Blind Elephant:
Web Application Fingerprinting & Vulnerability Inferencing

Patrick Thomas
Qualys
10/26/10
Outline

• 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
Existing Fingerprinting Approaches

• Basically: Regex
• Human judgment 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…
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

Web App Versions
(eg, Joomla-* .zip)

What versions will a path give me info on?

If I want to confirm or rule out a version/versions, what’s a path that will do that?
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

Best Candidates to Identify the Version

- `/.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/combobox.js`
- `/.language/en-GB/en-GB.mod_search.ini`
- `/.templates/rhuk_milkyway/css/template.css`
- `/.media/system/js/switcher.js`

- 14 hashes/31 versions, fitness=15.0
- 14 hashes/20 versions, fitness=14.64
- 13 hashes/20 versions, fitness=13.64
- 10 hashes/28 versions, fitness=10.90
- 8 hashes/28 versions, fitness=8.90
- 8 hashes/20 versions, fitness=8.64
- 8 hashes/20 versions, fitness=8.64
- 8 hashes/20 versions, fitness=8.64
- 8 hashes/20 versions, fitness=8.64
- 7 hashes/20 versions, fitness=7.64
- 7 hashes/20 versions, fitness=7.64
- 7 hashes/20 versions, fitness=7.64
- 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
- ...

QUALYS

DEFCON 18
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
Fingerprinting

Best Candidates

'/htaccess.txt'
'/language/en-GB/en-GB.ini'
'/language/en-GB/en-GB.com_content.ini'
'/configuration.php-dist',
'/includes/js/joomla.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/combobox.js'
'/language/en-GB/en-GB.mod_search.ini'
'/templates/rhuilmilkyw/css/template.css'
'/media/system/js/switcher.js'

3.0.4-RC4, 3.0.4

2.0.1, 2.0.2...
3.0.4-RC4,
3.0.4

2.5.1, 2.3.16...
3.0.4-RC4,
3.0.4

3.0.4-RC4,
3.0.4, 3.5

3.0.4-RC4,
3.0.4, 3.5.1
Winnowing

3.0.0, 3.0.1
3.0.2, 3.0.3,
3.0.4-RC1,
3.0.4-RC2

Darn, Not Enough Data

(Confirm or rule out versions)
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}

Want a small set of files with at least one present in every release
App Discovery / App Guessing

Indicator Files

```json
{ "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`

- Hit `http://laws.qualys.com/mt-static/js/tc/client.js`

- Hit `http://laws.qualys.com/mt-static/css/main.css`

- File produced no match. Error: Error code: 404 (Not Found)
Does it work?

  

  

  

  
Does it work?


- Hit [http://laws.qualys.com/mt-static/js/tc/focus.js](http://laws.qualys.com/mt-static/js/tc/focus.js)

Interlude

This is what matters!
Does it work?

- Hit http://laws.qualys.com/mt-static/css/simple.css

- Hit http://laws.qualys.com/mt-static/mt JA.js

- Hit http://laws.qualys.com/mt-static/js/tc/gestalt.js

- 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! Let's Play

- App ID & Fingerprinting on 1,084,152 hosts

- \textbf{\~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)

- \textbf{\~50k} and \textbf{\~1M} host random sample of 87M .com domains
  - Stats on accuracy and net-wide webapp population are from these
Version Distribution: Drupal
(June 18, 2010)

Affected by A Critical Vulnerability: 69%
Version Distribution: Joomla
(June 18 2010)

Affected by A “High” Vulnerability: 91%
Version Distribution: MovableType
(June 18, 2010)

Affected by a Moderate Vulnerability: 77%
Affected by a Critical Vulnerability: 57%
Precision

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

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

Average Versions Produced: 3.06 versions
Fingerprinting Time
(Quicker is better)

# Hosts

Time To Fingerprint (seconds)
Fingerprinting Time
(Quicker is better)

Average Time to Fingerprint: 6.4 seconds
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://blinđelephant.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
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.
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.
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.
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
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
  - 503 percent of domains are “active”
  - ~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.
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.