Improving Content Delivery Using Provider-aided Distance Information

Authors: Ingmar Poese, Benjamin Frank, Bernhard Ager, Georgios Smaragdakis, Anja Feldman

Presenter: John Rula
Show more than half of total traffic can be delivered from multiple servers at diverse network locations

- Propose PaDIS as a content location recommendation system within an ISP
- Quantify Content Delivery performance improvement of PaDIS
Content Delivery Networks

- Also known as CDNs deliver content through a worldwide network of ‘edge servers’
- Host content from content providers and serve it to clients from the “best” possible server (closest or lowest latency).
- Most popular use DNS Redirection to find closest server.
What is DNS

- Domain Name System
  - A distributed database of mappings from domain names to IP addresses
  - There are different levels of DNS Servers
    - Authoritative name servers
      - Cache servers
        - Subdomains can be run by different Name Servers
  - A Query is called by a client and is recursively resolved
A Recursive DNS Query

Query is first sent to root nameserver

Then is recursively sent down the hierarchy of name servers until it is resolved
CDNs and DNS

- CDNs utilize this recursive structure to first encode information into a DNS query

  - A CDN DNS query is resolved to the edge server which has been determined by the CDN to be the “best” server (for the initially requesting DNS nameserver)

- Typically the ISP’s DNS Server

- DNS queries are cached across the network, and as such, the TTL (Time-To-Live) value is used (abused) to make sure the queries are not cached long.
Content Delivery Networks (CDNs)
Dataset

Table 1: Summaries of anonymized traces.

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Start date</th>
<th>Dur</th>
<th>Size</th>
<th>Application Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAR10</td>
<td>Packet</td>
<td>Thu 04 Mar’10 2am</td>
<td>24 h</td>
<td>&gt;5 TB</td>
<td>&gt;3 TB HTTP, &gt;5 GB DNS</td>
</tr>
<tr>
<td>HTTP-14d</td>
<td>log file</td>
<td>Wed 09 Sep’09 3am</td>
<td>14 d</td>
<td>&gt;200 GB</td>
<td>corresponds to &gt;40 TB HTTP</td>
</tr>
<tr>
<td>DNS-5d</td>
<td>packet</td>
<td>Wed 24 Feb’10 4pm</td>
<td>5 d</td>
<td>&gt;25 GB</td>
<td>&gt;25 GB DNS</td>
</tr>
</tbody>
</table>

*Based on limited dataset using payload inspection
DNS Analysis

- MultiQuery - Returns multiple IP addresses within a single DNS response
  - CrossQuery – Can return different IP addresses for repeated queries and this perform DNS redirection

- Internal DNS Resolver is ISP’s, External is something like OpenDNS or GoogleDNS.
  - Found that 95% of clients use ISP’s DNS Resolver as default
  - Use of external DNS decreases performance of CDN delivery
DNS Analysis cont...
Impact of Traffic Localization

- Labeling hostnames
  - Using external DNS amounts to 12% of traffic within same AS
  - Using ISP DNS amount to 25% of traffic in AS
  - Potential is if any IP found within same AS is used all the time

<table>
<thead>
<tr>
<th>Metric</th>
<th>ISP DNS observed</th>
<th>ISP DNS potential</th>
<th>OpenDNS observed</th>
<th>OpenDNS potential</th>
<th>GoogleDNS observed</th>
<th>GoogleDNS potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>IPs</td>
<td>12.3 %</td>
<td>24.2 %</td>
<td>5.8 %</td>
<td>16.0 %</td>
<td>6.0 %</td>
<td>9.7 %</td>
</tr>
<tr>
<td>requests</td>
<td>14.9 %</td>
<td>33.2 %</td>
<td>4.7 %</td>
<td>18.8 %</td>
<td>4.8 %</td>
<td>6.4 %</td>
</tr>
<tr>
<td>volume</td>
<td>23.4 %</td>
<td>50.0 %</td>
<td>12.0 %</td>
<td>27.7 %</td>
<td>12.3 %</td>
<td>13.4 %</td>
</tr>
</tbody>
</table>
What is PaDIS?

- Provider-aided Distance Information System
- Can be used by client, CDN, and ISP
- Acts as a middle layer between an ISP and a Content Delivery Network.
- Allows ISP to optimize CDN traffic for their benefit (cost).
How does PaDIS work?

- Consists of a Data Management System and a Query Processing System

  - Data Management System
    - Information Retrieval
      - Network Map Generator
  
  - Query Processing System
    - Request Translator
    - Query Manager
    - Path Ranker
    - Frequent hitter Detector
PaDIS Use Cases

- **Clients**
  - Can have all DNS replies sent to PaDIS server for re-ranking and returning IP address of ISP’s preference

- **CDNs/CDPs**
  - Collaborate with ISP through PaDIS by contacting them before returning server choices to the DNS resolver

- **ISP**
  - Enhance DNS Resolver to contact its PaDIS server and reorder the IP addresses if needed before returning any answer to the client
PaDIS Evaluation

![Graph 1: Download Time](image1)

![Graph 2: Download Time](image2)

![Graph 3: Probability for Choosing Link](image3)
Problems with PaDIS

- Potential for abuse by “bad” ISPs.
  - Unable to correctly implement the low TTL times used by CDNs.
Questions.....?