Measurements of root server traffic in DITL 2008

Sebastian Castro
secastro@caida.org

CAIDA / NIC Chile

10th CAIDA/WIDE workshop – August 2008
Overview

• DITL 2008
• General statistics
• Query characteristics
  – Query rate comparison
  – Client rate comparison
  – Query types
  – Distribution of queries/clients
• Invalid traffic
  – Comparison with 2007
  – Exploration of sources
    • Recursive queries
    • A-for-A
    • Invalid TLD
DITL 2008

• Particularly successful in terms of variety of DNS traffic
  – 8 root servers
  – 2 old root servers
  – 2 ORSN servers
  – 5 TLD (1 gTLD, 4 ccTLD)
  – 2 RIR
  – 7 instances of AS112
  – Cache traces from SIE and University of Rome

• Also includes traces and measurements
# General statistics

<table>
<thead>
<tr>
<th></th>
<th>DITL 2007</th>
<th>DITL 2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset duration</td>
<td>24h</td>
<td>24h</td>
</tr>
<tr>
<td>Dataset start</td>
<td>Jan 9, noon (UTC)</td>
<td>Mar 19, midnight (UTC)</td>
</tr>
<tr>
<td>Number of queries</td>
<td>3.84 billion</td>
<td>8.00 billion</td>
</tr>
<tr>
<td>Number of unique clients</td>
<td>~2.8 million</td>
<td>~ 5.6 million</td>
</tr>
<tr>
<td>Recursive queries</td>
<td>17.04%</td>
<td>11.99%</td>
</tr>
<tr>
<td>TCP Bytes</td>
<td>1.65%</td>
<td>N/A</td>
</tr>
<tr>
<td>TCP Packets</td>
<td>2.67%</td>
<td>N/A</td>
</tr>
<tr>
<td>TCP Queries</td>
<td>~700K</td>
<td>N/A</td>
</tr>
<tr>
<td>Queries from RFC1918 addresses</td>
<td>4.26%</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Query rates

Variation of query rates along the years

- Between 2007 and 2008, the qrate grew:
  - C: 40%
  - F: 13%
  - K: 33%
  - M: 5%

- Between 2006 and 2008:
  - C: 139%
  - F: 71%
  - K: 74%
Client rate

- Follows the same pattern of query rates:
  - A, C, F, K and M with similar behavior
  - E and H
  - L

Average client rate per second (2008)
The highest fraction of queries are A queries (slightly below 60%).

Important increase on AAAA queries (pink): from around 8% in 2007 to 15% in 2008.

Reduction of MX queries (purple): K-root drop from 13% to 4%.
Distribution of clients/queries

DITL 2008

Leftmost column: ~2.8% of the queries are sent by ~86.4% of clients

Rightmost column: 1200 clients generated ~54.3% of the queries.
Distribution of fraction of queries/clients

- Are the clients sending their queries evenly to the roots?
  - The answer is NO
  - We excluded the clients sending just one query.

- L-root receives low fractions of traffic per client (80% of clients sent 20% or less of their queries)
- Followed by H and E (unicast servers)
- The rest are more on the same level.
Invalid queries analysis

• To prepare the invalid queries analysis we required to split the traces per source address.
  – We sampled 10% of the unique source addresses observed on each root

• Each query could fit in nine categories of invalid queries
  – The match was done sequentially
  – If none matched, was counted as valid query
Invalid queries categories

- Unused query class:
  - Any class not in IN, CHAOS, HESIOD, NONE or ANY
- A-for-A: A-type query for a name is already a IPv4 Address
  - <IN, A, 192.16.3.0>
- Invalid TLD: a query for a name with an invalid TLD
  - <IN, MX, localhost.lan>
- Non-printable characters:
- Queries with ‘_’:
  - <IN, SRV, _ldap._tcp.dc._msdcs.SK0530-K32-1.>
- RFC 1918 PTR:
  - <IN, PTR, 171.144.144.10.in-addr.arpa.>
- Identical queries:
  - a query with the same class, type, name and id (during the whole period)
- Repeated queries:
  - a query with the same class, type and name
- Referral-not-cached:
  - a query seen with a referral previously given.
Query validity (the graph)

Query Validity (2008)

Fraction of queries

Query rate interval

- Unused query class + non-printable char + queries with underscore + RFC 1918 PTR
- A–for–A
- Invalid TLD
- Identical queries
- Repeated queries
- Referal not cached
- Legitimate
### Query validity (the numbers)

<table>
<thead>
<tr>
<th>Category</th>
<th>A</th>
<th>C</th>
<th>E</th>
<th>F</th>
<th>H</th>
<th>K</th>
<th>L</th>
<th>M</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unused</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>A-for-A</td>
<td>1.6</td>
<td>1.9</td>
<td>1.2</td>
<td>3.6</td>
<td>2.7</td>
<td>3.8</td>
<td>2.6</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>Invalid TLD</td>
<td>19.3</td>
<td>18.5</td>
<td>19.8</td>
<td>25.5</td>
<td>25.6</td>
<td>22.9</td>
<td>24.8</td>
<td>22.9</td>
<td>22.0</td>
</tr>
<tr>
<td>Non-print char</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Queries with _</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>RFC 1918 PTR</td>
<td>0.6</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>Identical queries</td>
<td>27.3</td>
<td>10.4</td>
<td>14.9</td>
<td>12.3</td>
<td>17.4</td>
<td>17.9</td>
<td>12.0</td>
<td>17.0</td>
<td>15.6</td>
</tr>
<tr>
<td>Repeated queries</td>
<td>38.5</td>
<td>51.4</td>
<td>49.3</td>
<td>45.3</td>
<td>38.7</td>
<td>42.0</td>
<td>44.2</td>
<td>43.9</td>
<td>44.9</td>
</tr>
<tr>
<td>Referral not cached</td>
<td>10.7</td>
<td>15.2</td>
<td>12.1</td>
<td>10.9</td>
<td>12.9</td>
<td>11.1</td>
<td>14.3</td>
<td>11.1</td>
<td>12.4</td>
</tr>
<tr>
<td><strong>Valid 2008</strong></td>
<td><strong>1.7</strong></td>
<td><strong>2.0</strong></td>
<td><strong>1.8</strong></td>
<td><strong>1.9</strong></td>
<td><strong>1.8</strong></td>
<td><strong>2.0</strong></td>
<td><strong>1.8</strong></td>
<td><strong>1.8</strong></td>
<td><strong>1.8</strong></td>
</tr>
<tr>
<td><strong>Valid 2007</strong></td>
<td><strong>4.1</strong></td>
<td><strong>2.3</strong></td>
<td><strong>1.8</strong></td>
<td><strong>1.8</strong></td>
<td><strong>4.4</strong></td>
<td><strong>2.5</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Query validity (the words)

- Based on our first graphs, we query load keeps increasing
  - So the pollution
- The fraction of valid traffic is decreasing
- The pollution is dominated by “invalid TLD”, repeated and identical queries.
Looking some of the sources of pollution

- We explored more details on the sources of pollution
  - Recursive queries
  - A-for-A queries
    - Including some evidence of address space scanning and a new type of trash.
  - Invalid TLD
- … and propose some solutions
Recursive Queries

- During 2008 the number of recursive queries reduced compared to 2007
  - 2008: 11.99%; 2007: 17.04%
- But the number of sources increased
  - 2007: 290K (11.3%)
  - 2008: 1.97M (36.4%)
- We propose:
  - If a server doesn’t have the recursion enabled, return a REFUSED instead of a referral
A-for-A: Address space scanning

- Took all QNAME and convert them to addresses
- Group them by /24 and /16
- 18270 sources sent queries for the 80/8 – 83/8
- 8845 sources sent queries for the 88/8 – 89/8
- 8115 sources in common
- Seemed coordinated: different sources sent queries for different partitions, iterating over the third octet.

![Graph showing distribution of QNAME's on A-for-A queries observed at the roots during DITL 2008.](image-url)
A6-for-A? AAAA-for-A?

• Originally this category included A-queries with a query name in the form of an IPv4 address
  – What about the other query types for addresses?
  – The result: 3.32% of this type of queries were for A6/AAAA queries

00:04:03.347275 IP 195.2.83.107.5553 > 12.0.0.2.53: 40248 [1au] A? 221.0.93.99. (40)
00:04:03.347392 IP 195.2.83.107.5553 > 12.0.0.2.53: 1887 [1au] AAAA? 221.0.93.99. (40)
00:04:03.347642 IP 195.2.83.107.5553 > 12.0.0.2.53: 2737 [1au] A6? 221.0.93.99. (40)
00:04:59.579904 IP 195.2.83.107.5553 > 6.0.0.30.53: 40723 [1au] A? 84.52.73.160. (41)
00:05:36.016886 IP 195.2.83.107.5553 > 11.0.0.8.53: 28473 [1au] A? 148.240.4.32. (41)
00:05:36.016902 IP 195.2.83.107.5553 > 11.0.0.8.53: 27782 [1au] AAAA? 148.240.4.32. (41)
00:05:36.016908 IP 195.2.83.107.5553 > 11.0.0.8.53: 1175 [1au] A6? 148.240.4.32. (41)
Invalid TLD

- Queries for invalid TLD represent 22% of the total traffic at the roots
  - 20.6% during DITL 2007
- Top 10 invalid TLD represent 10.5% of the total traffic
- RFC 2606 reserves some TLD to avoid future conflicts
- We propose:
  - Include some of these TLD (local, lan, home, localdomain) to RFC 2606
  - Encourage cache implementations to answer queries for RFC 2606 TLDs locally (with data or error)

<table>
<thead>
<tr>
<th>TLD</th>
<th>Percentage of total queries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2007</td>
</tr>
<tr>
<td>local</td>
<td>5.018</td>
</tr>
<tr>
<td>belkin</td>
<td>0.436</td>
</tr>
<tr>
<td>localhost</td>
<td>2.205</td>
</tr>
<tr>
<td>lan</td>
<td>0.509</td>
</tr>
<tr>
<td>home</td>
<td>0.321</td>
</tr>
<tr>
<td>invalid</td>
<td>0.602</td>
</tr>
<tr>
<td>domain</td>
<td>0.778</td>
</tr>
<tr>
<td>localdomain</td>
<td>0.318</td>
</tr>
<tr>
<td>wpad</td>
<td>0.183</td>
</tr>
<tr>
<td>corp</td>
<td>0.150</td>
</tr>
</tbody>
</table>
Conclusions

• The traffic grows, the pollution grows
• We don’t know much about the sources of unwanted traffic
  – But we do know a little bit more every time
  – And we will continue looking for answers
    • Minas will talk about more about this
• Some root servers seem to be underused