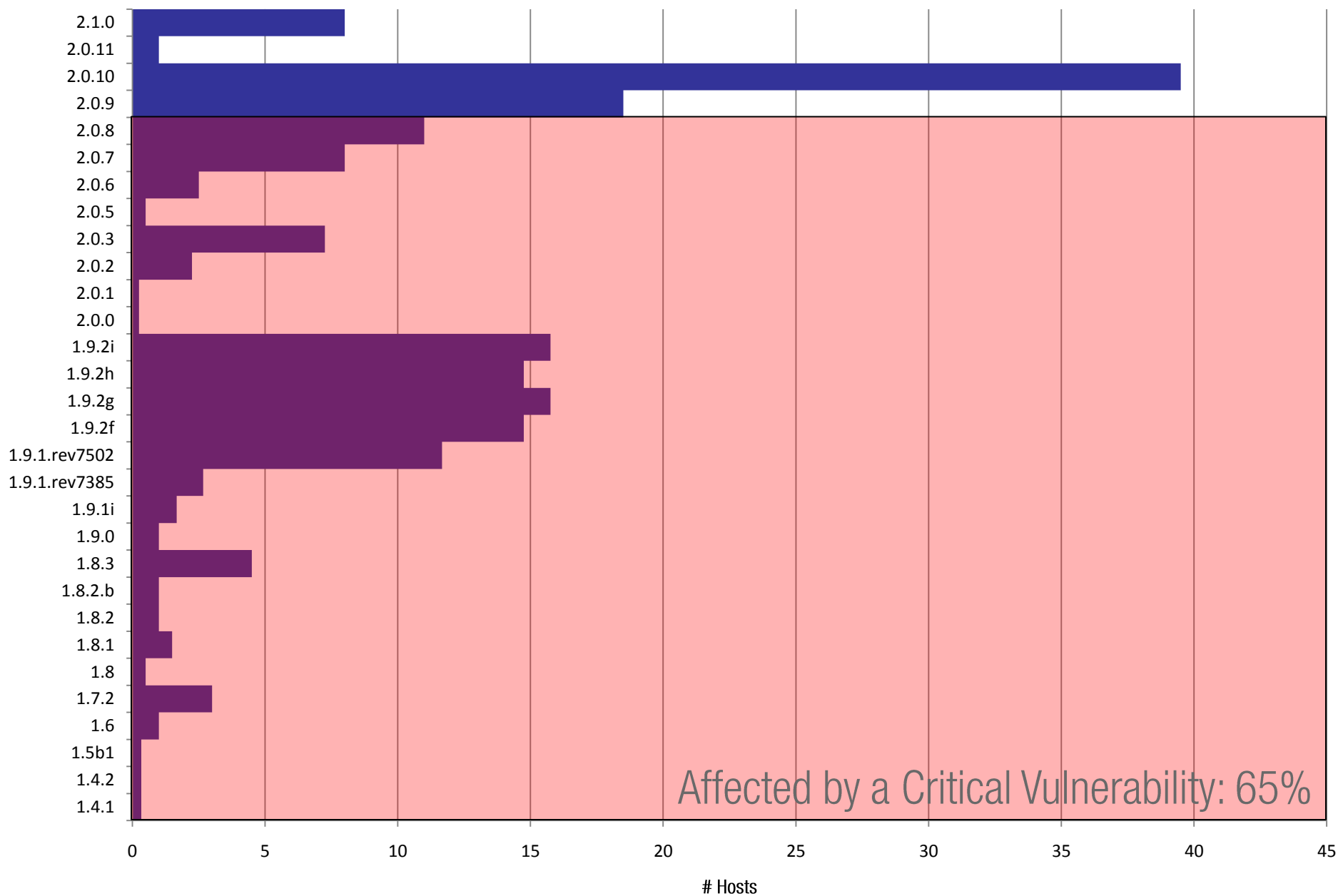
Version Distribution: phpMyAdmin

(June 18, 2010)



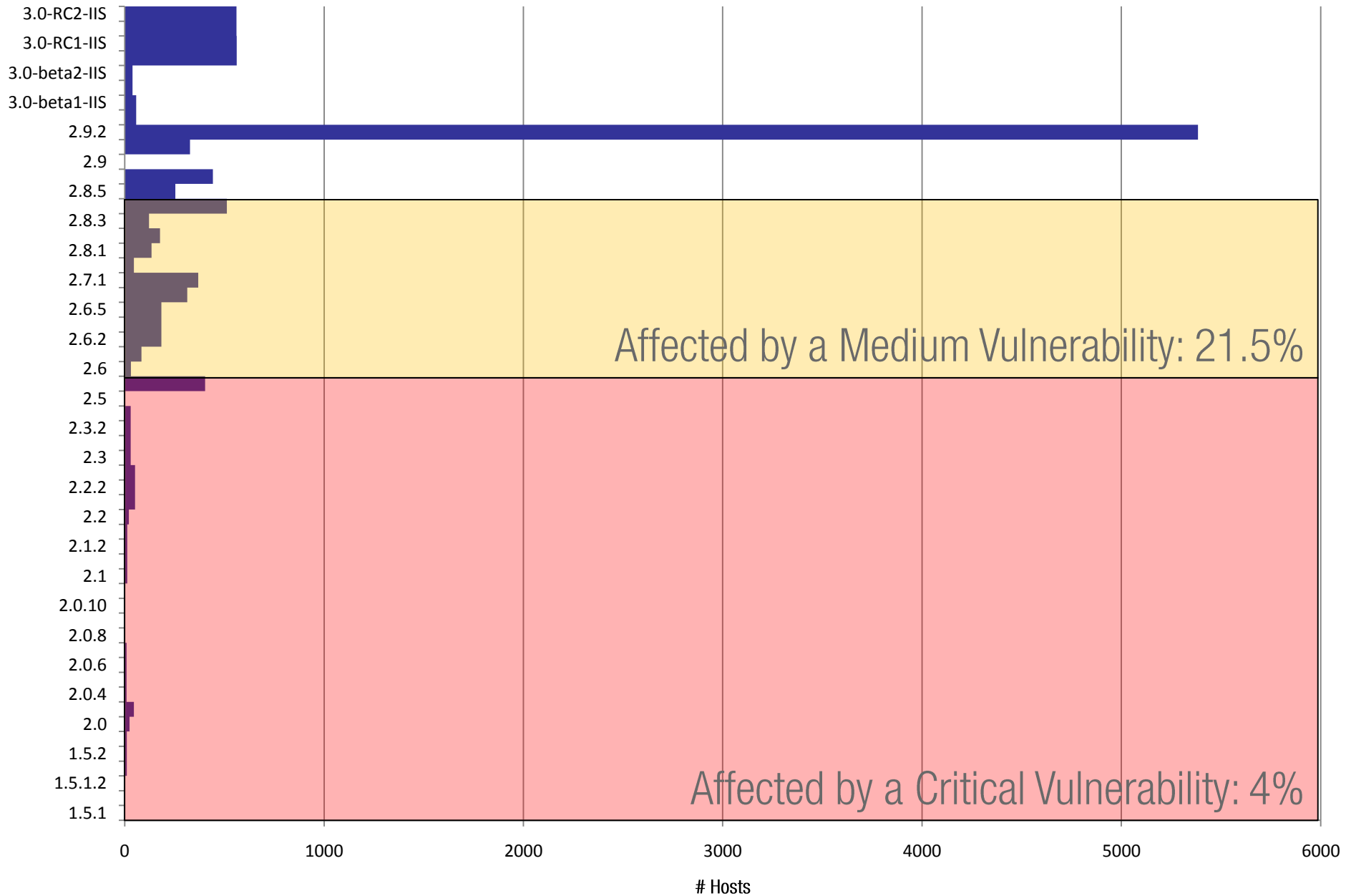
Version Distribution: SPIP

(June 18, 2010)

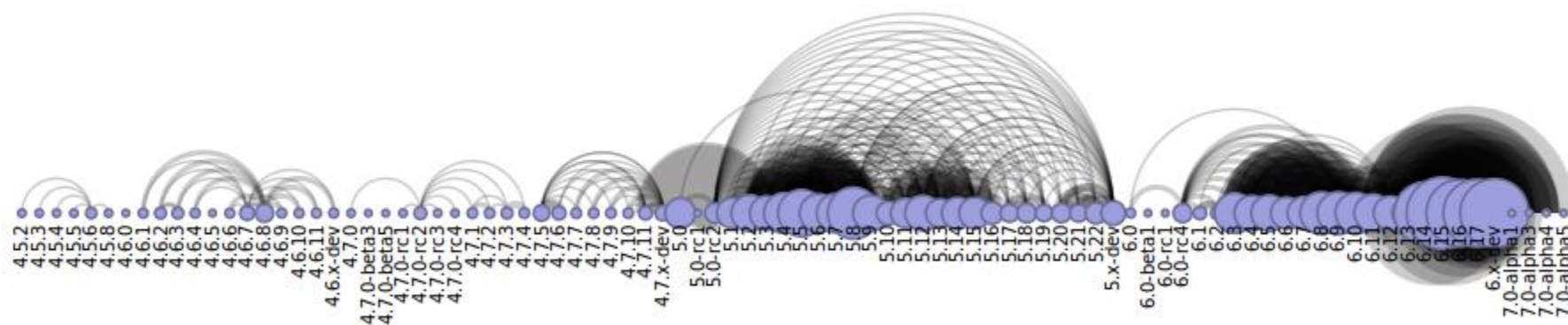


Version Distribution: Wordpress

(June 18, 2010)

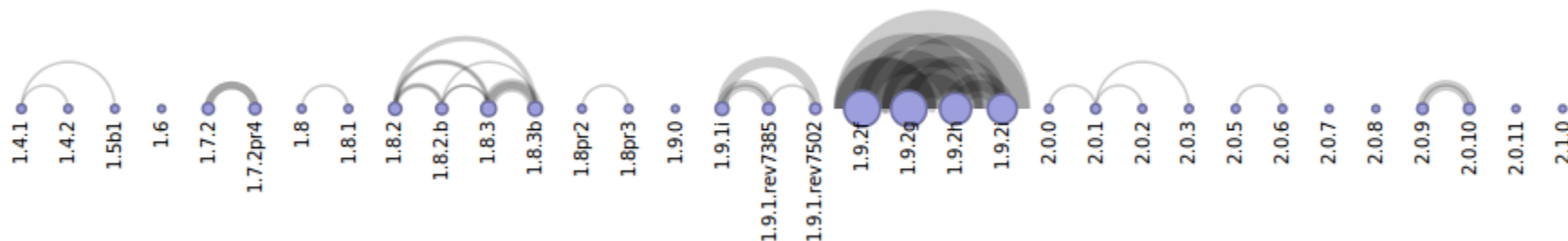


Version Clusters: Drupal

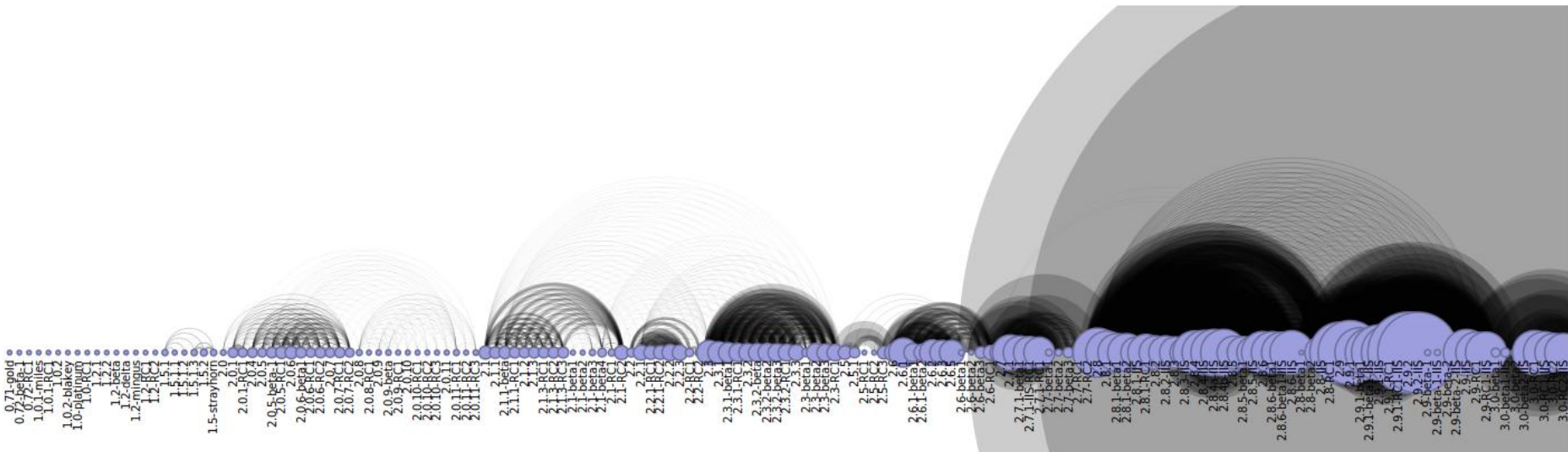


Connections between nodes (versions) indicate that a fingerprint job produced those versions and was unable to differentiate. Thicker connections indicate the number of times this confusion occurred (two particularly difficult to distinguish versions), while many connections among adjacent nodes indicates a family of difficult to distinguish versions.

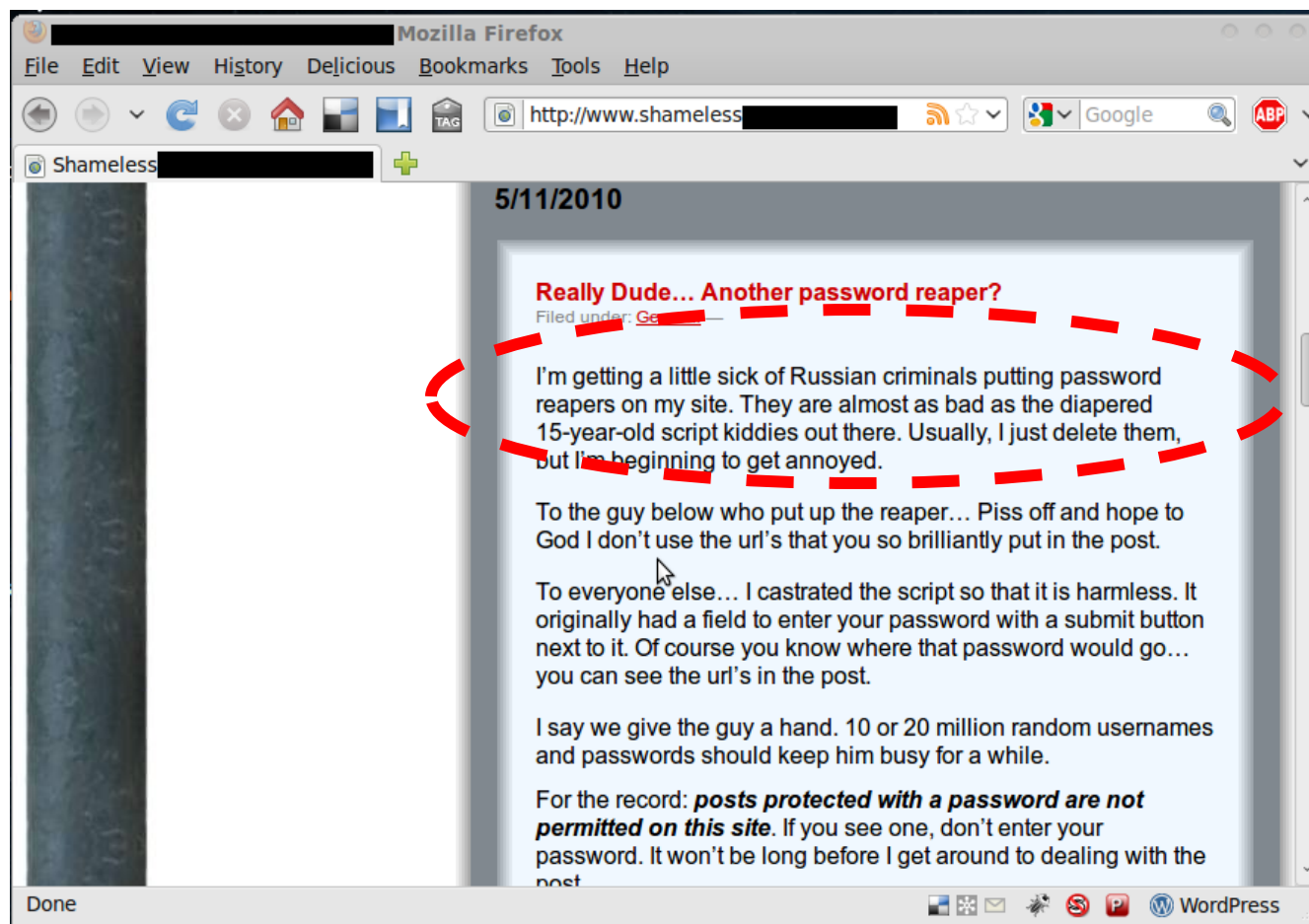
Version Clusters: SPIP



Connections between nodes (versions) indicate that a fingerprint job produced those versions and was unable to differentiate. Thicker connections indicate the number of times this confusion occurred (two particularly difficult to distinguish versions), while many connections among adjacent nodes indicates a family of difficult to distinguish versions.



Lost: a Clue



Lost: A Clue



5/11/2010

Really Dude... A Russian criminal putting password reapers?
Filed under: [General](#)

I'm getting... Russian criminals putting password reapers... They are almost as bad as the diapered 15-year-old kiddy out there. Usually, I just delete them, but I'm starting to get annoyed.

...ss off and hope to... put in the post.

...hat it is harmless. It... with a submit button... sword would go...

I say we give the guy a hand. 10 or 20 million random usernames and passwords should keep him busy for a while.

For the record: **posts protected with a password are not permitted on this site.** If you see one, don't enter your password. It won't be long before I get around to dealing with the post.

```
Fingerprinting resulted in
1.2.2
1.2-mingus

Best Guess: 1.2-mingus
```

He's only 6 years and 60 releases behind...

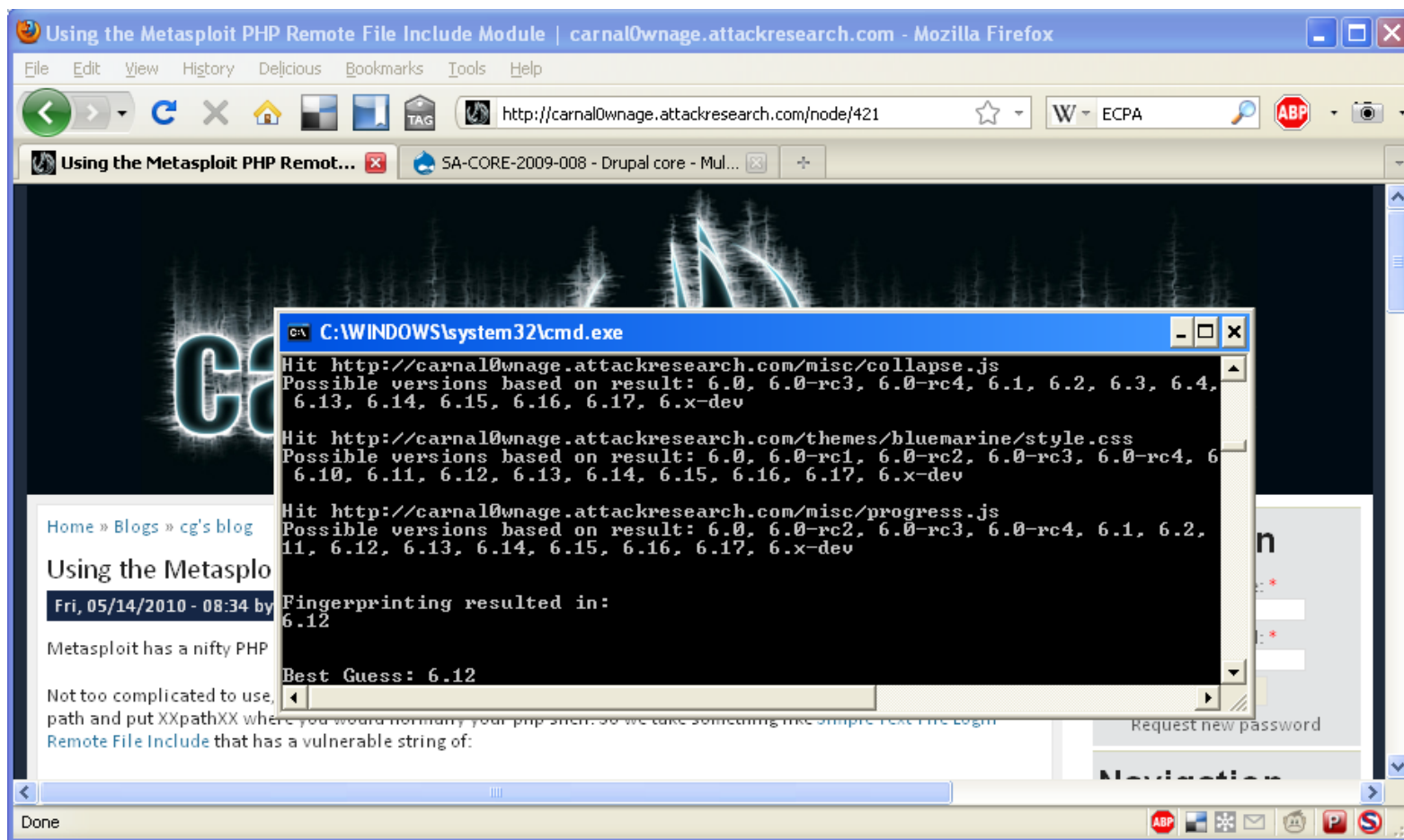
Done

WordPress

Sorry Guys...



Sorry Guys...



Sorry Guys...



Sorry Guys...



SA-CORE-2009-008 - Drupal core - Multiple vulnerabilities | drupal.org - Mozilla Firefox

File Edit View History Delicious Bookmarks Tools Help

http://drupal.org/node/579482

Using the Metasploit PHP Remote File I... SA-CORE-2009-008 - Drupal core ...

An implementation error allows a user to access the account of another user when they share the same OpenID 2.0 provider.

This issue affects Drupal 6.x only.

Wha-whaaaaaa

File upload

File uploads with certain extensions are not correctly processed by the File API. This may lead to the creation of files that are executable by Apache. The .htaccess that is saved into the files directory by Drupal should normally prevent execution. The files are only executable when the server is configured to ignore the directives in the .htaccess file.

This issue affects Drupal 6.x only.

Session fixation

Drupal doesn't regenerate the session ID when an anonymous user follows the one time login link used to confirm email addresses and reset forgotten passwords. This enables a malicious user to fix and reuse the session id of a victim under certain circumstances.

This issue affects Drupal 5.x only.

Versions affected

Done

- Contributions
- Play bug bingo!
- Drupal core
- Contributions
- Mailing list archives
- Drupal.org webmasters
- Drupal.org server administrators
- Drupal.org CVS applications
- Web links
 - Planet Drupal
 - Drupal talk
 - Drupal dojo

Observations

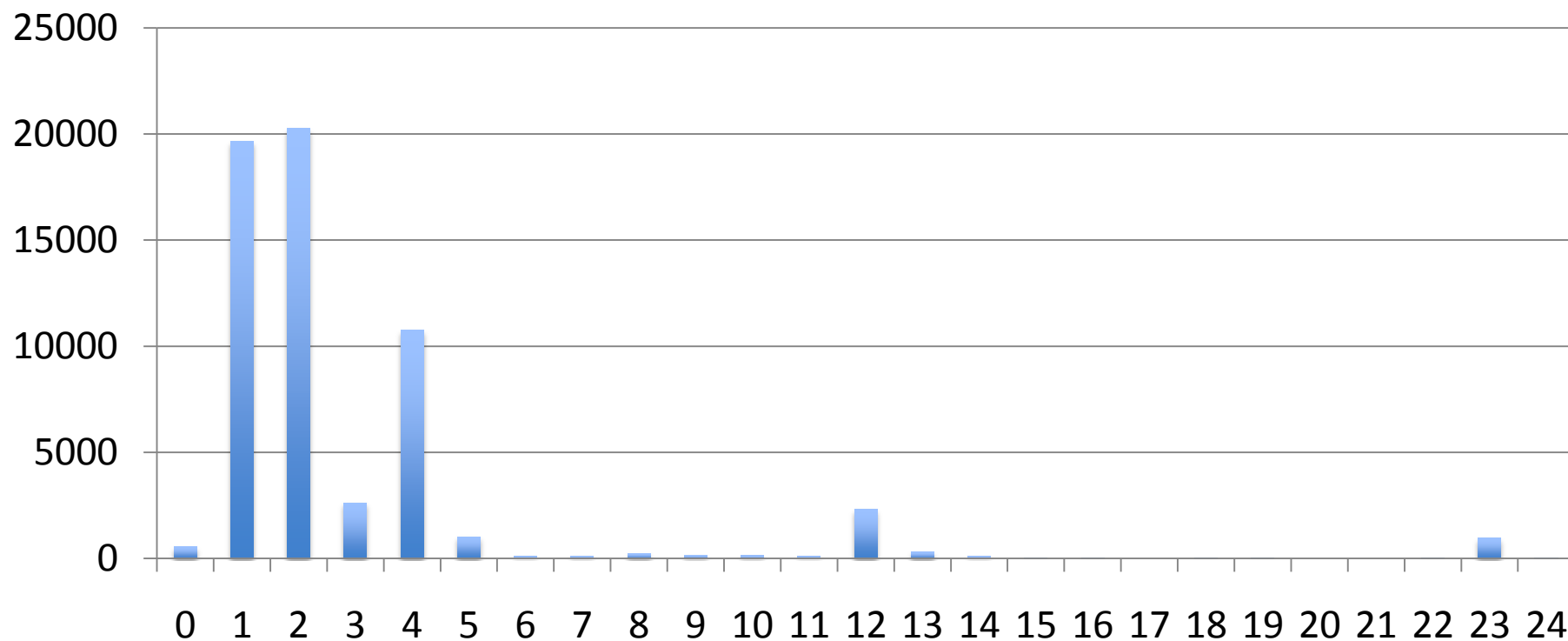


- Webapps actually doing pretty well update-wise
 - ...but not quite good enough
- Huge spike at version provided by package managers and hosting services
 - If you're trusting either to keep you up to date, you're probably behind
- Improperly removed webapps abound
 - Switch from CMS A to CMS B, but leave A lying around
 - Net-visible test/QA sites

Precision



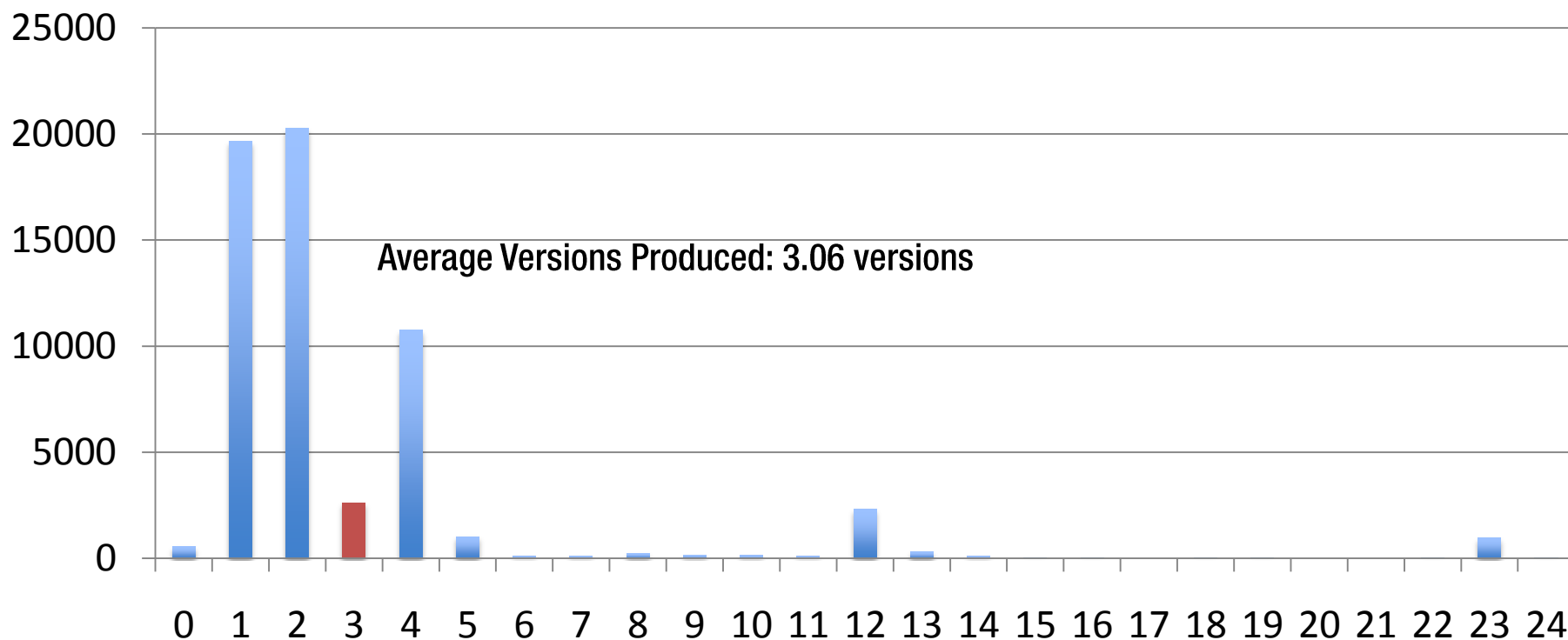
Fingerprint Precision
(# Versions Resulting from a Fingerprint (1 is best))



Precision



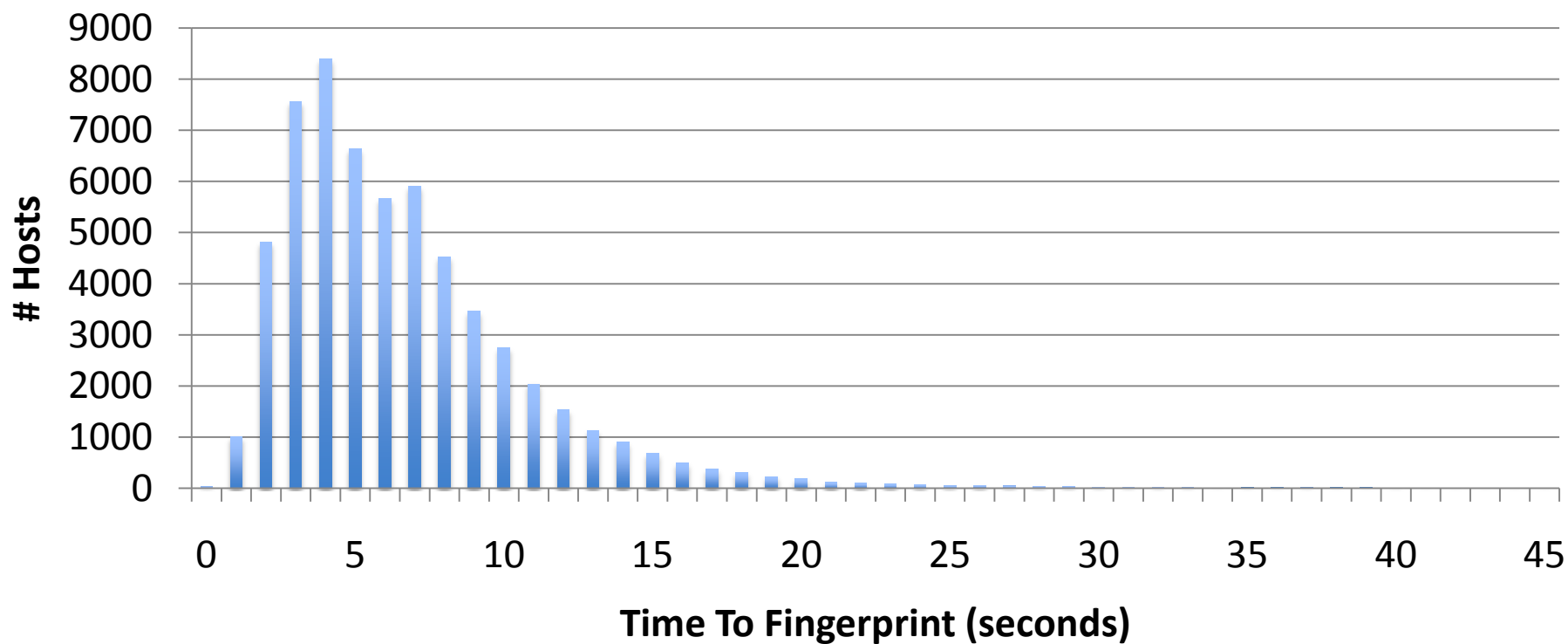
Fingerprint Precision
(# Versions Resulting from a Fingerprint (1 is best))



Speed



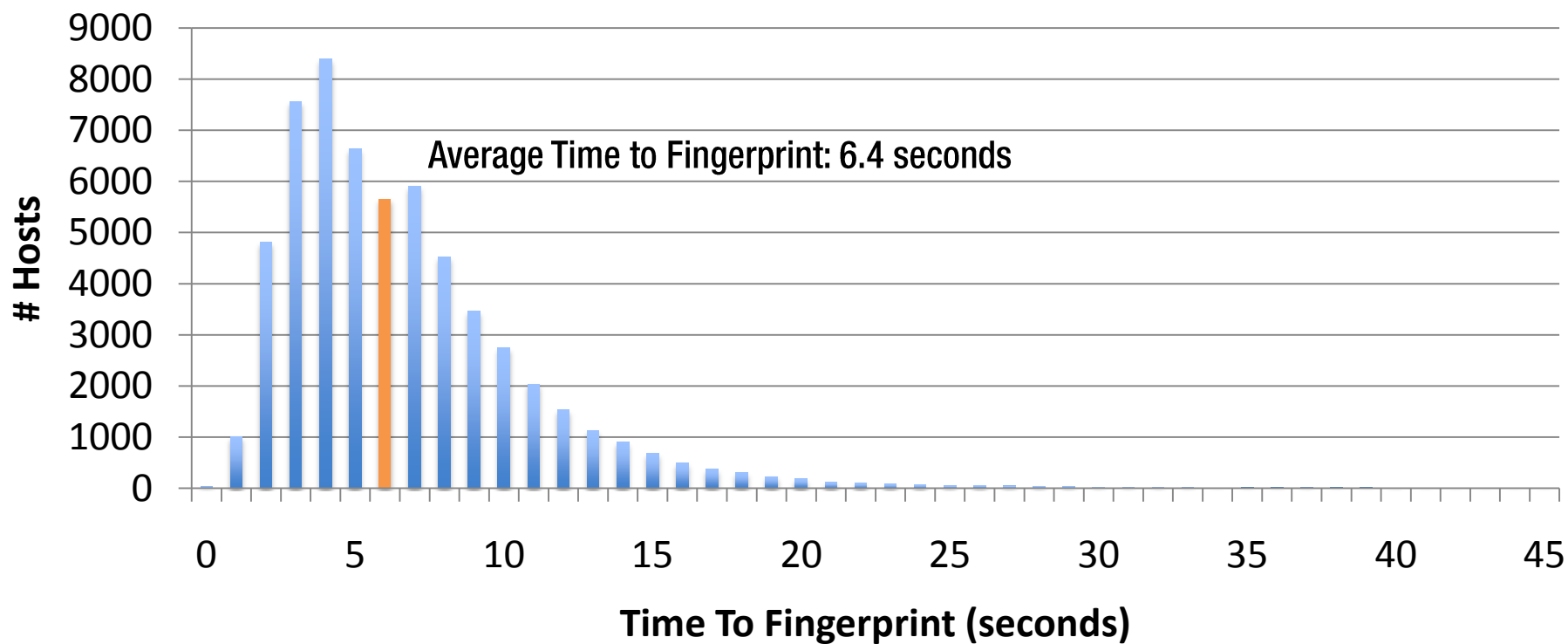
Fingerprinting Time
(Quicker is better)



Speed



Fingerprinting Time (Quicker is better)



BlindElephant Scorecard



- Very Generic
 - Same code for all apps & plugins
- Fast
 - 1-10 sec, based on host (Avg 6.4)
- Low resources
 - Avg 354.2 Kb to fingerprint
- Accurate
 - Avg 3.06 versions & ID 98.0% of sites
- Resistant to hardening/banner removal
 - Yes
- Easy to support new versions/apps
 - ~2 hours to support all available versions of a new app (1 if they're packed nicely)

Sources Of Error



- WebApp Incompletely Removed
- Partial/Manual Upgrades
 - We tend to catch these though
- Changed App Root
- Static hosting on alternate domain (eg, Wikipedia)
- Forked Project (osCommerce, phpNuke)
- Fails completely if static files are trivially modified
 - But guess what? People don't do it (yet)

Release the Kra... Elephant



<http://blindelephant.sourceforge.net/>

To Do



- Web App Developers
 - Help us create fingerprint files to recognize your app!
 - But also think about default deployments that resist fingerprinting
- Site Administrators
 - Fingerprint yourself – know what the attackers know
 - Harden to resist fingerprinting
 - Just... stay up to date
- Everyone Else
 - Try it out
 - Report bugs, contribute signatures, implement a pet feature...

Questions?

pthomas@qualys.com

@coffeetocode





Extras

Theory of Fingerprinting



- Find some characteristic(s) that is...
 - ...always the same for a particular individual (implementation/version/person)
 - ...always different from other members of the population
- If there's one piece of info that fulfills both, great
 - If not, take several that pin it down
 - Tons of [interesting reading in information theory and entropy](#)
- OS & HTTP Server Fingerprinting: Lots of protocol-aware checks that rely on subtle differences in implementation

The Question That Started This All



- What % of (active) sites on the net are running a well-known webapp?

- Not counting Parked/ad-only, down, or blank/40x
- Only examined the root of the domain
- Sample set is from a list of 87M .coms

The Question That Started This All



- What % of active sites on the net are running a well-known webapp?

- 23% Parked
- + 5.8% Ads only
- + 7.9% No Content/40x
- + 13.1% Down
- ~49.7% of the web is *junk**

■ **That's all? Hush you.*

The Question That Started This All



- What % of active sites on the net are running a well-known webapp?

- 4.4% of domains had a supported app
- $\div 503$ percent of domains are “active”
- $\sim 8.8\%$

It Only Goes Up



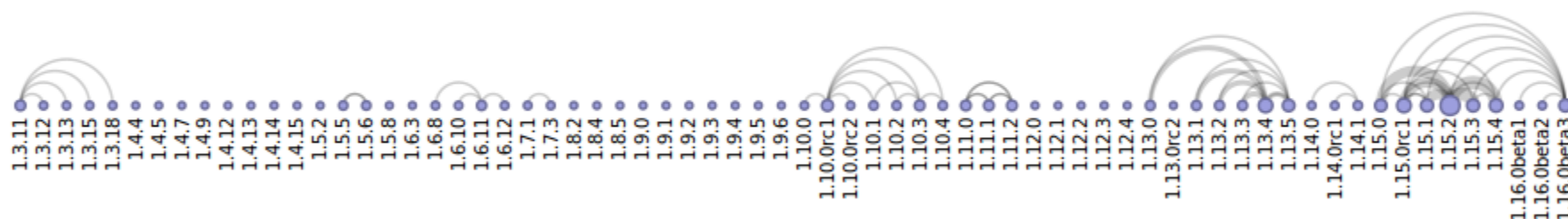
- 8.8% is *definitely* a lower bound
 - Support for more apps
 - Could test /blog, /wiki, /forum and subdomains
 - Improvements in app guessing (was tuned for false negatives)
- What % of web applications are a “well-known” webapp?
 - I don’t know... I’d like to find out though

Beyond Hashing



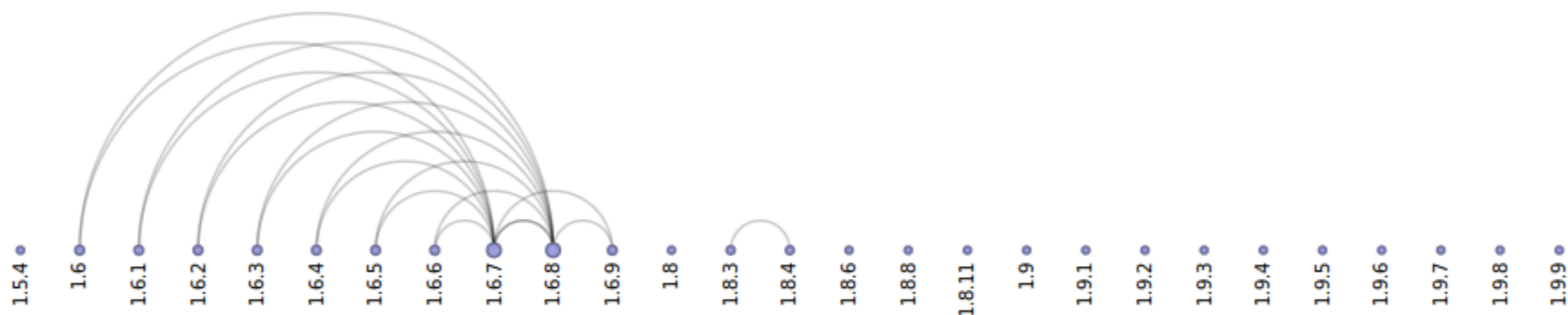
- Nearest neighbor search
- Rolling hashes
- Version trajectory
- Error tolerant hashing?

Version Clusters: Mediawiki



Connections between nodes (versions) indicate that a fingerprint job produced those versions and was unable to differentiate. Thicker connections indicate the number of times this confusion occurred (two particularly difficult to distinguish versions), while many connections among adjacent nodes indicates a family of difficult to distinguish versions.

Version Clusters: Moodle



Connections between nodes (versions) indicate that a fingerprint job produced those versions and was unable to differentiate. Thicker connections indicate the number of times this confusion occurred (two particularly difficult to distinguish versions), while many connections among adjacent nodes indicates a family of difficult to distinguish versions.