DNS over TCP as seen from the authoritative servers

DNS-OARC 26 Workshop
May 14, 2017
Madrid

NS1.

Jan Včelák · Software Engineer

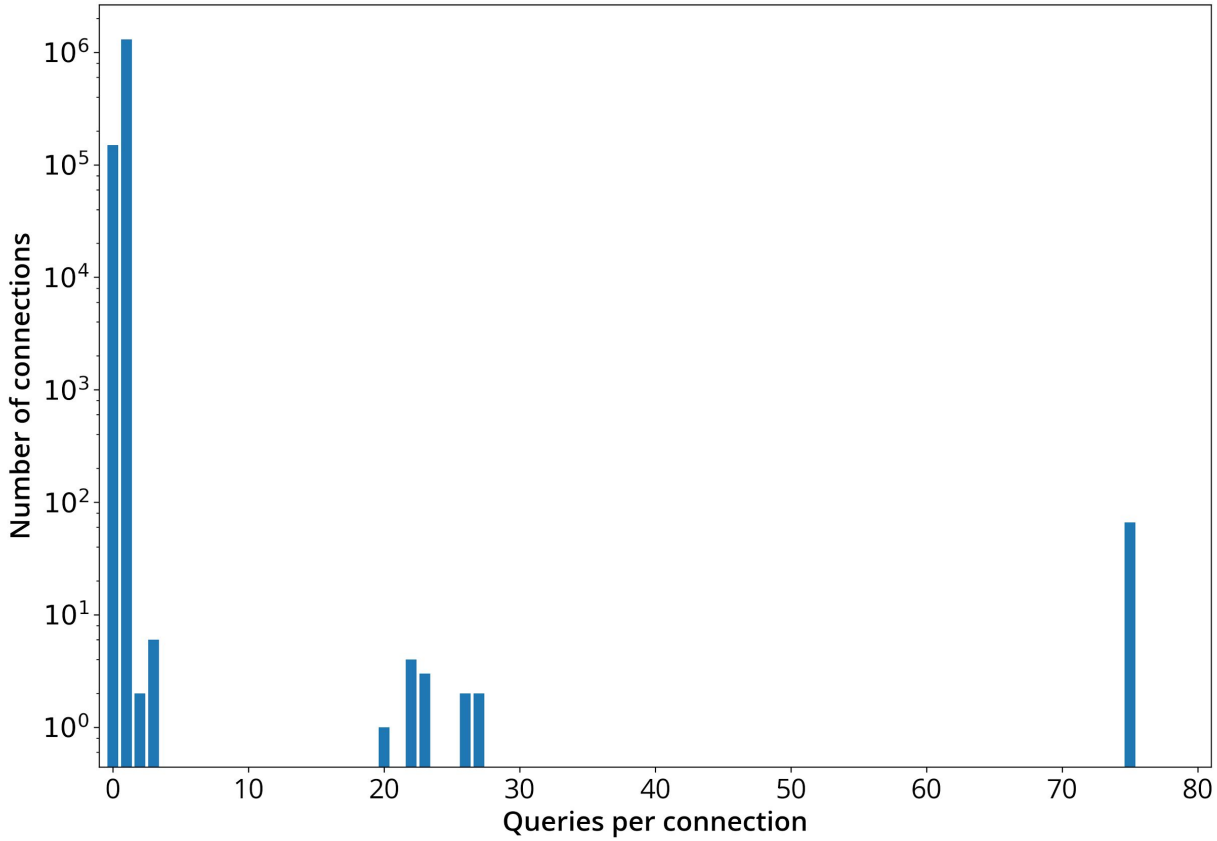
JVCELAK@NS1.COM | @FCELDA @NSONEINC

Motivation to investigate DNS over TCP

- Re-architecturing part of our DNS platform
- No deep understanding how resolvers behave if TCP is used
- Planning ahead (DNS over TCP, TLS, HTTP, etc.)
- Desire to understand benefits of TCP during attacks on DNS
- We have data, so why not to take a look?

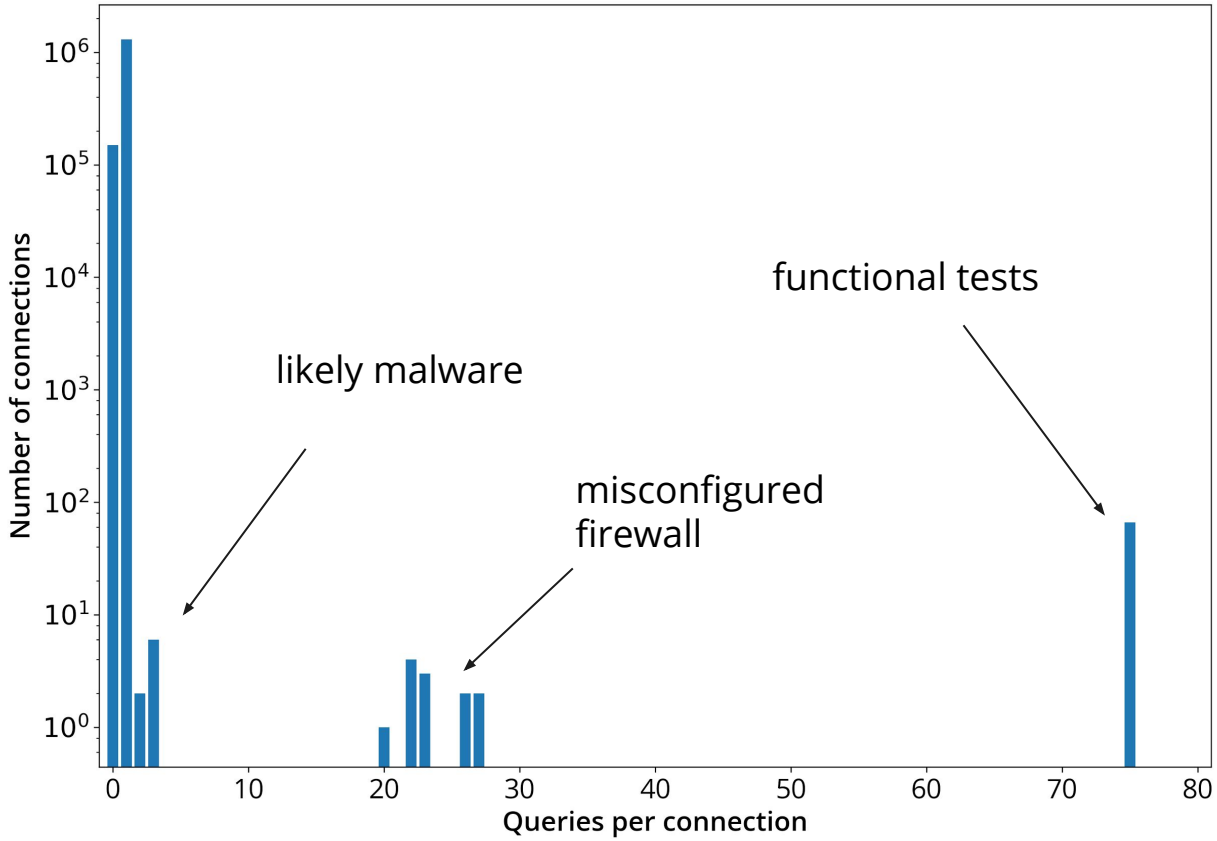
First peek at the data

- Sample of TCP queries from a few servers in our managed network
- Do the servers reuse TCP connections?
- “Looks good, let’s make a talk!”



First peek at the data

- Sample of TCP queries from a few servers in our managed network
- Do the servers reuse TCP connections?
- “Looks good, let’s make a talk!”



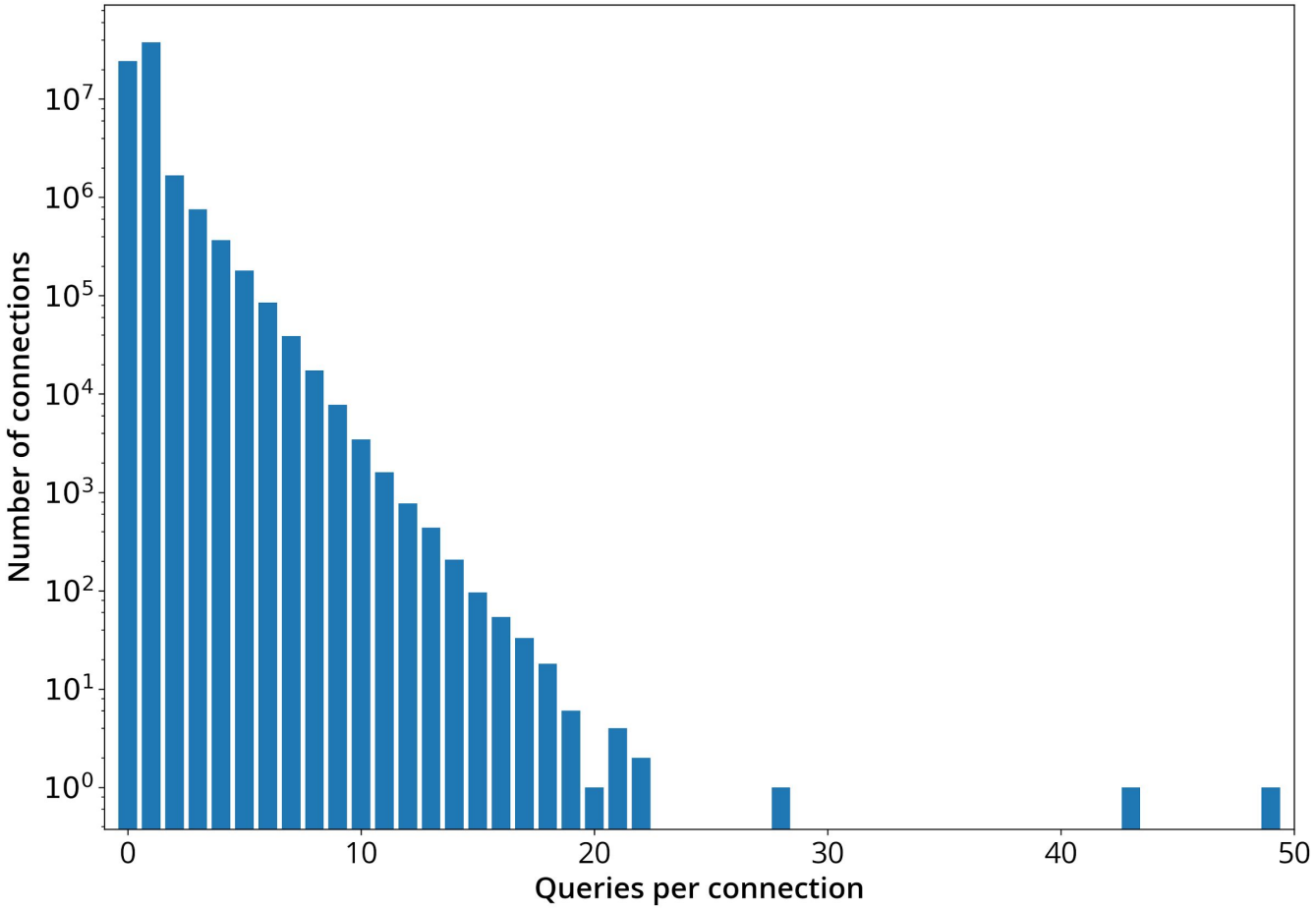
Data source for my research

- DITL (Day In The Life of the Internet) by OARC
- 2016, April 5 - 7
- Selected sample:
  - root servers — C, E, J, K, E, I, L
  - TLD — CIRA (.ca), SWITCH (.ch), NIC Chile (.cl), CZ.NIC (.cz), EIS (.ee)
  - RIR — AFRINIC
  - AS112 — WIDE Project
- ~67 million DNS queries in ~85 million TCP sessions

Questions asked

- Is TCP used only as a fallback protocol?
- Do the resolvers reuse existing TCP sessions efficiently?
- What is the resolvers' policy on keeping the connection open?
- When will TCP perform better than UDP?
- …

TCP connection reuse



	Number of connections	Number of DNS queries
0	24,314,732 (37.3044 %)	
1	37,740,432 (57.9025 %)	37,740,432 (80.6964 %)
2	1,668,266 (2.5595 %)	3,336,532 (7.1342 %)
3	753,228 (1.1556 %)	2,259,684 (4.8316 %)
4	367,293 (0.5635 %)	1,469,172 (3.1414 %)
5	179,896 (0.2760 %)	899,480 (1.9233 %)
6	84,910 (0.1303 %)	509,460 (1.0893 %)
7	38,740 (0.0594 %)	271,180 (0.5798 %)
8	17,397 (0.0267 %)	139,176 (0.2976 %)
9	7,752 (0.0119 %)	69,768 (0.1492 %)
10	3,451 (0.0053 %)	34,510 (0.0738 %)

Clients talking TCP

	queries	sessions	avg queries/session
AS 15169 (Google Inc.)	9,025,277 (99.686 %)	3,122,913 (99.956 %)	2.89
AS 4134 (Chinanet)	1,644 (0.018 %)	822 (0.026 %)	2
AS 4808 (China Unicom Beijing Province Network)	186 (0.002 %)	93 (0.003 %)	2
AS 57356 (Highland Network Ltd)	13,264 (0.147 %)	67 (0.002 %)	197.97
AS 8605 (University of Latvia)	12,582 (0.139 %)	63 (0.002 %)	199.71
AS 16276 (OVH SAS)	86 (0.001 %)	43 (0.001 %)	2
AS 4812 (China Telecom (Group))	74 (0.001 %)	37 (0.001 %)	2
AS 4847 (China Networks Inter-Exchange)	54 (0.001 %)	27 (0.001 %)	2
AS 15076 (Delgado Industries, LLC)	51 (0.001 %)	23 (0.001 %)	2.22
AS 3356 (Level 3 Communications, Inc.)	24 (0.000 %)	12 (0.000 %)	2

The numbers above exclude TCP connections that delivered < 2 queries.

Clients talking TCP

- Legitimate clients:
  - Google Public DNS
  - Custom tools to mass check if domains are available
  - "`dig +keepopen`"-like tools
  - …
- Broken clients:
  - Clients retransmitting every query several times
  - Clients reflecting responses back to servers
  - …

Can TCP perform better than UDP?

- UDP is a natural choice because it's stateless
- UDP is cheaper for clients
- UDP is cheaper for servers (unless processing is expensive)

- TCP handshake
- TCP head of line blocking
- TCP source address is unlikely to be spoofed
- TCP has congestion control
- TCP has higher throughput (Nagle)

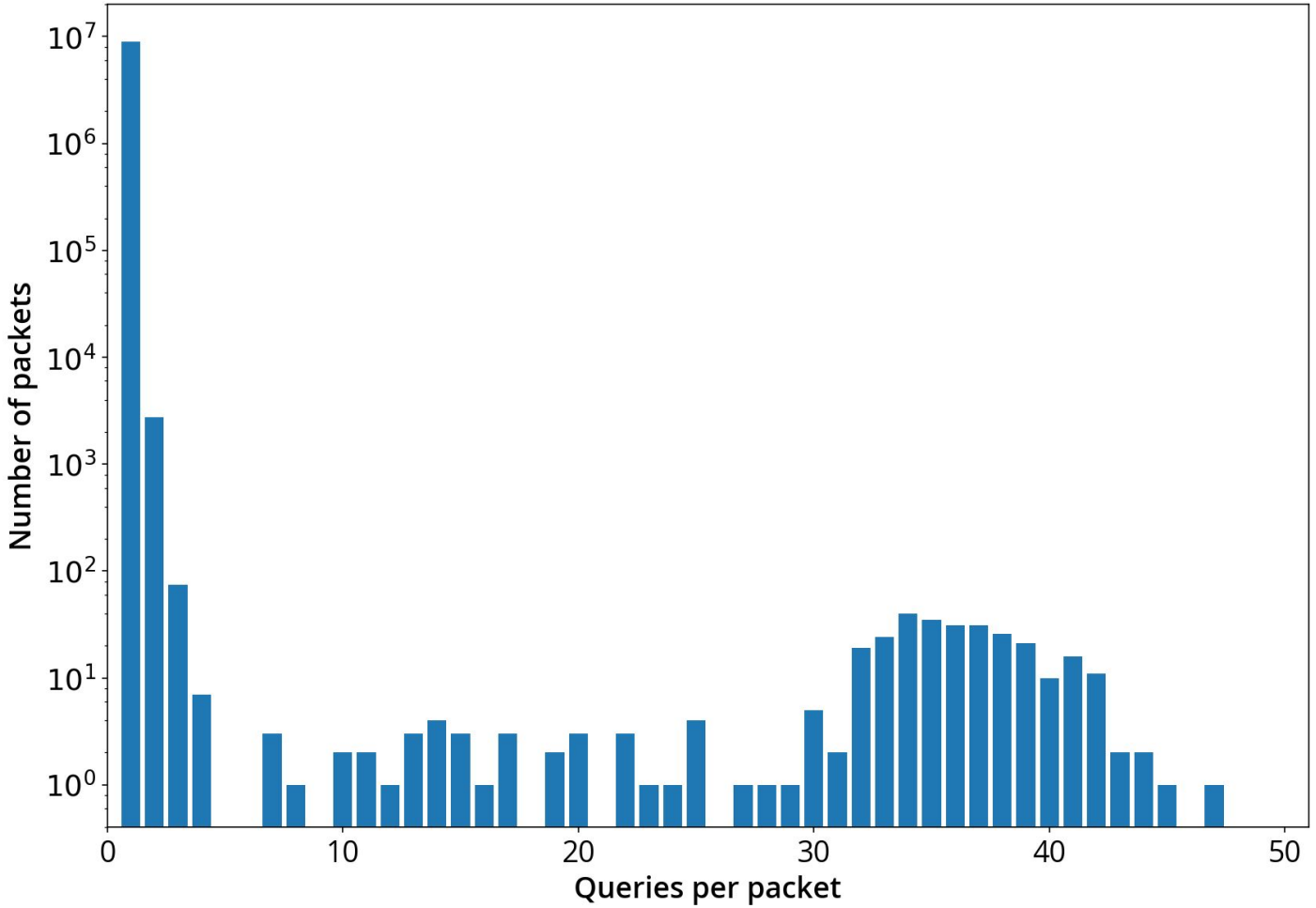
Nagle's algorithm (1)

- Gather small writes into a single packet:

```
if there is data to send:
    if window size >= MSS and available data >= MSS:
        send complete MSS segment
    else if there is unacknowledged data in flight:
        enqueue data and wait for acknowledgment
    else:
        send data immediately
```

- Example with spherical cows:
  - 20 DNS queries, 50 bytes each, Ethernet, IPv6
  - UDP: 20 * (18 + 40 + 8 + 50) = 2320 bytes (in 20 packets)
  - TCP: 18 + 40 + 20 + 20 * (50 + 2) = 1118 bytes (in 1 packet)

Nagle's algorithm (2)

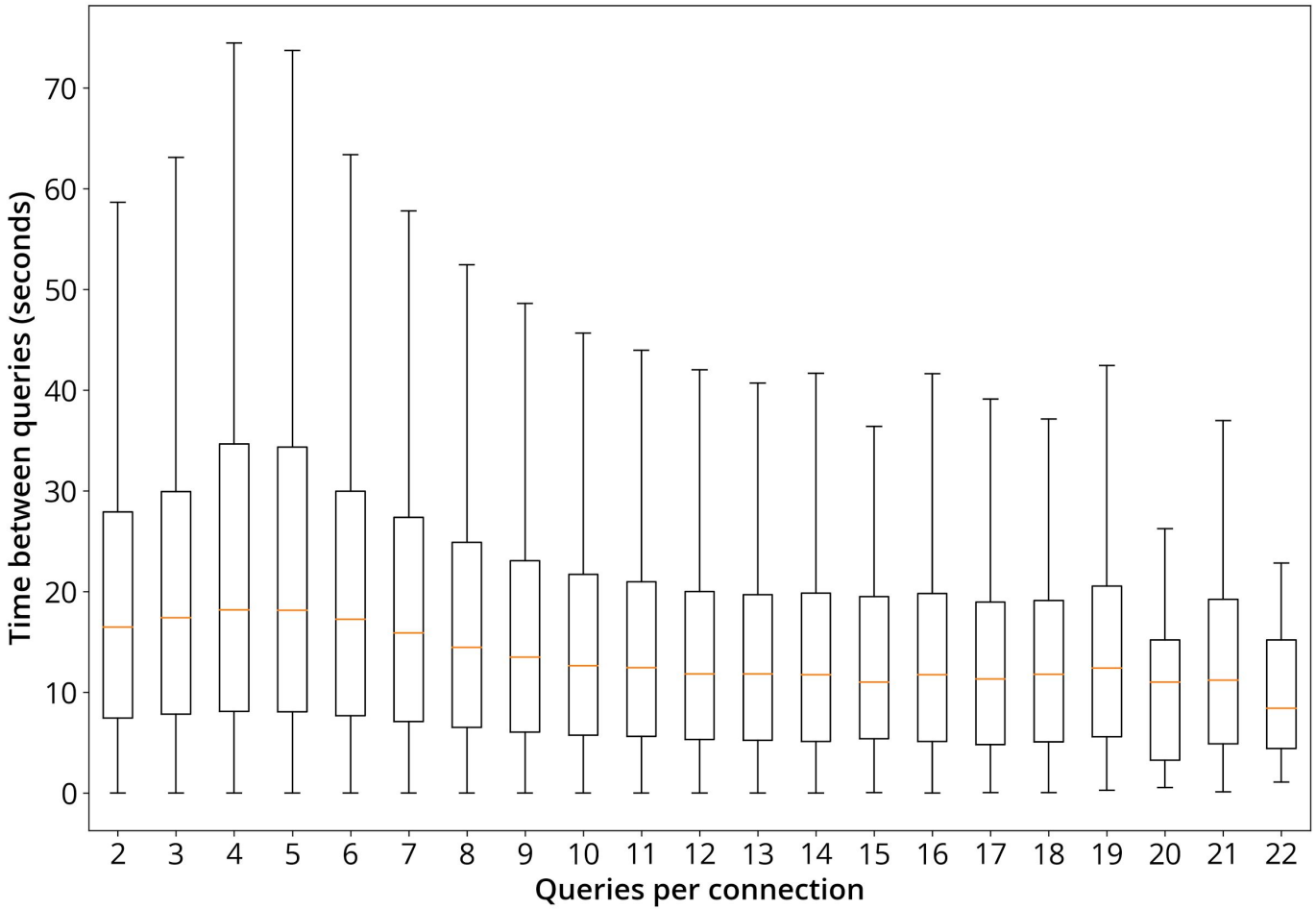


1	9,037,834 (99.818 %)
2	2,767 (0.061 %)
3	75 (0.002 %)
4	7 (0.000 %)
5	0 (0.000 %)

< 5 qpp: "real" Nagle

> 5 qpp: domain availability check

TCP preference (1)

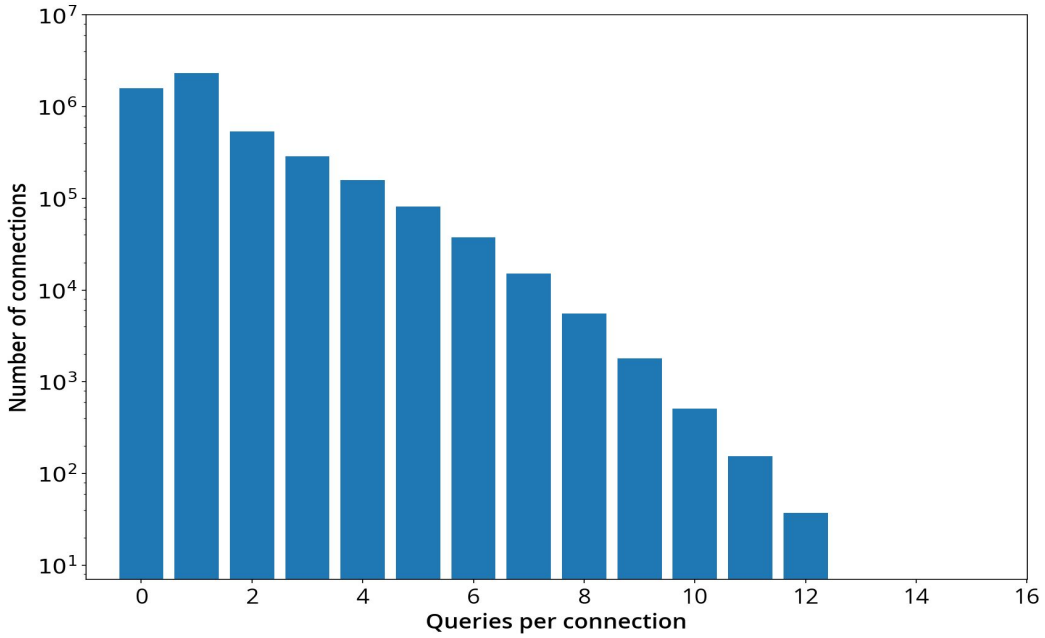


Use of TCP seems to be independent on the number of queries.

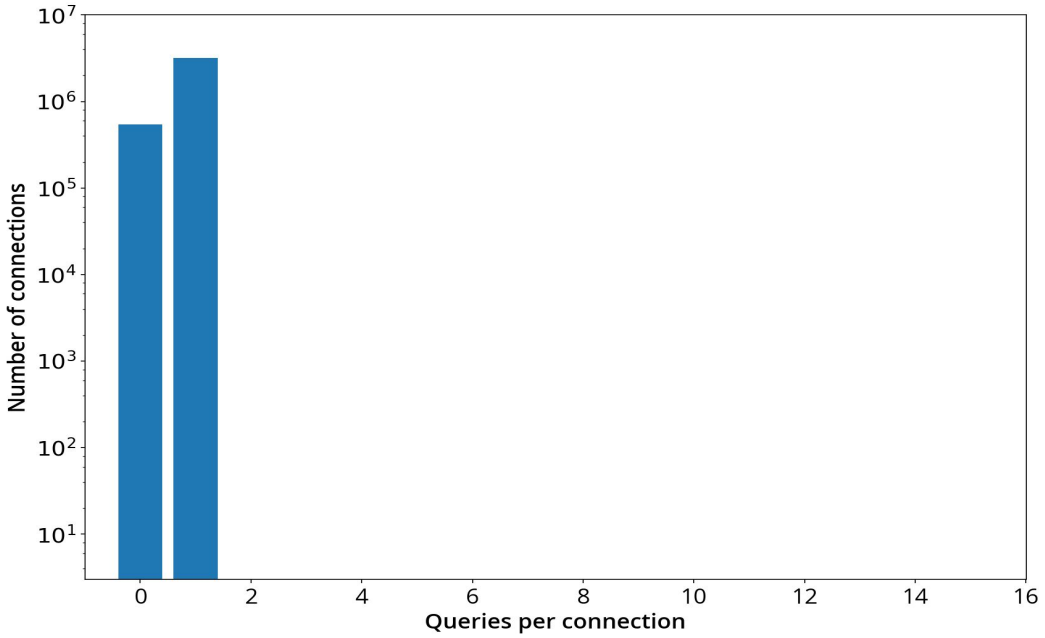
The mean time between queries is relatively high (10-20 seconds).

TCP preference (2)

There is a point where reusing existing TCP connections starts to be preferred.



E root in Atlanta (5,690,859 queries)



E root in New York (3,718,454 queries)

Open questions

- Why is ~28 % of connections closed without sending a single query?
- Are there “network failure conditions” where TCP will perform better?
- What is retransmitted ratio on UDP vs TCP?
- What is the effect of TCP congestion control on DNS?
- Are the queries sent by resolvers grouped if related?
- …

Used tools

- Available at https://github.com/fcelda/dns-tcp-stats
- Conservative toolset (tshark, python, shell scripts)
  - Reads pcaps, extracts TCP sessions to a CSV format
  - Removes retransmits or invalid packets
  - Removes server initiated sessions and zone transfers
  - Attempts to remove garbage queries

THANK YOU.

BUILD A SMARTER INTERNET™

NS1.

Jan Včelák · Software Engineer

JVCELAK@NS1.COM | @FCELDA @NSONEINC