DNS PREFETCHING: WHEN GOOD THINGS GO BAD

Srinivas Krishnan and Fabian Monrose

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL
Information quest

Timeline

1980 1990 2000 2010
Information quest

Timeline

1980 1990 2000 2010
Latency: Hours Minutes Seconds
Information quest

Timeline

Google!

1980  1990  2000  2010

Latency: Hours  Minutes  Seconds

Google Search  I'm feeling lucky
Information quest

Google

Timeline

1980
Latency: Hours

1990
Minutes

2000
Seconds

2010
Milliseconds

Latency: Hours
Minutes
Seconds
Milliseconds
Browser Wars
Browsing and DNS

- www.unc.edu

DNS Server:
- root.
- dmtns07.turner.com
- cnn.com
- ns2.unc.edu
- bristol.cs.unc.edu
- cs.unc.edu

Cache:
- unc.edu NS 86400 ns2.unc.edu
- ns2.unc.edu A 86400 152.2.253.100
- unc.edu A 86400 152.19.240.120

<domain> <A, CNAME, NS> <TTL> <meta>
DNS Optimization

• Proactive DNS pre-resolutions

• Two basic approaches:
  • Guess as the user types
  • Fetch <href> links from a rendered page

• Focus on reducing user perceived latency
DNS PRE-RESOLUTION

Gambling Addiction

Google Search  I'm Feeling Lucky

DNS Server

Cache

www.google.com CNAME 586186 www.l.google.com

www.l.google.com A 60 www.l.google.com
DNS PRE-RESOLUTION

Gambling Addiction

Google Search  I'm Feeling Lucky

DNS Server

sac.edu

Cache

www.google.com CNAME 586186 www.l.google.com

www.l.google.com A 60 www.l.google.com
DNS PRE-RESOLUTION

Gambling Addiction

Google Search  I'm Feeling Lucky

DNS Server

www.google.com CNAME 586186 www.l.google.com
www.l.google.com A 60 www.l.google.com
sac.edu A 73136

sac.edu

Cache
DNS PRE-RESOLUTION

Gambling Addiction

Google Search  I'm Feeling Lucky

sac.edu

DNS Server

www.google.com CNAME 586186 www.l.google.com
www.l.google.com A 60 www.l.google.com
sac.edu A 73136

gamblersanonymous.org. A 73416
casinogambling.about.com.CNAME 900
treatment-centers.net. CNAME 3600
robertperkinson.com. A 86400
en.wikipedia.org. CNAME 1052
ncpgambling.org. A 73416,
helpguide.org. A 73340
gamblingaddiction.org. A 3600

Cache

Prefetching
Privacy Threat

• Reconnaissance of an enterprise

• Ability to track users

• Exploit:
  • Ability to probe a DNS server to infer cache hits.
  • Online probes with target search
  • Offline probe with no prior knowledge
Online Probing

Was a target search performed by a client?

• Build a profile of target search
• Use cache snooping
• Check for presence of profile
• Report
Building a Profile

ama-assn.org
learn.genetics.utah.edu.
www.humancloning.org.
www.ornl.gov.
en.wikipedia.org
www.globalchange.com
www.ncsl.org
Building a Profile

Domains
- howstuffworks.com
- ama-assn.org
- genetics.utah.edu
- humancloning.org
- time.com
- ornl.gov
- en.wikipedia.org
- globalchange.com
- ncsli.org

MinTTL

Decay Curve
Building a Profile

Domains

- ama-assn.org.
- genetics.utah.edu.
- humancloning.org.
- ornl.gov.
- globalchange.com
- ncsl.org

MinTTL

Decay Curve
# Building a Profile

## Domains

- ama-assn.org
- genetics.utah.edu
- humancloning.org
- ornl.gov
- globalchange.com
- ncsl.org

## MinTTL

<table>
<thead>
<tr>
<th>Domain</th>
<th>MinTTL</th>
</tr>
</thead>
<tbody>
<tr>
<td>ama-assn.org</td>
<td>1800</td>
</tr>
<tr>
<td>genetics.utah.edu</td>
<td>3600</td>
</tr>
<tr>
<td>humancloning.org</td>
<td>3600</td>
</tr>
<tr>
<td>ornl.gov</td>
<td>86400</td>
</tr>
<tr>
<td>globalchange.com</td>
<td>600</td>
</tr>
<tr>
<td>ncsl.org</td>
<td>86400</td>
</tr>
</tbody>
</table>

## Decay Curve

The graph shows the decay curve for human cloning over time in the cache.
Building a Profile

Decay Curve

Get Scan Rate

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Time in Cache</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>5 Mins</td>
</tr>
<tr>
<td>90%</td>
<td>10 Mins</td>
</tr>
<tr>
<td>80%</td>
<td>20 Mins</td>
</tr>
<tr>
<td>75%</td>
<td>30 Mins</td>
</tr>
<tr>
<td>50%</td>
<td>60 Mins</td>
</tr>
</tbody>
</table>

Human Cloning

Accuracy

Time in Cache
Probe

Attacker
ama-assn.org.
genetics.utah.edu.
humancloning.org.
ornl.gov.
globalchange.com
ncsl.org

DNS Server

Cache Hit

genetics.utah.edu ?
Probe

Attacker

ama-assn.org.
genetics.utah.edu.
humancloning.org.
ornl.gov.
globalchange.com
ncsl.org

DNS Server

ama-assn.org.
genetics.utah.edu.
humancloning.org.
ornl.gov.
globalchange.com
ncsl.org
Probes

Confidence = % of Elements with same age

<table>
<thead>
<tr>
<th>Domain</th>
<th>Current TTL</th>
<th>Auth TTL</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>ama-assn.org</td>
<td>1498</td>
<td>1800</td>
<td>302</td>
</tr>
<tr>
<td>genetics.utah.edu.</td>
<td>3298</td>
<td>3600</td>
<td>302</td>
</tr>
<tr>
<td>humancloning.org</td>
<td>3301</td>
<td>3600</td>
<td>299</td>
</tr>
<tr>
<td>ornl.gov</td>
<td>86099</td>
<td>86400</td>
<td>301</td>
</tr>
<tr>
<td>globalchange.com</td>
<td>298</td>
<td>600</td>
<td>302</td>
</tr>
<tr>
<td>ncsl.org</td>
<td>86101</td>
<td>86400</td>
<td>299</td>
</tr>
</tbody>
</table>
And if we had access to logs?

• Can we extract all searches?
DNS Cache: privacy leaks

Goal: Reconstruct Search Term from DNS Cache

Cluster By Age
- steroid.com 600s
- steroidsinbaseball.net 598s
- baseballsteroidera.com 602s

Extract Keywords
- steroid
- steroid, baseball
- steroid, baseball, era

Search Term
- Rank
  - (1) steroid
  - (2) baseball
  - (3) era
- n-Suggest
  - steroid
  - baseball
  - steroid baseball
  - baseball steroids
  - steriod baseball era
Case I: Preliminary Results

~500 Clients

- 50 queries
- Over 4 hours
- Variable scan rate
Case I: Preliminary Results

~500 Clients

Inject Queries

Target DNS Server

Control DNS Server

Build Profile

- 50 queries
- Over 4 hours
- Variable scan rate
Case I: Preliminary Results

~500 Clients

Inject Queries

Target DNS Server

Control DNS Server

Probe Server @Scan Rate

• 50 queries
• Over 4 hours
• Variable scan rate
Selected Results

<table>
<thead>
<tr>
<th>Scan Rate</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 Mins</td>
<td>90%</td>
</tr>
<tr>
<td>30 Mins</td>
<td>85%</td>
</tr>
<tr>
<td>60 Mins</td>
<td>65%</td>
</tr>
</tbody>
</table>
Case II: Preliminary results

- ~500 Clients
- 50 queries
- Over 24 hours
Case II: Preliminary results

- ~500 Clients
- Inject Queries
- 50 queries
- Over 24 hours

Target DNS Server

Cache Snapshot @5 mins

Disk

Reconstruct
## Snapshot of Results

<table>
<thead>
<tr>
<th>Actual Query</th>
<th>First Guess</th>
<th>Second Guess</th>
<th>Third Guess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gambling Addiction</td>
<td>gambling addiction</td>
<td>gambling age</td>
<td>addict</td>
</tr>
<tr>
<td>Alcohol Withdrawal</td>
<td>alcohol withdrawal symptoms</td>
<td>alcoholics anonymous</td>
<td>alcohol poisoning</td>
</tr>
<tr>
<td>Syndrome</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gun Control</td>
<td>gunbroker</td>
<td>guns for sale</td>
<td>-</td>
</tr>
<tr>
<td>Racism In America</td>
<td>racism america</td>
<td>racism today</td>
<td>racism facts</td>
</tr>
<tr>
<td>Biological Weapons</td>
<td>biological warfare</td>
<td>weapons</td>
<td>-</td>
</tr>
</tbody>
</table>
Limitations

• Current profiles are non-adaptive, hence searches on “hot topics” will lead to high false negatives

• Similarly, if majority of prefetched domains do not have identifiable keywords, search reconstruction will fail
Summary

• Wide-scale study required to fully gauge the effect of DNS prefetching (w.r.t. its privacy implications)
  • Effect on DNS server load remains unclear
  • Reduction of user-perceived latency at the cost of privacy
• Primary focus is to foster discussion on the effects of DNS prefetching
Questions