Fragmentation, truncation, and timeouts: are large DNS messages falling to bits?

Giovane C. M. Moura[1], Moritz Müller[1,2], Marco Davids[1], Maarten Wullink[1], Cristian Hesselman[1,2]

1: SIDN Labs, 2: University of Twente

DNS OARC 35
Virtual conference
2021-05-07

Introduction

- *this presentation is from a paper presented at PAM2021*
  - PDF: `http://shorturl.at/iqtB0`
- The DNS is one of the core protocols on the Internet [5]
- Every web page visit requires DNS queries
- DNS uses both UDP and TCP [4]:
  - DNS/UDP: super fast (1 RTT)
  - DNS/TCP: zone transfer and UDP-fall back

The problem: large messages over DNS/UDP

- Transport limits:
  - Vanilla DNS/UDP: max 512 bytes
  - DNS/TCP: <no strict limit>
  - **The issue: DNS/UDP with EDNS-0 [2]: up to 65k bytes**
- If a response is too large:
  - For the **network MTU**: packets will be either **FRAGMENTED** [1] or DISCARDED: may lead to **unreachability**
  - For the **server**: then **TRUNCATE** it, and client should ask via TCP
- **Question**: how big is this of a problem on DNS?

SIDN LABS UNIVERSITY OF TWENTE.

We investigate the issue in production traffic

- Analyze traffic to a ccTLD (The Netherlands' `.nl`)
  - 3 months of data (2019 and 2020)
  - **164 billion queries** from 3M unique IPs and 46k ASes

	July 2019		July 2020		October 2020	
	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6
Queries/responses	29.79B	7.80B	45.38B	15.87B	48.58B	16.62B
UDP	28.68B	7.54 B	43.75B	15.01B	46.94B	15.87B
UDP TC off	27.80B	7.24B	42.06B	13.88B	45.49B	14.93B
UDP TC on	0.87B	0.31B	1.69B	1.14B	1.44B	0.93B
Ratio (%)	2.93%	3.91%	3.72%	7.15%	2.96%	5.59%
TCP	1.11B	0.25B	1.63B	0.85B	0.36B	0.20B
Ratio (%)	3.72%	3.32%	3.59%	5.37%	3.17%	5.09%
Resolvers						
UDP TC off	3.09M	0.35M	2.99M	0.67M	3.12M	0.62M
UDP TC on	0.61M	0.08M	0.85M	0.12M	0.87M	0.13M
TCP	0.61M	0.08M	0.83M	0.12M	0.87M	0.13M
ASes						
UDP TC off	44.8k	8.3k	45.6k	8.5k	46.4k	8.8k
UDP TC on.	23.3k	4.5k	27.6k	5.4k	28.2k	5.6k
TCP	23.5k	4.3k	27.3k	5.2k	27.9k	5.4k

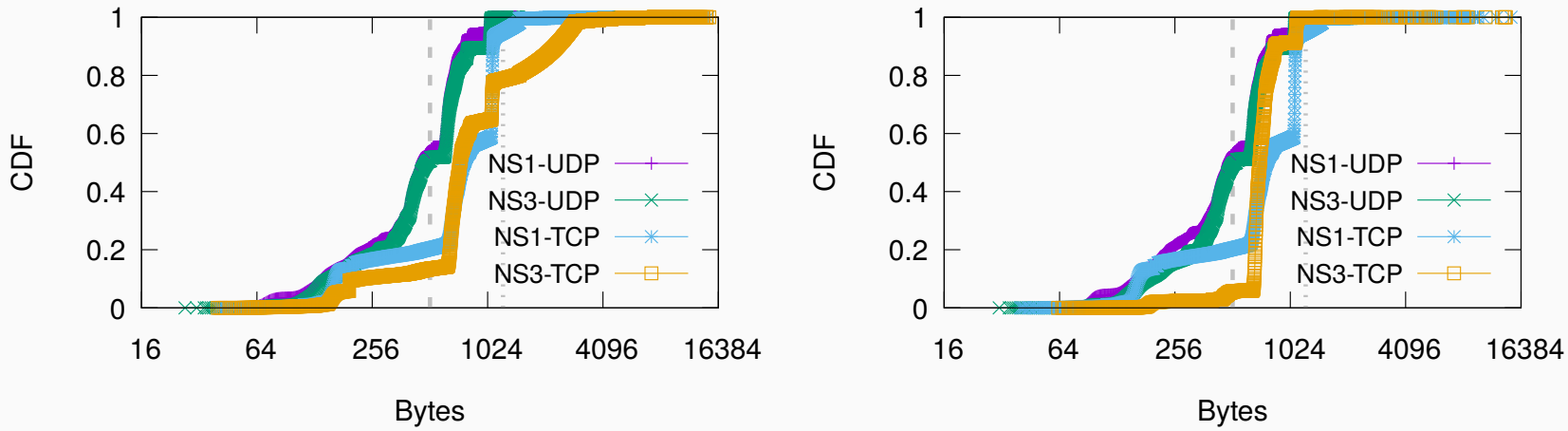
Table 1: Evaluated datasets of `.nl` zone

Research questions

1. How common are large DNS responses?
2. How common is DNS truncation and server-side fragmentation?
3. Do resolvers fall back to TCP after truncation?
4. Impact of DNS Flag day 2020 on buffer configurations

SIDN LABS UNIVERSITY OF TWENTE.

How common are large responses?



(a) 2019: IPv4 **(b)** 2019: IPv6

Figure 1: Response size CDF for `.nl`: July 2019

- **99.99% of responses from `.nl` are smaller than 1232 bytes**
- No need to FUD. Google Public DNS says 99.7% are smaller than 1232 bytes.

How often *server-side* fragmentation occurs?

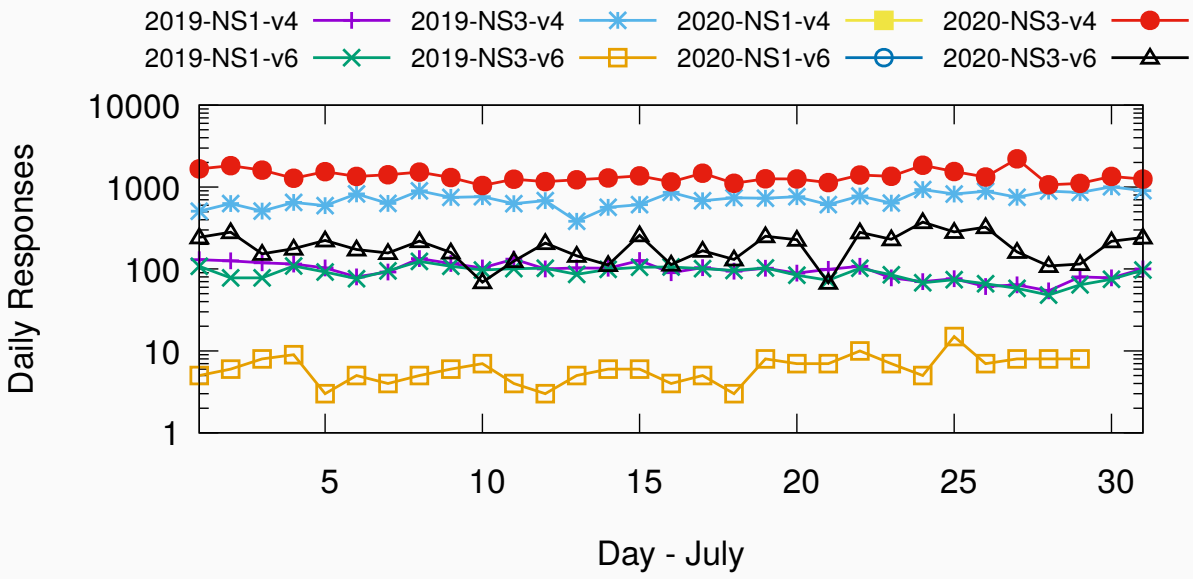


Figure 2: UDP fragmented queries for `.nl` authoritative servers.

- **Rarely**: <10k queries/day (from 2.2B/daily)

What about in-network fragmentation?

	Large	Small
EDNS0 buffer	4096	512
Query	ANY NS .nl	A ns1.dns.nl
Target	ns3.dns.nl	
Response Size	1744	221
Protocol/IP	UDP/IPv4	
Active Probes	9323	9322
∩	8576	
Queries	557047	555007
∩	512351	510575
OK	473606	497792
timeout	38745(6.9%)	12783 (2.5%)

Table 2: Atlas measurements for large and small responses. Datasets:[6]

What about in-network fragmentation?

- It only occurs for IPv4
- Our vantage point (authoritative servers) allow to see if clients received responses
- We then measure with Ripe Atlas: 8500 probes over 1 day
  1. 2.5% of small responses timeout (221 bytes)
  2. 6.9% of large responses (1744 bytes) timeout
  3. (similar figures with previous works [3, 7])

How common is DNS truncation?

	July 2019		July 2020		October 2020	
	IPv4	IPv6	IPv4	IPv6	IPv4	IPv6
Queries/responses	29.79B	7.80B	45.38B	15.87B	48.58B	16.62B
UDP	28.68B	7.54 B	43.75B	15.01B	46.94B	15.87B
UDP TC off	27.80B	7.24B	42.06B	13.88B	45.49B	14.93B
UDP TC on	0.87B	0.31B	1.69B	1.14B	1.44B	0.93B
Ratio (%)	2.93%	3.91%	3.72%	7.15%	2.96%	5.59%

Table 3: Evaluated datasets of `.nl` zone

In the paper: .

- most queries truncated to 512 bytes
- Large EDNS0 buffers size don't prevent truncation

SIDN LABS UNIVERSITY OF TWENTE.

So resolvers fall back to TCP after truncation?

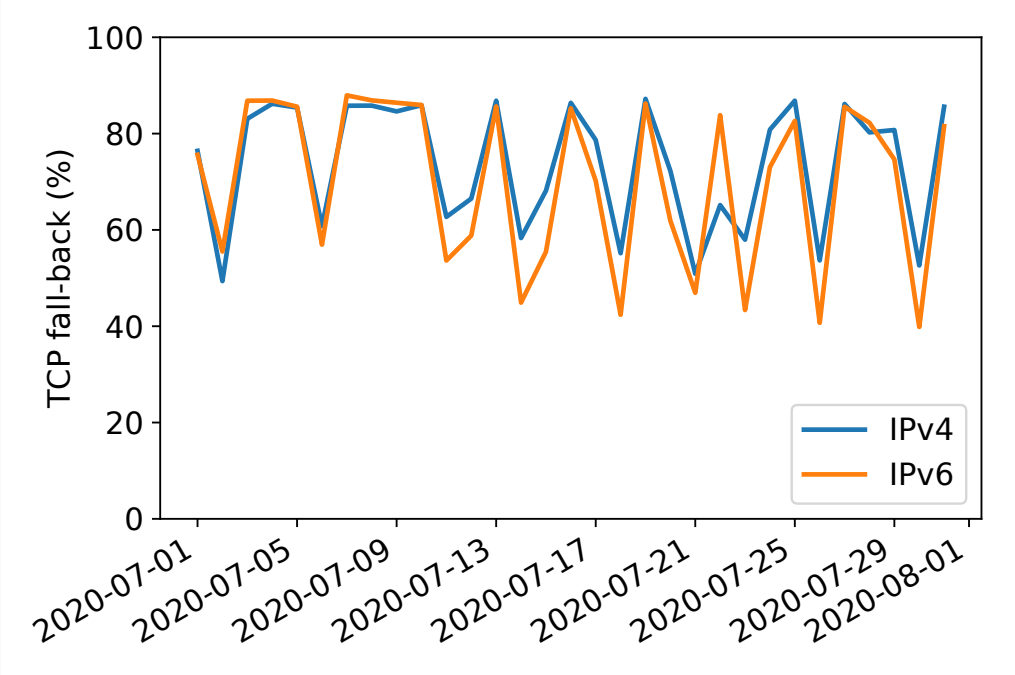
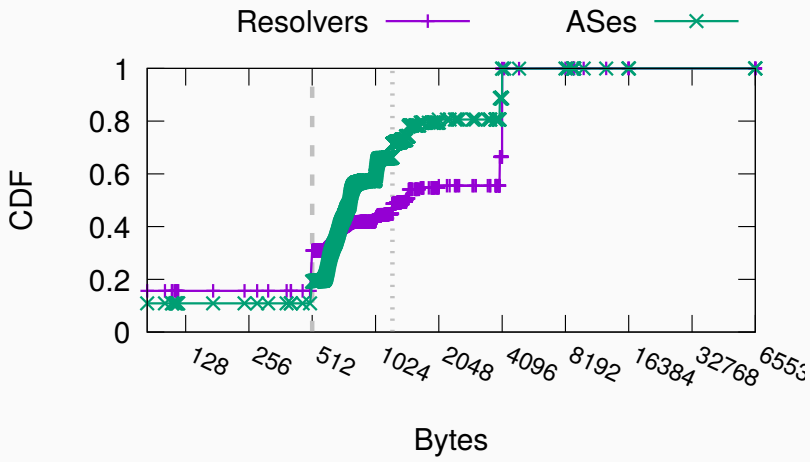


Figure 3: TC replies with TCP retries

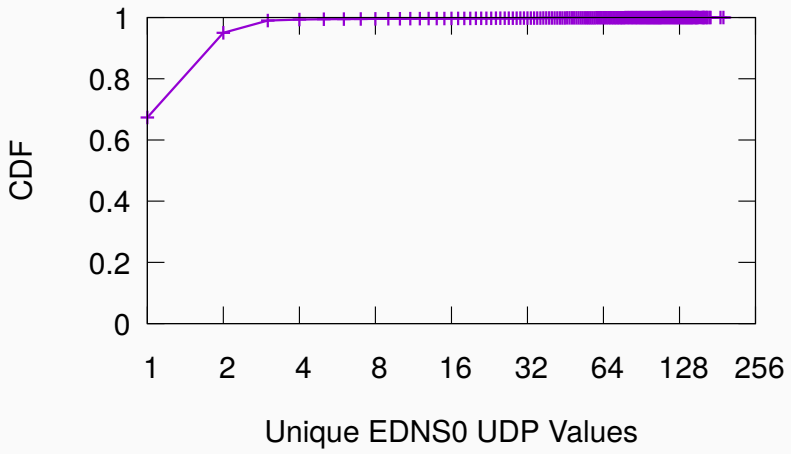
79-85% of truncated responses are followed by TCP

SIDN LABS UNIVERSITY OF TWENTE.

What are the most common EDNS0 values



(a) EDNS0 Values distribution

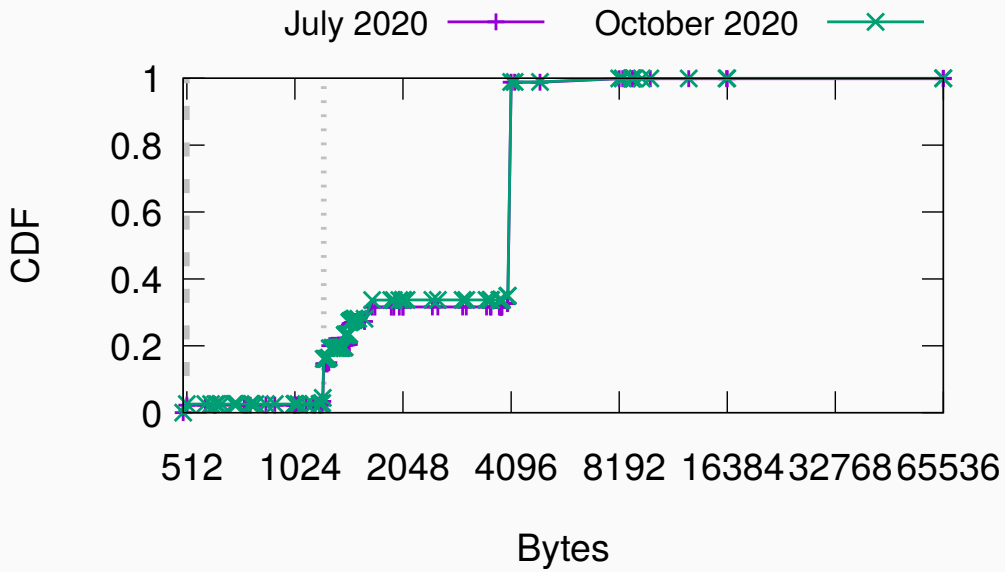


(b) Unique EDNS per resolver

Figure 4: EDNS0 per resolver and values: July 2020

DNS Flag day 2020

- To avoid fragmentation, member of DNS community proposed 1232 byte limit for DNS/UDP
- Resolvers can advertise this value as their EDNS0 value
- What was the uptake? (not much)



SIDN LABS UNIVERSITY OF TWENTE.

DNS Flag Day 2020

	July 2020	October 2020
Resolvers	3.78M	3.84M
∩	1.85 M	
UDP Queries	60.3B	62.81B
∩	117.54 B	

(a) Before and After Datasets

Resolvers	**11338**
from 4096 bytes	7881
from 1680 bytes	1807
from 512 bytes	1252
rest	398
ASes	958
Queries	3.01B

(b) EDNS0 1232 resolvers

Table 4: DNS Flag Day datasets and Changing Resolvers

Summary

Are DNS responses falling to bits?

1. Most DNS responses are small, so little fragmentation risk
2. Server-side fragmentation is minimal
3. 2–7% of `.nl` UDP responses are truncated
4. 79–85% are followed by a TCP query
5. DNS Flag Day 2020 uptake was not very noticiable yet

References i

[1] Bonica, R., Baker, F., Huston, G., Hinden, R., Troan, O., and Gont, F.

IP Fragmentation Considered Fragile.

RFC 8900, IETF, Sept. 2020.

[2] Damas, J., Graff, M., and Vixie, P.

Extension Mechanisms for DNS (EDNS(0)).

RFC 6891, IETF, Apr. 2013.

[3] Huston, G.

Dealing with IPv6 fragmentation in the DNS.

https://blog.apnic.net/2017/08/22/dealing-ipv6-fragmentation-dns/, Aug. 2017.

SIDN LABS UNIVERSITY OF TWENTE.

References ii

[4] Mockapetris, P.

Domain names - concepts and facilities.

RFC 1034, IETF, Nov. 1987.

[5] Moura, G. C. M., de O. Schmidt, R., Heidemann, J., de Vries, W. B., Müller, M., Wei, L., and Hesselman, C.

Anycast vs. DDoS: Evaluating the November 2015 root DNS event.

In *Proceedings of the ACM Internet Measurement Conference* (Santa Monica, California, USA, Nov. 2016), ACM, pp. 255–270.

SIDN LABS UNIVERSITY OF TWENTE.

References iii

[6] RIPE NCC.

RIPE Atlas measurement IDS.

https://atlas.ripe.net/measurements/ID, Oct. 2020.

, where ID is the experiment ID: large:27759950, small:27760294.

[7] Van Den Broek, G., van Rijswijk-Deij, R., Sperotto, A., and Pras, A.

DNSSEC meets real world: dealing with unreachability caused by fragmentation.

IEEE communications magazine 52, 4 (2014), 154–160.