



VERISIGN™

Analysis of query traffic to .com/.net name servers

Duane Wessels

Principal Researcher, Verisign Labs

Matt Larson

Director, Verisign Labs
VP, Research, Verisign, Inc.

DNS WG, RIPE 66, Dublin, Ireland, May 2013



Our Infrastructure

- Operator of A root (6 sites), J root (~70 sites)
- Registry for *.com/.net* (among other TLDs)
 - 17 “large” DNS resolution sites at large Internet exchanges
 - 67 additional “smaller” sites
- SPAN (switched port analyzer) feeds from large sites
 - Traffic backhauled to central analysis environment across our private backbone
 - Started with four sites, but numbers have varied



Our Data Source

- Data since March 2011
 - At first from four routers: dfw, nyc (2), sfo
 - Later, additional feeds: ams, chi, fra, iad, lax, lon, sea
- Homegrown libpcap app *srcip-count* runs on a server connected to backhailed SPAN feeds
- Counts queries, recording source IP, destination IP, query counts by type, DO, RD, EDNS buffer size, and min/max IP TTL
- Writes a data file every 60 seconds

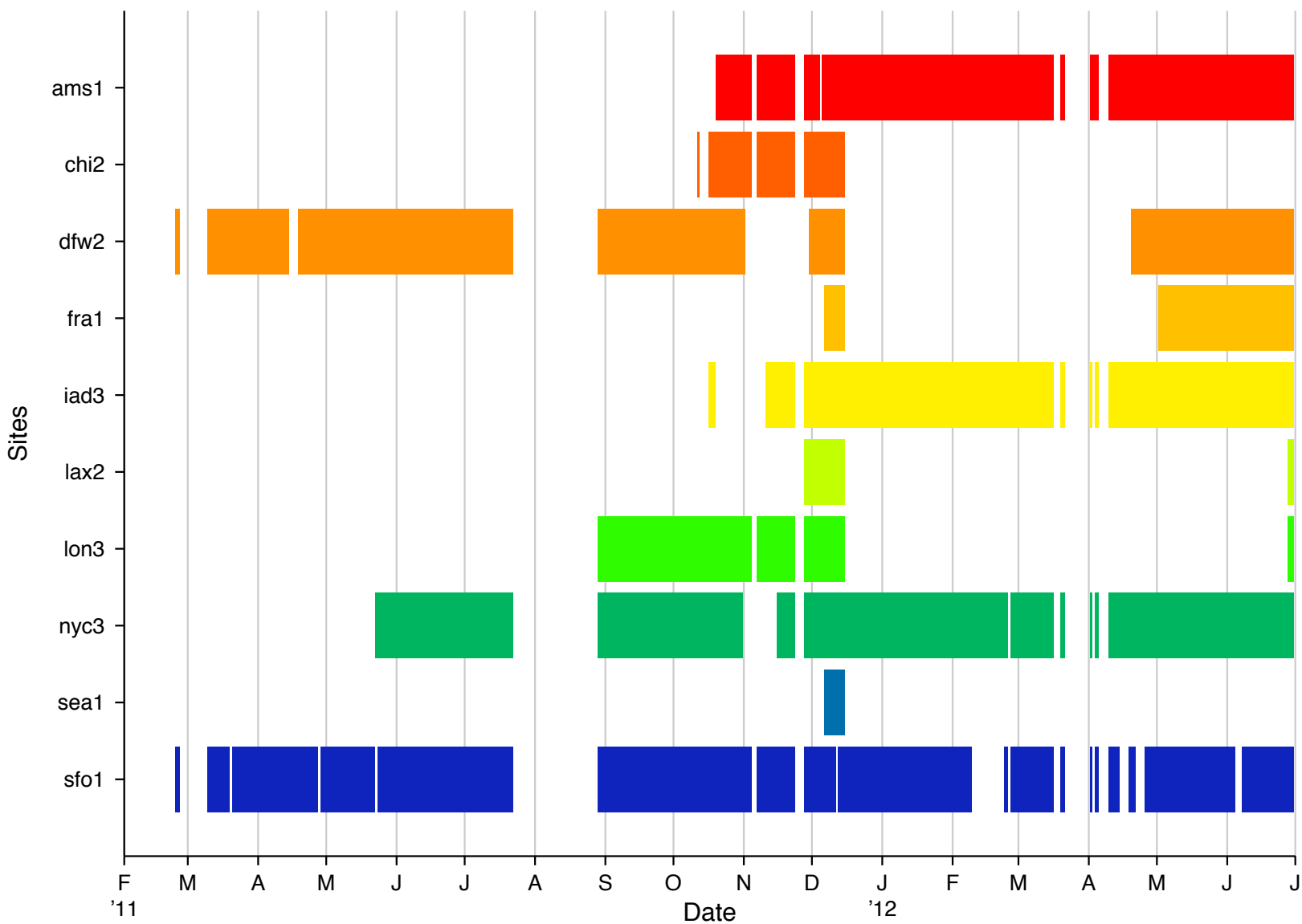
```
# File format version: 1.0 ($Id: srcip-count.c 47 2011-05-12 20:25:18Z mlarson $)
# Interval: 1306932420 1306932480 (2011-06-01 12:47:00 -- 2011-06-01 12:48:00)
# Columns: src dst bufsize rd do aaaaa mx ptr ds dnskey ns other min_IP_ttl max_IP_ttl
200.201.17.20 192.55.83.30 0 0 0 1 0 0 0 0 0 0 0 115 115
184.154.234.2 192.55.83.30 4096 0 11 4 4 0 0 0 0 3 0 55 55
85.13.135.58 192.55.83.30 4096 0 1 0 0 1 0 0 0 0 0 57 57
216.206.32.22 192.55.83.30 0 0 0 1 0 0 0 0 0 0 0 53 53
200.27.131.51 192.55.83.30 0 0 0 3 0 0 0 0 0 0 0 116 116
93.244.27.60 192.55.83.30 0 6 0 4 0 2 0 0 0 0 0 116 116
```



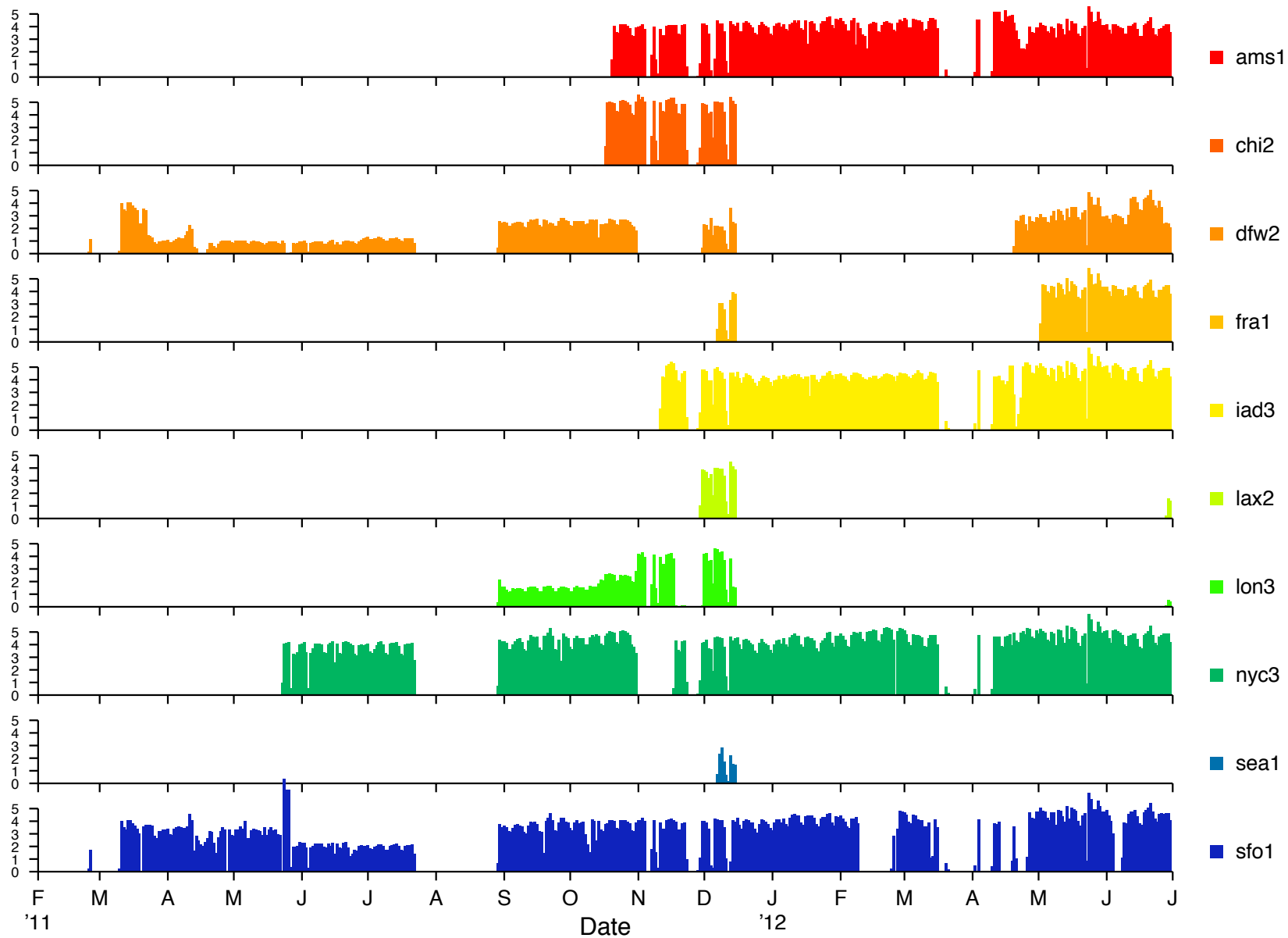
Our Three Data Sets For This Talk

- 1-Month: *srcip-count* data for the month of June, 2011
- 15-Month: *srcip-count* data from March, 2011 to June, 2012
 - Minus August, 2011
 - Minus parts of March/April, 2012
 - Between 4-10 sites, coming in from SPAN feeds
 - Queries to *gtld-servers.net* name servers only
 - I.e., authoritative servers for *.com* and *.net*
- Monthly 24-hour pcap snapshot files of “all” traffic

Site Coverage



Queries Per Day (Billions)





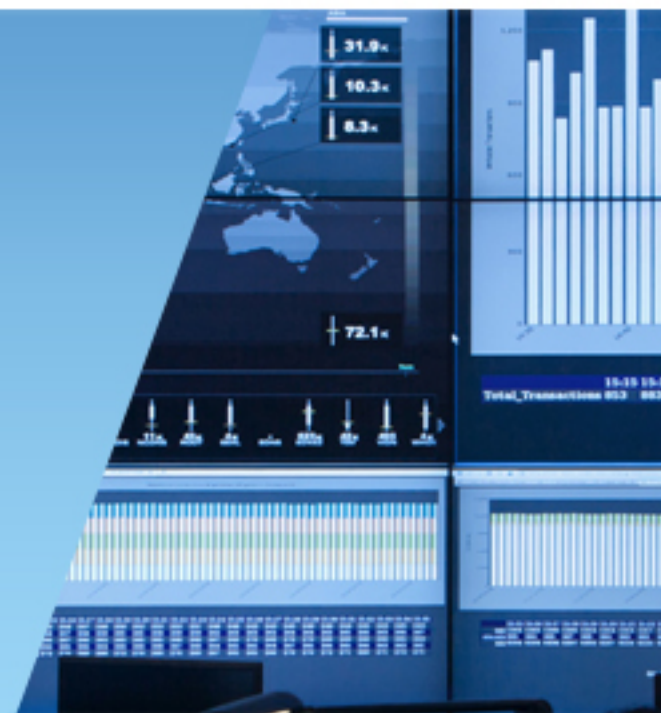
Agenda

- Number of Clients
- Query Type Distribution
- AAAA Queriers
- Characterizing Top Talkers
- RD=1 Queries
- Observations



VERISIGN™

Number of Clients

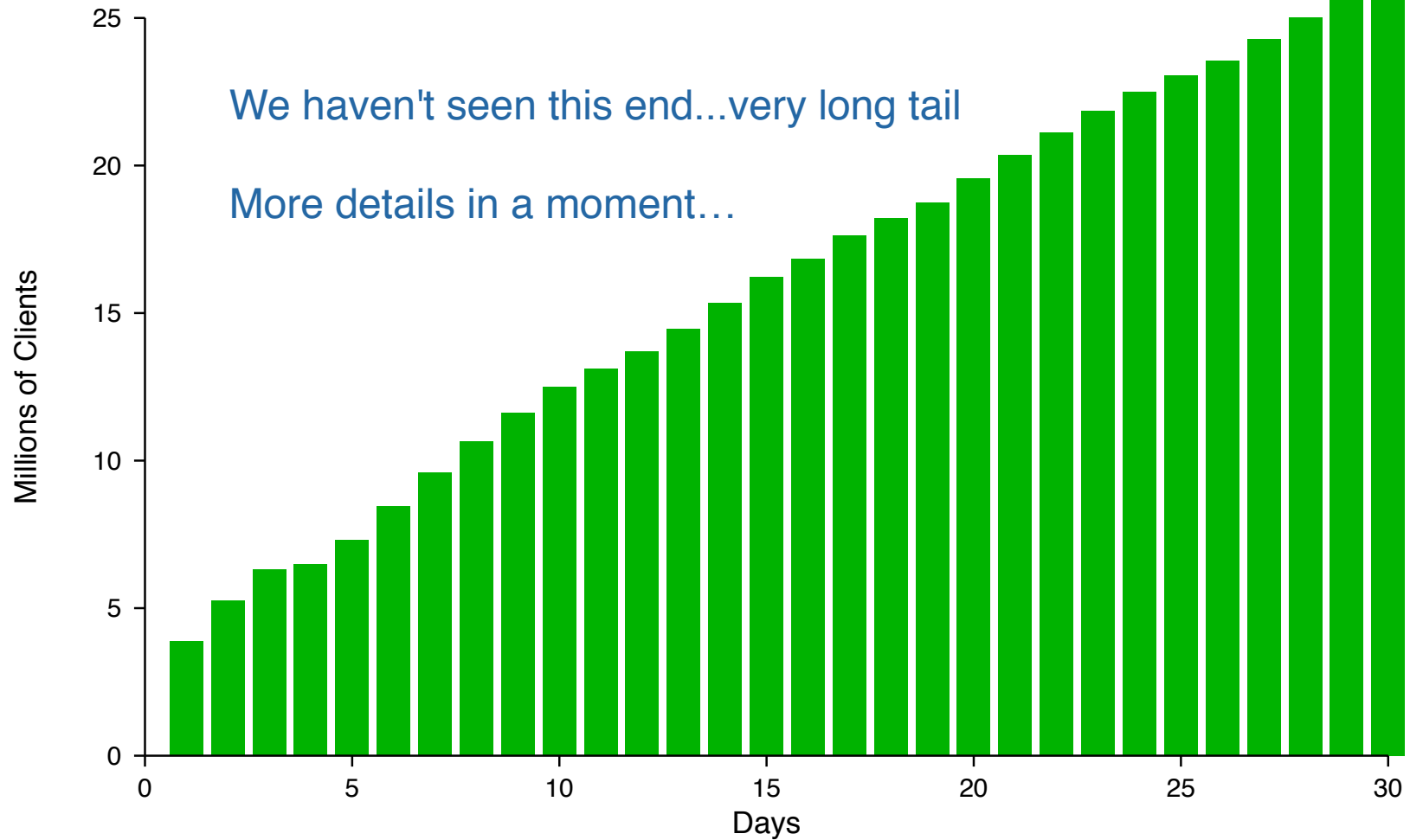




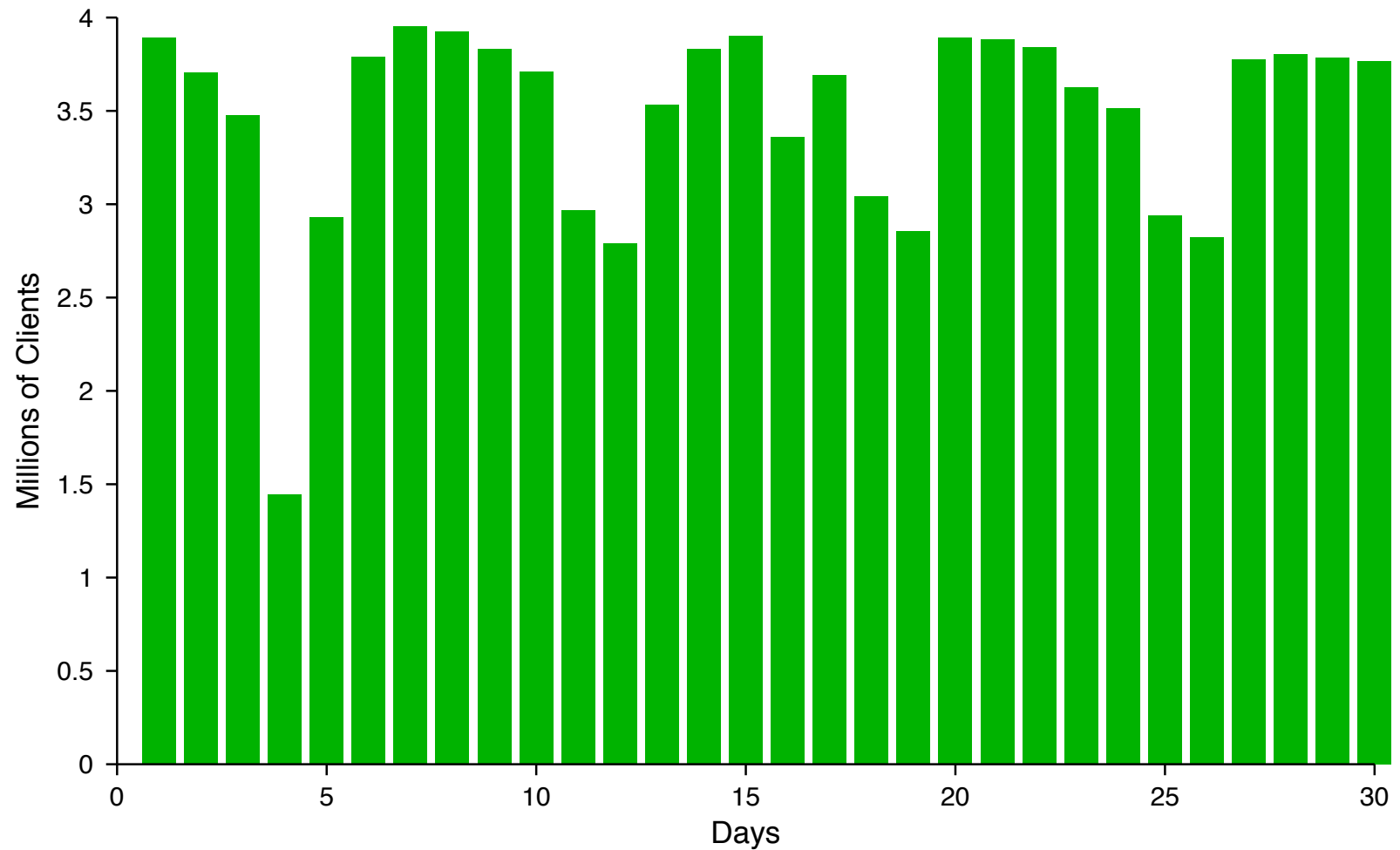
How Many Unique IPs Queried *.com/.net* in June 2011?

26,437,427

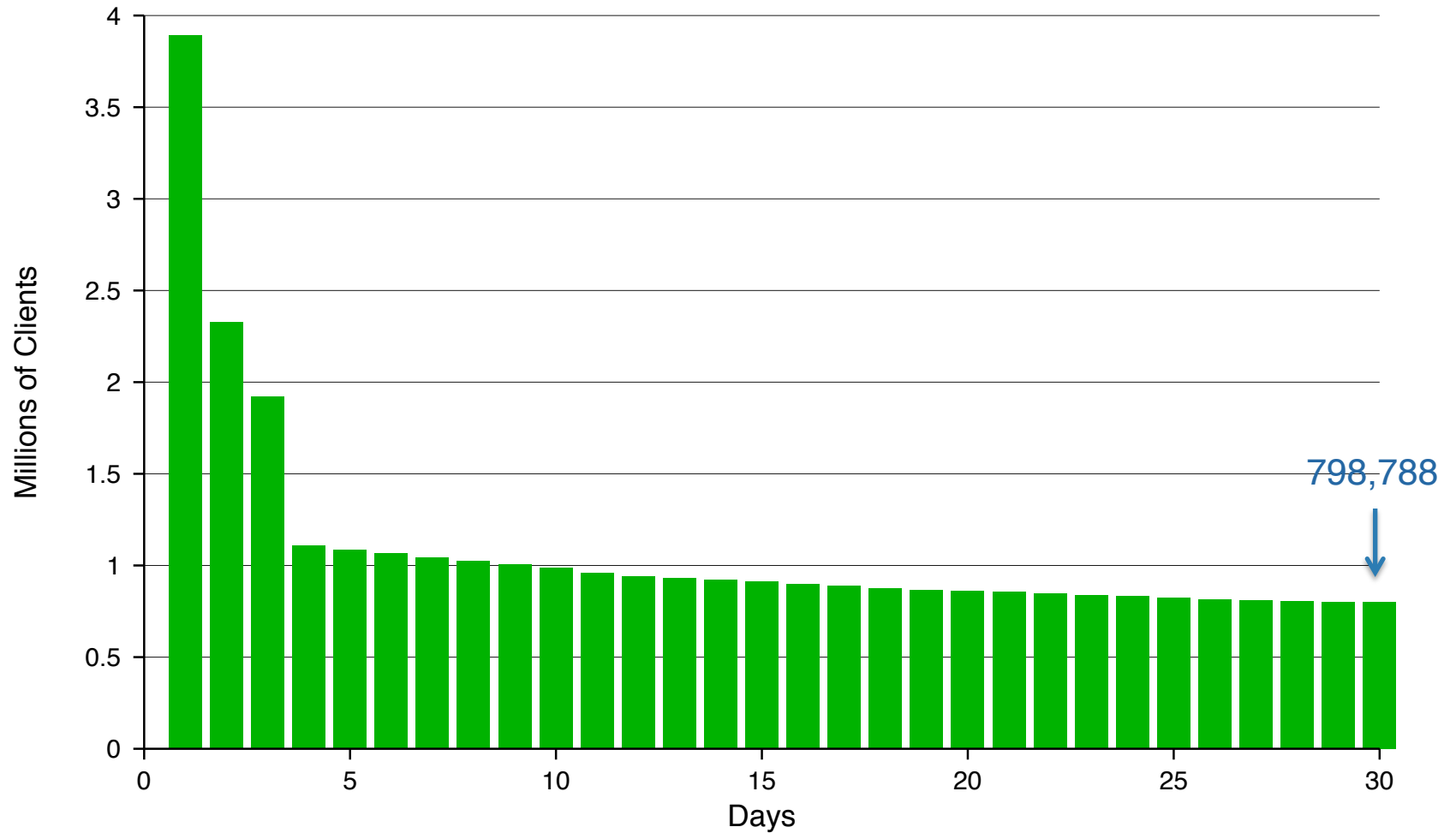
Cumulative Number of Unique Clients (June 2011)



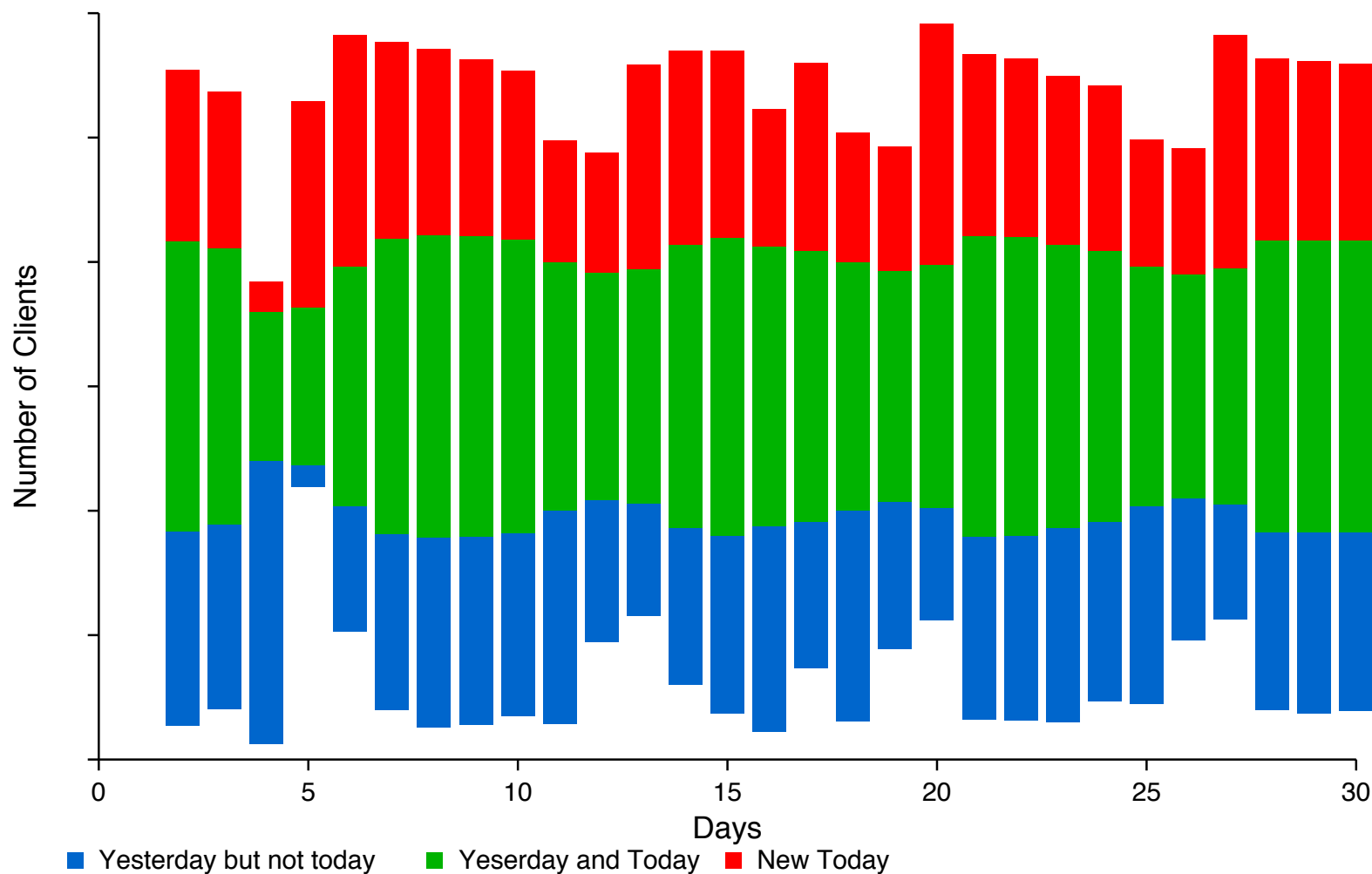
Unique Clients Per Day (June 2011)



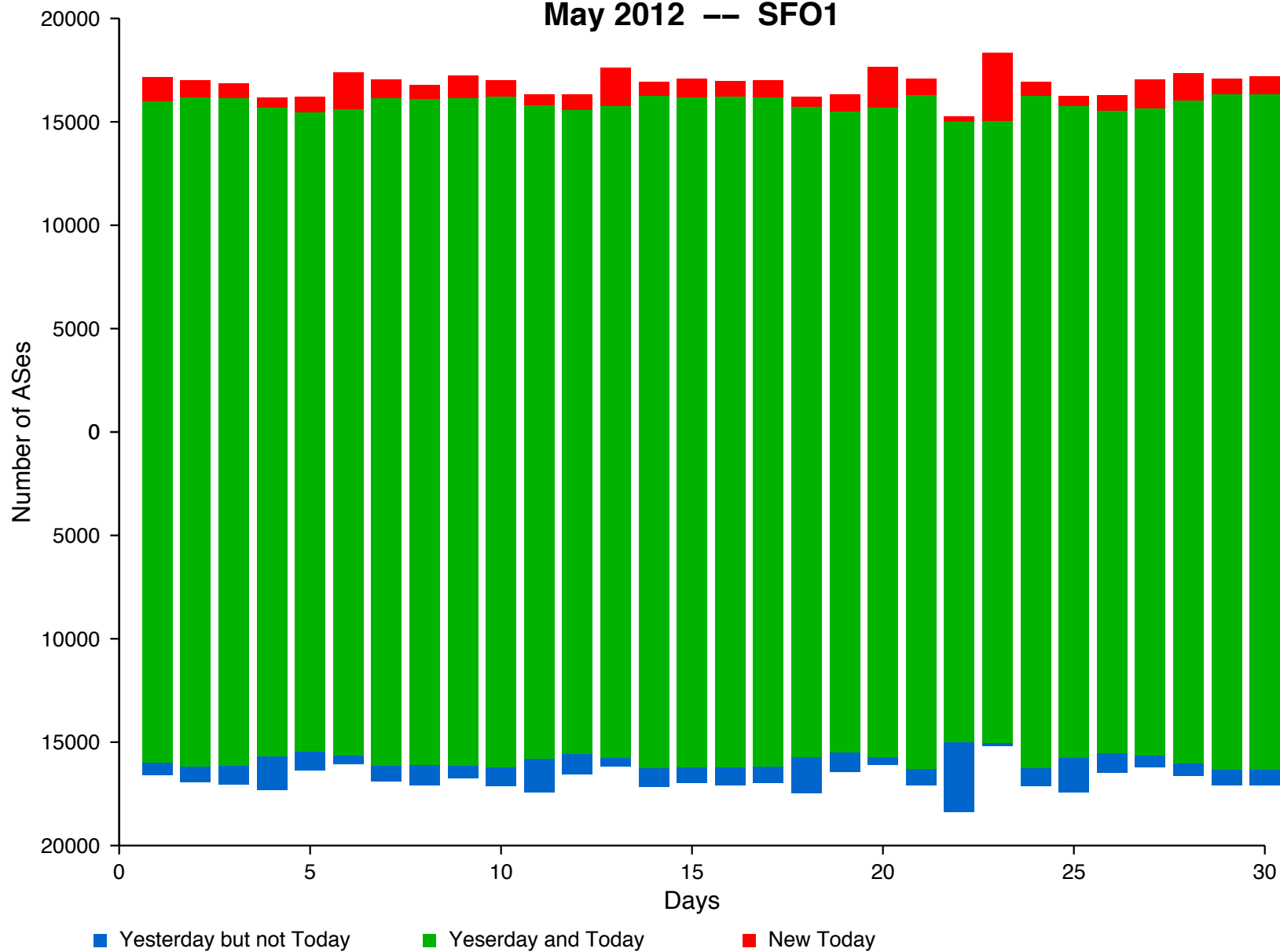
Number of Clients Seen Every Day Since The Start (June 2011)



How Clients Change Over Time (June 2011)



Daily Changes in ASes Seen May 2012 -- SFO1

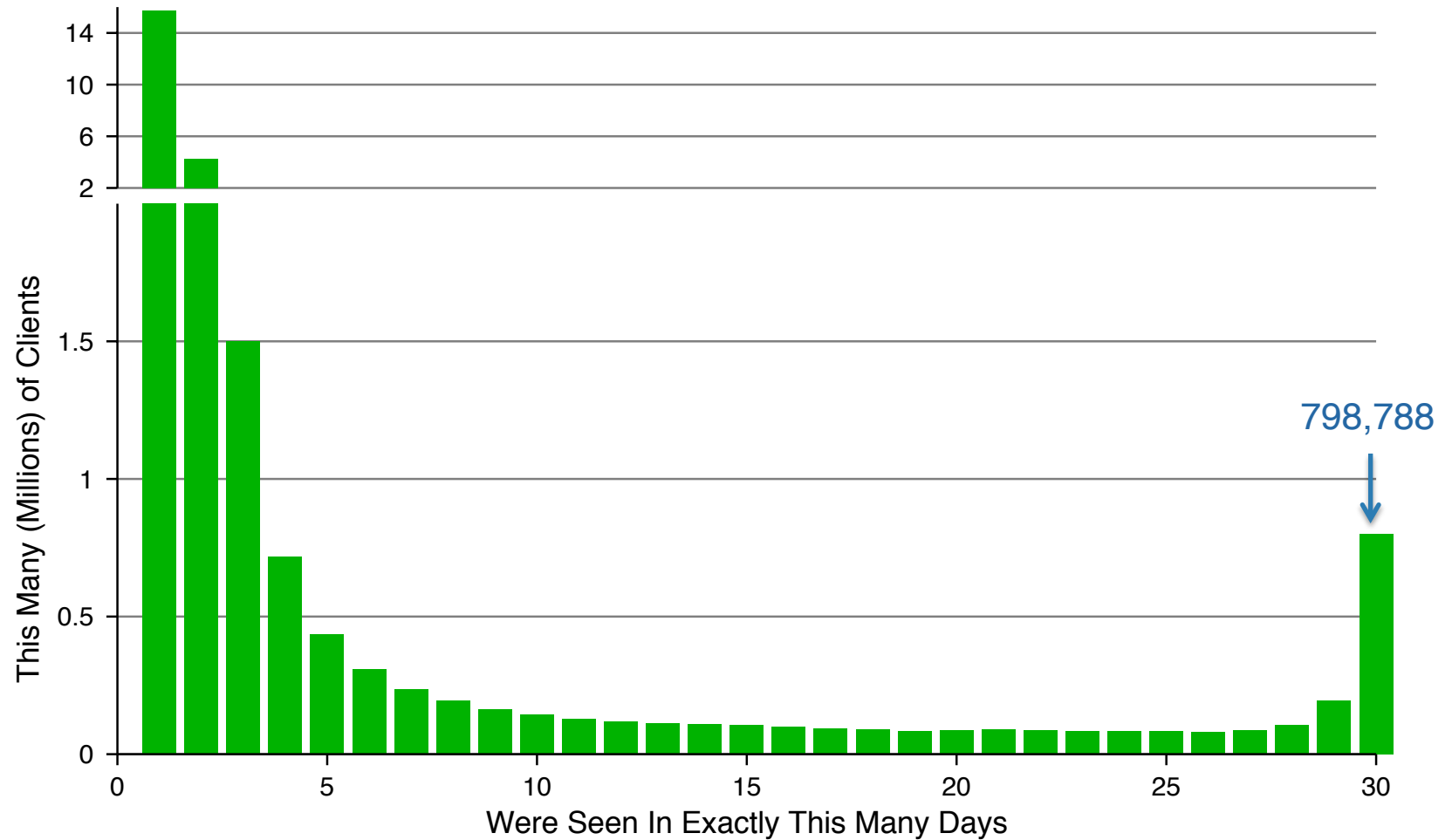




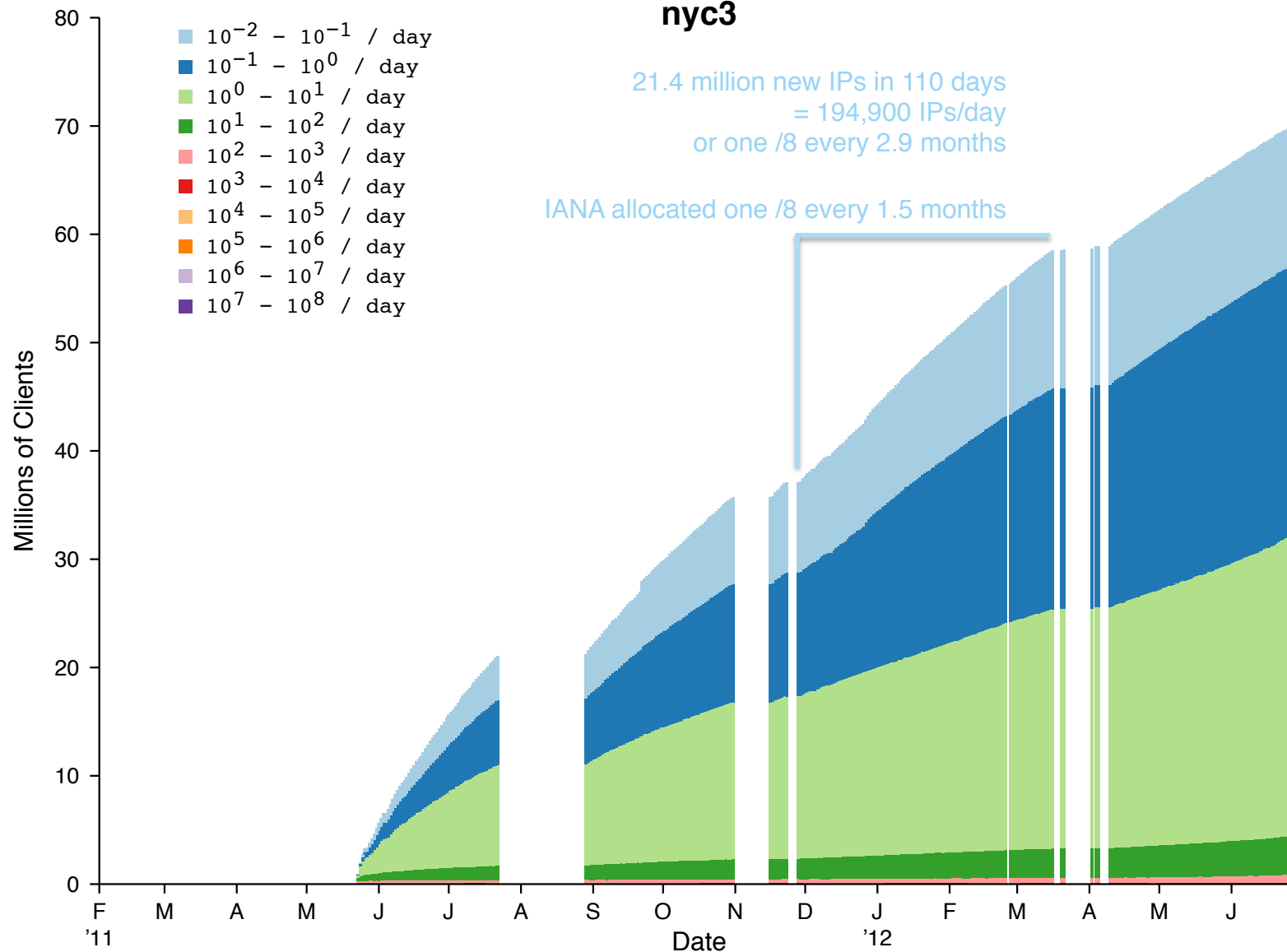
How Many IPs Query Less Than Once Per Day?

15,626,031

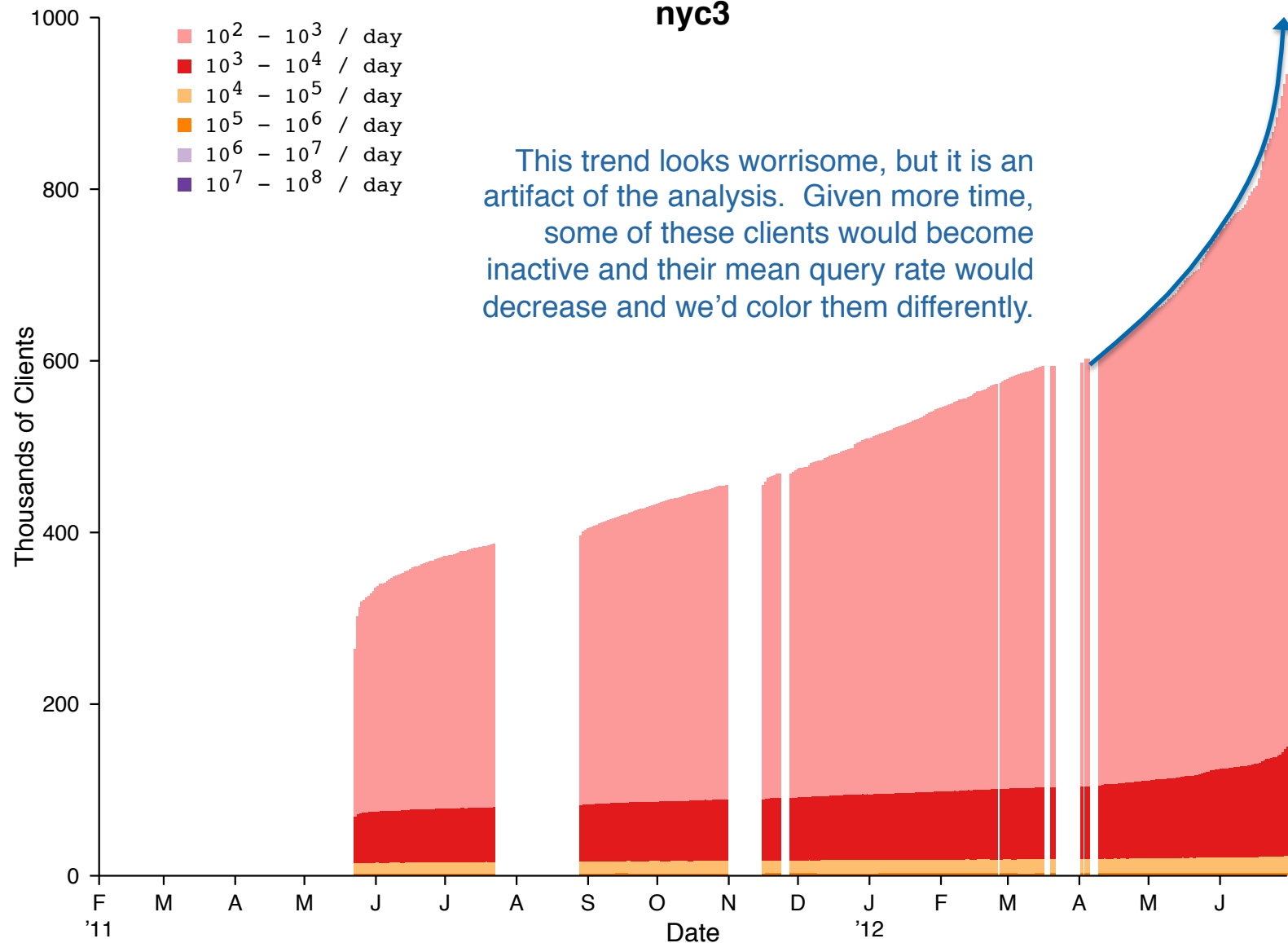
Histogram of Clients Seen per Number of Days (June 2011)



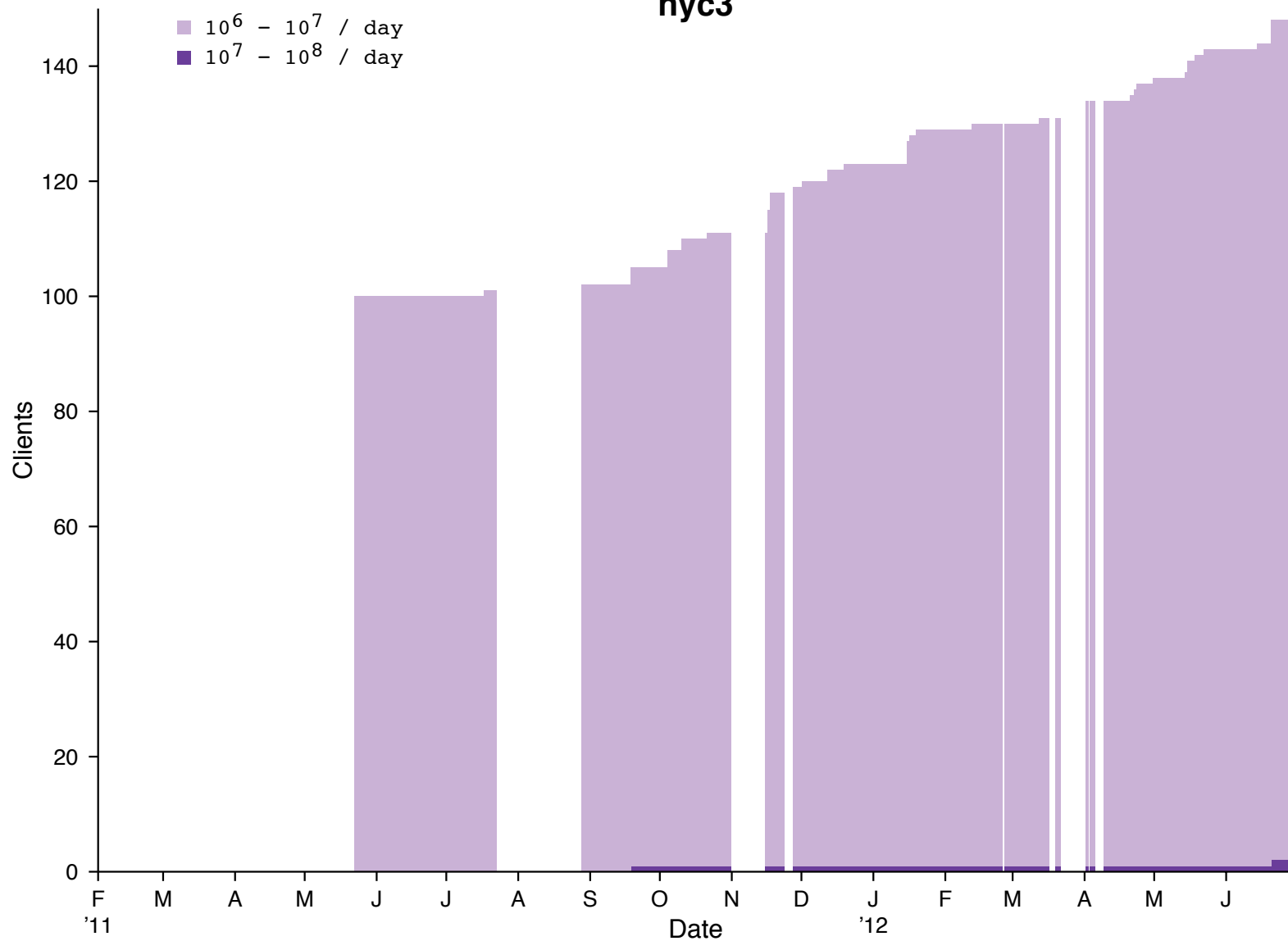
Cumulative Client Count nyc3



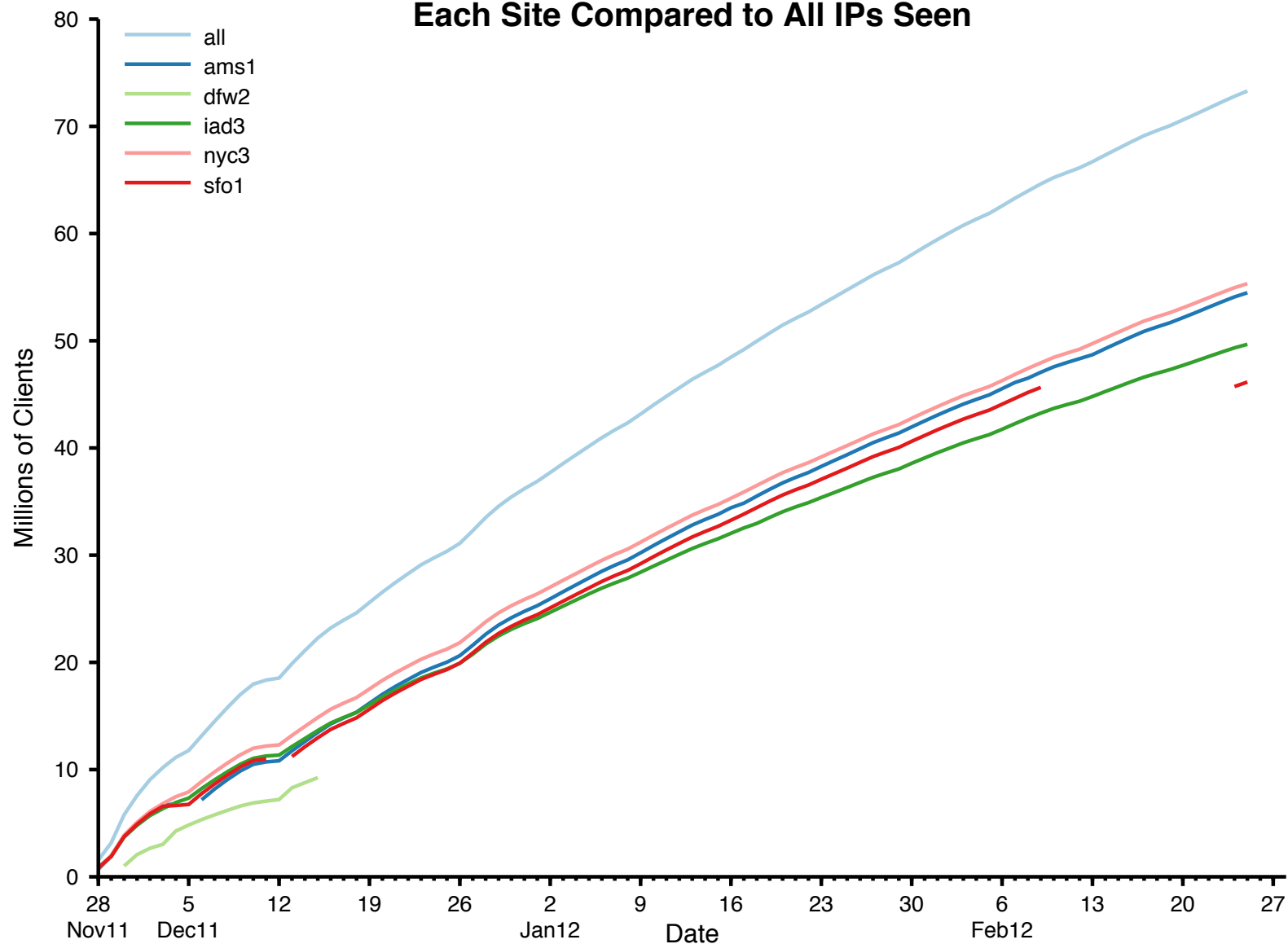
Cumulative Client Count nyc3



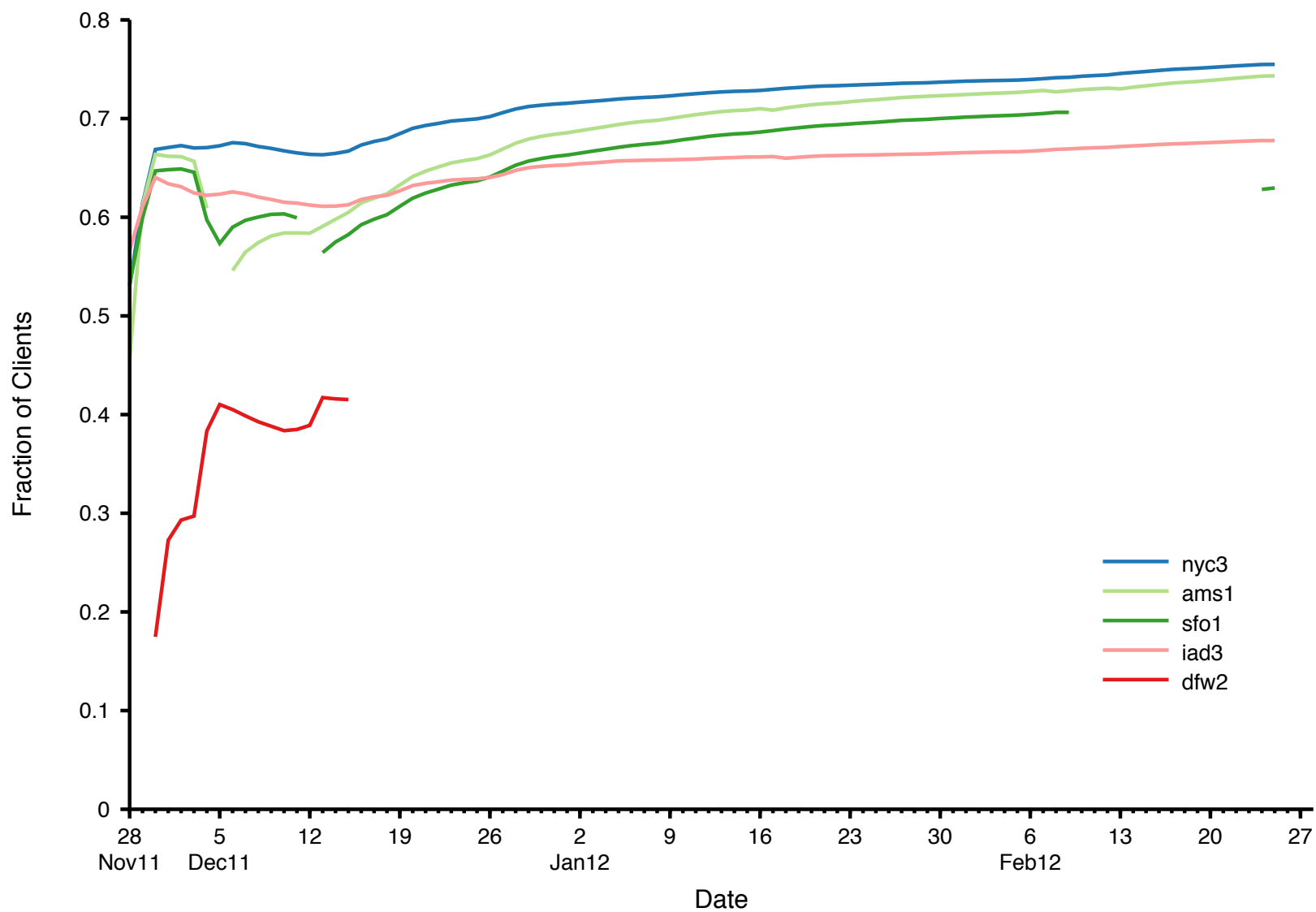
Cumulative Client Count nyc3



Cumulative Client Count Each Site Compared to All IPs Seen



Cumulative Client Count Each Site Compared to All IPs Seen





VERISIGN™

Query Type Distribution



Query Type Profiles

- Want to create profiles of resolvers based on the types of queries they send
- Define a profile as the fraction of:
 - A, AAAA, MX, PTR, NS, DS, DNSKEY, Other
 - The profile values sum to 1
- Treat profiles as points in 8-D space
- Then use some clustering techniques to identify the common profiles

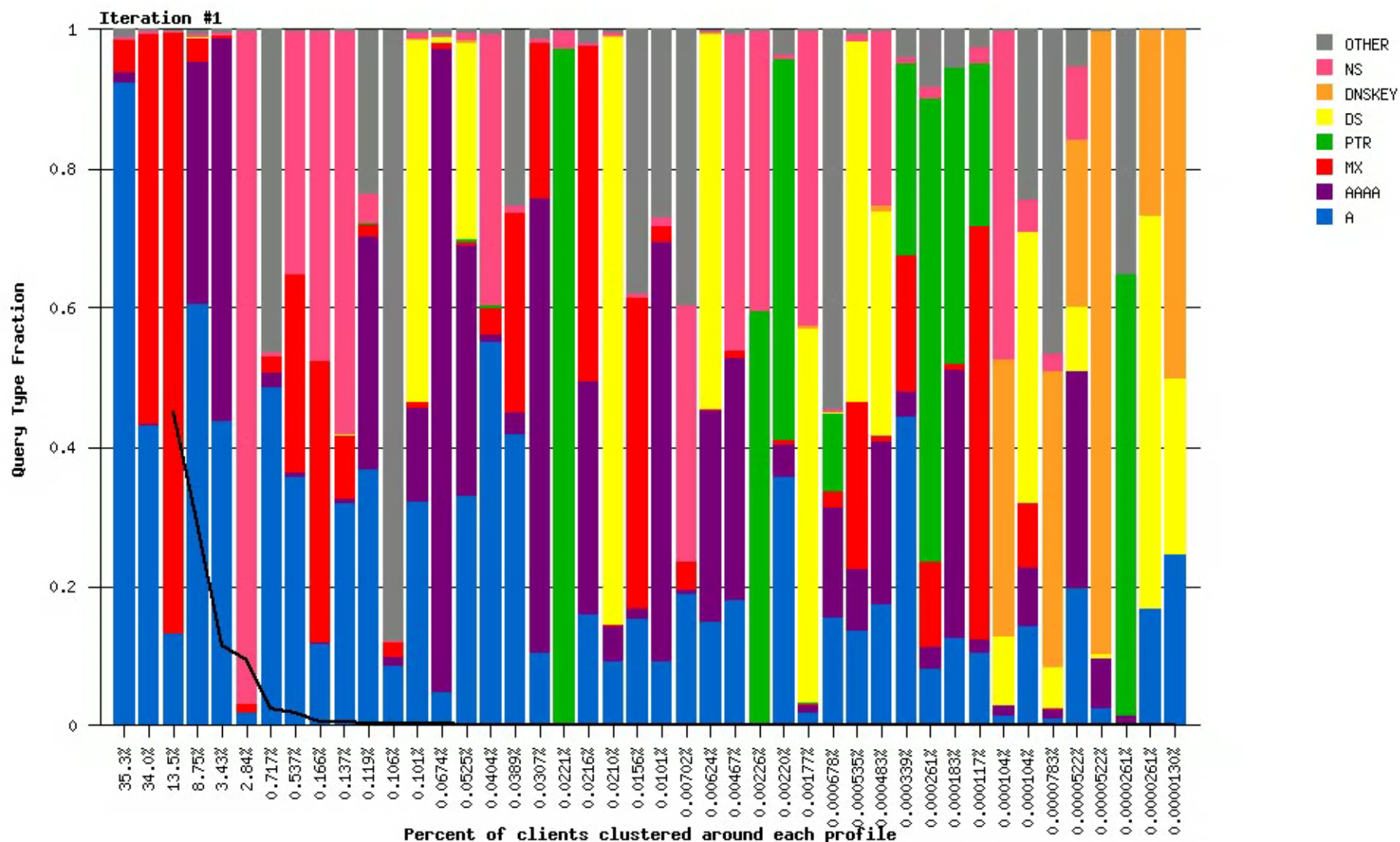
K-Means Clustering

- Select K random initial “centers”
- For each point in the data, find its closest center
- Recalculate centers as mean of all points closest to it
- Repeat
- Stop when centers stabilize according to some metric
- Results are good if different random starting points lead to the same final points

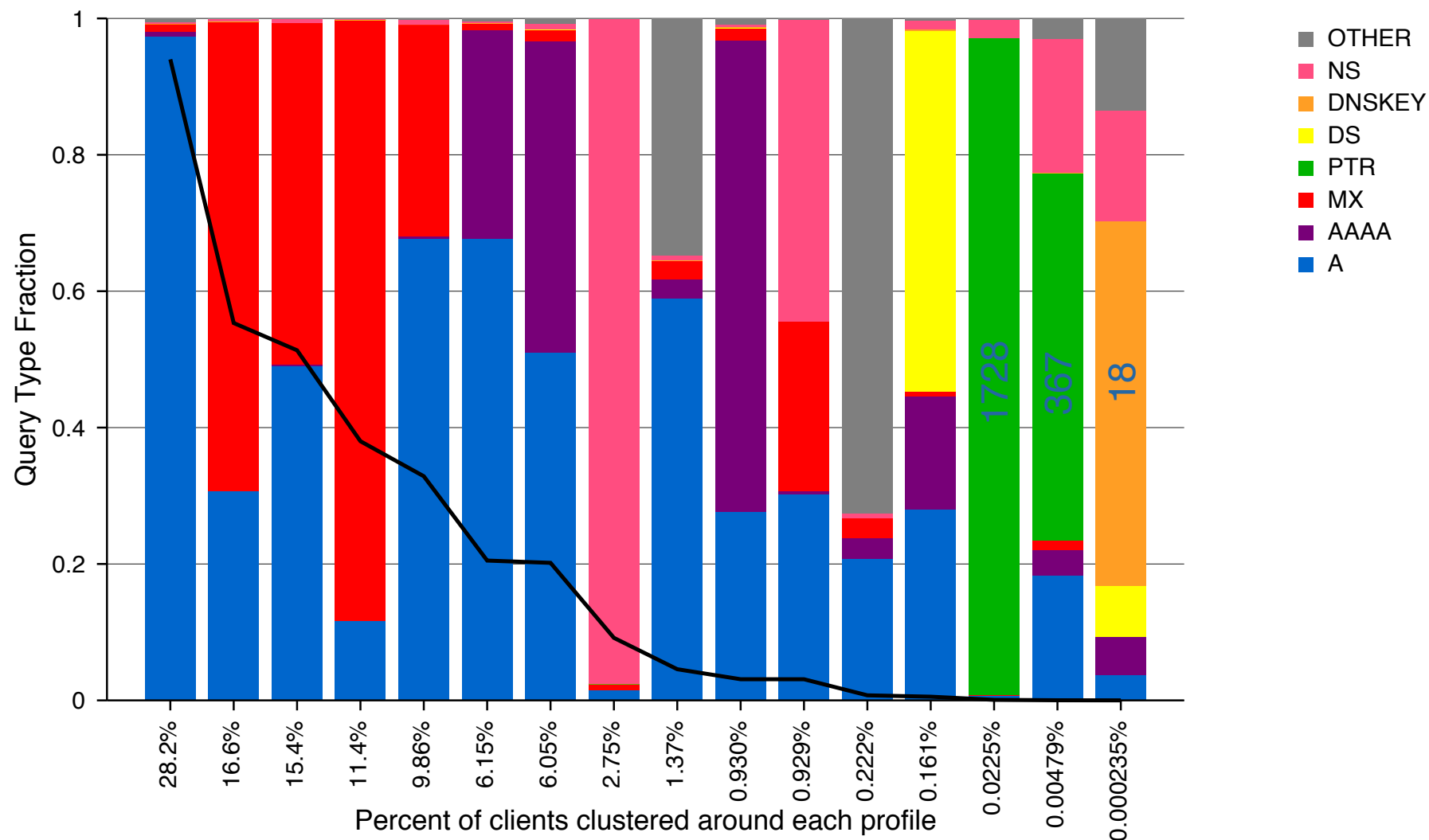
K-Means Clustering to Find Profiles

- Begin with approx 60 random centers subject to following constraints:
 - Center must lie on the 7-D surface described by
$$C_A + C_{AAAA} + C_{MX} + C_{PTR} + C_{NS} + C_{DNSKEY} + C_{DS} + C_{Other} = 1$$
 - Centers must be separated by 0.4 units
- Ignore clients that send less than 100 queries (per month)
 - Data includes 7,664,000 of 26,437,000 clients (29%)
- Every 10 iterations, remove the center that is closest to all other centers
- Stop when there are 16 centers remaining

Finding Clusters of Query Type Profiles



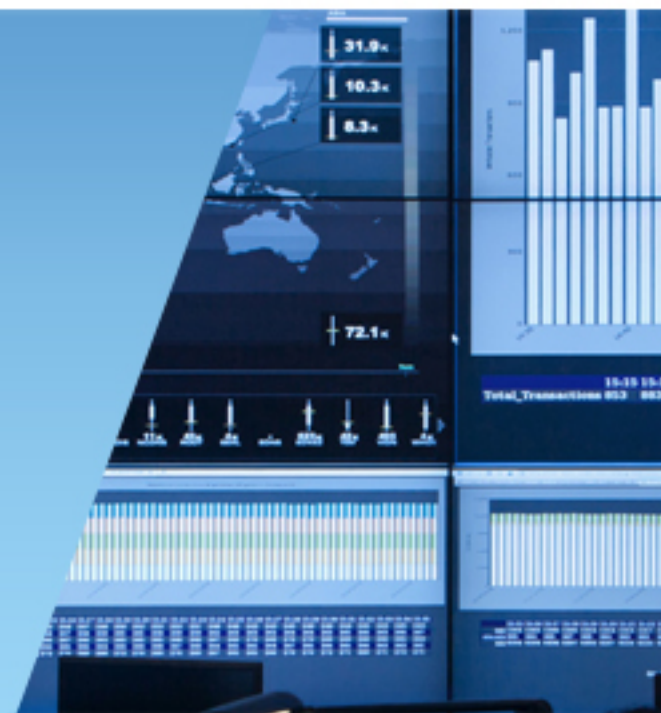
Query Type Profile Clusters





VERISIGN™

AAAA Queriers



Queries for A and AAAA

- One measure of IPv6 deployment is queries for AAAA
 - How many source IPs send at least one AAAA query vs. at least one A query?
 - How many total queries for AAAA vs. total queries for A?

SOURCES

6/2011

A	91.4%	19,925,511
AAAA	14.7%	3,214,272
ALL	100.0%	21,811,360

6/2012

A	95.1%	31,020,570
AAAA	18.0%	5,869,714
ALL	100.0%	32,607,318

QUERIES

6/2011

A	68.7%	93,413,090,254
AAAA	12.3%	16,701,014,767
ALL	100.0%	135,914,964,320

6/2012

A	66.3%	492,837,069,754
AAAA	11.8%	87,371,872,688
ALL	100.0%	742,991,989,695

Queries for A and AAAA

- One measure of IPv6 deployment is queries for AAAA
 - How many source IPs send at least one AAAA query vs. at least one A query?
 - How many total queries for AAAA vs. total queries for A?

SOURCES

6/2011

A	91.4%	19,925,511
AAAA	14.7%	3,214,272
ALL	100.0%	21,811,360

6/2012

A	95.1%	31,020,570
AAAA	18.0%	5,869,714
ALL	100.0%	32,607,318

QUERIES

6/2011

A	68.7%	93,413,090,254
AAAA	12.3%	16,701,014,767
ALL	100.0%	135,914,964,320

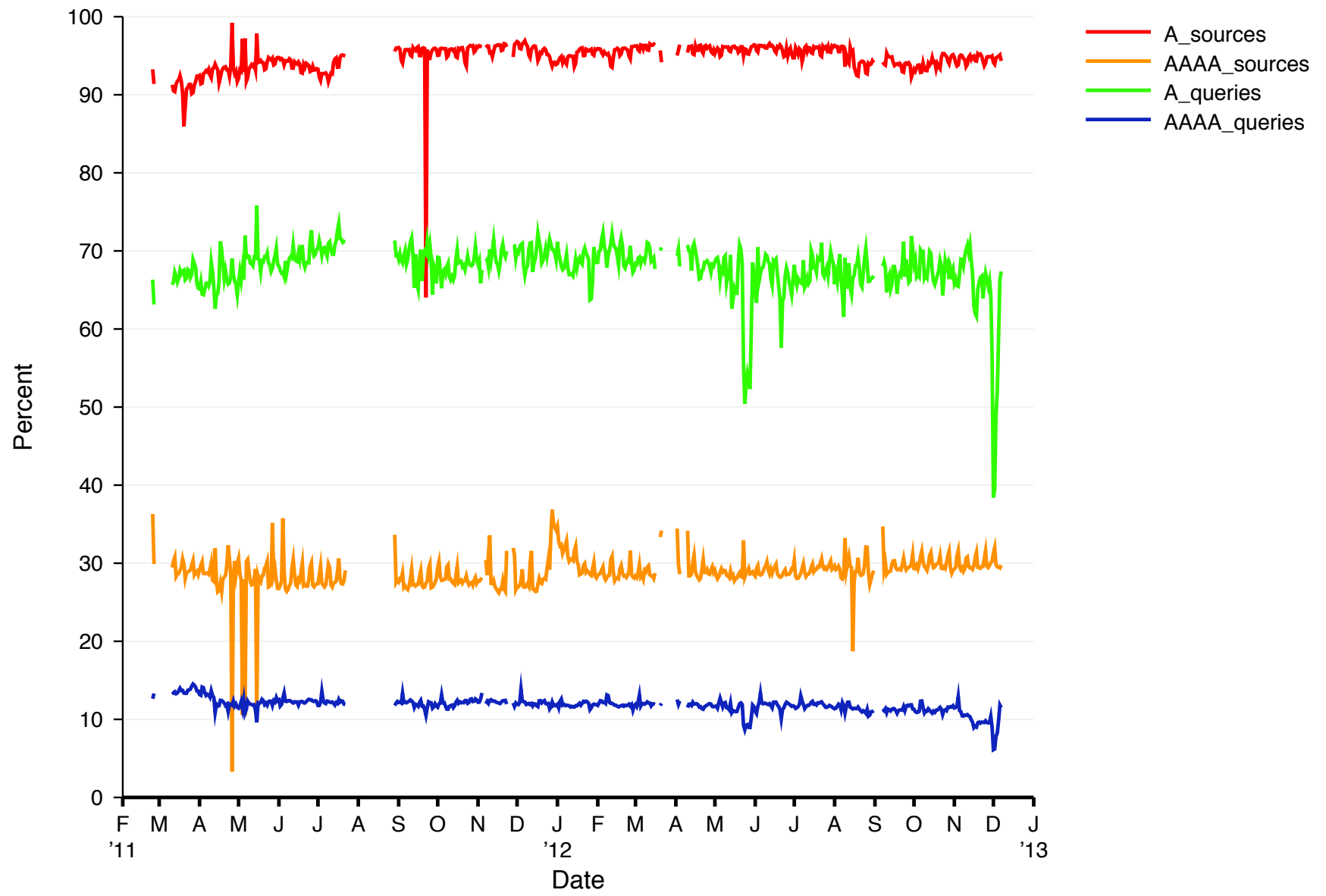
6/2012

A	66.3%	492,837,069,754
AAAA	11.8%	87,371,872,688
ALL	100.0%	742,991,989,695



VERISIGN™

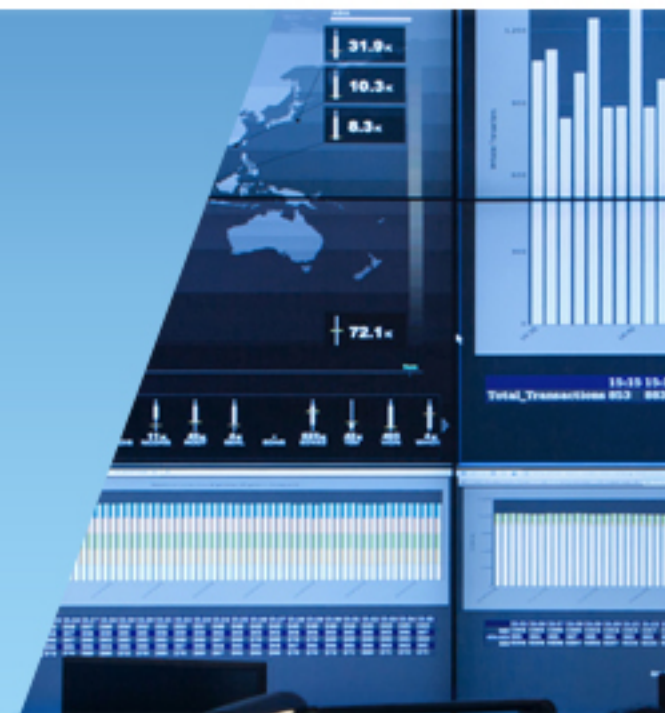
A vs AAAA Queries



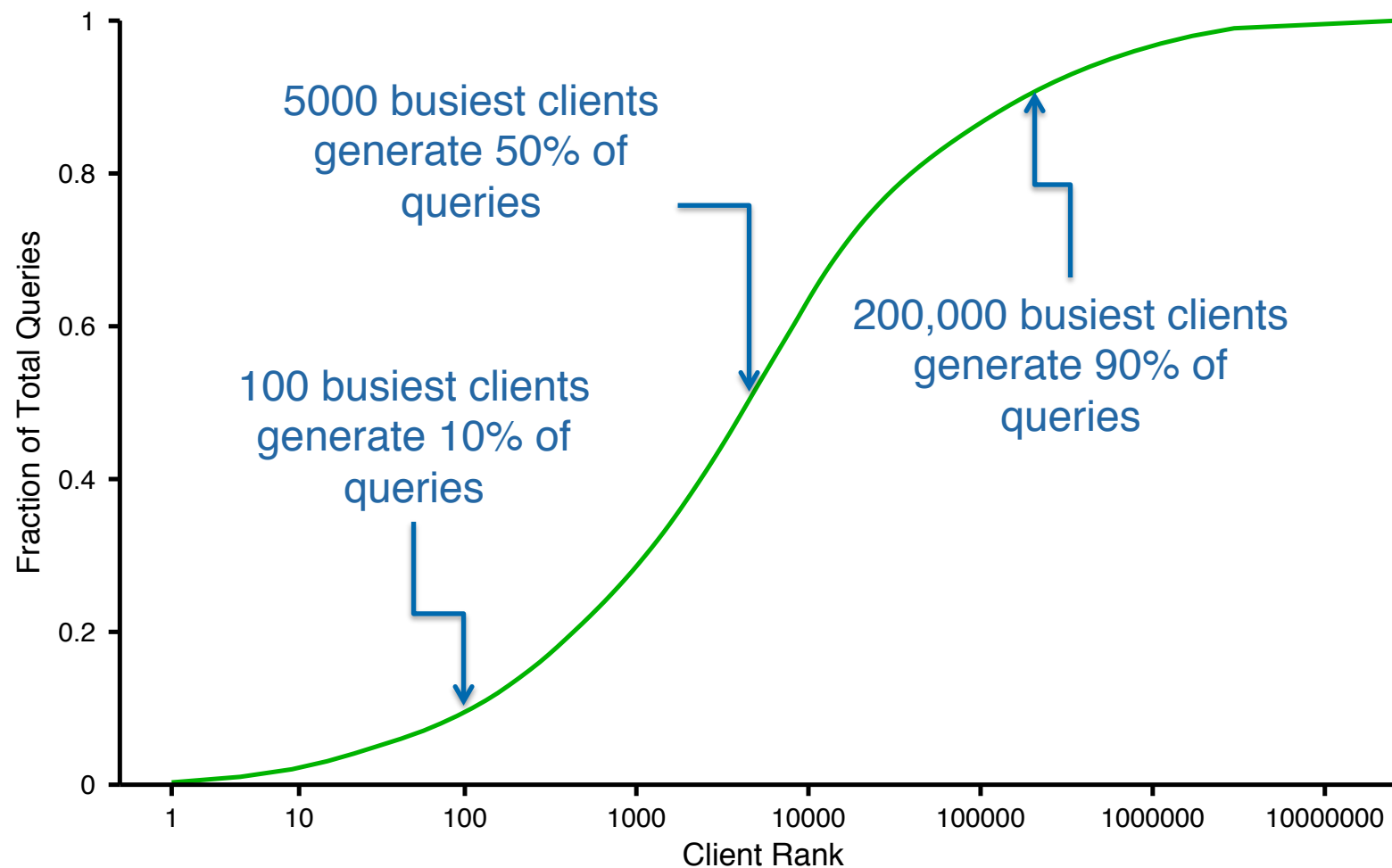


VERISIGN™

Characterizing Top Talkers



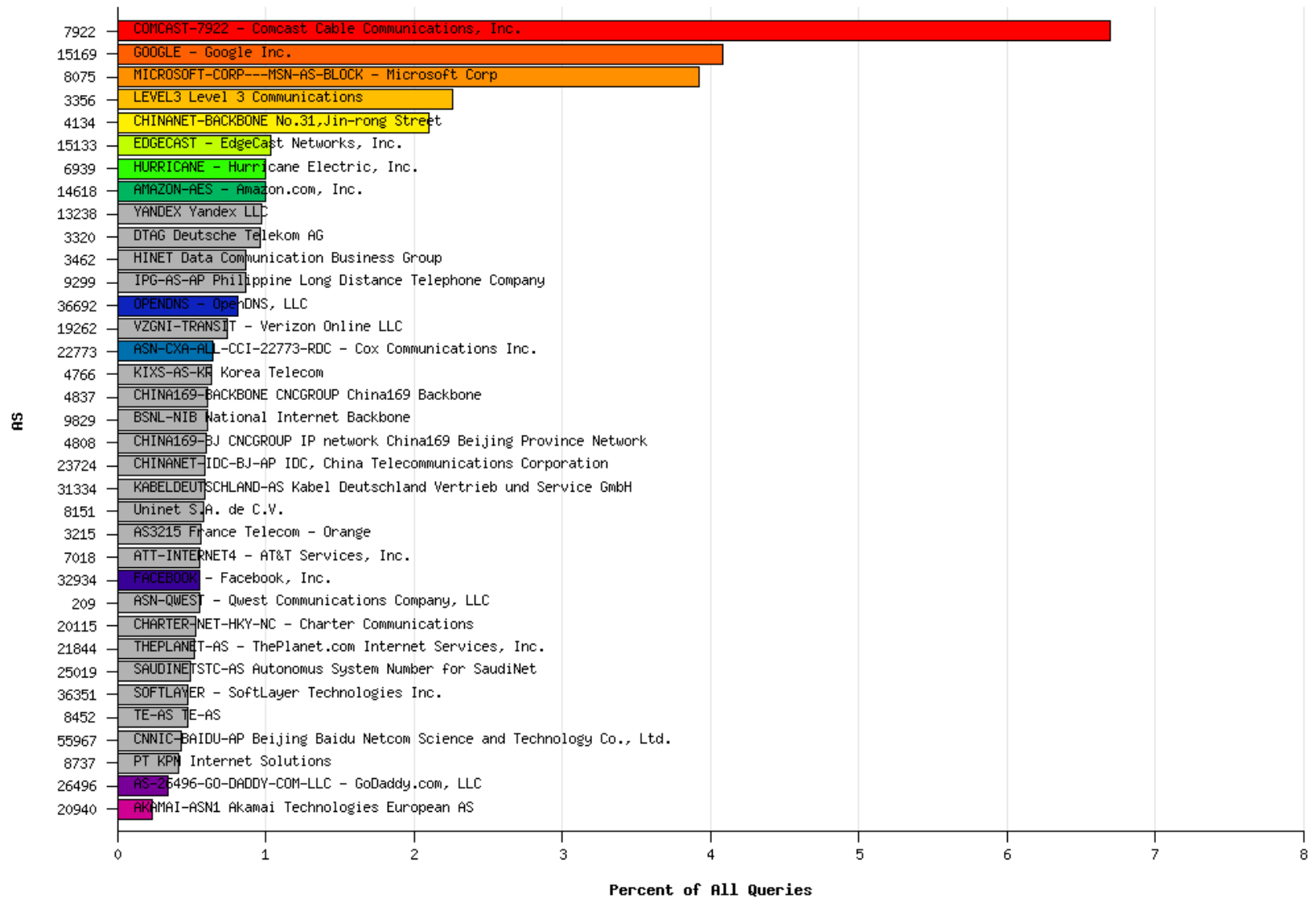
Ranked Clients CDF (June 2011)





VERISIGN

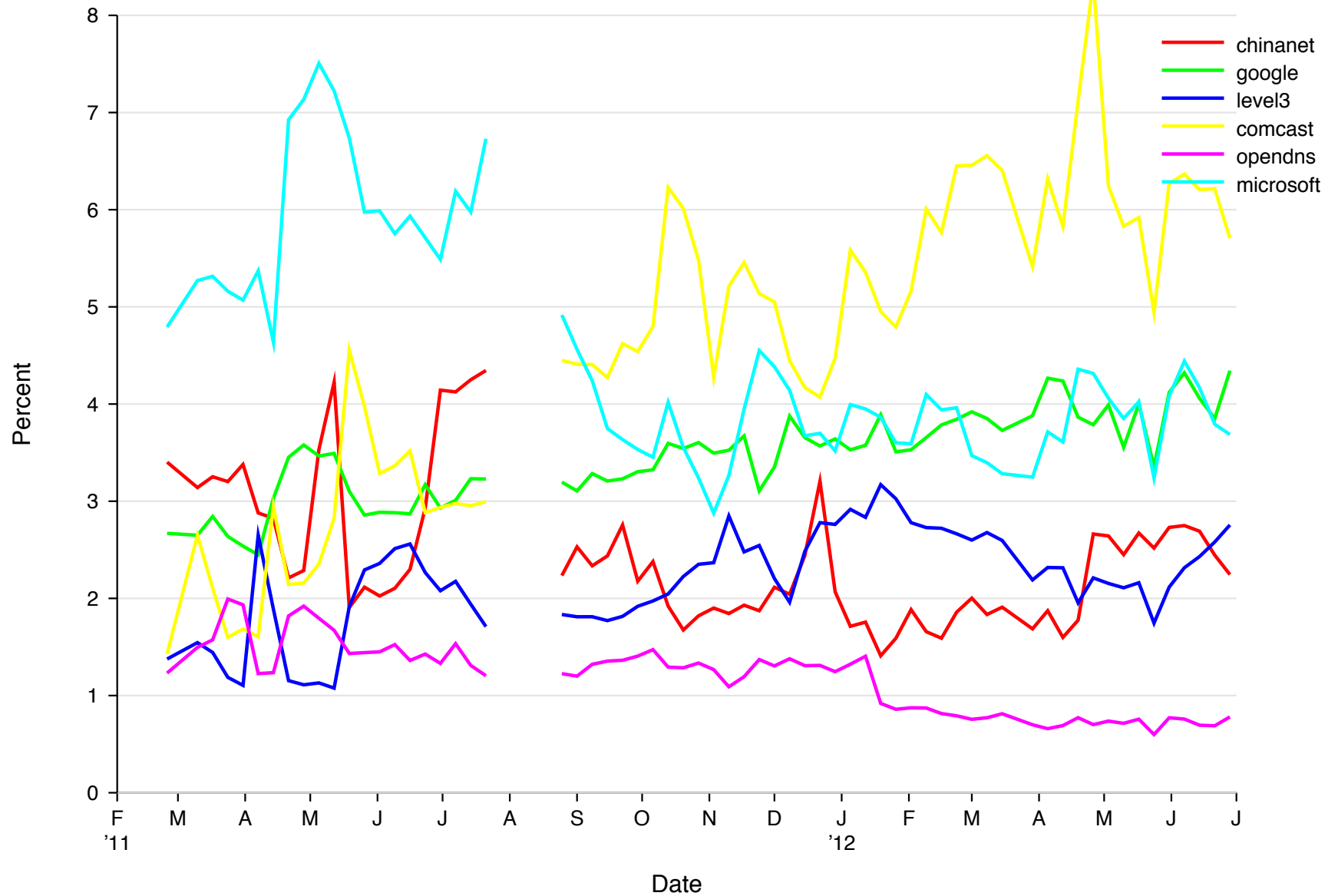
Interesting and Top ASes 20120627-all





VERISIGN™

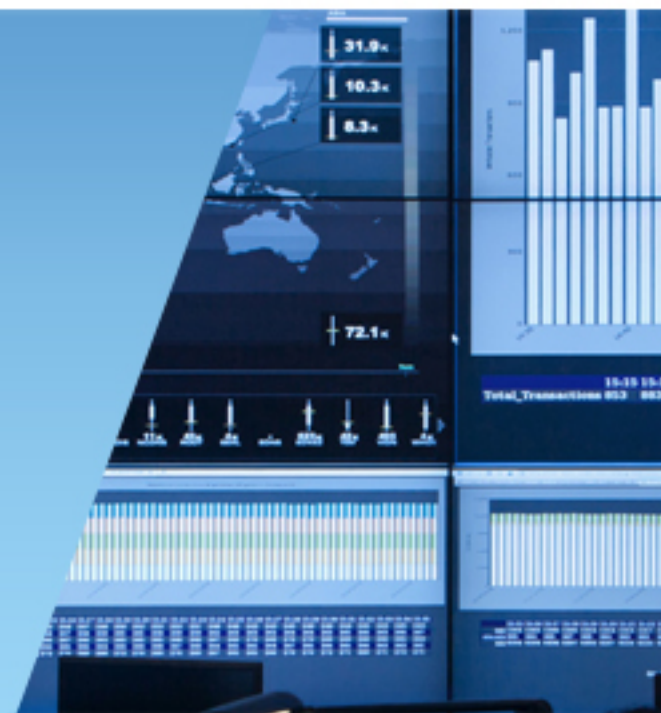
What Percent of All Queries Come These Companies?





VERISIGN™

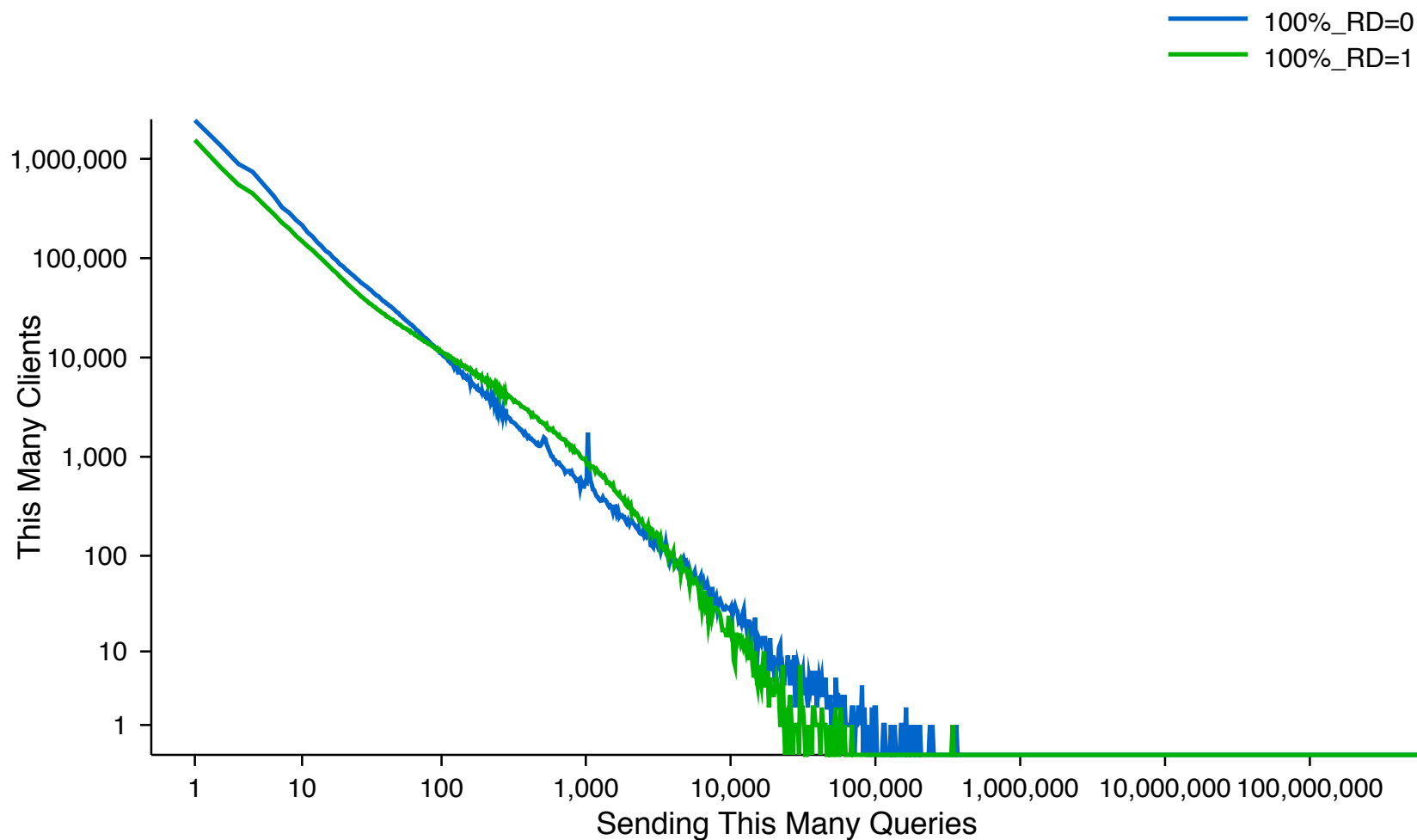
RD=1 Queries



Recursion Desired

- The RD bit is set by dumb *stub resolvers* that rely on smarter *iterative resolvers* to do the hard work of following referrals
 - Iterative resolvers are found in *recursive name servers*
- In theory stub resolvers should not be sending queries to the *.com/.net* name servers.
 - In other words, most queries we receive should have RD=0
- Some debugging tools, like dig, set RD=1 by default, so we can expect to see some RD=1 queries
- What do we really see? (June 2011)
- **6.79%** of all *queries* have RD bit set
- **53%** *clients* RD=0, **42.5%** RD=1, **4.5%** mix

How Many Queries Do 100% RD=1 and RD=0 Clients Send? (June 2011)



Let's Look at Sources Sending Only RD=1

- Examine our monthly archived pcap files from January through June 2012
 - We save 24 hours of traffic one day per month and retain permanently
- From four sites (ams1, iad3, nyc3, sfo1), generate a list of sources with 100% RD=1 queries and order by query count
- From the list of sources, select a logarithmically based subset of sources to examine in full
 - About 11 sources per site per date.
- Analyze each source



RD=1 Analysis

- Most common query type
- Most common query name (at least 5% of queries)
- Most common second level domain (at least 5%)
- Source port entropy (bits)
- Query ID entropy (bits)
- Similarity between first-components of names

RD=1 Results

Source	Site	Query Count	Popular Type	Source Port Bits	Query ID Bits	Name Similarity	Popular SLD	Popular Name
69.43.160.96	nyc3	1,627,963	NS [99%]	8.9	11.4	66.4	--	--
216.69.191.248	sfo1	1,256,491	MX [80%]	11.7	11.7	5.0	aardvarkpest.com [56%]	aardvarkpest.com [56%]
91.102.117.68	dfw2	456,432	DNSKEY[100%]	0.0	-2.0	0.0	. [100%]	. [100%]
59.109.76.18	ams1	385,263	TXT [94%]	13.4	13.4	0.2	osvr.com [93%]	osvr.com [93%]
220.130.119.236	sfo1	302,248	A [83%]	13.7	13.8	1.4	spamcop.net [15%]	--
220.130.119.230	ams1	311,759	A [83%]	13.7	13.7	1.4	spamcop.net [14%]	--
220.13.119.232	ams1	279,425	A [83%]	13.8	13.9	1.4	spamcop.net [14%]	--
173.201.193.177	sfo1	274,864	MX [74%]	13.9	13.9	8.6	tessag.com [63%]	tessag.com [63%]
41.215.54.6	sfo1	15,644	MX [92%]	13.6	2.1	6.8	--	--
113.167.125.110	ams1	195	A [61%]	14.2	7.5	24.4	zenon.net [10%]	dns1.zenon.net [5%]

How Did We Calculate #Bits?

- For a uniform discrete distribution, the variance (square of standard deviation) is

$$\frac{n^2 - 1}{12}$$

- Where n is the largest possible value in the distribution
- For a given sample of integers, we can calculate its standard deviation, and thus its variance, and invert the above equation to estimate n
- For example, the variance of (1,2,3,4,5,6) is 3.5

$$\sqrt{(3.5 \cdot 12) + 1} = 6.557$$

$$\ln_2 6.557 = 2.7 \text{ bits}$$



Name Similarity Metric

- Sort all first-label names seen from the client
- Calculate the fraction of similar characters in sorted word pairs.
 - Similar to Hamming Distance
- Average all the per-pair fractions.

MX Queries with 8-bit IDs are very common

```
03:00:00.005742 IP 14.41.125.110.52013 > 192.42.93.30.53: 110+ MX? abb-es.com. (28)
03:00:00.019290 IP 14.41.125.110.52016 > 192.42.93.30.53: 111+ MX? dfaslfd.com. (29)
03:00:00.022266 IP 14.41.125.110.52020 > 192.42.93.30.53: 165+ MX? awfde.com. (27)
03:00:00.133709 IP 14.41.125.110.52053 > 192.42.93.30.53: 148+ MX? afdcd.com. (27)
03:00:00.195615 IP 14.41.125.110.52076 > 192.42.93.30.53: 141+ MX? jyhsdfne.com. (30)
03:00:00.216858 IP 14.41.125.110.52083 > 192.42.93.30.53: 101+ MX? mailnaitor.com. (32)
03:00:00.341819 IP 14.41.125.110.52115 > 192.42.93.30.53: 23+ MX? sdvckjsx.com. (30)
03:00:00.388041 IP 14.41.125.110.52130 > 192.42.93.30.53: 125+ MX? dahakjh.com. (29)
03:00:00.555571 IP 14.41.125.110.52187 > 192.42.93.30.53: 10+ MX? asfdsgf.com. (29)
03:00:00.615890 IP 14.41.125.110.52205 > 192.42.93.30.53: 176+ MX? fdsa2432354.com. (33)
03:00:00.630916 IP 14.41.125.110.52216 > 192.42.93.30.53: 56+ MX? dfhde.com. (27)
03:00:00.842527 IP 14.41.125.110.52280 > 192.42.93.30.53: 255+ MX? dwawdfa.com. (29)
03:00:00.866784 IP 14.41.125.110.52289 > 192.42.93.30.53: 242+ MX? lsdajflkasjfd.com. (35)
03:00:00.967920 IP 14.41.125.110.52319 > 192.42.93.30.53: 205+ MX? asdfkasjfhk.com. (33)
03:00:01.232021 IP 14.41.125.110.52400 > 192.42.93.30.53: 24+ MX? argerg.com. (28)
03:00:01.287871 IP 14.41.125.110.52416 > 192.42.93.30.53: 2+ MX? kahsgf.com. (28)
03:00:01.351182 IP 14.41.125.110.52434 > 192.42.93.30.53: 29+ MX? aosdfl.com. (28)
03:00:01.354212 IP 14.41.125.110.52435 > 192.42.93.30.53: 178+ MX? tbrwpyeo.com. (30)
03:00:01.446287 IP 14.41.125.110.52466 > 192.42.93.30.53: 42+ MX? aodfasd.com. (29)
03:00:01.496147 IP 14.41.125.110.52481 > 192.42.93.30.53: 206+ MX? awfde.com. (27)
```

54,914 queries from this source


poor source port randomness

poor query ID randomness

mostly MX queries

About $\frac{3}{4}$ of the sources fit this pattern

Spam Blocker?



```
20:00:10.143218 IP 220.130.119.230.20438 > 192.54.112.30.53: 54918+ A? ns.rz-ip.net. (30)
20:00:12.079813 IP 220.130.119.230.35565 > 192.54.112.30.53: 44970+ MX? aol.com. (25)
20:00:12.114577 IP 220.130.119.230.8898 > 192.54.112.30.53: 59222+ TXT? 82.135.251.60.sa-accredit.habeas.com.
20:00:12.452343 IP 220.130.119.230.60079 > 192.54.112.30.53: 39609+ A? rbldns7.sorbs.net. (35)
20:00:12.527095 IP 220.130.119.230.42007 > 192.54.112.30.53: 1276+ A? rbldns3.sorbs.net. (35)
20:00:12.657687 IP 220.130.119.230.31136 > 192.54.112.30.53: 4219+ A? psbl.primerelay.net. (37)
20:00:12.685739 IP 220.130.119.230.63657 > 192.54.112.30.53: 54006+ A? blns55.spamcop.net. (36)
20:00:12.976421 IP 220.130.119.230.52606 > 192.54.112.30.53: 1780+ A? ltms2.rpdns.net. (33)
20:00:12.981781 IP 220.130.119.230.56831 > 192.54.112.30.53: 19812+ TXT? 96.240.231.61.bl.spamcop.net. (46)
20:00:13.042276 IP 220.130.119.230.28790 > 192.54.112.30.53: 58166+ A? rbldns8.sorbs.net. (35)
20:00:13.208830 IP 220.130.119.230.64167 > 192.54.112.30.53: 61461+ A? cmtu.mt.ns.els-gms.att.net. (44)
20:00:13.390708 IP 220.130.119.230.11808 > 192.54.112.30.53: 24332+ A? 137.227.55.202.dnsbl.sorbs.net. (48)
20:00:13.500984 IP 220.130.119.230.29280 > 192.54.112.30.53: 48006+ A? blns69.spamcop.net. (36)
20:00:13.642464 IP 220.130.119.230.50186 > 192.54.112.30.53: 9826+ A? ns4.apnic.com. (31)
20:00:13.835682 IP 220.130.119.230.56969 > 192.54.112.30.53: 37367+ A? blns65.spamcop.net. (36)
20:00:13.898614 IP 220.130.119.230.50892 > 192.54.112.30.53: 40955+ MX? ms9.hinet.net. (31)
20:00:13.952499 IP 220.130.119.230.18620 > 192.54.112.30.53: 21670+ A? auth3.ns.gin.ntt.net. (38)
20:00:14.978141 IP 220.130.119.230.17202 > 192.54.112.30.53: 52089+ A? 190.232.151.118.psbl.surriel.com. (50)
```

The lower query rate and numerous references to RBLs makes us think this is a spam-blocking application (with custom DNS resolution).

230.119.130.220.in-addr.arpa domain name pointer mx17.pchome.com.tw.

CNAME looker-uppers?

```

20:00:14.223659 IP 119.252.144.128.25969 > 192.41.162.30.53: 16918+ CNAME? gmail.com. (27)
20:00:15.606876 IP 119.252.144.128.27757 > 192.41.162.30.53: 8764+ CNAME? gmail.com. (27)
20:00:16.045133 IP 119.252.144.128.16218 > 192.41.162.30.53: 9440+ CNAME? vsnl.net. (26)
20:00:16.942806 IP 119.252.144.128.15833 > 192.41.162.30.53: 54415+ CNAME? asp.net. (25)
20:00:16.971148 IP 119.252.144.128.4652 > 192.41.162.30.53: 20801+ CNAME? signet-india.com. (34)
20:00:18.942625 IP 119.252.144.128.60833 > 192.41.162.30.53: 36385+ CNAME? hotmail.com. (29)
20:00:32.887576 IP 119.252.144.128.61287 > 192.41.162.30.53: 31603+ CNAME? yahoo.com. (27)
20:00:45.315219 IP 119.252.144.128.35529 > 192.41.162.30.53: 30117+ CNAME? tbwt.com. (26)
20:01:01.646496 IP 119.252.144.128.13418 > 192.41.162.30.53: 59616+ A? mm9india-com.mail.eo.outlook.com. (50)
20:01:02.795366 IP 119.252.144.128.30100 > 192.41.162.30.53: 18016+ CNAME? yahoo.com. (27)
20:01:02.895947 IP 119.252.144.128.47887 > 192.41.162.30.53: 54328+ CNAME? yahoo.com. (27)
20:01:04.836620 IP 119.252.144.128.61499 > 192.41.162.30.53: 14047+ A? mx.connectivity.com. (37)
20:01:06.046199 IP 119.252.144.128.43349 > 192.41.162.30.53: 11379+ MX? gl00on.net. (28)
20:01:06.447223 IP 119.252.144.128.60663 > 192.41.162.30.53: 5183+ CNAME? gmail.com. (27)
20:01:09.252072 IP 119.252.144.128.60082 > 192.41.162.30.53: 16551+ A? mmin4-b.corp.bfl.yahoo.com. (44)
20:01:09.700854 IP 119.252.144.128.50417 > 192.41.162.30.53: 35704+ CNAME? gmail.com. (27)
20:01:09.772103 IP 119.252.144.128.26468 > 192.41.162.30.53: 40784+ CNAME? hotmail.com. (29)
20:01:14.748022 IP 119.252.144.128.9391 > 192.41.162.30.53: 31236+ CNAME? gmail.com. (27)
20:01:14.827309 IP 119.252.144.128.59594 > 192.41.162.30.53: 52160+ CNAME? yahoo.com. (27)
20:01:15.263412 IP 119.252.144.128.5200 > 192.41.162.30.53: 12721+ CNAME? retailmart.web10.eperiumindia.com.
20:01:18.885109 IP 119.252.144.128.59223 > 192.41.162.30.53: 43757+ CNAME? fascom.com. (28)
20:01:27.325840 IP 119.252.144.128.44852 > 192.41.162.30.53: 30563+ CNAME? gmail.com. (27)
20:01:55.949103 IP 119.252.144.128.20997 > 192.41.162.30.53: 6081+ CNAME? aakronrule.com. (3)

```

128.144.252.119.in-addr.arpa domain name pointer host-119-252-144-128.rediffdns.com.

rediff = “India’s leading portal which covers India news, Hindi Movies, Photos ...”



Trolling for registrations?

23:59:24.167869 IP 86.97.17.116.61693 > 192.41.162.34.53: 61060+ NS? mzmslcbpcbpz.cc. (33)
23:59:27.003021 IP 86.97.17.116.61708 > 192.41.162.34.53: 61068+ NS? smmyuhxlt.cc. (30)
23:59:32.939380 IP 86.97.17.116.61744 > 192.41.162.34.53: 6044+ NS? sfhbkoesbf.tv. (31)
23:59:34.052852 IP 86.97.17.116.61750 > 192.41.162.30.53: 6047+ NS? garrtchx.net. (30)
23:59:35.091427 IP 86.97.17.116.61755 > 192.41.162.30.53: 61090+ NS? garrtchx.net. (30)
23:59:36.878694 IP 86.97.17.116.61765 > 192.41.162.34.53: 61095+ NS? cwnizwzd.tv. (29)
23:59:40.744679 IP 86.97.17.116.61785 > 192.41.162.34.53: 6064+ NS? zoipmnwr.tv. (29)
23:59:49.153031 IP 86.97.17.116.61832 > 192.41.162.34.53: 61128+ NS? xtnouxrcav.cc. (31)
23:59:49.870170 IP 86.97.17.116.61836 > 192.41.162.30.53: 61130+ NS? fvzkoky.net. (29)
23:59:50.256616 IP 86.97.17.116.61838 > 192.41.162.34.53: 61131+ NS? mcpzonsstpqx.tv. (33)
23:59:51.968548 IP 86.97.17.116.61847 > 192.41.162.34.53: 6096+ NS? gnjbyaoolbc.tv. (32)
23:59:53.521128 IP 86.97.17.116.61855 > 192.41.162.34.53: 6100+ NS? wkylsjrw.tv. (29)
23:59:54.080899 IP 86.97.17.116.61858 > 192.41.162.30.53: 61141+ NS? qeezwogo.com. (30)
23:59:56.892956 IP 86.97.17.116.61869 > 192.41.162.30.53: 6107+ NS? wwgdswgnnek.net. (32)
23:59:59.684794 IP 86.97.17.116.61885 > 192.41.162.34.53: 61155+ NS? egbmbdey.tv. (29)

Observations

- *Number of Clients:* We see queries from about 2% of all IPv4 addresses over the course of a year
 - Probably higher if we analyzed more sites
- *Query Type Distribution:* Most sources send an expected mix of A and MX queries
- *AAAA Queriers:* IPv6 queries are significant part of the workload now
- *Characterizing Top Talkers:* “Big resolvers” each contribute 3-5% to the total query load
- *RD=1 Queries:* We receive a *lot* of RD=1 queries, with a majority of the sources sending almost exclusively MX queries



Questions?



VERISIGN™

© 2013 VeriSign, Inc. All rights reserved. VERISIGN and other trademarks, service marks, and designs are registered or unregistered trademarks of VeriSign, Inc. and its subsidiaries in the United States and in foreign countries. All other trademarks are property of their respective owners.