Measuring DNS over TLS from the Edge: Adoption, Reliability, and Response Times

Trinh Viet Doan(✉), Irina Tsareva, and Vaibhav Bajpai

Technical University of Munich, Munich, Germany
doan@in.tum.de, irina.tsareva@tum.de, bajpaiv@in.tum.de

Abstract. The Domain Name System (DNS) is a cornerstone of communication on the Internet. DNS over TLS (DoT) has been standardized in 2016 as an extension to the DNS protocol, however, its performance has not been extensively studied yet. In the first study that measures DoT from the edge, we leverage 3.2k RIPE Atlas probes deployed in home networks to assess the adoption, reliability, and response times of DoT in comparison with DNS over UDP/53 (Do53). Each probe issues 200 domain name lookups to 15 public resolvers, five of which support DoT, and to the probes' local resolvers over a period of one week, resulting in 90M DNS measurements in total. We find that the support for DoT among open resolvers has increased by 23.1% after nine months in comparison with previous studies. However, we observe that DoT is still only supported by local resolvers for 0.4% of the RIPE Atlas probes. In terms of reliability, we find failure rates for DoT to be inflated by 0.4–32.2 percentage points when compared to Do53. While Do53 failure rates for most resolvers individually are consistent across continents, DoT failure rates have much higher variation. As for response times, we see high regional differences for DoT and find that nearly all DoT requests take at least 100 ms to return a response (in a large part due to connection and session establishment), showing an inflation in response times of more than 100 ms compared to Do53. Despite the low adoption of DoT among local resolvers, they achieve DoT response times of around 140–150 ms similar to public resolvers (130–230 ms), although local resolvers also exhibit higher failure rates in comparison.

1 Introduction

The Domain Name System (DNS) faces various privacy-related issues such as fingerprinting or tracking [11,22,10,36,23] that affect DNS over UDP/53 (Do53). Consequently, DNS over TLS (DoT) was standardized in 2016 [19] to upgrade the communication [35]: The protocol establishes a TCP connection and TLS session on port 853, so that DNS messages are transmitted over an encrypted channel to circumvent eavesdropping and information exposure. DoT has gained increasing support since its standardization; e.g., it is supported on Android devices as "Private DNS" since Android 9 (August 2018) [24]. Similarly, Apple

The final authenticated version is available online at https://doi.org/10.1007/978-3-030-72582-2_12.

supports DoT and DNS over HTTPS (DoH) on their devices and services with the recent iOS 14 (September 2020) and MacOS Big Sur (November 2020) [38].

Previous work [8,26,17] has studied the support and response times of DoT (and DoH). However, the studies performed response time measurements from proxy networks and data centers, which means that results might not appropriately reflect the latency of regular home users: The measured response times are likely overestimated due to the incurred latency overhead of proxy networks or underestimated due to the usage of well-provisioned data centers. We close this gap by measuring DoT from the end user [28] perspective for multiple DoT resolvers as the first study to do so, using 3.2k RIPE Atlas home probes deployed at the edge across more than 125 countries (§ 3). We issue DNS queries to 15 public resolvers, five of which support DoT, to analyze and compare the reliability and response times of Do53 and DoT resolvers. Our main findings are:

DoT support (§ 2)**:** We find DoT support among open resolvers to have increased by 23.1% compared to previous studies [8,26]. TLS 1.3 support [31,15] among these resolvers has increased by 15 percentage points, while support for TLS 1.0 and 1.1 is increasingly dropped. For RIPE Atlas (§ 4), we only find 13 (0.4%) of 3.2k home probes to receive responses over DoT from their local resolvers.

DoT failure rates (§ 4)**:** While overall failure rates for Do53 are between 0.8–1.5% for most resolvers, failure rates for DoT are higher with 1.3–39.4%, i.e., higher by 0.4–32.2 percentage points for individual resolvers. Failure rates are more varying across the continents for DoT, ranging from $\leq$1% up to >10%, with higher values primarily seen in Africa (AF) and South America (SA). On the other hand, Do53 failure rates are more consistent across most resolvers and continents (roughly 0.3–3%). Most failures occur due to timeouts (no response within 5 seconds), which we suspect is due to intervening middleboxes on the path that blackhole the connections by dropping packets destined for port `853`.

DoT response times (§ 5)**:** Comparing response times between Do53 and DoT, we find that most DoT response times are within roughly 130–230 ms, and are, therefore, slower by more than 100 ms, largely due to additional TCP and TLS handshakes. For most samples of well-known DNS services (such as Google, Quad9, or Cloudflare), response times of for Do53 are consistent across the continents, while other resolvers show larger regional differences. For DoT, only Cloudflare exhibits consistent response times across regions, whereas the remaining resolvers have highly varying response times. In cases where the local resolver does support DoT, response times are comparable to those of the faster public resolvers (140–150 ms) and similarly inflated compared to Do53.

We discuss limitations (§ 7) and compare our findings to previous work (§ 6) before concluding the study (§ 8). To facilitate reproducibility of our results [1], we share the created RIPE Atlas measurement IDs, analysis scripts, and auxiliary/supplementary files[1]. The measurements do not raise any ethical concerns.

[1] Repository: `https://github.com/tv-doan/pam-2021-ripe-atlas-dot`

2 DoT Background: Adoption and Traffic Share

DoT adoption among open resolvers. Deccio and Davis [8] study and quantify the deployment of public DoT resolvers as of April 2019. Note that in the context of their study, a resolver refers to an IP endpoint, which may, therefore, include a replicated or anycasted service. They identify 1.2M open DNS resolvers in the public IPv4 address space, out of which 0.15% (1,747) support DoT. Of the DoT resolvers, 97% (1,701) support TLS 1.2 and 4.5% (79) support TLS 1.3, whereas older TLS versions (TLS 1.0 and 1.1) are not supported by 4.6% (80) of the resolvers. A similar number of open DoT resolvers (1.5k) was found by Lu *et al.* [26] (2019).

We repeat this scan from a research network at Technical University of Munich (TUM) in January 2020 (i.e., nine months after Deccio and Davis [8]) for the same set of open DNS resolvers. We find that the number of open resolvers supporting DoT has increased to 2,151, i.e., an increase by 23.1%. The share of resolvers supporting TLS 1.2 has increased to 99.9% (2,149 resolvers), while the percentage of TLS 1.3-supporting resolvers has increased to 20% even (433). Older versions of TLS are not supported anymore by 508 resolvers (24%), which altogether indicates that the adoption of DoT and newer TLS implementations is increasing.

DoT traffic share. To assess the usage of DoT in terms of traffic, we analyze public traffic traces collected from samplepoint-F of the WIDE backbone [7], which monitors a research network link in Japan. We aggregate the daily traffic traces of 2019 by month and inspect the traffic share of DoT, i.e., traffic on TCP/`853`. We observe that DoT accounts for roughly 2M out of 11.8B flows in the dataset, which means that DoT accounts for around 0.017% of all flows. On the other hand, the traffic share of Do53 is more than 135 times as much with 271.5M flows (2.3%), which indicates that DoT only contributes a very negligible amount of traffic overall.

3 Methodology

Measurement platform and probes. We use RIPE Atlas [32] to measure reliability and response times of Do53 and DoT from distributed vantage points; DoT measurements are performed over TLS 1.2, as RIPE Atlas probes do not fully support TLS 1.3 yet. For our experiment, we first select probes that are IPv4-capable and resolve `A` records correctly through the RIPE Atlas API. We exclude anchor probes to capture the Do53 and DoT behavior for end users more accurately. As older versions of RIPE Atlas probes (V1 and V2) exhibit load issues [2,14], we only consider V3 probes, ultimately finding 5,229 probes in total. For the analysis, however, we only take residential probes into account: We use RIPE Atlas user tags [3] for the identification of residential networks. Additionally, we issue `traceroute` measurements to an arbitrary public endpoint from all probes over IPv4: If the IP address of first hop on the path is private [30] and the IP address of the second hop is in the public address space (i.e., the probe is directly connected to the home gateway), we also identify the probe as residential. Combining the set of probe IDs determined from both these approaches, we identify 3,231 *home probes* overall. As the number of dual-stacked residential probes is significantly lower (roughly 700 globally), we decide to not perform measurements over IPv6: The low number of IPv6-capable probes overall limits the regional analysis, since such probes are primarily deployed in Europe (EU) and North America (NA), which would leave other continents largely underrepresented. Thus, we focus on IPv4 measurements exclusively in our study, although we suggest to repeat the measurements over IPv6 with increased deployment of probes having native IPv6 connectivity.

DNS resolvers. We issue the resolution of 200 domains (`A` records) to 15 selected IP endpoints of different public DNS services once a day, repeated over a period of one week (July 03–09, 2019). Out of the 15 public DNS services, listed in Table 1, five support DoT: CleanBrowsing, Cloudflare, Google, Quad9, and UncensoredDNS. For these services, we additionally issue the same DNS lookups to the same IP endpoints using DoT for comparison. Moreover, we query the same 200 domains using the DNS resolvers provided by a probe's network configuration, which we will refer to as *local resolver* (typically operated by the ISP and assigned via DHCP) in the following; this allows us to study the support of DoT among ISPs. Note that probes may use multiple IP endpoints when resolving domains locally. In particular, probe hosts may use public resolvers as their local resolvers; thus, we exclude all occurrences of these public resolvers from the local resolver measurements, including alternative IP endpoints which these public DNS services may use. Among the 2,718 probes that receive at least one successful Do53 response from a local resolver, we find 2,257 probes to use an endpoint in their private network as local resolver (e.g., a CPE) and 572 probes to use an ISP resolver (public IP address) for local name resolution. However, as we do not see significant differences in terms of response times at the 5^{th} percentiles of each probe (9.5 ms for CPE, 9.8 ms for ISP resolver), we do not further distinguish between both groups.

Domains. The 200 queried domains consist of 150 websites from Alexa Top 1M [33]: We split the Top 1M list into 10 equally-sized bins of 100k each (by rank order) and select the 15 first domains of each bin, resulting in 150 popularity-focused domains. The remaining 50 domains are selected from the country-based Alexa Toplists, for which we determine 10 countries across the continents with high numbers of probes (US, DE, GB, RU, NL, IT, JP, NZ, ZA, BR). We then pick 5 website domains from each Alexa Toplist of the associated Top-Level Domain (`.us`, `.de`, `.co.uk`, etc.), resulting in 50 region-focused domains. Note that sampling the entire 1M domains does not improve representativeness, since we repeat the measurements over a period of one week and expect records to be cached. Also, the known instability of the Alexa Toplist [33] does not substantially influence our measurements: We construct the list of overall 200 domains (from July 01, 2019) to investigate whether there are larger differences between bins of more popular and less popular domains, or in terms of Top-Level Domain (TLD) and probe location. However, we do not find any significant devi-

Table 1. Overview of measured resolvers together with the number of failed requests, total requests, and failure rates for both Do53 and DoT. Failure rates for DoT are higher compared to Do53 for each resolver, with failure rates also being lower for public DNS services than local resolvers. Highlighted cells are referred to in § 4.

Resolver Name	Do53			DNS over TLS		
	# Failures	# Total	Failure Rate	# Failures	# Total	Failure Rate
1) CZ.NIC ODVR	44,942	4,269,957	1.1%	—	—	—
2) CleanBrowsing	37,681	4,273,000	0.9%	430,401	4,163,095	10.3%
3) Cloudflare 1.1.1.1	107,841	4,273,000	2.5%	122,932	4,157,033	3.0%
4) Comodo Secure DNS	65,849	4,272,976	1.5%	—	—	—
5) DNS.WATCH	43,349	4,272,960	1.0%	—	—	—
6) Google Public DNS	38,670	4,272,587	0.9%	53,059	4,157,354	1.3%
7) Neustar UltraRecursive	4,190,474	4,269,365	98.2%	—	—	—
8) OpenDNS	34,826	4,273,051	0.8%	—	—	—
9) OpenNIC	61,077	4,266,712	1.4%	—	—	—
10) Oracle + Dyn	46,247	4,272,609	1.1%	—	—	—
11) Quad9	51,292	4,272,979	1.2%	110,404	4,157,340	2.7%
12) SafeDNS	37,291	4,269,648	0.9%	—	—	—
13) UncensoredDNS	62,175	4,269,656	1.5%	4,039,111	4,157,277	97.2%
14) VeriSign Public DNS	36,644	4,269,638	0.9%	—	—	—
15) Yandex.DNS	53,581	4,269,591	1.3%	—	—	—
16a) Local Resolver without DoT support	573,514	5,108,671	11.2%	—	—	—
16b) Local Resolver with DoT support	2,356	32,649	7.2%	13,737	34,839	39.4%
Total	**5,487,809**	**69,209,049**	**7.9%**	**4,769,644**	**20,826,938**	**22.9%**

ations in terms of response times, neither regarding popularity rank nor TLD. Thus, we do not further distinguish between individual domains in the analysis.

With this experiment setup, we collect measurements for around 90M DNS requests from home probes in total (see Table 1).

4 Reliability

We investigate the reliability of Do53 and DoT by analyzing the *failure rate*, which we define as the relative number of failed queries to the total number of queries. A query is defined as failed if the domain lookup could not be sent to the resolver or the probe did not receive a response; in both cases, the RIPE Atlas API will return an error. Table 1 shows the overall failure rate, as well as the failure rate by resolver, for both Do53 and DoT. Note that we exclude 33 probes which failed nearly all of their DoT measurements (see *error analysis* below) from all following analyses. Further, only 2,718 probes of the 3.2k home probes successfully receive a Do53 response from local resolvers, i.e., the remaining probes cannot resolve a domain using a local resolver (but can with a public resolver). Considering DoT, we find that only 13 probes receive responses from their local resolver via DoT, which means that DoT is only supported by 0.4%

of the local resolvers. We exclusively see these DoT-supporting local resolvers (discussed in more detail in § 5) in EU (11 probes) and NA (2 probes). As such, we separate the queries to local resolvers (by probes with and without DoT-supported local resolvers) in Table 1 and this subsection.

Overall failure rates. The overall failure rate for Do53 is 7.9%, with individual failure rates of 0.8–1.5% for most resolvers, whereas the overall failure rate for DoT is much higher at 22.9%, i.e., a difference of 15.0 percentage points (p.p.). However, the total failure rates are heavily influenced by a few resolvers exhibiting particularly high failure rates of close to 100%: For instance, 98.2% of the Do53 requests to Neustar UltraRecursive fail, accounting for 76.4% of the Do53 failure rate in total. For DoT, UncensoredDNS accounts for 84.7% of all DoT failures with an individual failure rate of 97.2%; local resolvers with DoT support have an overall DoT failure rate of 39.4%.

Individually, the Do53 failure rate is between 0.8% and 2.5% for all public resolvers when disregarding Neustar. Local resolvers encounter failures in 11.2% of the cases instead (7.2% for probes with DoT-supported local resolvers).

We observe an inflation of failure rates when moving from Do53 to DoT for all DoT resolvers: Inflations range from 0.4 and 0.5 p.p. for Google and Cloudflare, over 1.5 p.p. for Quad9 and 9.4 p.p. for CleanBrowsing, to 95.7 p.p. for UncensoredDNS; local resolvers with DoT support show an inflation toward the higher end with 32.2 p.p.. Overall, these numbers suggest that DoT support on the paths is still experimental and, therefore, varying concerning reliability.

Error analysis. Regarding the respective error messages, we find that most failures are attributed to timeouts (5 seconds), socket errors, and `connect()` errors (connection refused/reset, network unreachable). For Do53, nearly all failed requests toward Neustar (>99.9%) are due to timeouts. DoT measurements show a significant amount of `TUCONNECT` errors, which are exclusive to DoT and suggest TLS negotiation errors. To further investigate this, we count the number of `TUCONNECT` errors for each combination of probe and public resolver; we exclude UncensoredDNS from this analysis due to its high failure rate overall (which indicates server-side issues). For all combinations of 3.2k probes × 4 resolvers, we find repeated `TUCONNECT` errors for 33 probes across all resolvers where the probes fail nearly all scheduled 1.4k DoT measurements (200 domains × 7 days). This indicates blackholing of DoT packets closer to these probe (home router or in the ISP network). Although the number of affected probes is negligible (≈1%), we have excluded the affected 33 probes from the previous and following analyses. We further investigate `TUCONNECT` errors and find a higher number of probes failing nearly all DoT measurements for Cloudflare in particular, which affects 99 probes. The differential of 66 probes between these two groups show no errors for the other resolvers, suggesting DoT blackholing closer to Cloudflare anycast instances that serve these probes, which in return causes a higher failure rate compared to other resolvers. CleanBrowsing, on the other hand, shows a similar failure rate regarding `TUCONNECT` errors as Google or Quad9; the majority of CleanBrowsing's overall DoT failures (10.3%) stem from timeouts instead.

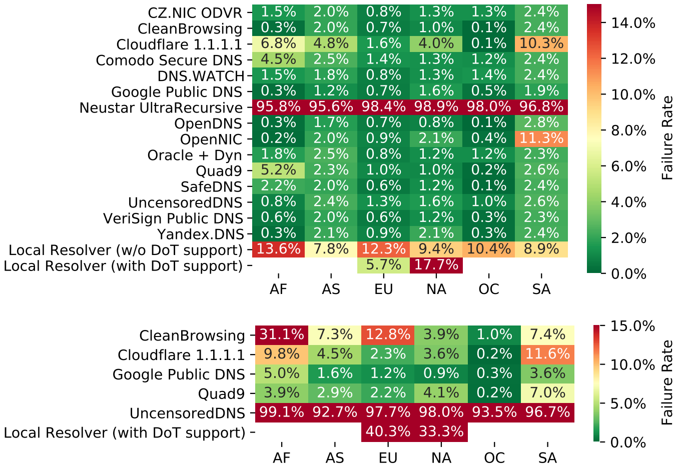

Fig. 1. Failure rates of resolvers by continent for Do53 (top) and DoT (bottom). Each cell represents the failure rate based on all failures relative to all queries for the specific resolver and continent. Most failure rates for Do53 are between 0.3–3%, whereas DoT failure rates are generally higher and more varying.

The inflated failure rates for DoT in comparison with Do53 are less surprising, as DoT was only standardized in 2016 [19]: As such, DoT likely still faces issues with middleboxes along the path [16,29], which intervene with DoT packets (TCP/853) and result in timeouts.

Regional comparison. To identify regional differences, Fig. 1 depicts the failure rates of Do53 (top) and DoT (bottom) by resolver and continent. Most resolvers exhibit similar Do53 failure rates across all continents, in the range of roughly 0.3–3%. Local resolvers show significantly higher failure rates (5.7–13.6%), which means that RIPE Atlas probes have less success in resolving domain names when using their local resolver (regardless of DoT support). Thus, Do53 resolutions are more reliable with public resolvers compared to local ones concerning RIPE Atlas measurements. Nevertheless, we find similarly high values for OpenNIC in SA (11.3%), and Cloudflare in AF (6.8%) and SA (10.3%). As mentioned, Neustar represents an outlier, as measurements fail in nearly all cases (95.6–98.9%). Probes in Oceania (OC) have the lowest failure rates for all resolvers when comparing different continents, with most resolvers having failure rates of at most 0.5%.

Regarding DoT, Google and Quad9 exhibit the lowest failure rates across all continents ($<$ 5% in most continents), although still higher than their respective Do53 failure rates. On the other hand, Cloudflare and CleanBrowsing show higher failure rates, especially in AF (9.8% and 31.1%) and SA (11.6% and 7.4%), with CleanBrowsing having a high failure rate in EU (12.8%) as well. Queries to UncensoredDNS fail in nearly all cases (92.7–99.1%). As multiple public DoT resolvers (even those with otherwise reliable services in other continents) have higher failure rates in AF and SA, these regions may be affected more heavily by ossification in terms of middleboxes. Local resolvers with DoT support also show high failure rates, with 40.3% in EU, and 33.3% in NA. In total, this indicates that the DoT reliability is highly dependent on the geographical location as well as the chosen DNS service.

5 Response Times

We aggregate the measurements by grouping distinct tuples of probe and resolver and, for each group, determine the 5th percentile in terms of response time (i.e., one value for each probe-resolver tuple across all measurements). We choose 5th percentiles to limit the analysis to responses for cached records, as those accumulate at the lower end of the distribution and represent best-case scenarios.

Background. Before discussing response times of the measurements, we elaborate on a technical limitation regarding DoT: By design, a DoT client would first establish a TCP connection and TLS session with the recursive resolver, then keep this session alive to reuse it for resolutions of multiple domains. Thus, the added delay due to the TCP and TLS handshake RTTs only apply once for as long as the connection and session stay alive. For RIPE Atlas probes, however, DoT measurements do *not* keep the connection/session alive in between different measurements, which means that the additional RTTs required for the TCP and TLS handshakes apply to every DoT measurement. We contacted the RIPE Atlas support regarding specific protocol details: RIPE Atlas probes do not use TCP Fast Open or other extensions, so establishing the TCP connection will add 1 RTT to the response time. Further, probes typically use TLS 1.2 (2 additional RTTs), though some probes may use TLS 1.3 (1 additional RTT); however, the DoT measurement results do unfortunately not provide any information about the used TLS version for validation. As such, DoT measurements include 3 additional RTTs (2 in the best case) on top of the DNS lookup (1 RTT).

Considering we focus on cached responses (5th percentiles, see above) exclusively in this section, we argue that the lookup times are negligibly small (since results are simply returned from the cache). Thus, the response times largely consist of the RTTs between probe and resolver. Consequently, Do53 measurements resemble roughly 1 RTT, which we consider as the baseline RTT (cf. overall response times below), whereas DoT measurements resemble roughly 4 RTTs in total, plus time for connection/session management and processing on both probe and resolver. For approximation, we calculate the ratio between the 5th percentiles of the DoT and Do53 response times per probe for each resolver, shown in Fig. 2; the vertical dashed lines represent the outlined ratio of 4 RTTs to 1 RTT (i.e., DoT to Do53).

The minimum ratio across all resolvers is 3.11, which suggests usage of TLS 1.3 in these cases (1 RTT less than with TLS 1.2). Yet, these cases are rare (only four probe-resolver pairs), as the median ratio among the public resolvers is 10.5 (25th percentile 7.5); this suggests that besides the approx. 4 RTTs required for the handshakes, most samples require at least around 4 more RTTs

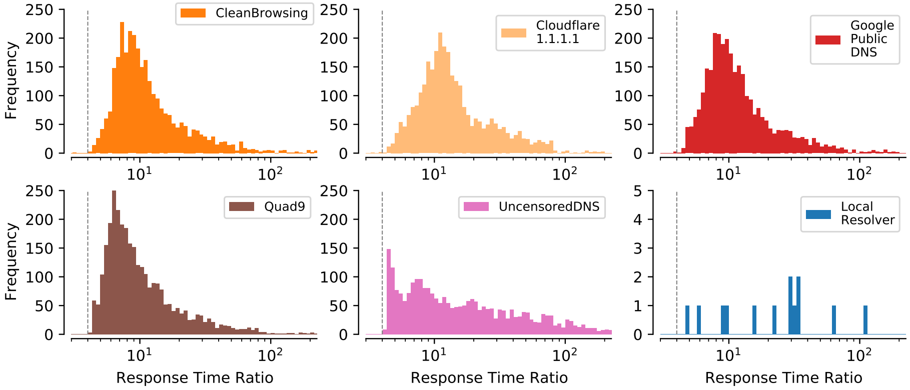

Fig. 2. Histograms of response time ratios (DoT to Do53) per probe for each resolver. The vertical dashed line represents the ratio of 4 RTTs for DoT (TCP handshake + TLS handshake + DNS lookup) to 1 RTT of Do53 (DNS lookup).

for processing of the DoT request on probe and resolver side. However, this processing overhead for DoT measurements cannot be accurately determined, as probes record the total response time only and, therefore, do not allow separation of different steps during the DoT lookup. Nevertheless, note that the handshake RTTs still account for a large fraction of the measured DoT response times overall. Recall that only 13 probes leverage DoT-supporting local resolvers, most of which have ratios toward the higher end (see Fig. 2, bottom right) due to very low Do53 response times (<10 ms) and likely early-stage DoT implementations.

Due to these limitations (also see § 7), the following analyses describe the DoT response times as measured by RIPE Atlas, i.e., incl. TCP/TLS handshakes; observed inflations will only apply when initiating connections to DoT resolvers and, thus, represent upper bounds of response times for cached records.

Overall response times. The distributions of the 5th percentile response times for Do53 are shown in Fig. 3 (left). The fastest resolvers with medians of less than 15 ms are Neustar (median 2.4 ms), local resolvers (9.3 ms), Cloudflare (10.8 ms), and Google (12.6 ms). However, note that the sample size of Neustar measurements is much lower due to its high failure rate (see § 4). Public resolvers that primarily serve clients of a specific country such as CZ.NIC (CZ, 41.2 ms) and Yandex (RU, 51.8 ms), as well as UncensoredDNS (44.9 ms) show response times toward the higher end. The remaining resolvers have response times in between (16–31.3 ms) over Do53.

On the other hand, response times for DoT (see Fig. 3, right) are much higher in comparison with Do53, as expected considering the additional RTTs. The medians for Google (129.3 ms), Cloudflare (131.9 ms), and local resolvers (147 ms) are in the same range of roughly 130–150 ms, whereas Quad9 (170.4 ms) and CleanBrowsing (227 ms) show higher response times, which indicates response time inflations of 150–200 ms when compared to Do53. The median for Uncen-

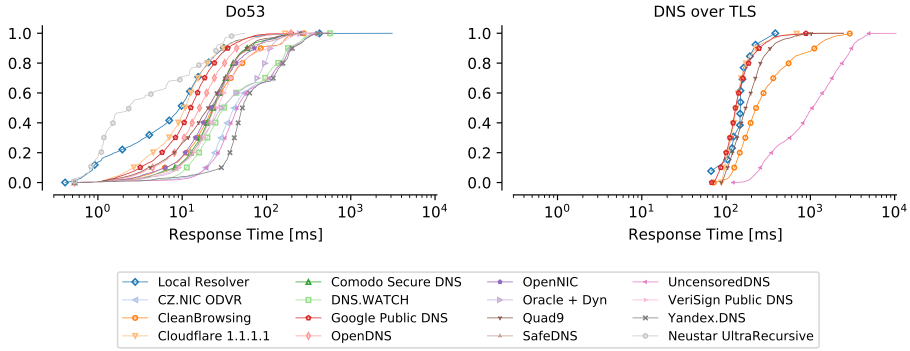

Fig. 3. CDF of resolver response time for successful Do53 (left) and DoT (right) requests (5th percentiles per probe). While most Do53 responses arrive within roughly 100 ms, the majority of DoT responses require more than 100 ms to return.

soredDNS is an outlier at 1.06 seconds; coupled with its high DoT failure rate, the measurements suggest that UncensoredDNS is less suitable as a DoT resolver at this stage. Despite the low support of DoT by local resolvers, the response times are comparable to (and in some cases even better than) well-known public resolvers such as Google, Cloudflare, and Quad9.

Regional comparison. Fig. 4 shows response times for each resolver and continent for Do53 (top) and DoT (bottom); each cell represents the median value for the respective continent-resolver pair, with the sample values being the 5th percentiles of the response times from Fig. 3.

For Do53, we observe that the lowest delays are measured in EU, where the responses arrive within 43.4 ms for all resolvers. For other continents, we see occasionally higher response times, especially in AF, Asia (AS), OC, and SA, where some resolvers take more than 100 ms (up to 339.2 ms) to respond to a Do53 request. Local resolvers exhibit the lowest response times by far, with values ranging between 7.1–12.4 ms, similar to Google (10.2–23.4 ms); again, note that Neustar shows very low response times but is not fully comparable due to its lower sample size. Overall, we observe that the performance of well-known resolvers (Google, Quad9, Cloudflare) is consistent when comparing response times between different continents, i.e., regional differences for resolvers are marginal, while for other resolvers (with fewer points of presence around the globe) regional differences are higher.

Considering DoT (Fig. 4 bottom), we again find response times to be substantially higher than their Do53 counterparts for all cells. However, differences between the continents are much more varying compared to Do53, with the exception of Cloudflare which shows the least varying median response times (128.1–147.7 ms) across all continents. On the other hand, samples for Google are in between 122.9–315.1 ms (showing high response times in AF and OC), which is comparable to DoT-supported local resolvers in EU and NA (148.1 and

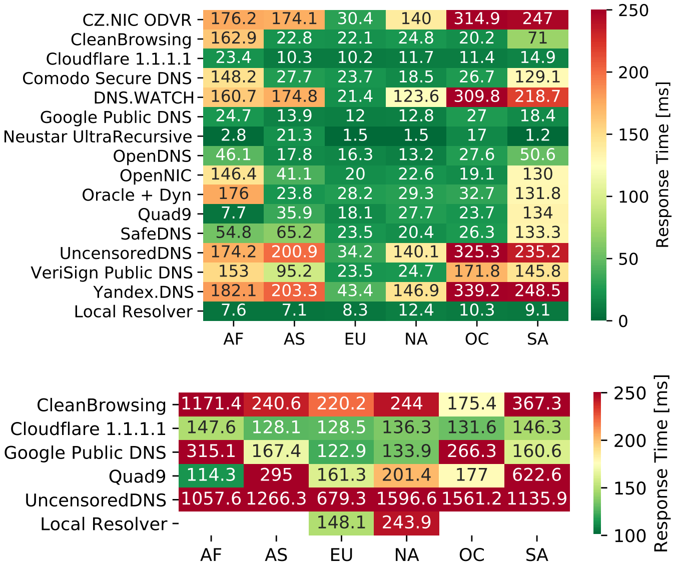

Fig. 4. Medians of the 5th percentile response times by continent and resolver for Do53 (top) and DoT (bottom). Do53 response times are mostly below 20–40 ms for most resolvers, whereas DoT response times are between roughly 120–180 ms instead.

243.9 ms). Quad9 (114.3–622.6 ms) and CleanBrowsing (175.4–1,171.4 ms) show higher variance across the regions; responses from UncensoredDNS even require more than 1 second in most cases. Overall, response times for DoT are much more varying across different continents when compared to Do53.

Response time inflations by individual probes. To further investigate the actual difference between Do53 and DoT in terms of response time, we only consider resolvers that offer both protocols in the following. We calculate the individual deltas between Do53 and DoT for each probe-resolver tuple (i.e., the probe's inflation in response time to a specific resolver) by subtracting the 5th percentile of a probe's DoT response times from the 5th percentile of its Do53 response times.

We find all deltas to be negative, which means that DoT is slower than Do53 in all cases. We observe the lowest inflations regarding response times to be around 66 ms (i.e., delta of -66 ms), although the interquartile range across all samples is [-285.6; -114.8] ms. The medians of the deltas are highly varying across the continents (EU -145 ms, NA -164.9 ms, OC -188.4 ms, AS -234.4 ms, SA -330.5 ms, AF -367.3 ms). Regarding resolvers, Google (median -115.9 ms), Cloudflare (-121 ms), local resolvers (-143.8 ms), and Quad9 (-149.8 ms) show similar inflations in the range of roughly 120–150 ms; on the other hand, CleanBrowsing (-202.8 ms) and UncensoredDNS (-910.3 ms) exhibit much higher response time differences between Do53 and DoT.

Overall, while the observed overheads of DoT differ depending on continent and resolver, we still see differences of more than 100 ms for almost all samples in favor of Do53.

Local DoT resolvers. To further examine local resolvers, we split the measurement of local resolvers with DoT support by individual resolvers. The 9 local resolvers that support DoT are operated by larger commercial ISPs, smaller associations that offer Internet services, cloud/DNS service providers, and academic institutions. However, note that they are only used by 13 probes (11 EU, 2 NA) in our study; DoT is not supported by *any* local probe resolver in AF, AS, OC, or SA. We find varying DoT response times for the different local resolvers in the range of 66.4–383.8 ms overall. XS4ALL (an ISP from NL) shows consistent response times (145.9–156.6 ms) for the five corresponding probes. Further, most of the remaining local resolvers respond within 104–223.2 ms; as such, the DoT response times of local resolvers are largely on par with those of public resolvers.

6 Related Work and Discussion

We contrast our results with those of recent studies: Deccio and Davis [8] find that DoT is supported by 0.15% (1.7k) of all publicly routable IPv4 resolvers, with most of them being assigned to CleanBrowsing (among some resolvers from Cloudflare, Google, and Quad9); our repetition of the experiment reveals that this number has increased by 23.1% within nine months (see § 2).

Lu *et al.* [26] find a similar number of open DoT resolvers (1.5k) and measure response times for DoT and DoH from two residential proxy networks, covering 123k vantage points in total (30k global, 85k in China). In terms of reachability, 99% of the global users in their study can reach a DoT resolver. In their example, Cloudflare is reachable by 98.9% of the users due to the DoT failure rate being 1.1% only; for our results, we observe Cloudflare to fail in 3.0% for all DoT measurements, whereas Google only fails in 1.3%, ultimately resulting in roughly similar numbers in terms of reachability. However, they find much lower failure rates for Quad9 (0.15%, compared to our 2.7%). To contrast this with DoH, they find DoH failure rates of less than 1% from their global proxy network; overall, they observe DoH to have about equal or higher reachability than DoT.

Regarding response times, they find median response times for DoT without connection reuse to range between 349–1,106 ms based on location for Cloudflare, Google, and Quad9 resolvers; this includes overheads for TLS session negotiation, which are in the range of 77–470 ms. These response times are higher by as much as factors of 1.75–5.5 compared to the DoT response times (§ 5) of our RIPE Atlas measurements (median of all probe-resolver response time medians at 201 ms). This indicates that the residential proxy networks add a significant amount of latency to the measurements, which does not reflect the actual response times for home users. Nevertheless, the authors [26] find that connection reuse improves the average response times substantially. This suggests that our measurements represent a rough upper bound for the average DoT response times of home users.

Hounsel *et al.* [17] measure Do53, DoT, and DoH from five global vantage points through Amazon EC2 instances, using Cloudflare, Google, and Quad9. They compare the effects of the different DNS protocols on loading times of webpages and take advantage of the aforementioned connection reuse. For their DoT queries from Frankfurt (FRA), they observe most responses to return within 100 ms for Google and Cloudflare, although results for Quad9 are much more varying (only around 20% within 100 ms). These numbers are much lower compared to the RIPE Atlas 5th percentiles of roughly 130–150 ms that we discuss (§ 5), although this difference is likely related to the connection/session reuse as well as usage of well-provisioned data centers as vantage points (rather than home networks). Nevertheless, while DoT and DoH response times for individual queries are higher compared to Do53, the overall page loading times are lower when reusing the connection and session, showing that a switch from Do53 to DoT or DoH might be beneficial in terms of response times already.

7 Limitations and Future Work

We restrict the set of probes to home and V3 probes exclusively; note that these probes are deployed in 1.1k different ASes, with the top 10 ASes (0.9%) accounting for roughly 27.6% of all home probes. Although there is a potential bias toward overrepresented ASes, we decide not to normalize by ASes since network conditions and, hence, measurements are not guaranteed to be uniform across an AS either: Sampling "representative" probes for each AS would, therefore, introduce another bias into the dataset and analysis.

Furthermore, we cannot directly control the caching behavior of the measured resolvers, though the 200 selected domains are likely cached due to being highly ranked in Alexa Toplists and repeated measurements. Regarding response times, we further limit the analysis to the 5th percentiles for each probe. Note that measurements over RIPE Atlas cannot be guaranteed to run simultaneously or back-to-back due to scheduling and load balancing on the probe. Therefore, we cannot (for instance) pair Do53 and DoT measurements for a head-to-head comparison, and instead rely on the entire distribution (reliability, § 4) and 5th percentiles (response times, § 5) of the measurements.

Moreover, as RIPE Atlas does not keep the TLS session alive for reuse between different measurements, the presented response times represent the initial delays for the first DNS request. Thus, they estimate the upper bounds for DoT response times which end users would experience since subsequent DNS requests through the same TLS session do not require additional handshakes and will have lower response times as a result. Further, applications typically resolve multiple domains concurrently in real use cases, while measurements from RIPE Atlas are performed sequentially.

In the future, we plan to study the impact of different TLS versions, or the benefit of TLS session reuse, but also to study changes over time by repeating the measurements, including measurements over IPv6. To further investigate issues with middleboxes, `traceroute` measurements over UDP/`53` and TCP/`853` can complement the failure analysis of DNS requests by comparison to see where packets are dropped in the network. With the increasing adoption of DNSSEC and larger DNS responses, DNS measurements over TCP/53 can provide further insight about the adoption, reliability, and response times of DNS over TCP. Lastly, DoH measurements (which are not yet possible with RIPE Atlas) from home networks can contribute to ongoing research, as response times and reliability of DoH from the edge have not been widely studied yet.

8 Conclusion

We present first measurement results that compare Do53 and DoT w.r.t. reliability and response times in the context of residential networks, based on 90M domain lookups over both protocols from 3.2k RIPE Atlas home probes. We study the support of DoT among the local resolvers of the probes, finding that only 13 probes (i.e., 0.4%) have DoT-capable local resolvers, which indicates that the adoption of DoT is still very low. When comparing the failure rates for resolvers that respond to both Do53 and DoT queries, we observe that the DoT failure rate is higher by 0.4–32.2 percentage points for these resolvers. In particular, the majority of failures occurs due to timeouts, which is likely seen due to middleboxes that drop packets associated with DoT on port 853. In terms of response times, we find that DoT is slower by more than 100 ms (in a large part due to connection and session establishment), with response times between 130–150 ms for the fastest resolvers and up to 230 ms when including slower ones. Although the support of DoT among local resolvers is low, some local resolvers achieve similar DoT response times (140–150 ms) to the faster public resolvers. Local resolvers further have the lowest latency over Do53, however, both their Do53 and DoT failure rates are higher compared to public resolvers.

With increasing support of DoT among mobile devices as shown by Android [24] and Apple [38], increasing support by local resolvers is important and necessary to avoid centralization of DNS traffic [27] to third parties besides the ISP: Although this can be worked around by cycling through several resolvers [12], this comes at the cost of higher resolution times (especially due to multiple connection and session establishments). As such, to reduce the information leakage through DoT [18] to additional parties while also keeping resolution times low, it is crucial for local resolvers to adopt encrypted DNS and be discoverable within home networks [6]; as seen, DoT response times are comparable between local and public resolvers.

Considering the issues with inflated failure rates for DoT due to ossification, one question that arises is whether to switch the development and deployment focus to DoH [13,5] instead: Just like HTTPS, DoH runs over TCP/443, which will make middlebox issues along the path less likely. Further, popular Web browsers such as Chrome [37] and Firefox [9] already support DoH. However, studies [34] have shown that DoH is more susceptible to fingerprinting attacks than DoT, and further drives centralization of DNS traffic [4,25,12,27]. As both DoT and DoH bring latency overheads, DNS over QUIC [20] might be another encrypted alternative with response times which are closer to Do53. Yet, legislation may discourage and hinder the deployment of encrypted DNS and similar protocols beyond the area of jurisdiction [21]. Thus, further advances and future follow-up studies on encrypted DNS are required to get a better understanding.

Acknowledgements. We thank Alexander Niedrist (TUM), Johan ter Beest and Philip Homburg (RIPE NCC), and the volunteering RIPE Atlas probe hosts for their valuable support regarding our measurement study. We also thank our shepherd Timm Böttger and the anonymous reviewers for their insightful feedback and suggestions.

References

1. Bajpai, V., Brunström, A., Feldmann, A., Kellerer, W., Pras, A., Schulzrinne, H., Smaragdakis, G., Wählisch, M., Wehrle, K.: The Dagstuhl Beginners Guide to Reproducibility for Experimental Networking Research. Computer Communication Review (CCR) **49**(1), 24–30 (2019), https://doi.org/10.1145/3314212.3314217
2. Bajpai, V., Eravuchira, S.J., Schönwälder, J.: Lessons Learned From Using the RIPE Atlas Platform for Measurement Research. Computer Communication Review (CCR) **45**(3), 35–42 (2015), https://doi.org/10.1145/2805789.2805796
3. Bajpai, V., Eravuchira, S.J., Schönwälder, J., Kisteleki, R., Aben, E.: Vantage Point Selection for IPv6 Measurements: Benefits and Limitations of RIPE Atlas Tags. In: Symposium on Integrated Network and Service Management (IM). pp. 37–44. IEEE (2017), https://doi.org/10.23919/INM.2017.7987262
4. Bertola, V.: Recommendations for DNS Privacy Client Applications. Internet-Draft draft-bertola-bcp-doh-clients-01 (Sep 2019), https://datatracker.ietf.org/doc/html/draft-bertola-bcp-doh-clients-01, Work in Progress
5. Böttger, T., Cuadrado, F., Antichi, G., Fernandes, E.L., Tyson, G., Castro, I., Uhlig, S.: An Empirical Study of the Cost of DNS-over-HTTPS. In: Internet Measurement Conference (IMC). pp. 15–21. ACM (2019), https://doi.org/10.1145/3355369.3355575
6. Boucadair, M., Reddy.K, T., Wing, D., Cook, N.: DHCP and Router Advertisement Options for Encrypted DNS Discovery within Home Networks. Internet-Draft draft-btw-add-home-09 (Sep 2020), https://datatracker.ietf.org/doc/html/draft-btw-add-home-09, Work in Progress
7. Cho, K., Mitsuya, K., Kato, A.: Traffic Data Repository at the WIDE Project. In: USENIX Annual Technical Conference (ATC), Freenix Track. pp. 263–270. USENIX (2000), http://www.usenix.org/publications/library/proceedings/usenix2000/freenix/cho.html
8. Deccio, C.T., Davis, J.: DNS Privacy in Practice and Preparation. In: Conference on Emerging Networking Experiments And Technologies (CoNEXT). pp. 138–143. ACM (2019), https://doi.org/10.1145/3359989.3365435
9. Deckelmann, S.: Mozilla Blog: Firefox continues push to bring DNS over HTTPS by default for US users (02 2020), https://blog.mozilla.org/blog/2020/02/25/firefox-continues-push-to-bring-dns-over-https-by-default-for-us-users/, [accessed 2021-Jan-13]
10. Greschbach, B., Pulls, T., Roberts, L.M., Winter, P., Feamster, N.: The Effect of DNS on Tor's Anonymity. In: Network and Distributed System Security Sympo-

sium (NDSS). ISOC (2017), https://www.ndss-symposium.org/ndss2017/ndss-2017-programme/e-effect-dns-tors-anonymity/

11. Herrmann, D., Banse, C., Federrath, H.: Behavior-based Tracking: Exploiting Characteristic Patterns in DNS Traffic. Computers & Security (COSE) **39**, 17–33 (2013), https://doi.org/10.1016/j.cose.2013.03.012
12. Hoang, N.P., Lin, I., Ghavamnia, S., Polychronakis, M.: K-resolver: Towards Decentralizing Encrypted DNS Resolution. In: Workshop on Measurements, Attacks, and Defenses for the Web (MADWEB) (2020), https://doi.org/10.14722/madweb.2020.23009
13. Hoffman, P.E., McManus, P.: DNS Queries over HTTPS (DoH). RFC **8484**, 1–21 (2018), https://doi.org/10.17487/RFC8484
14. Holterbach, T., Pelsser, C., Bush, R., Vanbever, L.: Quantifying Interference between Measurements on the RIPE Atlas Platform. In: Internet Measurement Conference (IMC). ACM (2015), https://doi.org/10.1145/2815675.2815710
15. Holz, R., Hiller, J., Amann, J., Razaghpanah, A., Jost, T., Vallina-Rodriguez, N., Hohlfeld, O.: Tracking the deployment of TLS 1.3 on the Web: A story of experimentation and centralization. Computer Communication Review (CCR) **50**(3), 3–15 (2020), https://doi.org/10.1145/3411740.3411742
16. Honda, M., Nishida, Y., Raiciu, C., Greenhalgh, A., Handley, M., Tokuda, H.: Is it Still Possible to Extend TCP? In: Internet Measurement Conference (IMC). pp. 181–194. ACM (2011), https://doi.org/10.1145/2068816.2068834
17. Hounsel, A., Borgolte, K., Schmitt, P., Holland, J., Feamster, N.: Comparing the Effects of DNS, DoT, and DoH on Web Performance. In: The Web Conference (WWW). pp. 562–572. ACM / IW3C2 (2020), https://doi.org/10.1145/3366423.3380139
18. Houser, R., Li, Z., Cotton, C., Wang, H.: An Investigation on Information Leakage of DNS over TLS. In: Conference on Emerging Networking Experiments And Technologies (CoNEXT). pp. 123–137. ACM (2019), https://doi.org/10.1145/3359989.3365429
19. Hu, Z., Zhu, L., Heidemann, J.S., Mankin, A., Wessels, D., Hoffman, P.E.: Specification for DNS over Transport Layer Security (TLS). RFC **7858** (2016), https://doi.org/10.17487/RFC7858
20. Huitema, C., Mankin, A., Dickinson, S.: Specification of DNS over Dedicated QUIC Connections. Internet-Draft draft-ietf-dprive-dnsoquic-01 (Oct 2020), https://datatracker.ietf.org/doc/html/draft-ietf-dprive-dnsoquic-01, Work in Progress
21. Internet Society: Internet Society: Russia's Proposal Would Weaken the Internet, Make It Less Secure (Sep 2020), https://www.internetsociety.org/news/statements/2020/internet-society-russias-proposal-would-weaken-the-internet-make-it-less-secure/, [accessed 2021-Jan-13]
22. Kirchler, M., Herrmann, D., Lindemann, J., Kloft, M.: Tracked Without a Trace: Linking Sessions of Users by Unsupervised Learning of Patterns in Their DNS Traffic. In: Workshop on Artificial Intelligence and Security (AISec). pp. 23–34. ACM (2016), https://doi.org/10.1145/2996758.2996770
23. Klein, A., Pinkas, B.: DNS Cache-Based User Tracking. In: Network and Distributed System Security Symposium (NDSS). ISOC (2019), https://www.ndss-symposium.org/ndss-paper/dns-cache-based-user-tracking/
24. Kline, E., Schwartz, B.: DNS over TLS support in Android P Developer Preview (2018), https://android-developers.googleblog.com/2018/04/dns-over-tls-support-in-android-p.html, [accessed 2021-Jan-13]

25. Livingood, J., Antonakakis, M., Sleigh, B., Winfield, A.: Centralized DNS over HTTPS (DoH) Implementation Issues and Risks. Internet-Draft draft-livingood-doh-implementation-risks-issues-04 (Sep 2019), https://datatracker.ietf.org/doc/html/draft-livingood-doh-implementation-risks-issues-04, Work in Progress
26. Lu, C., Liu, B., Li, Z., Hao, S., Duan, H., Zhang, M., Leng, C., Liu, Y., Zhang, Z., Wu, J.: An End-to-End, Large-Scale Measurement of DNS-over-Encryption: How Far Have We Come? In: Internet Measurement Conference (IMC). pp. 22–35. ACM (2019), https://doi.org/10.1145/3355369.3355580
27. Moura, G.C.M., Castro, S., Hardaker, W., Wullink, M., Hesselman, C.: Clouding up the Internet: how centralized is DNS traffic becoming? In: Internet Measurement Conference (IMC). pp. 42–49. ACM (2020), https://doi.org/10.1145/3419394.3423625
28. Nottingham, M.: The Internet is for End Users. RFC **8890**, 1–10 (2020), https://doi.org/10.17487/RFC8890
29. Papastergiou, G., Fairhurst, G., Ros, D., Brunström, A., Grinnemo, K., Hurtig, P., Khademi, N., Tüxen, M., Welzl, M., Damjanovic, D., Mangiante, S.: De-Ossifying the Internet Transport Layer: A Survey and Future Perspectives. Communications Surveys & Tutorials (COMST) **19**(1), 619–639 (2017), https://doi.org/10.1109/COMST.2016.2626780
30. Rekhter, Y., Moskowitz, B.G., Karrenberg, D., de Groot, G.J., Lear, E.: Address Allocation for Private Internets. RFC **1918**, 1–9 (1996), https://doi.org/10.17487/RFC1918
31. Rescorla, E.: The Transport Layer Security (TLS) Protocol Version 1.3. RFC **8446**, 1–160 (2018), https://doi.org/10.17487/RFC8446
32. RIPE NCC: RIPE Atlas: A Global Internet Measurement Network. In: Internet Protocol Journal (IPJ) (2015), http://ipj.dreamhosters.com/wp-content/uploads/2015/10/ipj18.3.pdf
33. Scheitle, Q., Hohlfeld, O., Gamba, J., Jelten, J., Zimmermann, T., Strowes, S.D., Vallina-Rodriguez, N.: A Long Way to the Top: Significance, Structure, and Stability of Internet Top Lists. In: Internet Measurement Conference (IMC). pp. 478–493. ACM (2018), https://doi.org/10.1145/3278532.3278574
34. Siby, S., Juárez, M., Díaz, C., Vallina-Rodriguez, N., Troncoso, C.: Encrypted DNS ⇒ Privacy? A Traffic Analysis Perspective. In: Network and Distributed System Security Symposium (NDSS). ISOC (2020), https://www.ndss-symposium.org/wp-content/uploads/2020/02/24301-paper.pdf
35. Sood, P., Hoffman, P.E.: Upgrading Communication from Stub Resolvers to DoT or DoH. Internet-Draft draft-pp-add-stub-upgrade-02 (Jun 2020), https://datatracker.ietf.org/doc/html/draft-pp-add-stub-upgrade-02, Work in Progress
36. Sun, M., Xu, G., Zhang, J., Kim, D.W.: Tracking You through DNS Traffic: Linking User Sessions by Clustering with Dirichlet Mixture Model. In: Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems. pp. 303–310. ACM (2017), https://doi.org/10.1145/3127540.3127567
37. The Chromium Projects: DNS over HTTPS (aka DoH): Auto-upgrade project (2020), https://www.chromium.org/developers/dns-over-https, [accessed 2021-Jan-13]
38. WWDC 2020 - Apple Developer: Enable encrypted DNS (2020), https://developer.apple.com/videos/play/wwdc2020/10047, [accessed 2021-Jan-13]