DESIGN AND DEVELOPMENT OF A WORLDWIDE-SCALE MEASUREMENT METHODOLOGY AND ITS APPLICATION IN NETWORK MEASUREMENTS AND ONLINE ADVERTISING AUDITING

by

PATRICIA CALLEJO PINARDO

A dissertation submitted by in partial fulfillment of the requirements for the degree of Doctor of Philosophy in

Telematic Engineering

Universidad Carlos III de Madrid

Advisor:
Rubén Cuevas Rumín

September 2020

Design and development of a worldwide-scale measurement methodology and its application in network measurements and online advertising auditing

Prepared by:
Patricia Callejo Pinardo
IMDEA Networks Institute, Universidad Carlos III de Madrid
contact: patricia.callejo@imdea.org

Under the advice of:
Rubén Cuevas Rumín
Universidad Carlos III de Madrid
Telematic Engineering Department, Universidad Carlos III de Madrid

This work has been supported by:

“I was taught that the way of progress was neither swift nor easy.”
— Marie Curie

“A person who never made a mistake never tried anything new.”
— Albert Einstein

Gracias Rubén por confiar en mí y hacerme crecer todos estos años.
No tengo suficientes palabras de agradecimiento,
soy afortunada por ser tu estudiante.

Gracias amigos *telemáticos* por enseñarme tanto
y por los buenos momentos entre cafés.
Este viaje juntos ha sido inmejorable.

Gracias amigos por seguirme allá donde vaya.

Gracias FAMILIA por estar siempre ahí.

Gracias mamá, papá, hermanito y abueli
por ser mis mayores admiradores.

Gracias Sergio, mi compañero de viaje,
por hacer más fáciles los días difíciles.

GRACIAS a todos.

PUBLISHED CONTENT

The content of this thesis have been published in the following conferences and journals:

1. **Patricia Callejo**, Rubén Cuevas, Ángel Cuevas and Mikko Kotila. Independent Auditing of Online Display Advertising Campaigns. Published in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks (ACM HotNets 2016)* `https://dl.acm.org/doi/10.1145/3005745.3005752`
   wide
   - This work is fully included and its content is reported in Chapter 1, Chapter 2, Chapter 5, Chapter 6, and Chapter 7.
   - The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.
2. **Patricia Callejo**, Conor Kelton, Narseo Vallina-Rodriguez, Rubén Cuevas, Oliver Gasser, Christian Kreibich, Florian Wohlfart, Ángel Cuevas. Opportunities and Challenges of Ad-based Measurements from the Edge of the Network. Published in *Proceedings of the 16th ACM Workshop on Hot Topics in Networks (ACM HotNets 2017)* `https://dl.acm.org/doi/pdf/10.1145/3152434.3152895`
   wide
   - This work is fully included and its content is reported in Chapter 1, Chapter 2, Chapter 4, Chapter 6, and Chapter 7.
   - The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.
3. **Patricia Callejo**, Rubén Cuevas, Narseo Vallina-Rodriguez, Ángel Cuevas. Measuring the Global Recursive DNS Infrastructure: A View From the Edge. Published in *IEEE Access 2019* `https://ieeexplore.ieee.org/iel7/6287639/8600701/08886568.pdf`
   wide

- This work is fully included and its content is reported in Chapter 1, Chapter 2, Chapter 4, Chapter 6, and Chapter 7.
- The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.

4. **Patricia Callejo**, Rubén Cuevas, Ángel Cuevas. An Ad-Driven Measurement Technique for Monitoring the Browser Marketplace. Published in *IEEE Access 2019* `https://ieeexplore.ieee.org/iel7/6287639/8600701/08932475.pdf`
   wide
   - This work is fully included and its content is reported in Chapter 1, Chapter 2, Chapter 4, Chapter 6, and Chapter 7.
   - The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.
5. **Patricia Callejo**, Antonio Pastor, Rubén Cuevas, Ángel Cuevas. Q-Tag a transparent solution to measure ads viewability rate in online advertising campaigns. Published in *Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (ACM CoNEXT 2019)* `https://dl.acm.org/doi/pdf/10.1145/3359989.3365434`
   wide
   - This work is fully included and its content is reported in Chapter 1, Chapter 2, Chapter 5, Chapter 6, and Chapter 7.
   - The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.
6. **Patricia Callejo**, Rubén Cuevas, Ángel Cuevas, Mercedes Esteban Bravo, Jose Manuel Vidal-Sanz. Tracking Fraudulent and Low-Quality Display Impressions. Published in *Journal of Advertising 2020* `https://doi.org/10.1080/00913367.2020.1749914`
   wide
   - This work is partially included in Chapter 1, Chapter 2, Chapter 5, and Chapter 7.
   - The author's role in this work is focused on the design, implementation and experimentation of the proposed methodology.

Other Publications and Submitted Content

1. Antonio Pastor, Matti Pärssinen, **Patricia Callejo**, Pelayo Vallina, Rubén Cuevas, Ángel Cuevas, Mikko Kotila, Arturo Azcorra. Nameles: An intelligent system for Real-Time Filtering of Invalid Ad Traffic. Published in *The World Wide Web Conference (ACM WWW 2019)*. https://doi.org/10.1145/3308558.3313601
2. **Patricia Callejo**, José González Cabañas, Pelayo Vallina, Rubén Cuevas, Ángel Cuevas, Antonio Fernandez Anta. How resilient is the online advertising market to the COVID-19 pandemic?. Submitted in *The Internet Measurement Conference (ACM IMC 2020)*.

RESUMEN

La publicidad online ha evolucionado hasta convertirse en un componente clave del Internet que conocemos hoy en día. Es un ecosistema muy complejo, que logra llegar a billones de usuarios en un corto período de tiempo. Tiene cobertura global, y es capaz de llegar a audiencias específicas basadas en aspectos demográficos, geográficos y de comportamiento. Las capacidades que ofrece el ecosistema de la publicidad online han abierto una nueva era en la investigación que ha atraído el interés de la comunidad científica.

Esta tesis, aprovecha la naturaleza de la publicidad online y construye una novedosa metodología capaz de insertar código JavaScript en un anuncio, que se ejecuta cada vez que se muestra en el dispositivo de un usuario. Esta metodología abre nuevas oportunidades para realizar medidas. En concreto, esta metodología se aplica para dos propósitos diferentes en esta tesis: (1) Realizar medidas de red desde la perspectiva del usuario final, y (2) Auditar la transparencia del ecosistema de la publicidad online desde la perspectiva de los anunciantes.

En el contexto de las medidas de Internet, esta metodología se implementa en una solución llamada AdTag. Se discute y evalúa su diseño -incluyendo factores técnicos, de despliegue y económicos- y su potencial para analizar una amplia gama de aspectos de la conectividad a Internet desde el navegador. Se realizan varios experimentos que prueban la capacidad de AdTag para llegar a millones de nodos en un corto período de tiempo. Además, también se demuestra la posibilidad de seleccionar los nodos de medidas en función de su ubicación geográfica. En esta tesis, mostramos la utilidad de AdTag para realizar medidas de red en dos casos de uso específicos.

Primero, estudiamos la infraestructura DNS, uno de los sistemas más críticos de Internet. Nuestro análisis aborda cuestiones como comprender la verdadera infraestructura DNS configurada por los ISP, y entender las opciones DNS de los usuarios finales, ya sea que utilicen los *resolvers* de los ISP privados o establezcan DNS *resolvers* de terceros, para mejorar la seguridad y el rendimiento de la web. Aprovechando la escala que ofrece el ecosistema de la publicidad online, se han lanzado dos campañas de publicidad que han conseguido más de 3 millones de resoluciones DNS, que permiten la identificación y el estudio de más de 76k DNS *resolvers* cubriendo

más de 25k ASes en 178 países. El análisis de los datos proporciona nuevos conocimientos sobre la infraestructura DNS, como las preferencias de los usuarios con respecto a terceros. Nuestros resultados indican que el 13% de los usuarios utilizan proveedores DNS de terceros (como Google, OpenDNS, Level 3, y Cloudflare). Además, esta investigación detecta diferentes decisiones de despliegue de muchos ISP, que proporcionan acceso a redes tanto móviles como fijas, que separan la infraestructura DNS que sirve a cada tipo de red de acceso.

El segundo caso de uso considerado consiste en analizar el escenario del mercado de navegadores mediante medidas activas. Aprovechamos AdTag para desarrollar una plataforma de medidas activa para obtener la marca y la versión del dispositivo que recibe el anuncio. Demostramos que la muestra obtenida con nuestra metodología es muy similar a la que ofrecen las técnicas de vanguardia basadas en medidas pasivas. Sin embargo, nuestra solución presenta algunas ventajas con respecto a las soluciones pasivas: la capacidad de llevar a cabo medidas dirigidas geográfica y demográficamente, además de su accesibilidad a un grupo más amplio de científicos y profesionales. El rendimiento, la precisión y las capacidades de esta metodología se analizan a través de experimentos reales que, en total, produjeron más de 6M de medidas.

La falta de transparencia en el ecosistema de la publicidad online motiva la segunda parte de esta tesis. En particular, hemos desarrollado Q-Tag, una novedosa metodología que sirve para auditar las métricas de calidad de la publicidad online para que los anunciantes puedan obtener información fiable sobre el desempeño real de sus campañas publicitarias.

La primera versión de Q-Tag fue desplegada en Google AdWords. Los resultados revelan que AdWords parece proporcionar información incompleta a los anunciantes. En particular, muestran que: (i) AdWords no informó sobre el 57% de los *publishers*, en los que se mostraron impresiones de anuncios de nuestras campañas, (ii) AdWords informa sobre una gran fracción de impresiones contextualmente significativas, basadas en criterios (no revelados) distintos del tema de los *publishers*, (iii) una mayor inversión en CPM no conduce a que las impresiones se entreguen a *publishers* más populares, (iv) AdWords no ofrece un control predeterminado sobre el *frequency cap* (límite de impresiones por usuario), (v) alrededor del 10% de las impresiones de anuncios en dos de las campañas se entregaron a IPs de *Data Centers*.

La segunda versión de Q-Tag fue desarrollada para medir la métrica de *viewability*. Esta métrica estándar sirve para evaluar si una impresión de un anuncio ha sido vista o no por un usuario. Q-Tag ha sido desplegado en producción por un *Demand Side Platform (DSP)* (Plataforma del lado de la demanda) para medir el índice de visibilidad de las campañas publicitarias. Aprovechando la infraestructura de este DSP, se ha comparado el rendimiento de Q-Tag con una solución comercial. Ambas técnicas informan de una *viewability* global similar del 50% (es decir, el 50% de las impresiones cumplen con el estándar de *viewability* y por lo tanto se consideran vistas). Sin embargo, Q-Tag es capaz de medir la métrica de *viewability* en el 93% de los anuncios servidos por el DSP a diferencia del 74% de los anuncios medidos por la solución comercial.

En resumen, la investigación realizada en esta tesis muestra el potencial de la metodología

de medidas a gran escala basada en anuncios, que ofrece un mayor rango de posibilidades más allá de las presentadas en esta tesis. Una metodología que puede desentrañar diferentes aspectos de la infraestructura y el rendimiento de Internet desde la perspectiva del usuario final, así como proporcionar una herramienta independiente para que los anunciantes midan la calidad de sus campañas publicitarias.

ABSTRACT

Online advertising has evolved into a key component of the Internet we know today. It is a very complex ecosystem that accomplishes to reach billions of users in a short period of time. It has global coverage, and it is able to target specific audiences based on demographic, geographic, and behavioral aspects. The capabilities offered by the online advertising ecosystem have opened a new era in research that has attracted the interest of the scientific community.

This thesis leverages the nature of online advertising and builds a novel methodology capable of inserting JavaScript code into an ad that runs every time it is displayed on a user's device. This methodology opens up new measurement opportunities. Specifically, this methodology is applied for two different purposes in this thesis: (1) Performing network measurements from the end-user perspective, and (2) Auditing the transparency of the online advertising ecosystem from the advertisers' perspective.

In the context of Internet measurements, this methodology is implemented in a solution referred to as AdTag. Its design - including technical, deployability, and economic factors - and its potential to analyze a wide range of aspects of Internet connectivity from the browser are discussed and evaluated. Several experiments are performed that prove the ability of AdTag to reach millions of nodes in a short period of time. Furthermore, the possibility of selecting the measurement nodes based on its geographical location is also demonstrated. In this thesis, we showcase the utility of AdTag to conduct network measurements in two specific use cases.

First, we study the DNS infrastructure, one of the most critical Internet systems. Our analysis addresses issues such us grasping the real DNS infrastructure configured by the ISPs, and understanding the end-users DNS choices, whether they use private ISPs' resolvers or establish third-party DNS resolvers, to improve security and web performance. Harnessing the scale offered by the online advertising ecosystem, two ad campaigns have been launched, triggering more than 3M DNS lookups, which allow the identification and study of more than 76k recursive DNS resolvers supporting more than 25k eyeball ASes in 178 countries. The data analysis provides new insights into the DNS infrastructure, such as user preferences towards third-parties. Our re-

sults indicate that 13% of users use third-party DNS providers (such as Google, OpenDNS, Level 3, and Cloudflare). Besides, this research detects different deployment decisions of many ISPs that provide both mobile and fixed access networks to separate the DNS infrastructure that serves each access technology type.

The second considered use case consists of analyzing the browser market landscape with active measurements. We leverage AdTag to develop an active measurement platform to obtain the brand and the version of the device receiving the ad. We prove that the landscape picture obtained with our methodology is very similar to that offered by state-of-the-art techniques based on passive measurements. However, our solution presents some advantages over passive solutions: the ability to conduct geographically and demographically targeted measurements and its accessibility to a larger group of scientists and practitioners. The performance, accuracy, and capabilities of this methodology are analyzed through real experiments that, in total, produced more than 6M measurements.

The lack of transparency in the online advertising ecosystem motivates the second part of this thesis. In particular, we have developed Q-Tag, a novel methodology that serves to audit reported quality metrics so that advertisers can obtain trustable information about the real performance of their advertising campaigns.

The first version of Q-Tag was deployed in Google AdWords. The results reveal that AdWords seems to provide incomplete information to advertisers. In particular, they show that: (i) AdWords did not report 57% of the publishers where ad impressions from our campaigns were delivered, (ii) AdWords reports a large fraction of contextually significant impressions based on (undisclosed) criteria other than publisher's theme, (iii) higher CPM investment does not lead to impressions being delivered to more popular publishers, (iv) AdWords does not offer default control of *frequency cap* (limit of impressions per user), (v) about 10% of ad impressions in two of the campaigns were delivered to IPs from Data Centers.

The second version of Q-Tag was developed to measure the *viewability* metric. This standard metric serves to assess whether an ad impression was viewed or not by a user. Q-Tag has been deployed in production by a Demand Side Platform (DSP) to measure the viewability rate of the ad campaigns. Taking advantage of the infrastructure of this DSP, the performance of Q-Tag has been compared with a commercial solution. Both techniques report a similar overall viewability rate of 50% (*i.e.,*, 50% of the ad impressions meet the viewability standard and thus are considered viewed). However, Q-Tag is able to measure the viewability metric in 93% of the ads served by the DSP, unlike 74% of the ads measured by the commercial solution.

In summary, the research conducted in this thesis showcases the potential of the proposed large-scale ad-based measurement. It offers a wider range of possibilities beyond those presented in this thesis. A methodology that can unravel different aspects of the Internet infrastructure and performance from the user perspective as well as provide an independent tool for advertisers to measure the quality of their advertising campaigns.

Table of Contents

List of Tables

List of Figures

ABBREVIATIONS

API Application Programming Interface

AS Autonomous System

CDN Content Delivery Network

CPC Cost Per Click

CPM Cost Per Mille

DNS Domain Name System

DSP Demand Side Platform

GDN Google Display Network

GPS Global Positioning System

HTML HyperText Markup Language

HTTP HyperText Transfer Protocol

IAB Internet Advertising Bureau

IP Internet Protocol

ISP Internet Service Providers

MRC Media Rating Council

NS Name Server

OS Operating System

QoE Quality of Experience

SSP Supply Side Platform

TCP Transmission Control Protocol

VPN Virtual Private Network

CHAPTER 1

Introduction

Advertising is in continuous growth, such that we have shifted from the traditional media, like newspapers, radio, or TV, to the big online world. The emergence of online advertising in 1994 marked a turning point in the Internet era. From the first time we saw an ad online to the present day, an ecosystem has been built capable of operating in real-time, automatically and working faster than the human eye can blink [1].

Online advertising is one of the most crucial factors in marketing campaigns, thanks to its ability to reach a vast customer base at a low cost. Many Ad-Tech companies make the argument that online advertisements provide an effective form of advertising and that such advertisements provide a plausible alternative to newspapers and other types of traditional advertising. As a result of this phenomenon, the revenue growth generated by online advertising surpasses the expectations yearly.

The Internet Advertising Bureau (IAB) publishes every year the Internet Advertising Revenue Report, a report reflecting the revenue generated by online advertising from all online services. The last report from 2018, states that online advertising generated in the United States revenue of $107.5 billion, with an annual growth of 21.8% [2]. In addition, it is worth noting that online advertising is arguably the main source of funding for the open and free web.

The online advertising ecosystem has access to over 4 billion users around the world on devices like mobile phones, computers, tablets, and televisions connected to both fixed and mobile networks. This ecosystem has set up one of the largest networking distributed architectures with worldwide-scale. This infrastructure may be used for other purposes than solely delivering ads.

This thesis proposes a versatile measurement methodology that leverages the aforementioned worldwide-scale distributed infrastructure set up by the online advertising ecosystem. In particular, we insert a lightweight custom measurement code in an ad, which is distributed (up) to millions of devices via advertising campaigns that leverage the online advertising infrastructure. Each time the ad is shown in a device, the code executes and runs the custom measurement. The code can be modified to conduct different types of measurements so that various campaigns can be used to measure different things. The use of online advertising infrastructure enables our methodology to run worldwide-scale measurement experiments in short periods of time (days). To showcase the great utility of the proposed methodology, in this thesis, we leverage it to address research questions in two different areas: Internet/Network measurements and transparency in online advertising.

On the one hand, our methodology allows networking researchers to conduct measurements from the end-user perspective at an unprecedented scale in time and coverage, experiments involving tens of millions of devices in a few days. This methodology opens a new way for network measurements in different aspects, such as network transparency, security, infrastructure, or performance.

On the other hand, the lack of transparency is (arguably) the most important problem in the online advertising ecosystem. We have adapted our methodology to build an open and independent tool to audit the transparency and accuracy of standard reports and quality metrics used nowadays in online advertising.

Our methodology is conceptually built on previous network measurement literature, which used ad-based measurements based on flash ads [3, 4, 5] as well as proprietary monitoring and tracking techniques used in the online advertising ecosystem [6, 7, 8]. However, further than the concept, these previous works whereof no use in practice. On the one hand, flash technology has been deprecated and is, in general, not supported in new OSes. On the other hand, the details and functionality of monitoring and tracking methods used in online advertising are unknown mainly due to their proprietary nature. Note that since the first time our methodology was publicly released, we found other research papers using a similar approach [9, 10].

In the remainder of the Chapter, we provide a more detailed overview of the details and findings of the usage of our methodology in the two considered use cases, network measurements, and online advertising auditing.

1.1 Measuring Network Transparency, Security and Performance from the End-User Perspective

Tens of thousands of Internet Service Provides (ISPs) offer Internet access to billions of customers from all over the world [11]. The Quality of Experience (QoE) perceived by Internet users is defined by myriad factors related to the ISPs' network design, regulatory policies, network configuration, and operational decisions. In addition, a large number of research studies

have revealed application-level and end-to-end connectivity violations, including traffic discrimination and network neutrality infringements [12, 13], DNS manipulations for profit [14], in-path TLS proxies [15], and traffic manipulation by in-path proxies [16], for example via HTTP header injection to facilitate advertising and user-tracking [17]. Revealing these manipulations, as well as identifying the culprits, is of significant interest to researchers, regulators, and end-users alike. This has motivated both the research community and practitioners to design and deploy tools to perform network measurements from the edge of the network. The resulting tools leverage dedicated testbeds, crowdsourced measurements, and Virtual Private Network (VPN) services to gather insights into the edge view of the network. While powerful, they all possess inherent drawbacks such as limited geographical and ISP coverage, or short-term experiment lifespan.

Despite years of network measurement and other studies conducted from the edge of the network, pervasive access to the network edge in order to facilitate measurements has remained elusive. To close this gap, this thesis propose AdTag, an approach that leverages online advertising to launch network measurements at a global scale, in a time- and cost-effective manner. The nature of online advertising services make them an ideal, yet underused, distribution channel for launching rich network measurements either globally, opportunistically or focused on specific regions using the targeting mechanisms provided by online advertising service.

The goal in this work is to take a step back and consider the experimental apparatus of JavaScript-enabled ad placement and explore its broader feasibility for network measurement. In Chapter 4, we disclose the details of the developed methodology, that makes it reproducible. Furhermore, we discuss the aspects and challenges inherent to the distribution channel (*i.e.,* ad networks), and the execution environment (*i.e.,* the browser). Demonstrating that AdTag provides a viable and promising alternative platform for conducting a wide range of network measurements at scale, driven by web-based JavaScript APIs.

1.1.1 Measure DNS Global Infrastructure

Internet users can leverage either the recursive DNS resolvers provided by their ISPs or those offered by third-party 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 [18, 19].

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 [20, 19, 21, 22, 23].

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.

For this purpose, we adapt AdTag 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 control. To gather empirical data at a global scale and in a timely manner, we distribute the 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 the 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 to identify 76k IP addresses hosting recursive DNS resolvers across 49k $/24$ IP prefixes and 14k ASes. [2]

The pool of IP addresses provides 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 the developed 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 third-party DNS providers rather than the resolvers offered by their ISPs.
2. We explore the recursive DNS resolvers providing service to users from 128 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 $\sim$13% of the DNS requests. We observe a notable increase in the use of third-party 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 third-party DNS resolvers in

[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`.

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 work is limited, this thesis demonstrate in Chapter 4 the potential of the proposed lightweight method to run global DNS measurements. Stakeholders – from regulators to researchers and industry – can 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.

1.1.2 Measure Browser Marketplace

In the current Internet, desktop computers interact with a large number of services, including the most popular ones (Online Social Networks, Video Portals, Streaming services, etc.), through browsers. Although this affirmation is not valid for mobile devices where most popular services run through proprietary applications; it is notably the importance of browsers in the mobile ecosystem as well. Arguably, we can assert that browsers are the most common online tool on the Internet, used every day by billions of users.

Having the control of a widely used browser helps to bring a technology company into a privileged position. For instance, a company with a dominant position in the browser market can (among other things):

- Influence the adoption of different web technologies (e.g, flash vs. HTML).
- Have access to the browsing history, and thus accurate information, about the interests of hundreds of millions of users. Such information is precious for digital marketing (a business generating a revenue of $107.5B in 2018 just in US [2]).

The most important technology companies (Apple, Google, and Microsoft) are aware of the importance of having a strong position in the browser market so that they dedicate a large amount of resources to develop their browsers. Measuring the browser market share is, for obvious reason, relevant for these browser development companies. However, it is also important for other businesses such as[3]:

- Software development companies including online gaming companies, e-commerce sites, plug-in development companies, benefit from knowing the browser marketplace so that they are aware of the most popular browser brand and version per region and demographic end-users profile (i.e., age and gender) and thus the most critical for their own business.
- The security bugs of a browser's version are typically reported and fixed in the next released version[24, 25], and thus vulnerable security browsers are typically associated with old

[3]Note that this is a non-exhaustive list of businesses benefiting from using a solution to measure the browser landscape market share.

versions. Online security companies commercializing products such as firewalls, antivirus, etc., know the security bugs and vulnerabilities of different browsers' brands and versions. However, they do not know how widespread are such vulnerable browsers or identify where they are located (e.g., country, IP prefix, Internet Provider). Therefore, a tool to quantify the presence of vulnerable browsers, and identify where they are, is of great value for these companies to control the damage these versions may cause.

- Digital marketing companies run display campaigns whose ads are shown in browsers and mobile apps. Knowing the browser market share per geographical region and end-users demographic profile (age and gender) would help them: i) from a marketing perspective define better targeting strategies based on the profile of users associated to different brands and versions; ii) from a technical perspective, they can make sure their scripts (creativities) run (render) correctly in those most popular browsers brands and versions.

In addition to these business reasons, there are other important arguments to develop these techniques, for instance:

- Having an independent and accurate estimation of the browser market share would allow regulators in different parts of the world to assess the presence of monopoly situations in such critical aspects as the browser landscape. Indeed, the European Union takes monopoly situations very seriously and have already sanctioned Google for abusing its dominant position [26].
- The utilization of different browsers has different associated implications. Browsers owned by companies related to the business of digital marketing (e.g., Google or Microsoft) may collect end-users information for their digital marketing products. Other browsers like Firefox are supported by non-profit foundations, i.e., Mozilla, with no commercial interest. Therefore, the use of these types of browsers provides, in principle, higher privacy guarantees[27].

A monitoring solution should account with three principal characteristics to offer the functionality described above:

1. *Scalability*: It should be able to retrieve a large scale sample with at least millions of data points in order to provide statistically representative results.
2. *Accessibility*: Any entity including small private companies, researchers, regulators should be able to use and obtain valid results from it.
3. *Geographical and Demographic Targeting Capability:* The solution should offer the possibility to run targeted measurements on specific geographical locations (e.g., a country or a region) and for specific demographic groups of users based on their age and their gender. By doing so, for instance, a company willing to lunch a new online software targeting a specific demographic group (e.g., male between 20 and 30) in a given geographical location (e.g., France) can know the browser market share of its target in advance and chose the best development strategy to follow.

Existing solutions to monitoring the browser market landscape rely on passive measurement

techniques [28, 29]. The monitoring company installs a tracking code in a pool of N websites to collect the browser id associated with each visit to these websites. N ranges between thousands and millions of websites. While this methodology provides the scalability property and thus it is undoubtedly valuable to provide a solid knowledge about the browser marketplace, it presents important limitations in the accessibility and targeting capabilities since (i) it is accessible to just few companies with capacity to monitor thousands of websites and (ii) its passive nature prevents targeted monitoring campaigns for specific geographical areas or browser ids/versions. In addition to commercial solutions, there are academic works that analyze the browsers' marketplace with a focus on security [30, 31]. However these works do not develop any specific technique to collect browser information, instead they use logs from Google, which is obviously proprietary and not accessible.

In this work, we use AdTag to monitor the browser market landscape. This approach overcomes the described limitations of traditional passive methodologies and offers the three main functional requirements mentioned above: scalability, accessibility and targeting capabilities. AdTag inserts a lightweight JavaScript code within display advertisements. When an impression of these instrumented ads is displayed on a website, the embedded JavaScript code collects the User-Agent and the IP address[4] of the device. The User-Agent reveals the browser brand and its version, whereas the IP address allows to map the device to a geographical region.

There is a large number of advertising providers that can serve as an appropriate infrastructure to execute AdTag. Moreover, the cost associated with it is low. Note that the price of a thousand impressions (a.k.a. CPM) can be as low as \$0.01 in some providers. For instance, the provider used in the experiments has a CPM starting at \$0.10. This shows that the described technique is accessible and affordable for any institution (small private companies, researchers, regulators, etc.) interested in monitoring the browsers marketplace. Moreover, it offers the required scalability, allowing to obtain millions of measurements per day with a low investment of tens to hundreds of dollars.

In addition, this proposal can leverage the targeting capacity of the online advertising ecosystem to set up targeted measurement campaigns based on geographical location and demographic properties (age and gender). Note that other targeting parameters are also available, e.g., Operating System, device type (mobile vs. desktop vs. tablet), etc.

Finally, to prove the efficiency of the aforementioned methodology, we have run general purpose as well as targeted experiments:

- The general purpose experiments were configured without targeting parameters. We collected more than 6M measurements. We compare the browsers' market share obtained from the measurements and the one reported by well-known companies using traditional passive techniques. The results indicate that the discrepancy between my results and those reported by companies using traditional techniques is in the same range as the discrepancy between

[4]Note that we use the IP address to extract metadata information. Afterward, we anonymize the IP using hashing techniques.

the results of these companies. Therefore, we conclude that AdTag methodology provides reasonably accurate results.

- To exemplify the targeting capacity of the methodology developed in this thesis, we run three targeted campaigns: First, a geographically targeted campaign for Albania, a location with no representation in the general dataset. Second, a targeted campaign focused on old versions of operating systems, which are likely to run outdated versions of browsers with security vulnerabilities. We are able to discover the presence of more than 30 thousand devices, out of 345k, using outdated browsers with security vulnerabilities within 3 days of duration of the campaign. Third, a demographically targeted campaign for people aged between 18 and 25 years old in Italy on mobile devices. This fact shows that the proposed methodology presents functionalities not available in the traditional passive technique.

Chapter 4 presents the active measurement approach and discusses its advantages and limitations, shows the empirical results obtained from applying the described methodology in general purpose and targeting use cases.

1.2 Auditing Quality Metrics of the Online Advertising

With thousands of vendor companies, helping advertisers place ads on millions of sites, to target over 4 billion Internet users, the online advertising ecosystem is far from transparent. Without transparency, it is not possible to truly establish if online advertising is as effective as a form of advertising as the total dollar investment in it suggests.

In particular, the lack of transparency of this market forces advertisers to rely in reports and metrics provided by different vendors such as Ad Networks, DSPs or Agency partners to assess the quality of their advertising campaigns. Some recent works have shown that, protected by this opacity, some vendors are providing inaccurate information to advertisers about their advertising campaigns [32]. These findings urge to define methodologies to allow advertisers to independently assess the quality of their online advertising campaigns as well as auditing the reports received from vendors.

The research community has contributed techniques to evaluate the efficiency of different vendors in the detection and filtering of fraud [33, 32, 34, 35]. Unfortunately, fraud is not the only one aspect of the transparency problem.

For filling this gap, using the measurement methodology developed in this thesis, we present Q-Tag, a lightweight and scalable methodology to audit the performance of display advertising campaigns. In essence, we propose to inject a light JavaScript code in the ads, a method which is typically used for collecting behavioral targeting data from a user that sees the ad. This code collects relevant information associated with each impression and sends it to a central server. Specifically, the JavaScript code obtains the User-Agent receiving the impression, the URL where the impression was shown and user interactions with the ad impression (mouse movements or clicks on the ad). Moreover, we use the connection established with the server to obtain the

IP address of the device receiving the ad impression as well as the timestamp associated to the impression. Finally, we estimate the exposure time of the ad impression as the duration of the connection.

Processing this information for an ad campaign, an advertiser would be able to objectively evaluate important quality aspects such as: (i) the potential exposition to *Brand Safety* violation episodes, (ii) the popularity and *contextual* relevance of publishers where ad impressions were delivered, (iii) the quality of delivered impressions as measured by de-facto standard metrics such as *viewability* or *frequency cap* and (iv) the exposure of the ad campaign to fraud.

We have tested the proposed methodology in 8 different campaigns set up using Google AdWords. In total these campaigns delivered around 160k ad impressions across more than 7k publishers. The obtained results indicate that the information reported by AdWords to advertisers is incomplete. In particular, this auditing methodology reveals the following insights: (i) AdWords did not report 57% of the publishers where ads from the campaigns were delivered. Without a complete list of publishers, an advertiser cannot optimize its *Brand Safety* protection; (ii) AdWords reports a large fraction of contextually relevant ad impressions based on (non-disclosed) criteria different from the publisher's thematic context; (iii) we configure campaigns with Cost-Per-Mille (CPM) investment ranging between 0.01 € and 0.30 € and conclude that, contrary to my expectation, a higher investment does not lead to impressions delivered to more popular publishers; (iv) AdWords does not impose any default *frequency cap*. This leads to hundreds of cases in the campaigns where a user receives the same ad more than 100 times with inter-arrival times between two consecutive ad impressions lower than 1 minute; (v) $\sim$10% impressions are served to IP addresses belonging to Data Centers in two of the campaigns. Note that the Ad-Tech industry considers Data Center traffic to be likely associated to fraud [36, 37].

In addition, we have developed a specific version of Q-Tag tailored to measure one of the most important quality metric in online advertising campaigns, defined as *viewability*. The ad-tech industry, under the guidance of the Internet Advertising Bureau (IAB) [38] and accreditation entities such as the Media Rating Council (MRC) [39] and JICWEBS [40], has defined the viewability standard [41, 42]. Based on this standard, for instance, a display ad impression is considered *viewed* by a user only if at least 50% of the pixels of the ad are visible to the user during at least 1 second (these requirements are slightly different for other ad formats). Then, ads shown below the fold, displayed in a different tab than the one currently visible, or hidden in the background, would not be considered *viewed*. Unfortunately, as it occurs with other metrics, reported viewability rates also suffer from the opacity of the ad-tech industry. Significant stakeholders, such as Google, Facebook, or Yahoo, directly measure the viewability rate to report it to its customers. Indeed, these large vendors have defined pricing schemes that only charge their advertisers for those ad impressions meeting the viewability condition characterized by the standard [43, 44, 45]. Conversely, smaller vendors rely on third-party companies referred to as *verifiers* (Integral Ad Science [7], Moat [6], DoubleVerify [8], etc.) specialized in quality assessment of ad campaigns. All these companies use proprietary techniques to measure the viewability. As a result, the performance

and limitations of such techniques are unknown. Different studies conducted by the industry and the research community have revealed episodes of inaccurate measurements of ad impressions' viewability [39, 46] as well as misreporting of different quality-related metrics [47, 48]. These findings question the performance of these opaque techniques and claim for the necessity of transparent and auditable mechanisms to measure viewability.

For all the stated above, in this work we modify Q-Tag for assessing if an individual ad impression meets the viewability standard criteria. The methodology can be used to compute the viewability of individual ad impressions as well as the viewability rate of ad campaigns. We have performed a thorough evaluation of the proposed solution through stress tests in a lab environment that report a high measurement accuracy of 93.4%.

Q-Tag has been deployed in production by a Demand Side Platform (DSP) and its performance compared in real ad campaigns with one of the most widely used viewability measurement solution in the ad-tech ecosystem. Q-Tag can measure viewability for 93% of the ad impressions in a campaign (on average). This represents a 19 percentage points of improvement over the commercial solution analyzed, which can measure viewability for only 74% of the ad impressions (on average). This substantial enhancement in the rate of measured ads may translate into an annual revenue increase in the order of millions of dollars for mid-size DSPs serving in the order of hundreds of millions of ads per day.

In summary, Q-Tag and its application to measure general ad campaigns metrics and more specifically the viewability, contribute a novel research methodology whose application in a real use case provides solid evidences about the inconsistency of reporting the quality metrics from vendors in the online advertising market, how this may affect the interests of advertisers, and how this may impact the revenue of online advertising intermediaries.

The full description of the methodologies and findings described in this Section can be found in Chapter 5.

1.3 Contributions

The main contributions of the thesis have been published in the following venues:

- Patricia Callejo, Conor Kelton, Narseo Vallina-Rodriguez, Rubén Cuevas, Oliver Gasser, Christian Kreibich, Florian Wohlfart, Ángel Cuevas. Opportunities and Challenges of Ad-based Measurements from the Edge of the Network. In Proceedings of the 16th ACM Workshop on Hot Topics in Networks (HotNets 2017).
- Patricia Callejo, Rubén Cuevas, Narseo Vallina-Rodriguez, Ángel Cuevas. Measuring the Global Recursive DNS Infrastructure: A View From the Edge. IEEE Access, 2019.
- Patricia Callejo, Rubén Cuevas, Ángel Cuevas. An Ad-Driven Measurement Technique for Monitoring the Browser Marketplace. IEEE Access, 2019.
- Patricia Callejo, Rubén Cuevas, Ángel Cuevas, Mikko Kotila. Independent auditing of online display advertising campaigns. In Proceedings of the 15th ACM Workshop on Hot

Topics in Networks (HotNets 2016).

- Patricia Callejo, Antonio Pastor, Rubén Cuevas, Ángel Cuevas. Q-Tag a transparent solution to measure ads viewability rate in online advertising campaigns. In Proceedings of the 15th International Conference on Emerging Networking Experiments And Technologies (CoNEXT 2019).
- Patricia Callejo, Rubén Cuevas, Ángel Cuevas, Mercedes Esteban Bravo, Jose Manuel Vidal-Sanz. Tracking Fraudulent and Low-Quality Display Impressions. In Journal of Advertising 2020.

1.4 Thesis Outline

The remainder of the thesis, detailing the contributions mentioned above, is organized as follows: Chapter 2 explains the concepts needed to understand this thesis, how the online advertising works, and why it is important in the context of this thesis. Chapter 3 describes the design and technical aspects of the methodology developed for this thesis. Chapter 4 discloses the details of AdTag and its applications to perform Internet measurements from the end-user perspective. Chapter 5 explains Q-Tag for auditing online advertising and showcases the usability of this methodology to assess the quality metrics of online advertising. Chapter 6 reviews the legal and ethical considerations implied in this thesis. Finally, Chapter 7 summarizes the findings obtained through this work and future research lines.

CHAPTER 2

BACKGROUND

The complex ecosystem of online advertising needs an extensive explanation. This Chapter is focused on the definition of online advertising and all the intermediaries that make it work. Furthermore, it highlights the main advantages of this ecosystem for network measurements.

2.1 Overview of the Online Advertising Ecosystem

The online advertising ecosystem is currently responsible for delivering around a trillion ads from hundreds of thousands of advertisers into tens of millions of websites and mobile apps every day. To this end, the ecosystem has evolved into what is referred to as *programmatic advertising*.

In *programmatic advertising*, the ad-spaces are available on a website or mobile app. The aggregated pool of ad-spaces is referred to as *ad inventory*, whereas the individual instance of an ad shown to a user is referred to as *ad impression*. Last, the content of the ad is referred to as *creativity*.

There are two main sides, the sell-side, and the buy-side. Publishers, Ad Networks, and Supply Side Platforms (SSPs) from the sell-side of the online advertising ecosystem since their main goal is selling *ad inventory*. Its counterpart, the buy-side, is formed by Demand Side Platforms (DSPs), agencies, and advertisers since they pay for the *ad impressions* to be in the publishers' websites/apps.

Figure 2.1 summarizes the advertising ecosystem described below. More detailed information regarding the operation of *programmatic advertising* can be found in [49].

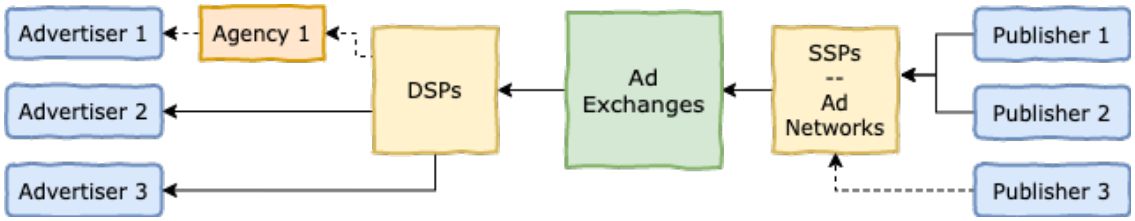

Figure 2.1: Overview of the programmatic online advertising ecosystem.

- **Publisher:** It is the website or mobile application that sells the ad-spaces. Publishers rely their ad-spaces on an SSP or Ad Network to handle the *ad inventory*.
- **SSP / Ad Network:** They are on the sell-side. They give service to websites or mobile apps from publishers to manage their available ad-spaces to help to sell them and receive revenue for it. It is an automated process where every time a user opens a website, the SSP /Ad Network receives that information and spread it to the buy-side.
- **Ad Exchange:** An Ad Exchange is in between the selling and buying sides; it is the default link between the SSP/Ad Network and the DSP. It groups the *ad inventory* sell by SSPs/Ad networks and offers it to DSPs in the buy-side. It is like a *central* market that makes the transaction easier. The inventory prices are negotiated in real-time by an auction process. The winning price is determined based on all the received bids from DSPs. It is the Ad Exchange who picks the highest auction price, and consequently, the winner. Algorithms and technologies automatically drive the process.
- **DSP:** The DSP is the counterpart of the SSP on the buy-side. DSPs allow advertisers (or agencies on behalf of advertisers) to set up their ad campaigns. DSPs will interact with Ad Exchanges to buy the *ad inventory* matching the specification of advertisers' campaigns. The purpose is to optimize the advertisers' needs and to simplify the interactions with the Ad Exchanges. Again, this process is held automatically in real-time.
- **Advertiser:** It is a company running ad campaigns to show their products, services, etc. to users through ad-spaces. Advertisers typically hire the services on an agency or a DSP to run their campaigns and achieve the best marketing results. Besides, the advertiser also sets up the price it is willing to pay to deliver an ad. There are two primary monetization schemes: CPM that indicates the price an advertiser is willing to pay by 1000 impressions of its ad; CPC indicates the price an advertiser is disposed to pay if the user clicks on the ad.

Finally, it is worth mentioning that the work presented in this thesis focuses on the buy-side, specifically in the DSPs, where the ad campaigns are configured. These campaigns are configured based on a specification, including geographical location, demographic information, users' preferences, etc., from the targeted audience.

2.2 Campaign Quality Metrics

There are two main types of campaigns referred to as *branding* and *performance* campaigns, respectively. Branding campaigns aim to reach a brand or product known so that their goal is to get as many ad impressions as possible *viewed* by users. Instead, performance campaigns aim to sell a product or service, so that their goal is to persuade the user to click on the ad, bring him to the product's website, and make a purchase.

Since both types of campaigns have different goals, the metrics to assess their performance are also different. In branding campaigns, *viewability* is the key performance metric since it determines whether the ad was sufficiently exposed to the user to have some marketing effect. In particular, the viewability standard defined by the IAB considers a display ad *viewed* if at least 50% of its pixels are exposed to the user during at least 1 second. The standard slightly differs for large display (video) ads where it is required that 30% (50%) of the pixels are shown to the user for at least 1 (2) second(s). In performance campaigns, there are two widely used metrics Return of Investment (ROI) and Click Through Ratio (CTR). ROI is defined as the ratio between the sales and the investment in an ad campaign, whereas CTR captures the fraction of ad impressions in a campaign that attracts a click. Note that ROI and CTR depend on the viewability rate since the higher is the viewability rate of a campaign, the more chances to get clicks and purchases.

However, those are not the unique metrics used to assess the quality of advertising campaigns. Other important metrics are:

- **Brand Safety:** It refers to *"practices and tools allowing to ensure that an ad will not appear in a context that can damage the advertiser's brand"* [50]. For instance, avoiding an ad from a toy brand to be displayed on a porn website. One of the "golden rules" for an advertising campaign is to preserve the advertiser's brand safety.
- **Context:** Advertisers are, in general, interested in displaying their ads with publishers whose content is topically relevant to the topic of the ad. For instance, a hotel ad is better placed on websites related to holidays or travel agencies than on websites related to job search. Note that recent forms of online advertising, such as Online Behavioural Advertising (OBA) [51, 52], have led to ad placements being based decreasingly in contextual relevance.
- **Publishers' popularity:** The popularity of a publisher indicates its capacity for attracting users. Together with other factors, it is widely used to assess the quality of a publisher. In general, advertisers pay higher CPM and CPC for impressions placed (or clicks occurring) in popular publishers. The term *premium inventory* is generally used to describe inventory from popular websites.
- **Frequency cap:** Defines the limit for the number of impressions of the same ad that should be shown to the same user in a given period of time [36, 53, 54].
- **Fraud indicators:** The World Federation of Advertisers defines advertising fraud as events *"associated with an activity where impressions, clicks, actions, or data events are falsely*

reported to criminally earn revenue, or for other purposes of deception or malice" [55]. The Interactive Advertising Bureau estimates that advertisers lose more than $8B annually directly to ad fraud in US [56].

- **Conversion Ratio:** The fraction of the sum of impressions that lead to the desired action (e.g., a seat booking from an airline ticketing site).

2.3 Assessment of the Online Advertising Capabilities

DSPs offer a wide range of capabilities that make advertising campaigns flexible and controllable to achieve the desired marketing effects for their clients. For instance, if a new ice-cream shop wants to reach people close to its location to advertise an opening offer (to get people to come physically and discover the location), it needs to target people geographically in an area around the shop. Otherwise, the offer will not have any effect. This example showcases how DSPs can target a specific audience based on location information. However, DSPs targeting capabilities are broader than location. In particular, DSPs can target audiences based on location, demographic information (age and gender), and users' interests and behaviors (e.g., users interested in sports). This provides significant flexibility in the definition of targeted audiences.

The potential user base that can be targeted from a DSP is, in theory, every user owning a digital device connected to the Internet. Some sources estimate that there are more than 4 billion people actively using the Internet [1]. This represents roughly 60% of the world's population and defines an upper limit of the number of users, which can be, in theory, reached by a DSP.

To enable the referred targeting capabilities based on demographic, geographic and behavioral information, the online advertising ecosystem relies on a sophisticated tracking subsystem able to obtain this information for individual users. Next, we describe in more detail each of the possible targeting options offered by a DSP.

1. Geographically
    - **Location-based targeting:** Some publishers or Ad Exchanges have access to the GPS of the users allowing them to run geographical campaigns with a fine-grained purpose. There are other cases where it is only needed the location at country or city level. For that purpose, the ads also have access to the users' IP address that, combined with a GeoIP database, can get that information.
    - **Operating System-based targeting:** Most ads run on websites or mobile apps. On the one hand, websites are rendered by browsers specific for each Operating System (OS). Therefore, the browser information reveals the specific OS. On the other hand, apps run on a specific mobile OS (*e.g.,*, Android, or iOS), and then they also reveal information about the OS.
    - **Device-based targeting:** OSes are typically associated to a type of device. For

[1] Digital around the world: `https://datareportal.com/global-digital-overview`

instance, mobile OSes are associated to mobile devices. Hence, leveraging the information related to the OS, campaigns can target specific device types.

2. Demographically
   - **Gender-based targeting:** Some stakeholders of the ad-tech ecosystem analyze the user profiles to infer the demographic (age and gender), interest/preference, and behavior information. This information is later on used to run targeted campaigns. In particular, one option is running gender-targeted campaigns.
   - **Age-based targeting:** In the same way, the age can be inferred and can be used to target specific age groups in the online advertising campaigns.

One of the contributions of this thesis introduces the use of online advertising to run network measurements on a global scale, in a time- and cost-effective manner. The nature of online advertising services makes them an ideal, yet underused, distribution channel for launching rich network measurements globally, opportunistically, or focused on specific regions, using the targeting mechanisms provided by online advertising service.

CHAPTER 3

METHODOLOGY OVERVIEW

The overreaching contribution of this thesis is the design of a general-purpose worldwide-scale measurement methodology that leverages the ad ecosystem infrastructure as a measurement platform. To showcase the utility of this methodology, we implement it for two use cases: network measurements (Chapter 4) and online advertising auditing (Chapter 5). However, all the use cases are based on the same methodology that relies on lightweight technologies, ensuring efficiency, scalability, and robustness.

In the remainder of this Section, we first present a high-level description of the rationale of the methodology. Then we describe the main building blocks of the methodology from a technology point of view.

3.1 Rationale of the Methodology

The methodology proposed in this thesis, explained at a high-level, introduces a custom code within an advertisement to obtain many measurement points, from the users' devices, with global coverage and in a short period of time. Its operation can be summarized in the process of a user who opens a web page with ad-spaces and receives our advertisement that is displayed on the screen. At that time, the code inserted executes the configured experiments. In turn, the code communicates with an external server under our control, where it sends the results of the measurements collected within the advertisement. In this way, repeating the process in the billions of users that can see our ad every day, we obtain a large number of measurements in a very

efficient way. This methodology allows us to make very different measurements, applicable to many different aspects. Although they can be categorized under two types of measures from the point of view of the custom code that is introduced in the ad:

1. Measurements directly in the ad, without using external sources or servers, that the necessary information is extracted from the ad using the available libraries and sent to our control server.
2. Measurements with external iterations, in this case, the ad needs to actively communicate with external services or with our control server, thus getting the additional information needed to complete the experiments.

In short, it is a methodology that can be shaped to any scenario, with high capabilities and opportunities. The technical details are explained below.

3.2 Building Blocks Description

In this Subsection, we explain the technical aspects of the proposed methodology. We detail the programming languages, the structure, the communication protocol between the ad and the server, and the database used. Figure 3.1 exemplifies the methodology built for this thesis.

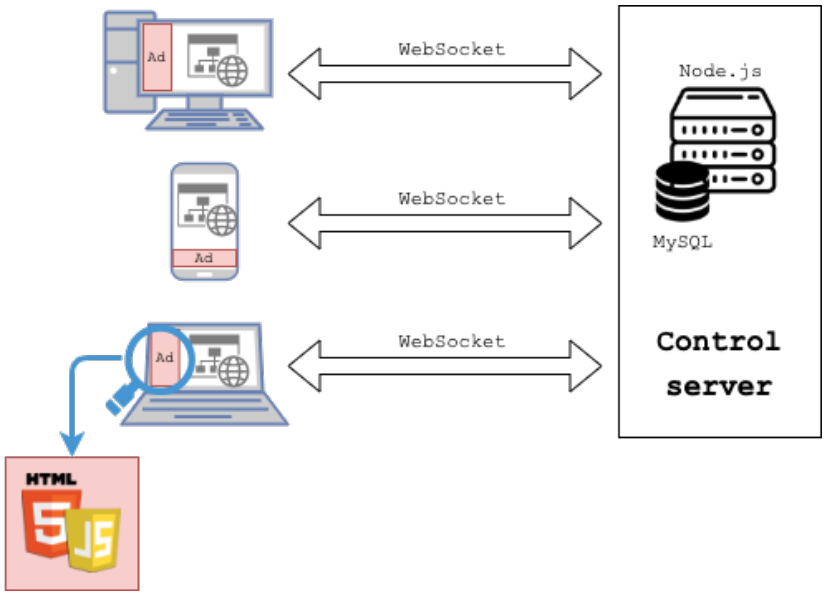

Figure 3.1: Schema of the infrastructure and technologies used by the developed methodology

- **Plain JavaScript** for local information collection from the ad impression. The support of JavaScript libraries in modern browsers, make the methodology flexible, being able to insert custom code for multiple purposes. Using plain JavaScript and avoiding the use of external libraries, we can control the number of code lines we use. Then, we can assure our methodology has a negligible impact on the process and overall overhead imposed in the device to display the ad.

- **WebSocket protocol** is a lightweight protocol, independent from HTTP, that operates over TCP. It was designed for communications between web browsers and a web server. Since the communication in our case will be initiated from a web browser, WebSocket perfectly suits our requirements. Note that the collected information is transmitted in the form of a string to the server [57].
- **Node.js JavaScript library** is a widely extended lightweight and efficient library. We use it on the server-side to receive and process the information received from the ad through a WebSocket connection. We have selected Node.js in front of other options due to its faster processing time [58].
- **MySQL database** The processed and collected information is stored in a MySQL database under our control [59]. However, in some of the experiments presented in this theses, the data have been stored in an external company.

Note that this Section has presented the overall rationale and architecture of our methodology. However, the implementation of the methodology is specific for each measurement purposes. For instance, the specific JavaScript running on the ad or the server-side implementation is ad-hoc for each measurement goal. Therefore, the specific implementation of our methodology for the different use cases addressed in this thesis is explained in its corresponding chapter: different types of network measurements in Chapter 4 and transparency analysis in the online advertising in Chapter 5.

CHAPTER 4

NETWORK MEASUREMENTS

4.1 Opportunities and Challenges of Ad-based Measurements from the Edge of the Network

For many years, the research community, practitioners, and regulators have used myriad methods and tools to understand the complex structure and behavior of ISPs from the edge of the network. Unfortunately, the nature of these techniques forces the researcher to find a balance between ISP-coverage, user scale, and accuracy. In this work we present AdTag, a network measurement paradigm that leverages the opportunistic nature of online targeted advertising to measure the Internet from the edge of the network. We discuss and formalize AdTag's design space—including technical, deployability and economic factors—and its potential to analyze a wide spectrum of Internet connectivity aspects from the browser. We run several experiments to demonstrate that AdTag can be tailored towards geographic and device-based user groups, finding also several challenges to be faced in order to maximize the number of samples. In a 7-day campaign, AdTag could access more than 20k ISPs at a global scale (185 countries) using millions of edge nodes.

4.1.1 Background

Existing edge-driven measurement techniques fall into four broad categories that we survey in this section. Table 4.1 summarizes our findings.

Project	Nodes†/IPs*	ASes	Countries	Time	Deployment strategy
AdTag	2,500,000*	20,700	185	7 days	Targeted ads
RIPE Atlas	9,300†	3,300	181	6 years	Testbed / Dedicated node
Archipelago	181†	146	60	10 years	Testbed / Dedicated node
Netalyzr	2,200,000*	14,500	196	6 years	Crowdsourcing / Mobile app, browser applet
Luminati	1,300,000*	14,700	172	5 days	P2P-based VPNs

Table 4.1: Comparison of a global AdTag campaign with previous studies in terms of network coverage, measurement duration, and deployment strategy. (*: number of sessions; †: number of nodes)

- **Dedicated testbeds:** Several dedicated measurement testbeds exist. RIPE Atlas [60], CAIDA's Archipelago (Ark) Measurements Infrastructure [61], the MONROE Mobile Broadband measurements platform [62], BISmark [63], and PlanetLab [64] are prominent examples. RIPE Atlas, Ark, and BISmark require dedicated hardware typically hosted by volunteers or academic institutions. As a result, these platforms typically possess limited geographical and ISP coverage due to their high deployment cost. Moreover, these platforms differ widely in openness and the types of tests one can execute.
- **Crowdsourcing:** Researchers have developed several user-friendly tools to help users to understand the behavior of their network. In exchange, the research teams collect valuable, oftentimes anonymized, real-world data about the access link. Examples include the ICSI Netalyzr [65], DASU [66], MobiPerf [67], and Encore [68], which embeds JavaScript code on popular landing pages, unbeknownst to users. These tools are available as apps for mobile devices, browser-based clients, command line clients, or plugins for BitTorrent clients. As opposed to measurements run on dedicated testbeds, measurement campaigns following a crowd-sourcing strategy allow researchers to maximize ISP and user coverage without necessarily sacrificing data accuracy and detail. Commercial products like Ookla's SpeedTest [69], and measurement campaigns run by regulators (*e.g.,* FCC's speedtest[70]) have also followed this model with great success. Unfortunately, the majority of these tools only provide a snapshot of the network at a given time when the user executes the tool. This limits their ability to run longitudinally, and to measure behavior at a point in time chosen by the researcher.
- **VPN-based studies:** A number of research efforts have leveraged VPN services to penetrate ISPs all over the world. One popular VPN service used by researchers is Luminati [71], a commercial VPN service that provides vantage points in more than 20M residential and enterprise IPs. Luminati has been used to detect traffic manipulations inflicted by in-path HTTP proxies [72] and end-to-end violations in the Internet [73]. Further, Luminati's low-end monthly price is $500 for 40GB of traffic. However, recent studies have questioned the ethical, privacy and security aspects of such VPN services [74], and it is unclear whether the egress points can also actively manipulate user's traffic. Other projects like ICLab have also used commercial VPN services to conduct censorship analysis [75]

at a global scale. Unfortunately, recent studies have questioned the ISP coverage of these services [76], which may bias the experimental results.

- **Targeted ads:** Ads have rarely been used for academic Internet measurements on a large scale. O'Neill *et al.* leveraged Flash-based ads to identify the presence of TLS proxies [15]. Since most modern browsers and ad networks move to deprecate or disable Flash, [77] it no longer offers a sustainable deployment mechanism. The same holds true for Java applets. Geoff Huston used advertising campaigns for APNIC Labs' IPv6 Measurement System [78], achieving good coverage by downloading a tracking pixel using JavaScript and Flash ads. A recent paper by Corner *et al.* proposes advertisement as a platform for large-scale network measurements. The authors demonstrate its ability to improve geo-IP databases, conduct bandwidth measurement and the identifiability of mobile users [9]. It corroborates our proposal of an advertisement-driven solution to edge measurement, but their study is focused solely on mobile measurements, namely device battery management and GeoIP databases.

4.1.2 AdTag

AdTag leverages ad networks for conducting network measurements at a global scale, in a time- and cost-effective manner. However, distributing complex network measurements through ad networks and running them on the browser poses several challenges which have not been systematically studied so far.

In this subsection we discuss AdTag's design space[1]. First, we describe the test distribution channels through ad networks. Then, we focus on understanding aspects inherent to ad networks such as the cost of launching campaigns, the ability to target specific user groups and platforms, and the available execution window. For these, we use empirical data that we obtained from a purposely-run advertising campaign launched through a Demand Side Platform (DSP).

4.1.2.1 Deploying Network Measurements

We deploy AdTag measurements using real advertising campaigns configured through a DSP. As explained in Chapter 2, the current online advertising ecosystem [79], typically called *programmatic advertising*, is a complex one, composed by multiple intermediaries. The ad spaces available in a publisher website are typically handled by Ad Networks or SSPs, those intermediaries are in charge of selling the ad spaces. From the buying side, the advertisers typically rely on agencies or DSPs to manage their campaigns. A DSP is an intermediary platform providing advertisers unified access to multiple vendors (Ad Exchanges), each selling ad spaces from a pool of websites and mobile apps. It also enables advertisers to configure targeting parameters for their campaigns (geographical location, device type, etc).

[1]The online advertising industry uses the term *ad tag* to refer to a piece of code typically used to monitor ad behavior.

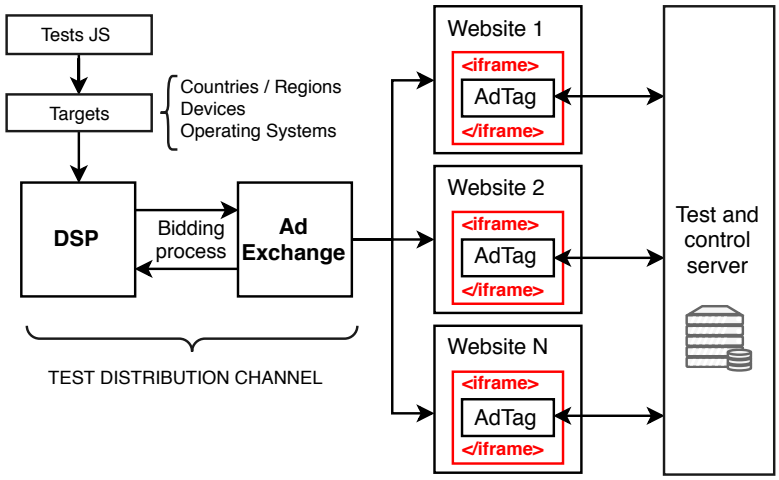

Figure 4.1: AdTag architecture, distribution channel and client-server components for measurements.

As a proof of concept, we run a 7-day campaign using 9 of the more than 20 ad networks provided by a DSP [2]. This campaign provides more than 3M measurements from 2.5M unique IP addresses covering 185 different countries. This rivals the number of sessions initiated by the crowdsourced Netalyzr [65] platform over a timespan of 6 years, underscoring the method's broad reach.

AdTag leverages HTML5-based ads [80, 81] to execute JavaScript-based active network measurements from the edge of the network. JavaScript allows embedding different pieces of code to conduct a wide range of network measurements, which will be distributed at a global scale through advertising campaigns, as illustrated in Figure 4.1. AdTag is constrained to the features and APIs provided by end-user browsers. It is important to remark that the DSP renders the ads in an iFrame, which sandboxes the JavaScript code. This prevents it from interacting directly with the parent window, including via cookies. Apart from those constraints, the DSP enables performing all the measurements explained in this subsection. Note that other limitations may apply depending on the DSP.

4.1.2.2 Targeting ISPs and Locations

Targeting measurements to specific ISPs and geographical locations allow researchers to precisely analyze and penetrate particular providers. This ability is determined by the accuracy of the targeting mechanisms provided by the DSP. Most DSPs allow targeting campaigns based on location, device type (*e.g.,* desktop vs. mobile), and even operating system. This feature is used to configure the campaigns to the experiment's needs and to target specific ISPs.

We perform several experiments to analyze the feasibility of targeting ISPs and platforms,

[2] By request of the DSP used for this work, its name cannot be shared.

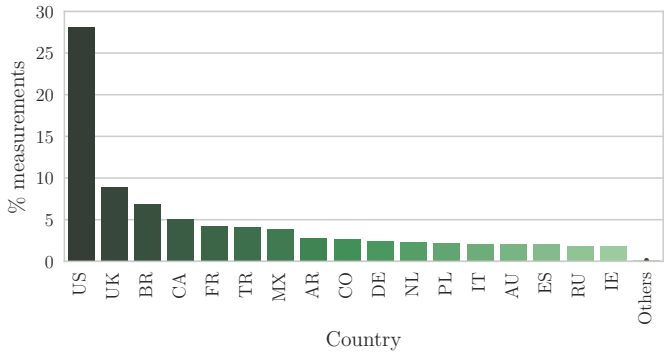

Figure 4.2: Distribution of user IPs around the world according to the results of the global campaign.

ISP Name	# samples	% samples
Comcast Cable	149.4k	17.7
CenturyLink	99.3k	11.7
Time Warner Cable	85.8k	10.1
AT&T U-verse	69.3k	8.2
Cox Communications	37.8k	4.5

Table 4.2: Top 5 most representative ISPs from the USA according to the results of the global campaign.

and evaluate the precision of the DSP's target mechanisms. MaxMind's database [82] is used to geolocate client IP addresses. While research has shown that the use of IP geolocation databases can introduce biases [83], we believe them still to be indicative of the overall deployment. The global ad campaign covers 185 countries, with the majority of measurements coming from clients in the US (28%), UK (8.8%), Brazil (6.8%), and Canada (5.1%). Figure 4.2 shows the overall geographical coverage obtained with our global campaign, which covers 185 different countries.

When geolocating US-based IP addresses, Table 4.2 showcase that most of the impressions come from large fixed-line and mobile ISPs like Comcast. However, the advertising campaign used for this work also allows accessing a fair number of small ISPs such as NTS Connections (AS46698) and Northwest Open Access Network (AS16713), both of them with at least one hundred samples.

These observed coverage distributions are expected as we did not use precise geographical targeting and thus received biases in impressions towards the US, where most of the websites are hosted, and towards ISPs with a large customer base. To target a particular ISP, the researcher can adapt the campaign using various features offered by DSPs. Some DSPs allow deployment on a country or city-level, this can be used to target the area where a desired ISP is known to operate, maximizing the number of valid samples. Other DSPs allow researchers act similarly by specifying IP ranges for deployment [9].

To validate these proposed solutions, we ran two 1-day 50k sample experiments targeting the USA and NYC, respectively. In the country-level experiment, 97% of the users had a US-based IP address. The rest of the samples came from a handful of countries, namely Canada (2% of the total samples). The results of the city-level experiment show similar accuracy.

4.1.2.3 Price

Running online advertising campaigns comes at a cost. However, it is possible to leverage different strategies to maximize the geographical coverage while keeping the budget under control. For instance, in our global campaign run we fixed the CPM budget. The used DSP allows CPMs starting at $0.10. Therefore, it is possible to launch campaigns at this minimum CPM cost and consider higher CPMs in order to increase geographical and ISP coverage when needed (*e.g.,*, to target under-represented geographical areas).

For the majority of network measurements, user clicks are irrelevant. User interaction may be only needed when their feedback is required, as in the case of QoE experiments. As a result, AdTag does not need to apply any campaign optimization based on CPC (Cost per Click), notably reducing the budget requirements to launch measurement campaigns.

As AdTag is running on a large number of heterogeneous systems and configurations, the measurements are subject to multiple sources of errors which can cause data loss, such as browser extensions preventing JavaScript (*e.g.,* ad-blockers [84, 85]), transient network disruptions, and limited browser API support. Overall, comparing the DSP reports and the data gathered by AdTag, there is a 15% data loss on average per campaign.

An estimation of the cost per campaign, assuming an average CPM of $\$0.10$[3] and a conservative efficiency ratio of 80%, resulted in approximately 1M measurements for a $125 budget, more cost efficient than previous research driven by ad placements ($5k for almost 3M successful measurements) [15].

We conclude that running measurements using online ads is 1) more flexible, 2) increases ISP coverage, and 3) is more economic than using VPN-based systems.

4.1.2.4 Execution Window

A website—including any embedded element, such as ads—may be active in the browser for only a short period of time: if the user opens a new website or simply closes the tab, the JavaScript code running AdTag tests will be immediately interrupted. As a result, it is important to know for how long the measurements can last, *i.e.,* the *execution window*.

We use the data provided by our global campaign to estimate the expected execution window. The results suggest that 75% of ads are active for more than $11s$, regardless of end-user platform, with a median time of $33s$. Table 4.3 shows the 25^{th}, 50^{th} and 75^{th} percentiles of the execution

[3]Paying the minimum CPM allowed by the DSP, as the goal is maximizing the number of impressions and not their quality.

Device	Percentiles		
	25th	50th	75th
Mobile	7.8s	30.1s	105.9s (>1min)
Desktop	14.3s	33.6s	110.7s (>1min)

Table 4.3: Execution time percentiles per device type.

window for desktop and mobile devices. It shows significant differences in the execution window depending on the platform: 75% of ads rendered on the desktop are active for at least 15s whereas this decreases to just 8s for mobile devices.

This analysis suggests that being time-conscious is critical to the experiment's design. Tests should launch and complete quickly, and should be scheduled opportunistically to make use of long-running ad displays.

4.1.3 Network Measurements in the Browser

Modern web browsers run powerful JavaScript engines that offer a rich suite of networking libraries to web developers. Many of the client-side APIs used in AdTag have been standardized by the web community:

- **XMLHttpRequest (XHR):** This API allows clients to communicate to servers via HTTP(s) protocols, allowing custom crafted methods, headers, and payloads [86].
- **WebSocket:** This standard allows delivering custom application-level data in a bi-directional manner between a client browser and a server over TCP [87].
- **Network Information API:** Most DSPs claim to be able to run ad campaigns restricted to mobile devices. However, mobile devices may not necessarily be connected over a cellular link: users can also access the Internet from their smartphones over WiFi. AdTag can use the Network Information API [88] supported by Firefox and Chrome browsers on Android to obtain ground-truth about the access link technology of the device.
- **WebRTC:** This API, not completely standardized yet by the W3C but already fully supported by most browsers [89], allows communicating custom application data (namely for video and audio) over a bi-directional UDP channel. WebRTC also provides access to many of the utilities required for establishing peer to peer connections, including methods to perform NAT traversal.

As opposed to programming languages with a full network stack like Java and Flash, JavaScript networking APIs have several technical constraints that limit our ability to implement certain network measurements. Restrictions on WebSocket and WebRTC do not allow the creation of data directly over TCP/UDP such that they could be used to exactly mimic and modify existing application-level protocols. Even though a WebSocket can carry arbitrary unencrypted data over TCP, it requires a connection phase between client and server using HTTP(s) before

proceeding with any data transfer. It also has its own custom headers, which encapsulate the data. WebRTC UDP is restricted in a similar manner, requiring DTLS encryption for any data channel and encapsulating the data channel within SCTP. As a result, AdTag will not be able to directly test certain UDP-based protocols and Internet sub-systems like DNS [14].

Nevertheless, the implications of what these APIs allow in terms of network measurements are still enormous as it is demonstrate in Section 4.1.3.2. As UDP traffic via WebRTC is delivered over SCTP at the application level, it provides a good balance between accuracy and efficiency for network measurements. This allows to choose whether SCTP data is guaranteed to be delivered in order, reliably, neither, or both. Consequently, performance reliant tests, such as latency or timeout tests, can be more accurate than those done over TCP, where overheads occur due to mandatory inclusion of reliable/in-order delivery and state maintaining.

Alternatively, tests where accuracy is the priority such as outbound port scans, can take advantage of the added utility of probing the lower levels with UDP flows while still producing reliable results at the application level.

4.1.3.1 Browser support

The advertising campaign is instrumented to measure browser's API support in the wild. Table 4.7 shows a breakdown of dominant browsers, according to their `User-Agent` field, that it is identified during the global campaign, ordered by the percentage of successful measurements run on each one of them over the total. For each browser and JS API, we report the minimum version supporting a given API. `n/a` indicates that a given browser does not support such technology yet. The percentage value for each technology reports the percentage of users for a given browser running at least the minimum browser version supporting this technology.

45% of the most common browsers (shown in Table 4.4) of our global campaign, were launched on browsers supporting the three networking APIs simultaneously. As it shows, most measurements come from Chrome users, which guarantees that a large number of tests will be executed on browsers with full API support. The analysis also reveals that mobile browsers provide more limited APIs than their desktop counterparts. Unfortunately, DSPs do not allow targeting

Browser	%	WebRTC		WebSocket		WebWorker	
		Version	%	Version	%	Version	%
Chrome	34.5	49	97	49.0	97.0	49.0	97.0
Mobile Safari	21.7	n/a	n/a	9.3	14.3	9.3	14.3
Chrome Mobile	19.8	59	56	59.0	56.0	59.0	56.0
Firefox	5.7	52	88	52.0	88.0	52.0	88.0
Safari	4.6	n/a	n/a	9.3	95.0	9.3	95.0

Table 4.4: Top 5 most common browsers in the global campaign and the minimum version supporting relevant JS APIs. The percentage value for each API is computed over the total number of browsers of a given kind.

end users according to API support. Therefore, understanding browser API support is key to plan complex measurement campaigns and adjust their budget accordingly by predicting how many impressions will be required to obtain statistically representative results.

4.1.3.2 Use cases

JavaScript libraries can be used to bootstrap a wide range of network measurements through AdTag. Some may require only instrumenting the client-side JavaScript. However, others may require interaction between the client and collaborative server, as illustrated in Figure 4.1. Next, we present a non-exhaustive list of interesting network measurements—some based on previous measurement tools using full-stack programming languages—that can be successfully ported to JavaScript.

- **Detecting middleboxes and traffic manipulation:** A careful instrumentation of both the client- and the server-side of AdTag can reveal the presence of HTTP and HTTPS middleboxes and if they perform any traffic manipulation. Using the WebSocket and XHR libraries, it can force the client and the server to speak custom variants of HTTP over TCP, a technique proved valid to identify and characterize HTTP(s) proxies [16, 17, 65].
- **NAT detection and characterization:** WebRTC allows performing STUN and TURN requests that can be used to study NATs at scale. In this case, a STUN/TURN server is required. Because of the direct access of the user to proper protocols over UDP for NAT traversal through STUN and TURN, the client can obtain data regarding its IP, probe for NAT existence, check for middlebox state and identify port allocation policies. These features were previously limited to Java-based frameworks like NAT-Analyzer [90] and Netalyzr [65, 91].
- **CDN performance:** CDN performance highly depends on the replica selection algorithm and DNS resolution. AdTag clients can fetch one (or more) small object(s) from a CDN provider hence providing detailed performance metrics such as the time-to-first byte (TTFB), and the location of the assigned replica.
- **IP classification:** AdTag-based tests can help to classify a given IP address along different dimensions: by network type (*i.e.,* residential, enterprise or mobile) and characteristics (*e.g.,* proxied or NATed). The mapping of an IP to User-Agents reveals the sharing condition of an IP address. This can complement existing IP intelligence datasets, helping to further contextualize the data provided by IP blacklists, WHOIS records, and GeoIP services [92].

4.1.4 Discussion

In this Subsection we have presented and discussed AdTag, a measurement platform that leverages online advertising to quickly conduct experiments at global scale. AdTag leverages ad

networks' ability to target specific client populations in order to analyze the Internet from the edge of the network. We discussed AdTag's design space, including its ability to target specific networks and devices, typical campaign costs, as well as technical challenges imposed by browser runtimes. Common JavaScript APIs can serve to detect and characterize middleboxes such as proxies and NATs, analyze CDN performance, or furnish the input for IP address classification. The empirical experiments placed ads in 9 ad networks and confirm the ability to target specific ISPs and geographic locations at low cost, facilitating large-scale data collection within days.

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

The 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 work, we modify AdTag, 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, leveraging the outreach capacity of online advertising networks can distribute lightweight JavaScript-based DNS measurement scripts. To showcase the potential of this technique, we launch two separate ad campaigns that triggered more than 3M DNS lookups, allowing 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 $\sim$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.

4.2.1 Background

Previous research efforts used three methods to study different aspects of the recursive DNS infrastructure at a global scale. Table 4.5 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 [19, 93, 94, 95] 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 work used proprietary telemetry provided by major CDN providers, ISPs, and DNS operators [96, 22, 94, 97]. 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 companies owning planetary-scale infrastructure. As a result, this data is often inaccessible for independent academic researchers and most practitioners.

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 [19]	151 / -	DNS Scans	Performance	No	13.6M	6k	1 month
M. Müller et al. [95]	- / 3.3k	RIPE Atlas	Both	Yes	9.7k	11k	5 days
M. Almeida et al. [96]	-/94	Mobile network		No	19M / 5k	-	1 month / 1.5 year
F. Chen et al. [97]	102 / -	CDN Telemetry	Both	No	3.6M	584k	15 days

Table 4.5: Comparison of our methodology with previous DNS measurement studies from the edge of the network.

DNS observatories have been recently developed and proposed by the research community and operators to enable the access to large-scale DNS data [98].

- **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 [99, 100, 21, 96, 101]. 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, Mao et al. [102] 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.

4.2.2 Measurement Method

As demonstrated in Table 4.5, 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, 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.

4.2.2.1 JavaScript-based DNS Measurements

We design and develop HTML5-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 recording in a lawful, privacy-preserving, and user-safe fashion IP-level information[4] of both the client (d) and the recursive resolver (R). Figure 4.3 details the process.

[4]It only records the $/24$ subnetwork of the user, and the public IP address of the recursive DNS resolver providing support to the user.

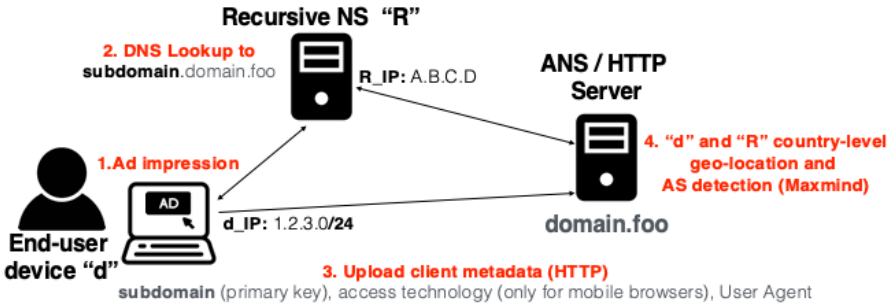

Figure 4.3: Method and data collection

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 the measurements in the background.

In parallel to the DNS resolution process, the JavaScript code opens a connection with the 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 [88] (*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 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 [82]. The random subdomain generated for each user allows identifying 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:

$$< R\text{'s IP address, } R\text{'s AS, } R\text{'s country geolocation, anonymized } d\text{'s public IP address, } d\text{'s AS, } d\text{'s country geolocation, } d\text{'s type of connection}^5 >$$

4.2.2.2 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 DSP or Ad Networks. Depending on the budget, [6] it is possible to obtain between millions to hundreds of millions of daily ad

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

[6] The cost of an ad display campaign is defined based on the CPM. CPM can be as low as $0.01.

impressions (i.e., DNS measurement samples) in real end-users' devices. A beneficial side effect of this distribution method is that ad providers allow setting 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).

4.2.2.3 Dataset

For running the DNS measurements we launch 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 the campaigns was $450 (average CPM ∼$0.12). Despite the 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. We compare the 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 the dataset offers a representative picture of the DNS subsystem. The two campaigns allowed unveiling the presence of 76k different DNS recursive servers distributed across 49k /24 IP prefixes in 14k ASes. The dataset is available to the community at `http://dns-analytics.netcom.it.uc3m.es:5000`.

4.2.2.4 Method Limitations

The current method and dataset present several limitations which is described 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). This natural bias can be tackled with a higher investment in targeted advertising campaigns to access underrepresented ASes and countries like the case of Africa and Oceania users.
2. The 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 [83]. RIPE IPmap[7] was also considered, but the response time and coverage of this service do not meet the requirements.
3. The 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 the 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 [91]. The presence of middleboxes can be inferred statistically – e.g., a significant

[7] `https://openipmap.ripe.net/`

large number of requests coming from a given IP address. For this purpose, dedicated targeted experiments can be run.

4.2.3 DNS Global Infrastructure

The first step in this 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 the explained methodology, sorted by the number of unique IP addresses recorded are shown in Figure 4.4. It states 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, it indicates so instead of the country.

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. This methodology does not allow distinguishing 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.

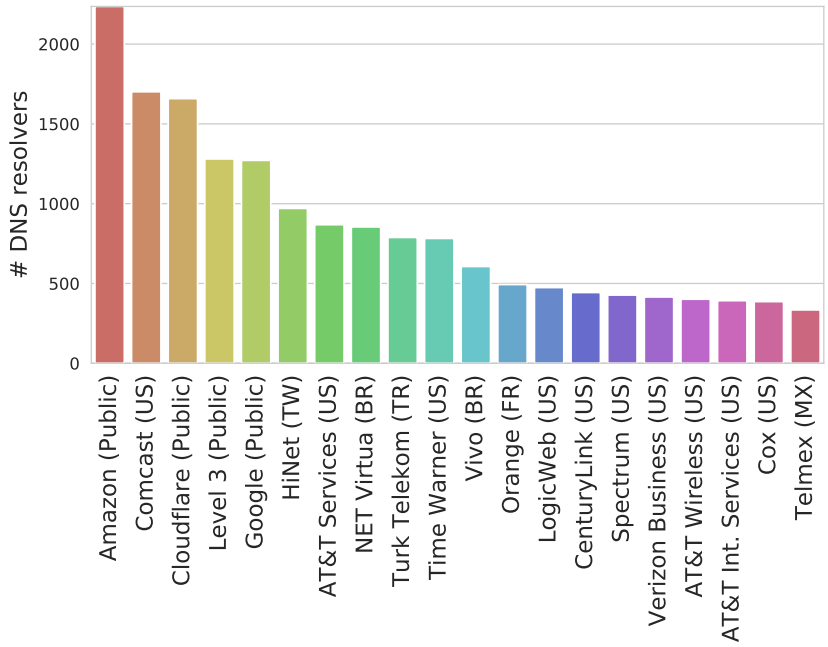

Figure 4.4: Top 20 organizations by the number of public IP addresses hosting recursive DNS resolvers

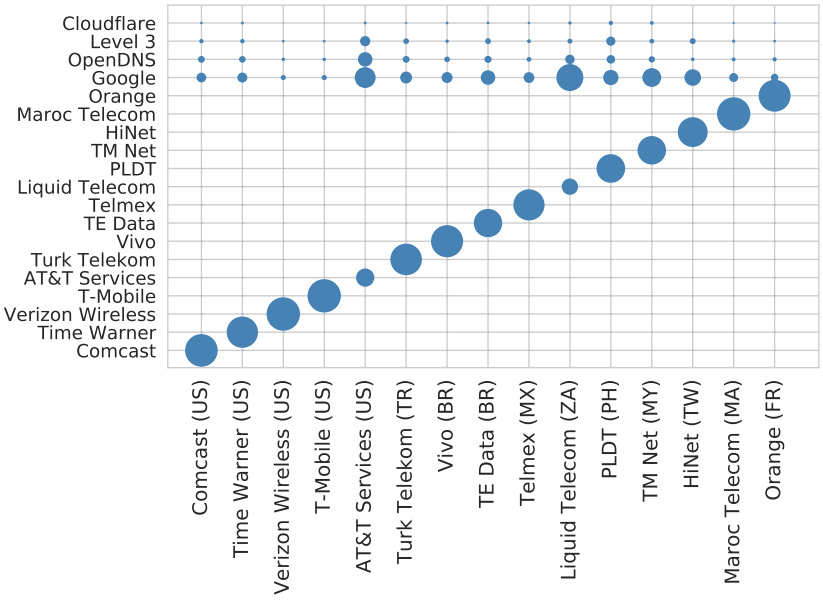

Figure 4.5: Fraction of DNS requests triggered by users from relevant ISPs (x-axis), served by machines hosted in their own infrastructure or third-party DNS providers (y-axis)

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 4.5 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 DNS providers (y-axis). The size of the circle shows the fraction of the DNS lookups triggered by each type of resolver.

It showcase that most of the DNS requests observed by the deployed 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.

4.2.4 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 [103]. 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. The results indicate that 99% of the eyeball ASes in the 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 [104] – 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. As mentioned above, we are not able to individually analyze cases in which multiple resolvers are hosted behind the same IP address. 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. The 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.

4.2.5 Third-Party DNS Providers

As presented in Section 4.2.3, third-party DNS providers play a relevant role in the DNS subsystem worldwide. Previous studies performed small-scale experiments to compare the performance of third-party DNS providers with ISP-provided ones [22]. 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 the measured coverage of the DNS infrastructure for these third-party providers compared with the /24 IP blocks publicly announced by Google DNS [105],OpenDNS [106], and Cloudflare [107], and we obtained 75%, 79%, and 43% overall coverage, respectively.

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

Table 4.6: DNS infrastructure metrics continent-based for the users using Public DNS resolvers

1. **Use of Third-Party DNS Providers:** 13% of the global DNS requests handled by the deployed NS come from 21% of the $/24$ IP prefixes in the collected dataset which belong to the four considered third-party DNS providers. Among the four providers, Google is the most popular one by attracting almost 75% of all the requests coming from third-party 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 [108]. Similarly, Geoff Houston showed that Google's DNS adoption was around 7% in 2013 [5]. Considering 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 4.6 shows the percentage of DNS requests handled by the deployed NS coming from third-party resolvers in each world continent. When analyzing at the country-level, developing countries tend to present the largest adoption of third-party providers.The research literature suggests that end-users resort to third-party DNS resolvers to obtain better performance and reliability [109], or to circumvent censorship and obtain better privacy protection [110]. We study whether the 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, the dataset suggest that the use of third-party DNS providers in developing countries may impair DNS and web performance. Table 4.6 shows the percentage of DNS lookups resolved by the NS coming from ISP-provided and third-party DNS resolvers per continent. It 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. 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 third-party 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 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 [109]. Then, hypothetically 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 DNS providers as seen by the NS with Reporters Without Borders' (RWB) World Press Freedom index per country[8]. 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 4.6. 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 third-party 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 third-party DNS providers to enhance their privacy and security, and avoid Internet censorship.

3. **Comparison of Third-Party DNS Providers:** We conclude this Subsection with a comparison of the IP infrastructure of third-party DNS providers using the metrics introduced in Section 4.2.4. 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). 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 the 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.

[8] `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

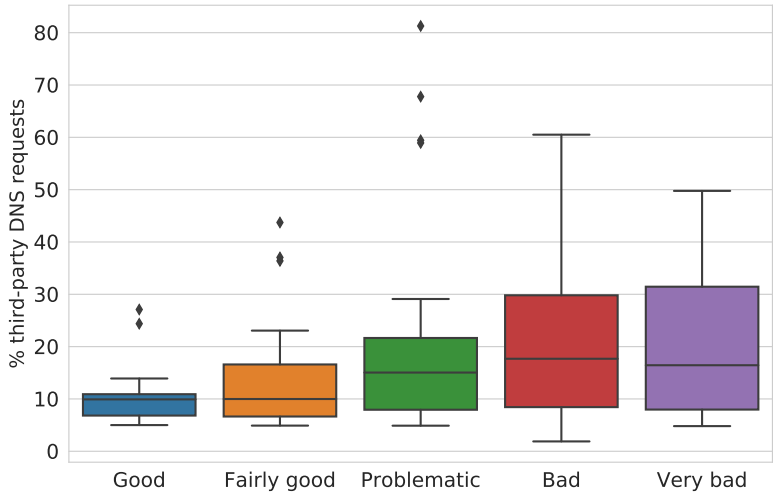

Figure 4.6: Distribution of the percentage DNS requests recorded by the NS as coming from third-party DNS providers across countries grouped by their Reporters Without Borders World Press Freedom index category.

4.2.6 Mobile vs. Fixed ISPs

The last part of this work is a comparative analysis of the DNS infrastructure provided by ISPs offering both mobile and fixed-line (*e.g.,* DSL and Cable) network access. Using 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 have 78% (22%) of fixed (mobile) connections in the dataset.

We compare the overall size of the DNS infrastructure allocated to serve mobile and fixed users. The 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. Therefore, we study more in depth the infrastructural commonalities for this type of dual ISP. To obtain statistically representative results, we restrict the 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 4.7 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 the dataset. Interestingly, it is remarkable a clear trend in which ISPs providing dual access tend to use dedicated DNS resolvers for their mobile and fixed networks. Some examples are

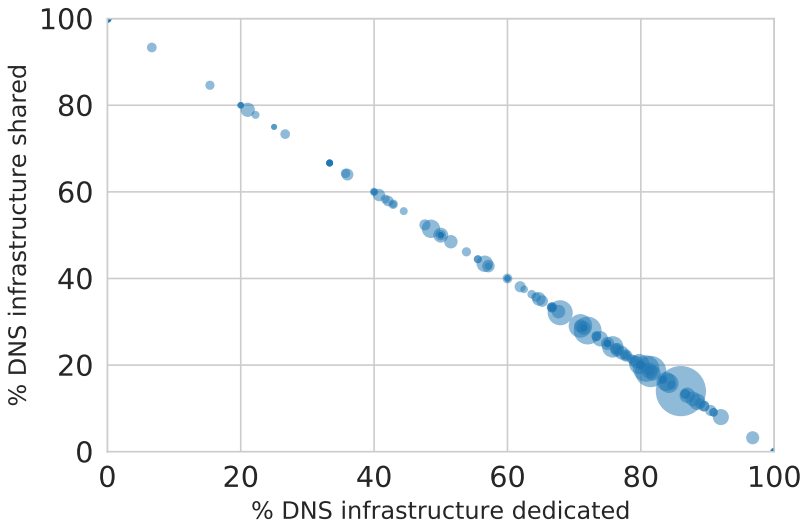

Figure 4.7: Percentage of recursive DNS resolvers for ISPs offering both fixed and mobile network access that serve both access technologies or just one of them.

NTT Docomo (JP), and Vivo (BR) where 97% and 85% of their DNS infrastructure seems to be dedicated according to the obtained 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 work, this analysis demonstrates the potential of the proposed methodology to perform large-scale analyses to identify new research questions.

4.2.7 Discussion

This work presents a reproducible, lightweight, and cost-effective measurement technique suitable to study the global DNS infrastructure and their usage by regular users. The 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.

By running two small measurement campaigns, 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. The empirical results indicate that 13% of the global DNS lookups are resolved by third-party DNS providers like Google DNS rather than by ISP-provided DNS resolvers. This 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. This work also show that ISPs providing both mobile and fixed access tend to decouple the DNS infrastructure serving each type of network access.

4.3 An ad-driven measurement technique for monitoring the browser marketplace

In this work we present a novel active measurement methodology for monitoring the browser market landscape. It leverages the display ads delivered through online advertising campaigns to collect the browser brand and version of the device receiving the ad. While providing a similar accuracy to traditional techniques based on passive measurements, this methodology offers some advantages: (i) a lower entry barrier for researchers and practitioners interested in measuring the browser marketplace; (ii) it allows targeted measurements, which can be useful to fix biases in the data sample or to analyze specific aspects of the browser market. In the next subsections we analyze the performance, accuracy, and capabilities of the proposed methodology through real experiments that overall produced more than 6M measurements.

4.3.1 Traditional Methodology for Monitoring the Browser Marketplace

The traditional methodology for monitoring the Browser Market landscape consists of installing a tracking code in a large number of websites. This code collects the User-Agent associated with each visit. The User-Agent serves as a browser identifier which reveals the browser's brand and version. Moreover, the tracking code can collect some other information such as the IP address of the device visiting the website. The IP address can be processed to retrieve the geolocation of the visit.

Several companies implement this traditional methodology: StatCounter, W3Counter, Net Applications, Wikimedia, etc. Each of them has access to a different set of websites and thus report their results based on an independent set of visits. These companies monitor between thousands (W3Counter [28] or Net Market Share [111]) and millions (StatCounter [29]) of websites. These companies provide datasets including up to 15B visits per month as in the case of StatCounter.

4.3.1.1 Limitations

Next we discuss the main limitations of the described traditional methodology:

- *High Entry Barrier*: The described traditional methodology presents very high entry barriers, limiting its use to a handful of companies able to install their monitoring code in (at least) thousands of websites. For instance, the research community is (in general) excluded from the use of this methodology since it is very unlikely that a research team can have access to a such large pool of websites.
- *Bias in data samples*: Traditional techniques register the visits to their monitoring websites as data samples. This means that a user visiting 50 times with the same browser a web page generates 50 data inputs. This is potentially a source of bias, since heavy visitors to the monitoring websites would have a higher weight in the final market distribution across

brands and versions. Furthermore, the reach of this methodology is limited to the users that visits those websites, which can not be the real distribution of the browsers' market share.

- *Other biases*: Despite each of the companies using the traditional methodology accounts with a large data sample, we find discrepancies among their reported browsers' market share. The reason is that the collected dataset may be affected by different biases: geographical biases (having a higher representation from users located in certain countries), OS biases (having a higher representation from users of a specific Operating System), Type of device bias (having a higher representation of mobile or desktop devices), etc. These biases are associated with the pool of websites used by each company. For instance, if the pool of websites is predominantly in a specific language (e.g., English), there will be a geographical bias towards countries speaking such language.
- *Does not allow targeted measurements*: The vision of the browser marketplace depends on the users that connect to the monitoring websites. This fact is out of the control of the company. Hence, even if a bias is identified in the sampled data (e.g., a geographical bias), the company has many difficulties for fixing it because they cannot modify the demographic or geographic properties of the user base connecting to the monitoring websites. Instead, if the company had some capacity to define the targeting population for their measurements, it could fix identified biases in a simpler manner.

4.3.2 Active Measurement Based Methodology for Monitoring the Browser Market Landscape

The goal is to define a methodology that overcomes the limitations of the traditional techniques for monitoring the browser marketplace discussed in Section 4.3.1. In particular, this methodology should meet the following requirements: (i) low entry barrier so that any person, company, or research team interested in monitoring the browser market can do it at a reasonable cost; (ii) it should allow targeted measurements. This will serve to fix identified biases in the data sample, but also to conduct specific analysis of the browser market landscape such as analyze the market share for a particular demographic group (based on age and/or sex), analyze the presence of insecure browser versions and identify its associated IPs, etc; (iii) it should guarantee that each browser instance represents a single data sample in the collected dataset.

To achieve these goals, we propose the utilization of AdTag, running active measurements, contrary to the passive measurements used so far. In particular, this approach relies on the online advertising ecosystem. Most websites have embedded ads, we propose to use these ads as vantage points to collect the User-Agent and IP address of the device connecting to webpages where an ad, under control, is shown. It is estimated that around a trillion ads are delivered every day. This number provides a solid basis to meet the required scalability to monitor the web browser marketplace. Moreover, the online advertising ecosystem offers the needed functionality to achieve the goals described above. First, it allows running targeted advertising campaigns based on different

parameters including: demographic characteristics (age and sex of the user), geographic location (country, region, and even cities), type of device (desktop vs. mobile), Operating System, etc. Second, any person or company can use one of the hundreds of available online advertising vendors to configure their own campaigns using the monitoring methodology proposed. Third, online advertising campaigns offer a configuration parameter referred to as *Frequency Cap*, which determines the maximum number of times an ad is shown to a specific user, i.e., browser in this case. By setting up the *Frequency Cap* equal to 1, only one a specific browser instance will contribute a single data sample to the dataset. Finally, the experiments have a low cost. As a reference, the cost of 1M measurements ranges between $10 and $100 approximately, depending on the vendor. Therefore, the entry barrier of the proposed methodology is significantly lower than the one imposed by the traditional methodology.

4.3.2.1 Details of the Methodology

Online advertising offers different forms of ads: video ads, display ads, search ads, etc. AdTag leverages display ads. These are the typical banner ads that appear on most websites. Display ads are currently developed in HTML5. Then, they can include JavaScript code. We take this opportunity to insert a custom JavaScript code to collect the User-Agent information.

We create our own HTML5 ad, which includes a JavaScript code for collecting the User-Agent information. We set up an advertising campaign with our instrumented ad in a DSP. Once this campaign is started each time an impression of our ad is delivered, the JavaScript code retrieves the User-Agent of the browser receiving the ad. Then, the JavaScript code establishes a TCP connection with a central server where we store the collected information. The server obtains the IP address of the device receiving the ad from this TCP connection[9].

Therefore, each time our ad is displayed we collect a tuple including the following information: <*timestamp, IP address, User-Agent*>.

Each of these tuples is processed. We use the GeoLite MaxMind database[10] to map the IP address to a geographical location (country and region) and two Python libraries, `user_agents`[11] and `httpagentparser`[12] to map the User-Agent to its browser brand and version as well as to obtain the OS and OS's version. After this, the IP address is anonymized using hashing techniques. Therefore the final tuple stored in a central database is: <timestamp, hashed IP address, country, region, browser's brand, browser's version, OS, Os's version>.

Finally, this methodology allows performing active targeted measurements. As we have described earlier, an advertiser can configure display ad campaigns targeting specific audiences, which are defined by a combination of geographical location, demographic characteristics, users' interests, device type, operating system, etc. These options are available in most DSPs.

[9] Note that an alternative and more lightweight manner of doing this is sending an HTTP GET message to the server. However, certain ad-tech providers block GET requests if they come from a third party.

[10] `http://www.maxmind.com`

[11] `https://pypi.org/project/user-agents/`

[12] `https://pypi.org/project/httpagentparser/`

4.3.2.2 Performance Evaluation

We have run our server code on a standalone machine (24 2.4GHz cores, 64GB RAM). Under this setting, it is able to handle over 100k simultaneous connections. Note that in case more resources are needed, multiple servers can be installed using load-balancing techniques to distribute the load among them. Therefore, the proposed methodology offers the necessary scalability to collect (at least) hundreds of millions of measurements every day.

Moreover, our methodology is meant to run in the wild through real ad campaigns. Hence, it may be affected by different type of errors, which may prevent collecting the information from some ad impressions: browser extensions preventing the deployment of ads or JavaScript code (e.g., ad blockers [112] or no-script [113]), network problems preventing the establishment of the connection, problems in the execution of the JavaScript code of our ad, etc. We have observed that on average our methodology was not able to collect information for 15% ad impressions. This rate was computed as the ratio between the number of ad impressions recorded with our methodology and the total number of ad impressions reported by the DSP used to run the ad campaigns. A careful analysis of these losses indicates that most of them are due to the fact that the server used in this work was running on an academic network that offers good performance, but it is not designed to support large-scale experiments receiving a large number of connections. Indeed, we have run our methodology for a different research project within the infrastructure of an ad-tech provider network (this one designed to handle a large number of connections) experiencing a much lower fraction of losses below 5%.

4.3.2.3 Limitations

In this Subsection we discuss the main limitations of our methodology with respect to the traditional passive measurement techniques.

- *Scalability:* Some well-established companies using the traditional methodology can reach in the order of millions to hundreds of millions measurements per day. However, just a handful of companies have the coverage and infrastructure to reach such scale. Our methodology has the theoretical capacity to achieve such magnitude, but it would require a recurrent high investment. If we assume a CPM of $0.10, obtaining 1M (100M) daily measurements would cost $100 ($10000). Therefore, reaching equivalent scalability as the one offered, for instance, by StatCounts seems unfeasible due to the high economic cost. However, reaching scalability in the order of a few millions of measurements per month is affordable for interested companies or research teams. As we will show in section 4.3.3, a few million measurements suffice to obtain results similar to those presented by companies using the traditional methodology accounting with billions of measurements every month.
- *Data sample biases:* As in the case of traditional measurements, our methodology is subject to suffer from biases in the obtained data sample (e.g., underrepresented geographical areas or demographic groups). However, once the bias is identified, our methodology al-

lows taking correction measures by defining complementary ad campaigns that target the underrepresented audiences. This is a clear advantage over the traditional technique, which cannot take straightforward countermeasures to existing biases in its data sample.

- *Device Resource Consumption:* Our methodology requires the device receiving the ads to devote some computation resources to execute the JavaScript code and some bandwidth to send the collected information to the central server. Contrary, the traditional methodology does not require to use any resource from the device, since it uses passive measurements. We have carefully evaluated the resource consumption in lab experiments. The computation resources used by our JavaScript code are negligible (executing a call to the browser API to retrieve the User-Agent and establishing a TCP connection). Moreover, our JavaScript code sends a message of 600 Bytes to the central server. Hence, the consumption of end-users' data is also minimal.

4.3.2.4 Implications for businesses

As discussed in the introduction, having an accurate solution to estimate the browsers' market share is important for several companies, including different types of software development companies, online security firms, companies operating in the digital marketing ecosystem, etc. One of the main problems of traditional solutions is that, as mentioned above, they require a monitoring infrastructure only available to a few companies. Indeed, most companies in the mentioned businesses (software development, online security, and digital marketing) do not have such infrastructure. However, any of them can use our proposed solution, since it does not require to have any pre-existing large-scale infrastructure and can be deployed on-demand through any of the hundreds of available providers. Therefore, our solution allows, for the first time, the democratization of the monitoring of the browser marketplace to any company regardless of its size.

Second, our technique provides targeting capabilities, not offered by traditional solutions. This offer companies the possibility to perform specific studies based on their own needs. For instance, as we will show in section 4.3.3, a security company can use these targeting capabilities to identify installations of vulnerable browsers and take appropriate protection measures. Also, we will show how these targeting capabilities would allow a company to perform an accurate study of the browser market share in underrepresented geographical areas (e.g., a country). This could be, for instance, useful for a software company developing a web application in one of such geographical areas in order to identify the most popular browser versions and make sure the web application works properly for them. Finally, the demographic targeting capabilities can be again leveraged by businesses to understand the browser market share across the targeted demographic group. For instance, a company developing an online game for males between 20 and 30 years in a specific geographical area (e.g., France), can use our solution to characterize the browsing marketplace in that region and optimize the performance of the game for those browsers most commonly used by the targeted demographic group.

In summary, companies interested in understanding and characterizing the browser marketplace can benefit from our solution for two main reasons: 1) its accessibility for companies of any type and size and 2) its targeting capabilities.

4.3.3 Experiments

In this Subsection, we evaluate the ability of the proposed methodology to monitor the web browser marketplace. To this end, we first present the results of running a large-scale general purpose non-targeted campaign. This campaign serves to assess the accuracy of the estimation of the browser market share reported by our methodology in comparison with the reports of companies using the traditional methodology. Afterward, we present the results of two targeted experiments, whose goal is to showcase the targeting capacity of our methodology. In particular, we run a geographically targeted experiment focused on Albania (a country underrepresented in the general purpose dataset). Secondly, we run a targeted experiment focused on identifying the presence of outdated versions of browsers presenting security vulnerabilities. To this end, we configure targeted campaigns to old versions of OSes, which are likely running obsolete versions of browsers. Finally, we run a demographic campaign to identify the most popular browsers used among a specific audience. To achieve this goal we configure a targeted campaign for people between 18 and 25 years old, in Italy, using mobile devices.

4.3.3.1 Measurement Platform and Experiment Setup

We configure our ad campaigns through a DSP, which allows us to set up the following targeting parameters: geographical location, device type, OS brand, OS version, specific User-Agent, demographical information, etc. From these, in this work, we use only three targeting parameters, the geographical location, the demographical information, and the OS brand and version. Next, we describe the specific set up of each of the four run experiments:

1. ***Large-scale non-targeted campaign:*** We configure a campaign in which we do not select any targeting parameter. This campaign is run through 9 well-known vendors including Google, AOL, Pubmatic, etc. We have run this experiment twice in May 2017 and Oct 2017, generating more than 3M measurements in each of the experiment. We present the combination of both datasets for the results of this work.

2. ***Geographic-targeted campaign:*** We use the large-scale dataset obtained in the previous experiment to identify underrepresented countries (i.e., countries with a very low number of samples) and chose one of them to run a targeted ad-campaign to show how our methodology (contrary to the traditional one) can take actions to correct biases. To this end, we configure a targeted campaign to deliver ads to Albania. We obtain a total of 3k measurements in a couple of days (note that in the large-scale dataset we only have 14 entries from Albania, in one week).

Browser	total
Chrome	2435760 (40.85%)
Chrome Mobile	1262945 (21.18%)
Safari Mobile	898924 (15.08%)
Firefox	428896 (7.19%)
IE	277716 (4.65%)
Safari	181277 (3.04%)
Edge	133241 (2.23%)
total	5618759 (94.22%)

Table 4.7: Browsers' market share obtained from our general purpose large-scale dataset. Results show the absolute number of samples and its equivalent percentage per OS.

3. ***OS-targeted campaign:*** Our goal is identifying outdated browsers with severe security vulnerabilities, which represent serious security threats. Note that our methodology would allow to identify the IP addresses associated to those browsers[13]. As a result, the Security responsible of the institution or provider hosting such an IP can be warned. To identify these type of browsers we configure ad campaigns targeting old versions of OSes, in particular we target Windows XP, Mac OS X v10.0 (known as *Cheetah*) and Linux v686. Instances of outdated browsers are likely to run in old version of OSes. As a result of this experiment we obtained a total of 345k measurements.

4. ***Demographic-targeted campaign:*** To showcase the demographic targeting capabilities of the proposed solution, we have configured a campaign targeting the following demographic group: young people (ages between 18-25) in Italy using mobile devices. We obtain a total of 13k data samples that provides an estimation of the mobile devices and browser market share across the targeted demographic group.

The overall cost of all these experiments was around $720 at an average CPM of $0.5. These numbers offer a cost reference that confirms that our methodology presents a low entry barrier in comparison with the traditional methodology that requires direct access to a large number of websites.

4.3.4 Results

4.3.4.1 *Accuracy of Estimation of Browser market share*

Using our large-scale dataset, we have computed the market share of different mobile and desktop browsers. Results are presented in Table 4.7. Moreover, Table 4.8 shows the list of countries where each of the three most common browser brands (Chrome, Safari, and Firefox) shows the highest presence. We merged the desktop and mobile platforms for this table.

[13]For ethical reasons we only stored an anonymized version of the IP address.

Browser	Country	total (%)
Chrome	United States	865299 (23.39)
	Brazil	289045 (7.82)
	United Kingdom	276000 (7.46)
	Turkey	221751 (5.99)
	Canada	191284 (5.17)
	Mexico	159655 (4.31)
	France	136857 (3.7)
	Argentina	118946 (3.2)
Safari	United States	432906 (40.07)
	United Kingdom	191143 (17.69)
	Canada	85938 (7.95)
	Australia	48396 (4.48)
	France	43137 (3.99)
	Ireland	32306 (2.99)
	Netherlands	27645 (2.56)
	Germany	23351 (2.08)
Firefox	United States	117141 (26.62)
	United Kingdom	39722 (8.91)
	Germany	38708 (8.68)
	France	31652 (7.10)
	Poland	24625 (5.52)
	Canada	17794 (3.99)
	Brazil	16844 (3.78)
	Spain	13130 (2.95)

Table 4.8: List of countries where each of the major browser brands (Chrome, Safari and Firefox) have highest presence.

Finally, Table 4.9 compares the market share reported by our methodology and other companies using the traditional methodology. If we take StatCounter as a reference for comparison (since it is the company accounting with a larger sample of data), we observe that our results present an average difference of 4 percentage points across the different browser brands. This difference is equivalent to that shown by other companies using the traditional methodology. In particular, StatCounter and NetMarketShare show an average difference of 3 percentage points.

The differences of reported results across different systems are caused by the distinct biases present in each dataset. In the lack of ground truth, it is not possible to conclude which report presents the closest results to such ground truth. However, all reports, including ours, show a high coherence, indicating that all of them are reasonably accurate.

Therefore, despite the fact that our methodology cannot reach, in practice, the volume of samples that some companies achieve using the traditional methodology, we can conclude that the results accuracy is expected to be similar as the one achieved by the traditional methodology.

Browser	StatCounter	W3Counter	NetMarketShare	AdTag
Chrome	62.7%	57.4%	63.88%	64.88%
Safari	15.89%	13.5%	17.64%	18.15%
Firefox	5.07%	6.8%	4.76%	7.22%
IE & Edge	4.68%	6.8%	5.72%	7.05%
Opera	2.55%	2.4%	0.89%	0.92%

Table 4.9: Market share reported by StatCounter, W3Counter, and NetMarketShare compared with our methodology, AdTag

4.3.4.2 *Geographic Targeting*

Our general purpose measurement campaign shows several underrepresented countries. One of them is Albania that contributes just 14 data samples out of the 3M. In the traditional methodology, such events cannot be easily addressed since the composition of the data sample is not under the control of the company issuing the measurements. To address this geographical bias under the traditional methodology, a company may try to add to its websites' pool the most popular pages in the underrepresented countries. This action may take time (it requires reaching the administrator of the website, establishing a negotiation, a (economic) compensation, etc.) and the success is not guaranteed. Instead, our methodology can straightforwardly address this type of biases since we can define targeted campaigns focused on specific geographical locations. The results of our *geographical-targeted campaign* proves it. We have run this campaign during 3 days, obtaining 3k data samples from Albania, addressing the under-representation problem of this country.

4.3.4.3 *Specific browser Targeting*

The described experiment produced a total of 345k data samples, distributed across pairs of browsers/versions we will analyze further. By the time we run this experiment, the stable versions of Chrome, Safari, and Firefox were 63, 59, and 11, respectively. For security reasons, all browsers recommend to update the engine to the latest version, and they have by default the option to update the version automatically. We consider that any browser with a version (at least) 4 years older than the mentioned version are likely linked to security vulnerabilities and thus, represent a security threat. Table 4.10 show the number of impressions served to browsers versions lower than the ones mentioned above.

While a general purpose measurement as the one in our large-scale experiment is able to identify some of these insecure browsers (25k out of the 5.2M of data samples from Chrome, Firefox. and Safari), a targeted study helps to unveiled a much larger number of them, as we demonstrated in this experiment. The traditional methodology does not have this capacity since its passive nature limits its ability to select a measurement target.

We are aware that some browser versions less than 4 years old are also vulnerable. However, the goal of our experiment is not to identify all possible browser versions presenting vulnerabili-

Browser	Stable version	Insecure version	# Impressions Insecure version
Chrome	62-63	31	13093
Firefox	58-59	30	11849
Safari	11	6	6829

Table 4.10: Number and percentage of impressions for old version usage of the most common browser brands, representing important security vulnerabilities

ties, but showcasing how our solution is valid to launch targeted experiments allowing to identify the presence of vulnerable browsers. Security researchers can use our solution to make a thorough analysis of the presence in the web of browsers with different type of vulnerabilities, even ranking them from most to least problematic vulnerabilities. However, such analysis is out of the scope of this work.

4.3.4.4 *Demographic Targeting*

The last experiment aims at showcasing the demographic targeting capabilities of our methodology. To this end, we configure a campaign with the following audience parameters: 1) Age: between 18 and 25 years old; 2) Country: Italy; 3) Device: Mobile. We run a 3-days campaign obtaining 13k data samples coming from Android and iOS (note that we got 126 data inputs from Windows Phone, which we consider negligible).

The results first indicate that Android dominates the mobile devices market in the considered demographic group since it accounts for 72% of the data samples, whereas iOS just account with 18%. In the case of browsers, the market share for the considered demographic group is as follows: 1) Chrome Android, 68,40% 2) Safari and Safari WebView, 20% 3) Android browser, 2.5% and 4) Chrome for iOS, 0.8%. These results indicate that young people in Italy prefer Android devices and Chrome browser.

4.3.5 Discussion

Measuring the browser marketplace is of interest for companies of different nature, researchers and regulators alike. Existing solutions, based on passive measurement techniques, offer great scalability. However, they present two main drawbacks: First, they require a large monitoring infrastructure, which makes them accessible to just a handful of companies. Second, their passive nature avoids them to offer targeting capabilities.

In this work, we have presented a novel solution from a technical perspective since it, for the first time, uses active measurements to monitor the browser marketplace enabling targeting capabilities. Moreover, our solution uses the online advertising infrastructure, which nowadays is a commodity used by tens of thousands of companies, as a measurement platform. This democratizes the measurement of the browser marketplace since, contrary to traditional solutions,

any interested company or researcher can use it. Finally, our solution offers sufficient scalability to measure the browser marketplace. However, we acknowledge that well-established companies using traditional solutions have reached a larger measurement scale than the one our methodology can achieve at a reasonable cost.

In conclusion, we believe our solution is more suitable for general use by companies and researchers due to its accessibility and targeting capabilities. However, those companies looking for immense scalability should opt for traditional solutions.

CHAPTER 5

APPLICATION IN ONLINE ADVERTISING

5.1 Independent Auditing of Online Display Advertising Campaigns

The reported lack of transparency of the online advertising market may seriously affect the interests of advertisers. In this section, we present a novel methodology that allows advertisers to independently assess the quality of display advertising campaigns. This methodology also serves to audit the accuracy and completeness of reports delivered by the vendor responsible for running a campaign. We have applied our methodology in 8 display ad campaigns configured in Google AdWords, which overall produced 160k ad impressions displayed in more than 7k publishers. Our results reveal that AdWords seems to provide incomplete information to advertisers. Specifically, we found that: (i) AdWords did not report 57% of publishers where ad impressions from our campaigns were delivered, (ii) AdWords reports a large fraction of contextually meaningful impressions based on (non-disclosed) criteria different from the publisher's theme, (iii) higher CPM investment does not lead to get impressions delivered to more popular publishers, (iv) AdWords does not offer default control of *frequency cap*, (v) around 10% ad impressions in two of our campaigns were delivered to IP's from Data Centers. The industry considers these IPs to be likely related to fraud. These findings should contribute to open a debate between advertisers and Ad Tech vendors to standardize the utilization of independent auditing methodologies as the one presented in this work.

5.1.1 Methodology

We have designed a methodology, that we refer to as Q-Tag, based on the methodology explained in Chapter 3. It is focused in HTML5 display ads. HTML5 allows creating ads using web technologies such as CSS or JavaScript. We leverage this opportunity by injecting a simple JavaScript code into HTML5 display ads that we buy trough an Ad Network. This code collects information about displayed ad impressions and sends it to a central server where it is properly stored in a database. The JavaScript code collects the following information: i) the URL of the webpage where the ad impression was displayed. Note that the domain part of the URL reveals the publisher; ii) the User-Agent receiving the ad impression; iii) user interactions with the ad. In particular, we collect mouse movements over the ad as well as click events. Moreover, we take advantage of the connection established between the device which received the ad impression and our server to obtain further information: iv) the IP address of the device receiving the ad, and thus, establishing the connection to our server[1]; v) the timestamp of the ad impression computed as the local UNIX time on the server at the instant of the connection establishment; iv) the *exposure time* of the ad computed as the duration of the connection measured at the server side.

We implement the described methodology employing widely used and lightweight technologies to guarantee efficiency, scalability and robustness. In particular, we use: (i) plain JavaScript for the code inserted in the ad; (ii) the WebSocket protocol [57] for transferring the information from the ad impression to the central server. Note that the information is transferred in the form of a string; (iii) Node.js JavaScript library [58] to parse and process the information received in the central server; (iv) MySQL and Python to store and process the collected datasets.

5.1.1.1 Limitations and Validation

The described methodology is directly applicable in ad formats that support JavaScript in a native manner, such as HTML5 ads. In other ad formats, such as images or video, this methodology would only work if the Ad Network allows to add a tracking pixel. Most Ad Networks and other trading platforms allow placement of 3rd-party javascript inside ads for collecting users' behavioural targeting data.

Moreover, most Ad Networks insert ads in a single (or a double) iFrame, therefore our JavaScript code will run inside this iFrame. There exists a widely extended security policy referred to as *Same-Origin* policy (SOP) [114], which avoids a code running as part of an iFrame tracking the activity in other parts of the webpage different from such iFrame. Hence the SOP avoids that our methodology collects information such as the upstream referrer (i.e., the website from where the user reached the current publisher). It also prevents us from collecting the position of the iFrame in the webpage, so that we cannot assess if the ad (or part of it) was shown in the

[1]Note that we use the IP address to extract meta-data information such as the Internet Service Provider association with a user. Afterwards, we anonymize the IP using hashing techniques.

Campaign ID	# Impressions	# Publishers	Dates (2016)	CPM	Targeted Keywords	Targeted Location
Research-010	5117	350	29-31 March	0.10 €	Research	Spain
Research-020	42399	1777	29-31 March	0.20 €	Research	Spain
Football-010	33730	1086	02-03 April	0.10 €	Football	Spain
Football-030	24461	1367	02-03 April	0.30 €	Football	Spain
Russia	4096	274	29-31 March	0.01 €	Research	Russia
USA	1178	136	29-31 March	0.01 €	Research	United States
General-005	8810	580	15-23 February	0.05 €	Universities, Research, Telematics	Spain
General-010	42357	1549	18-23 February	0.10 €	Universities, Research, Telematics	Spain

Table 5.1: Description of the 8 AdWords campaigns used to test our auditing methodology.

visible part of the screen. This limited our methodology, at the time we did this work, to measure an upper bound of the *viewability* metric presented in Section 2.2. This is, whether the ad was displayed more than 1 sec, but without knowing if (at least) 50% of it was shown.

We have tested our methodology in a lab controlled environment and confirmed its capacity to retrieve all the data described above. However, our methodology is expected to run in operational network environments and thus it is subject to different errors. Then, we cannot guarantee to retrieve information from every ad impression. Errors happening in the browser (e.g., untrusted JavaScript code not allowed to run due to the browser configuration or by an antivirus software), the network, our server, or in the connection establishment process would result in the affected ad impression(s) not being logged in our central server.

5.1.2 Ad Network and Datasets

We have applied our auditing methodology to campaigns configured in Google AdWords, which uses Google Display Network (GDN) to deliver display ads. We have selected this Ad Network due to the following two reasons: First, GDN is the most widely used Ad Network worldwide. It spans over 2 million publishers that reach over 90% of Internet users [115]; Second, GDN allows to run low budget campaigns, starting at few dollars. Then, using AdWords/GDN, we can test our methodology while respecting our budget restrictions. The main reason why we did not test our methodology in other Ad Networks is that they typically request an initial investment in the order of few thousands dollars prior to running the first campaign. This exceeded our available budget for this research.

To test our methodology we have run 8 different display advertising campaigns using Google AdWords. Overall we registered around 160k ad impressions distributed across approximately 7k publishers. We set-up campaigns with different duration, different CPM values as well as different targeted keywords and geographical locations. This diversity aims at reducing the chances that observed results are due to a specific campaign set-up. Table 5.1 summarizes the main properties of each campaign.

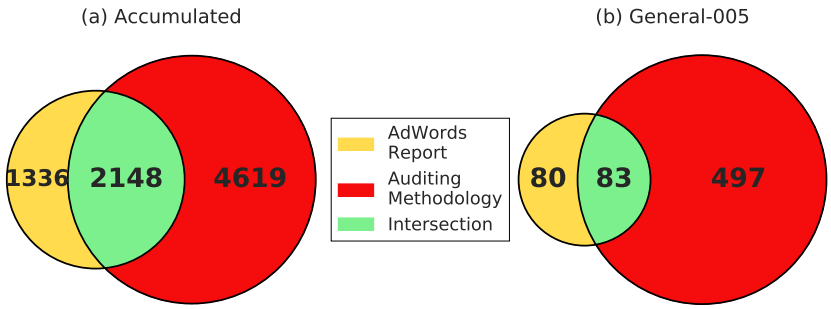

Figure 5.1: Venn diagram showing the number of publishers exclusively reported by our auditing methodology (red), exclusively reported by AdWords (yellow) and reported by both (green) for all our campaigns and campaign *General-005*.

5.1.3 Assessment of Quality Metrics

In this Subsection, we prove the validity of our methodology to first, perform a quality assessment for our 8 display ad campaigns and, second, audit the ad campaign reports from AdWords. To this end we study the different quality aspects presented in Section 2.2: *Brand Safety*, *Context*, *Publishers' popularity*, *Quality of Impressions* and *Fraud Indicators*. Our campaigns were configured based on CPM. The *conversion* analysis is out of the scope of this work, so we leave this for future work.

Note that the results presented in the rest of this section, except for the cases of *Brand Safety* and *Context*, are obtained from the analysis of the datasets resulting from our research without considering the information available in AdWords reports.

5.1.3.1 Brand Safety

To define an efficient *Brand Safety* strategy, an advertiser must know every publisher where ad impressions are displayed in its campaigns. For each one of the 8 ad campaigns, we have compared the list of publishers where ad impressions were displayed as reported by our methodology vs. reported by AdWords. Figure 5.1 shows a Venn diagram representing the total number of publishers exclusively reported by AdWords (in yellow), exclusively reported by our methodology (in red) and those reported by both (in green). In particular, the figure presents results for a specific campaign (*General-005*) as well as the aggregate results across all campaigns. The aggregate results reveal that AdWords did not report 57% of the publishers where ads from our campaigns were delivered[2]. This number can increase for individual campaigns up to 75%, as in the case of *General-005*.

Part of the impressions reported by AdWords are associated with *"anonymous.google"*. These entries correspond to impressions served through Google Ad Exchange to publishers or inventory partners that want to preserve their anonymity [116]. Our results show that it is invalid to

[2]Note that our methodology was not able to log 16.5% of the publishers.

Campaign ID	Auditing Methodology (% impressions)	AdWords Report (% impressions)
Research-010	2.50%	2.66 %
Research-020	3.75%	3.05 %
Football-010	64.12%	100 %
Football-030	46.66%	100 %
Russia	4.10%	7 %
USA	6.28%	10.73 %
General-005	4.96%	7.36 %
General-010	6.63%	56.65 %

Table 5.2: Fraction of impressions delivered to contextually meaningful publishers as reported by AdWords vs. our auditing methodology.

argue that publishers which Adwords did not report, correspond to those associated to *"anonymous.google"*. For instance, in *General-005*, AdWords registers only 425 impressions whose associated publisher is labelled as *"anonymous.google"*, however, 497 publishers identified by our methodology were not reported by AdWords. Then, even if these 425 impressions had been distributed across 425 publishers, still 72 (14.5%) publishers had not been reported by AdWords, in this specific campaign.

Therefore, *"anonymous.google"* is not the only source explaining this discrepancy. We have verified with a major Ad Tech company that this discrepancy is most likely explained by the fact that AdWords just report viewable impressions rather than all delivered impressions. Note, that this decision may have important implications for the brand safety protection of an advertiser as we argue next. An Ad Network may display an ad impression in a potentially harmful publisher for an advertiser. Whether the ad is seen or not is out of the control of the Ad Network and depends exclusively on the user's actions. If this ad is not seen by the user, then it is not reported to the advertiser. In this situation, there exists the risk that the algorithm of the Ad Network will deliver ads to that publisher again, and as a result the user may end up seeing the ad, thus leading to a brand safety violation episode. If advertisers would have access to the complete list of publishers where ads have been placed (regardless if the ad was reported to be seen or not), they could effectively identify potentially harmful sites and blacklist them. This would help prevent potential *Brand Safety* violation episodes in the future.

5.1.3.2 Context

AdWords support guidelines indicate that campaigns configured based on *audiences* would follow a *user-targeting* strategy. Instead, campaigns configured based on *keywords*, as it is the case with our campaigns, would follow a *contextual* strategy. This means that AdWords tries to display ads within publishers whose content is related to the targeted keyword(s), and thus contextually meaningful for the campaign. In addition, AdWords may use other factors to determine if a publisher is contextually relevant to the campaign such as the recent browsing history of a user

[117]. We have leveraged our auditing methodology to assess whether the context of a publisher is relevant to the keywords defined for a given campaign. In particular, we have obtained the keywords and topics that AdWords assigns to each publisher with at least 1 logged ad impression in our dataset. Then, we consider a publisher contextually meaningful if 1) any of its keywords match any of the campaign's keywords or 2) any of the publisher's topics are semantically similar to any of the keywords of the campaign. For this purpose we use the Leacock-Chodorow semantic similarity as described in [52].

Table 5.2 shows the fraction of impressions delivered to contextually meaningful publishers, as reported by AdWords vs. our auditing methodology, for our 8 campaigns. AdWords reports a notably higher fraction of ads delivered to contextually meaningful publishers compared to our methodology in most campaigns. This difference is likely due to the fact that Ad Words deliver contextual-driven impressions using other factors in addition to the publisher's theme.

5.1.3.3 Publishers' popularity

The popularity of a publisher indicates its capacity to attract users and thus, it is one of several factors affecting the perceived quality of a publisher. In general CPMs are higher with more popular publishers, which led to our assumption that campaigns configured with a higher CPM are expected to deliver ads to more popular publishers.

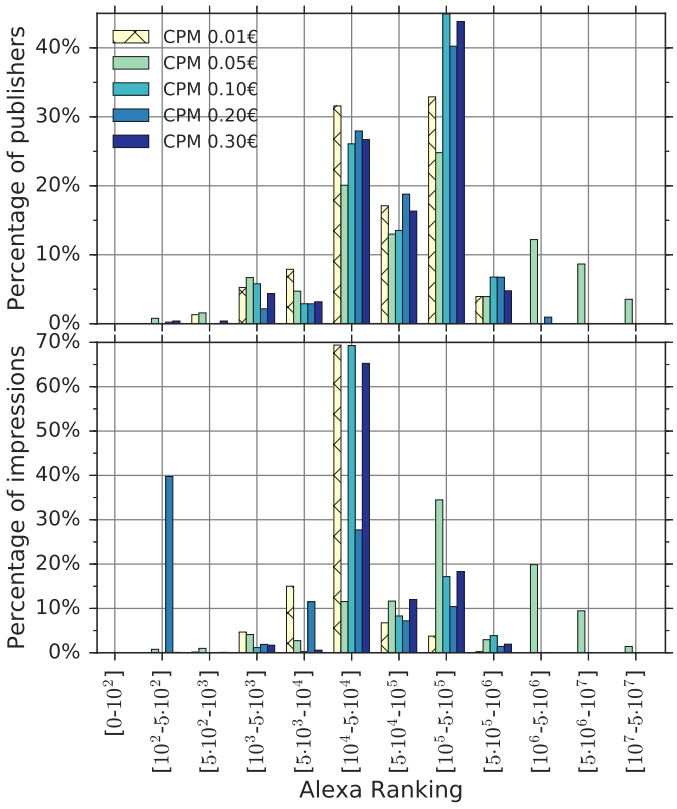

Figure 5.2: Distribution of publishers (top) and ad impressions (down) across the Alexa Ranking for 5 campaigns configured with different CPM investment.

Campaign ID	View $\geq 1s$
Research-010	56.18 %
Research-020	52.21 %
Football-010	79.89 %
Football-030	82.80 %
Russia	62.69 %
USA	71.13 %
General-005	75.13 %
General-010	55.03 %

Table 5.3: Fraction of impressions fulfilling the upper bound *viewability* criteria for each campaign.

Figure 5.2 shows the distribution of publishers and impressions across the Alexa ranking for 5 of our campaigns with CPMs ranging between 0,01€ and 0,30€. Specifically, we have defined logarithmic buckets and computed the fraction of publishers and impressions that fall in each bucket for each campaign. The results indicate that contrary to our expectation, higher CPMs do not lead to increase in impressions with popular publishers. The campaign with a CPM equal to 0,01€ seems to achieve higher than average performance with roughly 46% publishers and 89% impressions accumulated in the Alexa Top 50k sites. In comparison the campaign configured with a CPM of 0,30€ , representing a 30$\times$ investment increase, shows just 35% publishers and 68% impressions in the Alexa Top 50k. This is an unexpected observation, which may be an indication of potential inefficiencies in the market place under investigation.

5.1.3.4 Quality of Impressions

In this Section we evaluate the quality of impressions of our 8 campaigns using the two metrics described in Section 2.2, *viewability* and *frequency cap*.

1. ***Viewability*:** Table 5.3 presents the fraction of impressions that fulfills the upper bound of the *viewability* standard, and that we can measure with our methodology. The values range between 52% and 85% across campaigns. Interestingly, the two campaigns presenting the highest fraction of "viewable" impressions are the ones targeting "football", whereas other campaigns targeting other keywords (e.g., research) achieve a significantly lower viewability rate. We conjecture that the targeted context is an important factor that modulates ads *viewablity*.

2. ***Frequency Cap*:** Our goal in this case is to assess whether AdWords implements any default control in the *frequency cap*. Note that AdWords is used by a large number of customers without expertise in digital marketing, which may not configure a *frequency cap* in their campaigns. Therefore, it would be desirable that AdWords (or any other Ad Network) defines a default *frequency cap* on behalf of their customers. Research studies in the literature [53] have shown that a *frequency cap* over 10 does not lead to better

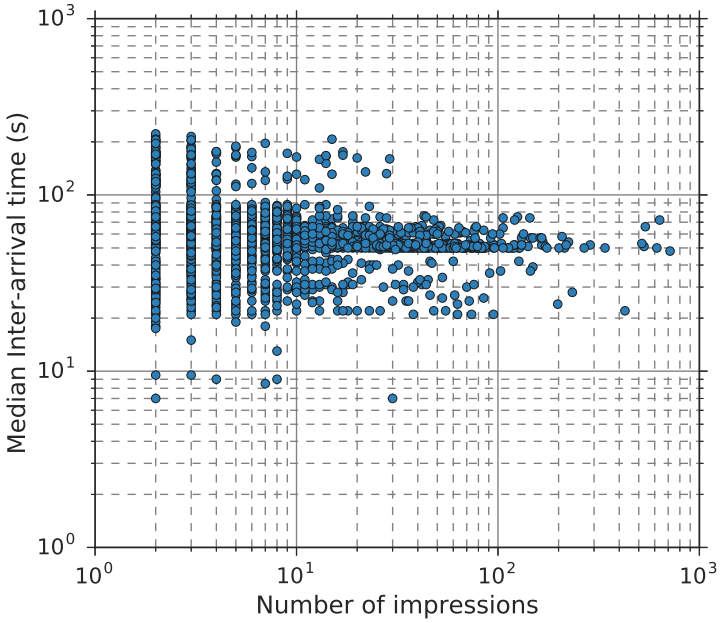

Figure 5.3: Number of ad impressions of a specific ad delivered to a user Vs. median inter-arrival time between impressions, considering all our campaigns.

conversion ratios. Based on this, 10 seems to be a reasonable reference value.

Figure 5.3 presents a scatter plot in loglog scale where the x-axis shows the number of impressions of a specific ad delivered to a user and the y-axis represents the median inter-arrival time between two consecutive impressions of that ad shown to the user. The figure presents aggregate results for all our campaigns. Note that we define a user as the combination of IP and User-Agent, so that two users behind a NAT using different User-Agents will be considered separately.

The results indicate that AdWords does not seem to use any default *frequency cap*. Indeed, 1720 (176) users receive more than 10 (100) impressions from the same ad. In addition, we observe that in many of these cases the inter-arrival time between impressions is rather small (below 1 min). In particular, there are extreme cases in which users receive hundreds of impressions with an inter-arrival time below 20 seconds. These observations suggest that unskilled or careless advertisers may experience inefficiencies in their campaigns performance due to the absence of a reasonable *frequency cap*.

5.1.3.5 Fraud Identification

Fraud is one of the primary threats to effectiveness in online advertising and causes direct losses of over $8B to advertisers in US [118]. Identifying, preventing and mitigating fraud is a complex and still unsolved problem which has only recently attracted the attention of the research community [33, 119, 120, 32, 34, 35]. In this subsection we show an example of how our auditing methodology can be used to identify one common ad fraud technique. The fraud technique in question consists of installing a bot on a server. This bot can be then sent to websites to view ads

Campaign ID	% of Cloud Providers IPs	% of Impressions delivered to Cloud IPs	% of Publishers showing ads to Cloud IPs
Research-010	3.39 %	4.42 %	8.62 %
Research-020	2.36 %	2.88 %	8.73%
Football-010	7.61 %	8.6 %	23.55%
Football-030	11.08 %	10.95 %	23.13%
Russia	0.52 %	0.27 %	2.58%
USA	1.03 %	0.68 %	5.56%
General-005	0.54 %	0.55 %	3.94%
General-010	0.42 %	0.58 %	2.59%

Table 5.4: Statistics on the volume of activity from Data Centers IPs for each campaign.

or perform other revenue generating actions. Associations responsible for defining the guidelines to fight fraud such as the Media Rating Council (US) and the JICWEBS (UK) both include Data Center traffic as a common source of invalid traffic (with some exceptions such as servers that are being used for providing VPN services) and recommends vendors to filter such traffic [36, 37].

Our methodology collects the IP addresses receiving ad impressions from a given campaign. Then, we identify which of the collected IPs belong to Data Centers (e.g., Cloud Providers or Hosting Providers). We use the following methodology for this purpose: First, we used MaxMind [82] to map each IP address in our dataset to its associated provider. Second, we identified the IPs from our dataset present in a list released by Botlab [121] including more than 130M IPs belonging to the top 100 Data Center providers worldwide. Finally, for the remaining IPs, we manually verified the website of its associated provider to assess whether it offered a Data Center service or not.

Table 5.4 presents the results of applying the previous methodology in each of our campaigns. Specifically, it shows: (i) the fraction of IPs located in Data Centers, (ii) the portion of ad impressions delivered to those IPs and, (iii) the fraction of publishers that served impressions to those IPs. We observe that using this methodology for detection, all our campaigns deliver ad impression to Data Center IPs. Specifically, "Football" campaigns present roughly 10% of the impressions delivered to Data Center IPs and 23% of publishers exposed to such impressions. For these particular campaigns we have verified that AdWords initially charged us for more than 1k impressions delivered to Data Center IPs. Later, we got a refund from AdWords. However, AdWords did not give details on the reasons for such refund and therefore we cannot assess if the previous impressions were part of it.

Finally, note that AdWords does not provide detailed information about the ad placement or publishers that are exposed to fraud, and thus an advertiser cannot currently assess its exposure to the analyzed type of fraud while running campaigns on Google AdWords.

5.1.4 Discussion

This work illustrates the lack of transparency and accurate information that advertisers are suffering from in the current online advertising ecosystem. This avoids advertisers from accurately assessing the efficiency and quality of their online campaigns. As a result they lack the required information to take decisions and actions to protect, for instance, their *Brand Safety*. These results should encourage advertisers to request the Ad Tech industry to standardize the use of independent measurements methodologies, as the one presented in this work. Doing so would allow advertisers to independently assess the quality of their online advertising campaigns as well as auditing the reporting practices of various vendors such as Ad Networks and DSPs.

5.2 *Q-Tag*: A Transparent Solution to Measure Ads Viewability Rate in Online Advertising Campaigns

Viewability is one of the most important metrics used in ad-tech to measure the performance quality of ad campaigns. The viewability standard defines the visibility conditions an ad impression must meet to achieve a sufficient marketing effect to be considered viewed. The ad-tech industry offers opaque measures of viewability whose performance is questionable. To address this issue, we propose a novel methodology for measuring viewability in ad campaigns. The disclosure of the functional details of this technique makes it reproducible and auditable. Our solution has been deployed in production by a Demand Side Platform (DSP) to measure the viewability rate of the ad campaigns. Leveraging the infrastructure of this DSP, we compare the performance of our methodology with a commercial solution. Both techniques report a similar overall viewability rate of 50%. However, our solution measured the viewability in 93% of the ads served by the DSP, unlike to 74% of the ads measured by the commercial solution. A rough estimation indicates that this increase in the measured rate may lead to a revenue increase of $3.5 million per year for a mid-sized DSP serving 100M of ads per day.

5.2.1 Measuring Viewability with Q-Tag

Our methodology is designed to measure the viewability metric for the most common types of ads, including display and video advertisements. These ads are typically embedded in an iframe (or a nested iframe). The vendor delivering the ad controls this iframe. In addition to the ad, vendors include in the iframe the so-called *ad tags* (a.k.a. *tracking pixels*). An *ad tag* is a piece of code (typically JavaScript) that allows the vendor, or other third parties, monitoring different aspects related to an ad impression shown to a user, such as: the URL where it was displayed, the type of device receiving the ad, if there was a click event, etc. The *ad tag* sends the collected information to a server for its subsequent analysis.

We have created our JavaScript *ad tag* to measure if an ad impression meets the viewability criteria defined by the standard. We refer to it as Q-Tag. The straightforward manner of measuring the viewability from an *ad tag* would be to retrieve the position of the iframe in the screen and based on that, compute which fraction of the iframe is in the *viewport*, i.e., the visible part of the screen. Unfortunately, this is not (in general) possible due to a widely extended security policy referred to as the *Same-Origin Policy* (SOP) [122]. This policy would avoid our *ad tag* to retrieve the position of the iframe in the screen, in most of the cases.

To address this limitation, we have used the ability of modern browsers to stop rendering an element out of the *viewport* determined by the refresh rate, e.g., when the content is located below the fold, in a non-active tab or in the background. The refresh rate in most devices is 60 (or more) fps [123]. When an element (i.e., a pixel) is in the viewport, browsers and apps use this refresh rate. However, when the element is not in the viewport, the refresh rate pass to be close to 0, thus optimizing the use of the CPU. Hence, monitoring the refresh rate of a pixel, we can infer if it is in

the viewport or not. In particular, we set up a threshold of 20 fps so that pixels refreshing at a rate equal or higher (lower) than this threshold are considered visible (not visible). We have chosen this conservative threshold to make our solution compatible in devices with overloaded CPUs that refresh at lower than 60fps rates. We have also tested our solution with thresholds of 30, 40, and 50 fps without noticing any major difference. To measure if an ad meets the viewability standard condition, we set up 25 monitoring pixels in the iframe embedding the ad and monitor the refreshing rate of each of them. The monitoring pixels are deployed in an *"X layout"* as shown in Figure 5.4.A: (i) ten in each diagonal (not including the central pixel), (ii) the central pixel, (iii) one pixel in each of the middle points of the four sides of the iframe ad-space (four in total). We compute the area associated with the visible monitoring pixels, and if this covers at least 50% of the area of the ad, a timer is started. If this visibility condition holds for 1 second, then we confirm that the viewability criteria has been met and the code sends an *in-view* message to the monitoring server indicating so. Contrary, if the visibility conditions change and less than 50% of the ad becomes visible before the timer reaches 1 second, an *out-of-view* event is triggered, which automatically stops the timer and restarts the process. Therefore, if the monitoring server does not receive the *in-view* message from our deployed Q-Tag, we conclude that the associated ad impression has not met the viewability criteria. Note that this explanation refers specifically to display ads. However, our tag can identify the type of ad (display, large display, or video) and measure the specific conditions defined by the standard for each type of ad.

5.2.2 Q-Tag Validation

To assess the correct functionality of our solution, first, we compute the theoretical error in measuring the visible area of an ad for the selected layout and compared it with alternative ones. Second, we replicate the tests that one of the most important accreditation agencies uses to certify viewability measurement solutions. Third, we run some additional tests to analyze, among other things, the ability of our solution to measure viewability in mobile in-app ads, and in the presence of adblockers.

5.2.2.1 Layout Validation

The viewability standard requires solutions that can accurately measure the viewable area of an ad and not just the viewability criteria. Based on that, the accuracy of our solution is directly associated to the selection of the number of monitoring pixels and their layout. In this subsection, we consider three different layouts: *"X layout"*, *"dice layout"*, and *"+ layout"*, whose specific deployment with 25 pixels is presented in Figures 5.4.A, 5.4.B and 5.4.C, respectively. Moreover, for each of these layouts, we consider deployments with a number of pixels ranging between 9 and 60. For each combination of layout and number of pixels, we compute the relative average error in the measurement of the viewable area of an ad for three scenarios: 1) *diagonal sliding*: the ad slides in the viewport diagonally; 2) *vertical sliding*: the ad enters in the screen from top to

bottom; 3) *horizontal sliding*: the ad slides in the viewport from left to right. Figure 5.4 shows the results. If we compare the layouts, we observe that the *dice layout* offers the worst performance. The *X layout* and + *layout* offer the same performance for the vertical and horizontal sliding, but the *X layout* is the best solution in the diagonal sliding case. If we analyze now the performance as a function of the number of pixels, we observe that the error decreases fast as we move from 9 to 21 pixels, and then the error reduction flattens. The activation of a large number of pixels requires a higher computational cost without offering significant reductions in the theoretical error. 25 pixels seem to be a good trade-off offering a low error with a minimal CPU overhead.

Finally, it is worth noting that in this subsection we analyze the error in the measurement of the viewable area of an ad, which is different from measuring the viewability standard criteria. As the results in the rest of this section show, our solution offers an extremely high accuracy measuring the viewability standard.

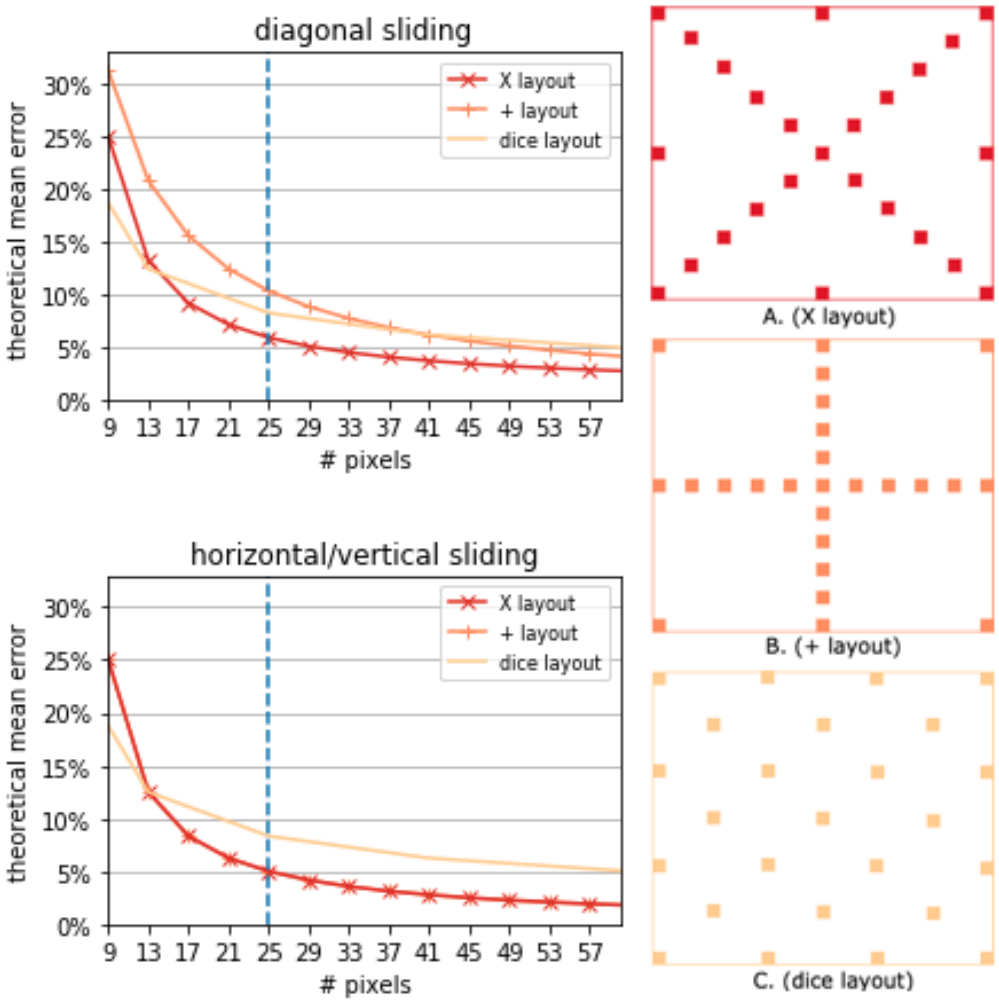

Figure 5.4: Comparison of possible layouts and the mean error given three scenarios for each layout.

5.2.2.2 Viewability Measurement Certification Tests

Mainly three entities define good practices in ad-tech: Media Rating Council (MRC) [39] operating in the US, JICWEBS [40] operating in the UK, and the Internet Advertising Bureau (IAB) [38] with international presence. Moreover, MRC and JICWEBS developed accreditation programs to certify the correct functionality of different solutions from ad-tech stakeholders. In particular, viewability measurement solutions are subject to certification by these entities, and the list of certified providers is publicly available [124, 125]. MRC does not disclose information about its accreditation process. JICWEBS relies on a third party (ABC) to develop the viewability certification process. ABC releases its Viewability Certification report every year [126], where they describe the tests conducted for the accreditation of viewability measurement solutions. These tests analyze whether a viewability measurement solution registers the *in-view* and *out-of-view* events properly in different scenarios. Table 5.5 describes each one of the tests as well as the expected result from them. ABC runs these tests for two types of ads (desktop banner and desktop video) and the following pairs of the browser-Operating System: Firefox-Windows, Chrome-Windows, IE11-Windows, Safari-macOS.

Note that these certification/accreditation processes are in practice accessible only for ad-tech stakeholders, and they are expensive. Therefore, it is not feasible to obtain an official certification for our solution. Instead, we replicate the ABC tests described in Table 5.5 in a lab environment and confirm with ABC that our tests are indeed similar to those used in their official accreditation process. In particular, we create a testing website and an ad creativity. We embed this ad inside two cross-domain iframes[3] included in our testing website.

Test	Description	Correct result
(1) Ad within cross-domain iframes	Ad served within multiple cross-domain iframes meeting the viewability standard criteria.	The ad is always in-view and thus the solution should register an *in-view* event once the viewability criteria is met
(2) Browser is resized	The browser page is enlarged so that the ad is always *in-view* thus meeting the viewability criteria.	
(3) Out of focus	The site with the ad becomes out of focus but it is always *in-view*.	
(4) Browser moved off-screen	The browser including an ad-space is moved off-screen after meeting the viewability criteria.	The solution should register an *in-view* event once the viewability criteria is met and when the ad-space moves out of view, it should register an *out-of-view* event.
(5) Page is scrolled	The browser page including an ad-space is scrolled after the ad impression meets the viewability criteria.	
(6) Browser is obscured	The user opens another app and the ad pass to background after it meets the viewability criteria.	
(7) Tab is obscured	The user switches to a new tab within the same browser after the ad impression meets the viewability criteria.	

Table 5.5: Description of the tests performed by Commercial Viewability Certification

[3]Note that a double cross-domain iframe is one of the most common scenarios faced by DSPs in the ad delivery process.

Finally, we deploy our *ad tag* for measuring viewability within the ad creativity. Note that, we run the 7 tests used in ABC accreditation, for the same two ad formats as ABC (desktop banner and desktop video). However, we consider 6 combinations of browser-OS (two more than ABC): Firefox (v67)-Windows10, Chrome(v75)-Windows10, IE(v11)- Windows10, Safari(v12)-macOS(v10.14), Firefox68-macOS(v10.14), and Chrome(v76)-macOS(v10.14). Hence, we consider 84 different scenarios (7 test types, 2 ad formats, 6 browser-OS combinations). Note that, these pairs browser-OS represent around 82% of the current browsers market share[127]. For each of these scenarios, we automate the test process and run 500 repetitions, using Selenium WebDriver[128], except for scenarios of *test type (6)*. For these scenarios, we manually run 10 repetitions. Overall, we perform more than 36k individual tests.

The results of this thorough validation are very satisfactory since 93.4% of the 36k individual tests produce a correct result. Note that the reported 6.6% wrong results occur in *tests type (4)* and *(5)*. In those specific instances of failed tests, we are not able to register any event (*in-view* and *out-of-view*). Since this only occurs in some instances but not always, and we could not identify any consistent pattern which could explain these failures, we hypothesize the failure might be associated with the automation process with Selenium WebDriver. To check our hypothesis, we manually perform several repetitions of these tests without using the automation process, in all of them, the *in-view* and *out-of-view* events are correctly registered. Hence, we conclude that errors are more likely due to the automation process rather than the viewability measurement solution.

In summary, these results are the first reliable indication of the correct functionality of our viewability measurement solution that, in the worst case, offers a 93% accuracy.

5.2.2.3 Other Tests

In this Subsection, we present some extra analyses, which extend the previous validation exercise.

- **In-view event accuracy:** We randomly place a double iframe including an ad creativity embedding Q-Tag in 10,000 positions on the testing website. Among them, there are all sorts of cases where the ad is wholly or partially visible on the screen as well as cases in which the ad is *out-of-view*. For each one of these cases, we know the exact position of the ad on the screen and, thus, whether the *in-view* event should be triggered or not by Q-Tag. The results show that our solution properly triggers the *in-view* event in the 10,000 analyzed cases.
- **Mobile in-app ads:** ABC does not evaluate in-app ads in its certification process. However, based on the information publicly released by MRC, it seems it analyzes this type of ad in its accreditation process. Hence, we set up a test to evaluate that our solution correctly measures viewability for mobile in-app ads. To this end, we use the Creative Preview App from Google [129], an application for previewing mobile in-app creatives. We use this app for testing the measurement accuracy of Q-Tag, in the case where the ad is displayed and

in-view in the mobile-app. We check two different creative sizes, and in both cases, Q-Tag notify the viewability measure correctly.

- **In-view event with adblockers and Brave:** Adblockers, as well as Brave [130], block the connection with third parties associated with ad-spaces in a webpage, and thus they block the ad delivery process. Since Q-Tag is only deployed if the ad is delivered, in the presence of adblockers, it should not be deployed. To confirm this, we install Adblock Plus [112] (the most popular ad blocker software) on Chrome in a lab environment and try to deliver three types of ad creativities (display, large display, or video) embedding Q-Tag to a testing website. We place ad-spaces in 50 random positions on the testing website for each ad type. In every test, all the connections are blocked as expected, and neither the ad nor Q-Tag is deployed. We reproduce the same test using Brave browser, and the ad and Q-Tag are not deployed, as expected.
- **Privacy-enhanced browsers:** We test our methodology in the latest Chrome, Safari, and Firefox versions (77, 13, and 69, respectively), which enable by default the prevention of cross-site tracking, i.e., blocking the third-party cookies. We reproduce the same test as in the case of AdBlock Plus and Brave. Q-Tag operates normally in these browsers since they block cookies while our methodology uses JavaScript code.

5.2.3 Deploying Q-Tag in Production

Q-Tag has been deployed and integrated within Sonata, a Digital DSP/DMP Platform engineered by TAPTAP Digital[131], a multinational company with presence in more than 10 markets within Europe, North America, South America, and Africa.Q-Tag has been instrumented to report the viewability measures to the distributed monitoring infrastructure of this DSP. Hence, our solution is ready to be activated in any ad campaign run by this DSP. In this work, we consider a dataset, including the viewability measures of more than 12M ads belonging to 99 ad campaigns that we monitor during a week. In addition to Q-Tag, the DSP allowed us to deploy the viewability measurement solution from one of the most important verifying companies in the ad-tech ecosystem[4] (also implemented as an *ad tag*). Note that the use of this verifying company has an associated cost. Due to budget limitations, we have run, both, the commercial solution and Q-Tag, in a subset of 4 ad campaigns including 1.89M ads.

Note that the ad campaigns considered in this work are a representative sample of the typical operation of a stakeholder, in this case, a DSP, in the ad-tech ecosystem: 1) each of the campaigns deliver ads through several Ad Exchanges including the most important ones (AppNexus, Axonix, DoubleClick, MoPub, OpenX, Rubicon, Smaato, Smart); 2) these campaigns belong to advertisers from different sectors (e.g., *Food & Drink*, *Personal Finance*, *Style & Fashion*, etc.) and countries (e.g., US, Mexico, Colombia, Spain, UK, Germany, etc.) and thus target different audiences and geographical regions; 3) we use different size of ads (*300x250* and *320x50*) across

[4]The name of the verifying company remain anonymous to meet the terms of an NDA with the DSP.

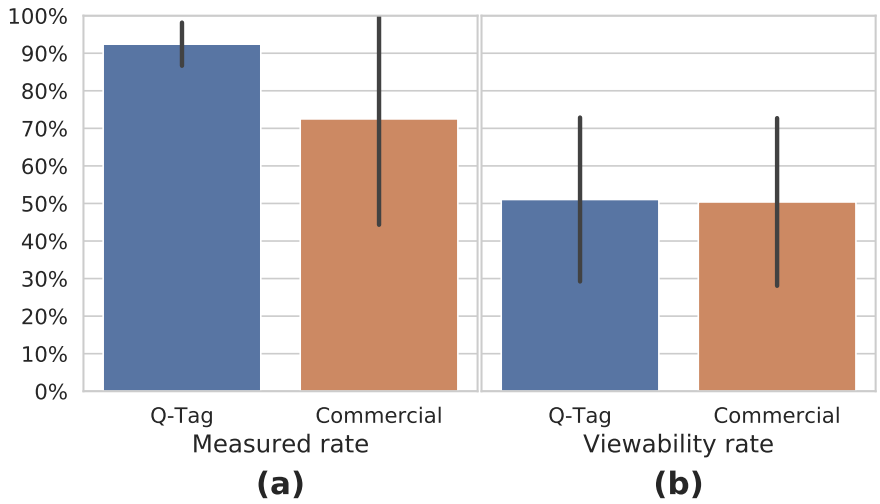

Figure 5.5: Comparison of the measured and viewable rate between our solution and the commercial one.

the ad campaigns. Based on this, we believe that the performance results of our viewability measurement solution are also representative.

5.2.4 Q-Tag vs. Commercial Solution

In this section, we compare the performance of Q-Tag and the mentioned commercial solution (one of the most widely used in the ad-tech ecosystem) using the data collected from real ad campaigns run by our DSP. In particular, we compare two performance metrics:

- **Measured rate**: This metric is defined as the fraction of ad impressions for which a solution can measure the viewability.
- **Viewability (or In-view) rate**: This metric is defined as the fraction of measured ad impressions that meet the viewability standard criteria.

Figure 5.5 shows the obtained results. In particular, Figure 5.5 (a) shows the measured rate for both solutions. The large bar shows the average, whereas the error bars show the standard deviation across the analyzed campaigns. Using this same representation, Figure 5.5 (b) shows the viewability rate results for both solutions.

First, we observe that both solutions offer similar average (roughly 50%) and standard deviation viewability rate. This fact indicates that our solution provides viewability rates in the same range as commercial solutions. This reinforces the conclusion regarding the high accuracy of our solution obtained through the exhaustive validation process presented in section 5.2.2.

Second, our solution offers significantly superior performance on the measured rate. Specifically, our solution can measure (on average) the viewability for 93% ads impressions, whereas the considered commercial solution can measure just 74%. An inspection of the data reveals that most of the measurement errors of the commercial solution come from impressions delivered to

Site type	OS	Q-Tag	Commercial Solution
App	Android	90.6%	53,4%
	iOS	97.0%	83.8%
Browser	Android	94.4%	86.7%
	iOS	94.6%	91.1%

Table 5.6: Q-Tag vs. commercial solution measured rate for site type and OS in mobile ad impressions

mobile devices. Table 5.6 shows a comparison of the measured rate obtained by Q-Tag vs. the commercial solution sliced by the OS (Android vs. iOS) and type of site (browser vs. app). While our solution offers in any case better measured rate than the commercial one, the most significant difference occurs in the viewability measurements for Android apps, where the commercial solution can measure just 53.4% of the impressions compared to 90.6% of Q-Tag.

5.2.4.1 Economic implications of a higher measured rate

Based on the obtained results, DSPs can obtain an important revenue increase using our solution instead of the referred commercial one. As we mentioned above, major vendors (Google, Facebook, etc.) have opted for a pricing model that only charges advertisers for *viewed* ad impressions. The rest of stakeholders are rapidly adopting this model, so that, it is expected that shortly it will be the de-facto viewability pricing model in the ecosystem. Under this pricing model, not measured ad impressions are not monetized. In this context, a DSP using Q-Tag instead of the considered commercial solution would be able to measure 19% more ads. Having a 50% viewability rate reported by both solutions, roughly half of these ads would be *viewed* so that a DSP opting for our solution would effectively monetize 9.5% more ads than using the referred commercial solution. If we consider a medium-size (large) DSP serving 100M (1B) ads per day at an average CPM of $1[5], this 9.5% extra measured *viewed* ads translate into $9.5k ($95k) revenue increase per day, i.e., roughly $3.5M ($35M) per year.

5.2.5 Related Work

The viewability standard was released in 2014 [41]. The wide adoption of this standard by the industry led to the development of proprietary solutions to measure viewability by verifying companies [7, 6, 8], whose performance and limitations are largely unknown. Despite the relevance of online advertising (a business generating a revenue of $107.5B in 2018 just in US [2]) and the importance of performance metrics, there is a lack of research literature addressing the viewability standard. This is probably due to the recent approval of the standard and its implementation by the ad tech industry. We could only find two theoretical studies orthogonal to our work. Chong Wang et al. have created a model to predict the viewability analyzing scroll depth for a given

[5]Note that a $1 average CPM is a realistic reference in the ad-tech ecosystem.

user and a page [132]. In a different work, David Bounie et al. [133] presented an analysis of the economic consequences of the investment in campaigns with low viewability rates.

From a measurement methodology perspective, we find previous works in the literature performing measurements from code embedded in ads. Some of these works use flash ads as a platform for measuring network properties and security aspects [4, 3, 15]. Note that most DSPs no longer support flash because it is deprecated in online advertising. More recent measurement works use JavaScript-based ad measurements for auditing the online advertising ecosystem [134], for measuring mobile devices network performance [9, 10], or for measuring DNS aspects [135, 136], among others. Note that none of these works present a methodology able to measure viewability as we do in this work.

5.2.6 Discussion

In this work, we have described, implemented, and evaluated Q-Tag, a new technique for measuring the viewability rate of online advertising campaigns, which offers a 93.4% measurement accuracy.

The release of functional details of our technique for measuring viewability makes it easily replicable by advertisers, agencies, or DSPs. In consequence, these stakeholders have *for first time* at their disposal an independent and auditable solution for assessing the viewability rate of their campaigns, without the need to rely upon opaque solutions offered nowadays by the industry.

Q-Tag has been deployed in production in a DSP. Using information from 12M measured ads served by this DSP, we compared the performance of our technique with one of the most important commercial solutions for viewability measurement. The comparison results show that Q-Tag can measure the viewability in 19% more ads than the commercial solution. A ballpark estimation reveals that these extra measured ads may lead to an annual revenue increase in the order of millions (tens of millions) of dollars for mid (large) size DSPs.

CHAPTER 6

Ethical and Legal Considerations

Ad-driven measurements such as AdTag and Q-Tag are likely to run *unbeknownst* to the crowd participant, placing particular responsibility on the measurement orchestrator. While previous work has demonstrated the community's sensitivity to this type of experiment [68], recent work continues to operate in similar fashion [137, 9]. In this Chapter, we review the ethical and legal aspects of this responsibility and position this thesis in this context.

We acknowledge and remind experimenters that ad-driven measurements bear the potential of harm to the client. Consider an experiment that collects client IP addresses together with HTTP request headers and their potential to profile individual users. While the ethical sensitivity of such experiments is evident, the experimenter also needs to be aware of potential legal constraints of the measurement, such as when an ad connects to websites deemed illegal in the user's country, or the local jurisdiction considers the collected information personally identifiable. The work on AdTag and Q-Tag has not and will not engage in practices that violate these concerns. We also followed the ethical guidelines defined by the community [138, 139].

In the context of ad-based measurements, informed consent [139] is difficult to obtain. The option of using ad-blocking software only offers blunt control over ad displays, and while the `Do Not Track` request header could serve as a possible signal to the experimenter, its applicability to arbitrary measurements remains unclear to both users and experimenters. Accordingly, we did not obtain informed consent from AdTag and Q-Tag's participants. For the purpose of this thesis, we did not collect any personal or sensitive information from the user, anonymizing collected data. The ads rendered in the campaign pointed to one ongoing research project [140], ensuring

that all connections were made to a safe and uncensored server under the authors' control. Finally, to the best of our knowledge, the tests also comply with the terms of use of the chosen DSPs.

Furthermore, the data used in the specific work of Q-Tag for measuring and comparing the viewability with a commercial solution has been collected by the referred DSP that has deployed *Q-Tag* in production. This DSP is compliant with the data protection legislation of those countries where it operates, including the recent EU data protection legislation (GDPR) [141]. Besides, the data we have received from the DSP does not include any personal information (PII) that can affect users' privacy. Finally, the deployment of our solution is compliant with the terms of service of all providers of the DSP.

CHAPTER 7

CONCLUSIONS

This thesis has presented a novel measurement methodology that introduces a custom JavaScript code inside the ad that can run a vast number of experiments at scale, with world-wide coverage and in a short period of time. The extensive capabilities of this methodology apply to different sectors. In particular, this thesis has used the methodology developed to two important research fields.

The first research field covered in this thesis is network measurements:

(i) A methodology has been designed to measure the Internet from the end-user perspective. We have named this methodology AdTag, and throughout this document, its design and potential for analyzing a wide range of aspects of Internet performance from the browser have been discussed and evaluated. Several experiments have proven the feasibility of AdTag capable of reaching millions of devices in a short period of time. Furthermore, it has also been demonstrated its ability to reach these devices based on their type (fixed vs. mobile) or their geographical location. With the use of standard JavaScript APIs, AdTag can be used to address numerous network performance and transparency issues such as: detect and characterize middleboxes, proxies, and NATs, analyze CDN performance, or furnish the input of IP address classification.

(ii) Besides, using AdTag, the global DNS infrastructure, and its use by normal users has been studied. Two small measurement campaigns have been carried out to demonstrate the potential of the proposed methodology and highlight its ability to gain new insights into the deployment strategies followed by ISPs around the world and user adoption decisions. The empirical results shown indicate that 13% of global DNS searches are resolved by commercial DNS providers such

as Google DNS, rather than the DNS provided by ISP resolvers. This study suggests that such adoption is not driven by performance gains, but is likely to be a mechanism for improving privacy and circumventing censorship and surveillance in oppressive countries. It has also been shown that ISPs that provide both mobile and fixed access tend to decouple the DNS infrastructure that serves each type of network access.

(iii) Another contribution of this thesis is the use of AdTag for measuring the browser market landscape. This study is of interest to companies of various kinds, researchers, and regulators. The existing solutions, based on passive measurement techniques, offer high scalability. However, they have two main drawbacks: Firstly, they require a large monitoring infrastructure, making them accessible to only a handful of companies. Second, their passive nature prevents them from offering targeting capabilities. The use of AdTag in this context addresses the previous limitations. As explained, it uses the online advertising infrastructure, which is now a commodity used by tens of thousands of companies, as a measurement platform. This democratizes the browser market's measurement since, contrary to traditional solutions, any interested company or researcher can use it. Besides, AdTag offers enough scalability to measure the browser market. AdTag also solves the limitation related to the lack of targeting capabilities of existing solutions. To this end, it leverages the targeting options of the online advertising infrastructure.

The second research area addressed in this thesis is transparency in online advertising:

(iv) We present Q-Tag, a specific implementation of our measurement methodology, which inserting JavaScript code inside the ads is able to measures different aspects of the quality metrics used to assess the efficiency of online advertising campaigns. We have run Q-Tag in Google AdWords (one of the most important advertising platforms at the time of doing the study) as a representative use case. The results indicate that advertisers are exposed to a lack of transparency and reported inaccurate information, which may have implications in terms of *brand safety* or exposure to fraud. The results should encourage advertisers to request the Ad Tech industry to standardize the use of independent measurement methodologies, like the one presented in this thesis.

(v) In subsequent work, the Q-Tag has been modified to measure the *viewabilty* standard metric of online advertising campaigns accurately, offering a measurement accuracy of 93.4%. Q-Tag has been deployed in production on a DSP. Using the information from the 12M measured ads served by this DSP, the performance of the proposed technique has been compared with one of the most important commercial solutions for viewability measurement. The comparison results showed that Q-Tag measured the viewability in 19% more ads than the commercial solution. A rough estimate reveals that these extra measured ads can lead to an increase in annual revenue in the order of millions (tens of millions) of dollars for medium (large) DSPs.

The proposed methodologies open up new avenues for investigating the infrastructure, robustness, and transparency of a wide range of aspects of the Internet. As future work, we plan to use this developed methodology for building two platforms. On the one hand, we will integrate AdTag into a network measurement platform and make this available to third parties. This platform

will incorporate the measurement use cases presented in this thesis but will add others, including: measuring CDN performance metrics, NAT-type deployment, and transparency analysis of middleboxes. On the other hand, we will integrate Q-Tag as part of an advertising campaign auditing tool that will be offered to online advertising stakeholders (Advertisers, Agencies, or DSPs) as an independent quality assessment platform.

References

[1] Oko Ad Management. The history of online advertising. (Last accessed on 16/06/2020). [Online]. Available: https://oko.uk/blog/the-history-of-online-advertising

[2] Internet Advertising Bureau. (IAB). Internet advertising revenue report. 2018 full year results. [Online]. Available: https://www.iab.com/wp-content/uploads/2019/05/Full-Year-2018-IAB-Internet-Advertising-Revenue-Report.pdf

[3] S. Zander, L. L. Andrew, G. Armitage, G. Huston, and G. Michaelson, "Investigating the ipv6 teredo tunnelling capability and performance of internet clients," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 5, pp. 13–20, 2012.

[4] S. Zander, L. L. Andrew, G. Armitage, G. Huston, and G. Michaelson, "Mitigating sampling error when measuring internet client ipv6 capabilities," in *Proceedings of the 2012 Internet Measurement Conference (IMC)*, 2012, pp. 87–100.

[5] G. Huston. Measuring Google's Public DNS. (Last accessed on 16/06/2020). [Online]. Available: https://labs.ripe.net/Members/gih/measuring-googles-public-dns

[6] Moat. Moat Analytics. (Last accessed on 16/06/2020). [Online]. Available: https://moat.com/analytics

[7] Integral Ad Science. (Last accessed on 16/06/2020). [Online]. Available: https://integralads.com

[8] DoubleVerify. (Last accessed on 16/06/2020). [Online]. Available: http://www.doubleverify.com

[9] M. D. Corner, B. N. Levine, O. Ismail, and A. Upreti, "Advertising-based measurement: A platform of 7 billion mobile devices," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 435–447.

[10] M. Corner and B. Levine, "Micromobile: Leveraging mobile advertising for large-scale experimentation," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2018, pp. 310–322.

[11] Internet Society. Global Internet Maps. (Last accessed on 16/06/2020). [Online]. Available: http://www.internetsociety.org/map/global-internet-report

[12] A. M. Kakhki, F. Li, D. Choffnes, E. Katz-Bassett, and A. Mislove, "Bingeon under the microscope: Understanding t-mobiles zero-rating implementation," in *Proceedings of the 2016 workshop on QoE-based Analysis and Management of Data Communication Networks*, 2016, pp. 43–48.

[13] A. Molavi Kakhki, A. Razaghpanah, A. Li, H. Koo, R. Golani, D. Choffnes, P. Gill, and A. Mislove, "Identifying traffic differentiation in mobile networks," in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 239–251.

[14] N. Weaver, C. Kreibich, and V. Paxson, "Redirecting DNS for Ads and Profit," in *The 2011 USENIX Workshop on Free and Open Communications on the Internet (FOCI)*, pp. 2–3.

[15] M. O'Neill, S. Ruoti, K. Seamons, and D. Zappala, "Tls proxies: Friend or foe?" in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 551–557.

[16] N. Weaver, C. Kreibich, M. Dam, and V. Paxson, "Here be web proxies," in *International Conference on Passive and Active Network Measurement*. Springer, 2014, pp. 183–192.

[17] N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, and V. Paxson, "Header enrichment or isp enrichment? emerging privacy threats in mobile networks," in *Proceedings of the 2015 ACM SIGCOMM Workshop on Hot Topics in Middleboxes and Network Function Virtualization*, 2015, pp. 25–30.

[18] 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 *Proceedings of the Internet Measurement Conference*. ACM, 2019, pp. 15–21.

[19] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, "Global measurement of DNS manipulation," in *26th USENIX Security Symposium (USENIX Security 17)*. Vancouver, BC: USENIX Association, aug 2017, pp. 307–323.

[20] B. Ager, W. Mühlbauer, G. Smaragdakis, and S. Uhlig, "Comparing dns resolvers in the wild," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*. ACM, 2010, pp. 15–21.

[21] N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, N. Weaver, and V. Paxson, "Beyond the radio: Illuminating the higher layers of mobile networks," in *Proceedings of the 13th*

Annual International Conference on Mobile Systems, Applications, and Services, 2015, pp. 375–387.

[22] 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 *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 523–536.

[23] M. Allman, "Comments on dns robustness," in *Proceedings of the Internet Measurement Conference 2018*. ACM, 2018, pp. 84–90.

[24] Mozilla. Security Advisories for Firefox. (Last accessed on 16/06/2020). [Online]. Available: https://www.mozilla.org/en-US/security/known-vulnerabilities/firefox/

[25] Chrome Releases. (Last accessed on 16/06/2020). [Online]. Available: https://chromereleases.googleblog.com/

[26] The Guardian. Google fined €1.49bn by EU for advertising violations. (Last accessed on 16/06/2020). [Online]. Available: https://www.theguardian.com/technology/2019/mar/20/google-fined-149bn-by-eu-for-advertising-violations

[27] J. Mayer and A. Narayanan. Deconstructing Google's excuses on tracking protection. (Last accessed on 16/06/2020). [Online]. Available: https://freedom-to-tinker.com/2019/08/23/deconstructing-googles-excuses-on-tracking-protection/

[28] W3Counter. (Last accessed on 16/06/2020). [Online]. Available: https://www.w3counter.com/globalstats.php

[29] GlobalStats. Statcounter. (Last accessed on 16/06/2020). [Online]. Available: http://gs.statcounter.com/browser-market-share

[30] S. Frei, T. Duebendorfer, G. Ollmann, M. May *et al.*, *Understanding the Web browser threat: Examination of vulnerable online Web browser populations and the insecurity iceberg*. ETH, Eidgenössische Technische Hochschule Zürich, Communication Systems Group, 2008.

[31] T. Duebendorfer and S. Frei, "Web browser security update effectiveness," in *International Workshop on Critical Information Infrastructures Security*. Springer, 2009, pp. 124–137.

[32] M. Marciel, R. Cuevas, A. Banchs, R. González, S. Traverso, M. Ahmed, and A. Azcorra, "Understanding the detection of view fraud in video content portals," in *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 357–368.

[33] L. Chen, Y. Zhou, and D. M. Chiu, "Analysis and detection of fake views in online video services," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 11, no. 2s, pp. 1–20, 2015.

[34] B. Stone-Gross, R. Stevens, A. Zarras, R. Kemmerer, C. Kruegel, and G. Vigna, "Understanding fraudulent activities in online ad exchanges," in *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, 2011, pp. 279–294.

[35] Q. Zhang, T. Ristenpart, S. Savage, and G. M. Voelker, "Got traffic? an evaluation of click traffic providers," in *Proceedings of the 2011 Joint WICOW/AIRWeb Workshop on Web Quality*, 2011, pp. 19–26.

[36] Ad.Product. Frequency Capping: Why Every Ad Server Should Implement It. (Last accessed on 16/06/2020). [Online]. Available: https://adzerk.com/blog/frequency-capping/

[37] Media Rating Council (MRC). Invalid Traffic Detection and Filtration Guidelines Addendum. (Last accessed on 16/06/2020). [Online]. Available: http://mediaratingcouncil.org/101515_IVT%20Addendum%20FINAL%20(Version%201.0).pdf

[38] Internet Advertising Bureau. (IAB). (Last accessed on 16/06/2020). [Online]. Available: https://www.iab.com

[39] Media Rating Council (MRC). (Last accessed on 16/06/2020). [Online]. Available: http://mediaratingcouncil.org/

[40] JICWEBS. (Last accessed on 16/06/2020). [Online]. Available: https://jicwebs.org/

[41] MRC and IAB. MRC Viewable Ad Impression Measurement Guidelines. (Last accessed on 16/06/2020). [Online]. Available: https://www.iab.com/wp-content/uploads/2015/06/MRC-Viewable-Ad-Impression-Measurement-Guideline.pdf

[42] JICWEBS. Viewability Product Principles. (Last accessed on 16/06/2020). [Online]. Available: https://jicwebs.org/wp-content/uploads/2018/07/JICWEBS_Viewability_Product_Principles_July_2018.pdf

[43] Wired. Facebook says it will not charge for an ad unless someone sees it. (Last accessed on 16/06/2020). [Online]. Available: https://www.wired.com/2015/02/facebook-an-ad-doesnt-count-unless-someone-sees-it

[44] Adage. Yahoo to charge some advertisers only for ads people actually see. (Last accessed on 16/06/2020). [Online]. Available: http://adage.com/article/digital/yahoo-charge-advertisers-ads-people/293857

[45] Adage. Google display network will charge only for ads that are viewed. (Last accessed on 16/06/2020). [Online]. Available: http://marketingland.com/google-display-network-will-charge-only-for-ads-that-are-viewed-144768

[46] Business Insider. Two of google's metrics have been suspended from a key accreditation service used to measure ads. (Last accessed on 16/06/2020). [Online]. Available: http://www.businessinsider.com/google-doubleclick-suspended-from-media-rating-council-accreditation-2016-10

[47] Business Insider. Facebook over-inflating its video view count is bad, but not as bad as it seems. (Last accessed on 16/06/2020). [Online]. Available: http://nordic.businessinsider.com/facebook-exaggerating-view-count-didnt-affect-advertiser-bills-2016-9

[48] S. Dhar. Mystery shopping inside the ad fraud verification bubble. (Last accessed on 16/06/2020). [Online]. Available: http://www.slideshare.net/ShailinDhar/mystery-shopping-inside-the-adverification-bubble

[49] IAB UK. Back to Basics Guide to Programmatic. (Last accessed on 16/06/2020). [Online]. Available: https://www.iabuk.com/standards-guidelines/back-basics-guide-programmatic

[50] Media Rating Council (MRC). MRC Supplement to IAB Guidelines for the Conduct of Ad Verification: Enhanced Content Level Context and Brand Safety. (Last accessed on 16/06/2020). [Online]. Available: http://mediaratingcouncil.org/MRC%20Ad%20Verification%20Supplement-%20Enhanced%20Content%20Level%20Context%20and%20Brand%20Safety%20(Final).pdf

[51] Online Behavioural Advertising. Advertising Standars Authority UK. (Last accessed on 16/06/2020). [Online]. Available: https://www.asa.org.uk/Consumers/What-we-cover/Online-behavioral-advertising.aspx

[52] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris, "I always feel like somebody's watching me: measuring online behavioural advertising," in *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies*, 2015, pp. 1–13.

[53] J. Chandler-Pepelnjak and Y.-B. Song. Optimal Frequency: The impact of frequency on conversion rates. Microsoft Advertising Institute. (Last accessed on 16/06/2020). [Online]. Available: https://studylib.net/doc/8870389/the-impact-of-frequency-on-conversion-rates

[54] Y. Yuan, F. Wang, J. Li, and R. Qin, "A survey on real time bidding advertising," in *Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics*. IEEE, 2014, pp. 418–423.

[55] WFA & The Advertising Fraud Council. Compendium of ad fraud knowledge for media investors. (Last accessed on 16/06/2020). [Online]. Available: https://ppcprotect.com/wp-content/uploads/2017/07/Ad_Fraud_Knowledge.pdf

[56] What Is An Untrustworthy Supply Chain Costing The U.S. Digital Advertising Industry? . IAB. (Last accessed on 16/06/2020). [Online]. Available: http://www.iab.com/insights/what-is-an-untrustworthy-supply-chain-costing-the-u-s-digital-advertising-industry/

[57] I. Fette and A. Melnikov, "The websocket protocol," Internet Requests for Comments, RFC Editor, RFC 6455, December 2011, http://www.rfc-editor.org/rfc/rfc6455.txt. [Online]. Available: http://www.rfc-editor.org/rfc/rfc6455.txt

[58] Node.js. (Last accessed on 16/06/2020). [Online]. Available: https://nodejs.org/

[59] Oracle. MySQL. (Last accessed on 16/06/2020). [Online]. Available: https://www.mysql.com/

[60] RIPE Atlas. (Last accessed on 16/06/2020). [Online]. Available: https://atlas.ripe.net

[61] CAIDA. Archipelago (ARK) Measurements Infrastructure. (Last accessed on 16/06/2020). [Online]. Available: http://www.caida.org/projects/ark/

[62] MONROE. Measuring Mobile Broadband Networks in Europe. (Last accessed on 16/06/2020). [Online]. Available: https://www.monroe-project.eu

[63] S. Sundaresan, S. Burnett, N. Feamster, and W. De Donato, "Bismark: A testbed for deploying measurements and applications in broadband access networks," in *2014 USENIX Annual Technical Conference (USENIX ATC 14)*, 2014, pp. 383–394.

[64] PlanetLab. (Last accessed on 16/06/2020). [Online]. Available: https://www.planet-lab.org

[65] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: Illuminating the edge network," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, 2010, pp. 246–259.

[66] M. A. Sánchez, J. S. Otto, Z. S. Bischof, D. R. Choffnes, F. E. Bustamante, B. Krishnamurthy, and W. Willinger, "Dasu: Pushing experiments to the internet's edge," in *Presented as part of the 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI 13)*, 2013, pp. 487–499.

[67] S. Rosen, H. Yao, A. Nikravesh, Y. Jia, D. Choffnes, and Z. M. Mao, "Mapping global mobile performance trends with mobilyzer and mobiperf," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, 2014, pp. 353–353.

[68] S. Burnett and N. Feamster, "Encore: Lightweight measurement of web censorship with cross-origin requests," in *Proceedings of the 2015 ACM conference on special interest group on data communication*, 2015, pp. 653–667.

[69] Ookla Speedtest. (Last accessed on 16/06/2020). [Online]. Available: http://www.ookla.com

[70] FCC. Measuring Broadband America. (Last accessed on 16/06/2020). [Online]. Available: https://www.fcc.gov/general/measuring-broadband-america

[71] Luminati. (Last accessed on 16/06/2020). [Online]. Available: https://luminati.io

[72] G. Tyson, S. Huang, F. Cuadrado, I. Castro, V. C. Perta, A. Sathiaseelan, and S. Uhlig, "Exploring http header manipulation in-the-wild," in *Proceedings of the 26th International Conference on World Wide Web*, 2017, pp. 451–458.

[73] T. Chung, D. Choffnes, and A. Mislove, "Tunneling for transparency: A large-scale analysis of end-to-end violations in the internet," in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 199–213.

[74] M. Ikram, N. Vallina-Rodriguez, S. Seneviratne, M. A. Kaafar, and V. Paxson, "An analysis of the privacy and security risks of android vpn permission-enabled apps," in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 349–364.

[75] A. Razaghpanah, A. Li, A. Filasto, R. Nithyanand, V. Ververis, W. Scott, and P. Gill, "Exploring the design space of longitudinal censorship measurement platforms," *arXiv preprint arXiv:1606.01979*, 2016.

[76] VPNscam.com. How to avoid VPN scams in 2017-2018. (Last accessed on 16/06/2020). [Online]. Available: http://vpnscam.com/how-to-avoid-vpn-scams-in-2017-2018/

[77] Google. Update your Flash ads. (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/adwords/answer/6249073

[78] G. Huston. APNIC Labs IPv6 Measurement System. (Last accessed on 16/06/2020). [Online]. Available: https://labs.apnic.net/?p=348

[79] Knowledge Bridge. Online advertising explained. (Last accessed on 16/06/2020). [Online]. Available: http://www.kbridge.org/en/online-advertising-explained-dmps-ssps-dsps-and-rtb/

[80] Internet Advertising Bureau (IAB). HTML5 for Digital Advertising v2.0. (Last accessed on 16/06/2020). [Online]. Available: https://www.iab.com/guidelines/html5-for-digital-advertising-guidance-for-ad-designers-creative-technologists

[81] Google. Build an HTML5 creative. (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/richmedia/answer/2672542?hl=en

[82] MaxMind. GeoIP Database. (Last accessed on 16/06/2020). [Online]. Available: https://www.maxmind.com

[83] I. Poese, S. Uhlig, M. A. Kaafar, B. Donnet, and B. Gueye, "IP geolocation databases: Unreliable?" vol. 41, no. 2. ACM New York, NY, USA, 2011, pp. 53–56.

[84] R. Nithyanand, S. Khattak, M. Javed, N. Vallina-Rodriguez, M. Falahrastegar, J. E. Powles, E. De Cristofaro, H. Haddadi, and S. J. Murdoch, "Adblocking and counter blocking: A slice of the arms race," in *6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 16)*, 2016.

[85] M. Malloy, M. McNamara, A. Cahn, and P. Barford, "Ad blockers: Global prevalence and impact," in *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 119–125.

[86] S. Hallvord, A. Van Kesteren, J. Aubourg, and J. Song, "XMLHttpRequest level 1," W3C Note, 2016. [Online]. Available: https://www.w3.org/TR/2016/NOTE-XMLHttpRequest-20161006/

[87] I. Hickson, "The WebSocket API," W3C Candidate Recommendation, 2012. [Online]. Available: http://www.w3.org/TR/2012/CR-websockets-20120920/

[88] W3C Community Group Draft. Network Information API. (Last accessed on 16/06/2020). [Online]. Available: http://wicg.github.io/netinfo/

[89] B. Aboba, C. Jennings, A. Narayanan, T. Brandstetter, D. Burnett, and A. Bergkvist, "Webrtc 1.0: Real-time communication between browsers," W3C Candidate Recommendation, 2019. [Online]. Available: https://www.w3.org/TR/webrtc

[90] A. Müller, F. Wohlfart, and G. Carle, "Analysis and Topology-based Traversal of Cascaded Large Scale NATs," in *Proceedings of the 2013 workshop on Hot topics in middleboxes and network function virtualization*, 2013, pp. 43–48.

[91] 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 *Proceedings of the 2016 Internet Measurement Conference*, 2016, pp. 215–229.

[92] Q. Scheitle, O. Gasser, P. Sattler, and G. Carle, "HLOC: Hints-Based Geolocation Leveraging Multiple Measurement Frameworks," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*. IEEE, 2017, pp. 1–9.

[93] ZMap. The ZMap Project. (Last accessed on 16/06/2020). [Online]. Available: https://zmap.io/

[94] K. Schomp, T. Callahan, M. Rabinovich, and M. Allman, "On measuring the client-side dns infrastructure," in *Proceedings of the 2013 conference on Internet measurement conference*. ACM, 2013, pp. 77–90.

[95] M. Müller, G. Moura, R. de O Schmidt, and J. Heidemann, "Recursives in the wild: engineering authoritative dns servers," in *Proceedings of the 2017 Internet Measurement Conference*. ACM, 2017, pp. 489–495.

[96] M. Almeida, A. Finamore, D. Perino, N. Vallina-Rodriguez, and M. Varvello, "Dissecting dns stakeholders in mobile networks," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*. ACM, 2017, pp. 28–34.

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

[98] P. Foremski, O. Gasser, and G. C. M. Moura, "Dns observatory: The big picture of the dns," in *Proceedings of the Internet Measurement Conference*. ACM, 2019, pp. 87–100.

[99] N. Weaver, C. Kreibich, B. Nechaev, and V. Paxson, "Implications of Netalyzr's DNS measurements," in *Proceedings of the First Workshop on Securing and Trusting Internet Names (SATIN), Teddington, United Kingdom*. Citeseer, 2011.

[100] OONI. Open Observatory of Network Interference. (Last accessed on 16/06/2020). [Online]. Available: https://ooni.torproject.org/

[101] M. Dhawan, J. Samuel, R. Teixeira, C. Kreibich, M. Allman, N. Weaver, and V. Paxson, "Fathom: A browser-based network measurement platform," in *Proceedings of the 2012 Internet Measurement Conference*. ACM, 2012, pp. 73–86.

[102] 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 *USENIX Annual Technical Conference, General Track*, 2002, pp. 229–242.

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

[104] R. Jain, D.-M. Chiu, and W. R. Hawe, *A quantitative measure of fairness and discrimination for resource allocation in shared computer system*. Eastern Research Laboratory, Digital Equipment Corporation Hudson, MA, 1984, vol. 38.

[105] Google. Google Public DNS servers locaiton. (Last accessed on 16/06/2020). [Online]. Available: https://developers.google.com/speed/public-dns/faq

[106] OpenDNS. OpenDNS DNS servers locaiton. (Last accessed on 16/06/2020). [Online]. Available: https://www.opendns.com/data-center-locations/

[107] Cloudflare. Cloudflare DNS servers locaiton. (Last accessed on 16/06/2020). [Online]. Available: https://www.cloudflare.com/ips/

[108] TurboBytes. Google DNS, OpenDNS and CDN performance. (Last accessed on 16/06/2020). [Online]. Available: https://www.cdnplanet.com/blog/google-dns-opendns-and-cdn-performance/

[109] 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 *Proceedings of the Internet Measurement Conference 2018*. ACM, 2018, pp. 252–264.

[110] S. Dickinson. DNS Privacy - The Problem. (Last accessed on 16/06/2020). [Online]. Available: https://dnsprivacy.org/wiki/display/DP/DNS+Privacy+-+The+Problem

[111] NetMarketShare. (Last accessed on 16/06/2020). [Online]. Available: https://netmarketshare.com

[112] AdBLock. (Last accessed on 16/06/2020). [Online]. Available: https://adblockplus.org/

[113] NoScript. (Last accessed on 16/06/2020). [Online]. Available: https://noscript.net/

[114] Same-Origin Policy. (Last accessed on 16/06/2020). [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy

[115] Google Support. About the Google Display Network. (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/adwords/answer/2404190?hl=en

[116] Google Support. What is anonymous.google in Display Campaigns? (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/google-ads/thread/1453540?hl=en

[117] Google Support. About contextual targeting. (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/adwords/answer/2404186

[118] Global entertainment and media outlook 2015-2019. PwC, Ovum. (Last accessed on 16/06/2020). [Online]. Available: http://www.pwc.com/gx/en/global-entertainment-media-outlook/assets/2015/internet-advertising-key-insights-1-advertising-segment.pdf

[119] V. Dave, S. Guha, and Y. Zhang, "Measuring and Fingerprinting Click-spam in Ad Networks," in *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, 2012, pp. 175–186.

[120] V. Dave, S. Guha, and Y. Zhang, "ViceROI: Catching Click-spam in Search Ad Networks," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, 2013, pp. 765–776.

[121] Botlab.io Deny-hosting IP List. (Last accessed on 16/06/2020). [Online]. Available: https://github.com/botlabio/deny-hosting-IP

[122] Mozilla. Same-originpolicy. (Last accessed on 16/06/2020). [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/Security/Same-origin_policy

[123] Google. Rendering Performance. (Last accessed on 16/06/2020). [Online]. Available: https://developers.google.com/web/fundamentals/performance/rendering/

[124] JICWEBS Signatories. (Last accessed on 16/06/2020). [Online]. Available: https://jicwebs.org/certification-process/signatories/

[125] MRC Accredited Services and Services Under Review. (Last accessed on 16/06/2020). [Online]. Available: http://www.mediaratingcouncil.org/Accredited%20Services.htm

[126] ABC with JICWEBS. Viewability certification. (Last accessed on 16/06/2020). [Online]. Available: https://www.abc.org.uk/images/Viewability_Report.pdf

[127] W3Counter. Browser & Platform Market Share. (Last accessed on 16/06/2020). [Online]. Available: https://www.w3counter.com/globalstats.php

[128] Selenium WebDriver. Browser Automation. (Last accessed on 16/06/2020). [Online]. Available: https://www.seleniumhq.org/projects/webdriver/

[129] Google Support. Preview ads on a phone or tablet. creative preview app. (Last accessed on 16/06/2020). [Online]. Available: https://support.google.com/richmedia/answer/2879163

[130] Brave. Secure, Fast & Private Browser. (Last accessed on 16/06/2020). [Online]. Available: https://brave.com/

[131] TAPTAP Digital. Location Intelligence for Marketing. (Last accessed on 16/06/2020). [Online]. Available: http://www.taptapdigital.com

[132] C. Wang, A. Kalra, C. Borcea, and Y. Chen, "Viewability prediction for online display ads," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 413–422.

[133] D. Bounie, V. Morrisson, and M. Quinn, "Do You See What I See? Ad Viewability and the Economics of Online Advertising," *Ad Viewability and the Economics of Online Advertising (March 1, 2017)*, 2017.

[134] S. Guha, B. Cheng, and P. Francis, "Challenges in measuring online advertising systems," in *Proceedings of the 10th ACM on Internet Measurement Conference*, 2010, pp. 81–87.

[135] W. Lian, E. Rescorla, H. Shacham, and S. Savage, "Measuring the practical impact of dnssec deployment," in *Presented as part of the 22nd USENIX Security Symposium (USENIX Security 13)*. Washington, D.C.: USENIX, 2013, pp. 573–588.

[136] G. Huston. Measuring the DNS from the Users' perspective. (Last accessed on 16/06/2020). [Online]. Available: https://www.potaroo.net/presentations/2014-05-14-dns-measurements.pdf

[137] P. Vines, F. Roesner, and T. Kohno, "Exploring ADINT: Using Ad Targeting for Surveillance on a Budget — or — How Alice Can Buy Ads to Track Bob," in *Proceedings of the 2017 on Workshop on Privacy in the Electronic Society*, 2017, pp. 153–164.

[138] C. Partridge and M. Allman, "Ethical considerations in network measurement papers," *Communications of the ACM*, 2016.

[139] D. Dittrich, E. Kenneally *et al.*, "The Menlo Report: Ethical principles guiding information and communication technology research," *US Department of Homeland Security*, 2012.

[140] J. González Cabañas, A. Cuevas, and R. Cuevas, "Fdvt: Data valuation tool for facebook users," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 2017, pp. 3799–3809.

[141] EUGDPR. EU GDPR News and Updates. (Last accessed on 16/06/2020). [Online]. Available: https://eugdpr.com/