2012 internet census

jeff crowell
cs558 – spring 2013
heatmap of all 460,000,000 ipv4 addresses

source: 2013 census
history
- 1988 First Internet Worm
- Morris Worm
- Goal => Gauge Size of the Internet
- Exploited debug hole in UNIX sendmail
- Infected 6,000 Machines
- Punishment
  - 3 years probation
  - 400 Hours Community Service
  - $10,000 Fine
2012 carna botnet

- Goal => map all ipv4 IP addresses
- A “last chance” at mapping all IP before IPv6 becomes standard
  - Ipv6 = 340 Sextillion addresses
  - IPv4 = “only” 3,706,452,992 public addresses
“be nice”

- Be as unobtrusive as possible
- Non-persistant
- Low priority
- Don’t capture traffic 😊
- “respect privacy”
how it works

- Part 1: scans given address space for “open” telnet connections
  - root:root, admin:admin, ...
- Find machine? Great! Drop Listener binary
- Part 2: Scanner manager, sends ip ranges to be scanned and uploads scan results to specified IP address. Stopped deployment after 30k machines.

infrastructure

- Unlike most botnets, not C&C, directly accessible from internet.
- “Middle Nodes”
  - Most powerful devices take client data, store it for the master server
- Each node gets “part id” “starting ip” “stepwidth” and “end ip” for coordination, addresses broken up to 240k jobs, each with 15k addresses
Software

- Binary - Portable! : 46-90kb (SMALL)
  - 9 Architectures, ARM/MIPS/x86/others
- Backend API
  - API, called by Python scripts
  - Web interface PHP
- Database
  - BIG DATA
  - Hadoop/PIG -> MapReduce
- No source released
Scans

- ICMP – *faaaaaaaaaast* – probe ipv4 in under a day
  - 52billion pings

- Reverse DNS
  - Who has <IP ADDRESS> to biggest 16 DNS Servers (Google, Level3,...) 10.5b records

```
{13-04-10 20:00} lostwoods:~ jeff% host 168.122.193.53
53.193.122.168.in-addr.arpa domain name pointer park509-0b01-dhcp53.bu.edu.
```

- Nmap
  - Heavier than ping/dns, only on the more powerful MIPS machines
  - Syn scan of top 100 ports, 85 service probes
  - Service Probes

- Traceroute
  - Targets can’t run linux/no shell, could only do ping/traceroute
  - Small, limited resources, not that useful.

```
{13-04-10 19:45} engridl:~ croweli% traceroute raxcity.com
traceroute to raxcity.com (168.122.193.53), 30 hops max, 40 byte packets
1  cumm024-0b08net-gw.bu.edu (128.197.115.1) 1.420 ms 1.405 ms 1.392 ms
2  comm595-core-aca01-gi2-2-cumm024-dist-aca01-gi5-2.bu.edu (128.197.254.205) 1.232 ms 1.224 ms 1.232 ms
3  comm595-core-aca01-gi2-2-cumm024-dist-aca01-gi5-2.bu.edu (128.197.254.74) 1.576 ms 1.589 ms 1.597 ms
4  park520-dist-res01-gi5-2-comm595-core-res01-gi2-4.bu.edu (128.197.254.246) 1.200 ms 1.246 ms 1.252 ms
5  park509-0b01-dhcp53.bu.edu (168.122.193.53) 0.991 ms 1.096 ms 1.324 ms
```
nmap -T4 -A -v raxcity.com

Retrying OS detection (try #2) against raxcity.com (168.122.193.53)
Initiating Traceroute at 20:17
Completed Traceroute at 20:17, 3.03s elapsed

NSE: Script scanning 168.122.193.53.
Initiating NSE at 20:17
Completed NSE at 20:18, 30.13s elapsed

Nmap scan report for raxcity.com (168.122.193.53)
Host is up (0.000s latency).
rDNS record for 168.122.193.53: park599-0b01-dhcp53.bu.edu

Not shown: 996 filtered ports

<table>
<thead>
<tr>
<th>PORT</th>
<th>STATE</th>
<th>SERVICE</th>
<th>VERSION</th>
</tr>
</thead>
<tbody>
<tr>
<td>22/tcp</td>
<td>open</td>
<td>ssh</td>
<td>OpenSSH 6.0p1 Debian 4 (protocol 2.0)</td>
</tr>
<tr>
<td>80/tcp</td>
<td>open</td>
<td>http</td>
<td>Apache httpd 2.2.22 ((Debian))</td>
</tr>
<tr>
<td>9001/tcp</td>
<td>open</td>
<td>tor-orport</td>
<td>Miniserv 1.620 (Webmin httpd)</td>
</tr>
<tr>
<td>10000/tcp</td>
<td>open</td>
<td>http</td>
<td>MiniServ 1.620 (Webmin httpd)</td>
</tr>
</tbody>
</table>

ERROR: Failed to get host information from server
Interesting Stats
- .NET = Most popular TLD for reverse DNS RRs
- Apache holds 20% of web servers in the world on Port 80
- HP LaserJet P2055 ~2.71% of all Web Connected Printers
- How Big?
  - 420M responded to pings
  - 36M with open ports that did not respond to pings
  - 450M “Definitely” in use
  - 141M closed ports/no ping, firewalled ranges, unknown if computer
  - 591M “in use”
  - 729M only have reverse DNS records, no probe response
- Data is free to analyze ~1TB
Geolocation of Ips from maxmind.com database