Is Your Caching Resolver Polluting the Internet?

Duane Wessels
The Measurement Factory, and
CAIDA
wessels@measurement-factory.com

September 2004
A Disclaimer

• This data comes from monitoring two instances of the “F” DNS root server
  – Introduces some biases
  – We may be missing something interesting that occurs elsewhere

• It would be nice if we had additional data from other sources, like
  – Authoritative TLD and SLD servers
  – A variety of caching resolvers

• I’ll try to not make this not all about the root servers.
What is DNS pollution?

- A-for-A queries
- A-for-. queries
- Queries and Updates for RFC 1918 Addresses
- Queries for Invalid TLDs
- Excessive Queries for [a–m].root-servers.net
- IPv6 Address Queries
- Repeated Queries
- Queries from Unroutable/Unreachable Sources
Why is this DNS pollution?

- Some queries are unanswerable because the server is not authoritative for the domain in question. e.g., lame delegation and “NXDOMAIN” replies.
- Some queries are unanswerable because the server cannot talk back to the client.
- Some represent local/private information that escapes onto the public Internet.
- Some queries are valid, but occur much more frequently than they would for properly configured systems.
A-for-A Queries

22:33:57.879179 152.68.35.117.49283 > F.53: 6958 A? 4.43.140.160.
22:33:57.903560 176.17.50.38.49902 > F.53: 33116 A? 149.109.60.44.

- Caused by buggy Windows NT DNS server
- Some resolvers (i.e., djbdns) recognize and answer these queries.
A-for-. Queries


- Caused by buggy resolvers that accept null query names?
- Why not have the resolver recognize and stop these?
Sites that use RFC 1918 addresses should configure their resolver to answer authoritatively for them.

The AS112 project servers take the bulk of this abuse.

- 20 anycasted servers authoritative for RFC 1918 space
- Certain popular operating systems enable dynamic DNS update by default
Invalid TLDs

• Common invalid TLDs: *localhost*, *local*, *corp*, *workgroup*, *domain*, *htm*, *txt*, *c*

• Negative caching not good enough to stop these
[a-m].root-servers.net

20:00:30.085662 80.41.136.212.33821 > F.53: 33822 A6? A.ROOT-SERVERS.NET.
20:00:30.085870 80.41.136.212.33821 > F.53: 31888 AAAA? A.ROOT-SERVERS.NET.
20:00:30.086014 80.41.136.212.33821 > F.53: 51435 A? A.ROOT-SERVERS.NET.
20:00:34.087164 80.41.136.212.33821 > F.53: 36477 A6? A.ROOT-SERVERS.NET.
20:00:34.087394 80.41.136.212.33821 > F.53: 45228 AAAA? A.ROOT-SERVERS.NET.
20:00:34.087663 80.41.136.212.33821 > F.53: 33202 A? A.ROOT-SERVERS.NET.
20:00:38.087294 80.41.136.212.33821 > F.53: 5231 A6? A.ROOT-SERVERS.NET.
20:00:38.087563 80.41.136.212.33821 > F.53: 54557 AAAA? A.ROOT-SERVERS.NET.
20:00:38.087669 80.41.136.212.33821 > F.53: 51685 A? A.ROOT-SERVERS.NET.
20:00:42.098064 80.41.136.212.33821 > F.53: 55265 A6? A.ROOT-SERVERS.NET.
20:00:42.098315 80.41.136.212.33821 > F.53: 112 AAAA? A.ROOT-SERVERS.NET.
20:00:42.098440 80.41.136.212.33821 > F.53: 57604 A? A.ROOT-SERVERS.NET.
20:00:46.091344 80.41.136.212.33821 > F.53: 47264 A6? A.ROOT-SERVERS.NET.
20:00:46.091467 80.41.136.212.33821 > F.53: 27313 AAAA? A.ROOT-SERVERS.NET.
20:00:46.091592 80.41.136.212.33821 > F.53: 25386 A? A.ROOT-SERVERS.NET.
[a-m].root-servers.net, cont

- Caching resolvers like to update/validate their “hints” when they startup.

- Some caches query for all 13 root servers, and for both IPv4 and IPv6 addresses.

- Excessive root-servers.net queries usually indicates a unidirectional communication channel.

- Certain versions of BIND sometimes pummel the roots with AAAA and A6 queries for [a-m].root-servers.net.
Query Spikes for root-server.net IPv6 addresses

- Sources are BIND 8.3.3 – 8.3.4
 IPv6 Address Queries

- Not necessarily pollution, but...

- BIND optimistically issues AAAA and/or A6 queries for other nameservers.

- A random sampling of 3150 authoritative nameservers found 17 (0.5%) with AAAA records, and none with A6 records.
Repeated Queries

• Repeated queries are not included in the current analysis.

• Repeat analysis requires keeping history or state.

• Our real-time analysis tools don’t (yet?) keep history.

• Repeated query analysis is complicated by cache poison-prevention techniques. That is, some resolvers start certain queries at the roots, rather than trust cached referrals.

• In an earlier study we found that up to 70% of F-root traffic is repeated queries.
Unroutable/Unreachable Sources

- Source address from RFC 1918 space
- No route back to source address
- DNS reply is blocked by misconfigured packet filter
- Source port is zero
DNS Pollution Seen at F-Root
Aug 20–27, 2004

SIGCOMM 2004 NetTs
The Measurement Factory
So What, Who Cares?

• The root servers are generally over provisioned and the amount of traffic we’re talking about is not that significant.
  – okay, but it’s not only about the roots

• Viruses, spam, and DDoS represent much, much more pollution than this.

• “The DNS works well enough for me anyway.”

• “I run up-to-date code and have smart people working for me. I’m sure my resolvers are well-behaved.”

• Your policy may be to not “spy” on your customers.
Why You Should Care

- The Internet is a public commons that we should all strive to keep clean.

- Less pollution makes it easier to differentiate good from bad traffic in the event of a real attack.

- Your DNS pollution may give away private or sensitive information.

- You may discover configuration errors that you didn’t know you had.

- As the Balkanization of the Internet continues you’ll find that others willingness to communicate with you depends on the amount of garbage leaving your network (DNSBLs, rfc-ignorant, Dshield).
Tools
**dnstop**

- *dnstop* is a curses-based Unix application that displays sorted tables of DNS queries

- Recent versions have filters to include only certain types of DNS pollution
  - A-for-A queries
  - RFC 1918 PTR queries
  - RFC 1918 UPDATEs
  - Unknown TLDs

- [http://dns.measurement-factory.com/tools/dnstop/](http://dns.measurement-factory.com/tools/dnstop/)
## Finding Invalid TLDs with dnstop

Wed Apr 14 19:14:08 2004
9 new queries, 3045 total queries

<table>
<thead>
<tr>
<th>TLD</th>
<th>count</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>local</td>
<td>414</td>
<td>13.6</td>
</tr>
<tr>
<td>localhost</td>
<td>251</td>
<td>8.2</td>
</tr>
<tr>
<td>txt</td>
<td>85</td>
<td>2.8</td>
</tr>
<tr>
<td>147</td>
<td>31</td>
<td>1.0</td>
</tr>
<tr>
<td>1</td>
<td>23</td>
<td>0.8</td>
</tr>
<tr>
<td>invalid</td>
<td>22</td>
<td>0.7</td>
</tr>
<tr>
<td>null</td>
<td>22</td>
<td>0.7</td>
</tr>
<tr>
<td>belkin</td>
<td>21</td>
<td>0.7</td>
</tr>
<tr>
<td>workgroup</td>
<td>20</td>
<td>0.7</td>
</tr>
<tr>
<td>212</td>
<td>19</td>
<td>0.6</td>
</tr>
<tr>
<td>50</td>
<td>18</td>
<td>0.6</td>
</tr>
<tr>
<td>c</td>
<td>17</td>
<td>0.6</td>
</tr>
<tr>
<td>105</td>
<td>17</td>
<td>0.6</td>
</tr>
<tr>
<td>iec</td>
<td>17</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>17</td>
<td>0.6</td>
</tr>
<tr>
<td>emails</td>
<td>17</td>
<td>0.6</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>0.5</td>
</tr>
<tr>
<td>8</td>
<td>15</td>
<td>0.5</td>
</tr>
<tr>
<td>56</td>
<td>15</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>0.5</td>
</tr>
<tr>
<td>lan</td>
<td>14</td>
<td>0.5</td>
</tr>
</tbody>
</table>
DSC: DNS Statistics Collector

- Kind of an “MRTG” for DNS servers
- Colorful, interactive graphs
- Distributed architecture
- Long-term data archival
- Currently “alpha release” quality/status
- http://dns.measurement-factory.com/tools/dsc/
DSC Sample Graph

Classification of Queries
Generated: Friday, Sep 03, 2004, 16:17:45 UTC

Query Rate (q/s)
2500
2000
1500
1000
500
0
16:00 18:00 20:00 22:00 0:00 2:00 4:00 6:00 8:00 10:00 12:00 14:00 16:00
Sep2
Time
[Click on legend entries for per-class graphs]
The End