DNS-OARC and the Open Knowledge Network

Keith Mitchell
DNS-OARC President
WIE-KISMET Workshop
December 2019
OARC's Mission Statement

The Domain Name System Operations Analysis and Research Center (DNS-OARC) is a non-profit, membership organization that seeks to improve the security, stability, and understanding of the Internet's DNS infrastructure.

DNS-OARC's mission is to:

• promote and conduct research with operational relevance through data collection and analysis
• offer useful services and tools
• build relationships among its community of members
• facilitate an environment where information can be shared responsibly
• enable knowledge transfer by organizing open workshops
• increase public awareness of the DNS's significance
OARC Governance

• Project founded by CAIDA and ISC in 2004
• Independent legal entity since 2008
• Diverse ~100 member base
• Financially stable and self-supporting
  • ~$800k annual revenue ~= expenses
• Self-governing, neutral
• Elected Board reflecting Member interests
• Contracted Executive Staff (~4 FTE)
• Volunteer workshop Programme Committee
• 501(c)3 non-profit public benefit corporation
What we provide to our Members and the Community

- DNS operational best-practice knowledge-sharing
- Development and maintenance of open-source DNS tools
- A range of online platforms and services to support the above
- DNS dataset collection, curation and sharing
- Collaboration venue between operators and researchers
- Workshops as a focus for all the above activities
  - “Global DNSNOG”
2019 Achievements

• The new home for DNSVIZ, dnsperf & dnsmeter
• New Member portal website operational
• OARC30 Bangkok biggest workshop ever, jointly with ICANN IDS
• New releases of dsc-datatool, dnsjit and drool tools
• Awarded Community Grant by ARIN to support open-source software maintenance
• Major dataset file store stabilization effort
• Administrative support of DNS Flag Day
OARC Workshops

- 2½ workshops per year, 2 days long
- 150-200 attendees
- Co-location with RIPE/NANOG/ICANN meetings

- OARC32
  - Feb 8\textsuperscript{th} 2020
  - San Francisco, CA
  - Co-located with NANOG78

- OARC33
  - May 9-10\textsuperscript{th} 2020
  - Paris, France
  - Co-located with ICANN IDS, GDD
Operator/Researcher Collaboration

- Our Members are a diverse mixture of operators, researchers, vendors, developers
- Operators have more data than cycles..
- Researchers have more cycles than data..
- While our Membership model is a great arena for collaboration, the resources required for supporting dataset storage and analysis are disproportionately funded from commercial Members' fees compared to researcher usage
OARC's DNS Dataset

- 230Tb, most of this DITL collections since 2004
- Other collections:
  - ZFR, Root zone archive, DSC, tester logs, resolver capture
- Data is mostly raw capture: minimal curation, metadata or semantic attributes
- Dataset is stand-alone, restrictions on export from OARC means it must be analyzed in-situ
- Remains a key resource for a core cadre of regular researchers, historical perspective has proven invaluable at various points
  - e.g. Name Collisions study
- Regular storage platform infrastructure upgrades have expanded raw storage with the dataset
  - but have not updated or scaling processing capacity to keep pace
Storage Infrastructure Challenges

- We had various issues with this in 2019, and have been unable to post-process several recent DITL collections as a consequence
- Various non-upgraded elements have aged
- While we are keeping the dataset in a stable state, it’s burning scarce sysadmin resource to keep it that way
- The nature of OARC’s dataset and current use policies do not make it amenable or economic to do storage/analysis in a 3rd-party cloud
- The resource drain of gathering and maintaining DITL data is impairing OARC’s ability to do many other activities
- Preserving the status quo is not necessarily cheaper nor safer than a major upgrade programme
What we are Planning

- New Ceph-based scalable storage architecture proposal:
  - https://indico.dns-oarc.net/event/32/contributions/736/attachments/702/1194/Filesystem-Clustering.pdf
  - with some bootstrap funding, running and growing this will be cheaper than our current infrastructure
- Surveying Members to identify resource needs and potential sources
- Seeking further funding to develop privacy-aware DNS tools
- Enhancements to DNSVIZ
- Board committee has been formed to update privacy policy:
  - meet post-Snowden/GDPR challenges
  - ideally enable use of cloud-based resources
What we would like to do

• Store our dataset in some kind of database to enable easier and more meaningful analysis

• Facilitate ongoing processing of new and existing data in ways that respect modern privacy models

• Restore realtime DNS telemetry sharing
  • e.g. DSC-Grafana

• Continue to facilitate equitable 2-way data ↔ knowledge sharing between operators and researchers

• Be a co-operative building block in a wider data sharing/analysis ecosystem
Questions/Discussion