2007 Day In The Life
What went wrong, Why, How to make it better next time

Duane Wessels
The Measurement Factory/CAIDA

WIDE+CAIDA Workshop #8
July 21, 2007

DITL 2007

- Day In The Life of The Internet. Okay, two days.
- 48 hour period: Jan 9 00:00:00 to Jan 10 23:59:59 UTC
- Primary focus is DNS and root servers, but other data was collected as well.

Problems Uploading

- These have already been covered by Keith
- OARC box ran out of disk space.
- Upload method does not preserve sender-side filename.
- Sender does not have shell access to fix mistakes.
- Receiving program did not handle partial uploads very well.
- Receiving program had bugs with the microsecond part of the first packet timestamp.
- No way to upload a "metadata" file.

Pcap file size and boundaries

- Inconsistent pcap file durations. Most pcap files are 1 hour long, except:
  - C-root: 5 minutes
  - K-root: random, other problems
- Inconsistent pcap file start times.
  - Some pcap files start at 1-hour boundaries.
  - C-, F-, and K-root start at random times.
- Consistent start time and lengths simplifies a number of tasks:
  - Selecting data for analysis
  - Merging pcap files together
  - Knowing if/when all data has been successfully uploaded

Clock Skew

- We sent queries with a timestamp-based query name to known anycast and unicast root servers.
- Found six nodes with skew greater than 3 seconds. One was off by 20 seconds and another by 17.
- Could affect anycast stability analysis?
- This technique is far from perfect.
  - Only hit 'a' nodes of F-root loadbalanced sites.
  - Did not have unicast addresses for C, K, M.
  - Hard to account for transmission delays.

Truncated Packets

- Most K-root instances have truncated packets (1500 vs 1514).
- b.orsn-servers.net has truncated packets (96 vs 1514).
- Probably only an issue for replies, rather than queries.

Unexpected Data

- Pcap files may contain packets for other servers, or even other protocols.
- For example, f-sfo2 pcap files also contain queries to ns-ext.isc.org and d.dns.br.
- We don't know the tcpdump command line and arguments used to capture the data.
- A simple analysis such as 'tcpdump -n -r - dst port domain — wc' may give incorrect results.
- pcap file may also contain queries sent *by* the server.
  - i.e., SOA queries for zone synchronization

Missing Data

- isc/f-dac1a: Missing about 6 hours
  - 2007-01-09 00:00 to 2007-01-09 06:00
  - although it includes an extra 6 hours of data after the end of the collection period.
- isc/f-muc1b: Missing about 22 hours
  - 2007-10-02:15 to the end
  - ISC had a hard time getting the pcap files from this node to the OARC server. Eventually they did upload 50 hours worth of data, but it seems to be shifted by 24 hours from the collection period.
- orsnb/b.root-servers.net: Missing 23 hours at the start, and 1 hour at the end.
- ripe/*: Much of RIPE's data is incomplete
  - Did not have enough local disk space for the capture files.
  - Some data given the wrong name ("poznan") when uploading. fixed?
  - DW accidentally deleted one of the ripe-brisbane files.
- wide/*: Most of the WIDE files are missing the final second or so of each hour from forgetting to call gzclose().

VLAN tags

- f-sfo2 is the only instance where packets are tagged with VLANs.
- A little bit annoying for people that write their own pcap readers.

Gzip Integrity

- Many uploaded pcap files fail a *gzip -t* test.
- Some software (e.g, Coral Reef) ignores the whole file if decompression fails.
- Should we keep the files as they are?
- Or re-compress them to remove the errors?
  - We re-compressed WIDE files

Pcap Integrity

- Some uploaded files encounter errors during pcap processing.
- Leave or fix?

Next Time?

- Intermediate storage sites to prevent data loss?
- Shell access for contributors?
- How much do we care about clock accuracy?
- *dnscap* will save us from truncated packets and other problems?
- Normalize pcap files after receipt by OARC?
  - start/end boundaries
  - remove pcap/gzip errors
  - remove VLAN tags
  - remove irrelevant packets

The End