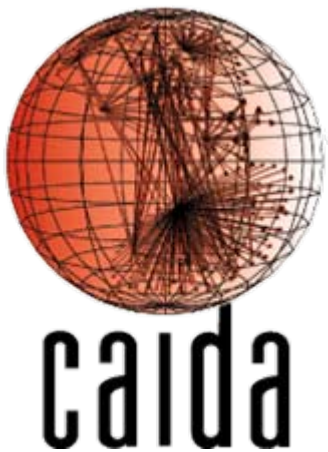
DNS: comparison of 2006 and 2007 snapshots

Sebastian Castro

secastro@caida.org

CAIDA



9th CAIDA/WIDE workshop – January 2008

Motivation

- DITL collections provides highly valuable data for researchers
- Root servers operators have actively participated on each collection
- The availability of traces from several root server instances provides the opportunity to know how is changing along the years.
- We prepared some graphs and analysis of the evolution of the DNS traffic using DITL 2006 and 2007 root servers traces.

Overview

- General statistics
  - Query rate
  - Client rate
- Stability parameters
  - Switching clients
  - Client persistence
- Query characteristics
  - Distribution of queries by query type
  - Distribution of source ports
  - Query validity
  - EDNS support
- Comparing with ORSN
  - Open Root Servers Network

General statistics

Collection	DITL 2006	DITL 2007
Time duration	47.2 hours	24 hours
Number of instances	C: 4/4 instances F: 34/37 instances K: 17/17 instances	C: 4/4 instances F: 36/40 instances K: 15/17 instances M: 6/6 instances

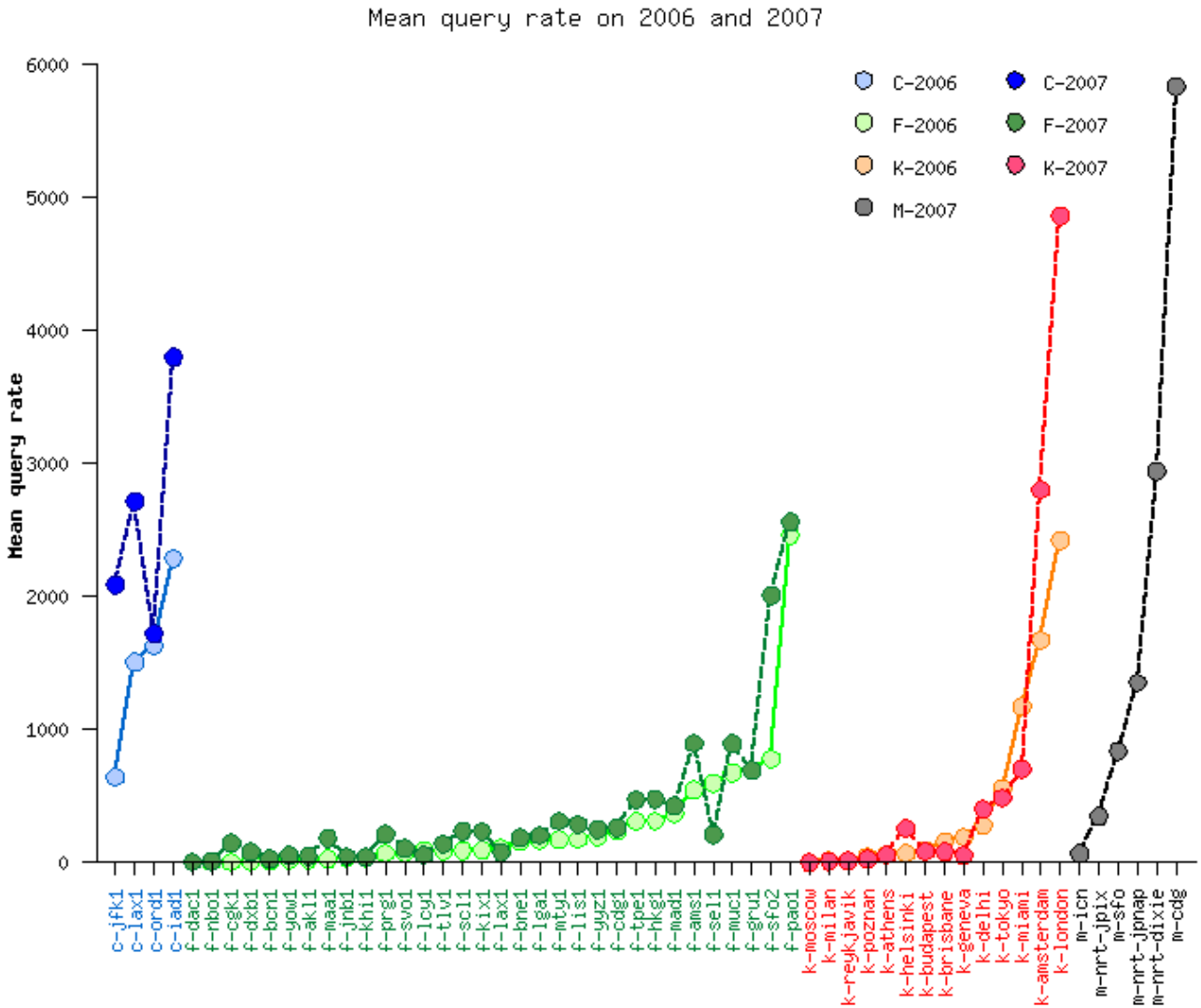
DITL 2007 collection includes additional DNS related traces coming from AS112 instances and ORSN servers.

Only the traces from AS112 were not included on the presented analysis.

General statistics

	DITL 2006 (C, F, K)	DITL 2007 (C, F, K)	DITL 2007 (C, F, K, M)
Number of queries	3.86 billion	2.8 billion	3.84 billion
Number of unique clients	~2.8 million	~2.2 million	~2.8 million
Recursive queries	4.02%	13.56%	17.04%
TCP			
Bytes	1.40%	1.24%	1.65%
Packets	2.26%	2.00%	2.67%
Queries	~221K	~500K	~700K
Queries from RFC1918 addresses	2.73%	2.83%	4.26%

Query rates



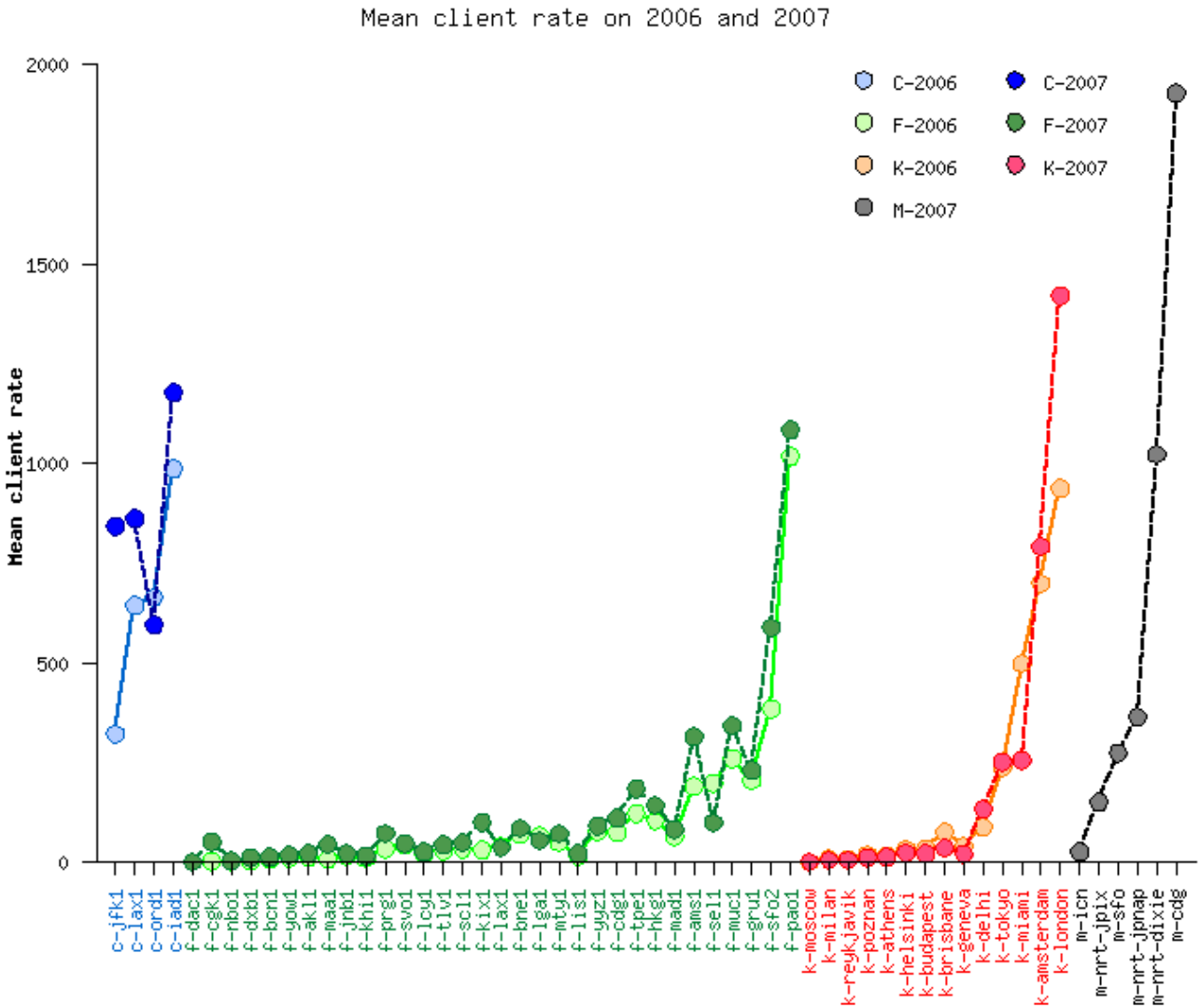
Used 50 instances of C, K and F common in DITL 2006 and 2007.

Ordered ascending by 2006 query rate.

24 instances saw an increase from 50% up to 2382% (f-cgk1).

13 saw a reduction up to 70%

Client rate

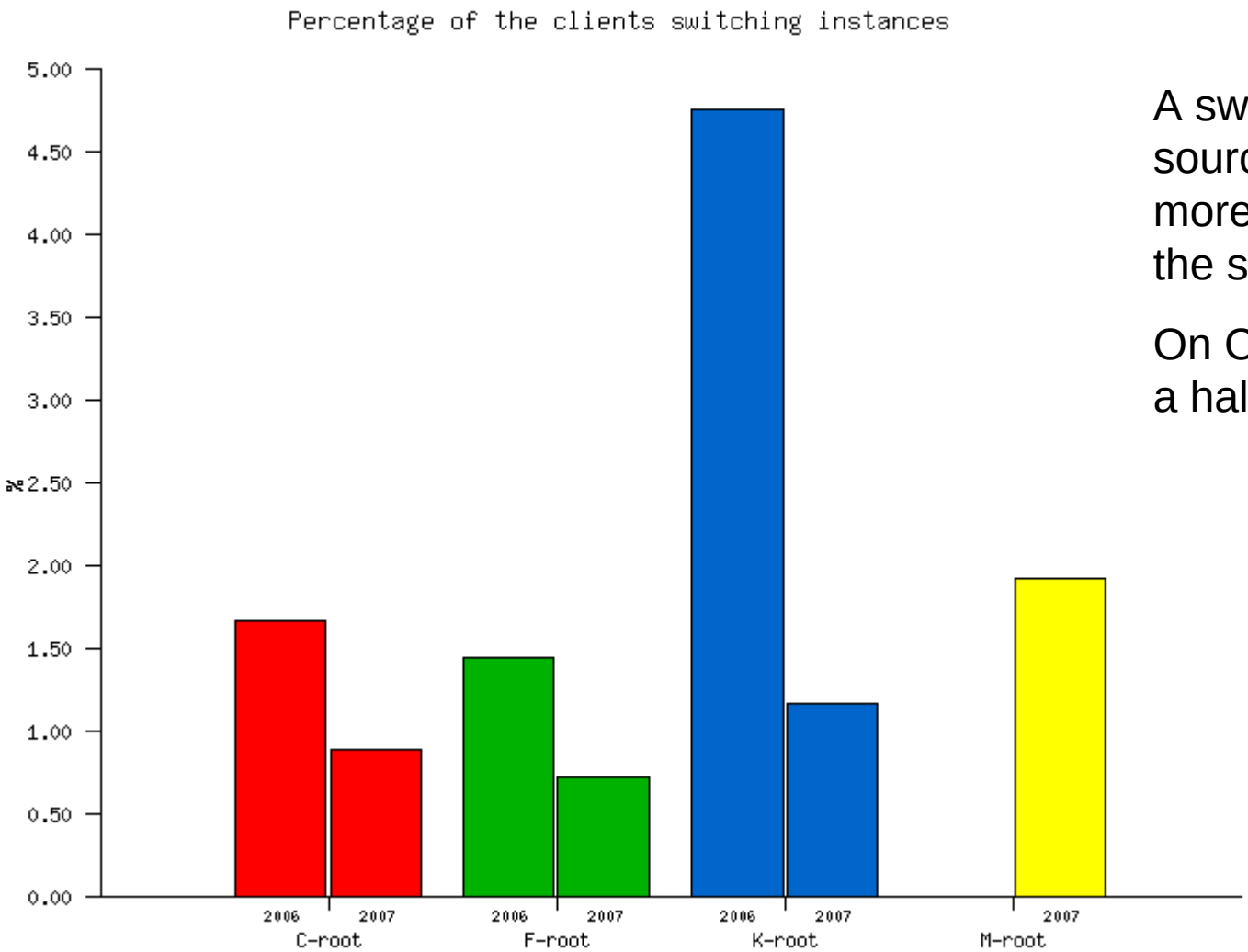


36 instances saw an increase. On 17 of them was at least of a 50%

The top increase is f-cgk1 with a 1344.6%

14 saw a reduction up to 51.3%

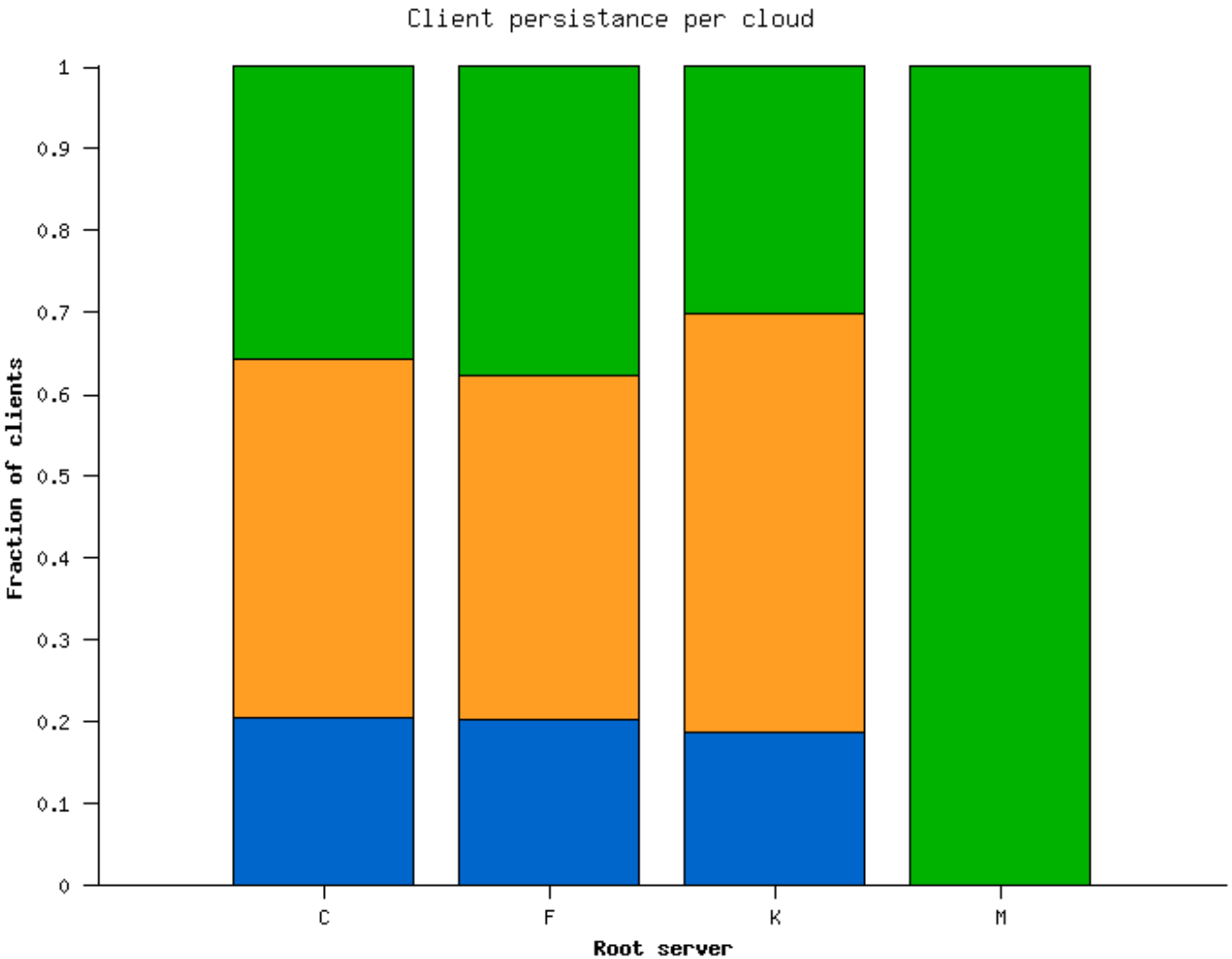
Switching clients



A switching client is any source address seen in more than one instance of the same root.

On C and F decreased to a half. K to a fourth.

Client persistence

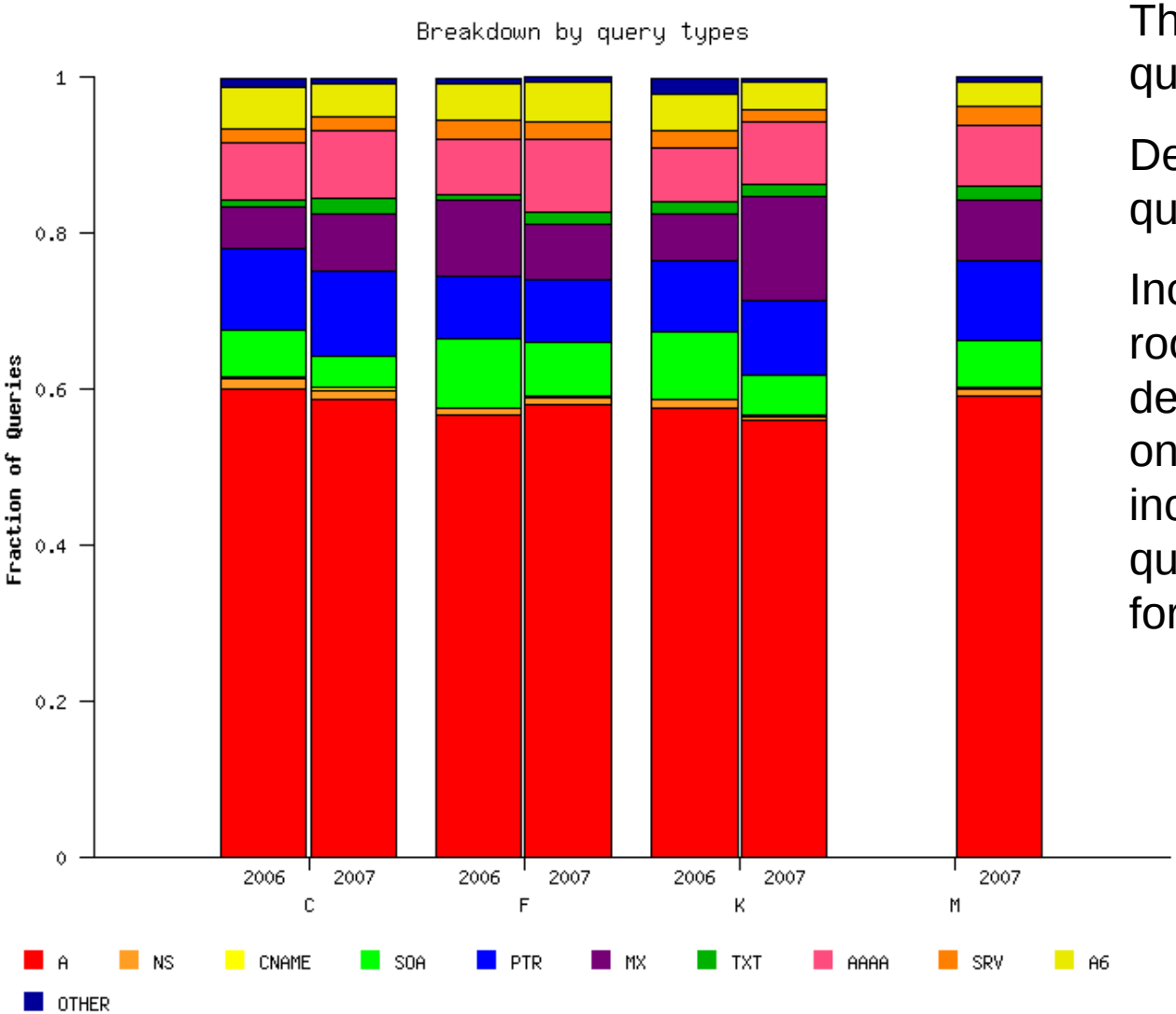


Using all source addresses present in 2006 and 2007.

Classified in three categories:

- Stable: Seen in both years.
- Only in 2006
- Only in 2007

Distribution of queries by query type

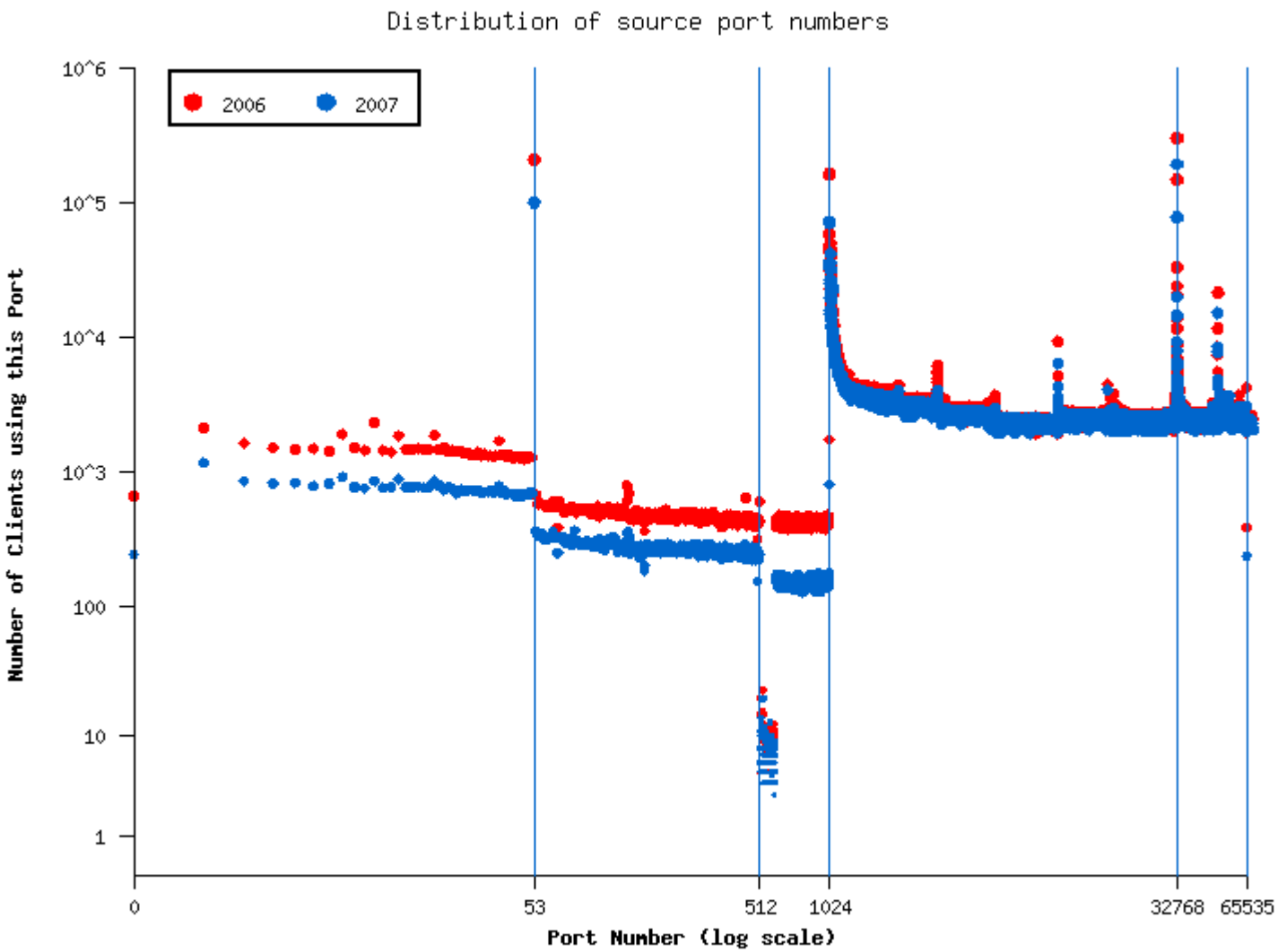


The highest fraction of queries are A queries.

Decreased fraction of SOA queries

Increase in MX queries for C-root and K-root but a decrease for the same type on F-root and a slight increase on fraction of AAAA queries on all roots available for the study.

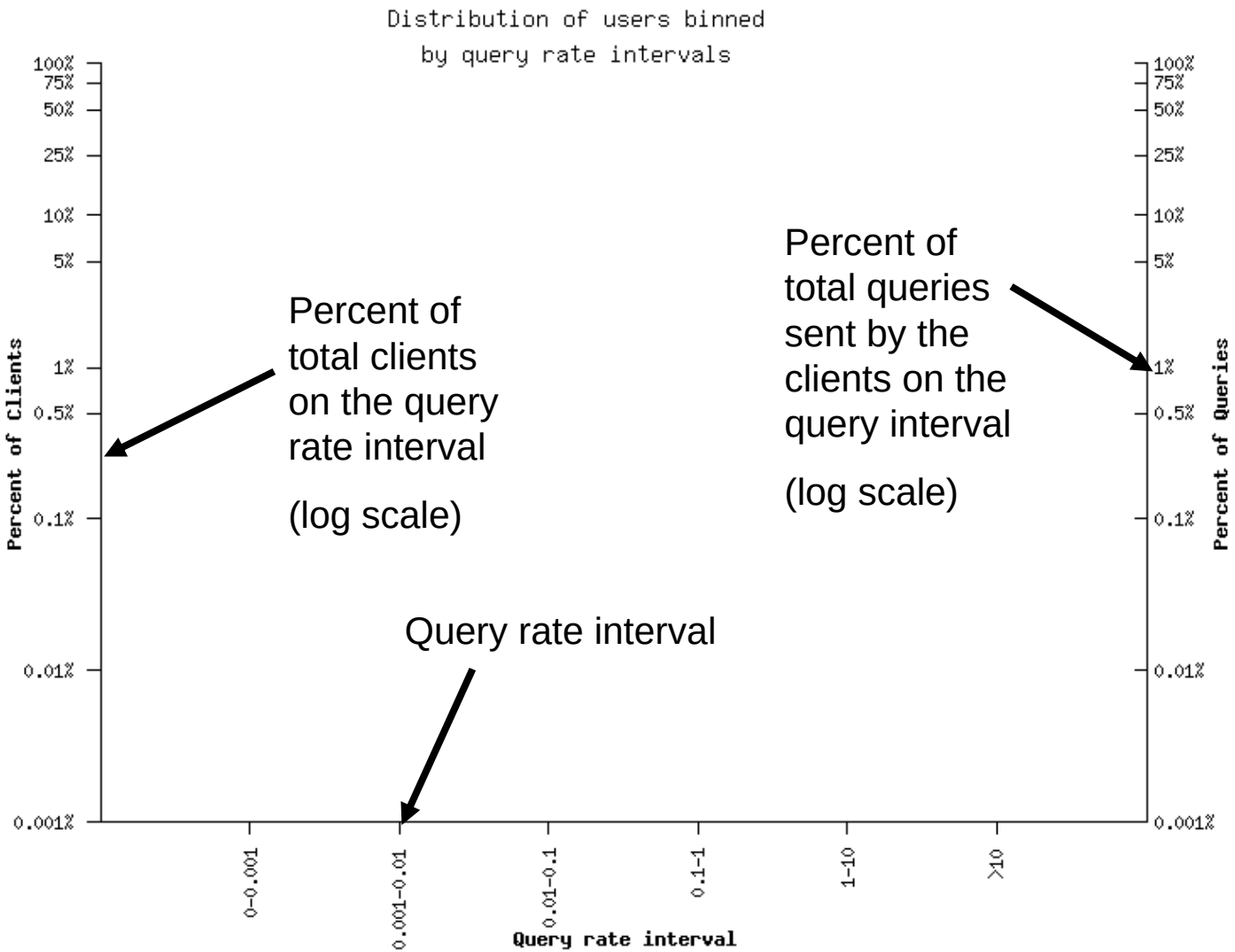
Source port distribution



Source port 0 shouldn't be used, but is allowed in case an answer is not expected.

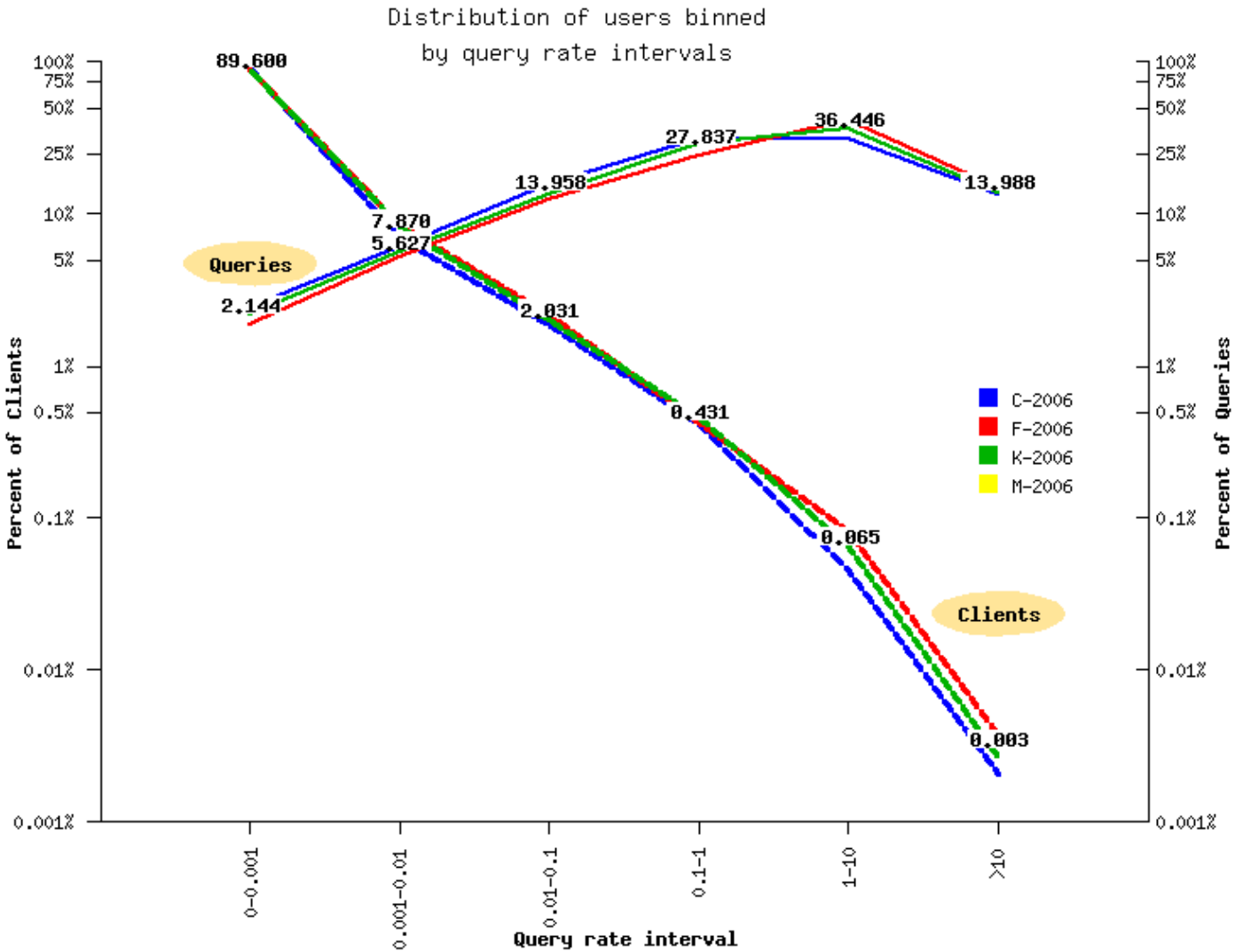
Source port 53 indicated the presence of old BIND 8 clients.

Distribution of clients/queries



Introduction to the graph

Distribution of clients/queries

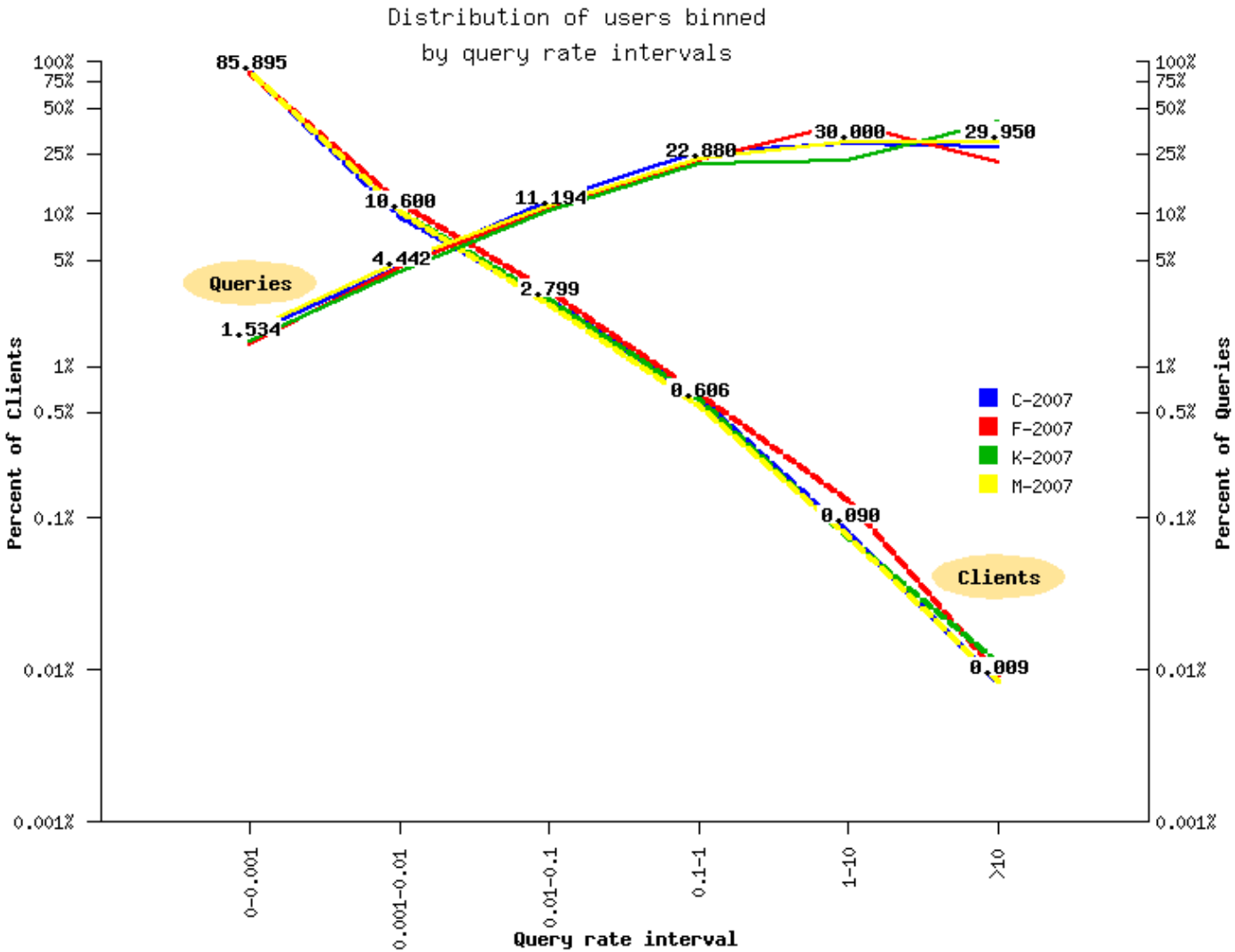


DITL 2006

Leftmost column: ~2.0% of the queries are sent by ~90% of clients.

Rightmost column: 145 clients(~0.003%) produced ~14% queries.

Distribution of clients/queries



DITL 2007

Leftmost column: ~1.5% of the queries are sent by ~85% of clients

Rightmost column: 548 (~0.009%) of clients generated ~30% of the queries.

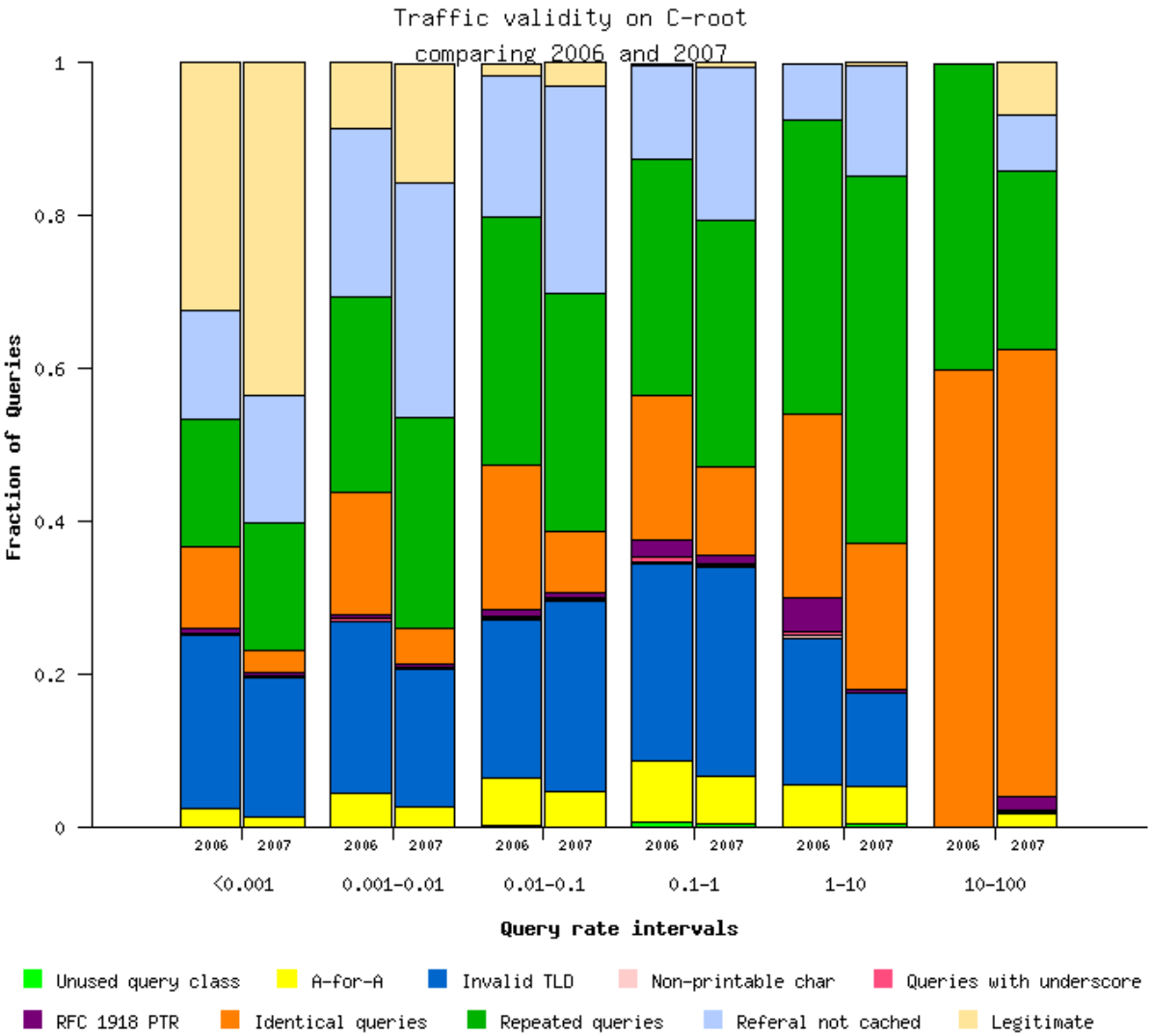
Invalid queries analysis

- To prepare the invalid queries analysis we required to split the traces per source address.
  - With more than 2 million sources, the effort would be enormous.
  - We sampled 10% of source addresses per root
- Each query could fit in nine categories of invalid queries
  - The match was done sequentially
  - If none matched, was counted as **valid query**

Invalid queries categories

- Unused query class:
  - Any class not in IN, CHAOS, HESIOD, NONE or ANY
- A-for-A: A-type query for a name is already a IPv4 Address
  - <IN, A, 192.16.3.0>
- Invalid TLD: a query for a name with an invalid TLD
  - <IN, MX, localhost.lan>
- Non-printable characters:
  - <IN, A, www.ra^B.us.>
- Queries with '_':
  - <IN, SRV, _ldap._tcp.dc._msdcs.SK0530-K32-1.>
- RFC 1918 PTR:
  - <IN, PTR, 171.144.144.10.in-addr.arpa.>
- Identical queries:
  - a query with the same class, type, name and id (during the whole period)
- Repeated queries:
  - a query with the same class, type and name
- Referral-not-cached:
  - a query seen with a referral previously given.

Query validity

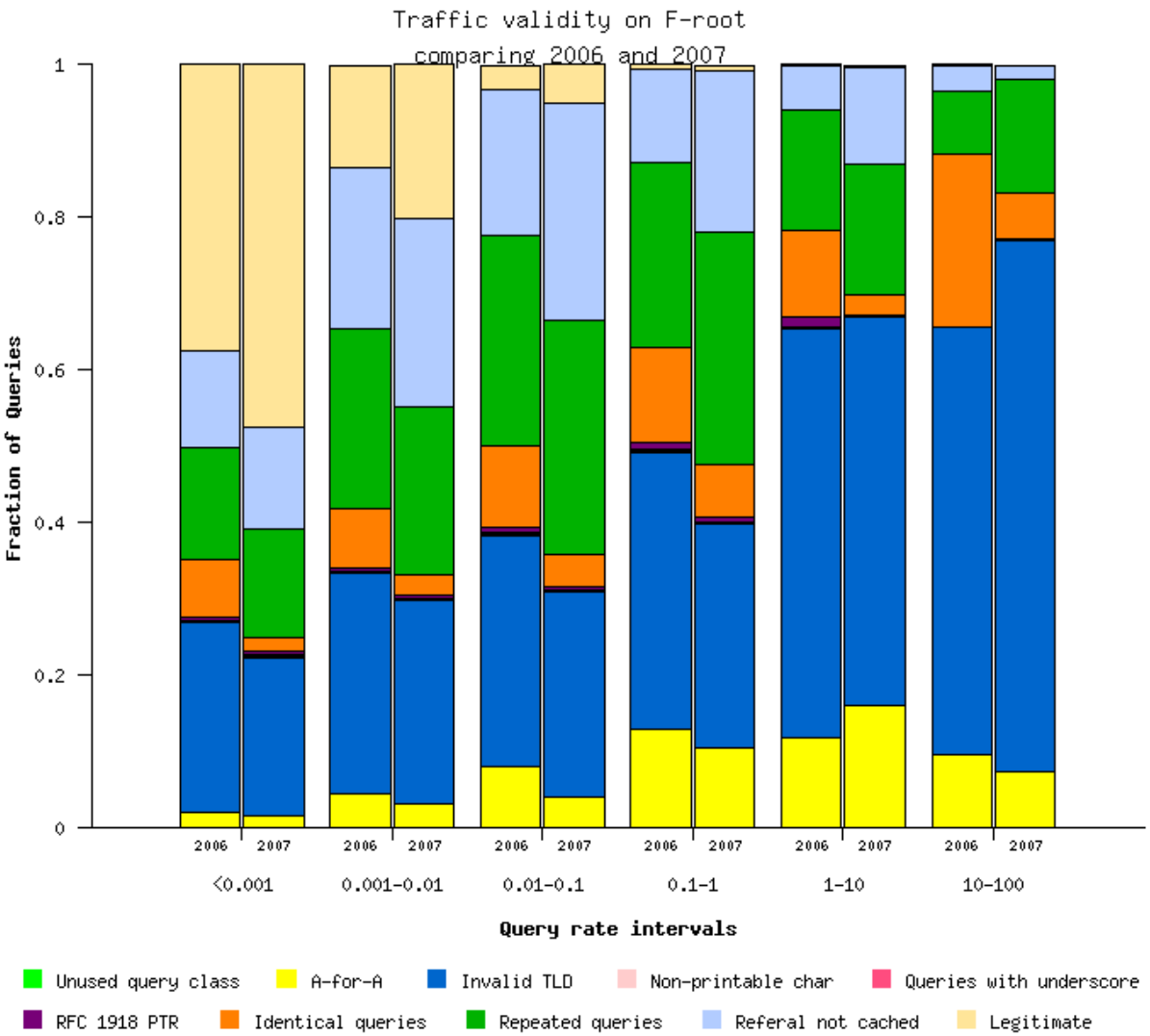


Fraction of valid/invalid queries seen on C-root

The higher the rate the lower the fraction of valid queries.

Exception on the rightmost column.

Traffic validity

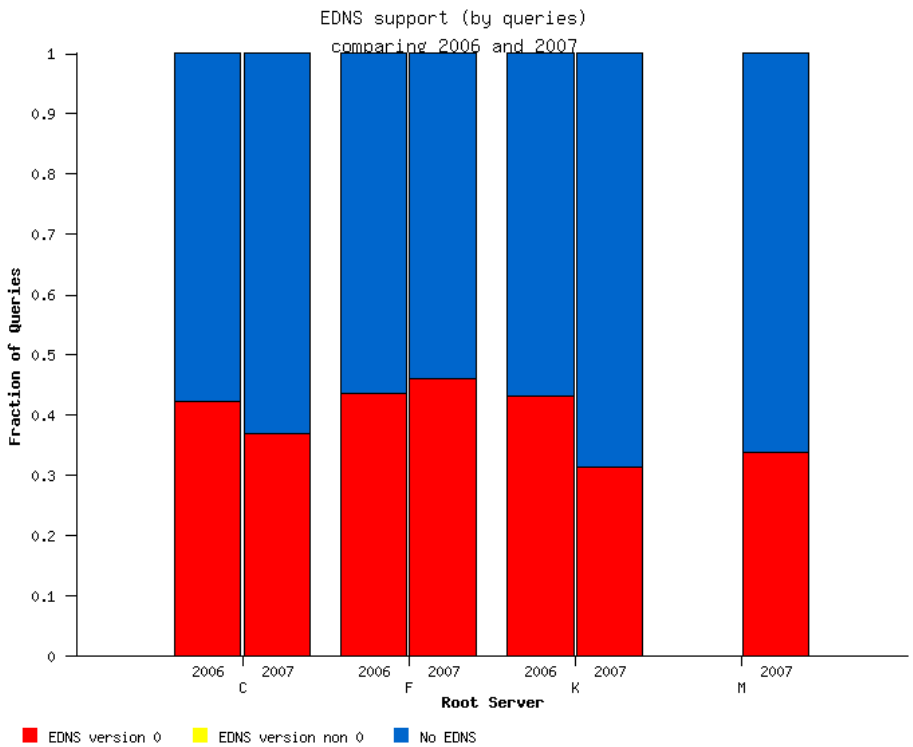


Fraction of valid/invalid queries seen on F-root

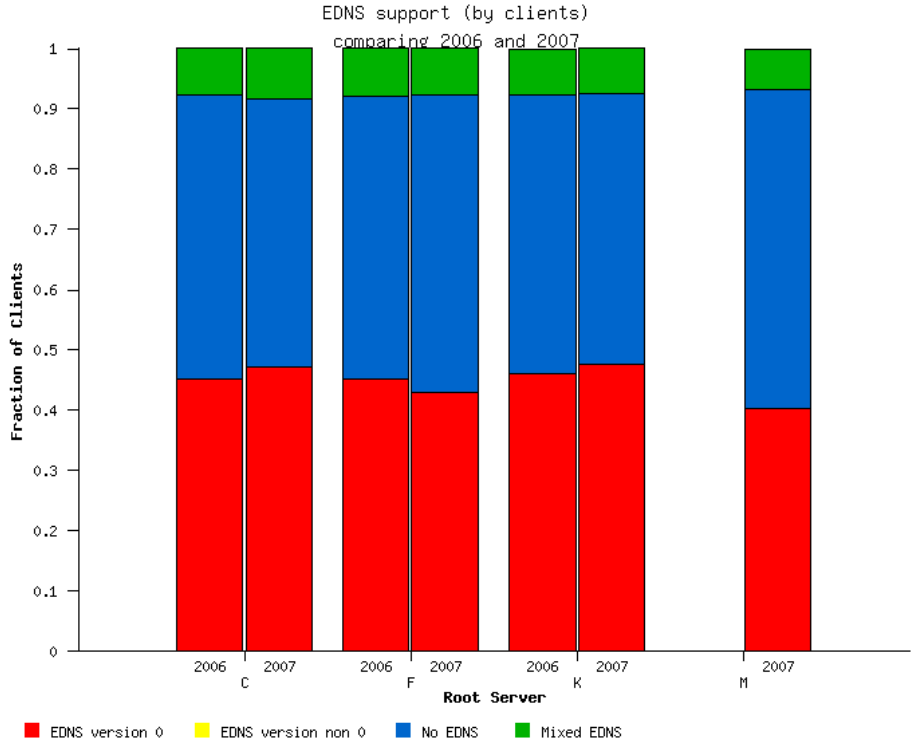
The same pattern for valid queries seen on C-root. K and M follow similar patterns.

Surprising proportion of queries for invalid TLD.

EDNS support



EDNS support by queries



EDNS support by clients.

Green represents clients with mixed EDNS support.

Open Root Server Network (ORSN)

- Created in Feb 2002 as an alternative for the ICANN-managed root servers.
- Europe centric (3 in Germany, 2 in Switzerland, one each in Austria, Slovenia, Denmark, Portugal, Greece, Netherlands, USA)
- Supports IPv6
- B (Vienna) and M (Frankfurt) contributed with traces on 2007.

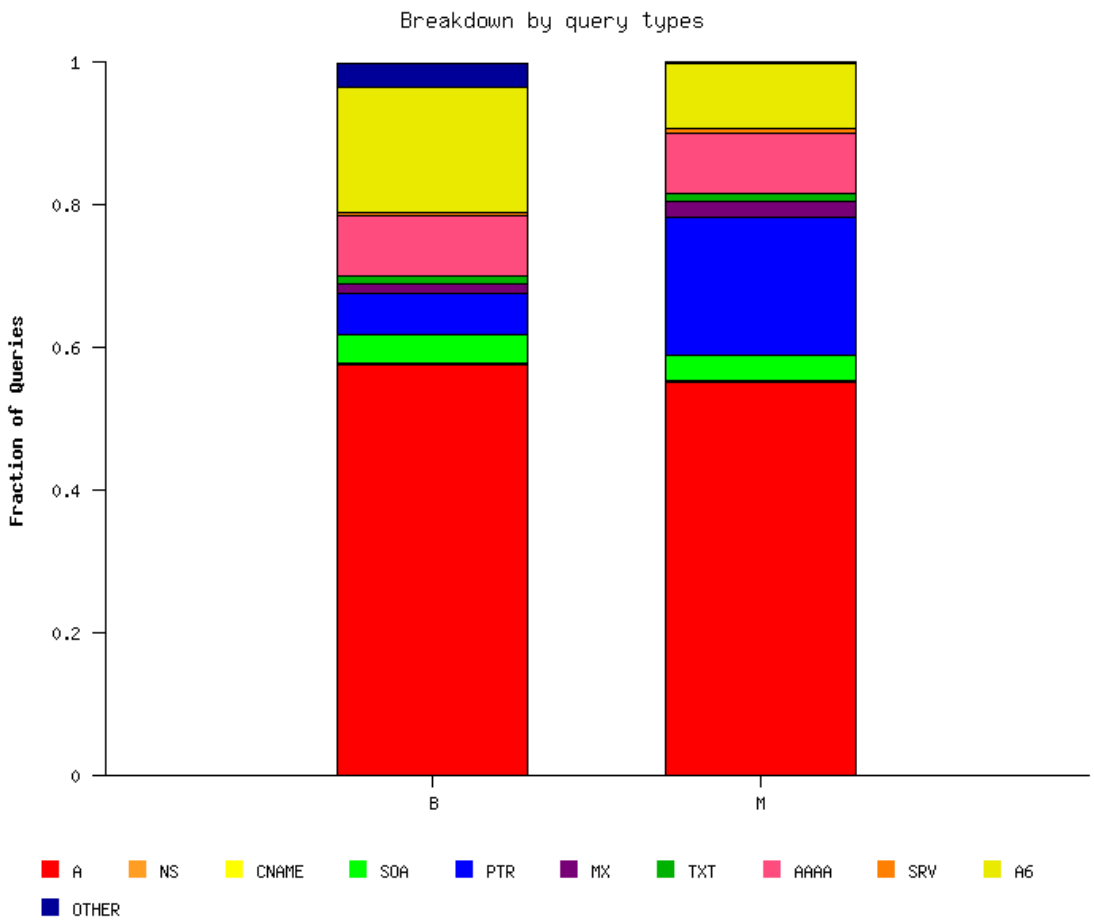
Number of queries	4.1 million
Number of unique clients	1 650
Recursive queries	11.59%
TCP	
Bytes	0.17%
Packet	0.22%
Queries	0.0118%
Queries from RFC1918 addresses	0.3%

General Stats

- Query rates
  - B-vienna: 3.3 queries per second, server side
  - M-frankfurt: 2.6 queries per second, server side
    - Comparable to the least busy root instances.
- Client rate
  - B-vienna: 2.28 clients per second
  - M-frankfurt: 2.53 clients per second
    - Similar to the client rates in f-ccs1 (2.1) or k-moscow (2.34). Higher than the lowest value found in f-dac1 (1.92).

ORSN

- Distribution by query type



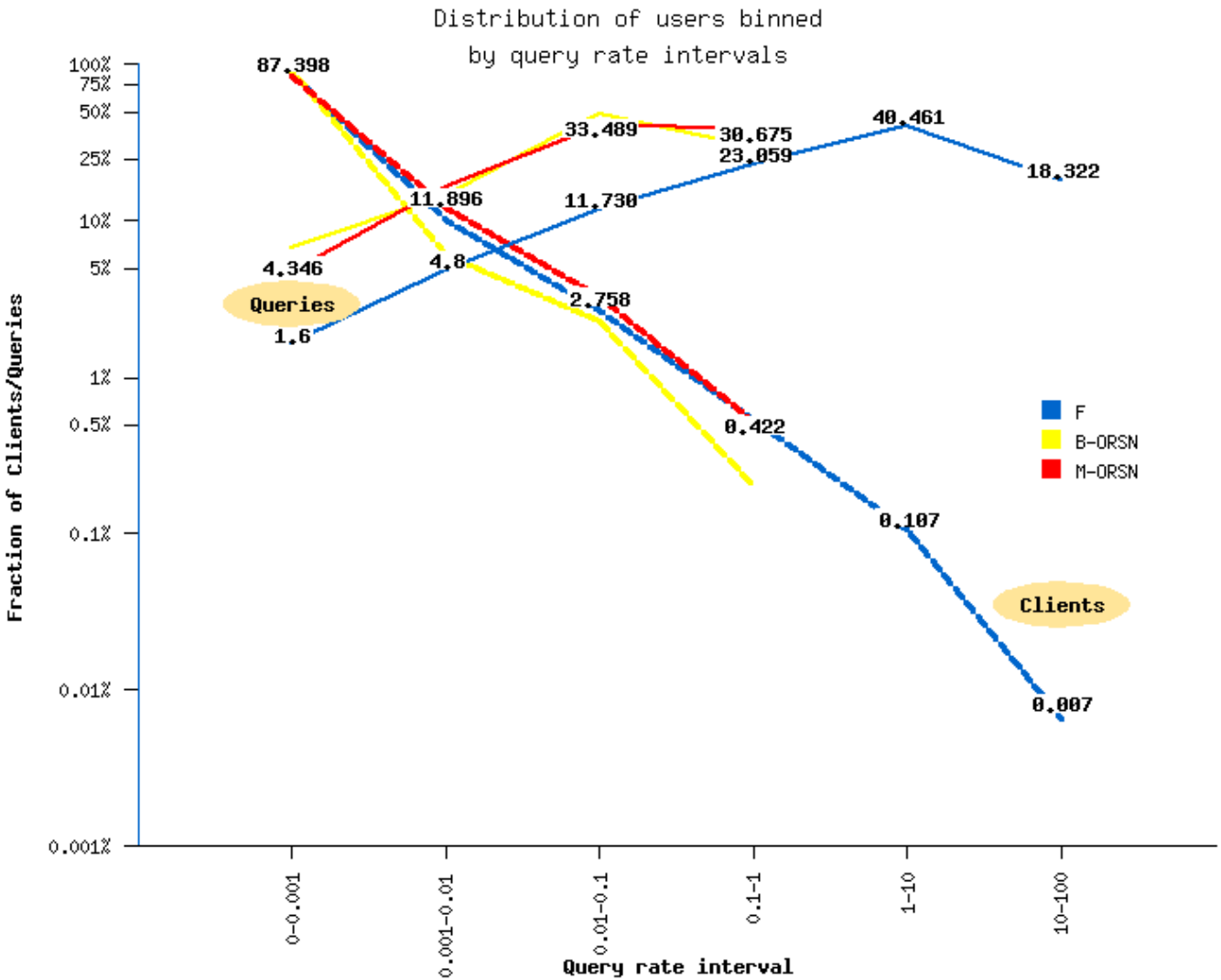
The fraction of A queries is slightly lower than the official roots: around 55%.

The A6 type queries have a more relevant presence: 18% in B and 9% in M.

Compared with 6% on roots

The fraction of AAAA queries is slightly higher: 8.5% against 7%.

Distribution of clients/queries

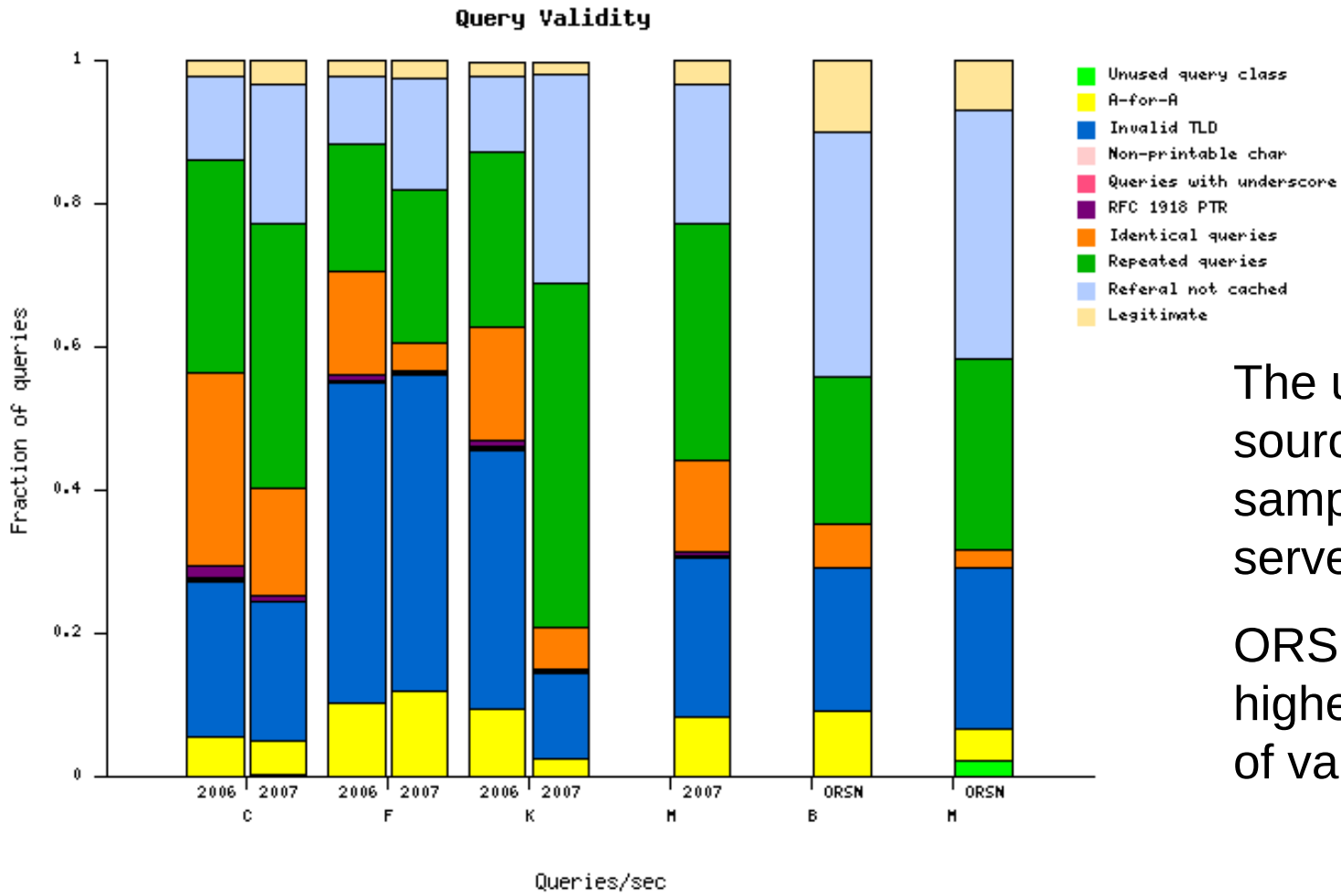


ORSN vs roots

The proportion of fraction clients/queries is similar.

ORSN has a difference of orders of magnitude in number of clients, queries and query rate.

ORSN



The used the all sources (not sampled as in root servers)

ORSN receives a higher proportion of valid queries.

Conclusions

- The query rate and client rate increased in some instances between 1.5-3 times
  - But very few instances had increments on both.
- The amount of invalid traffic hitting the roots in still high
  - Some sources could be mitigated by the approval and adoption of new RFC (local zones)
- ORSN servers are subject to similar anomalies seen on the official roots
  - Moderated by the reduced client space served.