DNS Research Update from CAIDA

Status and Recent Experiences

kc claffy

RSSAC
March 22, 2009

CAIDA DNS Research Goals and Support

- CAIDA supplies the research community with DNS measurement data, tools, models, and analysis methodologies for use by DNS operators and researchers.

- CAIDA receives support for DNS research from NSF grant SCI-0427144 "Improving the Integrity of Domain Name System (DNS) Monitoring and Protection". (ends Aug 09, final report by Sept)

DNS Related Measurements and Activities

- Measurement of traffic at DNS root name servers
  - DITL 2006, 2007, 2008, and prep for 2009
  - a comparison of traffic from DITL 2006 and 2007
  - in search of “heavy hitters”
  - analysis of DITL data

- DNS Research Community Interaction

- DNS Names for IPv4 Routed /24 Topology Dataset

DNS Related Measurements and Activities

- DNS DITL measurements
- Ongoing Open Resolver Surveys (Duane, OARC)
- Realistic Simulation of BGP (Riley08, Castro09)
- DNSNames, DNS traffic (CAIDA’s Ark project)
- DNSParse (Nevil Brownlee)

Traffic Measurement at DNS root servers

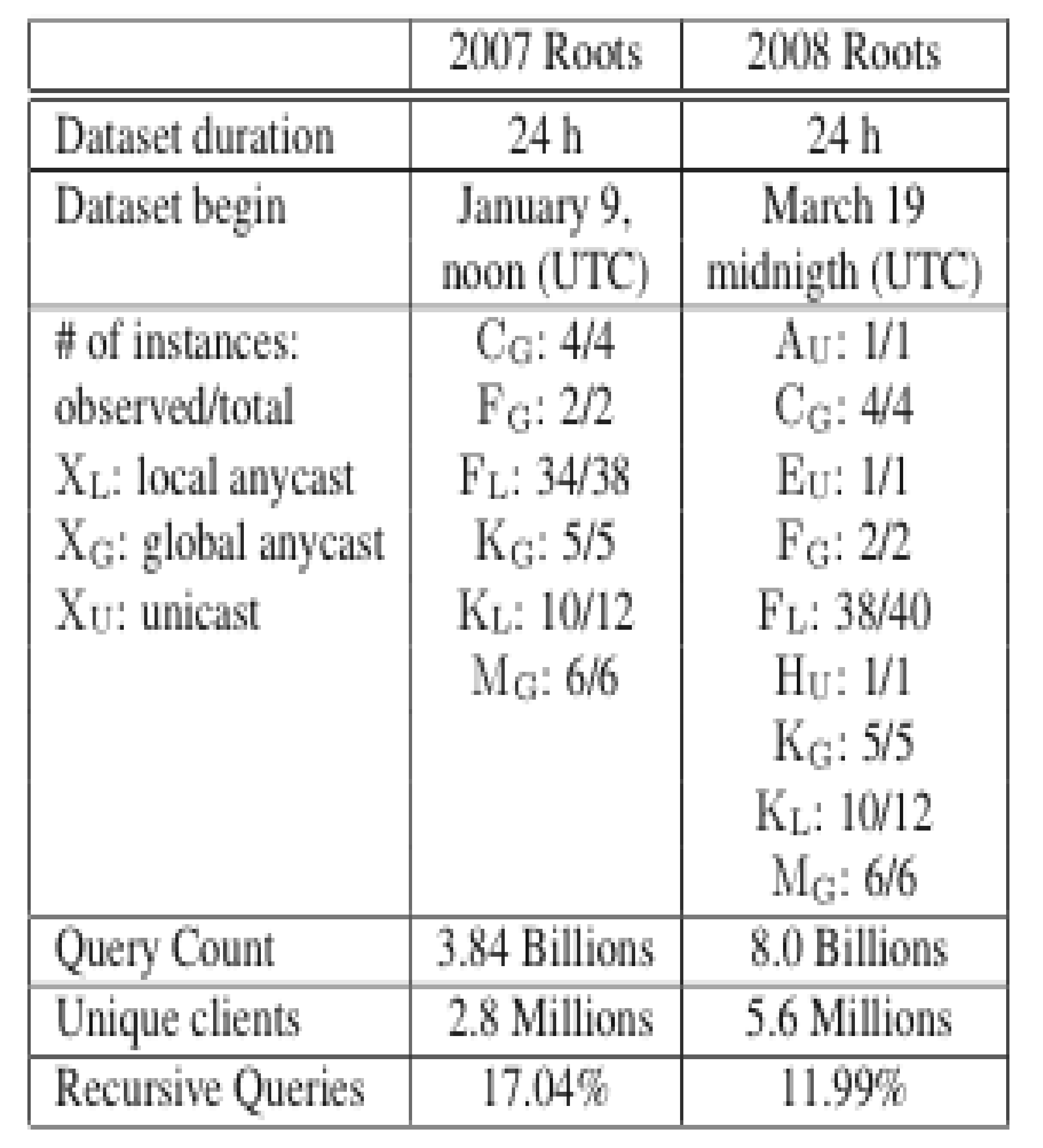
- In 2006, 2007, and 2008, CAIDA with DNS-OARC coordinated several collection events nicknamed a Day In The Life of the Internet. For 2008, Duane presented at NANOG42
  - During the March 2008 DITL, 8 of 13 root operators, 5 top level domains, 2 regional registries, 6 AS112 nodes, 2 open root server nodes, and 2 caching resolvers participated.
  - Over 2TB of traces in pcap format
  - http://www.caida.org/projects/ditl/

Comparison of DITL 2006 and 2007 traffic

- comparison of several parameters, including query rate, client rate, query type distribution, percentage of clients switching between anycast instances, client persistence, traffic validity, EDNS support and EDNS buffer size.
- Conclusions: anycast appears stable, efficient, and responsive, overall traffic is growing, 98% of queries are pollution, 40% support EDNS, ORSN see similar traffic.

Analysis of DITL Data

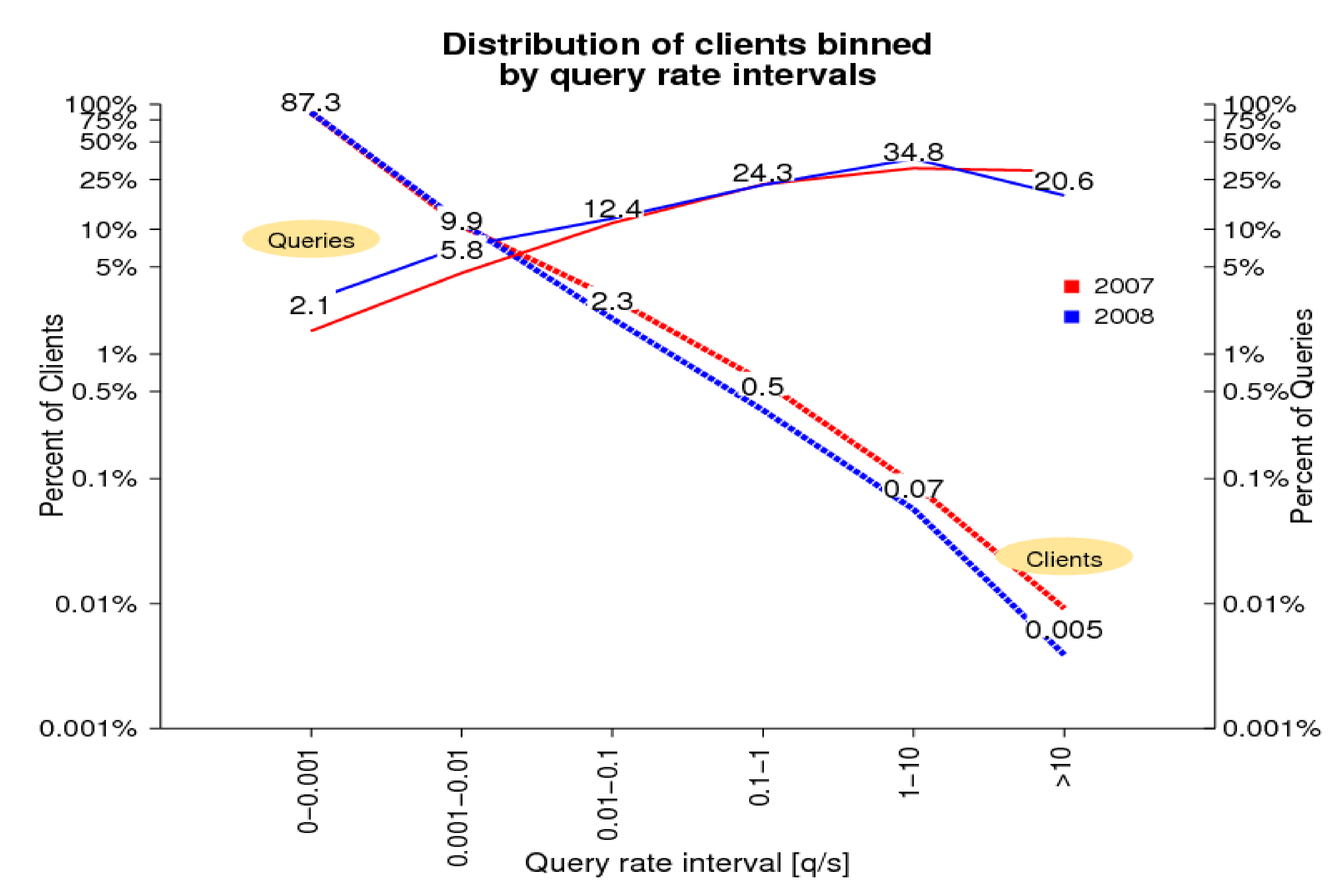
- Published “A Day at the Root of the Internet” in Computer Communications Review, Oct 2008
  - used 2006, 2007, 2008 DITL data
  - a huge number of clients sending a few queries only to the roots
  - an increase in AAAA queries
  - higher the query rate the lower the fraction of valid queries
  - fraction of traffic due to invalid TLD is huge! (25%)
  - confirmed 98% of traffic to roots should not be there at all



	2007 Roots	2008 Roots
Dataset duration	24 h	24 h
Dataset begin	January 9, noon (UTC)	March 19 midnigth (UTC)
# of instances: observed/total X_L: local anycast X_G: global anycast X_U: unicast	C_G: 4/4 F_G: 2/2 F_L: 34/38 K_G: 5/5 K_L: 10/12 M_G: 6/6	A_U: 1/1 C_G: 4/4 E_U: 1/1 F_G: 2/2 F_L: 38/40 H_U: 1/1 K_G: 5/5 K_L: 10/12 M_G: 6/6
Query Count	3.84 Billions	8.0 Billions
Unique clients	2.8 Millions	5.6 Millions
Recursive Queries	17.04%	11.99%

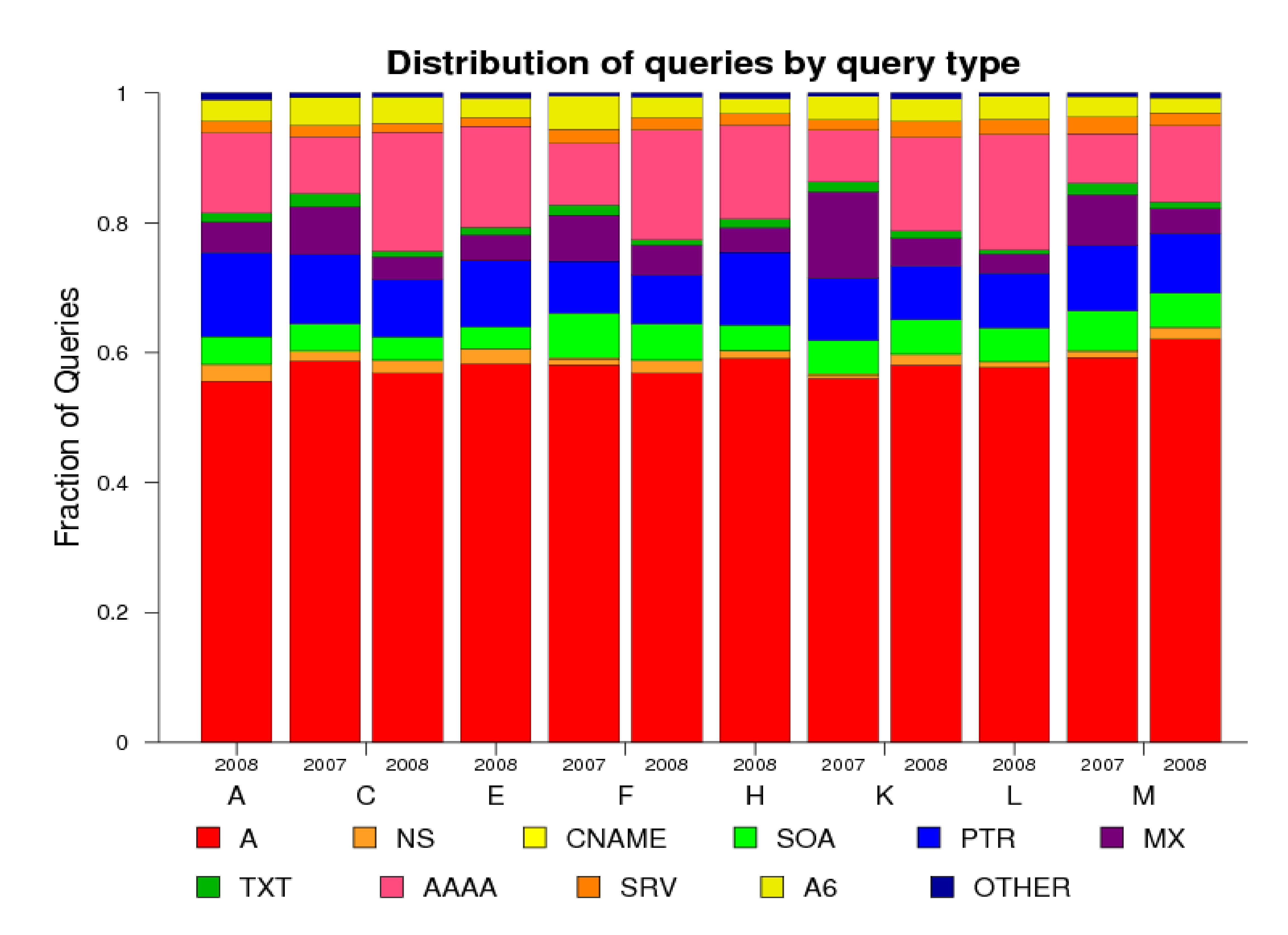
Table 1: General statistics of the 2007 and 2008 datasets

“Days at the Root of the Internet”



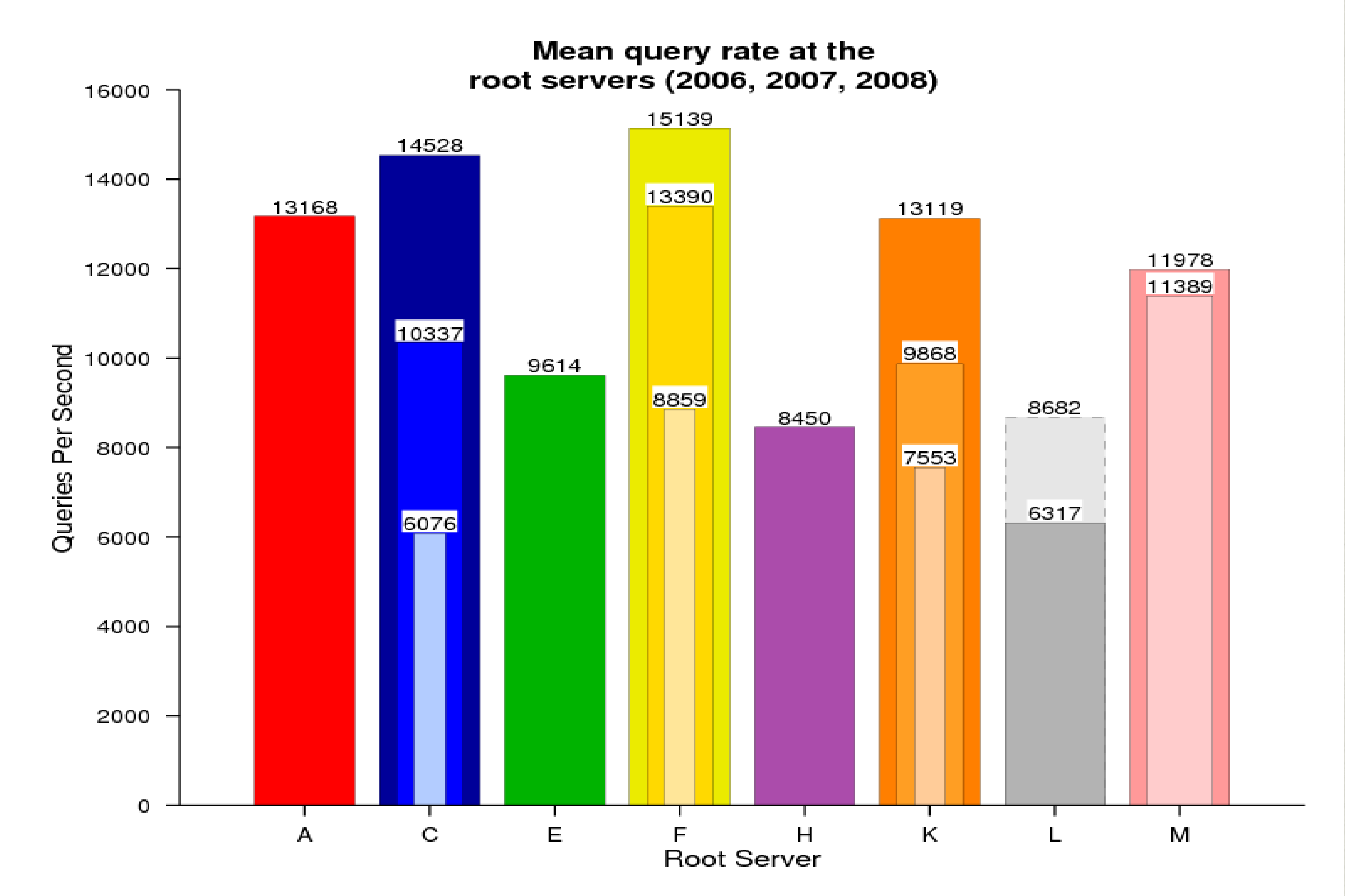
Just a few clients (two rightmost categories) responsible for > 50% of queries to observed root servers.

“Days at the Root of the Internet”



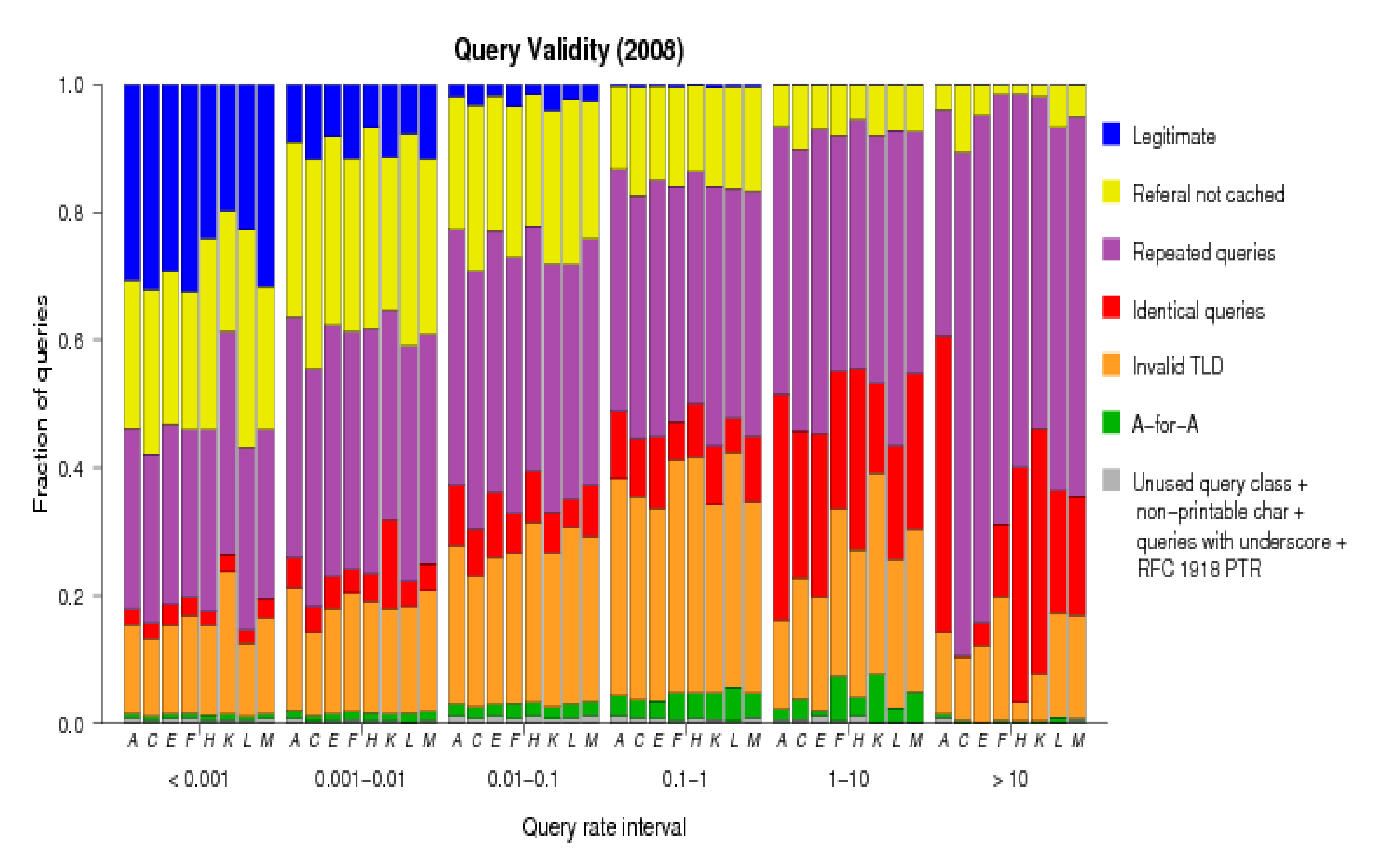
AAAAs growing due to IPv6 deployment

“Days at the Root of the Internet”



Light to dark are average query rates in 2006, 2007, 2008

"Days at the Root of the Internet"



Strong correlation between query rate and pollution rate

In Search of “Heavy Hitters” (polluters)

- slides presented at DNS-OARC workshop June 2008
- wanted to understand nature of clients making *many* queries (more than 10 q/s per root)
- “super heavy hitters” sending > 40 q/s per root
- sources of high traffic change with time
- active probing closer in time to collection may help
- need better methods of analyzing the data, e.g., machine learning

In Search of “Heavy Hitters” (polluters)

- CAIDA visiting graduate student, “Reducing pollution at DNS Root servers” did not get far. (lost student)
  - compared polluting addresses with blacklists
    - Spamhaus and UCEProtect
  - find no significant correlation between address lists. In general polluting addresses do not overlap with blacklists.

DNS Pollution: Related Work

- Daniel Sanchez & Joost Pijnaker, two students at the University of Amsterdam (UvA) – pollution at root servers http://staff.science.uva.nl/~delaat/sne-2006-2007/p21/report.pdf

  - suggestions for fixing pollution at the roots:
    - Install and use stable secure, patched applications (reduce A-for-A, priv,iden)
    - Use stable hardware and update firmware (reduce invalid)
    - Configure software appropriately: local names in hostfile, firewalls allow responses (reduce local/repeated queries)
    - Configure DNS server correctly. appropriate TTL reduction of the 'no caching of TLDs' queries.
    - contact software, hardware vendors to fix problems.
    - access lists on/near root servers (of heavy polluters)
    - uRPF (?)
    - contact polluters, ask them to stop
    - *** overprovision ****

Traffic Measurement at DNS roots (cont)

- Lots of lessons learned re: data management
  - timing is critical; early notice increases participation
  - data supplier agreement needs to be simple
  - System requirements critical:
    - plan for disk,
    - dry-run collection and upload,
    - adjust for local configurations,
    - allow participants to easily track upload status,
    - log MD5 checksums of files, and
    - maintain local copies of data, space allowing

DITL 2009 Collection Event

- DNS-OARC leading 2009 effort
  - https://www.dns-oarc.net/ditl/2009
- CAIDA helping to develop tools and dry run
  - dnscap – added TCP support
  - upload tracking: http://caida-oarc.caida.org/ditl_200903/
  - *(ditl analyst sebastian castro visiting caida from nic.cl and uchile is moving to .au in july)*

DNS Research Community Interaction

- DNS-OARC: The DNS Operations, Analysis, and Research Center (DNS-OARC) brings together key operators, implementors, and researchers on a trusted platform so they can coordinate responses to attacks and other concerns, share information and learn together.
  - incident response
  - operational characterization
  - testing
  - analysis
  - outreach

DNS Research Community Interaction (cont)

- CAIDA-WIDE Workshop Series
  - twice a year meetings
  - informal research discussions, ops feedback
  - open to roots
  - synergy with DNS-OARC workshops

DNS Names

- automated ongoing DNS lookup of IP addresses seen in the Routed /24 Topology traces
  - all intermediate addresses and *responding* destinations
  - using our in-house bulk DNS lookup service (HostDB)
    - can look up millions of addresses per day
- 240M hostnames since March 2008
- http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml

DNS Traffic

- tcpdump capture of DNS query/response traffic
  - only for lookups of Routed /24 Topology addresses
  - continuous collection of 3-5M packets per day
  - can download most recent 30 days of pcap files
- a broad sampling of the nameservers on the Internet due to the broad coverage of the routed space in traces
- how many nameservers have IPv6 glue records? DNSSEC records? support EDNS? typical TTLs?

Open Resolver Surveys (Duane Wessels)

- identifies nameservers that provide recursive name resolution for clients outside of their administrative domains.
- ongoing active measurements since June 2006
- http://www.caida.org/research/dns/surveys/open-resolvers-surv
- plan is to donate software/surveys to OARC

Realistic BGP Simulation (riley@gatech)

- Published “*Realistic Topology Modeling for the Internet BGP Infrastructure*” in Modeling, Analysis and Simulation of Computers and Telecommunication Systems, 2008. IEEE MASCOTS 2008.
  - Used DITL data for validation
  - Metrics of interest: BGP convergence and churn
  - Limitations of model
    - little known about internal topologies of tier 1 and 2 ISPs, link speeds, speed of light delays, filtering
- abstract on www.caida.org/publications/papers/

Realistic BGP Simulation (sebastian@nic.cl)

- Masters thesis at U.Chile, “Modeling, Analysis and Simulation of Anycast”
  - Used .cl data for validation
  - Metrics of interest: anycast stability
  - Limitations of model
    - See previous slide
- Paper to be submitted in may 2009

DNSParse **(Nevil Brownlee)**

- New DNS monitoring initiative
- Global mesh of DNS sensors feed a central collector/database at U. Auckland
- 8 sensors
- Database contains 800,000 entries

Summary of Milestones

- Collected traces from nearly all anycast instances of A, C, E, F, H, old-J, K, L, old-L, and M root servers and from two alternative Open Root Server Network (ORSN) servers on March 18-19, 2008.
- Published a paper on "A Day at the Root of the Internet" by S. Castro, D. Wessels, M. Fomenkov, and k claffy in Computer Communications Review in October, 2008 . http://www.caida.org/publications/papers/2008/root_internet/
- Presented at NANOG42 a Day In The Life of the Internet 2008 Data Collection Event. http://www.caida.org/publications/presentations/2008/nanog_dw_ditl/
- Presented an analysis of the 2008 data at the DNS-OARC 2008 DNS Ops Workshop in Brooklyn. http://www.caida.org/publications/presentations/2008/oarc_castro_ditlanalysis/
- Indexed the DITL 2007 and DITL 2008 data into DatCat.

Summary of Data Released

- OARC DNS root traces for January 10-11, 2006, January 9-10 2007, and March 18-19, 2008.
  http://imdc.datcat.org/collection/1-00BC-Z=OARC+DNS+root+traces+January+10-11%2C+2006
  http://imdc.datcat.org/collection/1-031B-Q=Day+in+the+Life+of+the+Internet%2C+January+9-10%20%2C+2007+%28DITL-2007-01-09%29
  http://imdc.datcat.org/collection/1-05MM-F=Day+in+the+Life+of+the+Internet%2C+March+18-19%2C+2008+%28DITL-2008-03-18%20%29

- five years of DNS RTTs from several campuses to root/gTLD servers
  http://www.caida.org/data/passive/dns_root_gtld_rtt_dataset.xml

- daily reports identifying open DNS resolvers
  http://dns.measurement-factory.com/surveys/openresolvers/ASN-re

- database of reverse DNS lookups
  http://www.caida.org/data/active/ipv4_dnsnames_dataset.xml

Future Work

- Help analyze DITL 2009 data -- workshop?
- sebastian’s masters thesis: simulation of anycast at .cl
- Final report on DNS-ITR project
- DHS-funded workshops (“belmont report”, network research agenda, best anonymization and data-sharing practices)
- Better data-sharing models (OARC, SIE, PREDICT, DatCat)