Large Scale Analysis of DoH Deployment on the Internet

Sebastián García[1], Joaquín Bogado[1(✉)], Karel Hynek[2,3], Dmitrii Vekshin[4], Tomáš Čejka[2,3], and Armin Wasicek[4]

[1] Faculty of Electrical Engineering, Czech Technical University in Prague, Prague, Czech Republic
sebastian.garcia@agents.fel.cvut.cz, joaquin.bogado@aic.fel.cvut.cz
[2] Faculty of Information Technology, Czech Technical University in Prague, Prague, Czech Republic
hynekkar@fit.cvut.cz
[3] CESNET, z. s. p. o., Prague, Czech Republic
cejkat@cesnet.cz
[4] Avast Software s.r.o., Prague, Czech Republic
{dmitrii.vekshin,armin.wasicek}@avast.com

Abstract. DNS over HTTPS (DoH) is one of the standards to protect the security and privacy of users. The choice of DoH provider has controversial consequences, from monopolisation of surveillance to lost visibility by network administrators and security providers. More importantly, it is a novel security business. Software products and organisations depend on users choosing well-known and trusted DoH resolvers. However, there is no comprehensive study on the number of DoH resolvers on the Internet, its growth, and the trustworthiness of the organisations behind them. This paper studies the deployment of DoH resolvers by (i) scanning the whole Internet for DoH resolvers in 2021 and 2022; (ii) creating lists of well-known DoH resolvers by the community; (iii) characterising what those resolvers are, (iv) comparing the growth and differences. Results show that (i) the number of DoH resolvers increased 4.8 times in the period 2021–2022, (ii) the number of organisations providing DoH services has doubled, and (iii) the number of DoH resolvers in 2022 is 28 times larger than the number of well-known DoH resolvers by the community. Moreover, 94% of the public DoH resolvers on the Internet are unknown to the community, 77% use certificates from free services, and 57% belong to unknown organisations or personal servers. We conclude that the number of DoH resolvers is growing at a fast rate; also that at least 30% of them are not completely trustworthy and users should be very careful when choosing a DoH resolver.

Keywords: DoH · Encrypted DNS · Network measurement · Network trends

V. Atluri et al. (Eds.): ESORICS 2022, LNCS 13556, pp. 145–165, 2022.
https://doi.org/10.1007/978-3-031-17143-7_8

1 Introduction

DNS over HTTPS (DoH) is a method of encrypting DNS [26] that has been in continuous deployment since 2017 [18]. Despite controversies over its impact on privacy and surveillance monopoly [14,16,25], many applications currently implement DoH and the transition to encrypted DNS is well underway.

Encrypted DNS is a fundamental part of our security and privacy, and DoH has emerged, together with DNS over TLS (DoT), as a standard for the community. With *standard* non-encrypted DNS, the decision about which DNS server to use was mainly based on performance. With DoH, users need to take into account other aspects of security, namely the capability to encrypt up to the DNS resolver vs. to the authoritative DNS resolvers, the ability to filter out the protocol, and the loss of visibility for lawful blocking.

Although some measurements on DoT adoption were made [10,27,35], there has not been large-scale measurements on DoH deployment. Most studies focus on DoT because by using port 853/TCP it is easy to find. As part of the advantages of DoH, the use of port 443/TCP makes it difficult to differentiate from web pages. Therefore, there is a lack of visibility on the amount of DoH resolvers, their features, and the type of organisations that deployed them. Without this knowledge, the security community lacks some perspective on the security of DoH.

This paper presents the first longitudinal measurement, comparison, and analysis of the deployment of DoH resolvers on the Internet from 2021 to 2022. We scanned the Internet for port 443/TCP, identified DoH resolvers, compiled a list of well-known DoH resolvers by the community, and verify the trustworthiness of the resolvers.

Results show a confirmed growing trend in the deployment of DoH between 2021 and 2022. The number of well-known DoH resolvers by the community increased from 234 to 262 (∼12%). The number of public DoH resolvers found on the Internet shows at least 350% increase in 2022, even when the difference in methodology between the two scans is taken into account. This is 28 times larger than the list of 262 well-known DoH resolvers of 2022, meaning that ∼94% of the public DoH resolvers are unknown to the community.

The contributions of this paper are (i) an updated and comprehensive dataset of well-known DoH providers by the community in 2021 and 2022, (ii) a dataset of all public DoH resolvers found by our global Internet scan, (iii) a new Nmap NSE script tool to scan and verify DoH resolvers, and (iv) a security overview of the organisations providing DoH resolution services.

2 Related Work

DoH is a relatively new protocol for encrypting DNS, already studied from multiple perspectives, such as performance [4], privacy [20] and deployment differences [24]. Since DoH shares port 443/TCP with the rest of HTTPS traffic, many studies tried to detect DoH in the network. Vekshin et al. [39] used a machine learning detection algorithm to detect DoH with 99% accuracy. With MontazeriShatoori et al. [30] achieving similar performance. However, these approaches

focused only on web browser traffic streams. None of these techniques works for a single DoH traffic query.

The detection of DoH *in general* is still an unsolved challenge. Furthermore, some well-established security software and appliances [9,36] rely on DNS queries to lawfully block access to certain sites at the host, enterprise, or ISP level. These security software work by filtering and blocking DNS based on rules. As DoH allows to bypass these network-based filters, security software can only block DoH by relying on domains and IP address [37]. The importance of accessing a comprehensive list of well-known DoH resolvers is then paramount for the correct functioning of this type of system.

DoH abuse was surveyed by Hynek et al. [22]. According to their study, DoH is already misused by malware creators and rogue users to hide their activities from network security defences. Furthermore, Hynek et al. defined several research challenges that need to be addressed to maintain network security at the current level. However, since these challenges are still not solved, the mass deployment of DoH has a significant impact on network security.

Deccio et al. [8] studied in 2019, the adoption of DoT and DoH by open resolvers. Their results show that the adoption was quite poor: From ∼1.2 million open DNS resolvers found on the Internet, only 9 (0.007%) supported DoH. However, since this study first scanned DNS resolvers and then asked for DoT/-DoH, it missed those resolvers that handle only DoH requests. The study by Lu et al. [27], in 2019, also scanned well-known open DNS resolvers from the Internet and checked their DoH support, finding only 17 DoH resolvers.

The previous techniques for finding DoH resolvers are insufficient to accurately estimate the population of DoH resolvers on the Internet. Contrary to previous studies, we searched the entire IPv4 Internet address space looking for DoH capable resolvers. Our measurement also found DoH-capable open DNS resolvers, which are not publicly known. To the best of our knowledge, no previous research has tackled the measurement of the DoH resolvers population across the Internet.

3 Background on DoH and Its Security Impact

The design of the DNS over HTTPS (DoH) protocol started in 2017 and was adopted as RFC 8484 [18] in October 2018. Currently, there are two significantly different implementations. The first implementation, compliant with RFC 8484, uses the DNS binary "wireformat" [29] to encapsulate DNS messages in HTTPS (GET or POST methods). The second implementation uses DNS messages encoded in JSON format, as described by RFC 8427 [17]. The JSON data is transferred through the HTTPS GET method. Most global DNS providers support both implementations [24]. However, in practise, all DoH-enabled Web browsers and most other performance-orientated DoH clients use wireformat messages with the HTTPS POST method.

The security community knows that encrypting DNS is one of the most required Internet features to protect user privacy and security. This is because

many surveillance and tracking organisations use DNS traffic to profile and monitor users [15], especially in countries without Internet freedom [5]. However, even though users can encrypt DNS traffic, the choice of DNS provider is still important because that provider will have access to the DNS traffic. Therefore, choosing a DNS provider that is trusted (encrypted or not) is a security decision.

This decision is also important because many protection tools rely on DNS, such as commercial DNS protection companies, DNS filters for policy enforcement in organisations, and antivirus tools. Moreover, many users choose and believe that using a third-party DNS resolver, instead of the DNS server provided by the local network or ISP, can better protect them from surveillance and monitoring [5].

The main difference between choosing a traditional DNS provider and an encrypted DNS provider is that for an encrypted DNS provider, the choice is largely dependent on the threat model of the user [11]. Users who suspect a domestic threat actor may prefer third-party encrypted DNS providers. But such provider may mean a dangerous centralisation of data.

The security problem of centralisation becomes more relevant as more users choose to use a small group of well-known DoH providers. Those providers have a privileged access to DNS requests for profiling and advertising [5]. These few providers are typically big tech giants and telecommunication providers (telcos), and effectively cut off smaller ISPs, small telcos, and even local administrators from accessing DNS. Such a centralisation affects some protection measures and puts our data in the hands of big tech companies.

Another essential aspect of DoH is that applications that enable DoH at the user level (such as a web browser) can bypass the DNS resolution of the Operating System (OS). This design decision was quite controversial, since resolving domain names is an action traditionally left to the OS due to its complexity and dependency on local policies.

Although users can choose any DoH resolver, most use the default settings in the applications. For instance, Firefox (since version 92.0) by default uses Cloudflare Inc.; the same as Opera Browser (from version 79.0.4143.50). Google Chrome offers a selection of five well-known DoH resolvers, but it can also detect if the system-defined DNS server supports DoH [3]. The decision to use the default settings has a double impact; first, it allows DoH to be used quickly and transparently by many users, and second, it allows these organisations to receive DNS requests by default.

One of the more important privacy features of DoH is to use HTTPS on port 443/TCP. This prevents to easily block DoH by using the port number. An alternative approach to block DoH may be to filter the domain in the SNI record using lists of well-known DoH resolvers. However, this filtering can be bypassed by (i) using a not well-known DoH provider, or (ii) by encrypting the SNI as described in the RFC draft [33]. Our research highlights the possibility of finding a not-so-well-known DoH provider.

In this security context, many questions regarding encrypted DNS, and DoH in particular, are asked. Is the number of DoH-enabled DNS servers growing? Who is implementing them? Can organisations successfully filter DoH by blocking the main third-party providers? Is the current centralisation of DoH providers

counterbalanced by new providers? Can users trust small and unknown DoH resolvers?

4 Methodology

The longitudinal analysis is composed of two exploration moments. The first in April 2021 and the second between January and April 2022. Each exploration consisted on the following methodology steps: (i) create a list of well-known DoH resolvers; (ii) scan all the host on the IPv4 Internet looking for servers with port 443/TCP open; (iii) discover which of those IPs are DoH resolvers; (iv) verify that they answer DoH correctly and compile a final list; (v) enrich the IP addresses of the discovered DoH resolvers with information from threat intelligence services; (vi) verify the use of SNI; (vi) estimate the number of organisations providing DoH resolution services.

4.1 Creation of the Well-known DoH Resolvers Lists

Each list of well-known DoH resolvers was created by aggregating all the resolvers available in public lists, reports, documents, and academic papers. The DoH resolvers were verified using our custom Python script described in Subsect. 4.4. There are some lists of DoH resolvers on the Internet, including the AdGuard list [1], and the curl tool list [19]. However, those lists are not comprehensive. The list of 2021 is called Known2021, and the list of 2022 is called Known2022. Both were published for this paper [12,21]. The exact sources used to create them are included in each dataset. Moreover, IP addresses from Known2021 which was working at the time of creation the list of 2022 were also added to the Known2022.

In the Known2022 list, the domain names were given so we decided not to include domain names acquired by reverse DNS queries (PTR). Moreover, reverse DNS domain names may belong to hosting providers or Virtual Private Server (VPS) providers, such as Amazon, Microsoft, or Google, and thus do not provide information relevant to organisation responsible of the DoH resolver.

4.2 Scan of Port 443/TCP on the Internet

We scanned the entire IPv4 address space on the Internet looking for servers with open port 443/TCP. It was done by dividing the IPv4 address space into 255 uniform A-class ranges in order to distribute the load among several scanning nodes. Each range was scanned from a different cloud virtual machine. The masscan tool was used to perform the scan [13] with a fixed rate of 2,000 packets per second. Masscan was also configured to retry each IP address three times. These parameters were chosen to avoid losing packets and connection errors[1].

[1] Masscan command example: `masscan -p 443 --range 20.0.0.0--29.0.0.0 --rate 2000 --retries 3`.

These parameters were both used in 2021 and 2022. Moreover, in both scans we used masscan feature to scanned the IP addresses in random order and limit the amount of packets per second sent to service providers.

4.3 DoH Service Discovery

Once the list of IP addresses with open port 443/TCP was collected, it was necessary to find which ones implemented the DoH protocol. To automate the process, we created a DoH Nmap script [34]. Nmap is a well-known multifunctional network scanner that implements the Nmap Script Engine (NSE) for users to develop their own scripts [28]. Our DoH script checks all six different DoH methods: HTTP/1 with GET, HTTP/1 with POST, HTTP/1 with JSON, HTTP/2 with GET, HTTP/2 with POST, and HTTP/2 with JSON. This scan was executed using the same cloud setup as for the scan of port 443/TCP.

In order to speed up the process, the script only checks the HTTP status code in the response. It does not parse the whole HTTP response, nor does it make any more DNS resolution. Therefore, false positives may occur, which were later filtered in the DoH verification stage 4.4. This verification stage was implemented in a separated script, in order to keep the Nmap script as simple and fast as possible.

The Nmap script sends six DoH requests with a DNS query asking for the `example.com` domain. This domain is managed and recommended by IANA for testing purposes. For all six methods, the script sends the same query endpoint `/dns-query`[2]. This endpoint is specified in RFC 8484 for the HTTP GET and HTTP POST DoH methods. Since the JSON method is not standardised by the RFC, the endpoint of DNS JSON API might differ between providers. However, many well-known providers, such as Cloudflare [7], AhaDNS [2], and Quad9 [32] use the same endpoint as defined in the RFC.

The Nmap parameters used for this stage in 2021 differ from the ones used in 2022. In 2021 we used Nmap with the most aggressive timing template (parameter-T5), allowing for a faster scan. However, this timing template is prone to packet loss, reducing the service discovery efficiency. In 2022, we used the normal Nmap timing template (parameter -T3) in order to obtain higher-quality results, minimising the packet loss. The Appendices Sect. 8.2 shows examples of the Nmap invocations used in both scans.

Therefore, to make a fair comparison between the two scans on the number of computers found, we estimated the number of resolvers lost in 2021. For this we re-scanned all the 2022 DoH resolvers using both timing parameters. Results show that the more aggressive parameters of 2021 indeed caused packet loss and resulted in a smaller number of detected DoH resolvers. From the 4,354 DoH resolvers found with normal timing parameters, the aggressive parameters found between 2,851 and 3,213 in repeated scans. the relative efficiency, then, of the *DoH Service discovery* in 2021 was between 65.5% and 73.8% compared to 2022.

[2] DNS query endpoint example: `https://1.1.1.1/dns-query?name=example.com`.

4.4 DoH Resolver Verification

The list of DoH resolvers found in the previous stage was verified to correctly implement DoH, in order to remove false positives. We implemented a Python script (available in [34]), which tests the correct support of three DoH methods (GET, POST, and JSON) via HTTP/1 and HTTP/2. Contrary to the Nmap DoH script, the Python script can parse the DoH responses and check that they are valid DNS responses. This step filtered out IP addresses that responded "HTTP 200 OK" to DoH requests, but the response did not contain DNS data. The result of this stage is a list of confirmed and validated DoH resolvers and DoH methods that they support. The same verification method was performed for 2021 and 2022.

At the end of this step, and from now on, the verified list of DoH resolvers of 2021 is called Scan2021, and the one from 2022 is called Scan2022.

4.5 IP Address Enrichment

The list of DoH resolvers was further enriched with related information about the discovered IP addresses. The enrichment consists of: (i) the TLS certificates, (ii) information from WHOIS service, (iii) information from VirusTotal threat intelligence feeds including downloaded samples and URLs related to malware samples associated to the IP addresses, (iv) passive DNS data with the referred domain names for the IP, (v) DNS server type, (vi) DNS server version identification, and (vii) information about the web page if there was any. In addition, a *suspicious* flag was included in case the IP address has a high probability of being *relate to a phishing campaign* according to a set of indicators used by the Avast Web Shield feature. This set of indicators consists of keywords, domain name structure, lexical analyses results, domain hosting information, and other indicators.

The information for points (v) and (vi) was obtained using the DoH inherited capabilities of traditional DNS. In DoH, as in DNS, it is possible to create a CHAOS record class with TXT requests and issue it into a `version.bind` query to identify which type of DNS software the server is using. Finally, the TLS certificate data of the DoH resolvers was analysed to detect anomalies, such as expired or self-signed certificates. Given that the IP address enrichment was implemented late in 2021, it was applied only to the DoH resolvers in 2022.

4.6 Verification of SNI Usage

The main limitation of our DoH scan is that it may not find DoH resolvers on servers that host multiple services on the same IP address. In such cases, to be successful, the query needs to send a Server Name Indication (SNI), or `HTTP Host` header, or `HTTP/2 :authority` header. When our DoH scan found an IP with an open 443/TCP port but we could not find its proper domain name, we could not verify whether the DoH resolver works but requires valid SNI only or DoH is not supported at all.

To investigate the severity of this limitation, we estimated how many DoH resolvers were not found by performing a test with the Known2022 list of well-known DoH resolvers, which have a domain name. The methodology was: (i) For each well-known DoH resolver in the Known2022 list with an IPv4 address, get its domain name; (ii) do all the six types of DoH queries providing the SNI, or HTTP/1, or HTTP/2 host headers; (iii) get the IPv4 address for that domain; (iv) do all the six types of DoH queries providing only the IPv4 address, without any SNI or HTTP header. By using these steps we obtained the share of well-known DoH resolvers, that require domain name for successful connection.

4.7 Estimation of the Number of Organisations

To estimate the number of organisations providing DoH resolution services in the Scan2021 and Scan2022 list, the following methodology was used: First, extract the reverse DNS of all the IPs in the Scan2021 and Scan2022 lists. Second, extract the effective second-level domain name for each IP and consider each unique effective second-level domain an organisation. Third, if the effective second-level domain was not available, extract the WHOIS organisation name and consider each *WHOIS organisation* an organisation. Fourth, if the *WHOIS organisation* was not available group the IP addresses by their /16 CIDR and Autonomous System Number (ASN), and consider each unique group as an organisation as used by Deccio et al. [8].

4.8 Methodology Limitations

The used methodology presents limitations that needs to be properly discussed and accounted for proper interpretation of our results. Faster scanning rate used in Scan2021 for DoH Service Discovery stage described in Sect. 4.3, can increase the number of missed hosts (false negatives) in the 2021 results. Therefore, the Scan2021 totals were scaled up to account for the reduced efficiency in Sect. 5.2 for proper comparison. The change in methodology also increases the time needed to finish the Scan2022. However, we argue that each IP address was scanned only once; thus, the longer period does not affect the comparison.

The DoH methods were tested only using the /dns-query API endpoint which is the standard endpoint except for the JSON method (which is not standardised). Therefore, our methodology cannot discover any DoH resolver using other endpoints. Moreover, Internet-wide scans can be blocked by service providers, which reduces efficiency of the scanning over time. Additionally, the methodology could not find resolvers that require domain names (in SNI or HTTP headers) for successful connection. Given that, the methodology described does not produce an exhaustive list of DoH servers. Accounting for these considerations, we say that the amount of DoH servers found in this work can be interpreted as a lower bound.

The IP address enrichment mainly includes information from commercial databases and freely available information found on the servers. The DNS server version was extracted using non-standard DNS requests. Therefore, the results

Table 1. Summary of well-known DoH resolvers in the Known2021 and Known2022 lists.

	Known2021	Known2022	Intersection	Increase
Total Unique Servers UP	234	262	157	11.9%
Total Unique IPv4 Servers	131	144	86	9.9%
Total Unique IPv6 Servers	103	118	78	14.5%
Unique Autonomous Systems	52	59	42	13.4%
Unique Domain Names	110	109	67	−0.1%

showing the DNS server version only include a fraction of found resolvers able to answer those requests.

Given that the number of organisations was inferred using second-level domain names, and that these names can be shared across virtual servers hosted by the same cloud provider, this number should also be considered a lower bound.

5 Results

This section shows the results of the Scan2021 and Scan2022 lists, and a comparison of these results with the well-known lists of DoH resolvers. Then, the estimated number of organisations that provide DoH resolution services is presented. Finally, the results of the threat intelligence feeds associated with the DoH resolvers are shown.

5.1 Results of Creating Well-Known DoH Resolvers Lists

Regarding the creation of DoH resolver lists that are well known by the community, Table 1 shows a summary of the main differences. The total number of well-known DoH resolvers between 2021 and 2022 increased by ∼12%. From the DoH resolvers found in 2021, only ∼67% remained active in 2022 (157 IP addresses from 234 IP addresses). In 2022 there was a ∼10% increase of IPv4 addresses, with ∼65% of them appearing in 2021 and 2022. This means that ∼35% of the IPv4 addresses of well-known DoH providers disappeared in 2022. Similarly, there was an increase of ∼13.4% of unique ASNs, and a ∼14.5% increase in the number of unique IPv6 addresses in 2022.

The number of unique domain names slightly decreased due to the different methodology of their collection. Contrary to the well-known DoH resolver list of 2021, the 2022 list does not contain domain names acquired by reverse DNS queries (PTR) as discussed in Sect. 4.1.

5.2 Results of DoH Scans

The port scan of 2021 found 41,022,969 IP addresses with port 443/TCP open on the Internet. Of these, 930 were verified to be actual DoH resolvers. Given

Table 2. Features of IP addresses of the discovered DoH resolvers.

Feature	Scan2021	Scan2022
Total number of unique IP addresses	930 (100%)	4,354 (100%)
IP addresses with domains	679 (73%)	4,197 (96%)
IP addresses without domains	251 (27%)	149 (3%)
Unique SLD	171	657
Unique /16 prefixes*	115	39
Unique Autonomous System*	72	27
Estimated number of unique providers	243–286	684–696

* Number calculated only from IP addresses for which we could not obtain domain name.

that this scan used a set of aggressive Nmap parameters, thus reducing its efficiency, this number of DoH resolvers could be underestimated. See Sect. 8.2 for a description. Given our tests, it can be concluded that the number of DoH resolvers in April 2021 was actually between 1,173 and 1,241. Our Scan2021 list contains 930 IP addresses.

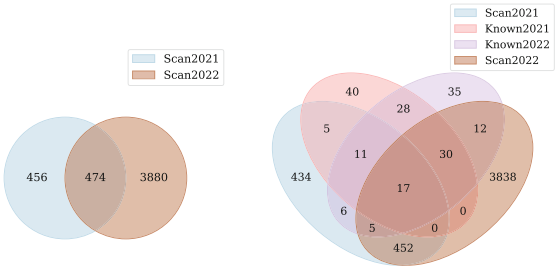
In 2022, the port scan found a total of 36,035,492 IP addresses with port 443/TCP open, which represents 87.84% of the IPs found during 2021. We attribute the smaller amount of IPs to the large variability in Internet scans (packet loss, bandwidth differences, geolocation filters, etc.) and not to an actual decrease of the amount of computers with port 443/TCP open. The number of verified IP addresses of DoH resolvers found during 2022 and contained in the Scan2022 list is 4,354. This number is ∼4.8 times larger than the amount of DoH resolvers of Scan2021.

Table 2 summarises the total number of resolvers discovered in both scans. In Scan2022, we found 4 times more unique IP addresses of DoH resolvers than during Scan2021. Even if the decreased efficiency of Scan2021 during the service discovery stage is taken into account, the difference in the discovered DoH resolvers with Scan2022 is statistically significant with p-value < 0.01. This result is based on a standard two sample one side T-Test for the mean of a distribution [38], and can be interpreted as a true increase in the effective number of public DoH resolvers.

Moreover, the number of organisations providing DoH resolution services in April 2022 is 2.5 times larger than in April 2021. Figure 1a shows that 474 DoH resolvers were found in both scans. However, almost half of the verified DoH resolvers found in Scan2021 were not found in Scan2022. Given that our methodology deals with the number of servers and does not track the DoH resolvers individually, we don't know if this DoH resolvers have been moved to another IP address or ceased operations. On the other hand, the decrease in the number of unique /16 prefixes can be explained by a slight increase in the efficiency of the IP enrichment process.

5.3 Comparison Between the Well-Known and DoH Scan Lists

The distribution of the DoH resolver IP addresses across all lists is shown in Sub-Figure 1b. Reading the figure from top to bottom, we find that 40 addresses



(a) Comparison between first and second scan **(b)** Comparison with Well-known lists

Fig. 1. Venn diagrams of DoH resolver IP addresses distribution.

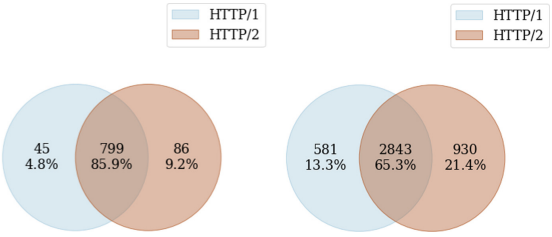
Table 3. Results of SNI Verification for finding DoH resolvers

Connected by	Successful	Successful %
Domain Name	93	100%
IP address	66	71%

are only present in the Known2021 list, 35 only in the Known2022 list, and 28 in both. None of these addresses was found in the Scan2021 or Scan2022. Meanwhile, five IP addresses present in the Known2021 list were only seen in that list, and 12 IPs present in the Known2022 list were only seen in that same list. From all IP addresses present in both well-known lists, 11 were found only during Scan2021, 30 only during Scan2022, and 17 were found during both scans. A total of 434 IP addresses were only found in Scan2021, 3,838 were only found in Scan2022, and 474 were present in both scans. But 452 of these IP addresses were not present in any of the well-known lists. Most DoH resolver IP addresses did not appear on any of the well-known lists. However, the well-known lists are evolving. There are 11 servers that were in Scan2021, which were not present in the Known2021 list but are included in the Known2022 list. However, only five of these servers appear to be still active on the Scan2022. The rest may have been moved to another address or stopped operations.

5.4 Results of the SNI Verification

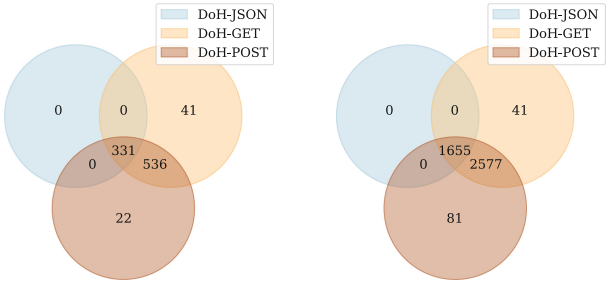
The results of the SNI verification are shown in Table 3. In the Known2022 list of DoH resolvers, there are only 93 that have a domain and an IPv4 address. It can be seen that of those 93, around 30% of the well-known DoH resolvers require an SNI or HTTP header to work successfully. This means that our Scan2021 and Scan2022 of DoH resolvers are a lower bound, and theoretically there could be *at least* 30% more DoH resolvers on the Internet. It should also be considered



(a) HTTP version support in Scan2021

(b) HTTP version support in Scan2022

Fig. 2. Venn diagrams of HTTP version support across DoH resolvers.



(a) DoH methods support in Scan2021

(b) DoH methods support in Scan2022

Fig. 3. Venn diagrams of supported methods across DoH resolvers.

that this test was performed on well-known resolvers, in which the use of an SNI may be different from others.

5.5 Capabilities of the DoH Resolvers Found

Since we have queried each DoH resolver multiple times, we can analyse the methods supported by the DoH resolvers. Figure 2 shows the HTTP version support. It can be seen that most DoH resolvers on both scans support both HTTP versions. In the 2022 scan, we notice an increased share of HTTP/2-only or HTTP/1-only resolvers with respect to the total.

The methods supported in the DoH resolvers are shown in Fig. 3. Most resolvers support the RFC 8484 compliant versions. Some resolvers support only DoH-GET or only DoH-POST, even though the RFC 8484 specifies that the resolver must implement both methods. The IP addresses of those DoH resolvers supporting only DoH-GET are the same in both scans. The JSON approach is supported by around one-third of all resolvers. None of the resolvers supports the JSON approach exclusively.

Table 4. DNS software identification of found DoH resolvers in Scan2022

Name	#	%	Name	#	%
a) empty	113	26.0	g) AkamaiVantioCacheServe	10	2.3
b) Unbound	88	20.2	h) Q9	8	1.8
c) PowerDNS	77	17.7	i) NominumVantioCacheServe	8	1.8
d) unknown	68	15.6	j) SDNS	1	0.2
e) Bind	48	11.0	k) I-Evolve DNS	1	0.2
f) Dnsmasq	13	3.0			

Table 5. Share of DoH provider categories

Name	#	%	Name	#	%	Name	#	%
a) unknown	280	41.9	e) other	24	3.5	i) security	16	2.3
b) DNS/ISP/Cloud	145	21.7	f) finance	22	3.2	j) government	11	1.6
c) personal webpage	92	13.7	g) software-provider	18	2.6	k) privacy	10	1.5
d) industry&business	34	5.1	h) education	16	2.3			

5.6 DNS Server Identification

Table 4 shows the results of the DNS software identification for all the DoH resolvers that answered the specialized version query correctly (only 435 or 10%). However, most of them replied with an empty string response. The Scan2022 IP addresses were also queried using traditional unencrypted DNS over port 53/UDP. We used `nslookup` software to query the Google.com address with a 10 s timeout and from 4,354 only 1,176 (∼27%) resolvers supported legacy DNS. We repeated the test three times with similar results.

5.7 Who Operates the DoH Resolvers

A total of 657 unique domain names from TLS certificates were analysed to find out who is offering the DoH resolution services. At first, we tried to use the domain classification service NetStar [31], however, only a negligible portion of domain names were classified. Therefore, we visited each of them manually via the web browser and classified domain names into one of 11 categories: **DNS/ISP/Cloud**—DNS providers, Internet service providers, hosting providers and cloud providers; **industry&business**— manufactures, e-shops, and other types of trade business; **finance**—banks, investment advisers, and insurance companies; **software-provider**—companies providing software development services; **education**—universities, research institutes, and libraries; **security**—computer security companies; **government**—governments and governmental organisations; **privacy**—companies that focus on privacy such as VPN providers and privacy enhancement software; **personal webpage**—domain names hosting personal web site portfolio or personal blogs; **other**—companies and institutions

Table 6. Share of TLS certification authorities across the found DoH resolvers in Scan2022. CA stands for Certification Authority, IIJ stands for Internet Initiative Japan Inc., ERDC stands for Engineer Research and Development Centre

CA Name	#	%	CA Name	#	%	CA Name	#	%
a) Let's Encrypt	1,703	39.1	d) Blue Coat	106	2.4	g) IIJ	63	1.4
b) ZeroSSL	1,654	38.0	e) Sectigo ltd	103	2.3	h) WoTrus CA ltd	36	0.8
c) other	545	12.5	f) Apple Inc.	100	2.3	i) ERDC	36	0.8

that did not fall into any other category; and **unknown**—domain names did not host website, or that could not be categorised it.

The share of each category among DoH providers is shown in Table 5. We were not able to categorise most of the resolvers. The web page hosted on these resolvers could not identify the owner of the website, or the server did not serve web pages. When the web server responds with a web page, it usually shows a login page. Around 20% of the domain names in the category "unknown" showed a log-in page on AdGuard Home DNS resolver. Two of the servers were misconfigured and showed a directory structure of private project files.

For identification, we did not use information directly from the domain names. Although some domain names suggested that the server is operated by an individual, we also categorised it into the "unknown" category since we could not verify it. For most domain names, we could not even estimate the owner, since sometimes they seemed to be randomly generated, such as `hhgasdygqwueysbjadasghds.com` or `kasldjflkasdjf.xyz`.

The second most common category is DNS/ISP/Cloud providers, which offer DoH. A significant share of these companies might be expected since these companies usually provide DNS resolution as part of their services. The third most common category is private web pages. Individuals operate these resolvers, and the website usually contains the portfolio of a freelance software developer, or it was a personal blog.

5.8 TLS Certificate Analysis

We analyse the TLS certificate data of found DoH resolvers in the Scan2022. The share of certificate authorities is written in Table 6. The most common certification authorities across found resolvers are Let's Encrypt and ZeroSSL. Most of the DoH resolvers provided valid and trusted certificates. We found 193 ($\tilde{4}$.5%) IP addresses with expired certificates. More than 57% of those expired certificates were certified by the Let's Encrypt Certification Authority. The expiration date of the invalid certificates was mainly 2021 and 2022 (in 81% of the cases). The certificates of 5 resolvers expired before the DoH standardisation in 2018.

5.9 Threat Intelligence Results

From the 4,354 IP addresses in the Scan2022 list surveyed with threat intelligence tools, 1,502 are considered *suspicious for phishing* according to the Avast Web Shield tool. This, does not mean that the IPs are malicious, but that they were associated to phishing activities during the studied period 2021–2022. Moreover, 105 of these addresses contain at least one reference to a site that the VirusTotal service considers to be malicious. VirusTotal also found 27 of them were used as a source of *downloaded malware samples*, that is they directly hosted malware.

6 Discussions on the Results

The results presented in the previous section confirm that the deployment of public DoH resolvers is increasing. The number of well-known resolvers in 2022 increased by 12% compared to 2021. However, only 67% of the well-known DoH resolvers in 2021 remained active in 2022.

A similar phenomenon is observed with the results of Scan2021 and Scan2022, where only 9.8% of the IPv4 addresses were found in both scans. A possible explanation for this discrepancy is that the missing servers were for testing purposes and, as such, have been moved to a definitive address or stopped operations. Furthermore, 88% of the DoH resolvers found in Scan2022 were not previously seen by any list, nor Known2021, Known2022, or Scan2021.

Approximately 55% of the well-known DoH resolvers in 2022 were not found in the Scan2022 performed in January 2022. This suggests that the combination of errors in finding open port 443/TCP on the Internet and the rate at which the DoH resolvers are added is enough for our methodology to miss half of them.

The fact that so many DoH resolvers could not be found one year later and that the number of DoH resolvers is increasing speaks of a great dynamism and casts doubts about the effectiveness of these kind of lists for filtering DoH resolvers or blocking them.

By comparing the results of the HTTP versions supported by the discovered resolvers of the two scans, as shown in Fig. 2, there were some changes in the support of the HTTP version. There is a decrease in the percentage of DoH resolvers that support both HTTP/1 and HTTP/2; however, we can also see an increase in HTTP/1 only resolvers, even though the RFC does not recommend it due to performance reasons.

Only 21% of the DoH resolvers found in Scan2022 belong to DNS/ISP/cloud providers, while 44.6% belong to unknown organisations, and 12.7% belong to personal web pages. Almost 35% of the IP addresses found in our study present indicators related to phishing campaigns, and 27 of 4,354 IPs were a source of malware. We expect the domain resolution service to be under constant security reviews, either if it is unencrypted through standard DNS or encrypted using DoH or some other protocols. The occurrence of DoH resolvers' IP addresses associated with malware or phishing shows that users' security and privacy could be already at risk or that these resolvers are misused for malicious purposes.

Leaving aside which of those groups can be considered trusted DoH resolvers, 77.3% of the certificates of DoH providers in Scan2022 were given by free services such as Let's Encrypt. This heterogeneity gave space for threat actors to hide and abuse DoH in ways that we will discover in the future.

The impact of widespread use of DoH by threat actors is still a matter of debate. DoH could be used with malicious intentions ranging from bypassing DNS filters, to use known DNS techniques for command and control and exfiltration, but with encrypted capacity. A very shallow treat intelligence analysis showed signs of malicious activities in a small, yet considerable percentage of the servers. Even if the question of which kind of malware is using DoH for communication or is hosted in DoH resolvers was not addressed in this work, the list of public DoH resolvers found could help the community to spot existing threats.

7 Conclusion

The choice of a particular DoH resolver can have an impact on the privacy and security of the user and the security policy of administrators. It can allow users to evade filters, censorship and surveillance; but then again it can deny security tools the opportunity to protect users, while proving threat actors a better tools to cover their tracks.

We studied the deployment of DoH on the Internet and evaluated their characteristics to answer: is the number of DoH resolvers growing? and who is implementing them?

This research is a longitudinal analysis (2021, 2022) studying the number of DoH resolvers on the Internet, how they implement DoH and their features as organisations.

Results show that there are at least 59% more DoH resolvers on the Internet in April 2022 than in April 2021, showing that the number of public DoH resolvers is growing. There are ∼28 times more public DoH resolvers on the Internet than those well-known in the community in April 2022. More than 95% of the resolvers found were unknown to the community and ∼30% were found to be suspicious.

The current practise to block DoH traffic is based on blocklists of IP addresses or SNI (e.g., Sophos [36] products). The blocklist of well-known DoH providers in April 2022 is slightly larger than the one from 2021, with a small intersection between them (28%). But we expect the lists to grow in the future, since there are many organisations trying the technology and developing new services. Measurements show that the number of unknown resolvers on the Internet and their rate of change are large enough to assume that the efficiency of blocklists could be very low, especially when someone intentionally wants to avoid the block. Thus, further studies are required to prevent breaches of security policies, malware abuse, or DoH data exfiltration.

The discovery of DoH resolvers linked to suspicious or malicious activities should put the information security community in alert, to better study and understand the threats posed by these resolvers.

From the user's privacy and security point of view, the selection of the DoH resolver is important, but it depends on the threat model of the user. While for most users a local DoH resolver may suffice, users in countries with Internet surveillance policies may prefer a third-party DoH resolver. However, these users will need to make the choice carefully, taking into account the organisation providing the service, the centralisation and surveillance by the third-party, the performance, and the possibility that the DoH resolver may be related to malicious activities. Moreover, as the DoH service is now controlled by applications, users can lack the ability to choose which DoH resolver to use, effectively bypassing any local protection based on filters implemented at the network level.

By knowing the population, distribution and characteristics of the public DoH resolvers on the Internet, we are better prepared to face the challenges of these new technologies.

Acknowledgment. This work was partially supported by Avast Software, the Ministry of Interior of the Czech Republic—project No. VJ02010024: "Flow-Based Encrypted Traffic Analysis," and also by the Grant Agency of the CTU in Prague—grant No. SGS20/210/OHK3/3T/18 funded by the MEYS of the Czech Republic.

8 Appendix

8.1 Ethical Considerations

Part of our research involved technical actions that require an ethical explanation and support.

Horizontal Port Scanning. of the Internet has many implications. Although in general considered an ethical practice [23], we analyse the implications of our actions. First, our horizontal port scan sent 3 packets per port to each IP address with a rate limit. This amount of packets is not enough to consume the bandwidth of any device, nor to force errors in the services, especially since our scan did not close the TCP handshake. Therefore, the technical risk of errors or problems in devices due to our scan is negligible. Higher rates of scanning or frequency of the scans, i.e. weekly scans can pose some threat to some services availability, and thus we limited the methodology accordingly. Some honeypot devices on the Internet detected our scan and report the source IP as an attacker; however, since the IP address was not really attacking, there was an impact of having the IP in block lists for some days.

The action of verifying the DoH protocol required us to connect to all ports 443/TCP and try to find out if they spoke DoH or not. It required the request for the TLS protocol handshake and then the DoH protocol. We measure the technical impact by testing our Nmap script against our own servers, and no server was impacted by our script, was taken down, or slowed in any way. We consider the script safe and with very low impact. The script made 6 connections in total to each server.

The action of analysing DoH resolvers implied a more thorough analysis of the responses and information found about this server on the Internet. We only performed this action with the few (order of thousands) found DoH resolvers and we continually verified that they were not affected by our DNS requests.

We consider our techniques to have very low impact on the servers scanned and without reason to suspect that our actions affected the servers contacted in any way.

Publishing the List of DoH Resolvers. can significantly impact the citizens of oppressive countries that use DoH to avoid surveillance or access censored websites from the free world. The oppressive government can misuse two outcomes of our research: 1) the list of DoH resolvers can be used for DoH blocking to enforce DNS surveillance and censorship, and 2) the methodology for creation and updates of such a list.

Nevertheless, as shown in our research, the IP addresses of DoH resolvers constantly change, making the efficiency of IP-based filtering limited as discussed in the Sect. 7. Regardless of the described methodology, we argue that the methodology presented in this work is not novel nor technically complex, and uses of the freely available tools. An oppressive regime interested in DoH blocking already could have its own DoH scanning and detection infrastructure.

Besides, DoH does not entirely bypass mass censorship or surveillance. For example, domain names transferred in TLS SNI are still visible and used by large censorship systems [6]. Therefore citizens living under an oppressive regime still need to use other privacy-preserving technologies such as Virtual Private Networks to avoid censorship.

Given that, we do not consider our research would contribute to oppression by authoritarian countries or decrease the Internet privacy. Instead, our study provides essential findings about DoH resolvers worldwide and points out security concerns arising from anonymous DoH resolvers.

8.2 Nmap Configuration

The Scan2021 used Nmap *insane* timing template and 1 maximum number of retries, to minimise scanning time.

```
nmap -n -iL data/ips.txt -v -T 5 --max-retries 1 -d -Pn -p443
--script=/data/dns-doh-check
```

The Scan2022 used Nmap *normal* timing template to minimise the number of packets lost.

```
nmap -n -iL data/ips.txt -v -d -Pn -p443 --script=/data/dns-doh-check
```

See Nmap timing templates for detailed timeout information of each mode. https://nmap.org/book/performance-timing-templates.html.

References

1. AdGuard software Limited: Adguard known DNS providers. https://kb.adguard.com/en/general/dns-providers. Accessed 25 May 2021
2. AhaDNS: DNSover https (DoH). https://ahadns.com/dns-over-https/
3. Baheux, K.: A safer and more private browsing experience with secure DNS (2020). https://blog.chromium.org/2020/05/a-safer-and-more-private-browsing-DoH.html. Accessed 17 Jan 2021
4. Borgolte, K., et al.: How DNS over HTTPS is reshaping privacy, performance, and policy in the internet ecosystem. In: Proceedings of TPRC47: The 47th Research Conference on Communication, Information and Internet Policy 2019. Elsevier BV (2019). https://doi.org/10.2139/ssrn.3427563
5. Callejo, P., Cuevas, R., Vallina-Rodriguez, N., Cuevas Rumin, A.: Measuring the global recursive DNS infrastructure: a view from the edge. IEEE Access **7**, 168020–168028 (2019). https://doi.org/10.1109/ACCESS.2019.2950325
6. Chandel, S., Jingji, Z., Yunnan, Y., Jingyao, S., Zhipeng, Z.: The golden shield project of china: A decade later-an in-depth study of the great firewall. In: 2019 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 111–119 (2019). https://doi.org/10.1109/CyberC.2019.00027
7. Cloudflare Inc: DNS over https – using JSON. https://developers.cloudflare.com/1.1.1.1/encryption/dns-over-https/make-api-requests/dns-json/
8. Deccio, C., Davis, J.: DNS privacy in practice and preparation. In: Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies, pp. 138–143. CoNEXT 2019, Association for Computing Machinery (2019). https://doi.org/10.1145/3359989.3365435
9. DNSFilter: DNSfilter AI-powered DNS security. https://www.dnsfilter.com/. Accessed 15 May 2022
10. Doan, T.V., Tsareva, I., Bajpai, V.: Measuring DNS over TLS from the edge: adoption, reliability, and response times. In: Hohlfeld, O., Lutu, A., Levin, D. (eds.) Passive and Active Measurement, pp. 192–209. Springer International Publishing, Cham (2021)
11. Fernando Gont: Introduction to DNS Privacy (2019). https://www.internetsociety.org/resources/deploy360/dns-privacy/intro/
12. García, S., Čejka, T., Valeros, V.: Dataset of DNS over HTTPS (DoH) Internet Servers (2021). https://doi.org/10.17632/ny4m53g6bw.2
13. Graham, R.: Masscan: the entire internet in 3 minutes (2013). https://blog.erratasec.com/2013/09/masscan-entire-internet-in-3-minutes.html
14. Grothoff, C., Wachs, M., Ermert, M., Appelbaum, J.: Toward secure name resolution on the internet. Comput. Secur. **77**, 694–708 (2018). https://doi.org/10.1016/j.cose.2018.01.018
15. Guha, S., Francis, P.: Identity trail: covert surveillance using DNS. In: Borisov, N., Golle, P. (eds.) PET 2007. LNCS, vol. 4776, pp. 153–166. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-75551-7_10
16. Herrmann, D., Banse, C., Federrath, H.: Behavior-based tracking: exploiting characteristic patterns in DNS traffic. Comput. Secur. **39**, 17–33 (2013). https://doi.org/10.1016/j.cose.2013.03.012
17. Hoffman, P.E.: Representing DNS Messages in JSON. RFC 8427 (2018). https://doi.org/10.17487/RFC8427. Accessed 25 May 2021

18. Hoffman, P.E., McManus, P.: DNS Queries over HTTPS (DoH). RFC 8484 (Oct 2018). https://doi.org/10.17487/RFC8484
19. curl DNS over HTTPS. https://github.com/curl/curl/wiki/DNS-over-HTTPS, Accessed 25 May 2021
20. Hynek, K., Cejka, T.: Privacy illusion: Beware of unpadded DoH. In: 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), pp. 621–628 (2020). https://doi.org/10.1109/IEMCON51383.2020.9284864
21. Hynek, K., García, S., Bogado, J., Cejka, T., Vekshin, D., Wasicek, A.: Dataset of DNS over https (DoH) internet servers (2022). https://doi.org/10.5281/zenodo.6517360
22. Hynek, K., Vekshin, D., Luxemburk, J., Cejka, T., Wasicek, A.: Summary of DNS over https abuse. IEEE Access **10**, 54668–54680 (2022). https://doi.org/10.1109/ACCESS.2022.3175497
23. Jamieson, S.: The ethics and legality of port scanning. Tech. rep., SANS Institute (2001). https://www.sans.org/white-papers/71/
24. Jerabek, K., Rysavy, O., Burgetova, I.: Measurement and characterization of DNS over HTTPS traffic (2022). https://doi.org/10.48550/ARXIV.2204.03975
25. Klein, A., Pinkas, B.: DNS cache-based user tracking. In: Proceedings 2019 Network and Distributed System Security Symposium. Internet Society (2019). https://doi.org/10.14722/ndss.2019.23186
26. Lioy, A., Maino, F., Marian, M., Mazzocchi, D.: DNS security. In: Proceedings of the TERENA Networking Conference, pp. 22–25 (2000)
27. Lu, C., et al.: An end-to-end, large-scale measurement of DNS-over-encryption: How far have we come? In: Proceedings of the Internet Measurement Conference, pp. 22–35. IMC 2019, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3355369.3355580
28. Lyon, G.F.: Nmap network scanning: The official Nmap project guide to network discovery and security scanning. Insecure, Com LLC (US) (2008)
29. Mockapetris, P.: Domain names - implementation and specification. RFC 1035 (1987). https://doi.org/10.17487/RFC1035. Accessed 25 May 2021
30. MontazeriShatoori, M., Davidson, L., Kaur, G., Habibi Lashkari, A.: Detection of doh tunnels using time-series classification of encrypted traffic. In: 2020 IEEE Intl Conference DASC/PiCom/CBDCom/CyberSciTech, pp. 63–70 (2020). https://doi.org/10.1109/DASC-PICom-CBDCom-CyberSciTech49142.2020.00026
31. NetSTAR Inc.: Netstar url/ip lookup. https://incompass-branch.netstar-inc.com/urlsearch. Accessed 15 May 2022
32. Quad9 Foundation: DoH with quad9 DNS servers. https://www.quad9.net/news/blog/doh-with-quad9-dns-servers/
33. Rescorla, E., Oku, K., Sullivan, N., Wood, C.A.: TLS Encrypted Client Hello. Internet-Draft draft-ietf-tls-esni-13, Internet Engineering Task Force (2021). https://datatracker.ietf.org/doc/html/draft-ietf-tls-esni-13
34. Sebastian, G., Hynek, K., Vekshin, D., Cejka, T., Wasicek, A.: DoH research scripts for cvut/cesnet/avast doh project (2022). https://github.com/stratosphereips/DoH-Research. Accessed 25 Jan 2022
35. Siby, S., Juarez, M., Diaz, C., Vallina-Rodriguez, N., Troncoso, C.: Encrypted DNS privacy? a traffic analysis perspective. In: Proceedings 2020 Network and Distributed System Security Symposium. Internet Society, Reston, VA (2020). https://doi.org/10.14722/ndss.2020.24301
36. Sophos Ltd: DNS over https (DoH) for web security. https://support.sophos.com/support/s/article/KB-000039056?language=en_US

37. Sophos Ltd: DNS over https (DoH) for web security. https://support.sophos.com/support/s/article/KB-000039056?language=en_US. Accessed 15 May 2022
38. The SciPy community: Scipy two sample t-test (2022). https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html. Accessed 15 May 2022
39. Vekshin, D., Hynek, K., Cejka, T.: DoH Insight: Detecting DNS over HTTPS by Machine Learning. In: Proceedings of 15th International Conference on Availability, Reliability and Security. ARES 2020, ACM, New York, NY, USA (2020). https://doi.org/10.1145/3407023.3409192