Interesting DNS Characteristics

Curtis R. Taylor
CS535
December 9, 2013
Outline

• Motivation and Goal
• Selected Related Work
• Approach
  – Zone files (nameserver data)
    • Offline data linking
  – DNS queries
    • PlanetLab
      – Online and offline collecting and linking
• Results and Discussion
• Lessons learned
Motivation and Goals

• DNS is ubiquitous
  – Performance is important
    • Recall redundancy paper

• Find characteristics about DNS that may lead to interesting research questions or uncover areas for improvement.
  – Track a domain’s authoritative nameserver (NS) over a year
  – Perform queries and examine responses
Selected Related Work

• Shaikh et al. looked at reduced TTLs values and using DNS to approximate geographical location [1].

• Cranor et al. used DNS traffic from backbone routers to try and identify DNS participants e.g., client, resolver, authoritative servers [2].

• Jung et al. found that reducing TTLs had little adverse effect on cache hit rates [3].

• Shue et al. used DNS characteristics to passively link clients to their DNS resolvers [4].
Approach – Tracking (1)

• TLD zone files for over a year (archived each day)
  – (.com, .org, .travel,) .net, .name, .info, etc.
    • Choice based on size

• Information in zone files
  – Domain, authoritative NS domain, IP of NS
    • Not in that order
Approach – Tracking (2)

• Un-archive all zone files
  – How long to track?
    • 1 year (weeks) – Sept. 1, 2012 – Aug. 31, 2013

• Link domain, NS, IP together
  – Reduces number of lookups
  – Cannot link separately (travel needs .com)
  – A script from fellow Grad student will link a given day
  – Time: 4.26 days (12 cores, 64GB RAM)
  – Data: ~44GB/day → 2.4TB total
  – Total domains: ~375 million
Approach – Tracking (3)

• Randomly sort domains and choose 15k from each TLD (45k total) on Sept. 1

• Store domain, NS, and IP
  – Domains may have multiple NS

• For each week, find domain, find NS, check IP
  – Each day’s file ~20GB
Approach – Queries (1)

- Choose another random 15k domains from each TLD (45k total)
  - Chosen from Aug. 31, 2013

- For each TLD
  - Ask NS for A record of domain and the following subdomains: www, web, ftp, mail
  - Capture all requests/responses
Approach – Queries (2)

• PlanetLab
  – Obtain slice
  – Add public key
  – Find active nodes
  – Add nodes to slice
  – CoDeploy to distribute software (distributed fashion – ended up dropping)
  – MultiQuery to execute software on machines (ended up dropping)
  – Most nodes at a Univ.
Approach – Queries (3)

• Required 15 nodes
  – 3 TLDs * 5 queries \rightarrow 4 subdomains + domain
  – 225k queries

• Each node runs a script
  – Install BIND utilities (dig)
  – Download list of domains from remote server
  – Start tcpdump for DNS traffic
  – Issue 15k requests (1 second sleeps)
Approach – Queries (4)

• Each node “phones home” when finished
  – Found many nodes never finished testing scripts from weeks ago. PL is “iffy”

• Copy packet captures locally for analysis
  – Most captures ~3.5MB
Implementation

• C++ for linking
  – Maps

• Perl scripts for everything else
  – Jim Clausing - SANS Institute
    • Library issues
Results – Tracking (1)

• Domains may have many nameservers
  – Made tracking more difficult. Attempted to track which ever NS domain/IP was chosen first. If not found for a given date, considered an NS change
Results – Tracking (3)

Cumulative Probability vs. Number of Namesever IP Changes

- com
- org
- travel
## Results – Queries (1)

<table>
<thead>
<tr>
<th>TLD</th>
<th>FQDN</th>
<th>Answers</th>
<th>Avg. A TTL (s)</th>
<th>Avg. CNAME TTL (s)</th>
<th>Avg. Suc. Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.com</td>
<td>X.com</td>
<td>12282</td>
<td>11111</td>
<td>7438</td>
<td>0.139</td>
</tr>
<tr>
<td>.org</td>
<td>X.org</td>
<td>12172</td>
<td>11909</td>
<td>16713</td>
<td>0.186</td>
</tr>
<tr>
<td>.travel</td>
<td>X.travel</td>
<td>11110</td>
<td>21175</td>
<td>14416</td>
<td>0.143</td>
</tr>
<tr>
<td>.com</td>
<td><a href="http://www.X.com">www.X.com</a></td>
<td>11461</td>
<td>11345</td>
<td>10783</td>
<td>0.124</td>
</tr>
<tr>
<td>.org</td>
<td><a href="http://www.X.org">www.X.org</a></td>
<td>10763</td>
<td>12941</td>
<td>11647</td>
<td>0.154</td>
</tr>
<tr>
<td>.travel</td>
<td><a href="http://www.X.travel">www.X.travel</a></td>
<td>10757</td>
<td>20082</td>
<td>19923</td>
<td>0.106</td>
</tr>
<tr>
<td>.com</td>
<td>web.X.com</td>
<td>5159</td>
<td>7737</td>
<td>13066</td>
<td>0.314</td>
</tr>
<tr>
<td>.org</td>
<td>web.X.org</td>
<td>4832</td>
<td>7706</td>
<td>14363</td>
<td>0.152</td>
</tr>
<tr>
<td>.travel</td>
<td>web.X.travel</td>
<td>2139</td>
<td>18947</td>
<td>27918</td>
<td>0.384</td>
</tr>
<tr>
<td>.com</td>
<td>ftp.X.com</td>
<td>8999</td>
<td>10355</td>
<td>10740</td>
<td>0.101</td>
</tr>
<tr>
<td>.org</td>
<td>ftp.X.org</td>
<td>9206</td>
<td>10856</td>
<td>11182</td>
<td>0.073</td>
</tr>
<tr>
<td>.travel</td>
<td>ftp.X.travel</td>
<td>1391</td>
<td>17015</td>
<td>32888</td>
<td>0.445</td>
</tr>
<tr>
<td>.com</td>
<td>mail.X.com</td>
<td>6212</td>
<td>13103</td>
<td>10402</td>
<td>0.121</td>
</tr>
<tr>
<td>.org</td>
<td>mail.X.org</td>
<td>5702</td>
<td>15212</td>
<td>10399</td>
<td>0.136</td>
</tr>
<tr>
<td>.travel</td>
<td>mail.X.travel</td>
<td>4083</td>
<td>23177</td>
<td>28401</td>
<td>0.179</td>
</tr>
<tr>
<td>.com</td>
<td>yammyhammy.X.com</td>
<td>4486</td>
<td>8125</td>
<td>14455</td>
<td>0.217</td>
</tr>
<tr>
<td>.com</td>
<td>Top 15K .com</td>
<td>13664</td>
<td>17872</td>
<td>9236</td>
<td>0.229</td>
</tr>
</tbody>
</table>
Results – Queries (2)

- Based on table wasn’t finding distinguishing characteristics

- Two new tests for .com
  - Attempt lookups for random subdomain (yammyhammy.X.com)
  - Top 15k via Alexa (allows top 1m CSV download)
  - Linked top 15k to 20130831 to find NS IP
    - There were top domains not listed in the database
Results – Queries (3)

Cumulative Probability vs. Response Time (s)

- www
- web
- ftp
- com
- mail
- yammyhammy
- top15k

Time intervals: 0.01s, 0.1s, 1s
Results – Queries (4)

Cumulative Probability vs. TTL (s)

- www
- web
- ftp
- com
- mail
- yammyhammy
- Top 15k com
Results – Queries (5)

• DNS server are using wildcards
  – web.X.org ~= yammyhammy.X.com
  – web doesn’t appear very popular
  – Tested wildcard functionality in BIND9. It might actually have security applications…?

• A few servers required TCP DNS request
  – Usually TCP due to size but were <300 bytes. Far less than UDP max

• Some response times were in the thousands (0.001 place)
Results – Queries (6)

• Found a max TTL in com at 2592000 seconds
  – 2592000s = 30 days

• Found way many more CNAMEs than expected
  – Maybe CNAME was wildcard to main domain
  – CNAME TTL != A record it points to. Problematic?
Lessons Learned (1)

- Dealing with large data sets is different
  - $O(n^2)$ logic in some places – bad

- Had Perl code that would have taken at least 12 hours to run. Implemented with C++ maps and finished in 2 minutes.

- Test thoroughly in small cases.
  - Overnight code crashed due to exception or didn’t capture data that I really needed.
  - Over capture and filter afterwards
Lessons Learned (2)

• POP3 or IMAP instead of mail

• Some versions of tcpdump limit the packet capture sizes unless you use “-s” flag.
  – Did captures twice…
Questions?


