Next generation web scanning
New Zealand: A case study
First presented at KIWICON III 2009

By Andrew Horton
aka urbanadventurer
NZ Web Recon

Goal: To scan all of New Zealand's web-space to see what's there.

Requirements:
– Targets
– Scanning
– Analysis

Sounds easy, right?
Targets

What does 'NZ web-space' mean?

It could mean:
• Geographically within NZ regardless of the TLD
• The .nz TLD hosted anywhere
• All of the above

For this scan it means, IPs geographically within NZ
Finding Targets

We need creative methods to find targets
DNS Zone Transfer

```bash
$ host -al nz
nz AXFR record query refused by ns2.dns.net.nz
nz AXFR record query refused by ns3.dns.net.nz
nz AXFR record query refused by ns7.dns.net.nz
nz AXFR record query refused by ns4.dns.net.nz
nz AXFR record query refused by ns5.dns.net.nz
nz AXFR record query refused by ns1.dns.net.nz
nz AXFR record query refused by ns6.dns.net.nz
No nameservers for nz responded

$ dig nz axfr

; <<>> DiG 9.5.1-P2 <<>> nz axfr
;; global options: printcmd
; Transfer failed.
```
Find IP addresses on IRC and by resolving lots of NZ websites

729,580,500 IPs. More than we want to try.
## IP address blocks in the IANA IPv4 Address Space Registry

<table>
<thead>
<tr>
<th>Prefix</th>
<th>Designation</th>
<th>Date</th>
<th>Whois</th>
<th>Status [1]</th>
</tr>
</thead>
<tbody>
<tr>
<td>000/8</td>
<td>IANA - Local Identification</td>
<td>1981-09</td>
<td></td>
<td>RESERVED</td>
</tr>
<tr>
<td>001/8</td>
<td>IANA</td>
<td></td>
<td></td>
<td>UNALLOCATED</td>
</tr>
<tr>
<td>002/8</td>
<td>RIPE NCC</td>
<td>2009-09</td>
<td>whois.ripe.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>003/8</td>
<td>General Electric Company</td>
<td>1994-05</td>
<td></td>
<td>LEGACY</td>
</tr>
<tr>
<td>201/8</td>
<td>LACNIC</td>
<td>2003-04</td>
<td>whois.lacnic.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>202/8</td>
<td>APNIC</td>
<td>1993-05</td>
<td>whois.apnic.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>203/8</td>
<td>APNIC</td>
<td>1993-05</td>
<td>whois.apnic.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>204/8</td>
<td>ARIN</td>
<td>1994-03</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>205/8</td>
<td>ARIN</td>
<td>1994-03</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>206/8</td>
<td>ARIN</td>
<td>1995-04</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>207/8</td>
<td>ARIN</td>
<td>1995-11</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>208/8</td>
<td>ARIN</td>
<td>1996-04</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>209/8</td>
<td>ARIN</td>
<td>1996-06</td>
<td>whois.arin.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>210/8</td>
<td>APNIC</td>
<td>1996-06</td>
<td>whois.apnic.net</td>
<td>ALLOCATED</td>
</tr>
<tr>
<td>211/8</td>
<td>APNIC</td>
<td>1996-06</td>
<td>whois.apnic.net</td>
<td>ALLOCATED</td>
</tr>
</tbody>
</table>

This list has 663,255,000 IPs. More than we want to try.
Failed methods to find targets

- DNS Zone transfers from top level domain name servers
- Learn IP address ranges for well known national websites and networks
- All IP addresses allocated to APNIC (Asia Pacific NIC)

We need new methods to find IP addresses and website hostnames for New Zealand
geoipgen and the MaxMind GeoIP database

Use MaxMind’s free database of IP to Country allocations
Homepage: www.morningstarsecurity.com/research/geoipgen

Produces 6,319,348 New Zealand IP addresses
Scanning for TCP Port 80 with nmap

Find the 75,964 web servers among 6 million IPs

```
nmap -i ./iplist -P0 -sT --open -n -p 80 --oG iplist.gnmap.log
```
Reverse Resolving IP addresses

Use adns-tools for fast, asynchronous resolving

```bash
~/projects/nzwide-whatweb$ cat iplist.port80.log | adnslogres | grep "^[a-z]"
122-57-247-79.jetstream.xtra.co.nz
dbsys2.digiweb.net.nz
222-152-235-159.jetstream.xtra.co.nz
118-92-112-95.dsl.dyn.ihug.co.nz
210-54-240-196.ipnets.xtra.co.nz
vcenter.vmware.solarix.net.nz
210-54-241-165.ipnets.xtra.co.nz
118-92-189-178.dsl.dyn.ihug.co.nz
60-234-220-15.bitstream.orcon.net.nz
203-114-179-201.dsl.sta.inspire.net.nz
125-239-232-161.jetstream.xtra.co.nz
h241-245.catalyst.net.nz
ns1.marketpulse.net.nz
ip-118-90-29-161.xdsl.xnet.co.nz
ip-119-47-113-133.cust.openhost.net.nz
```

31,973 IPs are resolved to hostnames
Search query ip:210.48.71.196

11,872 IPs are indexed by bing.com which have 89,265 virtual hosts.

urbanadventurer (Andrew Horton)
www.morningstarsecurity.com
Usage: ./gggooglescan [OPTION]... <QUERY>
by Andrew Horton (urbanadventurer)
-c=CC  Search within a country, eg. au, uk or nz
-d=NUM  Depth of results, 0 = 1st page, 1 = 2nd page. Default: 5
-g=IP  IP or hostname of a Google search appliance. Default: 210.55.180.157
-l=FILE Log file, output is appended if the file already exists
-o  Only print hostnames, not urls
-v  Verbose output

:-/projects/nzwide-whatweb$ ./gggooglescan -c nz -d 2 -o kiwicon
wellington.geek.nz
wellington.geek.nz
computerworld.co.nz
computerworld.co.nz
atta.cked.me
pressf1.pcworld.co.nz
coffee.geek.nz
www.trademe.co.nz
pressf1.co.nz
www.geekzone.co.nz
There is a common misconception that Google scraping is no longer possible and is halted by Google’s bot detection.

It is possible to search for a wide set of search terms and to retrieve a shallow set of the each result, i.e. 3 pages.

searching aaa through to zzz found 58,602 hostnames

searching every word in /usr/share/dict/words found 116,052 hostnames

126,408 unique NZ hostnames found with Google
### DNS Zone Transfers Revisited

```plaintext
; DiG 9.5.1-P2 @dns1.canterbury.ac.nz canterbury.ac.nz axfr
; (1 server found)
; --- global options: printcmd

canterbury.ac.nz. 86400 IN SOA dns1.canterbury.ac.nz. soa.canterbury.ac.nz. 2009111001 1800 3600 604800 86400

canterbury.ac.nz. 86400 IN NS dns1.canterbury.ac.nz.
canterbury.ac.nz. 86400 IN NS dns2.canterbury.ac.nz.
canterbury.ac.nz. 86400 IN MX 10 mx1.canterbury.ac.nz.
canterbury.ac.nz. 86400 IN MX 10 mx2.canterbury.ac.nz.
canterbury.ac.nz. 86400 IN TXT "University of Canterbury, Christchurch."
!webmail.canterbury.ac.nz. 86400 IN CNAME ucatmail2.canterbury.ac.nz.
*.canterbury.ac.nz. 86400 IN MX 10 mx1.canterbury.ac.nz.
*.canterbury.ac.nz. 86400 IN MX 10 mx2.canterbury.ac.nz.
sipfederationtls._tcp.canterbury.ac.nz. 86400 IN SRV 10 10 5061 sip.canterbury.ac.nz.canterbury.ac.nz.
sip._tls.canterbury.ac.nz. 86400 IN SRV 10 10 443 sip.canterbury.ac.nz.canterbury.ac.nz.
acad-hsm40.canterbury.ac.nz. 86400 IN A 132.181.223.83
acad-lvo11.canterbury.ac.nz. 86400 IN A 132.181.223.146
access.canterbury.ac.nz. 86400 IN A 132.181.106.24
accom.canterbury.ac.nz. 86400 IN A 132.181.2.10
acis.canterbury.ac.nz. 86400 IN MX 10 mx1.canterbury.ac.nz.
acis.canterbury.ac.nz. 86400 IN MX 10 mx2.canterbury.ac.nz.
www.acis.canterbury.ac.nz. 86400 IN A 132.181.190.13
www.adulteducation.canterbury.ac.nz. 86400 IN A 132.181.153.124
afis.canterbury.ac.nz. 86400 IN A 132.181.190.1
afis.canterbury.ac.nz. 86400 IN MX 10 mx1.canterbury.ac.nz.
afis.canterbury.ac.nz. 86400 IN MX 10 mx2.canterbury.ac.nz.
student.afis.canterbury.ac.nz. 86400 IN A 132.181.190.1
```
DNS Zone Transfers Revisited

Extracting domainnames

```bash
~:/projects/nzwide-whatweb$ ./basedomainname
basedomainname 0.1 by Andrew Horton (urbanadventurer) www.morningstarsecurity.com
Usage: basedomainname [-h|--help]|[-tld|--ext|--domain|--host] [-i <input-file>]
If input-file is not specified it reads from STDIN

Examples: --tld
www.wolves.mobi => mobi, www.panda.cn => cn
Examples: --ext
www.mice.co.uk => co.uk, k.iwi.nz => iwi.nz
Examples: --domain
www.dogs.co.nz => dogs.co.nz, dev12.wlg.cats.com => cats.com
Examples: --host
www.kangaroo.com.au => www, nose.shark.int => nose

~:/projects/nzwide-whatweb$ cat hostnames_combined_nz.log | ./basedomainname --domain | more
e-xpert.co.nz
e-xpert.co.nz
e-xpert.co.nz
e-xpert.co.nz
007films.com
01b.co.nz
01.co.nz
01dev.co.nz
01.net.nz
021builder.co.nz
jordansurfshapes.co.nz
021extras.com
040trainsnmodes.co.nz
0508pizza.com
0508pizzas.com
```
DNS Zone Transfers Revisited

Results

Attempt a DNS zone transfer for each domain

135,591 unique domain names were found with reverse resolving IPs, Bing, and Google scanning.

Tool: dns-enum.pl

Found 560,352 hosts in 70,475 domains.
DNS Brute Forcing – Not Implemented

Guessing subdomains, eg. test.example.com, www2.example.com, intranet.example.com
Final Target List

- 699,413 unique hostnames found with reverse resolving, Google, Bing and zone transfers
- Resolve the hostnames to IPs
- Keep only the hostnames with IPs in the port scanned list of 75,964 IPs found with nmap
- 75,964 IPs + 274,989 hostnames = 350,953 virtual hosts to test
Scanning

– Targets
– Scanning
– Analysis
Traditional Web Scanners

Nikto and Nessus

- Time. Nikto takes too long because it guesses 1000s of URLs.
- Impolite. Nikto has a big footprint with 1000s of lines in each web servers logs and it increases web server load.
- Law. Some Nikto tests will attempt to exploit vulnerabilities so it is not suitable for use without permission.
- Information. Pretty good

Nmap

- Time. Nmap is fast
- Impolite. Nmap is polite, it makes only a few connections
- Law. Unquestionable
- Information. Scarce
WhatWeb

• Time. Fast
• Polite. Doesn't trigger NIDS
• Law. Unquestionable
• Information. Rich

• Instead of guessing URLs to identify systems, make better use of the information provided by the web server during an HTTP transaction.
WhatWeb

Discover what powers websites by identifying:

• content management systems (CMS)
• blogging platforms
• stats/analytics packages
• javascript libraries
• HTTP servers

• Written in Ruby for Linux
• OpenSource License
• Plugin architecture
WhatWeb

• Passive and aggressive plugins
• Passive plugins use information from:
  – The HTML page
  – HTTP headers
  – Cookies
  – URL
• Lightweight like a search engine crawler
• A single GET / HTTP/1.0 request
WhatWeb

• Aggressive plugins use information from:
  – Testing for URLs and identifying patterns in the HTML
  – Testing for URLs and recognising the MD5 hash of the response
  – Testing for URLs and simply noting they exist or return an HTTP status 200 code.

• Can return an exact version of a CMS, can discover installed modules or plugins

• Uses multiple HTTP requests
WhatWeb

`~/projects/whatweb$ ./whatweb`

WhatWeb - Discover what powers websites.
Version 0.3 by urbanadventurer (Andrew Horton)
Usage: whatweb [options] <URLs>

---input-file=FILE, -i
Identify URLs found in FILE

---aggression, -a
1 passive - on-page
2 polite - follow on-page links if in the extra-urls list (default)
3 impolite - try extra-urls when plugin matches (smart, guess a few urls)
4 aggressive - try extra-urls for every plugin (guess a lot of urls)

---recursion, -r
Follow links recursively. Only follows links under the path (default: off)

---depth, -d
Maximum recursion depth (default: 3)

---max-links, -m
Maximum number of links to follow on one page (default: 25)

---list-plugins, -l
List the plugins

---run-plugins, -p
Run comma delimited list of plugins. Default is to run all

---info-plugins, -I
Display information about a comma delimited list of plugins. Default is all

---example-urls, -e
Add example urls for each plugin to the target list

---color=[WHEN], --color=[WHEN]
Control whether color is used. WHEN may be `never`, `always`, or `auto`

---log-full=FILE
Log verbose output

---log-brief=FILE
Log brief, one-line output

---user-agent, -U
Identify as user-agent instead of WhatWeb/VERSION.

---max-threads, -t
Number of simultaneous threads identifying websites in parallel (CPU intensive).

Default is 5.

---help, -h
This help

---verbose, -v
Increase verbosity (recommended), use twice for debugging.
WhatWeb Examples

```bash
:~/projects/whatweb$ . /whatweb research.elabs.govt.nz
a8246faa8a17de13d57789cc90], meta-generator[WordPress 2.5.1], server-hea
der[Apache/2.2.9 (Debian) PHP/5.2.6-1+lenny3 with Suhosin-Patch mod_ssl/
2.2.9 OpenSSL/0.9.8g], title[Research e-Labs &raquo; web trends, open
source and technology in government], uncommon-headers[x-pingback], x-p
owered-by-header[PHP/5.2.6-1+lenny3]
```
Passive & Aggressive Tests

With aggressive tests it identifies the Joomla CMS version by retrieving a handful of URLs and recognising the MD5 hashes.
Aggressive Tests

phpBB forum

```
~:/projects/whatweb$ ./whatweb forum.letterboxer.org.nz
http://forum.letterboxer.org.nz [200] md5[9cfae166b2b4db6c6aad8e9da9613
ee], phpBB[3], server-header[Apache], title[forum.letterboxer.org.nz &bu
ll; Index page]
~:/projects/whatweb$
~:/projects/whatweb$ ./whatweb -a 3 forum.letterboxer.org.nz
http://forum.letterboxer.org.nz [200] md5[336067e5c258b61448244632b14972
e7], phpBB[3,3.0.4], server-header[Apache], title[forum.letterboxer.org.
nz &bull; Index page]
~:/projects/whatweb$
```

/docs/CHANGELOG.html
## Plugins available

<table>
<thead>
<tr>
<th>Acclipse</th>
<th>Advanced-Guestbook</th>
<th>BlogSmithMedia</th>
<th>Blogger</th>
<th>DiBos</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drupal</td>
<td>EarlyImpact-ProductCart</td>
<td>Echo</td>
<td>GoAhead-Webs</td>
<td>Google-Analytics-GA</td>
</tr>
<tr>
<td>Google-Analytics-urchin</td>
<td>IIS-SiteNotFound</td>
<td>IIS-UnderConstruction</td>
<td>ISP-Config</td>
<td>Jquery</td>
</tr>
<tr>
<td>Joomla</td>
<td>Lightbox</td>
<td>Mailto</td>
<td>Mambo</td>
<td>Minify</td>
</tr>
<tr>
<td>Moodle</td>
<td>MovableType</td>
<td>NovellGroupwise</td>
<td>OSCommerce</td>
<td>Oce</td>
</tr>
<tr>
<td>Plesk</td>
<td>Plone</td>
<td>Prototype</td>
<td>Quantcast</td>
<td>Scriptaculous</td>
</tr>
<tr>
<td>Siemens-SpeedStream-Router</td>
<td>TypePad</td>
<td>VSNS-Lemon</td>
<td>Windows-SBS</td>
<td>WordPress</td>
</tr>
<tr>
<td>WordPressSpamFree</td>
<td>Antiboard</td>
<td>apache-default</td>
<td>asp-nuke</td>
<td>belkin-modem</td>
</tr>
<tr>
<td>bing-searchengine</td>
<td>citrix-metaframe</td>
<td>Comersus</td>
<td>Coppermine</td>
<td>Cpanel</td>
</tr>
<tr>
<td>Formmail</td>
<td>index-of</td>
<td>invasion-power-board</td>
<td>ispCP-omega</td>
<td>mailsite-express</td>
</tr>
<tr>
<td>Md5</td>
<td>meta-generator</td>
<td>mnoGoSearch</td>
<td>oki-pbx</td>
<td>php-cake</td>
</tr>
<tr>
<td>phpBB</td>
<td>redirect-location</td>
<td>server-header</td>
<td>snom-phone</td>
<td>Title</td>
</tr>
<tr>
<td>toshiba-printer</td>
<td>uncommon-headers</td>
<td>Vbulletin</td>
<td>vp-asp</td>
<td>Webguard</td>
</tr>
<tr>
<td>x-asnet-version-header</td>
<td>x-powered-by-header</td>
<td>xtra-business-hosting</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Making Plugins is Easy

urbanadventurer (Andrew Horton)
www.morningstarsecurity.com

```python
Plugin.define "Plone" do
author "Andrew Horton"
version "0.1"
description "CMS http://plone.org"
examples %w| www.norden.org www.trolltech.com www.plone.net www.smeal.psu.edu|

matches [
	{:name=>"meta generator tag",
	:probability=>100,
	:regexp=>/<meta name="generator" content="[^>]*http://\$/plone.org" \/>/},

	{:name=>"plone.css",
	:probability=>100,
	:regexp=>/\(/igkeit url|text|css\["]*/porta\_css\./.*plone.*css\(/\")/},

	{:name=>"plone.javascript",
	:probability=>100,
	:regexp=>/\["\] *ploneScripts\[0-9]+.js\"/},

	{:name=>"div class="visualIcon contenttype=plone-site"",
	:probability=>100,
	:regexp=>/<div class="visualIcon contenttype=plone-site"/>},

	{:name=>"div tag, visual-portal-wrapper",
	:probability=>75,
	:regexp=>/<div id="visual-portal-wrapper"/>},
]}

def passive
m=[]
X-Caching-Rule-Id: plone-content-types
X-Cache-Rule: plone-content-types
m = [{'name="X-Caching-Rule-Id: plone-content-types", :probability=>100 } if @meta["x-caching-rule-id"] =~ /plone-content-types/i
m = [{'name="X-Cache-Rule: plone-content-types", :probability=>100 } if @meta["x-cache-rule"] =~ /plone-content-types/i
m
end
```
The Scan
Analysis – What did I find?

- Targets
- Scanning
- Analysis
TLDs & SLDs hosted within NZ

Extn

- co.nz
- net.nz
- org
- govt.nz
- info
- tv
- iwi.nz
- travel
- ws
- name
- us
- asn.au
- ca
- com
- school.nz
- gen.nz
- geek.nz
- co.uk
- cri.nz
- eu
- si
- mobi
- com.fj
- nl
- nu
- org.nz
- ac.nz
- com.au
- biz
- maori.nz
- net.au
- org.au
- cc
- mil.nz
- co.za
- me
- aero
- to
NetCraft’s Top HTTP Servers

What I expected to find

Count

Apache
Microsoft
qq.com
nginx
Google
lighttpd

urbanadventurer (Andrew Horton)
www.morningstarsecurity.com
Top 10 HTTP Server Versions
What I found

Count

- Apache
- Microsoft-IIS
- cpanel
- WebServerX
- Netscape-Enterprise
- ZealdWeb
- NOYB
- Sun Java
- Zeus
- Lotus-Domino
Scale is based on the average worldwide traffic of *joomla* in the last 12 months. Learn more.
CMS Showdown

NZ Results

Count

- Joomla
- Wordpress
- Drupal
- Plone
- Movable Type

urbanadventurer (Andrew Horton)
www.morningstarsecurity.com
WordPress

- Is WordPress representative of other CMS’s?
- 89% are not patched and up to date. < 2.8.6
- 53% are at high risk of exploitation. <= 2.8.2
- An internet worm is currently exploiting WordPress installations with versions of 2.8.2 and prior.

http://www.securityfocus.com/bid/27669/info
What else is on the web?

• Websites but not as you know them
• Web interfaces to cameras, printers, phones, etc.
• Many of these devices should not be available through websites on public, internet IP addresses
• Insecure vs Unsecured. Many devices are not protected by any authentication mechanism

• This presentation contains a subset of the screenshots in the full presentation
Cameras
Printers

Xerox 9700 [1977]
Warning: A change in the IP Address will result in loss of connectivity to the browser.

IPv4 Configuration

**IP Configured By:** DHCP
**Host Name:** NP17B7699
**Domain Name:**

**IP Address Configuration**

- **Automatic IP**
  - DHCP
  - BootP
  - AutoIP

**Manual IP Address:** 192.168.15.102
**IP Subnet Mask:** 255.255.255.0
**Manual Default Gateway:** 192.168.15.254
Phones
TV devices
GoAhead

- Don’t judge a website by it’s HTTP Server Name
- Many different types of devices are powered by the GoAhead embedded HTTP server.
- *Most* of the following devices are shown to display the variation of devices, not because they have a lack of authentication.
Welcome to the web interface for the GRX1200 GG Pro
Heavy Equipment

- Air conditioning
- Industrial process sensors
- Data centres
Analysis Notes

• A high percentage of content management system websites are insecure due to poor updating. 53% are at high risk.

• Unsecured devices discovered include cameras, printers, phones, TV units, intranets (not shown in this version of the slides), air conditioning systems and industrial process sensors. These should be behind a firewall or secured with a password.
Tools Used

• Nmap – Network scanner.
  – Used to port scan to test IPs for web servers on TCP port 80

• Dnsenum – DNS enumeration
  – Used to execute zone transfers

• adns-tools
  – Used for fast reverse DNS resolving

• Geoipgen
  – Used to produce a near complete set of IP addresses in New Zealand. This is a MorningStar Security tool.
New Tools Developed

• **WhatWeb**
  – Used to identify websites with a light scan

• **Gggooglescan**
  – Find website hostnames by searching with Google.
  – Scan wide and shallow.

• **bing-ip2hosts**
  – Find all websites indexed by Bing on NZ IP addresses

• **Basedomainname**
  – Used to extract the domainnames of hostnames

Download these tools from [www.MorningStarSecurity.com](http://www.MorningStarSecurity.com)
53% at high risk of exploitation. WTF?

Check out www.morningstarsecurity.com for your freshest blend of IT security news each morning