DITL2008 and DatCat

Emile Aben <emile@caida.org>
10th CAIDA-WIDE workshop
Aug 15-16 2008, Marina Del Rey, CA, US
DITL2008

- Day in the Life of the Internet (DITL)
- 48hr Internet Measurement event
- March 18-19 2008 (UTC)
  - 8 of 13 DNS roots participating
  - Compared to 2007:
    - Double the number of participants
    - Double the volume of DNS data
DatCat

- Internet Measurement Data Catalog (DatCat): http://imdc.datcat.org
  - Facilitate searching for and sharing of data among researchers
  - Enhance documentation of datasets via a public annotation system
  - Advance network science by promoting reproducible research
- No data storage, only meta-data:
  - How many IPv6 pkts in this pcap file?
  - How was this pcap file created?
  - How to get access to data
Lesson learned from DITL2007 and DatCat

• Researchers need data described in great detail:
  • Find out if data is useful for them or not
  • Find the most useful data
  • Can do research on meta-data
• Data contributors usually don’t have time to describe data in detail
(Temporary) Solution

- Add a curator-role:
  - Fills out general information about the data
  - Extracts metadata when data is accessible (e.g. OARC)
- Data contributor checks the general information, corrects it, fills in the blanks, adds extra info
- Curator processes this into a submission
Result

- Less work for contributor
  - Only needs to check
- Extra role: curator
  - Create initial descriptions
  - Extract metadata
- Takes time/coordination
Interface for curator and contributor to collaborate
**Result in DatCat**

**Collection: NIC Chile, .CL ccTLD DNS trace DITL 2008**

DNS PCAP traces containing DNS queries, collected at 4 .CL nameservers.

Jump to: Description | Annotations | Citation | Record Details

**Collection Contents**

- **data objects:** 192 directly contained, 192 total

**Collection Details**

**Summary**

DNS packet traces collected during the DITL 2008 event. Includes queries from 4 authoritative nameservers for .CL ccTLD. The trace files in this collection contain IPv4 traffic with DNS queries (no replies), and a total of 75.27 million packets.

**Motivation**

This data was collected as part of the DITL project, in which DNS and other Internet data is collected for a 2-day period. As a member of OARC, NIC Chile participated for the first time.

**Data Start Time**

2008-03-17 17:00 PDT (-0700)

**Data End Time**

2008-03-19 18:59:59.999 PDT (-0700)

**Data Duration**

1 days 23:59:59.999 (172799.999 s)

**Creators**

Mauricio Vergara, Sebastian E. Castro, Duane Wessels

**Primary contact**

NIC Chile DNS Admin

**Keywords**

cctLD, DITL, DITL-2008-03-19, DNS, OARC, passive, trace

**Used in publications**

- directly contained by
  - Day in the Life of the Internet, March 18-19, 2008 (DITL-2008-03-18)

**Description**

The traces in this collection were collected at four authoritative nameservers for .CL located in Chile. Three of them are members of the same anycast cloud (tucapel, santiago, valparaiso), the other is a unicast node (ns.nic.cl). The unicast node serves other Latin American ccTLDs, but this traffic was not captured. The collection procedures were thoroughly tested on the days previous to the collection event, so no problems of disk space, clock skew were present. Only an additional step was added to the process: all traces, once collected, were sent to a transfer server for upload, due to bandwidth constraints on the nodes where the collection ran.

**Annotations**

Add a note...

**Citation**

Please use the following BibTeX citation to cite this collection. Some parts are optional or may need to be edited. To use the \texttt{\url{...}} command for nice URL formatting, you must call \texttt{\usepackage{url}} in the LaTeX preamble.

Saturday, August 16, 2008
Current status

• Positive reactions from contributors to reduced workload
• Contributors replace strings of text when asked to, but generally don’t add to text
• 17 datasets indexed in DatCat
• 11 datasets currently in process
• At least 7 OARC datasets and 3 others to be done
Longer Term Solution

- Have ability for DatCat contributor to describe a collection of data, but not necessarily all files in it
- Contributor can add detailed description of data later (at the file level)
- ... but probably needs incentives to do so
## Links

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DITL2008 in DatCat</td>
<td><a href="http://imdc.datcat.org/collection/1-05MM-F">http://imdc.datcat.org/collection/1-05MM-F</a></td>
</tr>
</tbody>
</table>