

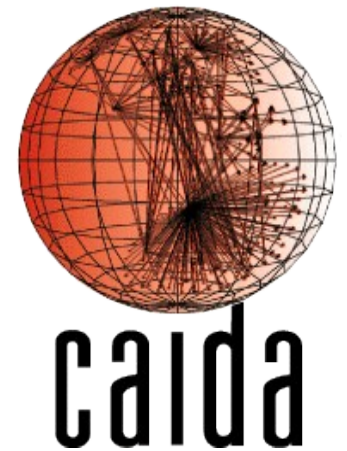


DITL

A Day in the Life of Internet

Sebastian Castro
sebastian@nzrs.net.nz

NZRS / CAIDA



Overview

- What's DITL and its motivation
- About CAIDA/DNS-OARC
- Evolution of the collection
- The DNS Root Servers data
 - Analysis
 - Trends
- Lessons learned
- What's happening now!
- Conclusions

What is DITL and its motivation

- U.S. National Academy of Science, 2001
 - Challenge to the research community
 - Network measurement
- CAIDA and DNS-OARC coordinated and organized the collection of data in the DNS root servers
 - Could be considered a prototype
 - Based on a trust relationship with operators
 - Been done yearly since 2006
 - Includes other sources of data

About CAIDA

- Cooperative Association for Internet Data Analysis
 - Research group based at the San Diego Supercomputer Center of UCSD
 - Long term involvement in Internet research and data collection
 - Macroscopic study of the Internet
 - Facilitate data collection, analysis and sharing
 - Help the development of informed public policies through data
 - <http://www.caida.org>

About DNS-OARC

- DNS Operations, Analysis, and Research Center
 - Non-profit, membership-based organization
 - Brings together key operators, implementors and researchers
 - Key functions
 - Information sharing
 - Operational Characterization
 - Workshops
 - Analysis
 - Tools and Services
 - <http://www.dns-oarc.net>

Evolution of the collection

	DITL 2006	DITL 2007	DITL 2008	DITL 2009
Participants	3 root servers	5 root servers 2 alternative root servers 1 AS112 instance 5 passive traces	8 root servers 2 alternative root servers 2 RIR 5 TLD 7 AS112 instances 6 passive traces 2 caching DNS servers	8 root servers 2 alternative root servers 3 RIR 8 TLD 6 AS112 instances 5 passive traces
Duration	46.2 hours	48 hours	48 hours	72 hours
Collection times	January 10-11	9-10 January	18-19 March	30 March – 1 April

How the data is collected

- How to collect
 - Port mirroring, network tap
 - Raw packets, usually pcap
- Tools
 - DNS traffic uses tcpdump/dnscap/ncap
 - For other traffic, depends
- Data Availability
 - DNS traffic is uploaded to central location
 - For other traffic, depends

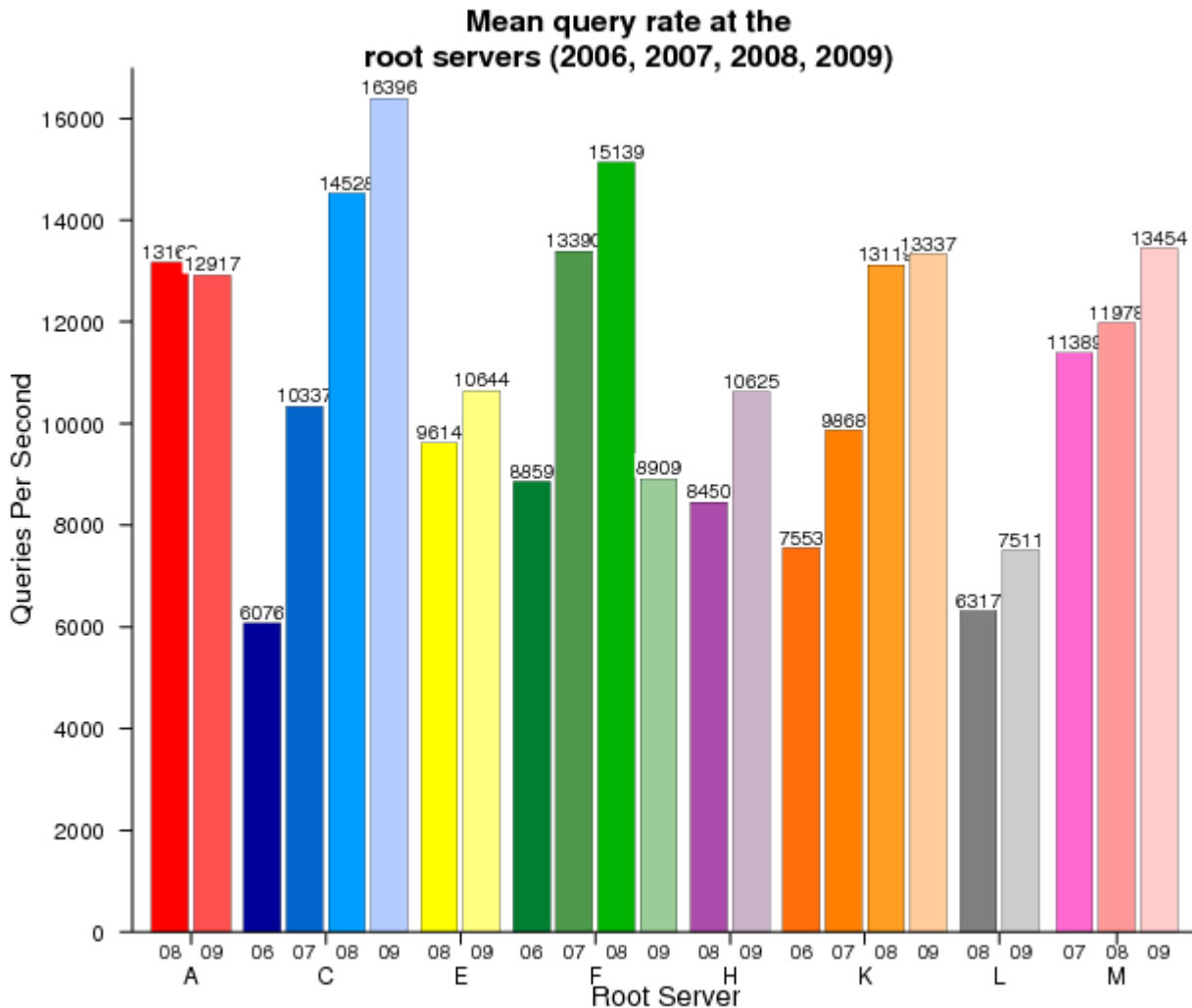
The DNS Root Servers data

- The DNS Root Servers
 - 13 names: [a-m].root-servers.net
 - Under the administration of 12 different organizations
 - Big footprint: one name can have 70+ instances
 - Key piece of Internet infrastructure
- Data
 - Mostly pcap files
 - Selected the best 24 hours for analysis

General Statistics table

	DITL 2007	DITL 2008	DITL 2009
Dataset duration	24h	24h	24h
Dataset start (UTC)	January 9, noon	March 19, midnight	March 31, midnight
Number of instances		A: 1/1 C: 4/4 E: 1/1 F: 36/40 H: 1/1 K: 15/17 L: 2/2 M: 6/6	A: 1/1 C: 6/6 E: 1/1 F: 36/48 H: 1/1 K: 16/17 L: 2/2 M: 6/6
Query count	3.84 billion	7.99 billion	8.09 billion
Unique clients	~2.8 million	~5.6 million	~5.8 million
Recursive queries	17.04%	11.99%	9.76%
Compressed size	164 G	278G	281G

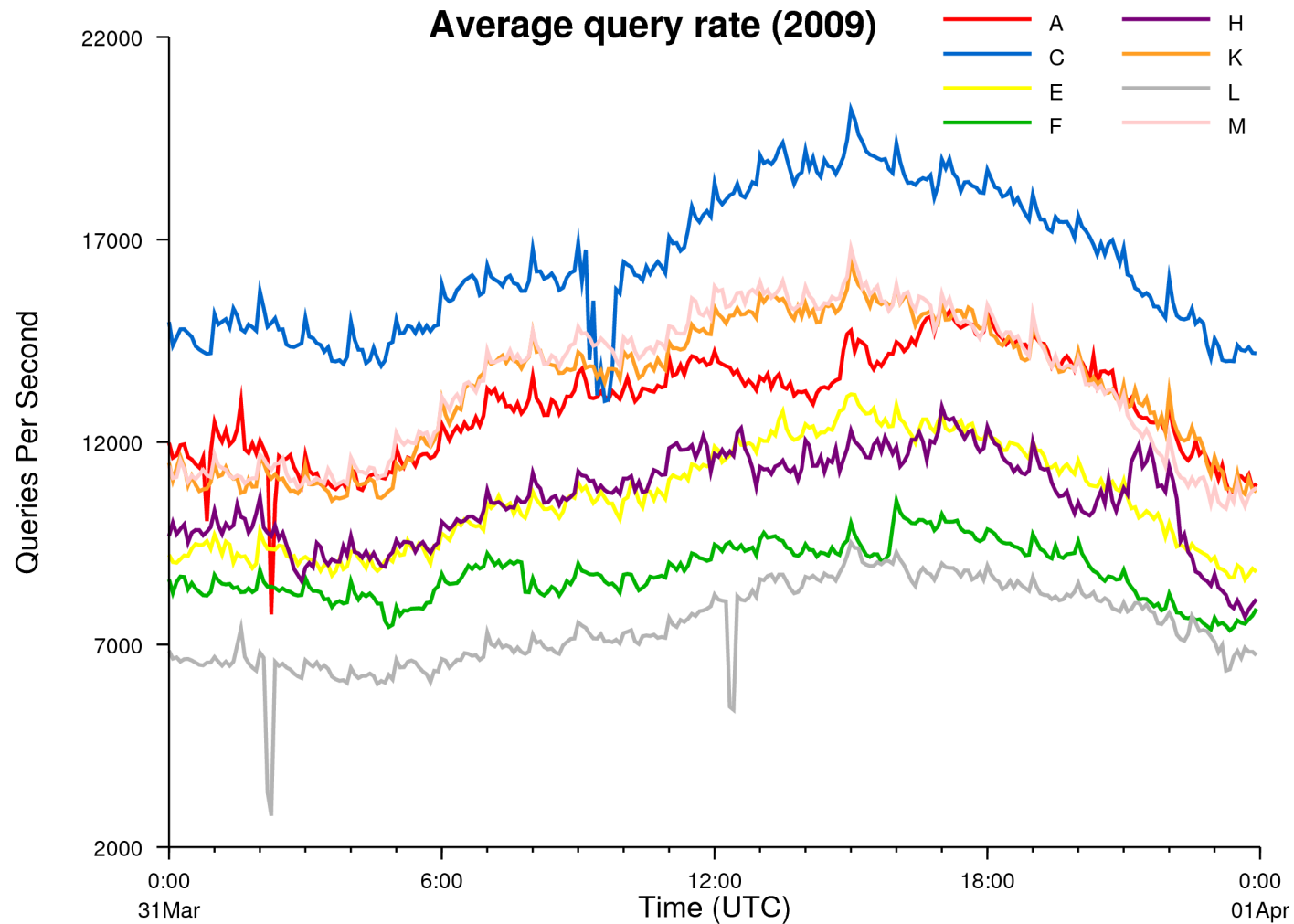
Mean query rate growth



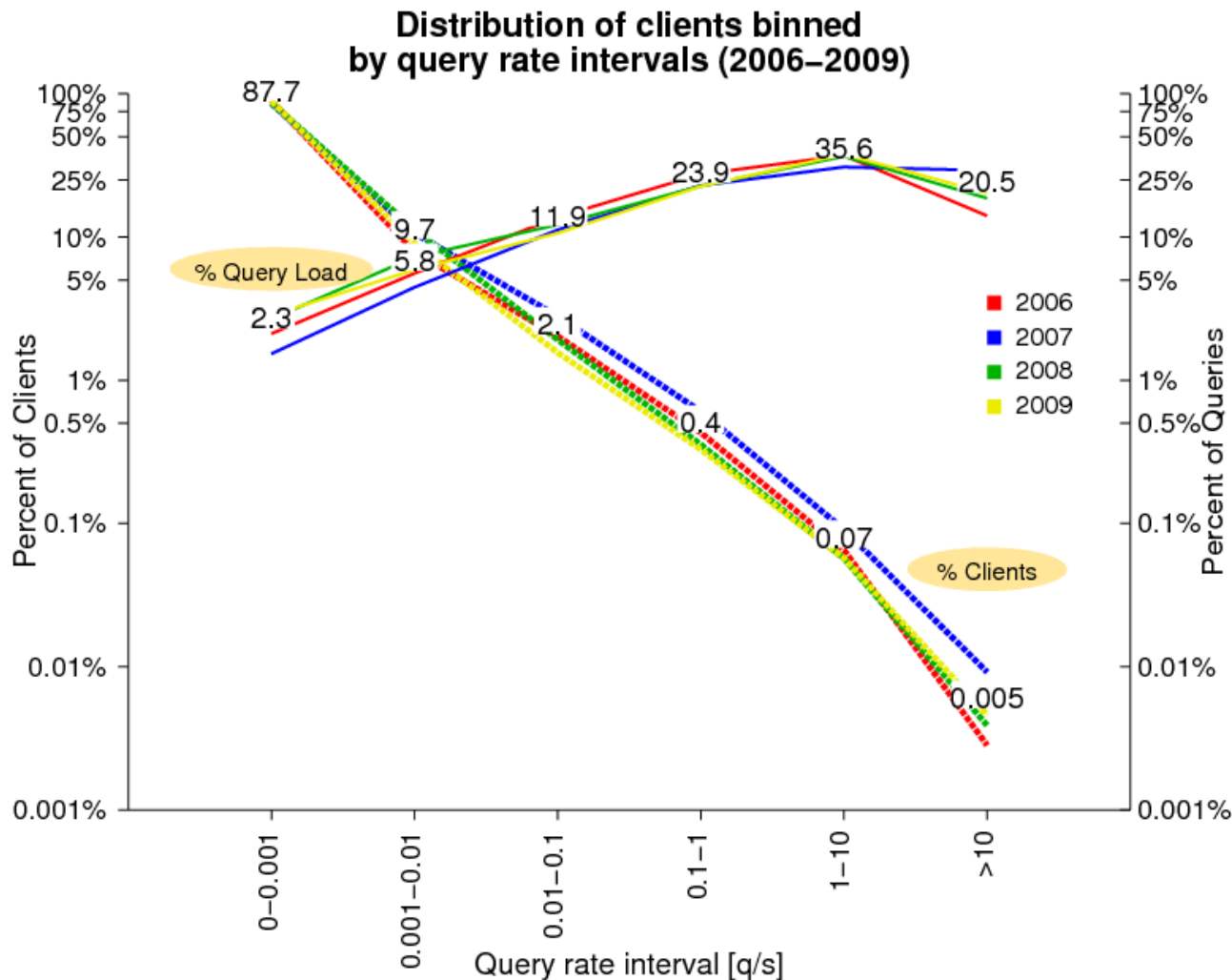
Root	Growth 2007-2008	Growth 2008-2009
A		-1.91%
C	40.54%	12.86%
E		10.71%
F	13.06%	-41.15% *
H		25.74%
K	32.94%	1.66%
L		18.90%
M	5.17%	12.32%

* Data from f-sfo (global instance) was not collected

Average query rate

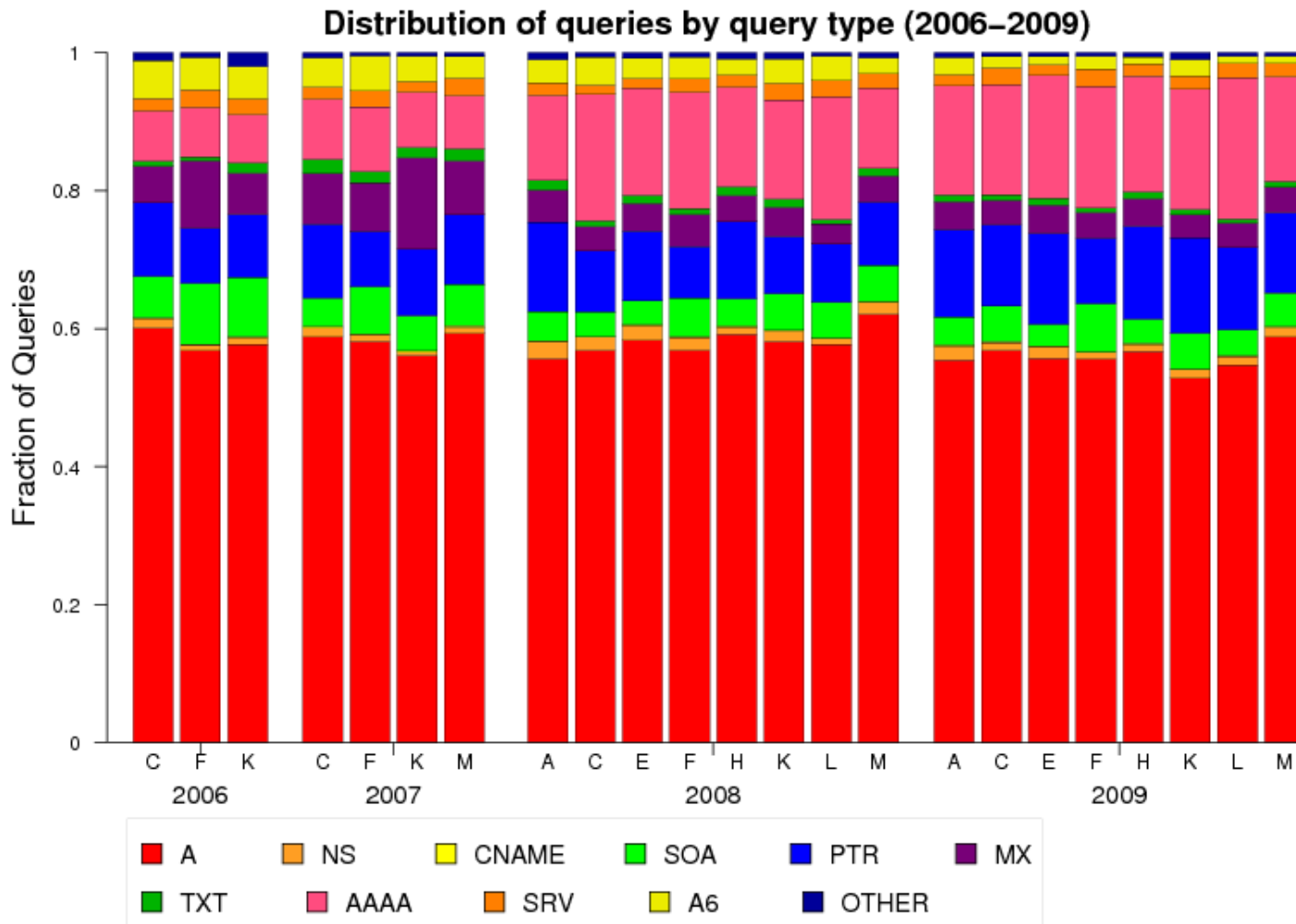


Query rate per source



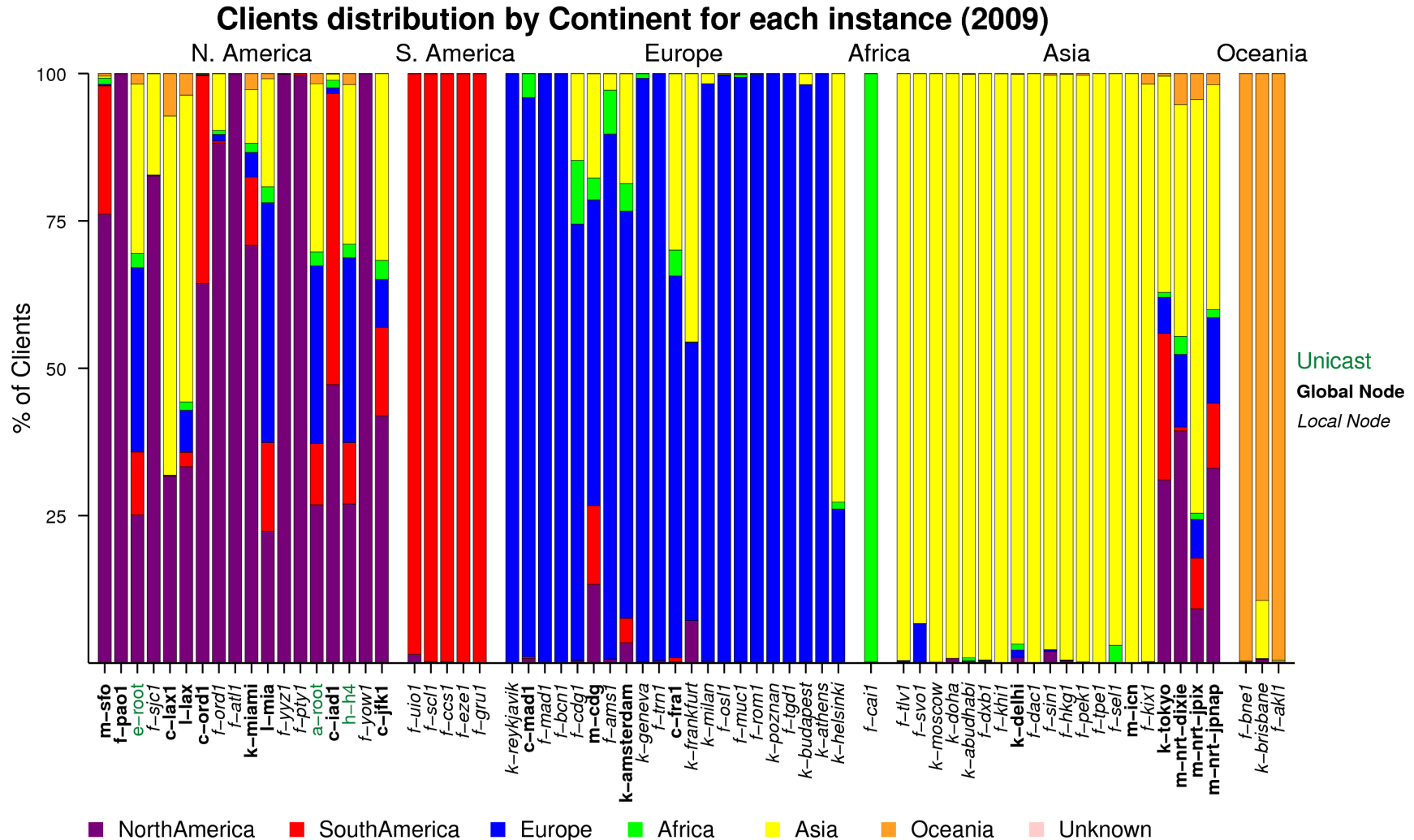
- For each source address we calculate the mean query rate
- Group the sources by their query rate in log bins
- Count the amount of traffic generated by each group
- Most of the clients generate a few queries
- Few clients generate most of the traffic

Query type distribution



- A-queries are still the most popular ~58%
- AAAA-queries has gained popularity
- A6 queries still there? (Deprecated in 2002)

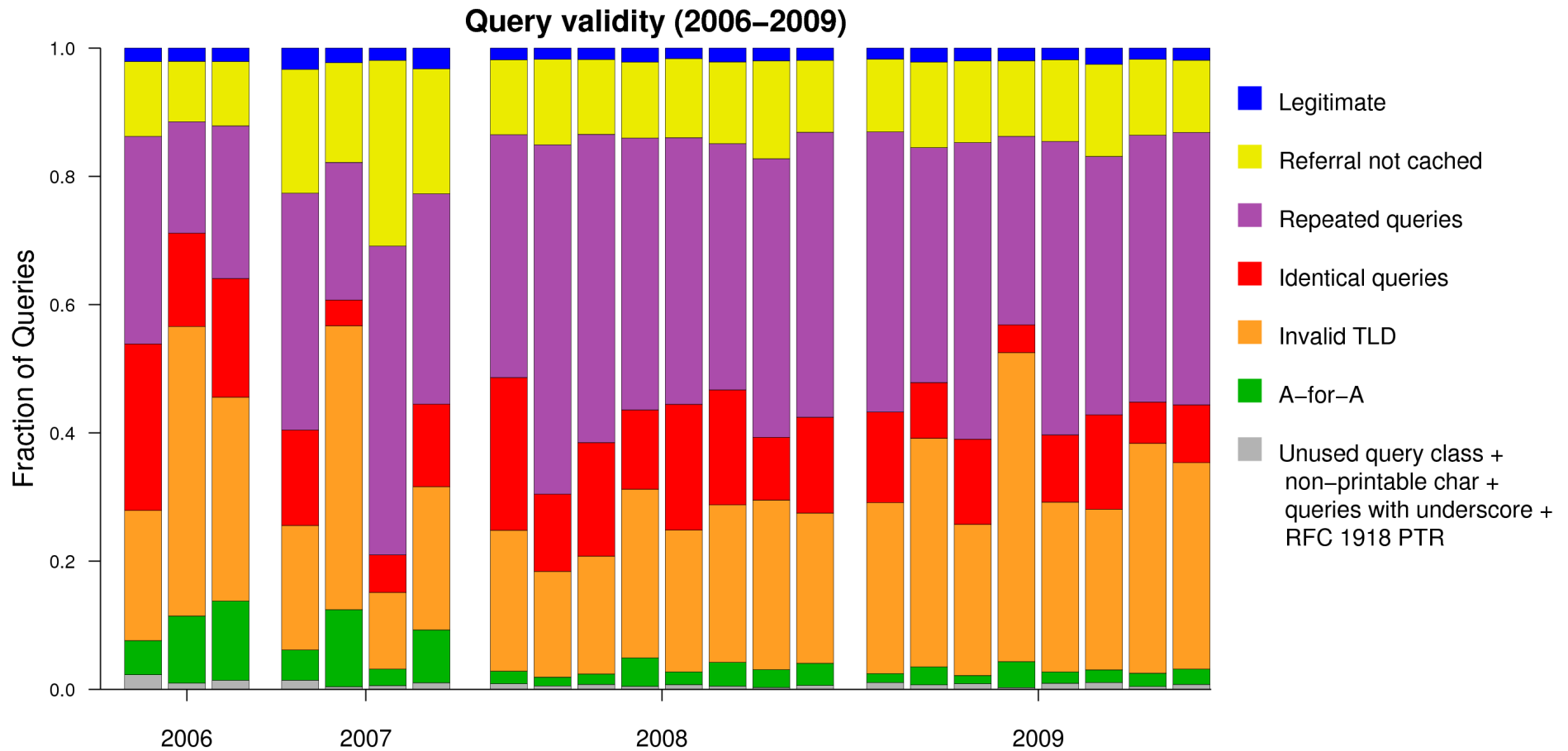
Client geography



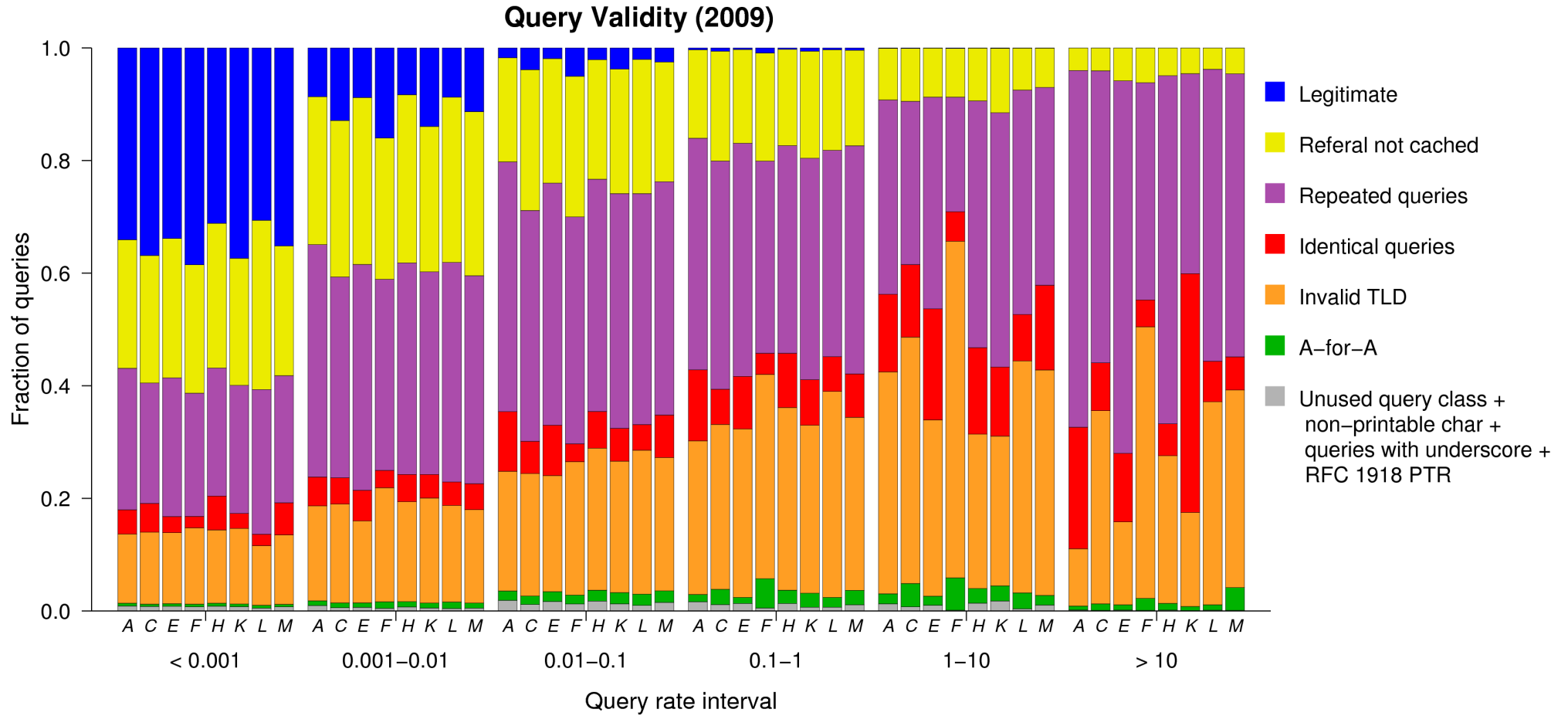
Query pollution

- Pollution: Certain types of queries that should not be seen
 - Invalid TLDs
 - Repeated/Identical queries
 - Referral not cached
 - A for A
 - Various other types
- Based on a paper by Wessels *et al* from PAM 2002

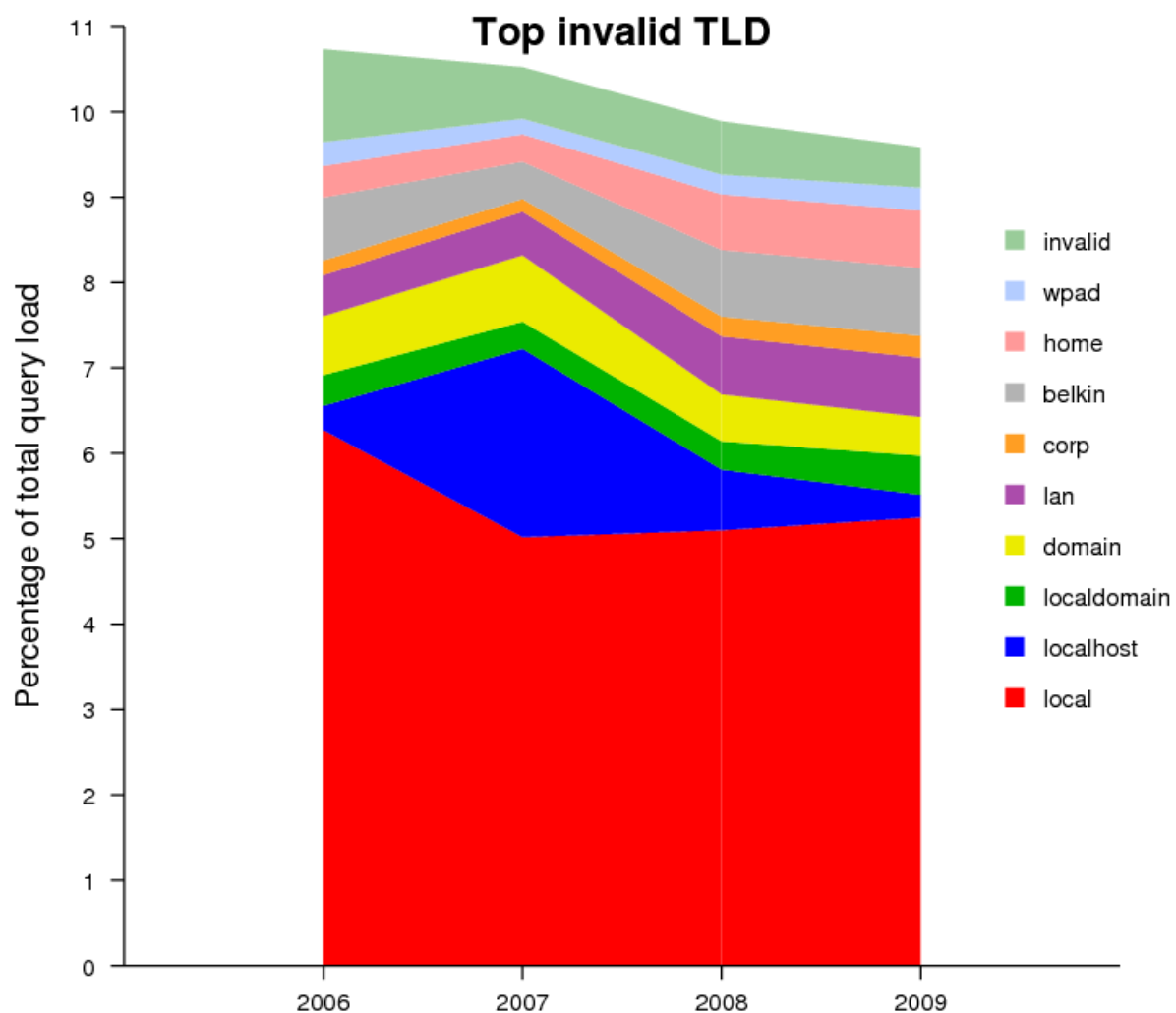
Query pollution



Query pollution



Traffic for invalid TLDs

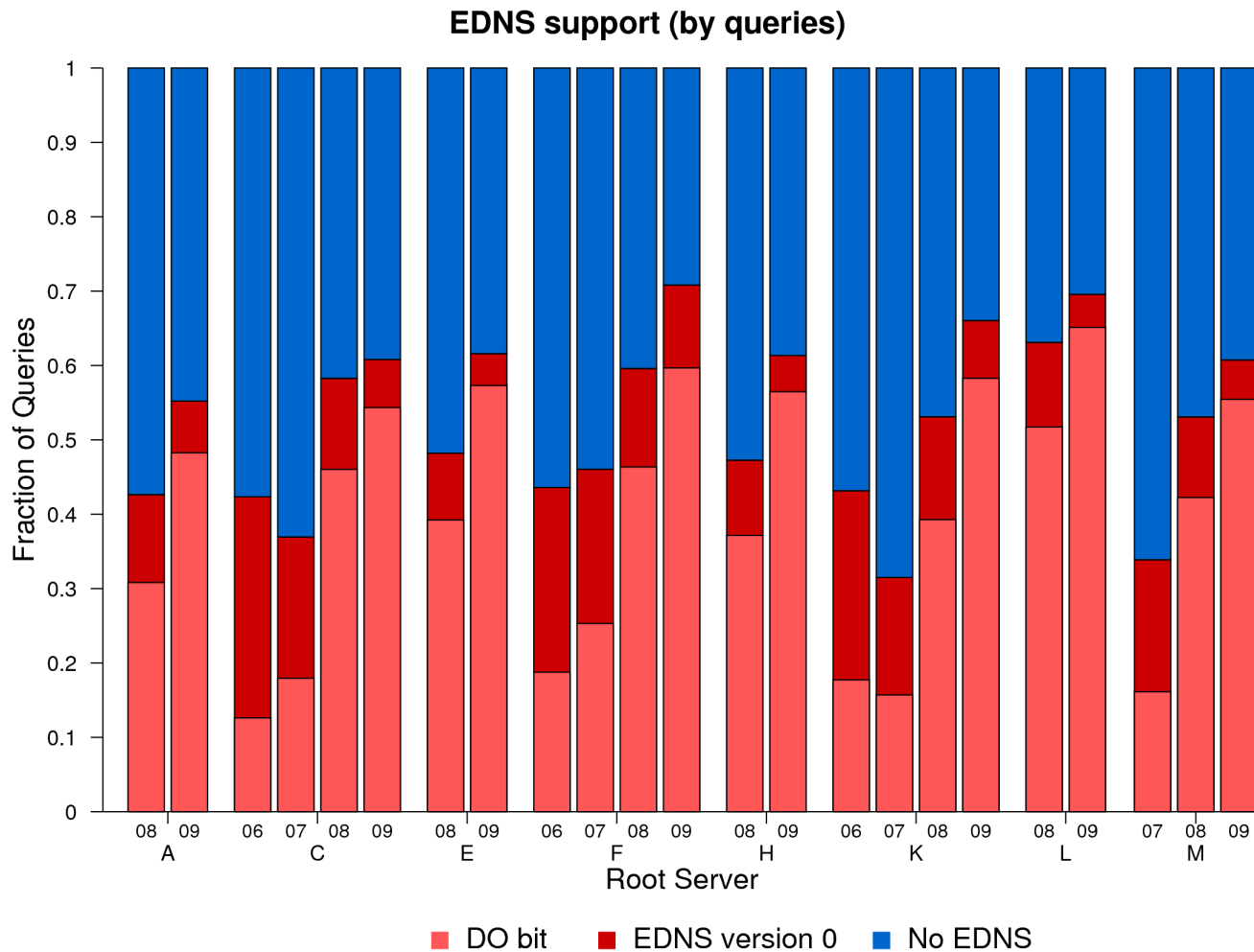


- 10 invalid TLDs represent 10% of the **total** query load at the root servers
- The TLD has not changed in the last four years (only the ranking)
- If all invalid TLDs are included, the percentage moves from 18% to 26%

Trends: EDNS

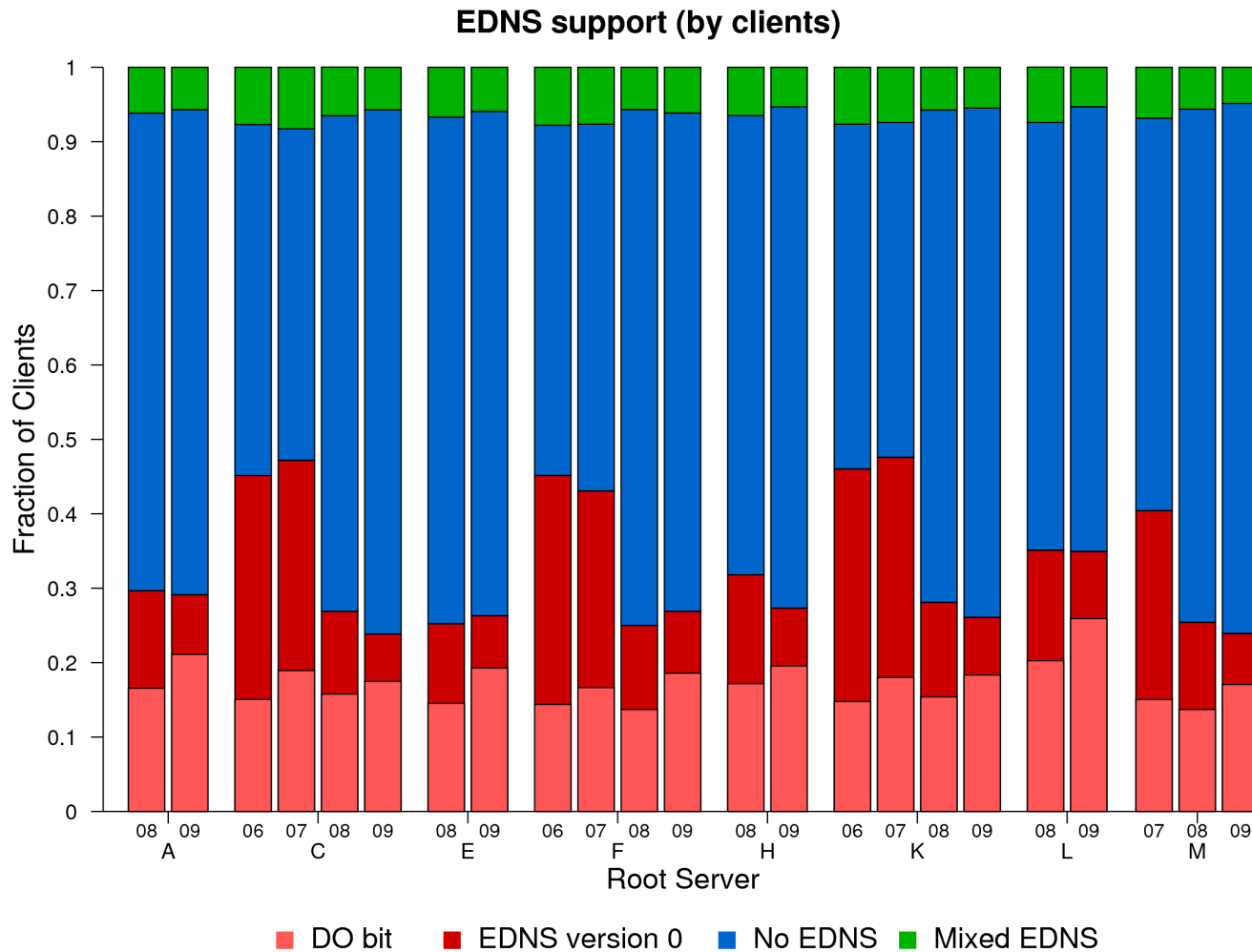
- EDNS: Enable longer responses (> 512 bytes) over UDP
- Pre-requisite for DNSSEC deployment
 - EDNS DO Bit: DNSSEC OK → enabled client
- Two ways to measure support
 - Per query
 - Per client (by checking all the queries from the same source address)

EDNS at the query level



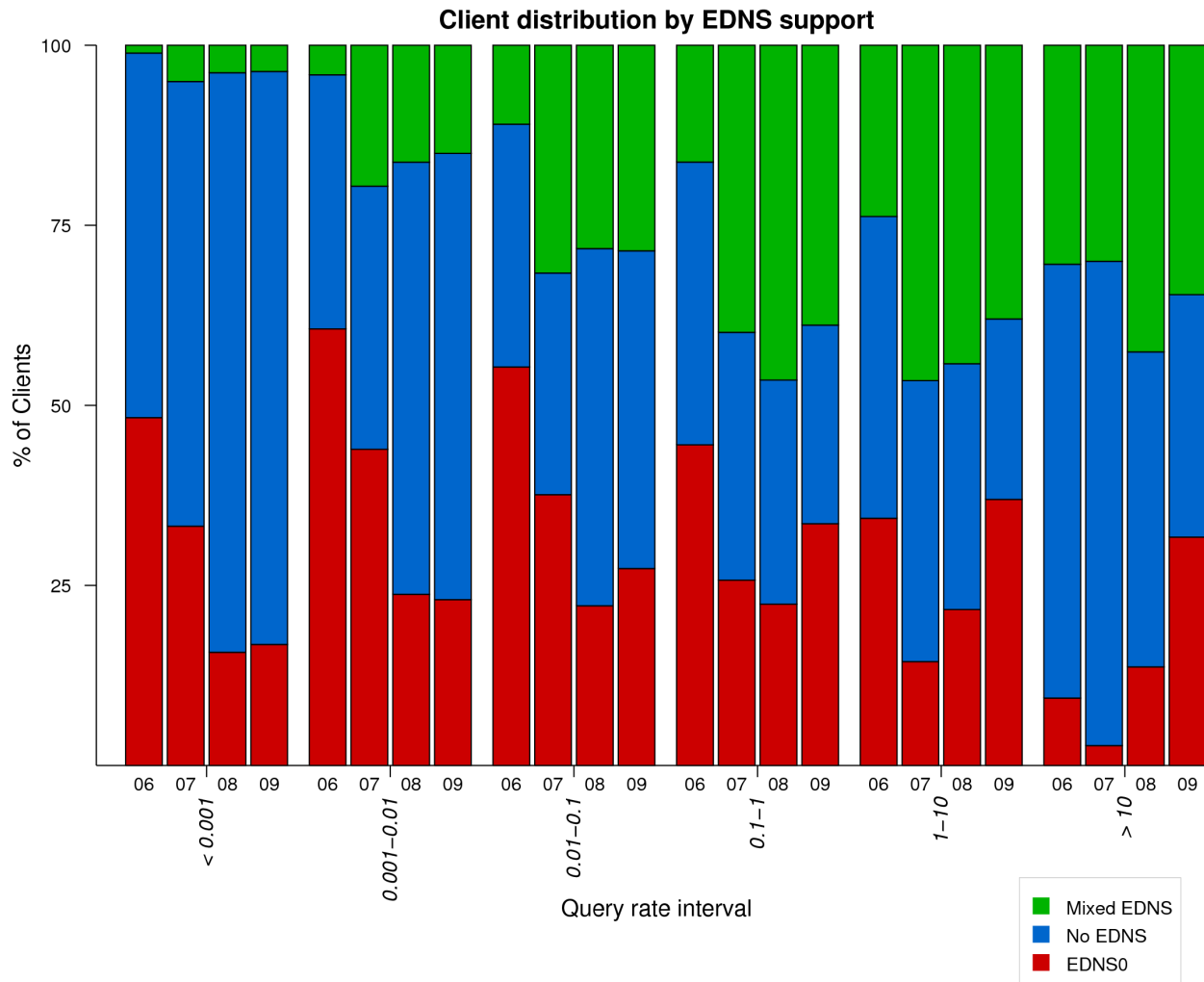
- Clear growth of EDNS, with a jump from '07 to '08
- Over 90% of the EDNS capable queries are DO enabled in 2009
- Good news, right?

EDNS at the client level



- At the client level, the situation is totally the opposite!
 - Reduced along the years
 - Around 30% support
 - The DO enabled/EDNS capable queries ratio is in the 60-70% range
 - How is this explained?
 - **The heavy hitters**

EDNS per query rate



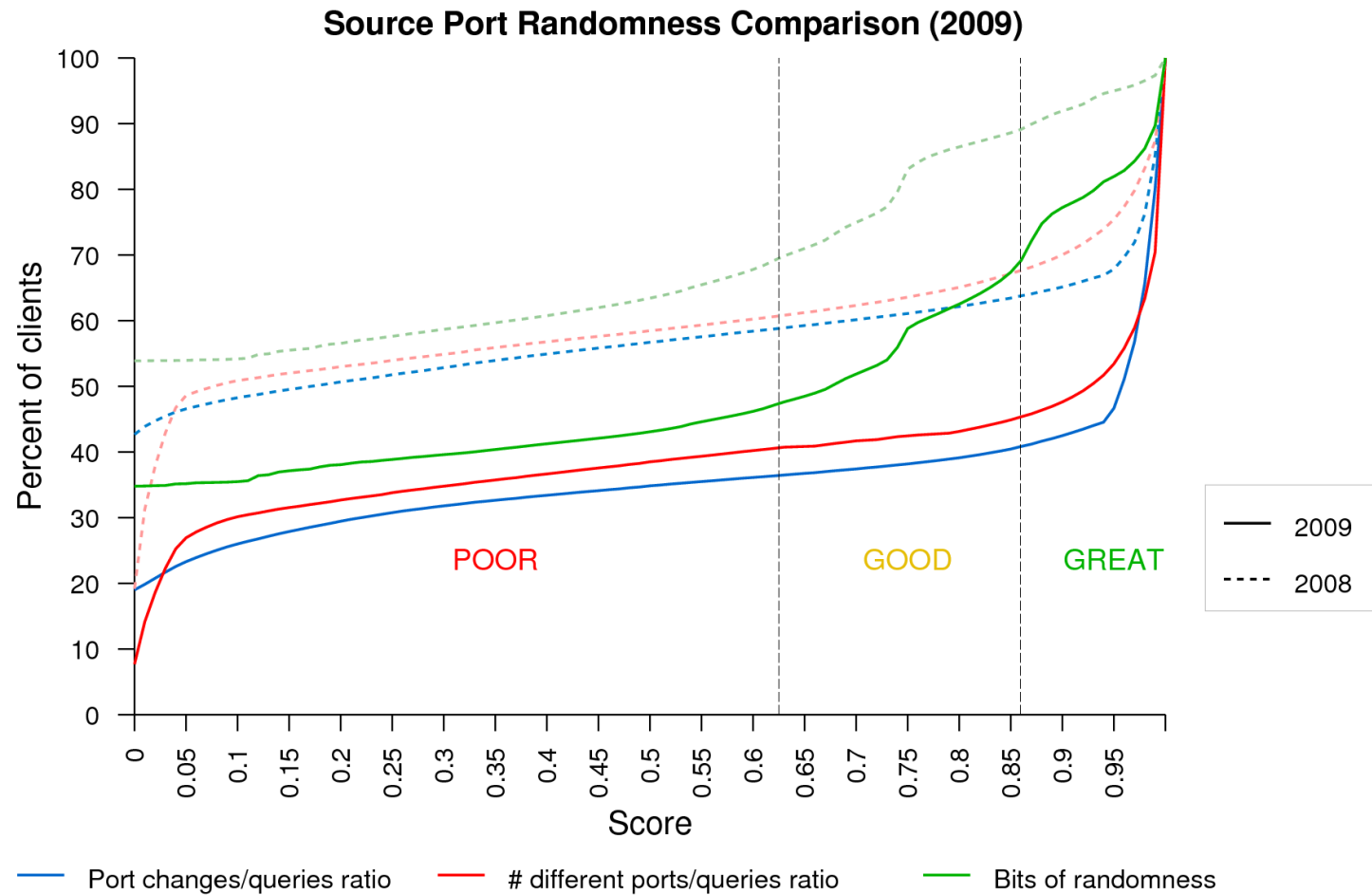
How we can explain the difference?

- We grouped the clients by their query rate
- Clients sending few queries present less EDNS support
 - And they represent most of the clients
- Client sending lots of queries present more EDNS support
 - Most of the queries (>50%) are generated by the two rightmost categories
 - Most of their queries are pollution :(

Source Port Randomization evolution

- Relevant since the Kaminsky bug
- Three scores for each client (if # queries > 20)
 - # of port changes/queries ratio
 - # different ports/queries ratio
 - These two proposed by nic.at
 - Bits of randomness
 - Proposed by Duane Wessels
- Based on the scores, each client is tagged
 - If score < 0.62, tagged as “Poor”
 - If score in [0.62, 0.86], considered “Good”
 - If score > 0.86, considered “Great”

SPR (cont)



Lessons Learned

- Data Collection
 - Consistency, e.g. clock skew, data loss, etc.
- Data Management
 - Preprocessing and formatting
 - Privacy
 - Curating, indexing, promoting the use of the data
- Data Analysis
 - Automate processing and analysis
 - Extend analysis to non-root servers

What's new

- DITL 2010 is happening **NOW!**
 - A signed root zone starts being served by B, C, F, G, and H root today.
 - All roots servers will be collecting packets!
- NZRS joins the DITL collection for the first time

In the end...

- The DITL Project

<http://www.caida.org/projects/ditl/>

- Interactive graph

<http://www.caida.org/research/dns/roottraffic/evolution/interactive-graphs/>