RSSAC-002

Archiving and Data Analysis
RSSAC-002 Overview

- ICANN's Root Server System Advisory Committee created an advisory to the community, RSSAC-002
- Baseline of metrics
- Recommendations on the collection of certain data plus format of the date
Data Formats

• Data is available in a variety of formats, with some consistency between collections
  – YAML is common to all while JSON and some PNGs generated by one root operator

• Primary data format is YAML

• Section 4 of RSSAC-002 defines format further
Data Metrics

• The data metric covered under the advisory are:
  – Load time
  – RCODE volume
  – Traffic size
  – Traffic Volume
  – Unique sources
  – Zone Size

• As defined by Section 2 in RSSAC-002
RSSAC-002 Mirroring

- DNS-OARC is actively mirroring all known and available RSSAC-002 data from the various root-operators, so you don’t have to
  - A, J, K and L root servers are covered
  - Assembled as one collection in one spot
- To be gentle to the web servers in question, the mirroring process itself can take about 6 hours to complete, depending upon the amount available
- Data is available to researchers and any other members via OARC’s analysis servers
- /mnt/oarc-pool3/rssac-002/
RSSAC-002 Coverage

• A
  – May 2013 and onwards
• J
  – May 2013 and onwards
• K
  – March 2015 and onwards
• L
  – June 2014 and onwards
• We will announce additions to the collection as we confirm they are available
Data Quality

• Thus far the data varies in quality – not of the data contained per se
  – Mirroring process was double-checked for errors it might be causing in its copy
• But instances of
  – Data not found in the right place (some day’s data found in next month’s directory tree)
  – Data missing (missing spans of 24h series) for some days here and there across a number of sources
• Section 4.7 defines a path scheme to the data which has been followed while base URLs (not part of the advisory) vary.
Finding Data

• Across 4 servers, let’s find the zone size for a given day. Let’s use 2015-03-20
  – Oops! Can’t. Only 2 of 4 servers have that info (neither J nor L have published this)
• How about RCODE data for 2014-06-01?
  – Careful! If looking by timestamp, some files are in another month, so look by filename
• It must be remembered: operational priority takes precedence over data collection and publishing
What does the data say?

• Stay tuned…
___END___