In the search of heavy hitters

Sebastian Castro

secastro@caida.org

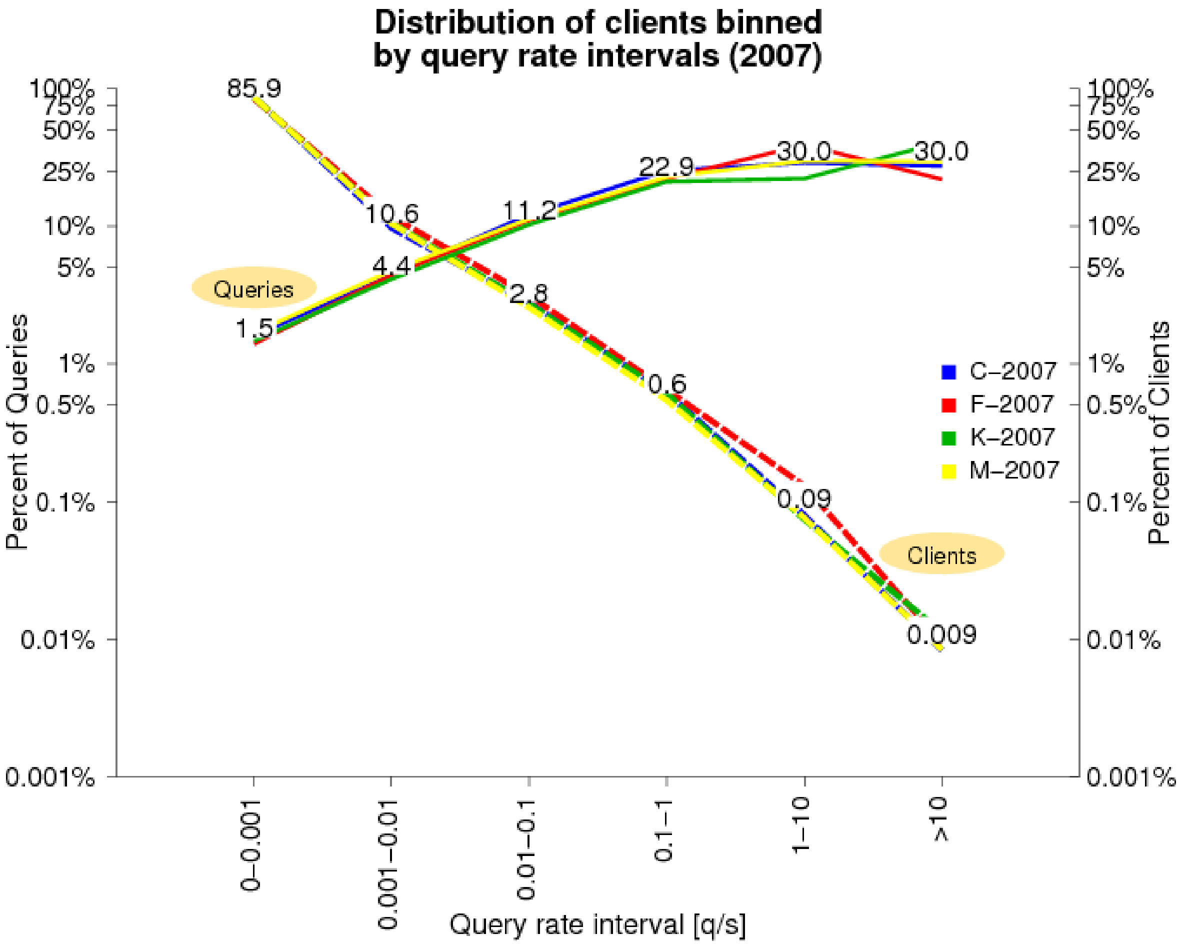
CAIDA / NIC Chile

DNS-Ops Workshop. June 4-5, 2008. Brooklyn, NY.

Motivation

- Using root servers traces (C, F, K, M) in DITL 2007 we found

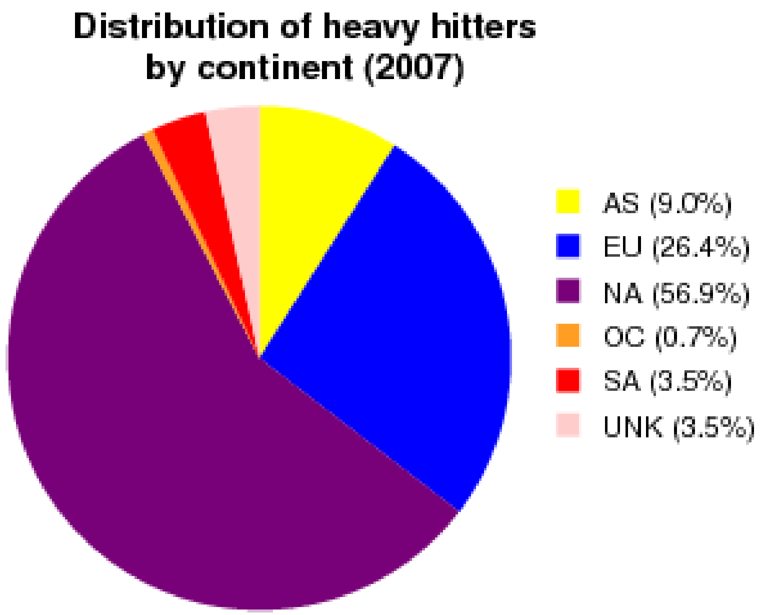
  - 510 unique source addresses generated 30% of the traffic
  - 12 of them sent more than 100 queries per second!
  - We wanted to find out as much as possible about them.

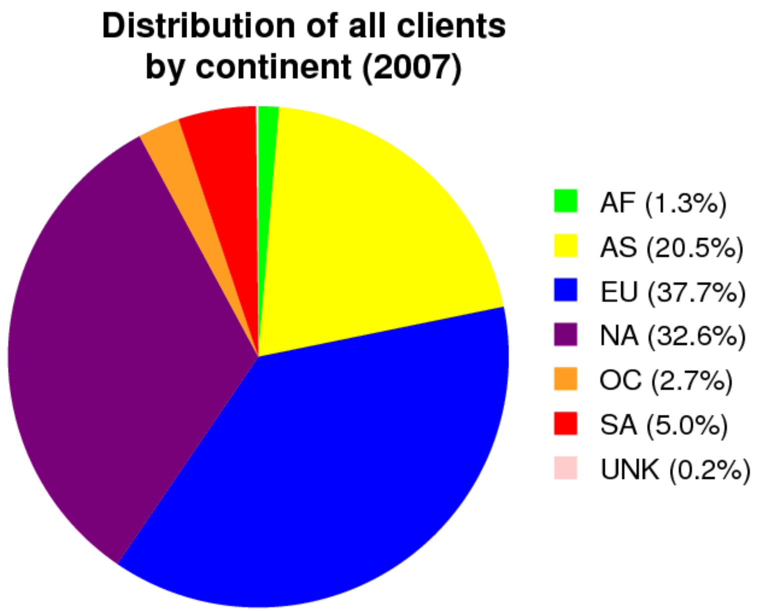


Heavy Hitters

- Initially we named “heavy hitter” to a unique IPv4 address sending more than 10 qps.
  - Having the DITL 2008 traces (with doubled the roots), the definition had to change.
- A “heavy hitter” is a unique IPv4 address sending more than 10 q/s per root.
  - 144 addresses matched this condition in 2007.
- A “super heavy hitter” is a subset sending more than 40 q/s per root.
  - 11 addresses are counted as super heavy hitter.

Geography

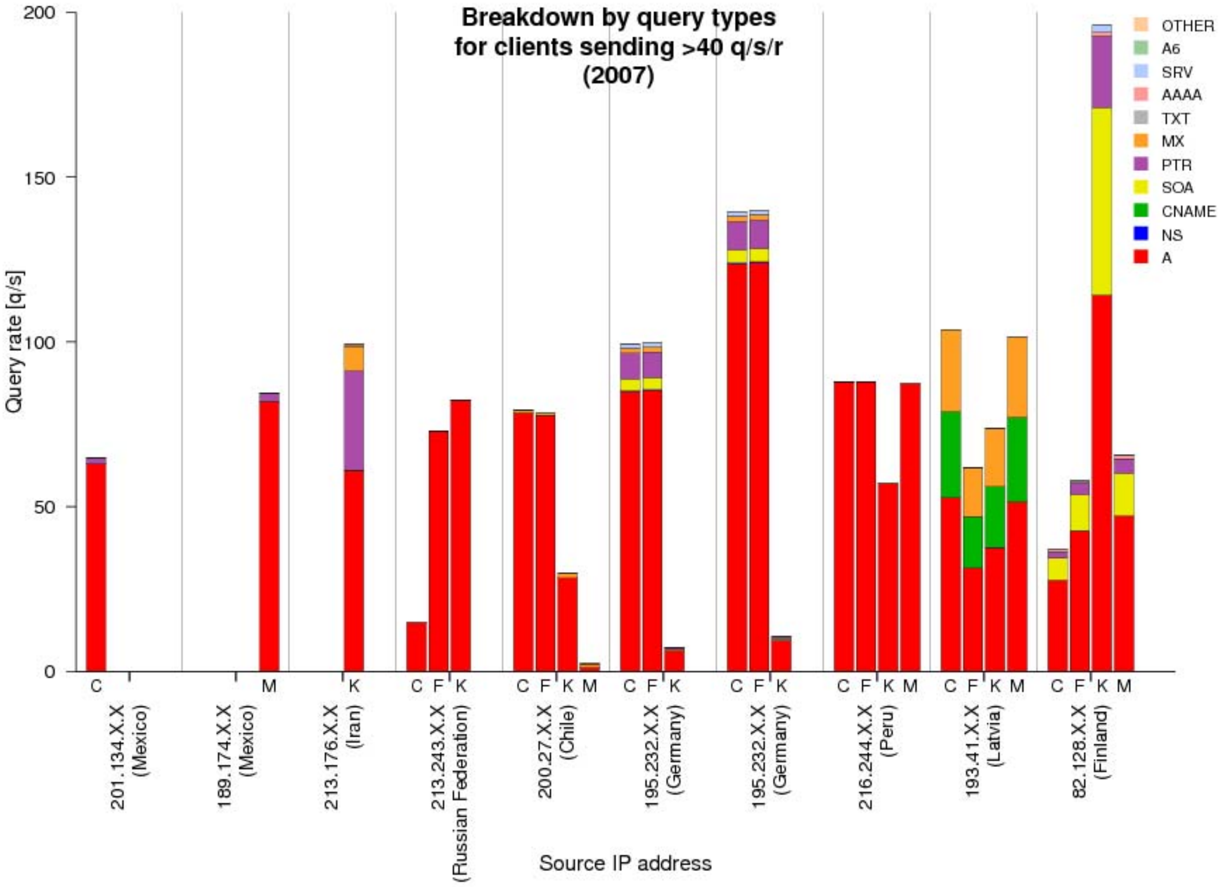




- The heavy hitters don’t have the same geographic distribution of the total clients.
  - It’s highly concentrated in the US

Super Heavy Hitters:
Detailed behavior by query type

- Selected the Top 10.
  - They generated 5% of the total query load.
- Ordered from left to right by query rate
- Distribution among roots is not balanced.
- The ninth client sent A, CNAME and MX queries.

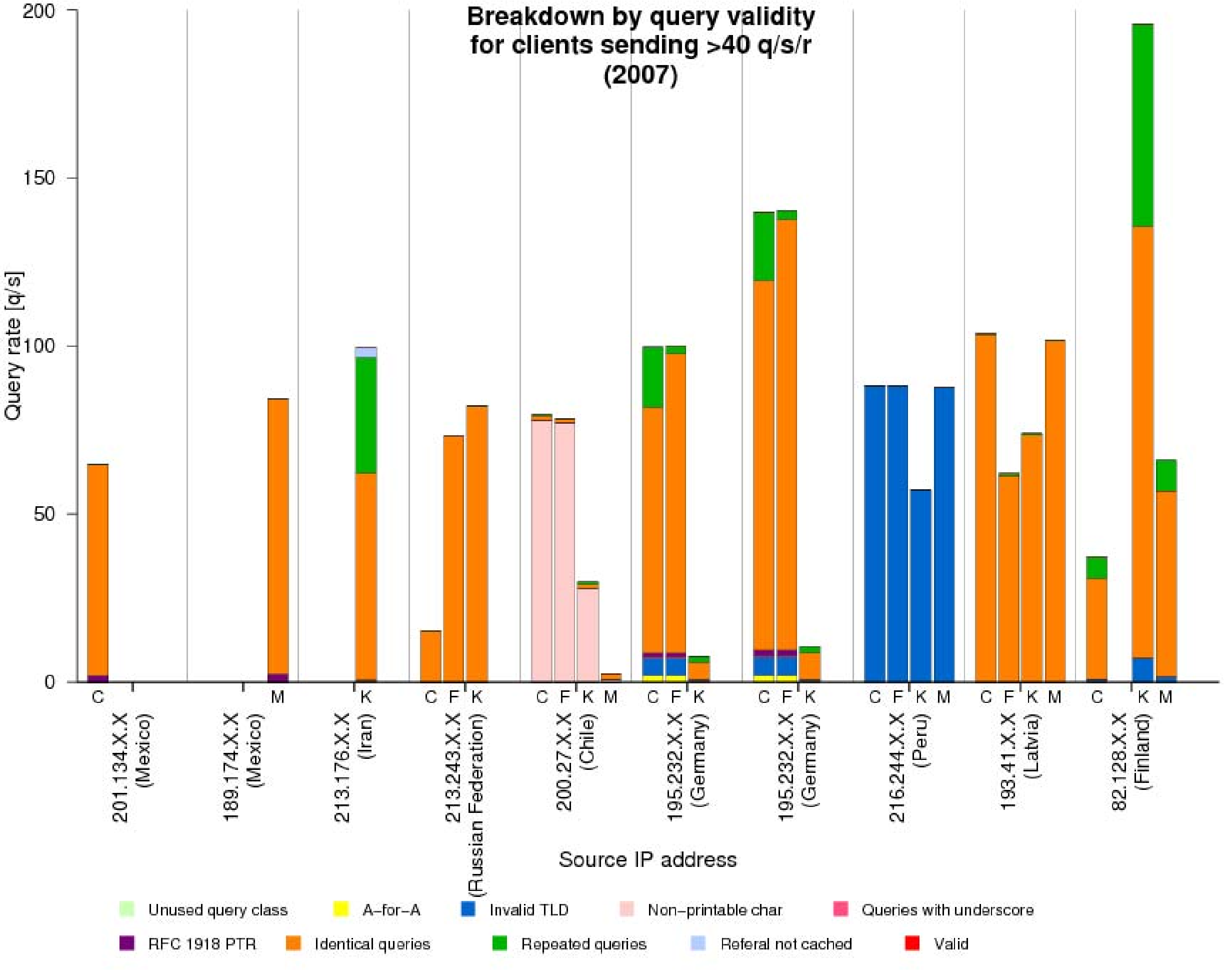


Query validity reminder

- Nine categories of invalid queries, evaluated sequentially
  - **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
  - **Non-printable characters**: a query for a name with characters not in [A-Z0-9\-] list
  - **Queries with '_'**: Special category for the invalid but widely used character.
  - **RFC 1918 PTR**: a PTR query for an IPv4 address in the private space
  - **Identical queries**: a query with the same class, type, name and id (during the 24 hours period)
  - **Repeated queries**: a query with the same class, type and name
  - **Referral-not-cached**: a query seen with a referral previously given.
    - If a client sent <IN, A, www.example.net> and later <IN, NS, ripe.net> the second query counts as "referral-not-cached" because a referral to "net" nameservers was answered.
    - A tolerance parameter of 2 seconds was included on this analysis
    - Root servers are authoritative for .arpa, .in-addr.arpa and root-servers.net zones, were included as special cases.
- No match means 'valid query'.

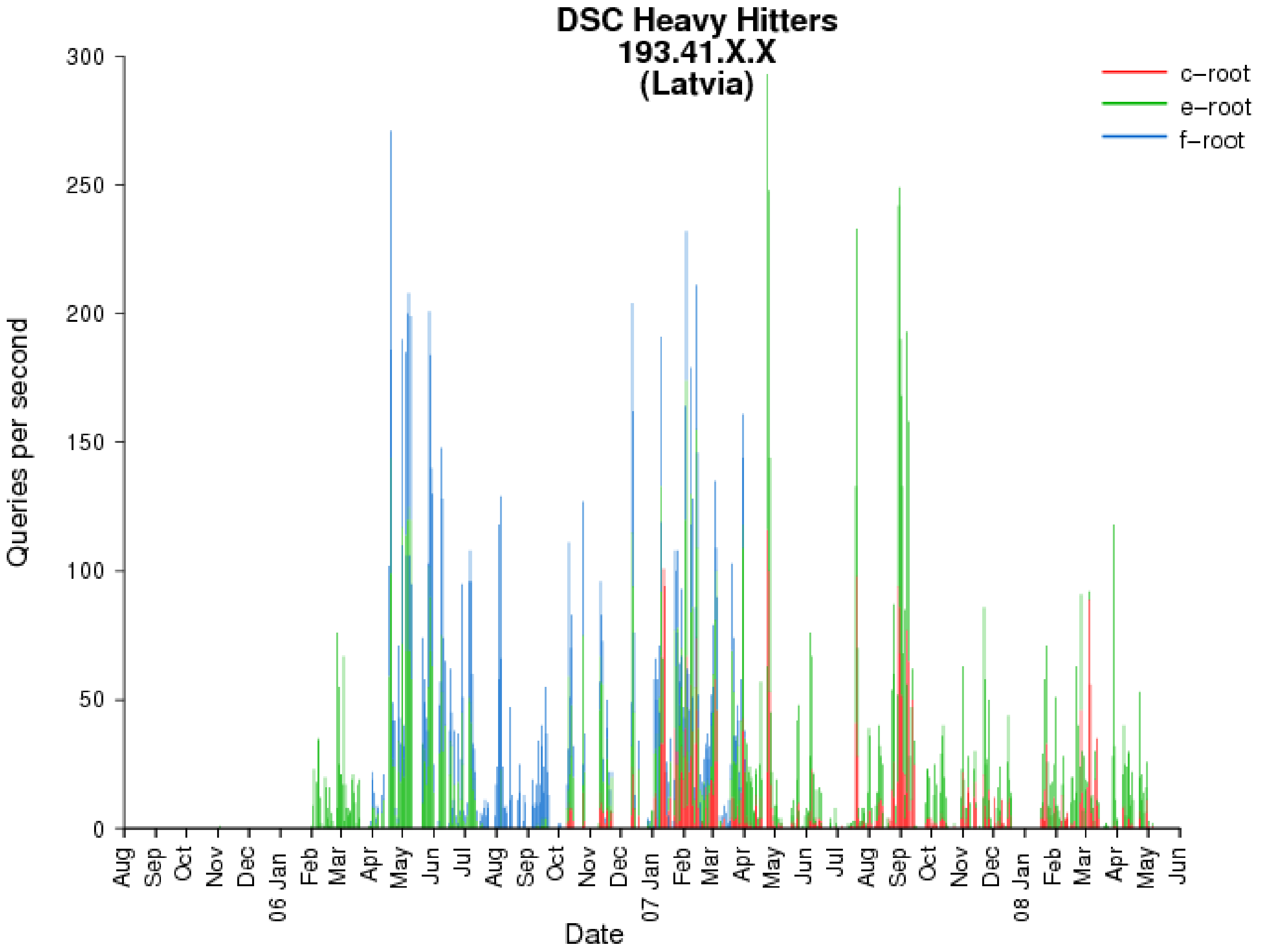
Super Heavy Hitters: Detailed behavior by query validity

- The same top 10
  - Generating 5% of the total query load…
  - … with 0.001% of their queries considered valid
- The fifth client has more than 96% of its queries asking for a hostname with two spaces on the name!
- The eighth client sent 98% of their queries for the 'localhost' TLD.



As seen by DSC

- 193.41.X.X, Top 2 in 2007



2007 Heavy Hitters in 2008

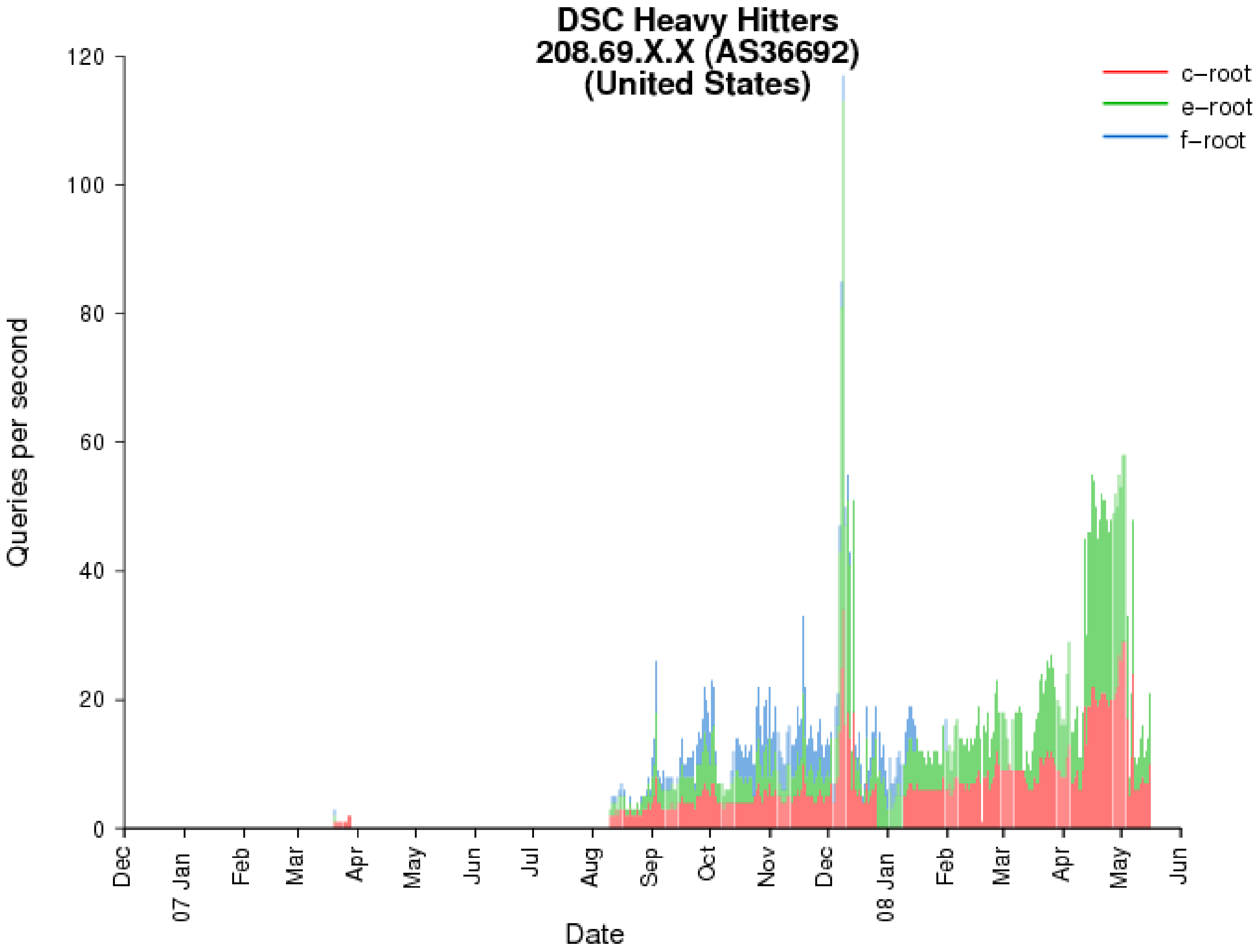
- Using the traces for the roots in DITL 2008, we observed
  - 112 (77.78%) were not present!
  - 29 (20.14%) decreased their query rate.
  - 3 (2.08%) increased their query rate.
- Let’s see the variation at the AS level

Grouping by AS

AS	AS Name	AS Country	Ranking			Percentage of total queries			Normalized query rate (queries/sec/root/client)		
			2007	2008	Var.	2007	2008	Var.	2007	2008	Var.
27595	INTERCAGE	US	1	1	-	7.60	3.52	-4.08	2.357	0.976	-1.381
9121	TTnet	TR	25	2	+23	0.54	2.54	+2.00	0.003	0.001	-0.002
3356	Level 3	US	24	3	+21	0.56	2.50	+1.94	0.011	0.015	+0.004
36445	Cernel	US	-	4	-	0	2.27	+2.27	0	7.971	+7.971
7132	AT&T Internet Services	US	5	5	-	1.61	2.09	+0.48	0.003	0.004	+0.001
4134	Chinanet	CN	3	6	-3	2.73	1.63	-1.10	0.009	0.004	-0.005
3320	Deutsche Telekom	DE	4	7	-3	1.88	1.54	-0.34	0.001	0.001	-
3215	France Telecom	FR	7	8	-1	1.41	1.53	+0.12	0.004	0.002	-0.002
36692	OpenDNS	US	176	9	+167	0.09	1.41	+1.32	0.378	4.599	+4.221
3352	Telefonica Data España	ES	6	10	-4	1.55	1.34	-0.21	0.008	0.003	-0.005
	TOTAL					17.97	20.37				

DSC shows

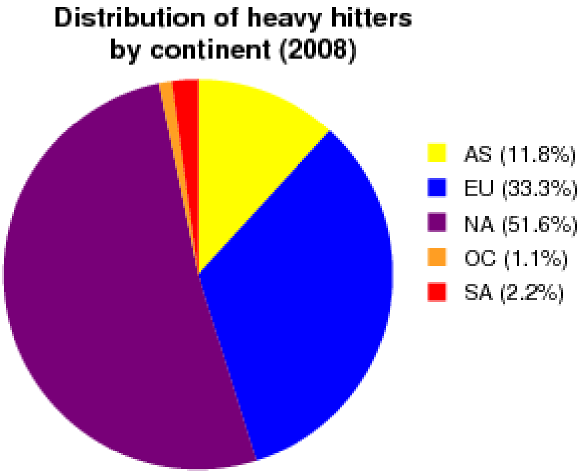
- One of the addresses from OpenDNS

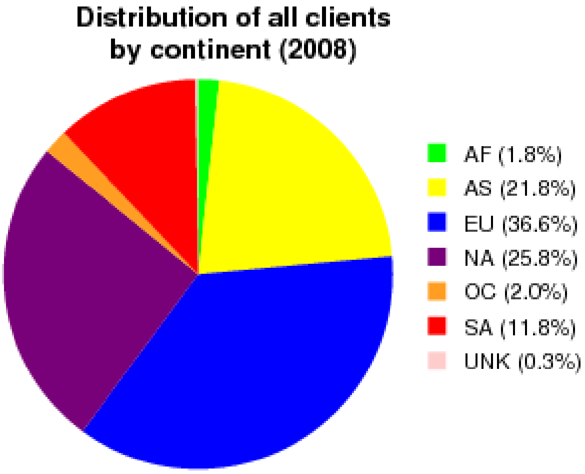


Heavy hitters in DITL 2008

	Heavy Hitters	Super Heavy Hitters
DITL 2007	144	11
DITL 2008	93	10

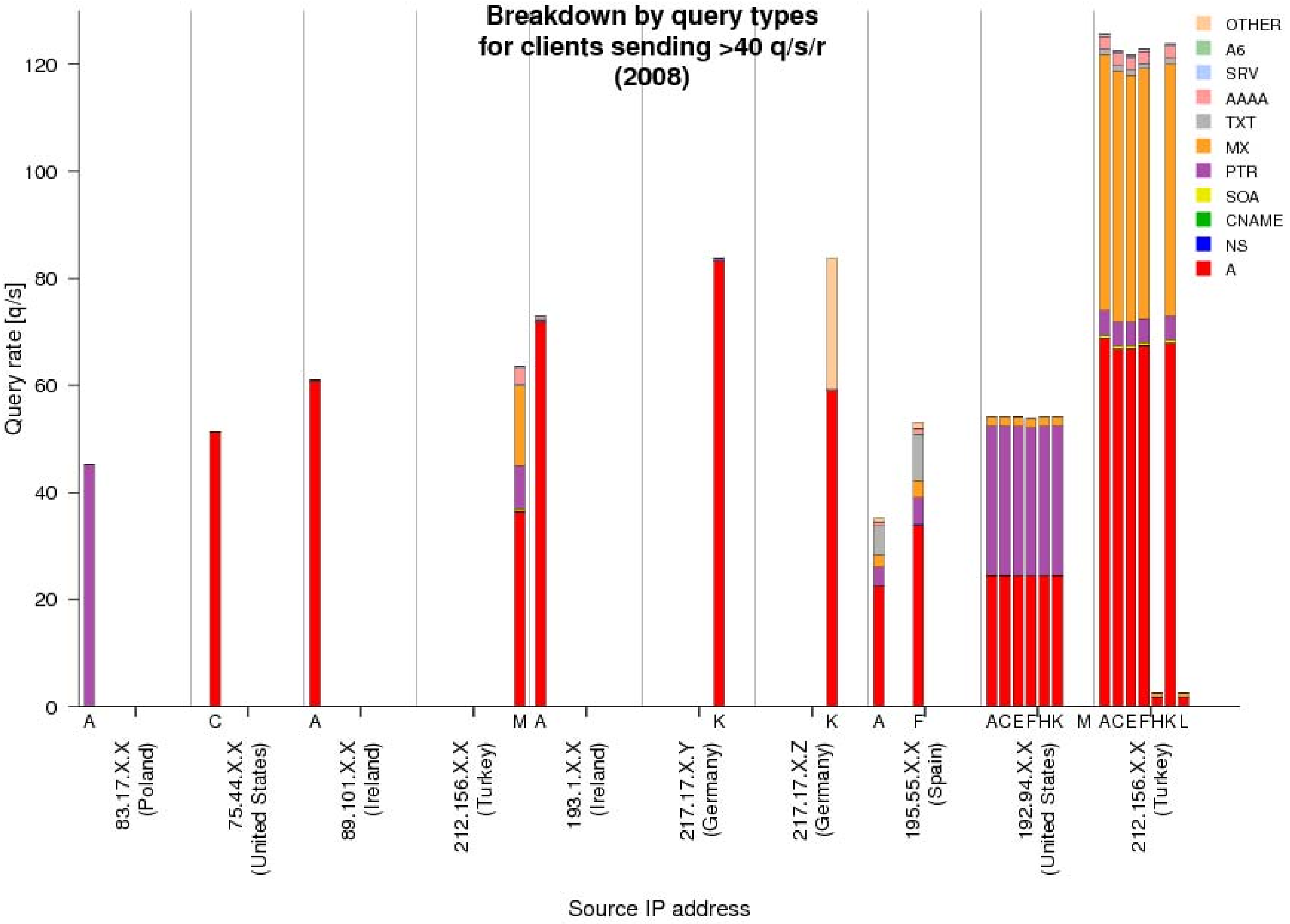
- We have less heavy hitters.
- The distribution by continent still highly concentrated in N. America





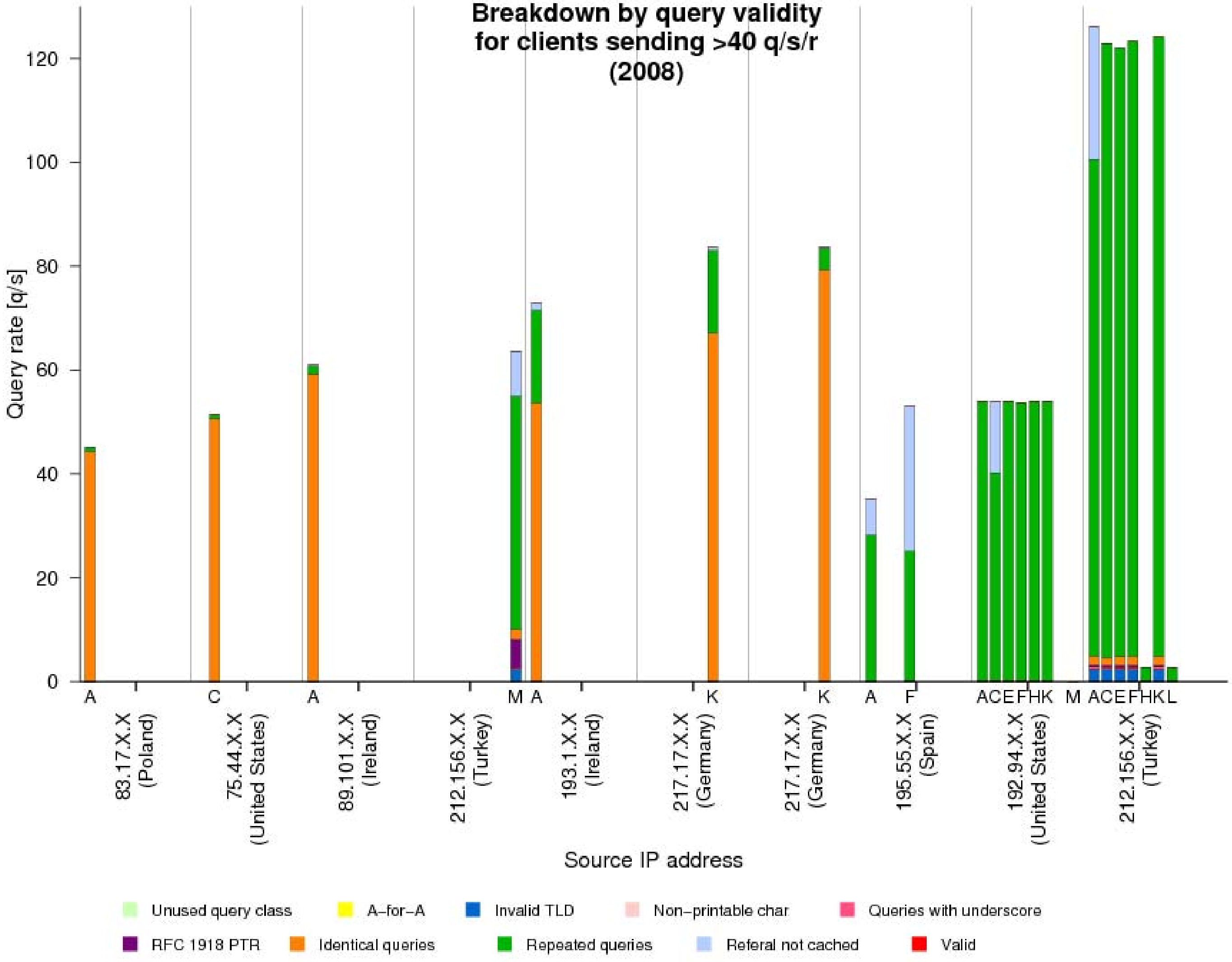
Super Heavy Hitters: Detailed behavior by query type

- Again the Top 10
  - In this case generated 1.738% of the total query load
  - And 0.004% of their queries are counted as valid.
    - Is it improving?

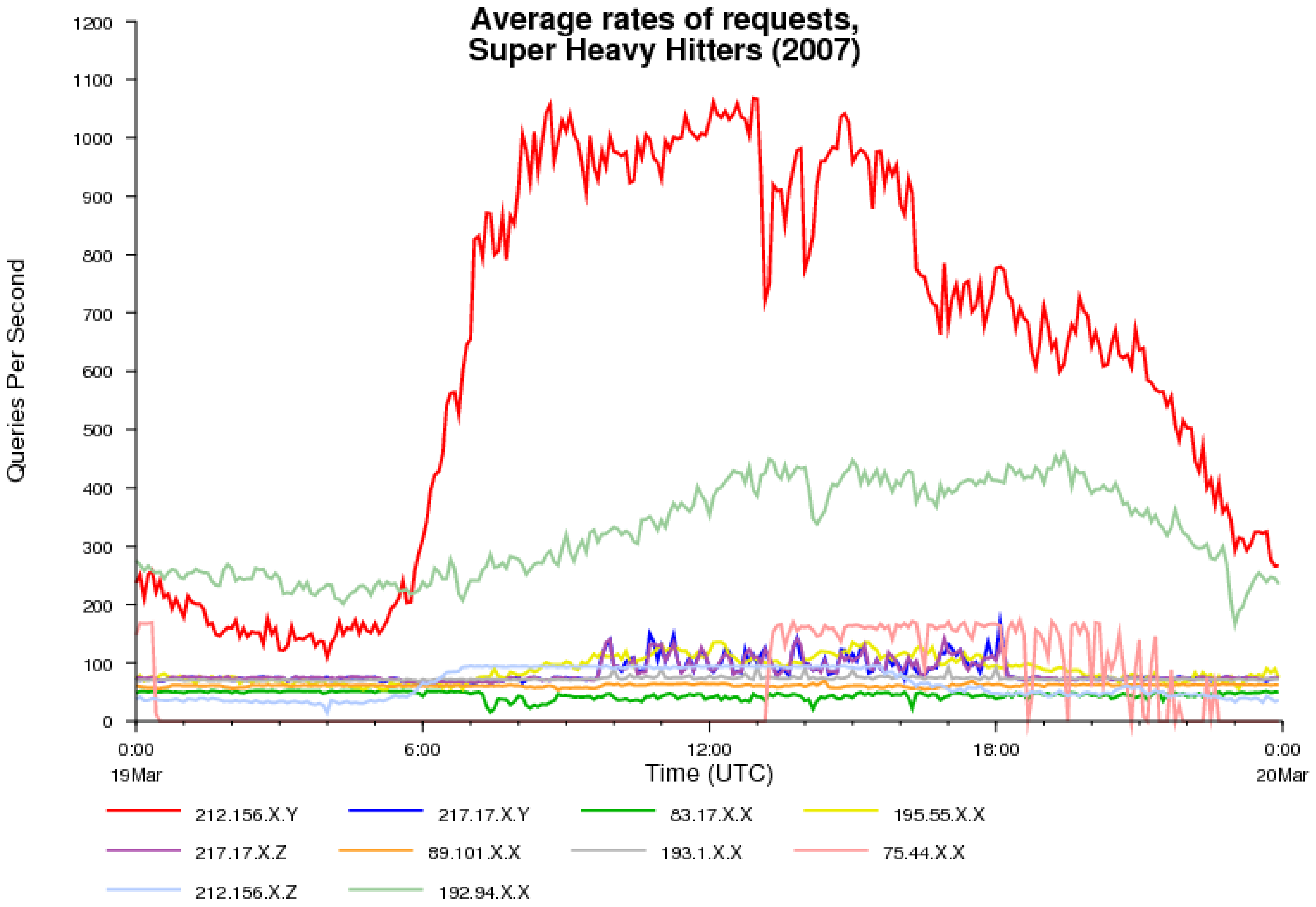


Super Heavy Hitters: Detailed behavior by query validity

- Comparing with 2007, most of the traffic is identical/repeated queries.
- No ‘special’ cases could be observed

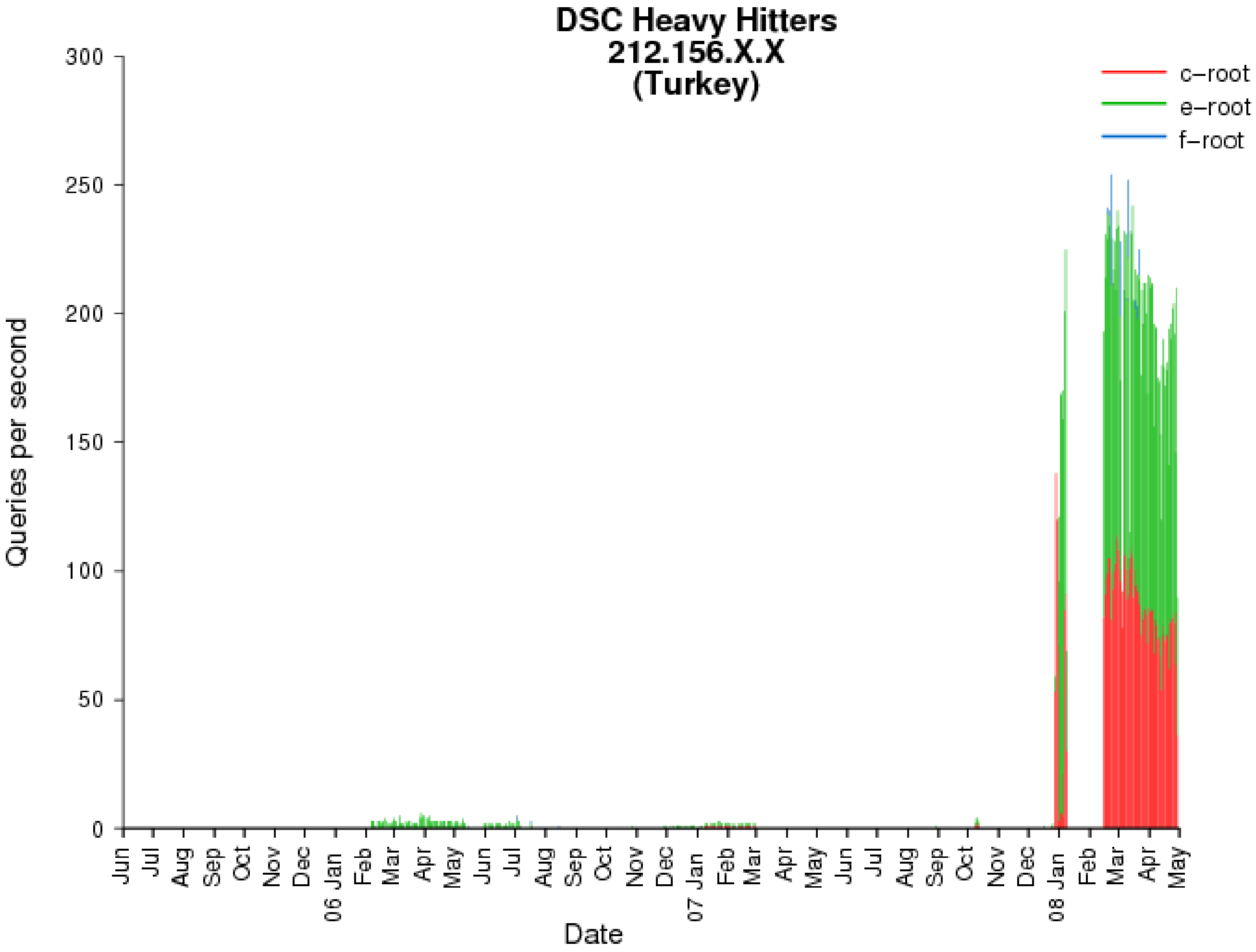


Query rates…


Average rates of requests,
Super Heavy Hitters (2007)
Queries Per Second
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
0:00
19Mar
6:00
12:00
18:00
0:00
20Mar
Time (UTC)
212.156.X.Y
217.17.X.Y
83.17.X.X
195.55.X.X
217.17.X.Z
89.101.X.X
193.1.X.X
75.44.X.X
212.156.X.Z
192.94.X.X

DSC shows

- Top 1: 212.156.X.X

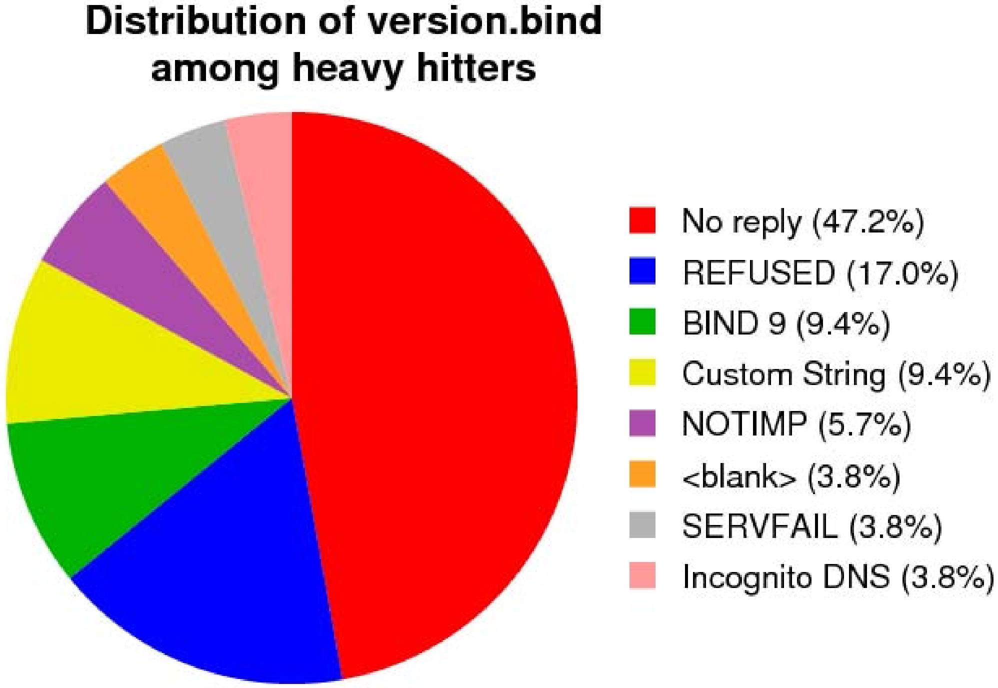


Fingerprinting

- Using the list of heavy hitters in 2008 and Duane’s DNS survey, we found:
  - 43% of the addresses didn’t have any information.

- For the ones with information, the distribution is:

- Unfortunately fpdns was unable to provide any further detail



Conclusions

- The sources of high traffic change with time
  - And we don’t have much clue about who’s behind them
- Active probing closer to the collection date would be helpful
- The use of smarter ways to analyze the data available (learning machine approach looking for patterns or sequences) could shed more light.