Analysis of queries from viewpoint of caching servers

Tsuyoshi Toyono, Haruhiko Nishida, and Keisuke Ishibashi
NTT Laboratories

2007 OARC Workshop
Outline

1. One-day summary of queries at caching servers
2. Bogus queries observed at caching servers
3. Summary & what can we do?
Focus

- DNS caching server in/out queries
  - User → Cache queries (recursive)
  - Cache → Authoritative (nonrecursive)

"Is Your Caching Resolver Polluting the Internet?", Duane Wessels, 2004
1) Summary of queries at caching servers
Data summary: day of cache data (May 11, 2007)

- Capturing queries sent to/from DNS cache servers (ISP’s commercial DNS traffic)
  - Handling billions of queries/day

- Cache query prevention percentage
  - About 82%
  
  Number of queries sent to authoritative servers / Number of queries from users
  
  $= 0.1740... \rightarrow 82.6%$ prevention
Qtype time series (queries from users/day)

Lunch time click
Qtype time series (queries to authoritative/day)

am. 4:00 PTR (web analyzer default setting)
Qtype percentages (from users)

User’s MX
• botnets
• SPAM sender
• Kinds of Netsky worms
Qtype percentages (to authoritative)

Distributions are similar to “from users”
Rcode-classified Qtypes (to authoritative)

NoError FormErr ServFail NXDomain NotImp other

(Ex) SRV _ldap._tcp.Default-First-Site-Name._sites.dc._msdcs.MS***.local
Queries sent to root servers

- Queries sent to 13 root servers
  - 8.23% of all nonrecursive queries
Using root servers (in Tokyo point)

queries/day

A
MX
PTR
NS
SOA
CNAME
AAAA
A6
ANY
SRV
TXT
other

F @Osaka
23% F

K @Tokyo
38% K

M @Tokyo
12% M
2) Bogus queries sent from users & sent to authoritative servers
Bogus queries: RFC1918(AS112)

- PTR queries for RFC1918
  - Ex. PTR "*.*.*.10.in-addr.arpa"

- From Users
  - RFC1918 PTR
    - 32.22% of all PTR queries
    - 4.42% of total queries
  - RFC1918 SOA
    - 32.27% of all SOA queries
    - 0.25% of total queries

- This is almost the same percentage of queries as that of our 2005 data.
- AS112 servers reply to these queries
Bogus queries: Invalid TLDs

- “localhost.”, “local.”, “localdomain.”, “workgroup.”, “wpad.”…

- From Users
  - 6.09% of total queries
  - Percentage more than that of RFC1918 PTR queries

- To Authoritative servers
  - 7.50% of all queries (of all nonrecursive)
    - 99.5% of all queries (to root servers)

- In the end, these queries are sent to root-servers
- These queries are not prevented by DNS cache server
Bogus queries: A for A queries

• A queries for IP addresses
  – e.g., A “10.0.0.1”

• From Users
  – 1.53% of A queries
  – 0.88% of total queries

• To Authoritative servers
  – 0.50% of A queries
  – 0.31% of all nonrecursive queries
    • 1.56% of total queries (to root-servers)
“TLD ranking” from user queries

<table>
<thead>
<tr>
<th>% of total</th>
<th>28.41%</th>
<th>25.98%</th>
<th>19.94%</th>
<th>6.04%</th>
<th>2.31%</th>
<th>2.15%</th>
<th>1.91%</th>
<th>0.98%</th>
<th>0.96%</th>
<th>0.84%</th>
<th>0.81%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PTR</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>AAAA</td>
<td>MX</td>
<td>MX</td>
<td>AAAA</td>
<td>AAAA</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td></td>
<td>arpa</td>
<td>jp</td>
<td>com</td>
<td>net</td>
<td>jp</td>
<td>com</td>
<td>jp</td>
<td>com</td>
<td>localhost</td>
<td>org</td>
<td>local</td>
</tr>
</tbody>
</table>

- Valid
- Suspicious
- Unnecessary
“TLD ranking” to authoritative queries

<table>
<thead>
<tr>
<th>% of total</th>
<th>Type</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.97%</td>
<td>A</td>
<td>com</td>
</tr>
<tr>
<td>15.75%</td>
<td>A</td>
<td>jp</td>
</tr>
<tr>
<td>15.61%</td>
<td>PTR</td>
<td>arpa</td>
</tr>
<tr>
<td>11.57%</td>
<td>A</td>
<td>net</td>
</tr>
<tr>
<td>5.76%</td>
<td>MX</td>
<td>com</td>
</tr>
<tr>
<td>5.43%</td>
<td>AAAA</td>
<td>localhost</td>
</tr>
<tr>
<td>2.12%</td>
<td>A</td>
<td>org</td>
</tr>
<tr>
<td>1.62%</td>
<td>AAAA</td>
<td>jp</td>
</tr>
<tr>
<td>1.51%</td>
<td>AAAA</td>
<td>com</td>
</tr>
<tr>
<td>1.32%</td>
<td>MX</td>
<td>jp</td>
</tr>
<tr>
<td>1.15%</td>
<td>ANY</td>
<td>jp</td>
</tr>
</tbody>
</table>
### “TLD ranking” to root-servers queries

<table>
<thead>
<tr>
<th>% of total</th>
<th>Domain</th>
<th>Suspicious</th>
<th>Unnecessary</th>
</tr>
</thead>
<tbody>
<tr>
<td>70.45%</td>
<td>localhost</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12.90%</td>
<td>local</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.56%</td>
<td>A_for_A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.09%</td>
<td>localdomain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.57%</td>
<td>LOCAL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.42%</td>
<td>arpa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.26%</td>
<td>Domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.22%</td>
<td>domain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.19%</td>
<td>not-defined</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.15%</td>
<td>**n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05%</td>
<td>valid-ccTLDs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03%</td>
<td>WORKGROUP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11.83%</td>
<td>Other (almost invalid)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Classified queries pie chart (from users)

- Netsky = MX repeat < 0.01sec
- (Err) Repeat = repeat < 1sec
- Ignore TTL = including not using SOA TTL

- legitimate: 44%
- invalid qtype: 1%
- invalid TLD: 6%
- ignore TTL: 39%
- RFC1918: 4%
- netsky: 0%
- Err Repeat: 5%
- Repeat: 1%
Query percentages pie chart (to authoritative)

- legitimate: 69%
- Repeat: 17%
- invalid qtype: 0%
- invalid TLD: 8%
- ignore TTL: 1%
- netsky: 2%
- Err Repeat: 3%
- RFC1918: 0%
“TLD ranking” to root-servers queries

- localhost: 71%
- local: 13%
- A_for_A: 2%
- localdomain: 1%

Other: 12%

valid-ccTLDs
not-defined
**n
WORKGROUP
Domain
domain
arpa
LOCAL
local
A_for_A
localdomain
localhost
Summary

• Caching servers cache about 82% of user queries
  – Cache servers work to the best of their ability

• Number of RFC1918 PTR queries has not decreased from 2005
  – About 4% of user queries are RFC1918 queries
  – AS112 servers work to the best of their ability
But...

- The number of “Invalid TLDs” queries is more than that of RFC1918 queries
  - 6% of user queries are “invalid TLD” queries
  - 85% of these queries comprise just 3 domains
    - “.localhost.”
    - “.local.”
    - “.localdomain.”

- Caching servers do not work effectively in some cases
  - 99.5% of to-root-servers queries are “invalid TLDs”
  - “Invalid TLDs comprise 15%~20% of F-root queries”
    - “Is Your Caching Resolver Polluting the Internet?”, Duane Wessels, 2004
What can we do?

• Is it appropriate to answer NX Domain to these queries immediately without sending authoritative queries?
  – "localhost." "local." "localdomain."
  – Cache server answers sent to users are same as before

• If it is, then who to answer NX Domains?
  – It's more effective if those queries are processed near end-nodes
What can we do? (cont.)

- **Root (& authoritative)**
  - Answer to those queries from AS112 servers

- **Caching**
  - Sinkhole, blackhole
    - Set up dummy authoritative servers for “.local.”
  - “default” configurations “Long negative cache TTL”
    - e.g. /etc/bind/zones.rfc1918
    - Like “draft-ietf-dnsop-default-local-zones-02”, M. Andrews

- **Stub & users**
  - The promotion of appropriate “default” configurations
  - The promotion of appropriate implementations
Any questions?