What's In the Cache?

Ralf Weber, Principal Architect, Special Projects

nominum

Introduction

About me:

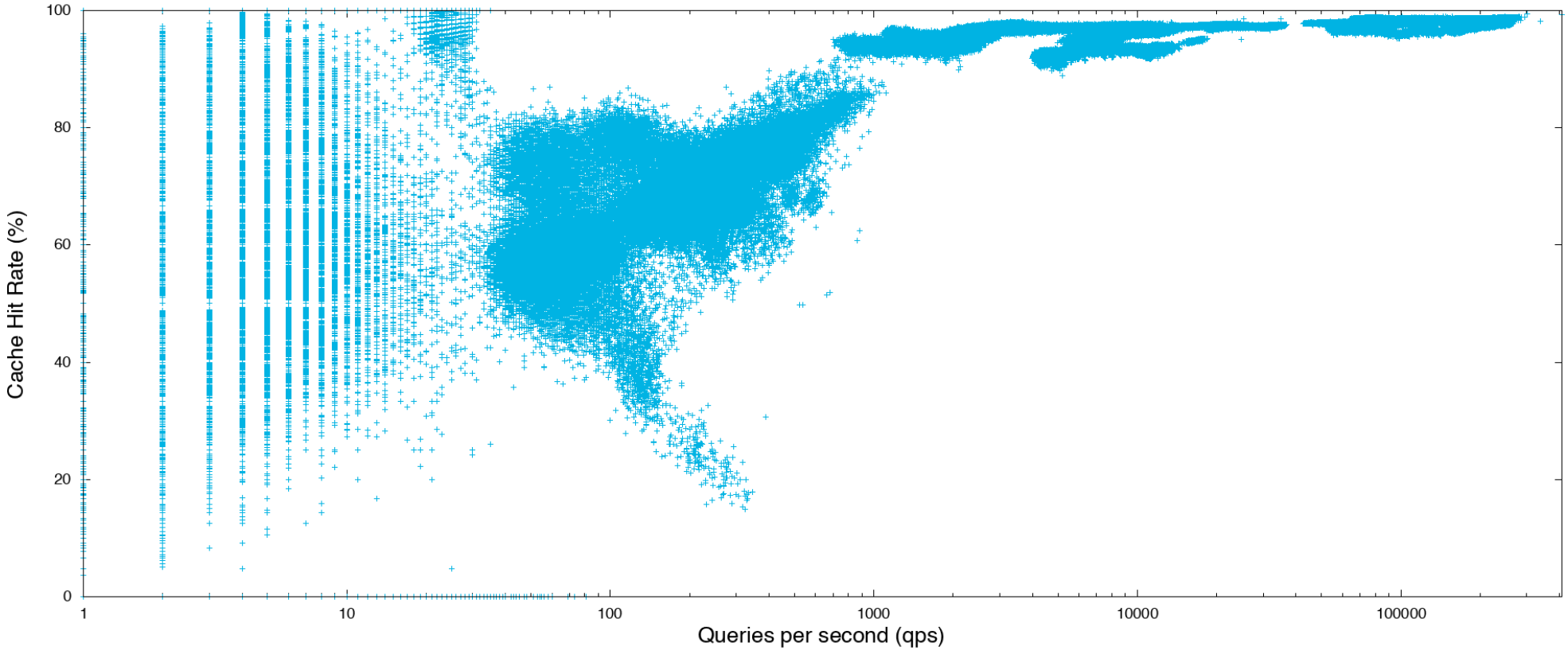
- Old grey bearded Unix sysadmin
- Ran DNS Caching servers for 15 years
  - Kevag Telekom
  - Colt Telecom
- Now help people deploy Caching servers
  - Nominum
- Obsessed with Cache Hits ;-)

About my work (Nominum)

- We do awesome DNS engines
- We also offer DNS policy enhanced applications
  - Security
  - Parental Controls
- To achieve success in the above we analyze DNS data
  - A subset of our customers DNS data
  - Anonymized client resolver data
  - Statistical data of the server (Telemetry)
  - All statistics of the server are emitted evey 5 seconds
  - Includes QPS and Cache Hit Rate

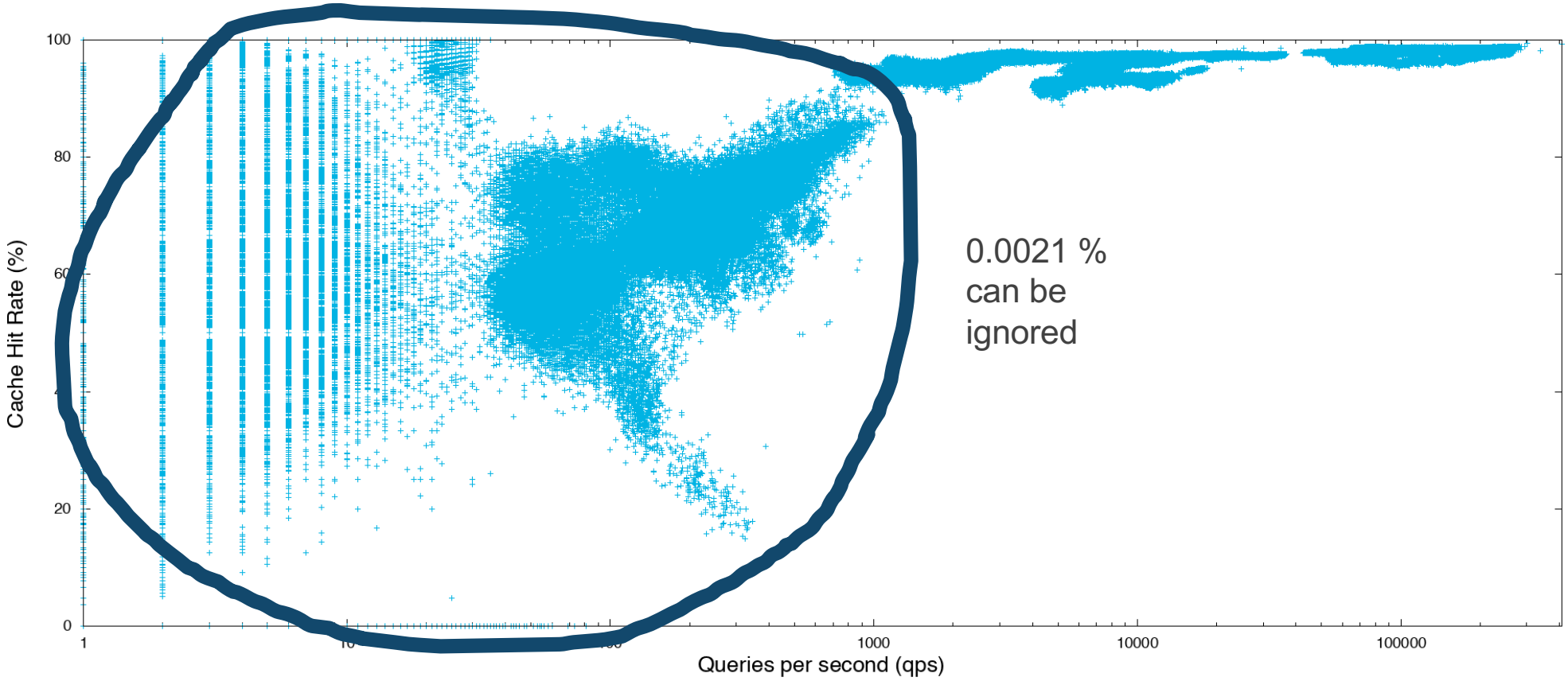
nominum

Cache Hit Rate in comparison to qps



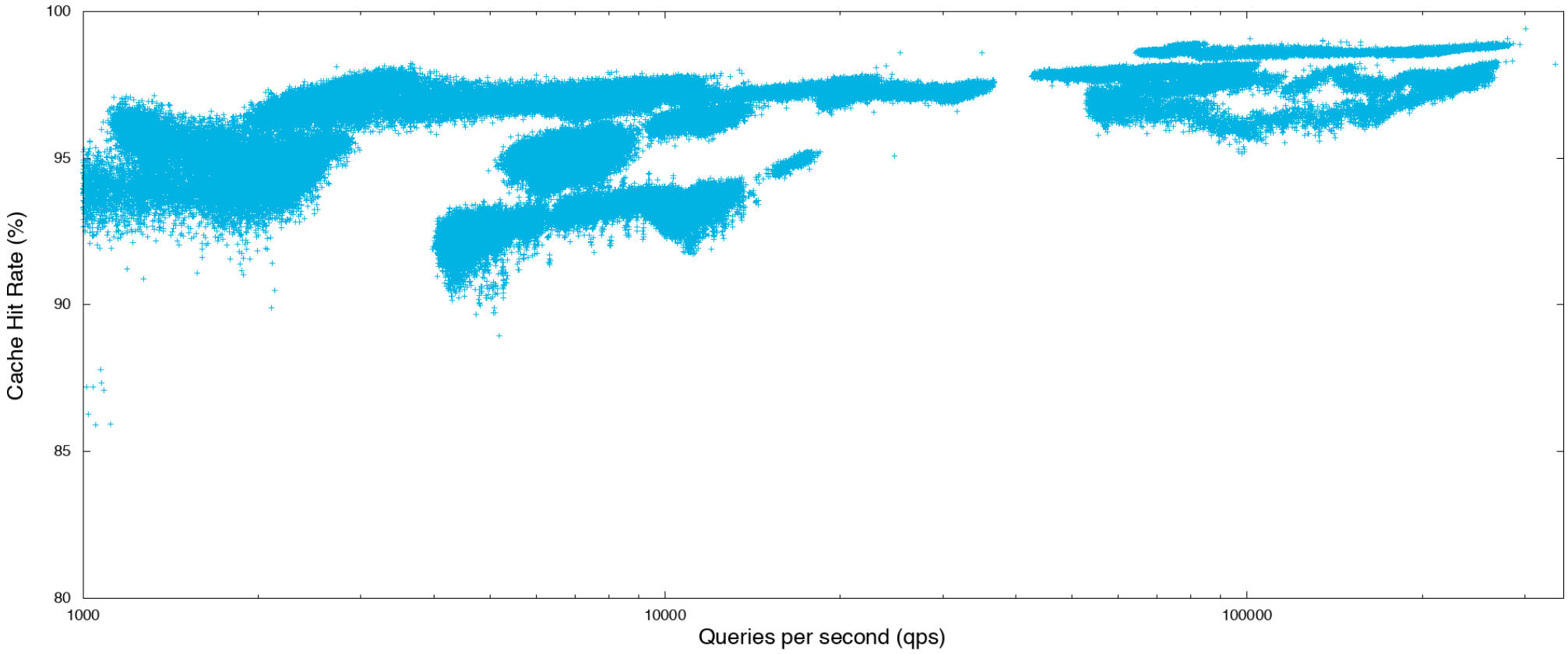
nominum

Cache Hit Rate in comparison to qps



nominum

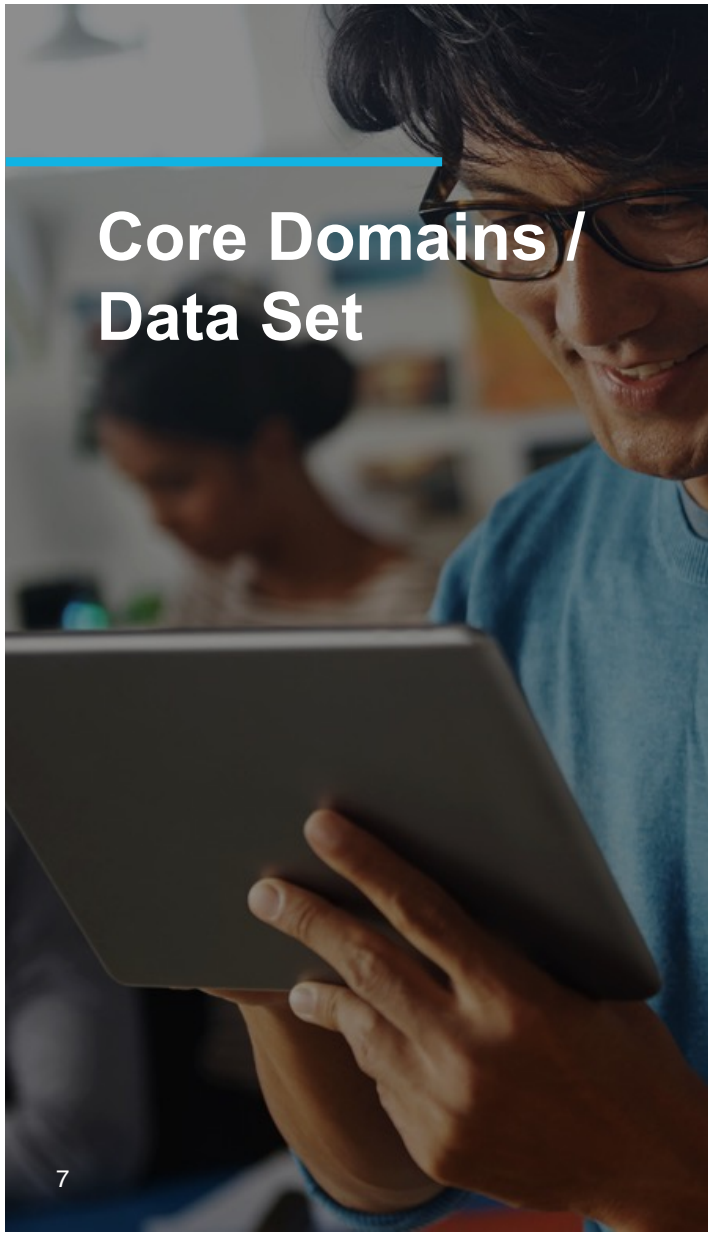
Zoom into Cache Hit Rate vs. QPS



nominum

Cache Hit Rate summarised

- Cache Hit rate is over 90% for normal servers
- Servers that serve <1000 qps are mostly lab servers and total traffic is negligible
- Loaded servers >40k qps have above 95% Cache Hit Rate

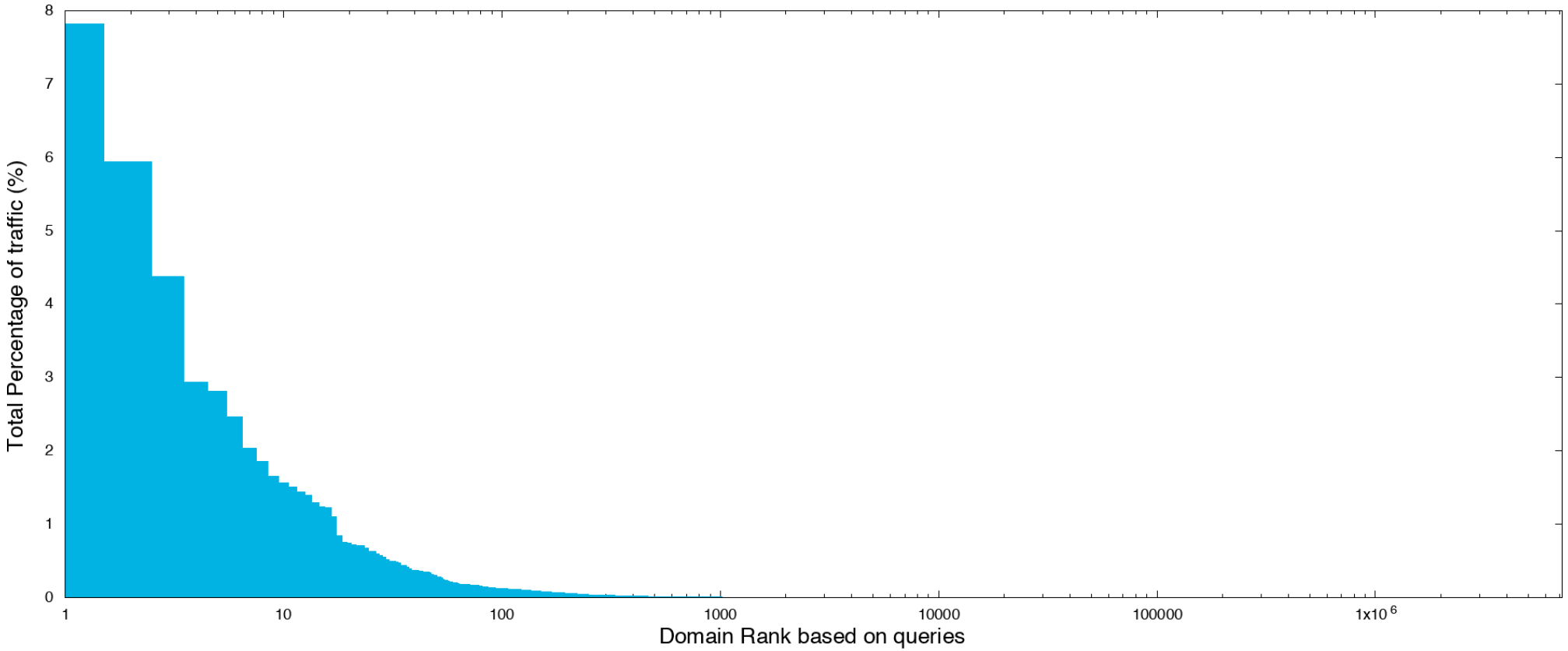


Core Domains / Data Set

- What is an core domain
  - Usually a second level domain
  - In some cases (co.uk) deeper
  - Delegated domains from Registries
  - Based on Mozilla Public Suffix List
- Why are we using it
  - Looking for a meaningful domain boundaries
  - Usually owned by one organization, though one organization can have multiple core domains
  - Easy to digest in the server (NS lookups e.g would be harder for xx millions of domains)
- Data set
  - One day of DNS data from customers around the world
  - Only select servers from customers mostly medium qps
  - 100 billion queries
  - 7 million core domains
  - Get used to logarithmic scales
  - Ratio A to AAAA = 2/3 to 1/3

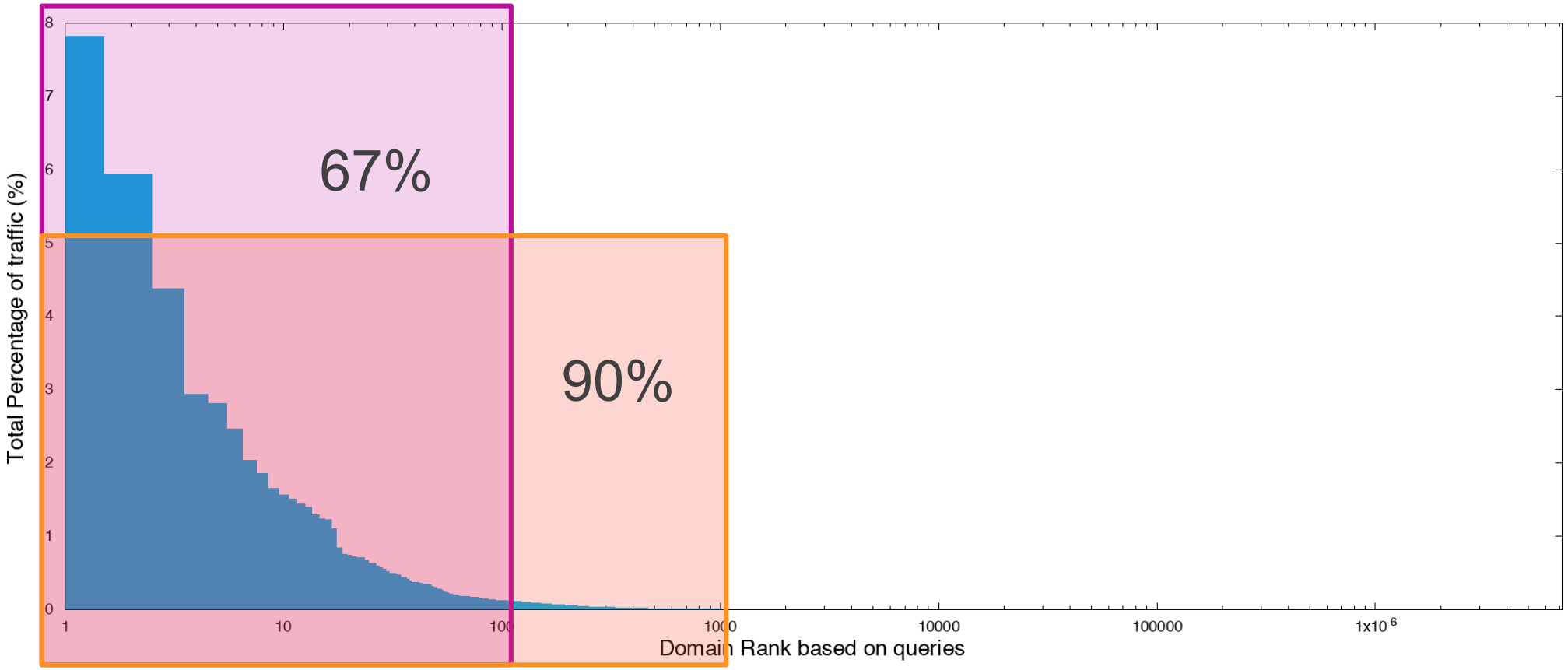
nominum

Core domain distribution



nominum

Core domain distribution



nominum

Core Domains

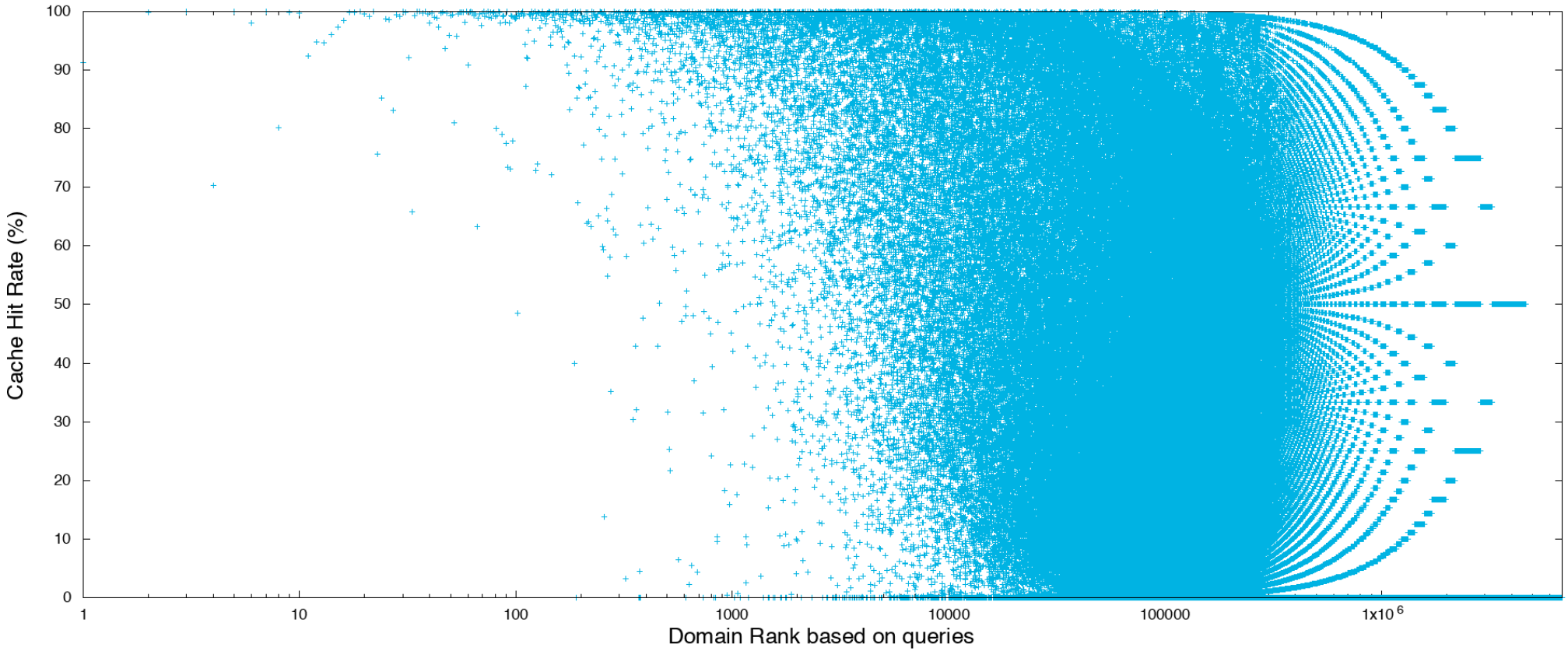
Domain	%	CHR%
apple.com	7.8	91.2
google.com	5.9	99.8
facebook.com	4.4	99.9
akadns.net	2.9	70.4
googleapis.com	2.8	99.9
akamaiedge.net	2.5	98.0
doubleclick.net	2.0	99.9
microsoft.com	1.9	80.1
whatsapp.net	1.7	99.9
netflix.com	1.6	99.8

- Top 100 core domains make 60% of traffic
- Top 1000 core domains make 90%
- Average hit rate 90% in top 1000
- 6 million core domains have less then 10 queries per day

- Lets look at the Cache Hit Rate distribution

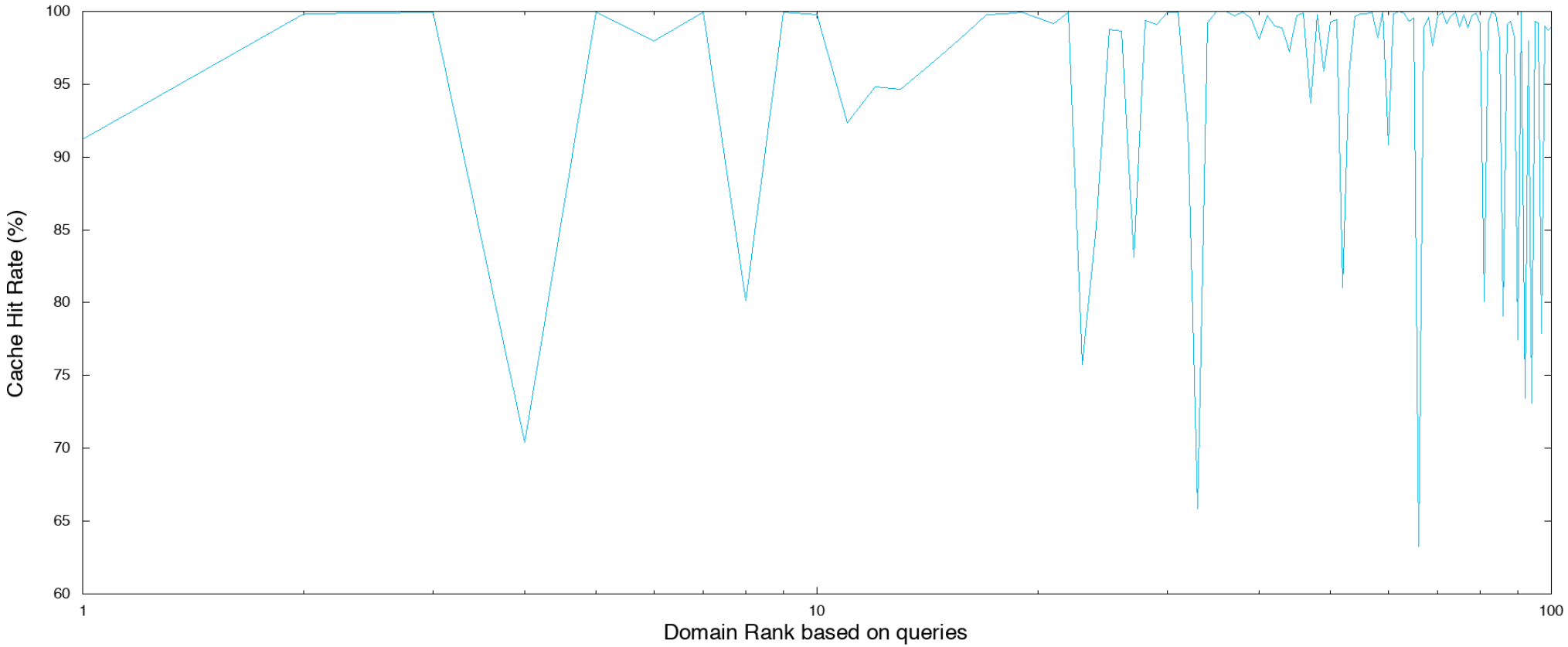
nominum

Cache Hit Rate Distribution



nominum

Zoom into Cache Hit Rate distribution



nominum

Cache Hit Rate overall

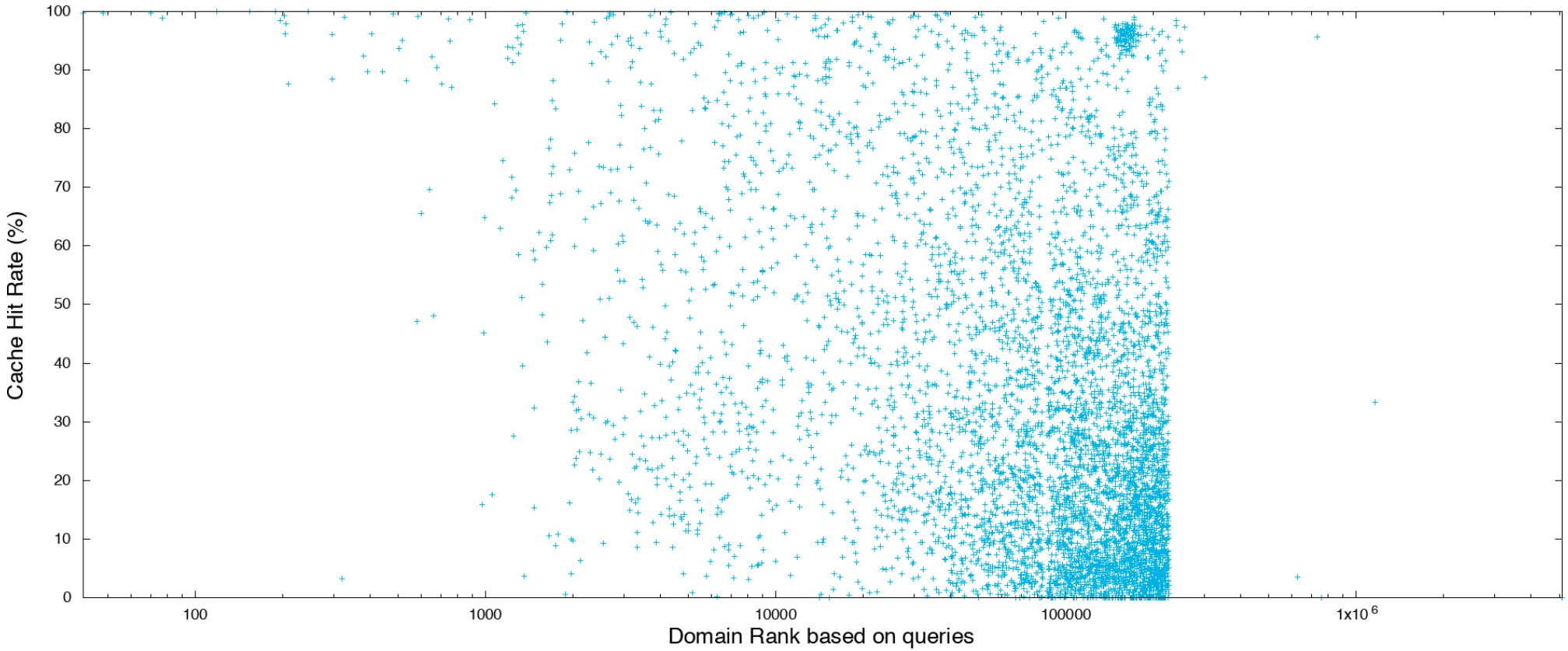
Rank	Domain	%	CHR
4	akadns.net	2.9	70.4
8	microsoft.com	1.9	80.1
23	adnxs.com	0.7	75.7
27	cloudfront.net	0.6	65.8
33	live.com	0.5	83.0
80	windows.com	0.2	80.0
86	dyndns.org	0.1	79.1
92	akamaihd.net	0.1	73.0

- Most populare domains have a good cache hit rate (>95%)
- Some cache hit rate seem low
  - More on that later
- Overall averagae Cache Hit Rate:
  - 91.46

- So lets look into DNSSEC

nominum

DNSSEC Cache Hit Rate



nominum

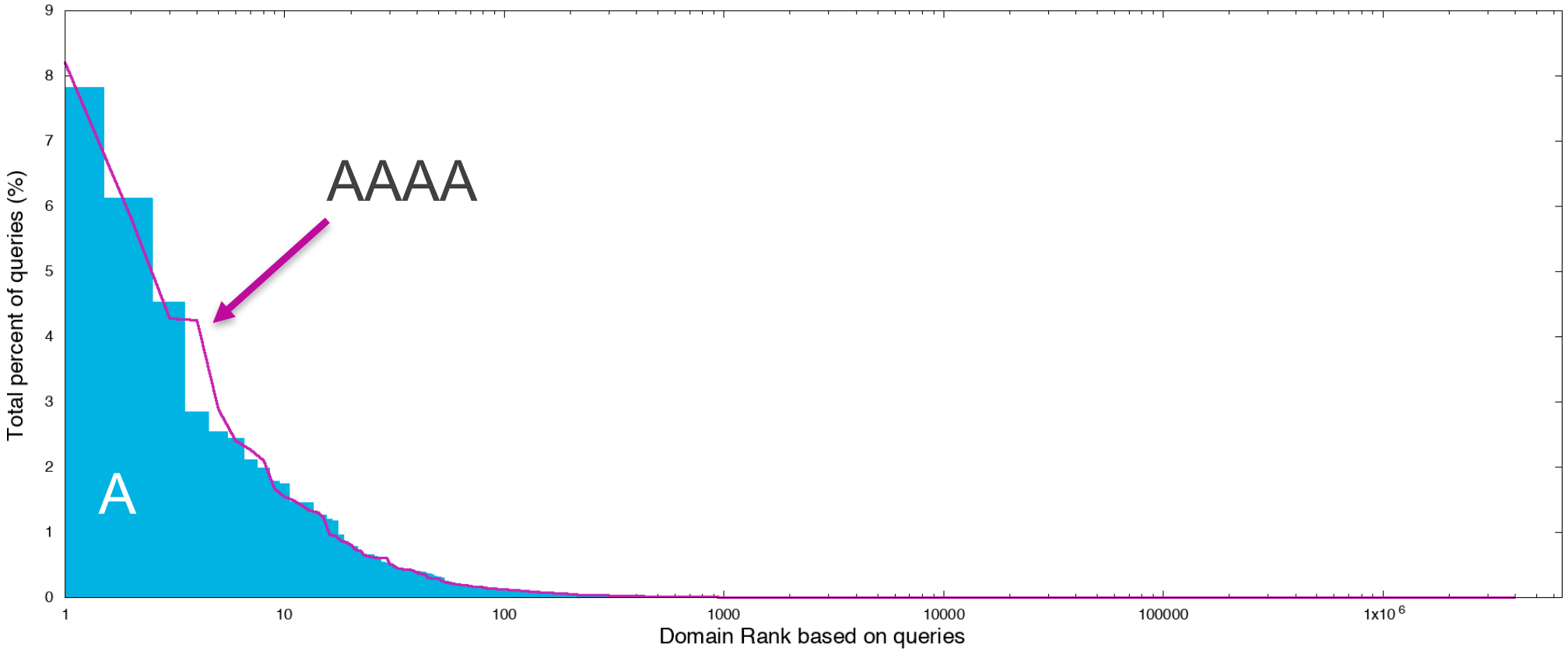
DNSSEC

#	Domain	%	CHR
41	vkcache.com	0.36	99.7
48	nist.gov	0.32	99.8
70	vmware.com	0.18	98.9
77	gmx.net	0.16	98.9
118	cpsc.gov	0.11	99.9
154	ioam.de	0.08	99.9
188	pki.goog	0.06	99.9
202	cloudflare.com	0.06	99.2
203	mozilla.org	0.06	96.1

- Not much DNSSEC domains
  - Total 5746 (0.0%)
  - Total traffic validatiable is 2.5 %
- Cache Hit Rate
  - Very good at popular domains
  - Long tail falls off
  - Average is 33%
- Lets look at IPv6
  - No real user traffic on v6
  - But A to AAAA distribution

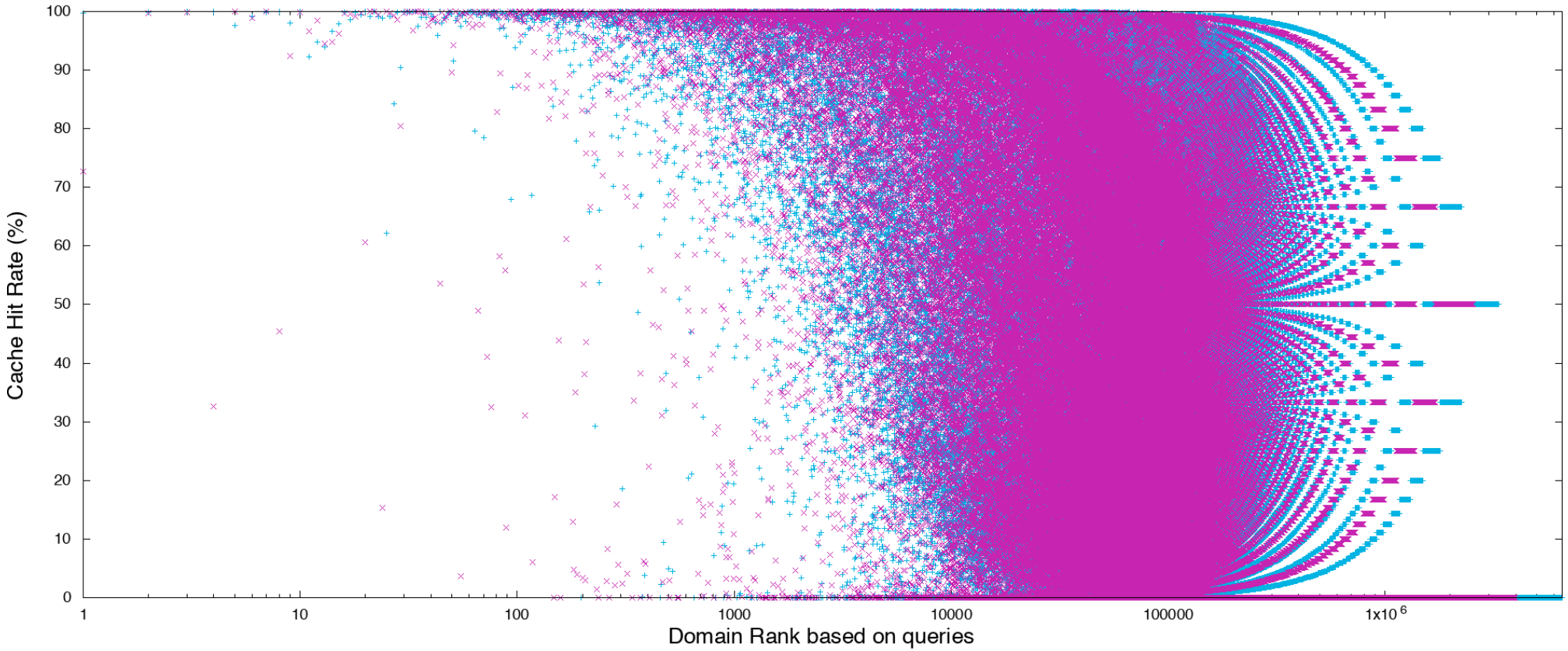
nominum

Core Domain Distribution A vs AAAA



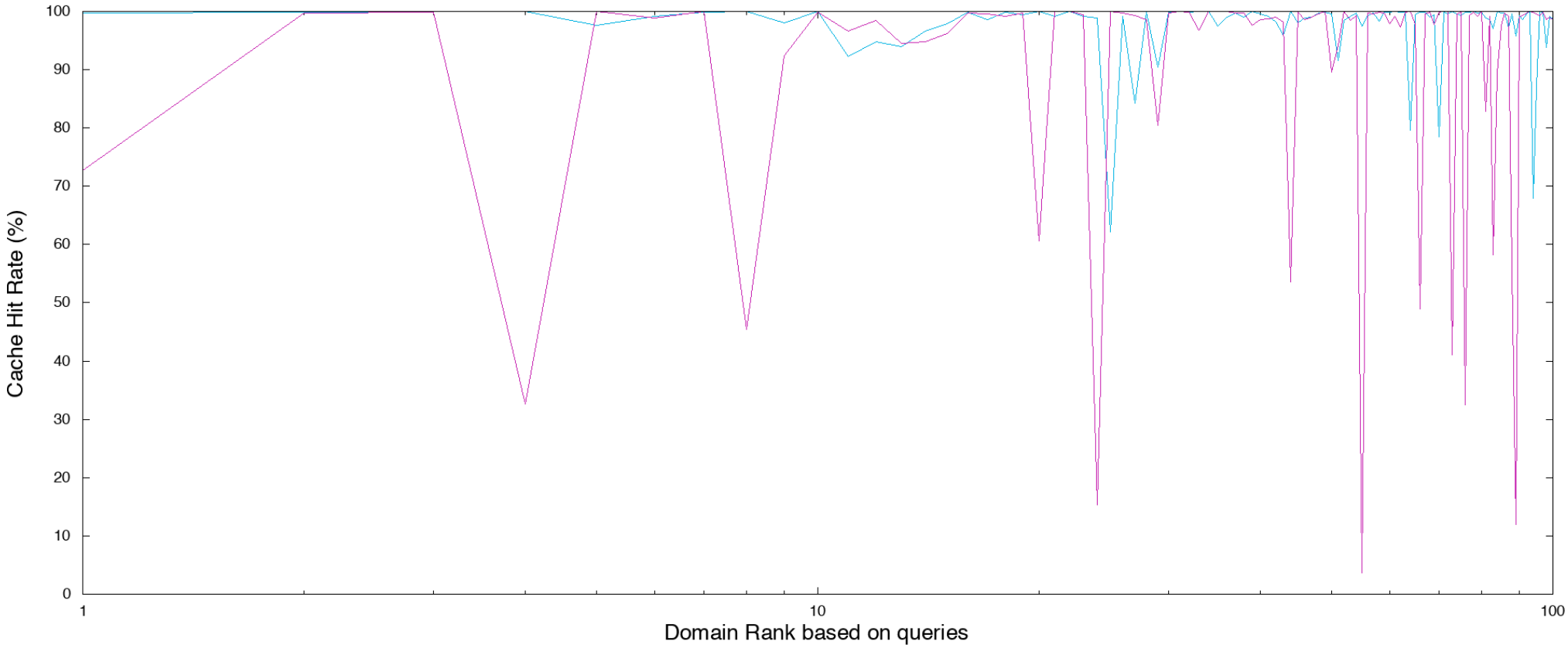
nominum

Cache Hit Rate A vs AAA



nominum

Zoom in Cache Hit Rate A vs AAAA



nominum

Core Domains A vs AAAA

Domain (A)	%	CHR (%)
apple.com	7.8	99.7
google.com	6.1	99.8
facebook.com	4.5	99.9
googleapis.com	2.9	99.9
akamaiedge.net	1.7	97.6
akadns.net	1.6	99.2
netflix.com	1.4	99.8
doubleclick.net	1.4	99.9
microsoft.com	1.2	98.0
whatsapp.net	1.2	99.9

Domain (AAAA)	%	CHR (%)
apple.com	8.2	72.7
google.com	5.8	99.7
facebook.com	4.3	99.9
akadns.net	4.2	32.6
googleapis.com	2.9	99.9
akamaiedge.net	2.4	98.8
doubleclick.net	2.3	99.9
microsoft.com	2.1	45.4
amazonaws.com	1.7	92.4
whatsapp.net	1.5	99.8

IPv6 (AAAA) Cache Hit Rate

#	Domain	%	CHR
1	apple.com	8.2	72.7
4	akadns.net	4.3	32.6
8	microsoft.com	2.1	45.2
24	live.com	0.6	15.4
44	skype.com	0.3	53.6
66	windows.com	0.2	49.0
89	mcafee.com	0.1	12.0

- Some domains especially in the top domains have absurdly low cache hit rate
- In general the IPv6 rates in the long tail are a bit lower, but nothing to what we see in the top
- Overall hit rate comparison
  - Top 100 A=97 AAAA=91
  - Top 1000 A=91 AAAA=87
- Overall traffic rate is 2/3 A and 1/3 AAAA
- So is IPv6 hindered by records expiring out of the cache
- Lets look into the data….

nominum

Normal query

```
; <<>> DiG 9.11.0-P3 <<>> +norec +nocookie dns-oarc.net @ns.dns-oarc.net. A
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 24774
;; flags: qr aa; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;dns-oarc.net. IN A

;; ANSWER SECTION:
dns-oarc.net. 3600 IN A 64.191.0.198

;; Query time: 5 msec
;; SERVER: 2620:ff:c000:0:1::65#53(2620:ff:c000:0:1::65)
;; WHEN: Fri Sep 29 16:38:37 PDT 2017
;; MSG SIZE  rcvd: 57
```

```
; <<>> DiG 9.11.0-P3 <<>> +norec +nocookie dns-oarc.net
@ns.dns-oarc.net. AAAA
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 15590
;; flags: qr aa; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL:
1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;dns-oarc.net. IN AAAA

;; ANSWER SECTION:
dns-oarc.net. 3600 IN AAAA 2620:ff:c000::198

;; Query time: 7 msec
;; SERVER: 2620:ff:c000:0:1::65#53(2620:ff:c000:0:1::65)
;; WHEN: Fri Sep 29 16:50:32 PDT 2017
;; MSG SIZE  rcvd: 69
```


nominum

No AAAA exist query

```
; <<>> DiG 9.11.0-P3 <<>> slashdot.org A
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 3584
;; flags: qr rd ra; QUERY: 1, ANSWER: 1, AUTHORITY: 5, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;slashdot.org. IN A

;; ANSWER SECTION:
slashdot.org. 300 IN A 216.34.181.45

;; AUTHORITY SECTION:
slashdot.org. 86400 IN NS ns3.dnsmadeeasy.com.
slashdot.org. 86400 IN NS ns4.dnsmadeeasy.com.
slashdot.org. 86400 IN NS ns0.dnsmadeeasy.com.
slashdot.org. 86400 IN NS ns1.dnsmadeeasy.com.
slashdot.org. 86400 IN NS ns2.dnsmadeeasy.com.

;; Query time: 21 msec
;; SERVER: 2620:0:ce0:100::12#53(2620:0:ce0:100::12)
;; WHEN: Fri Sep 29 16:37:06 PDT 2017
;; MSG SIZE  rcvd: 162
```

```
; <<>> DiG 9.11.0-P3 <<>> slashdot.org AAAA +multiline
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 334
;; flags: qr rd ra; QUERY: 1, ANSWER: 0, AUTHORITY: 1,
ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;slashdot.org. IN AAAA

;; AUTHORITY SECTION:
slashdot.org. 300 IN SOA ns0.dnsmadeeasy.com.
hostmaster.slashdotmedia.com. (
2016045443 ; serial
14400      ; refresh (4 hours)
600        ; retry (10 minutes)
604800     ; expire (1 week)
300        ; minimum (5 minutes)
)

;; Query time: 25 msec
;; SERVER: 2620:0:ce0:100::12#53(2620:0:ce0:100::12)
;; WHEN: Fri Sep 29 17:02:31 PDT 2017
;; MSG SIZE  rcvd: 121
```


nominum

RFC2308

5 - Caching Negative Answers

Like normal answers negative answers have a time to live (TTL). As there is no record in the answer section to which this TTL can be applied, the TTL must be carried by another method. This is done by including the SOA record from the zone in the authority section of the reply. When the authoritative server creates this record its TTL is taken from the minimum of the SOA.MINIMUM field and SOA's TTL. This TTL decrements in a similar manner to a normal cached answer and upon reaching zero (0) indicates the cached negative answer MUST NOT be used again.

Akadns

```
; <<>> DiG 9.11.0-P3 <<>> gsp64-ssl.ls-apple.com.akadns.net.
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 26202
;; flags: qr rd ra; QUERY: 1, ANSWER: 2, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;gsp64-ssl.ls-apple.com.akadns.net. IN A

;; ANSWER SECTION:
gsp64-ssl.ls-apple.com.akadns.net. 234 IN A 17.142.169.200
gsp64-ssl.ls-apple.com.akadns.net. 234 IN A 17.142.169.199

;; Query time: 25 msec
;; SERVER: 2620:0:ce0:100::12#53(2620:0:ce0:100::12)
;; WHEN: Fri Sep 29 17:08:03 PDT 2017
;; MSG SIZE  rcvd: 94
```

```
; <<>> DiG 9.11.0-P3 <<>> gsp64-ssl.ls-apple.com.akadns.net.
AAAA
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 30864
;; flags: qr rd ra; QUERY: 1, ANSWER: 0, AUTHORITY: 0,
ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 4096
;; QUESTION SECTION:
;gsp64-ssl.ls-apple.com.akadns.net. IN AAAA

;; Query time: 42 msec
;; SERVER: 2620:0:ce0:100::12#53(2620:0:ce0:100::12)
;; WHEN: Fri Sep 29 17:08:23 PDT 2017
;; MSG SIZE  rcvd: 62
```


nominum

RFC2308

Negative responses without SOA records SHOULD NOT be cached as there is no way to prevent the negative responses looping forever between a pair of servers even with a short TTL.

Live.com

```
; <<>> DiG 9.11.0-P3 <<>> @134.170.104.152 cid-24e99ee8e7b6bdf0.users.storage.live.com
; (1 server found)
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 32191
;; flags: qr aa; QUERY: 1, ANSWER: 1, AUTHORITY: 0, ADDITIONAL: 1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 512
;; QUESTION SECTION:
;cid-24e99ee8e7b6bdf0.users.storage.live.com. IN A

;; ANSWER SECTION:
cid-24e99ee8e7b6bdf0.users.storage.live.com. 600 IN A 157.55.109.224

;; Query time: 54 msec
;; SERVER: 134.170.104.152#53(134.170.104.152)
;; WHEN: Fri Sep 29 17:27:38 PDT 2017
;; MSG SIZE  rcvd: 88
```

```
; <<>> DiG 9.11.0-P3 <<>> @134.170.104.152 cid-
24e99ee8e7b6bdf0.users.storage.live.com AAAA +multiline
; (1 server found)
;; global options: +cmd
;; Got answer:
;; ->>HEADER<<- opcode: QUERY, status: NOERROR, id: 5415
;; flags: qr aa; QUERY: 1, ANSWER: 0, AUTHORITY: 1, ADDITIONAL:
1

;; OPT PSEUDOSECTION:
; EDNS: version: 0, flags:; udp: 512
;; QUESTION SECTION:
;cid-24e99ee8e7b6bdf0.users.storage.live.com. IN AAAA

;; AUTHORITY SECTION:
storage.live.com. 0 IN SOA geodns.storage.skyprod.akadns.net.
msnhst.microsoft.com. (
2007010101 ; serial
43200      ; refresh (12 hours)
21600      ; retry (6 hours)
86400      ; expire (1 day)
0          ; minimum (0 seconds)
)

;; Query time: 58 msec
;; SERVER: 134.170.104.152#53(134.170.104.152)
;; WHEN: Fri Sep 29 17:29:52 PDT 2017
;; MSG SIZE  rcvd: 158
```


nominum

- This is clearly bad behaviour
  - A lot of domains don't have AAAA records yet
  - I doubt that it does matter to bring them online in seconds when they get them
  - Defeats the purpose of negative caching
- However my analysis was based on client query data without answers, so can't quantify how bad
- Unfortunately this data only exists in the recursor to authoritative data which is hard to correlate with the client data
- So lets dig into that data without correlation
  - V4 = queries that asked for A
  - V6 = queries that asked for AAAA
  - V4no = NoAnswer/NoError for A
  - V6no = NoAnswer/NoError for AAAA
  - V4nc = Non cacheable NoAnswer/NoError for A
  - V6nc = Non cacheable NoAnswer/NoError for AAAA

nominum

Trying to quantify the problem

Domain	CHRv6	v4	v6	v4no	v4nc	v6no	v6nc
apple.com	72.7	67.0	32.4	7.9	1.1	53.6	0.8
akadns.net	32.6	5.5	94.5	0.0	0.3	0.0	97.8
microsoft.com	45.2	20.5	79.3	9.9	0.8	2.0	95.2
live.com	15.4	62.7	37.2	0.2	0.7	0.2	95.9
skype.com	53.6	73.9	21.6	7.0	0.5	5.0	0.1
windows.com	49.0	59.1	40.7	0.1	0.3	98.2	0.2
mcafee.com	12.0	46.7	51.7	70.1	5.5	2.3	97.2

Conclusion

- Caching works
  - Even in IPv6 stuff is served from Cache
  - Unless authorities decide not to cache
- Rather then work on protocol extensions for less then 10% (for getting more than one record at a time) we should:
- Work on DNSSEC deployment (get more domains signed)
- Work on getting current standards deployed as intended
  - Negative Caching
  - Empty non terminals
  - EDNS0

- Questions???

nominum