IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received September 9, 2019, accepted October 3, 2019, date of publication October 29, 2019, date of current version December 3, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2950325

Measuring the Global Recursive DNS Infrastructure: A View From the Edge

PATRICIA CALLEJO[1,2], RUBÉN CUEVAS[2,3], NARSEO VALLINA-RODRIGUEZ[1,4], AND ÁNGEL CUEVAS[2,3]

[1]IMDEA Networks Institute, 28918 Leganés, Spain
[2]Telematic Engineering Department, Universidad Carlos III of Madrid, 28911 Leganés, Spain
[3]UC3M-Santander Big Data Institute, 28903 Getafe, Spain
[4]ICSI, Berkeley, CA 94704, USA

Corresponding author: Patricia Callejo (patricia.callejo@imdea.org)

This work was supported in part by the Spanish Grant TIN2017-88749-R (DiscoEdge), in part by the Region of Madrid EdgeData-CM Program under Grant P2018/TCS-4499, in part by the Ministerio de Economía y Empresa, Spain, under Project TEC2016-76795-C6-3-R and Grant RyC-2015-17732, and in part by the European H2020 Project SMOOTH under Grant 786741.

ABSTRACT The Domain Name System (DNS) is one of the most critical Internet subsystems. While the majority of ISPs deploy and operate their own DNS infrastructure, many end users resort to third-party DNS providers with hopes of enhancing their privacy, security, and web performance. However, bad user choices and the uneven geographical deployment of DNS providers could render insecure and inefficient DNS configurations for millions of users. In this paper, we propose a novel and flexible measurement method to (1) study the infrastructure of recursive DNS resolvers, including both ISP's and third-party DNS providers' deployment strategies; and (2) study end-user DNS choices, both in a timely manner and at a global scale. For that, we leverage the outreach capacity of online advertising networks to distribute lightweight JavaScript-based DNS measurement scripts. To showcase the potential of our technique, we launch two separate ad campaigns that triggered more than 3M DNS lookups, which allow us to identify and study more than 76k recursive DNS resolvers giving support to more than 25k eyeball ASes in 178 countries. The analysis of the data offers new insights into the DNS infrastructure, such as user preferences towards third-party DNS providers (namely, Google, OpenDNS, Level3, and Cloudflare recursive DNS resolvers account for ∼13% of the total DNS requests triggered by our campaigns), and into deployment decisions of many ISPs providing both mobile and fixed access networks to separate the DNS infrastructure serving each type of access technology.

INDEX TERMS Internet measurements, DNS, and online advertisements.

I. INTRODUCTION

Internet users can leverage either the recursive DNS resolvers provided by their ISPs or those offered by third-party commercial DNS providers such as Google, OpenDNS, or CloudFlare. In many cases, users resort to third-party DNS providers hoping to enhance their performance, security or to avoid censorship and surveillance. However, their choices can render, in some cases, insecure and inefficient DNS configurations [1], [2].

Understanding the global infrastructure of recursive DNS resolvers, their behavior, and users' DNS choices is critical to identify common mistakes and inefficient deployment strategies that can degrade users' web experience, security, and privacy. The research community has devoted important efforts to study infrastructural and performance aspects of the DNS subsystem [2], [6]–[9]. However, previous measurement methods failed to reach the fundamental scale, openness, global coverage, and reproducibility requirements to characterize the DNS infrastructure from the edge of the network.

In this paper, we adapt AdTag [10], an open, lightweight, and flexible measurement methodology to overcome the limitations found in previous DNS measurement methods. For that, we use the rich suite of networking APIs and capabilities offered by modern browsers to develop JavaScript-based DNS measurement scripts that trigger a DNS resolution process with an authoritative Name Server (NS) under our control. To gather empirical data at a global scale and in

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Marozzo.

TABLE 1. Comparison of our methodology, *AdTag*, with previous DNS measurement studies from the edge of the network. (*: active measurements).

Platform	Coverage (Countries / ASes)	Method/Dataset	Scope	Data Availability	# Vantage Points	# DNS resolvers	Measurement time
AdTag	178 / 25k	Ad network	Infrastructure	Yes	2.5M	76k	14 days
Iris [2]	151 / -	DNS Scans	Performance	No	13.6M*	6k	1 month
M. Müller et al. [3]	- / 3.3k	RIPE Atlas	Both	Yes	9.7k	11k	5 days
M. Almeida et al. [4]	-/94	Mobile network	Performance	No	19M / 5k	-	1 month / 1.5 year
F. Chen et al. [5]	102 / -	CDN Telemetry	Both	No	3.6M	584k	15 days

a timely manner, we distribute our scripts through online advertising campaigns which also enables performing targeted experiments in regions of the world that were typically underrepresented in previous studies.

We distribute our JavaScript-based measurements using two small ad campaigns. With a $450 USD budget —a relatively low amount for online advertising campaigns— we could run 3.8M DNS measurements from 2.5M public IPs (including both mobile and desktop users) distributed across 1M /24 IP prefixes from 25k ASes and 178 countries.[1] These experiments allowed us to identify 76k IP addresses hosting recursive DNS resolvers across 49k /24 IP prefixes and 14k ASes.[2] Our pool of IP addresses provides us with large-scale data of the global DNS infrastructure deployed both by ISPs and third-party DNS providers, as well as unique information about end users' DNS configurations. Specifically, we use our methodology to analyze three aspects of the global DNS infrastructure:

1) We quantify the use of third-party DNS providers around the world. We revisit end users' motivations to use commercial DNS providers rather than the resolvers offered by their ISPs.
2) We explore the recursive DNS resolvers providing service to users from 178 countries, including their deployment strategies and global presence.
3) Finally, we compare the DNS infrastructure deployed by ISPs that serve users connecting over mobile and fixed networks.

The analysis of these aspects reveal new findings about DNS recursive resolvers not reported so far:

- A significant percentage of Internet users resort to third-party DNS providers. Namely, Google, OpenDNS, Level3, and CloudFlare are responsible for ∼13% of the DNS requests. We observe a notable increase in the use of third-party commercial DNS providers by users accessing the Internet from countries reported to implement state-level censorship and mass surveillance. These results suggest that end-users may perceive the use of third-party DNS providers as a useful resource for avoiding censorship despite the fact that regular DNS traffic is being sent in the clear.
- For users accessing the Internet from outside of Europe and North America, third-party DNS providers are more likely to assign DNS resolvers located far from the user (i.e., resolvers placed in other continents). The concentration of commercial DNS resolvers in North America and Europe may have an impact in the web experience of users accessing the Internet from other world regions.
- Most ISPs providing both mobile and fixed-line access tend to decouple the DNS infrastructures used to serve their mobile and fixed networks. However, a few ISPs deploy a single DNS infrastructure to serve both types of services.

While the number of features studied in this paper is limited, they successfully demonstrate the potential of the proposed lightweight method to run global DNS measurements. We are confident that stakeholders – from regulators to researchers and industry – will benefit from this technology to survey DNS usage trends, and to identify deployment and performance problems, both at the granularity of specific ASes and at a global scale.

II. RELATED WORK

Previous research efforts used three methods to study different aspects of the recursive DNS infrastructure at a global scale. Table 1 compares some of the most relevant studies and techniques across four dimensions: scale and coverage (*i.e.,* ASes coverage, number of vantage points/measurement nodes, and temporal length), measurement method, openness, and scope (*i.e.,* infrastructure, or performance studies).

- **Dedicated measurements infrastructure:** Several studies relied on dedicated vantage points provided by measurements platforms such as PlanetLab or RIPE Atlas [2], [3], [11], [12] to run active DNS scans. However, these studies are constrained by the actual physical deployment of vantage points, and they are unable to capture organic behavior from real end-user devices.
- **Proprietary large-scale datasets:** The only studies with comparable scale and longitudinal coverage to the one achieved by the measurement method proposed in this paper used proprietary telemetry provided by major CDN providers, ISPs, and DNS operators [4], [5], [8], [12]. While the results obtained with this approach contributed to extend our understanding of the DNS subsystem, these experiments can only be performed by (or with the help of) a handful of

[1]We define an "eyeball AS" as any type of network in which an online advertisement has been rendered. This might include commercial ISPs, enterprise networks, or VPN services.

[2]The dataset is available to the community at http://dns-analytics.netcom.it.uc3m.es:5000.

companies owning planetary-scale infrastructure. As a result, this data is often inaccessible for independent academic researchers and most practitioners. DNS observatories have been recently developed and proposed by the research community and operators to enable the access to large-scale DNS data [13].

- **Crowdsourcing tools:** Previous research studies developed crowdsourcing measurements platforms to execute active measurements through the proactive participation of the user. These studies leveraged different techniques such as purpose-built java-applets, mobile apps, or browser extensions [4], [7], [14]–[16]. Due to the crowdsourcing nature of these tools and the limitations of each platform, the data collected by these studied is sparse both in space and time. Similar to our proposed method, Godfrey *et al.* [17] developed a JavaScript-based method to evaluate the proximity between end-user clients and their recursive DNS that runs in the background. They insert their JavaScript code in websites, thus the coverage of their study is constrained by the number of users visiting the collaborating websites.

III. MEASUREMENT METHOD

As we can see in Table 1, all previous DNS studies used methods that fall short at meeting simultaneously the geographical and temporal scope, reproducibility, and openness requirements. To overcome the limitations of the state-of-the-art, we adapt AdTag [10], a flexible JavaScript-based methodology to measure the global infrastructure of recursive DNS resolvers and users' DNS choices at a global scale and in a short timescale. Our web-based measurement method leverages a purpose-built lightweight JavaScript code which probes the configured recursive DNS resolver from the browser. Inspired by previous measurement efforts [18]–[21], we embed our JavaScript code into display online ads to perform opportunistic network measurements from the vantage point of the user.

A. JavaScript-BASED DNS MEASUREMENTS

We design and develop JavaScript-based online advertisements to study the recursive DNS infrastructure of the user by inserting a JavaScript code in the online ad that triggers a DNS request to *subdomain.dnserv.es*. As there are no JavaScript DNS-specific libraries to perform DNS lookups, our code opens a new HTTP connection to trigger a DNS lookup to our server in the same way as any regular advertisement. Both the authoritative Name Server (NS) and the HTTP server for *dnserv.es* are under our control. This allows us to record in a lawful, privacy-preserving, and user-safe fashion IP-level information[3] of both the client (d) and the recursive resolver (R). Figure 1 details the process.

[3]We only record the /24 subnetwork of the user, and the public IP address of the recursive DNS resolver providing support to the user.

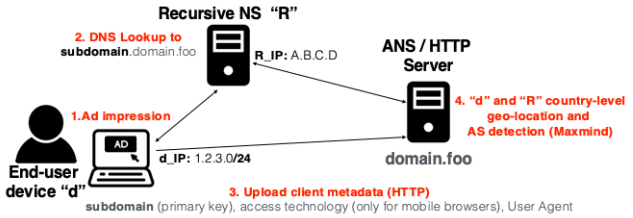


FIGURE 1. Method and data collection.

To guarantee that no caching along the whole DNS chain occurs, either on the client DNS resolver or the recursive DNS resolver, the *subdomain* of the domain resolution is a unique string randomly generated in run time for each user. Upon the reception of the DNS request from R, our Authoritative NS: (1) reports R's IP address and its subdomain to our log server; and (2) responds with the A record to R, containing the IP address of *dnserv.es*. The initial HTTP request will finalize the process with a 404 error, as it will not be able to find the random URL we inquire. Note that this HTTP response is orthogonal to the ad being rendered, which is fetched and displayed while we perform our measurements in the background.

In parallel to the DNS resolution process, our JavaScript code opens a connection with our log HTTP server (using a different domain, yet hosted in the same machine) and uploads a message that includes: the *subdomain* (to identify the session), the User-Agent (UA) of the device, and (for mobile devices only) the type of connection used as reported by the Network Information API[4] (*e.g.,* cellular or WiFi). The public IP address of the end device is obtained from the socket connection on the server side so that we can (1) geolocate at the country level, and (2) identify the network operator (at the AS level) for both the user and the recursive DNS resolver using MaxMind[5]. The random subdomain generated for each user allows us to identify a unique session and merge the data obtained from the NS and HTTP servers. In short, the final tuple obtained for each DNS measurement contains the following fields:

$$< R\text{'s IP address, } R\text{'s AS, } R\text{'s country geolocation, anonymized } d\text{'s public IP address, } d\text{'s AS, } d\text{'s country geolocation, } d\text{'s type of connection}^{6} >$$

B. RUNNING DNS TESTS AT A GLOBAL SCALE

To obtain DNS infrastructure data and usage telemetry at a global scale, we distribute our JavaScript-based tests using online advertising campaigns. Such campaigns can be configured and distributed through different kinds of ad-tech providers like Demand Side Platforms (DSP) or ad networks.

[4]http://wicg.github.io/netinfo/

[5]https://www.maxmind.com

[6]The connection of devices using desktop browsers or using WiFi are classified as *fixed*. Otherwise the connection is classified as *mobile*.

Depending on the budget,[7] it is possible to obtain between millions to hundreds of millions of daily ad impressions (i.e., DNS measurement samples) in real end-users' devices. A beneficial side effect of this distribution method is that ad providers allow us to set up targeted ad campaigns defining, for instance, a geographical location (country, region, or city) or a specific device type or platform (mobile or desktop), at any given time. As a result, the data collected through this method is independent of volunteering users, and their DNS configurations (including provider, transport method, or platform).

C. DATASET

We run our DNS measurements by launching two ad campaigns (27-04-2018 and 04-06-2018) without using the location- and device-level targeting capabilities of modern ad networks. The total cost of our campaigns was $450 (average CPM[8] ∼$0.12). Despite our limited budget, we successfully obtained 3.8M DNS measurement samples from 2.5M IP addresses, covering 1M /24 IP prefixes from 25k different ASes in 178 countries.[9] The two campaigns allowed us to unveil the presence of 76k different DNS recursive servers distributed across 49k /24 IP prefixes in 14k ASes. We made the dataset available to the community at http://dns-analytics.netcom.it.uc3m.es:5000.

D. METHOD LIMITATIONS

Our current method and dataset present several limitations which we describe below along with potential mitigation mechanism.

1) The lack of targeting in the configured ad campaigns results in a representative bias towards large ASes with millions of customers (e.g., in the US). We can tackle this natural bias with a higher investment in targeted advertising campaigns to access underrepresented ASes and countries like the case of Africa and Oceania users.
2) Our IP geolocation effort is subject to the Maxmind's geo-mapping accuracy, which previous studies have reported as good enough at the country granularity for the majority of the cases [22]. We have also considered RIPE IPmap,[10] but the response time and coverage of this service do not meet our requirements.
3) Our NS records the public IP address of the recursive resolver and end user connecting, but it only supports IPv4. Additionally, we only record the public IP address reported by our server, so we are unable to pinpoint the actual location of those DNS resolvers located behind a firewall, a DNS proxy, cascading DNS deployments, or a Carrier-Grade NAT [23]. The presence of middleboxes can be inferred statistically – e.g., a significant large number of requests coming from a given IP address. For this purpose, dedicated targeted experiments can be run.

E. ETHICAL CONSIDERATIONS

We are aware that the proposed methodology may raise ethical concerns that we take seriously. We followed the ethical guidelines defined by the network measurements [24] and ICT research communities [25] when designing and conducting the experiments. The two main ethical concerns associated with the experiments relate to: *i*) the retrieval of the public IP address of the end-user, which is considered PII by the EU legislation [26]; and *ii*) the consumption of data and energy in end user devices to run the measurements. As it is not feasible to obtain explicit user consent, we anonymize the IP address of the end-user to their /24 equivalent after extracting its meta information (*i.e.,* AS provider and country-level geolocation). No other personal or sensitive data is collected.

The data volume cost of running a test is limited to 200 KBytes, a negligible data volume in comparison with the overall data consumption associated with loading a web page. Additionally, to the best of the authors' knowledge, the conducted experiments are compliant with the Terms of Service of our ad provider and any applicable law. The experiments and data collection practices described in this paper have been supervised and approved by our institutional Ethical Committee. Note that we have not defined any specific targeting set-up in our ad campaigns that might cause any privacy damage to end users.

IV. DNS GLOBAL INFRASTRUCTURE

The first step in our empirical study is understanding the infrastructure of the recursive DNS resolvers used by millions of Internet users from all over the world. The Top-20 organizations hosting recursive DNS resolvers, based on the sample obtained by our methodology, sorted by the number of unique IP addresses recorded are shown in Figure 2.[11] According to the data, large commercial ISPs dedicate a large IP pool for hosting their recursive DNS infrastructure. However, the data also reveals that many Internet subscribers from all over the world tend to modify the DNS configuration of their devices to use third-party recursive DNS resolvers, namely CloudFlare, Google, Level3, and OpenDNS. Many other users seem to rely on recursive DNS resolvers hosted in cloud providers such as Amazon. Our methodology does not allow us to distinguish whether these cases are associated with individuals and organizations deploying their own DNS infrastructure, or if they are commercial DNS providers using Amazon's EC2 services to deploy their infrastructure.

[7] The cost of an ad display campaign is defined based on the cost per thousand ad impressions (a.k.a CPM). CPM can be as low as $0.01.

[8] CPM: Cost per thousand ad impressions.

[9] We compare our dataset coverage with the RSSAC02 metrics provided by RIPE's K-root DNS server (http://www-static.ripe.net/dynamic/rssac002-metrics/2019/). This platform observes around 3M unique IP addresses daily, so we can conclude that our dataset offers a representative picture of the DNS subsystem.

[10] https://openipmap.ripe.net/

[11] We state the name of the organization and in parenthesis the number of DNS resolvers' public IP addresses recorded and the country where the organization operate. In the case of Public providers, we indicate so instead of the country.

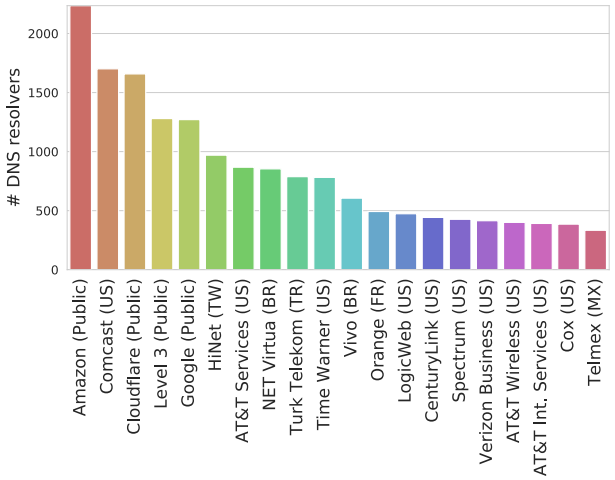


FIGURE 2. Top 20 organizations by the number of public IP addresses hosting recursive DNS resolvers.

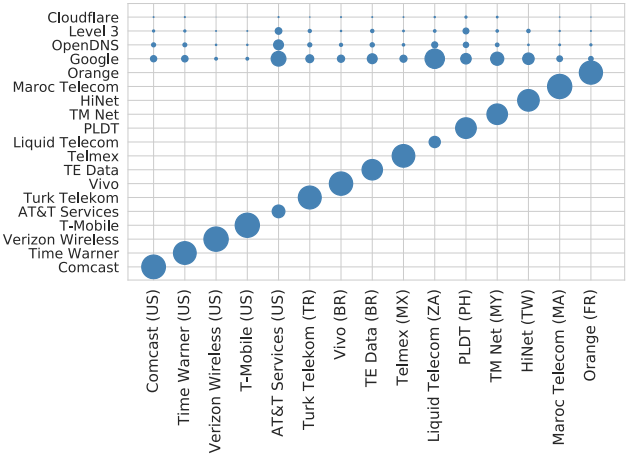


FIGURE 3. Fraction of DNS requests triggered by users from relevant ISPs (x-axis), served by machines hosted in their own infrastructure or third-party DNS providers (y-axis).

These preliminary observations suggest that the actual picture of the DNS infrastructure providing support to end customers per AS is diverse. The scatter plot in Figure 3 shows for subscribers of 15 hand-picked representative ASes/ISPs (x-axis) the ratio of DNS requests served by recursive DNS resolvers deployed by each selected AS versus the number of requests served by relevant third-party commercial DNS providers (y-axis). The size of the circle shows the fraction of the DNS lookups triggered by each type of resolver.

We can observe that most of the DNS requests observed by our NS come from within the AS providing network access to the user. However, there are differences in the ratio of requests served by third-party DNS providers across ASes. For instance, while over 80% of the subscribers of ISPs like Comcast (US), or Orange (FR) use the ISP-provided DNS infrastructure, over 50% of AT&T subscribers resort to Google DNS. Similar patterns are observed for users from ISPs in countries such as South Africa. Specifically, 19% and 58% of Liquid subscribers use the ISP-provided and Google DNS resolvers, respectively. It is worth noting that customers from Mobile Network Operators (MNOs) such as Verizon (US) and T-Mobile (US) rely (almost) exclusively in the recursive DNS infrastructure provided by their operator. This might be due to the tight control over network configurations enforced by mobile operators and mobile platforms.

A. AS-DEPLOYED INFRASTRUCTURE

In this section, we study and compare high-level properties of the DNS infrastructure for 14k commercial ASes.

1) GEOGRAPHICAL DISTANCE BETWEEN END-DEVICES AND DNS RESOLVERS

The larger the distance between the end-user and the recursive DNS resolver, the worst the customers' Quality of Experience is likely to be [27]. To investigate potential topological problems, we geolocate the IP addresses of end-user devices and recursive DNS resolvers at the country-level using Maxmind GeoIP database. Rather than measuring potentially inaccurate geographical distances, potentially inaccurate due to geo-location errors, we compute the fraction of DNS requests that are handled by DNS resolvers located (1) within the same country as the device generating the request; (2) in a different country but within the same continent; and (3) in a different continent. Note that DNS resolutions processed in a different country and specifically in a different continent are likely to produce significant delays. We observe that 99% of the eyeball ASes in our dataset resolve more than 95% DNS requests within the same country.

2) LOAD BALANCING STRATEGY

ISPs and other organizations may deploy multiple recursive DNS resolvers to cope with users' traffic demands. To study to what extent the eyeball ASes in our dataset implement a load balancing strategy, we compute the Jain Fairness Index – a metric to determine the fair share allocation of the servers, bounded between 0 and 1 –, of the distribution of DNS requests across the N recursive DNS resolvers deployed in a given AS I.[12] We refer to this metric as $JFI(I, N)$. In this analysis, we remove non-representative ASes and resolvers to minimize statistical bias. Therefore, we consider over 2k ASes whose deployed recursive DNS servers have resolved at least 50 requests, and also over 2k individual DNS resolvers that have received at least 10 DNS requests. Our results show that 57% of the ASes present a $JFI(I, N) \geq 0.8$ whereas just 2% present $JFI(I, N) \leq 0.4$. This observation confirms that most ASes commonly implement load balancing strategies.

[12] As mentioned above, we are not able to individually analyze cases in which multiple resolvers are hosted behind the same IP address.

TABLE 2. DNS infrastructure metrics continent-based for the users using Public DNS resolvers.

Continent	# DNS lookups	% third-party DNS providers	% cross-continent third-party DNS lookups	% cross-continent ISP DNS lookups
Africa	122,906	20.10	98.80	0.04
Asia	249,407	16.55	38.66	0.37
Europe	1,170,267	9.47	6.21	0.04
North America	1,428,735	12.40	4.30	0.05
Oceania	21,742	9.65	84.57	0.14
South America	855,875	14.95	30.68	<0.01

B. THIRD-PARTY DNS PROVIDERS

As presented in section IV, third-party DNS providers play a relevant role in the DNS subsystem worldwide. Previous studies performed small-scale experiments to compare the performance of commercial DNS providers with ISP-provided ones [8]. We now present a large-scale study of the use and infrastructure of popular third-party DNS providers; namely Google, OpenDNS, Level3, and CloudFlare. In particular, we computed our coverage of the DNS infrastructure for these third-party providers compared with the /24 IP blocks publicly announced by Google DNS,[13] OpenDNS,[14] and Cloudflare,[15] and we obtained 75%, 79%, and 43% overall coverage, respectively.

1) USE OF THIRD-PARTY DNS PROVIDERS

13% of the global DNS requests handled by our NS come from 21% of the /24 IP prefixes in our dataset which belong to the four considered third-party commercial DNS providers. Among the four providers, Google is the most popular one by attracting almost 75% of all the requests coming from commercial providers. These figures contrast with previous results. In 2012, TurboBytes reported that 8% of users (at the IP level) use Google and Open DNS resolvers [28]. Similarly, Geoff Houston showed that Google's DNS adoption was around 7% in 2013 [21]. If we consider these reported numbers as a reference, our results suggest that in around 5 years the userbase (i.e., IP addresses) using third-party DNS providers has increased by 85%.

2) MOTIVATION FOR USING THIRD-PARTY DNS PROVIDERS

There are significant geographic differences in the adoption of third-party DNS resolvers. Table 2 shows the percentage of DNS requests handled by our NS coming from commercial resolvers in each world continent. When analyzing at the country-level, we can observe that developing countries tend to present the largest adoption of commercial providers. The research literature suggests that end-users resort to commercial DNS resolvers to obtain better performance and reliability [29], or to circumvent censorship and obtain better privacy protection [30]. In this section, we study whether our dataset supports these adoption motivations:

- **Performance:** One may interpret that the poor performance offered by the recursive resolvers deployed by ISPs may motivate their users to use third-party providers. However, our dataset suggest that the use of third-party DNS providers in developing countries may impair DNS and web performance. Table 2 shows the percentage of DNS lookups resolved by our NS coming from ISP-provided and third-party DNS resolvers per continent. We also show the percentage of requests served by resolvers — both ISP and third-party DNS resolvers— hosted in a different continent than that of the end user. As we can see, the percentage of DNS queries coming from ISP-provided DNS resolvers hosted in a different continent than that of the end user is consistently below 0.5%, regardless of the continent. However, when users resort to third-party DNS providers, this percentage varies greatly from one to another. Over 84% and 98% of the DNS queries served by third-party DNS resolvers for African and Oceanian users are resolved by servers hosted in a different continent, respectively (even for providers supporting IP anycast). For European and North American users, this percentage never exceeds 7% of the total queries resolved by our NS. This result suggests that due to the concentration of commercial DNS resolvers in Europe and North America, a significant number of users accessing the Internet from technologically and economically developing regions are likely to experience a higher DNS lookup time, and as a result, a poorer web experience. Therefore, the argument of performance improvement does not seem to justify the use of third-party commercial DNS providers.
- **Censorship:** The research literature suggests that Internet censorship and mass surveillance may incentivize the utilization of third-party DNS providers by end users [29]. We hypothesize that the use of third-party DNS providers is, consequently, higher in countries restricting Internet freedom and human rights. To validate this, we compare the use of third-party commercial DNS providers as seen by our NS with Reporters Without Borders' (RWB) World Press Freedom index per country.[16] RWB's freedom index groups countries into 5 categories, *Good*, *Fairly Good*, *Problematic*, *Bad* and *Very Bad*, according to their degree of media and press freedom as shown in Figure 4. We only consider 94 world countries for which we have successfully recorded at least 100 DNS lookups. This analysis reveals that the median use of commercial DNS providers is over 10% in countries qualified as *Good* and *Fairly Good* by RWB's freedom index. However, for those categorized as *Problematic*, *Bad*, and *Very Bad*, the median usage is over 16%. This observation suggests that many users from all over the world resort to commercial DNS

[13] https://developers.google.com/speed/public-dns/faq
[14] https://www.opendns.com/data-center-locations/
[15] https://www.cloudflare.com/ips/

[16] https://rsf.org/en/ranking. RWB's World Press Freedom uses six indicators to estimate the degree of press and media freedom worldwide: pluralism, media independence, censorship, legislative framework, transparency, and infrastructure.

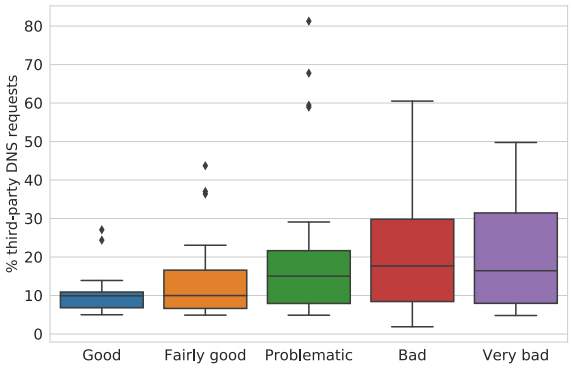


FIGURE 4. Distribution of the percentage DNS requests recorded by our NS as coming from third-party DNS providers across countries grouped by their Reporters Without Borders World Press Freedom index category.

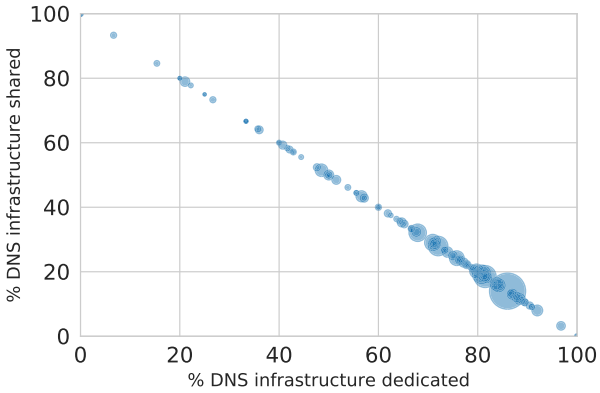


FIGURE 5. Percentage of recursive DNS resolvers for ISPs offering both fixed and mobile network access that serve both access technologies or just one of them.

providers to enhance their privacy and security, and avoid Internet censorship.

3) COMPARISON OF THIRD-PARTY DNS PROVIDERS

We conclude this section with a comparison of the IP infrastructure of third-party commercial DNS providers using the metrics introduced in section IV-A. We observe that Cloudflare's infrastructure offers the best replica assignment based on geographic distances (78% requests resolved within the country) and an almost perfect load balancing across its resolvers (JFI = 0.94). On the other hand, Google DNS resolves 71% (10%) of the requests in other countries (continents) and presents an unbalanced load across its servers (JFI = 0.28). We conjecture that these results might be due to two causes: 1) the overall traffic load of the provider – in particular, Google handles 86 times more requests than Cloudflare in our dataset (364k vs. 4.2k requests), particularly from developing countries; and 2) the notorious difference in the business models of these providers – as opposed to Google DNS, Cloudflare's DNS service is associated with its CDN services. Exploring in depth each one of these aspects would require conducting further experiments which we leave for future work.

C. MOBILE VS. FIXED ISPS

We conclude our paper with a comparative analysis of the DNS infrastructure provided by ISPs offering both mobile and fixed-line (*e.g.,* DSL and Cable) network access. We use the mobile browser's Network Information API to distinguish mobile from fixed subscribers. Using this signal, for those subscribers that provide the information required, we can tell that we have 78% (22%) of fixed (mobile) connections in our dataset.

We compare the overall size of the DNS infrastructure allocated to serve mobile and fixed users. Our results reveal that 98% and almost 16% of the observed IP addresses hosting recursive DNS resolvers serve both fixed and mobile subscribers. Only 84% and 2% of the recursive resolvers' IPs are exclusively dedicated to fixed and mobile networks, respectively.

Most of the large ISPs like Telefonica, Orange, Verizon and AT&T provide both mobile and fixed services. We, therefore, study more in depth the infrastructural commonalities for this type of dual ISP. To obtain statistically representative results, we restrict our analysis to the set of 202 ISPs for which our NS has recorded at least 20% of DNS requests coming from the least representative type of network access technology (*i.e.,* mobile or fixed), as reported by the Network Information API. Then, for each ISP, we compute the percentage of recursive DNS resolvers that are *shared* (*i.e.,* they serve requests from both the mobile and fixed networks) and *dedicated* (*i.e.,* they serve requests exclusively from either mobile or fixed network).

Figure 5 shows, for each one of the considered ISPs the percentage of *shared* (y-axis) and *dedicated* (x-axis) recursive DNS resolvers. Each ISP is represented by a circle in the figure and its diameter is proportional to the number of IPs hosting recursive resolvers in the eyeball AS in our dataset. Interestingly, we observe a clear trend in which ISPs providing dual access tend to use dedicated DNS resolvers for their mobile and fixed networks. Some examples are NTT Docomo (JP), and Vivo (BR) where 97% and 85% of their DNS infrastructure seems to be dedicated according to our measurements, respectively.

Finally, it is worth mentioning that only a few ISPs such as Telecom Italia and Skynet Belgium deploy a single DNS infrastructure to serve both types of users: 79% and 73% of their DNS resolvers serve both fixed and mobile users, respectively. While answering whether this deployment strategy is motivated by economic or performance reasons is outside of the scope of this paper, our analysis demonstrates the potential of our methodology to perform large-scale analyses to identify new research questions.

V. CONCLUSION

This paper presents a reproducible, lightweight, and cost-effective measurement technique suitable to study the global DNS infrastructure and their usage by regular users. Our JavaScript-based methodology leverages the outreach potential of online advertising networks to distribute and run lightweight DNS tests at a global scale, in a timely manner, and from the vantage point of the user.

We run two small measurement campaigns to demonstrate the potential of the proposed methodology and highlight its ability to gain new insights into the deployment strategies followed by ISPs from all around the world, and user adoption choices. Our empirical results indicate that 13% of the global DNS lookups are resolved by third-party commercial DNS providers like Google DNS rather than by ISP-provided DNS resolvers. Our study suggests that such adoption is not driven by performance gains, but likely as a mechanism to enhance privacy, and circumvent censorship and surveillance in oppressive countries. We also show that ISPs providing both mobile and fixed access tend to decouple the DNS infrastructure serving each type of network access.

The proposed methodology opens new avenues for investigating the infrastructure, robustness and user trends in the DNS subsystem at a global scale and, as a result, inform operators, standardization bodies, and regulators.

REFERENCES

[1] T. Böttger, F. Cuadrado, G. Antichi, E. L. Fernandes, G. Tyson, I. Castro, and S. Uhlig, "An empirical study of the cost of DNS-over-HTTPs," in *Proc. Internet Meas. Conf.*, 2019, pp. 15–21

[2] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, "Global measurement of DNS manipulation," in *Proc. 26th USENIX Secur. Symp. Secur.* Vancouver, BC, Canada: USENIX Association, Aug. 2017, pp. 307–323. [Online]. Available: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/pearce

[3] M. Müller, G. C. M. Moura, R. de O. Schmidt, and J. Heidemann, "Recursives in the wild: Engineering authoritative DNS servers," in *Proc. Internet Meas. Conf.*, 2017, pp. 489–495.

[4] M. Almeida, A. Finamore, D. Perino, N. Vallina-Rodriguez, and M. Varvello, "Dissecting DNS stakeholders in mobile networks," in *Proc. 13th Int. Conf. Emerg. Netw. Exp. Technol.*, 2017, pp. 28–34.

[5] F. Chen, R. K. Sitaraman, and M. Torres, "End-user mapping: Next generation request routing for content delivery," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 45, no. 4, pp. 167–181, 2015.

[6] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig, "Comparing DNS resolvers in the wild," in *Proc. 10th ACM SIGCOMM Conf. Internet Meas.*, 2010, pp. 15–21.

[7] N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, N. Weaver, and V. Paxson, "Beyond the radio: Illuminating the higher layers of mobile networks," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2015, pp. 375–387.

[8] J. S. Otto, M. A. Sánchez, J. P. Rula, and F. E. Bustamante, "Content delivery and the natural evolution of DNS: Remote DNS trends, performance issues and alternative solutions," in *Proc. Internet Meas. Conf.*, 2012, pp. 523–536.

[9] M. Allman, "Comments on DNS robustness," in *Proc. Internet Meas. Conf.*, 2018, pp. 84–90.

[10] P. Callejo, C. Kelton, N. Vallina-Rodriguez, R. Cuevas, O. Gasser, C. Kreibich, F. Wohlfart, and Á. Cuevas, "Opportunities and challenges of ad-based measurements from the edge of the network," in *Proc. 16th ACM Workshop Hot Topics Netw.*, 2017, pp. 87–93.

[11] ZMap. (2018). *The ZMap Project*. [Online]. Available: https://zmap.io/

[12] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman, "On measuring the client-side DNS infrastructure," in *Proc. Conf. Internet Meas. Conf.*, 2013, pp. 77–90.

[13] P. Foremski, O. Gasser, and G. C. M. Moura, "DNS observatory: The big picture of the DNS," in *Proc. Internet Meas. Conf. (IMC)*, 2019, pp. 87–100.

[14] N. Weaver, C. Kreibich, B. Nechaev, and V. Paxson, "Implications of Netalyzr's DNS measurements," in *Proc. 1st Workshop Securing Trusting Internet Names (SATIN)*, Teddington, U.K., 2011.

[15] OONI. (2018). *Open Observatory of Network Interference*. [Online]. Available: https://ooni.torproject.org/

[16] M. Dhawan, J. Samuel, R. Teixeira, C. Kreibich, M. Allman, N. Weaver, and V. Paxson, "Fathom: A browser-based network measurement platform," in *Proc. Internet Meas. Conf.*, 2012, pp. 73–86.

[17] Z. M. Mao, C. D. Cranor, F. Douglis, M. Rabinovich, O. Spatscheck, and J. Wang, "A precise and efficient evaluation of the proximity between Web clients and their local DNS servers," in *Proc. USENIX Annu. Tech. Conf., Gen. Track*, 2002, pp. 229–242.

[18] M. O'Neill, S. Ruoti, K. Seamons, and D. Zappala, "TLS proxies: Friend or Foe?" in *Proc. ACM IMC*, 2016, pp. 551–557.

[19] M. D. Corner, B. N. Levine, O. Ismail, and A. Upreti, "Advertising-based measurement: A platform of 7 billion mobile devices," in *Proc. ACM MOBICOM*, 2017, pp. 435–447.

[20] W. Lian, E. Rescorla, H. Shacham, and S. Savage, "Measuring the practical impact of DNSSEC deployment," in *Proc. 22nd USENIX Secur. Symp. Secur.*, Washington, DC, USA, 2013, pp. 573–588.

[21] G. Huston. (2013). *Measuring Google's Public DNS*. Accessed: Apr. 9, 2019. [Online]. Available: https://labs.ripe.net/Members/gih/measuring-googles-public-dns

[22] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "IP geolocation databases: Unreliable?" *ACM SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 2, pp. 53–56, Apr. 2011.

[23] P. Richter, F. Wohlfart, N. Vallina-Rodriguez, M. Allman, R. Bush, A. Feldmann, C. Kreibich, N. Weaver, and V. Paxson, "A multi-perspective analysis of carrier-grade NAT deployment," in *Proc. Internet Meas. Conf.*, 2016, pp. 215–229.

[24] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Commun. ACM*, vol. 59, no. 10, pp. 58–64, 2016.

[25] D. Dittrich and E. Kenneally, "The Menlo Report: Ethical principles guiding information and communication technology research," US DHS, Washington, DC, USA, Tech. Rep., 2012.

[26] EUGDPR. (2018). *The EU General Data Protection Regulation*. [Online]. Available: http://www.eugdpr.org/

[27] J. Pan, Y. T. Hou, and B. Li, "An overview of DNS-based server selections in content distribution networks," *Comput. Netw.*, vol. 43, no. 6, pp. 695–711, 2003.

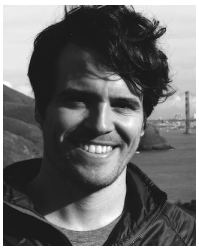
[28] TurboBytes. (2012). *Google DNS, openDNS and CDN Performance*. Accessed: Apr. 9, 2019. [Online]. Available: https://www.cdnplanet.com/blog/google-dns-opendns-and-cdn-performance/

[29] T. K. Yadav, A. Sinha, D. Gosain, P. K. Sharma, and S. Chakravarty, "Where the light gets in: Analyzing Web censorship mechanisms in India," in *Proc. Internet Meas. Conf.*, 2018, pp. 252–264.

[30] S. Dickinson. (2018). *Dns Privacy—The Problem*. Accessed: Apr. 9, 2019. [Online]. Available: https://dnsprivacy.org/wiki/display/DP/DNS+Privacy+-+The+Problem

PATRICIA CALLEJO received the B.Sc. degree in audiovisual systems engineering and the M.Sc. degree in telematics engineering from the University Carlos III of Madrid, in October 2015. She is currently pursuing the Ph.D. degree in telematics engineering with IMDEA Networks. She was granted by RIPE Academic Cooperation Initiative (RACI) on RIPE 76 that took place in Marseille, France, in 2018. After that, she did an internship with the International Computer Science Institute (ICSI), UC Berkeley, as part of her Ph.D. She is also the author of conferences, such as ACM HotNets, ACM CoNEXT, and WWW. She has worked in EU H2020 projects. Her areas of interest include the Internet measurements, online advertising, privacy, and web transparency.

RUBÉN CUEVAS received the M.Sc. degree in telematics engineering, the M.Sc. degree in telecommunications engineering, and the Ph.D. degree in telematics engineering from the University Carlos III of Madrid, Spain, in 2005, 2007, and 2010, respectively, and the M.Sc. degree in network planning and management from Aalborg University, Denmark, in 2006. In 2012, he was a Courtesy Assistant Professor with the Computer and Information Science Department, University of Oregon. He is currently an Associate Professor and the Secretary of the UC3M-BS Big Data Institute, University Carlos III of Madrid. He has coauthored over 70 articles in prestigious international journals and conferences, such as ACM CoNEXT, WWW, Usenix Security, ACM HotNets, the IEEE Infocom, ACM CHI, IEEE/ACM TON, the IEEE TPDS, CACM, PNAS, *Nature Scientific Reports*, *PlosONE*, or *Communications of the ACM*. He has been the PI of ten research projects funded by the EU H2020 and FP7 Programs, the National government of Spain and private companies, and in overall participated in 24 research projects. His research in filesharing piracy, online social networks, online advertising fraud, and Web transparency has been featured in major international and national media, such as The Financial Times, BBC, The Guardian, The Times, New Scientist, Wired, Corriere della Sera, O'Globo, Le Figaro, El Universal, El Pais, El Mundo, ABC, Cadena Ser, Cadena Cope, TVE, Antena3, and La Sexta. His main research interests include online advertising, Web transparency, personalization and privacy, online social networks, and the Internet measurements.



NARSEO VALLINA-RODRIGUEZ is currently an Assistant Research Professor with IMDEA Networks, where he has been leading the Internet Analytics Group, since 2016. He is also a part-time Research Scientist with the Networking and Security Team, International Computer Science Institute (ICSI), Berkeley. His research interests fall in the area of network measurements, privacy, and security. He has received the Best and Distinguished Paper Awards at international peer-reviewed conferences like USENIX Security 2019, ACM IMC 2018, ACM HotMiddlebox 2015, and ACM CoNEXT 2014 (short-paper). He has received the ACM IMC Community Contribution Award, in 2018, the IETF ANRP Award, in 2016, and the Qualcomm Innovation Fellowship, in 2012. His research in mobile privacy has a significant impact on industry practices and regulation, and has been covered by international media outlets, including New York Times, The Guardian, Washington Post, Le Figaro, ArsTechnica, Wired, and Financial Times.

ÁNGEL CUEVAS received the B.Sc. degree in telecommunication engineering, the M.Sc. and Ph.D. degrees in telematics engineering from the Universidad Carlos III de Madrid, in 2006, 2007, and 2011, respectively. He is currently a Ramón y Cajal Fellow (tenure-track Assistant Professor) with the Department of Telematic Engineering, Universidad Carlos III de Madrid, and an Adjunct Professor with the Institut Mines-Telecom Sud-Paris. He is also a coauthor of more than 50 articels in prestigious international journals and conferences, such as the IEEE/ACM TRANSACTIONS ON NETWORKING, the *ACM Transactions on Sensor Networks*, *Computer Networks* (Elsevier), the IEEE NETWORK, the *IEEE Communications Magazine*, WWW, ACM CoNEXT, and ACM CHI. His research interests focuses on the Internet measurements, Web transparency, privacy, and P2P networks. He was a recipient of the Best Paper Award at ACM MSWiM 2010.