4 Years of EU Cookie Law: Results and Lessons Learned

Martino Trevisan*, Stefano Traverso, Eleonora Bassi, and Marco Mellia

4 Years of EU Cookie Law: Results and Lessons Learned

Abstract: Personalized advertisement has changed the web. It lets websites monetize the content they offer. The downside is the continuous collection of personal information with significant threats to personal privacy. In 2002, the European Union (EU) introduced a first set of regulations on the use of online tracking technologies. It aimed, among other things, to make online tracking mechanisms explicit to increase privacy awareness among users.

Amended in 2009, the EU Directive mandates websites to ask for informed consent *before* using any kind of profiling technology, e.g., cookies. Since 2013, the ePrivacy Directive became mandatory, and each EU Member State transposed it in national legislation. Since then, most of European websites embed a "Cookie Bar", the most visible effect of the regulation.

In this paper, we run a large-scale measurement campaign to check the current implementation status of the EU cookie directive. For this, we use CookieCheck, a simple tool to automatically verify legislation violations. Results depict a shady picture: 49 % of websites do not respect the Directive and install profiling cookies *before* any user's consent is given.

Beside presenting a detailed picture, this paper casts lights on the difficulty of legislator attempts to regulate the troubled marriage between ad-supported web services and their users. In this picture, online privacy seems to be continuously at stake, and it is hard to reach transparency.

Keywords: Privacy, Web Tracking, Cookie Law, ePrivacy Directive.

DOI 10.2478/popets-2019-0023

Received 2018-08-31; revised 2018-12-15; accepted 2018-12-16.

***Corresponding Author: Martino Trevisan:** Politecnico di Torino, E-mail: martino.trevisan@polito.it

Stefano Traverso: Politecnico di Torino, Ermes Cyber Security SRL, E-mail: s.traverso@ermessecurity.com

Eleonora Bassi: Politecnico di Torino, E-mail: eleonora.bassi@polito.it

Marco Mellia: Politecnico di Torino, Ermes Cyber Security SRL, E-mail: marco.mellia@polito.it

1 Introduction

When browsing the web, users encounter the so-called "online trackers". They build their business on the massive collection and brokerage of personal data. Used by websites to monetize the content they offer via personalized advertisement, online tracking is perceived as a threat to users' privacy [43, 59]. To regulate the usage of tracking technologies in the web, the EU Commission issued already in 2002 the ePrivacy Directive. It includes one of the first and strictest regulations on the usage of online tracking mechanisms [21]. As amended in 2009 [22], it requires websites to ask "prior informed consent for storage or for access to information stored on a user's terminal equipment". In other words, a website must ask the visitor to authorize the storage and retrieval of data sent through cookies and similar tracking mechanisms *before* delivering and installing them. Since 2013, the so-called "Cookie Law" provided by the ePrivacy Directive has been adopted by EU Member States (EU MSs). Since then, its implementation has become evident to end-users because of the presence of a "Cookie Bar" on most of websites. This bar informs users about the presence of tracking mechanisms and asks for their consent to the usage of them. The ePrivacy Directive has impact also outside European boundaries, as any web service having users in EU must be compliant.

Despite the Directive has been in force for more than four years, only a few studies aimed to understand whether web services actually respect it. All these studies build on experiments conducted on a very small scale, without properly quantifying the extent to which the ePrivacy Directive is actually implemented. This is also explained by the fact that, to the best of our knowledge, there exists no scalable tool to automatically check if a website violates the Directive with regard to the use of cookies. In this paper, we design and use CookieCheck [3], a simple tool to automatically perform this check. Given a website, CookieCheck visits it as a "new user", and analyses installed cookies. It then checks the presence of cookies violating the ePrivacy Directive, i.e., the so-called *profiling* cookies, which, as defined in [12],

are used by services to identify users, and could be installed only after user consent is obtained.

We use CookieCheck to conduct a large measurement campaign in April 2017. Results are shady: 49% of popular websites violate the ePrivacy Directive, with some categories (e.g., "News and Media") where violations top to 86%. This is an optimistic estimate, as the definition of profiling cookie we use throughout this paper is stringent and conservative.

The ePrivacy Directive has been already criticized as a case of regulatory failure: it impairs user browsing experience, becoming ineffective in increasing the users' awareness and in their privacy [26, 45, 51, 57]. Here, we show that the Directive is a failure from the enforcement perspective too. This testifies the struggle of legislators to regulate online privacy in a complex ecosystem dominated by advertisers where web services need to monetize the content they offer. Indeed, the new General Data Protection Regulation (GDPR), entered in to force in May 2018 with a significant media coverage, is only marginally impacting website habits that still leverage profiling cookies without user consent. As of November 2018, no substantial change emerges, possibly because it is to early to measure the outcomes of the new GDPR.

The contributions of this paper are:
(1) We present results obtained by running large scale experiments on more than 35,000 websites and show that almost half of them violates the Directive with regard to the use of cookies (even in its less strict interpretation). A considerable fraction of violations are caused by online trackers managed by big players of the web and aimed to support advertising services.
(2) We conduct further experiments to investigate the behavior of websites under different scenarios, e.g., when consent to the usage of cookies is given, or changing browser settings and client location. The only major difference is an increase of tracking after consent is given.
(3) We deeply discuss the weak points of the ePrivacy Directive and the reasons behind its failure. We identify the critical points the new ePrivacy Regulation proposals [24, 32, 37] should discuss, and, in the perspective, the difficulties of regulating the troubled marriage between ad-supported services and users' privacy.

The whole dataset used throughout the paper is available for download [2], and we offer CookieCheck as free and open source [3]. We provide a running demonstrator of CookieCheck available at [3] for everyone to use and share its code at [4].
We truly believe the figures we present are useful to policy-makers, web industry players and researchers in the debate on how to monitor and enforce privacy policies on the web.

The remainder of the paper is organized as follows. Section 2 summarizes the EU web privacy regulatory framework. In Section 3 and Section 4, we describe the measurement methodology and campaigns, respectively. Section 5 presents the results. In Section 6, we summarize the related work. Section 7 provides a thorough description of the ePrivacy Directive, with particular attention to the most controversial points, and, Section 8 discusses the ePrivacy Directive issues in light of the new Proposals for the ePrivacy Regulation currently under revision. Finally, Section 9 concludes the paper.

2 Overview of The European Regulatory Framework

In this section, we give the reader a summary of the current European legislation that governs the usage of tracking technologies. We provide a thorough description of the regulatory framework on ePrivacy in Section 7, in which we also analyze the most controversial points. Readers who are not experts may read Section 7 in advance.

We focus on the European Union, as it provides a strict regulatory framework with a broad territorial application. The European regulation is contained in the 2002 ePrivacy Directive [21], updated in 2009 by the Dir. 2009/136/EC [22]. In few words, it mandates that websites obtain users' informed consent before using any kind of tracking technology. More precisely, the directive mandates that a website can store or gain access to information on user's device only if the user is provided with clear and comprehensive information, and is offered the right to refuse such processing. Consent *must* be acquired before proceeding. Being HTTP cookies the most popular means to implement web tracking [10, 31, 49], it follows that they cannot be installed on user's device without prior consent. Exceptions hold, for instance for the so called technical cookies which are strictly required for the service, such as session cookies used to implement e-commerce websites. In this work, we focus on *profiling cookies* – i.e., cookies installed by web trackers to identify users, that are clearly subject to consent according to the ePrivacy Directive [21]. Section 3 provides a definition for *profiling cookie*, that we motivate later in Section 5.3, based on experimental results. A detailed description of which cookies must be

subject to consent according to the current regulation is provided in Section 7.2.

In short: *The ePrivacy Directive specifies that the usage of profiling cookies must be authorized by the user using an opt-in mechanism, and the website must provide the user specific information about the privacy policy. As such, compliant websites typically implement an "Accept Cookie Bar" which offers the user the possibility of consenting cookies the first time the website is accessed.*

3 Measurement Methodology

We aim to detect whether a website violates the requirements on the use of cookies contained in the ePrivacy Directive. We assume a violation occurs if the website installs, without prior user consent, any profiling cookie. For this, in our experiments we do not provide any form of consent to cookies.

We build on automatic browsing of web pages, and use two different tools: CookieCheck [3], a custom tool we engineer to provide scalability through experiment parallelization, and WebPageTest [8] which enables higher configurability.

CookieCheck is a scalable system based on Docker technology and Google Chrome. It is lighter than OpenWPM [30] which embeds many tests that are useless to our purposes (tracking of Javascript calls, response body content, etc.). Instead, CookieCheck is fast enough to visit websites in a short amount of time, without sacrificing accuracy. Its code can be downloaded at [4], and a working demonstrator of the tool is accessible online [3]. CookieCheck instruments Google Chrome to visit a web page, and collects cookie-related statistics using the Chrome DevTools Protocol [1]. Thanks to Docker technology, CookieCheck provides a reliable and lightweight means to isolate multiple browser instances running in parallel.

WebPageTest from Google [8] is designed for flexibility. It offers more configuration options than CookieCheck, but it is significantly slower and does not allow parallel testing, thus considerably increasing the collection time. More precisely, WebPageTest allows us to select different language and locale settings and to emulate different browsers used on different devices (e.g., Safari on Apple iPhone or Firefox on Android). It does this by changing the `User-Agent` HTTP header and screen resolution. It also allows to define custom browser profiles to emulate past browsing activity by the user.

Both tools take as input a set of web pages, and visit each. When the page is fully loaded (i.e., the `OnLoad` event is triggered) or a 90s timeout fires, they dump to file the HTTP Archive (HAR) [55], a JSON-formatted structure that summarizes navigation data and reports statistics about services and downloaded objects. For each HTTP transaction, the HAR logs the headers of the corresponding request/response. If not otherwise specified, we take care of erasing the browser cache and cookies stored before each visit. No user action is performed on the page. Hence, we emulate the behavior of a new user accessing the web page for the first time, and not having provided any prior consent to the installation of cookies. We do not scroll the page, nor click on any link. As a consequence, we do not provide any form of implied consent, allowed in some interpretations of the ePrivacy Directive (see Section 7.2.2).

We next analyze the cookies installed at the end of each visit:

1. We look at all HTTP responses with `Set-Cookie` header in the HAR file.
2. We match the domain of the service installing the cookie against the domain of the visited website to identify third-party cookies.
3. We check if *profiling* cookies are installed. These are cookies exclusively by third-party domains categorized as *advertising trackers* by both Ghostery [41] and Disconnect [27], two popular tracker-blockers. We further tighten the classification by considering only those cookies having a lifetime greater than 1 month. The robustness of our results to the employed tracker list will be discussed in Section 5.
4. At the end of this process, we tag a web page as violating the ePrivacy Directive if at least one profiling cookie was installed.

The result is a conservative choice, since at step three we detect a violation only for those cookies which are third-party, have long lifetime, and are classified as privacy offending by popular anti-tracker blacklist.

4 Measurement campaigns

We run two measurement campaigns, both conducted in April 2017 at the premises of our University Campus in Italy. The first campaign aims to check the presence of profiling cookies at scale. The second investigates how profiling cookies differ depending on device, browser settings, location of client, and when prior consent has been

given. At last, we analyze measurements collected periodically during the last 4 years. Table 1 summarizes all datasets.

To perform these measurements, we profit from a Linux machine equipped with an Intel® Xeon CPU with 12 cores and 32GB RAM and connected to the Internet through a 1Gb/s network. For the analysis we use a Hadoop cluster running Apache Spark and Python, whose `Cookie` module allows us to parse cookies.

4.1 Large scale measurement campaign

We rely on SimilarWeb [7], a website ranking service analogue to Alexa to obtain per-country and per-category website ranks. We pick the most popular websites in 25 countries (21 EU MSs, plus 4 non-EU countries).[1] We consider sites from 25 different categories. The list of countries and categories can be deduced from Figure 5. For each country and category, we pick the 100 most popular websites. In total, we consider $25 \times 25 \times 100$ entries, corresponding to $35,862$ unique websites to visit (as the lists overlap). In total we performed $179,310$ visits (5 visits per website) using 24 CookieCheck instances over a period of 15 days. We collected more than 195 GB of HAR files. Less than 5% of visits failed for timeout intervention, likely due to websites being temporarily offline or slow loading speed. In the remainder of the paper we refer to this dataset with *Large-Dataset*.

4.2 Specific measurement campaigns

We employ WebPageTest to run experiments on the 100 most popular websites for three main European countries: France, Germany and Italy. In total we count 241 unique websites. We visit each of them, changing the configuration of the tool, mimicking different browsers and devices, and analyzing historical data.

4.2.1 Impact of giving consent to cookies

We want to evaluate the impact of providing consent to the installation of cookies. To this end, we first manually visit the homepage of each website in the list to give consent to the usage of cookies (whenever the Cookie Bar is offered). We save the resulting browser profile, that we then use later on to visit websites again. This lets us verify whether websites actually install profiling cookies only upon user consent has been provided. We refer to this dataset as *AcceptCookie-Dataset*.

4.2.2 Impact of different browsers

We want to investigate if the use of different browsers may affect the number of installed cookies. For instance, mobile devices typically download simpler pages with less objects. We run tests using all browsers available within WebPageTest: Microsoft Internet Explorer, Mozilla Firefox and Google Chrome, and we emulate mobile browsers by changing both `User-Agent` and screen resolution; we consider an Android smartphone, a tablet of the Nexus series, an Apple iPhone6 and an iPad2. We call this dataset *Browser-Dataset*.

4.2.3 Impact of client's country

We want to study if and how websites change behavior when visits come from different countries. To comply with different local regulations, a website could react differently and install different sets of cookies depending on the client location (as obtained from the client IP address or language settings). To validate this hypothesis, we use HTTP-proxies to change the client IP address and country. For this experiment, we follow the same approach employed in [50]. We use 9 proxies located in 9 different countries (8 in Europe, one in the USA), changing the locale settings accordingly. In this case we use the desktop version of Google Chrome. We refer to this dataset as *Country-Dataset*.

4.2.4 Impact of time

Finally, we want to analyze how the number of profiling cookies varied over the past 4 years. As our measurement campaigns started in 2017, we profit from a publicly available dataset collected by the HttpArchive team [5]. They used WebPageTest to visit to a large set ($>100\ k$) of websites every two weeks. The experiments

1 SimilarWeb provides per-country ranks for 21 EU MSs out of 28.

Dataset	Websites	Size	Experiment Goal
Large-Dataset	35,862	195.5 GB	Large-scale analysis varying country and category
AcceptCookie-Dataset	241	633 MB	Analyzing the impact of giving consent to cookies
Browser-Dataset	241	2.2 GB	Analyzing the impact of varying browser
Country-Dataset	241	2.9 GB	Analyzing the impact of varying location
LongLasting-Dataset	241	15.1 GB	Analyzing how situation changed over 4 years

Table 1. Description of datasets.

were performed by test servers located in USA.[2] The dataset is publicly available [6], and contains the HAR files resulting from each visit. In this work, we use the data collected from January 2015 to November 2018. All the 241 websites in our specific measurement campaigns are part of the HTTP Archive data. Even if such data was not collected under our control, it has been collected by WebPageTest. Manual inspection suggests a level of reliability comparable to our testbed. Yet, HttpArchive experiments lack of diversity in terms of measurement location, employed browsers and website list. We refer to this dataset as *LongLasting-Dataset*.

4.3 Limitations

Despite we followed the best practices in designing our experiments, our methodology and measurement campaigns are subject to limitations. We briefly discuss them in the following.

• Our strict definition of profiling cookies might lead to false negatives – i.e., we might incorrectly ignore situations in which violations happen. This is the case of systems exploiting advanced tracking mechanisms, such as Flash cookies, or fingerprinting techniques. We may also incur false positives if we incorrectly label a cookie as profiling. However, our requirements (domain installing cookie must appear in both Ghostery and Disconnect tracker lists, and cookie must have a lifetime longer than one month) make this case very unlikely. Furthermore, websites typically set tens of offending cookies, mitigating the impact of isolated classification errors (see Section 5.5 and Section 5.6).

• Considering the methodology, we adopt solutions developed in previous studies, as our goal is not to design new classification methodologies, but rather to thoroughly quantify violations.

• Considering data collection from smartphones (*Browser-Dataset* campaign), changing `user-agent` and screen size may be insufficient to emulate different classes of devices. Indeed, installed cookies may vary accordingly to inner peculiarities of browsers, devices, or Javascript engines.

• At last, the choice of the top-100 websites per country and category is not representative of overall services. For this, we offer CookieCheck as a flexible tool that one can use to check websites of interest.

5 Results

This section presents our results. We first look into how popular third-party cookies are, and which are the third parties installing them. Next, we investigate the features we use for classifying profiling cookies, unveil their usage in *Large-Dataset*, and quantify the robustness of our results. We investigate in details the impact of providing consent, and changing browser, device or country on cookie usage. Finally, we show how violations varied across the past 4 years.

5.1 Quantifying the usage of third-party cookies at scale

We start by analyzing the usage of third-party cookies across web pages before the user consent is obtained. While this does not necessarily represent a violation of the ePrivacy Directive in its strictest interpretation (see Section 7.2.1), it helps quantifying the extent of the phenomena. For this we leverage *Large-Dataset*.

First, we notice that on average 74% of websites install third-party cookies. Even categories which one may expect not to embed any advertising –"Law and Government"– shows 39% of websites embedding third-party cookies. Interestingly, Adult websites come second. They likely offer little (or specializes) ads, thus install fewer third-party cookies. On the opposite side, most of

2 As seen in Section 5.6, the location and country of the measurement server do not affect the number of installed profiling cookies.

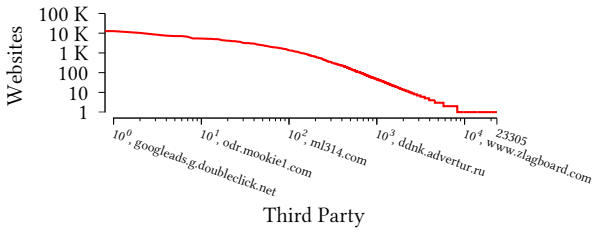


Fig. 1. Third-party services installing cookies in *Large-Dataset* websites. Domains are ranked based on the number of websites they cover. Notice the log-scale on both axes.

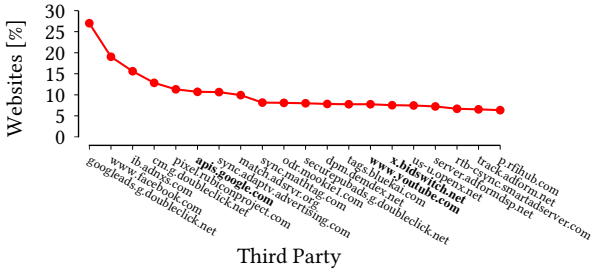


Fig. 2. Rank of the 20 third-party services installing the largest number of cookies in websites of *Large-Dataset*. Services are ranked based on the fraction of websites they cover.

websites in "News and Media" install third-party cookies (93% on average). This is no surprise as these websites typically build on advertisements, as well as analytics services. In fact, the per-website average of third-party cookies ranges from 2.64 for websites in "Law and Government" category to 24.71 for "News and Media".

If we consider per-country results, we observe a rather uniform picture. Values span from 60% (Croatia) to 80% (UK). Looking outside Europe, websites tend to install more third-party cookies at the first visit, with Russian (94%) and US (81%) websites on top. Details can be found on Appendix.

Take away: *On average 74% of websites install cookies from third parties before any user consent. The scenario is rather flat across countries, while we observe substantial differences across website categories.*

5.2 Third-party services installing cookies

In this section we dig into the ecosystem of third-party services that install cookies. We start by plotting in Figure 1 the number of domains (intended as hostnames extracted from URLs) of third parties installing cookies in *Large-Dataset* websites. Third parties are ranked based on the number of websites they are contained in. As shown, the rank is long-tailed, with a total of 23,305 third-party systems installing at least one cookie. The most popular ones are present in thousands websites. For instance, *googleads.g.doubleclick.net* installs its cookies on more than 10,000 websites.

Let us then focus on the top-20 third-party services installing cookies. Figure 2 reports percentage of websites they cover. Domains not in bold correspond to systems that are found in both Ghostery's and Disconnect's lists of trackers.[3] Considering the whole rank, 12.1% of them are labeled as *trackers*. We leverage Ghostery's website to collect further information, resumed in Table 2. In particular, we show the name of the company managing the domain, the country in which such company is established and a brief description of the service activity. First, we observe that, according to Ghostery, most prominent third parties installing cookies are devoted to *behavioral advertising* activities (A). The only exception is *www.facebook.com* which belongs to "Social Media" (B). Few offer services to end users (C), and all of these share data with other third-party systems (D). Second, 16 out of 20 are established in the US, where laws regulating the usage of users' personal data are more permissive than in EU. Finally, about half of these domains are managed by big players of the web.

Take away: *Most pervasive third parties managed by big players of the web are responsible for installing cookies in tens of thousands of websites. The majority of these is devoted to personalized advertising activities, whose mother companies are established outside Europe.*

5.3 Distinguishing profiling cookies

The scenario depicted above demonstrates that most of websites popular among European users install third-party cookies at the first visit, even before the user is given the option to accept them. The most diffused third-party cookies are from domains classified as advertising trackers from popular tracker-blockers. However, this is not conclusive to understand to what extent the ePrivacy Directive's principles are violated, as not all third-party cookies are installed to track users. Therefore, we detail in the following the methodology we use to distinguish *profiling* cookies, i.e., those cookies used for profiling and tracking users that violate the ePrivacy Directive if installed before obtaining user consent.

3 We use the intersection of the two lists to reduce the probability of misclassifying a third-party domain as a tracker.

Third Party	Company	Company Country	Description	Pervasiveness
googleads.g.doubleclick.net	Google DoubleClick	US	A,D	27.0 %
www.facebook.com	Facebook	US	B,D	19.0 %
ib.adnxs.com	AppNexus	US	A,D	15.6 %
cm.g.doubleclick.net	Google DoubleClick	US	A,D	12.8 %
pixel.rubiconproject.com	Rubicon	US	A,D	11.3 %
apis.google.com	Google	US	C,D	10.7 %
sync.adaptv.advertising.com	Advertising.com	US	A,D	10.6 %
match.adsrvr.org	TradeDesk	US	A,D	9.9 %
sync.mathtag.com	MediaMath	US	A,D	8.1 %
odr.mookie1.com	Media Innovation Group	US	A,D	8.1 %
securepubads.g.doubleclick.net	Google DoubleClick	US	A,D	8.0 %
dpm.demdex.net	Adobe	US	A,D	7.8 %
tags.bluekai.com	Oracle	US	A,D	7.7 %
www.youtube.com	Google	US	C,D	7.7 %
x.bidswitch.net	IPONWEB	UK	A,D	7.5 %
us-u.openx.net	OpenX	US	A,D	7.4 %
server.adformdsp.net	Adform	Denmark	A,D	7.2 %
rtb-csync.smartadserver.com	SMART AdServer	France	A,D	6.7 %
track.adform.net	Adform	Denmark	A,D	6.5 %
p.rfihub.com	Rocket Fuel	US	A,D	6.3 %

Table 2. Details of the top 20 third-party services installing the largest number of cookies in websites of *Large-Dataset*. Description is derived from Ghostery privacy tool [41]. A: Behavioral Advertising, B: Social Media, C: Offers page contents, D: Shares data with 3rd parties.

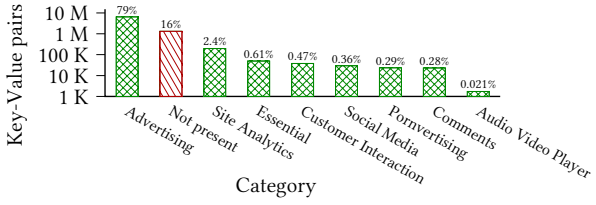


Fig. 3. Ghostery's classification of the third-party cookies (in terms of *key-value pairs*) in websites of *Large-Dataset*. Notice the log-scale on the y-axis.

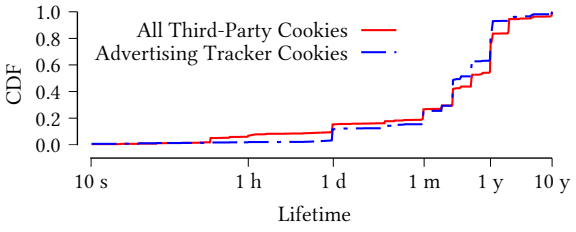


Fig. 4. Empirical distribution of lifetimes of third-party cookies in *Large-Dataset*. Notice the log-scale on the x-axis.

We leverage Ghostery's and Disconnect's lists of trackers. We perform a first analysis looking at third-party cookies encountered in *Large-Dataset* in terms of unique key-value pairs contained in them, i.e., possibly new values assigned for each "new" visit, and match the corresponding domain in the Ghostery's list. Results are shown in Figure 3. It details the category of the third parties according to Ghostery classification. For instance, more than 7M key-value pairs are installed by services falling in the advertising trackers category. Indeed, each time we encounter an advertising tracker, it sees us as a "new" customer. Hence, it assigns us a new identification value. We manually verify that keys that assume a very large number of unique values are used for profiling. For example the key "`id`" in the *googleads.g.doubleclick.net* cookie has been set 46,305 times, always with different values. Similarly "`anj`" key from *ib.adnxs.com* and "`rpx`" key from *pixel.rubiconproject.com* have been found respectively 109,254 and 85,737 times, each with a unique value. We omit to report the results for Disconnect as they are very similar.[4]

Next, out of all the domains classified as trackers by the intersection of Ghostery and Disconnect, we restrict our focus to only those appearing in the advertising category. We adopt this approach for two reasons: First, the high number of key-value pairs supports the intuition that those third-party cookies are actually used for profiling. Second, advertising trackers are known to be the most pervasive and prone to threat users' privacy [10, 49]. At the end, out of the total 2,870 unique entries in the union of the lists, we obtain 738 domains.

We want to make our definition of profiling cookie even more stringent by considering cookie lifetime. A similar approach was used by the authors of [31]. In-

4 Assessing the accuracy of the tracker lists provided by Ghostery and Disconnect is out of the scope of this paper.

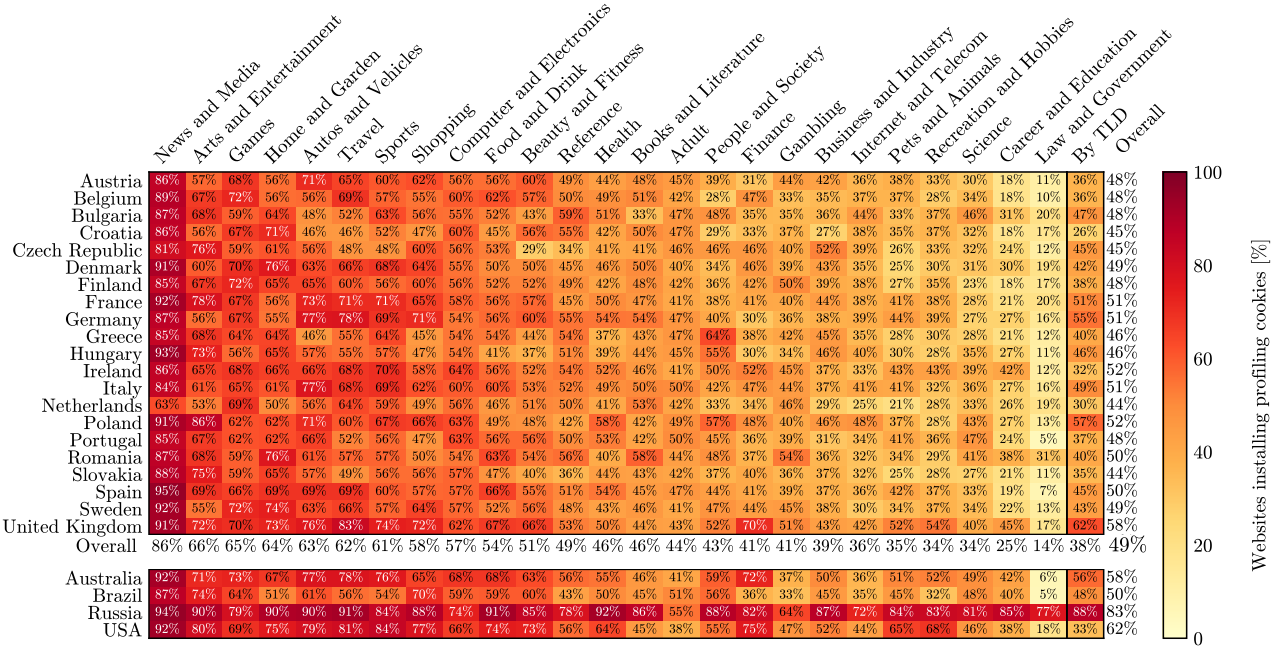


Fig. 5. Per-country and per-category fractions of websites installing at least one profiling cookie. At the end of each row (column) it is reported the country-(category-)wise average.

deed, cookies containing a far-in-time expiration date are more likely to be used for tracking purposes. To define a threshold, we compute the cumulative distribution of cookie lifetimes observed in *Large-Dataset*. Figure 4 shows the results. The solid line is calculated considering all third-party cookies, while the dashed one refers only to those installed by advertising trackers present in Ghostery's and Disconnect's lists intersection. As shown, 80% of third-party cookies last 1 month or more. The two curves mostly overlap, showing that trackers seldom use short-lived cookies. Based on this result, we add a further condition to consider a cookie to be "profiling cookie": it must exhibit a lifetime equal or greater than 1 month.

Take away: *We tag a cookie as profiling if installed by a third-party domain classified as an advertising tracker by both Ghostery and Disconnect, and if it exhibits a lifetime greater or equal than 1 month. We use this definition in the following.*

5.4 Assessing violations at scale

In this experiment we are interested in understanding whether websites users violate the requirements of the Directive on the use of profiling cookies. Leveraging the definition of profiling cookie described above, we analyze their usage across web pages of *Large-Dataset*. We offer a breakdown of this analysis for each country and category. We report results in Figure 5, where each cell details the fraction of websites that installed at least one profiling cookie during any of the 5 visits. Each cell cosniders 100 websites. Each row refers to a country, and each column to a category. The penultimate column (*By TLD*) considers websites by country code top-level domain (e.g., only **.fr* websites for France). The last column reports the country-wise average fraction; last row reports the average across EU countries per each category. Columns report the category-wise average fraction for EU countries only, and are sorted based on the average number of installed profiling cookies. Despite our definition of profiling cookies is very stringent, we observe the average number of websites using them is worryingly high: on average 49% of websites do (all in all average in Figure 5). Recall those are installed without prior user consent, thus violating the ePrivacy Directive. These figures thus constitute a lower bound on the violation fractions. To provide a possibly upper bound, we compute the fraction of websites installing any thidrparty cookie. Details are in the Appendix. Here the overall violations top to 74%.

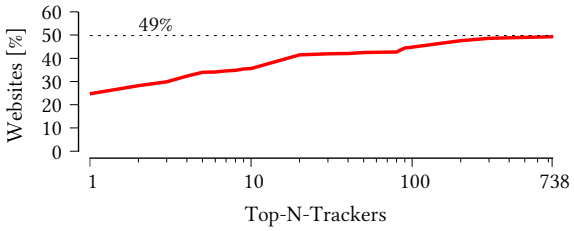


Fig. 6. Fraction of websites violating the Directive when considering profiling cookies installed by the N most pervasive trackers.

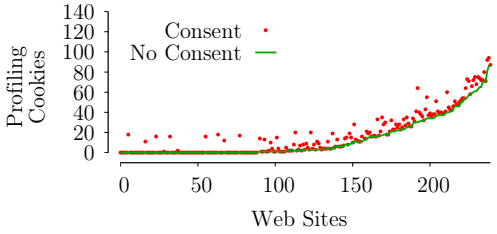


Fig. 7. Per-website number of profiling cookies installed before and after giving consent.

Country	Banner		No Banner	
	No Refresh	Refresh	No Cookie	But Cookies
France	69	2	11	**18**
Germany	31	0	18	**51**
Italy	53	14	15	**18**

Table 3. Number of websites showing the Cookie Bar, and refreshing the page once user consent is given.

Recalling that our measurements neglect cookies installed via Javascript code and other tracking mechanisms, the overall picture is startling. Even the category with the lowest fraction of non-compliant websites –"Law and Government"– shows 14% of violations, that grows to 38% considering all websites installing third-party cookies. This is due to the presence of ad-supported websites beside institutional portals. Websites in "News and Media" get the largest fractions with 86% (92%) of violations on average. This is no surprise as these websites typically offer lots of advertisements, which install their cookies to profile users without any consent.

Focusing on profiling cookies per-country results, EU countries show very similar results. Values span from 44% (Slovakia and the Netherlands) to 58% (UK). It worths remarking that the ePrivacy Directive had a positive effect in the Netherlands which transposed the Directive rather strictly: we record 63% of violations in "News and Media" –the lowest for this category– and 44% of violations in general. Looking outside Europe, websites tend to install more profiling cookies at the first visit, with Russian (83%) and US (62%) websites being negative examples. This demonstrates the ePrivacy Directive is having some positive effect if compared to countries with weaker or no regulatory frameworks such as US and Russia.

Finally, we consider websites separately per country code top-level domain (*By TLD* column). As such, the resulting lists never overlap, allowing an analysis not biased by the popularity of transnational portals (social networks, search engines, etc.). The overall considerations still hold, with a slightly higher variability among countries.

To complement above results, we investigate to what extent our results are robust to the employed tracker list. We run other experiments aiming at understanding how violation fraction varies when we change the list of considered trackers. In particular, we consider the top-N most popular third-party trackers. In Figure 6 we report the fraction of websites that use profiling cookies while we increase the number of considered trackers. If we consider the most pervasive tracker alone, *googleads.g.doubleclick.net*, we obtain an overall fraction of violations equal to 22%. The number gets close to 40% if we consider the top 10 trackers. Clearly, this number saturates to 49% (the all in all average in Figure 5) when we consider all trackers.

Take away: *A clear trend emerges: 49% of websites popular among European users violate the ePrivacy Directive as they embed profiling cookies installed by third-party trackers without any prior user consent. Most popular trackers are responsible for the large fraction of violations. In particular, one of the main advertising trackers is managed by Google, and causes websites to violate the ePrivacy Directive on more than 20% of cases.*

5.5 Installed cookies upon consent

To respect the Directive websites must ask for consent to install cookies by means of a button embedded in a Cookie Bar, as explained in [13]. When clicked, the website refreshes (or updates) the page to deliver new and enriched content with new objects that trigger the installation of cookies.

We consider the *AcceptCookie-Dataset*, and count the websites which lawfully implement this procedure. We consider three main EU countries, namely, Germany, France and Italy which transposed the ePrivacy Directive into their regulations and exhibit the same fractions

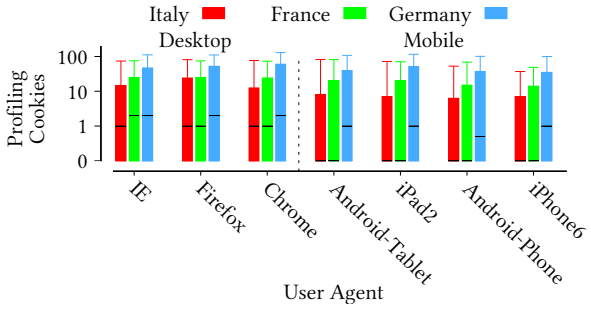


Fig. 8. Number of profiling cookies set with different browsers and devices. Notice the log-scale on the y-axis.

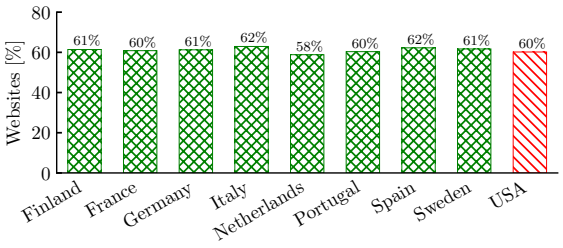


Fig. 9. Fraction of websites violating the Directive when changing user location. No significant difference is observed.

of violations in Figure 5. As reported in Table 3, we observe that dozens of websites in *AcceptCookie-Dataset* do not even provide a Cookie Bar, but regularly set profiling cookies! In particular, in Germany more than a half of websites do. Overall, 67 out of 241 websites in *AcceptCookie-Dataset* do not provide a Cookie Bar but set profiling cookies. Considering country ranks, in France and Italy, 71 and 67 out of 100 websites embed a Cookie Bar, respectively. For Germany, only 31 websites do that. Furthermore, the Cookie Bar consent button triggers a page refresh in only 14 cases for Italian websites, just in 2 cases for French, and in no cases for German.

We consider the same 241 websites to count the number of profiling cookies which are installed *before* and *after* we provide consent. We report results in Figure 7, where websites are sorted by the number of profiling cookies installed *before* consent (green line). To this baseline we then sum new cookies installed *after* consent has been given (red dots). As shown, only 43 websites do not install profiling cookies before obtaining consent. Among those, 34 never install cookies, while 11 correctly wait for user consent before installing them. All other websites, i.e., 80.5% install profiling cookies before consent, and possibly install more after that.

Take away: *67 out of 241 websites do not provide a Cookie Bar to let users provide consent, but install some profiling cookies anyway. Among the remaining ones, only 7% wait for user's consent before installing profiling cookies.*

5.6 Impact of device and location

We study *Browser-Dataset* dataset to check whether installed profiling cookies vary when changing browser, device, or country.

Figure 8 reports the number of installed profiling cookies when using different browsers and devices, separately for Italian, French and German websites. Box plots span from the 1^{st} to the 3^{rd} quartile, while whiskers report the 10^{th} and the 90^{th} percentiles; black strokes represent the median. A quite large amount of profiling cookies is installed independently on the kind of browser or device, and the number only slightly decreases for mobiles. This is likely due to the simpler structure of pages served to mobile devices. Looking at considered countries, German websites install more profiling cookies than French and Italian, respectively.

Finally, considering the *Country-Dataset*, we visit websites from 9 countries (Finland, France, Germany, Italy, Netherlands, Portugal, Spain, Sweden and US). We employ 8 proxies (in addition to the client in our country). Figure 9 shows the results. The number of profiling cookies does not change, as well as the fraction of websites violating the ePrivacy Directive. Figure 9 clearly shows that violations occur in $\approx 60\%$ of cases for whichever location. We conclude that websites do not adapt the set of profiling cookies to install if the country of the visitor implements the Directive in a different way.

Take away: *The set of profiling cookies installed by websites before users' consent is the same independently from the client's browser or device. Also, user location does not change the picture, even when outside EU.*

5.7 A view over the last 4 years

In this section, we investigate how the scenario evolved over the last years. We use the *LongLasting-Dataset* to study how the fraction of violating websites varied from 2015 to 2018. As done for previous experiments, we tag websites violating the ePrivacy Directive whenever they install at least one profiling cookie without obtaining prior consent. For this analysis we consider data col-

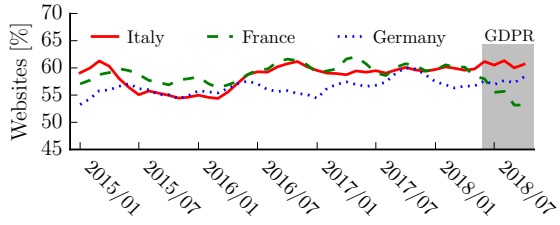
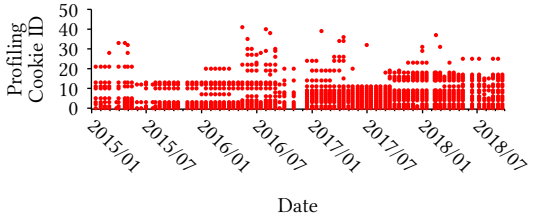
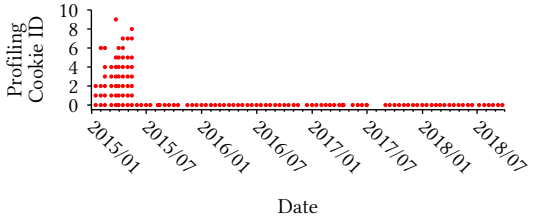


Fig. 10. Fraction of websites violating the Directive over 4 years.



(a) *www.tf1.fr*



(b) *www.repubblica.it*

Fig. 11. Profiling cookies for two websites over the last 4 years. Each line represents the presence/absence of a particular cookie.

lected from January 2015 to November 2018. Results are reported in Figure 10, separately for Italy, France and Germany. The picture depicts a quite flat scenario, and values fall in [52-63]% range. For Italy, we observe a small decrease around May 2015, time at which the Italian Data Protection Authority issued a new regulation that establishes sanctions for ePrivacy Directive violations [39, 40]. The entry into force of the new GDPR in May 2018 (highlighted in gray in the figure) seems to not generate changes yet, even if we observe a negative trend in France. The overall fraction of violating websites was 57% in January 2015 and stands at 56% in November 2018.[5] We conclude that all in all the picture is rather consistent, and the ePrivacy Directive has not diminished users' exposure to tracking technologies in the last four years.

We complement the above result with two examples of common behavior. Figure 11 shows the profiling cookies installed by two news websites for the entire duration of the *LongLasting-Dataset*. The figure reports measurements collected every two weeks from 2015 to 2018. The x-axis represents time, while the y represents a specific profiling cookie found in the dataset. The presence of a red point indicates that such cookie was installed at that time. Looking at *www.tf1.fr* in Figure 11a, we notice that dozens of profiling cookies are regularly installed before consent. Occasionally other cookies appear, due to the dynamicity of the ecosystem related with advertisement. A different consideration holds for *www.repubblica.it* in Figure 11b. Up to 9 different profiling cookies were regularly installed in the first half of 2015. Starting from June 2015, the number drops to one. This can be explained by the new national regulation entered into force in May 2015. Since that date, the website has introduced a new Cookie Bar, and removed all profiling cookies but one.[6] The profiling cookie that still appears after June 2015 belongs to *imrworldwide.com* a well-known tracking platform. Understanding why this profiling cookie keeps being installed is out of the scope of this study.

Take away: *The overall fraction of violating websites is constant over the last 4 years, and demonstrates that the ePrivacy Directive did not reach the goal of diminishing users' exposure to tracking technologies over the last years.*

6 Related Work

To the best of our knowledge, few works specifically address the problem of understanding whether the ePrivacy Directive is respected by websites. Leenes *et al.* [46] focused their study on 100 Dutch websites manually visited and observed that most of them do not respect the ePrivacy Directive. Borghi *et al.* [17] targeted 200 UK websites, concluding that violations are present in 84% of cases. In another study, Carpineto *et al.* [18] analyzed Italian Public Administration websites, and claimed that violations of the ePrivacy Directive occur in 6.6% of cases. Similar observations emerged as a side effect of our previous work about tracker-blocker effectiveness [58]. However, none of the previous works builds

5 Such values do not match the ones of Figure 5 as considered websites are different.

6 Once obtained the user's consent, the page is reloaded and other cookies get installed.

on a dataset large enough to achieve solid conclusions. Here we are the first to build a methodology to address this problem at scale. We present results obtained from hundreds of thousands of visits, and present statistically robust observations.

Related to the usage of cookies in the wild, it worths mentioning Gonzales *et al.* [42]. Their work presents a large-scale characterization of cookies derived from traffic traces. Some of presented findings are tangent to this work: Gonzales *et al.* show that some trackers, e.g., *krxd.net*, use the cookie name to concatenate various user information, making the number of unique identifiers associated to a user explode. However, our results (Figure 3) are not affected, as we observe a bunch of trackers actually encapsulating very dynamic information in the cookie name, but these are not sufficient to introduce any bias in results. Another prominent work related to ours is [30] from Englehardt *et al.* They addressed the problem of characterizing online tracking on very large number of websites. The dataset authors collect contains the same information we build upon to understand if a website violates the ePrivacy Directive, but they did not consider that the "first visit" may be different due to regulations in place. Other works [10, 31, 49] study how cookies are used as identifiers by trackers, and may expose sensitive information to eventual eavesdroppers [19]. Englehardt *et al.* [31] observe that cookies alone are enough to reconstruct up to 73% of users' browsing history, while, similarly to what we observe, Li *et al.* [49] find that 30% of Alexa top 10,000 websites use one of the top 5 most pervasive trackers. Finally, other recent studies [15, 54] show that users are continuously tracked when online, and tracking technologies have evolved dramatically beyond cookies: HTML5 Local Storage, Etags, Flash and fingerprinting [10, 28, 60]. Trackers are widespread even in commercial emails via similar methods such as embedded pixels [29].

Regarding the Cookie Law, some other works studied it from a legal perspective [20, 45, 47, 51]. All concluded that this is a case of regulatory failure. Koops [45] argues that the Directive creates the illusion that individuals have control over their data, and it is difficult for webmasters to respect it. Differently, Markou [51] pinpoints that the cause of the failure lies in resistance of advertisement ecosystem and in the lack of privacy awareness among users. A similar remark is made by Cofone [20]. Leenes *et al* [47] highlight four *"different flavours"* of the cookies regulatory failure: (i) the opt-out failure; (ii) the coexistence of different national implementations of the ePrivacy Directive; (iii) the lack of interest and attention of users and general public; (iv) the lack of a concrete enforcement. As a result, the cookie legal saga shows the difficulty of the European legislator in managing the complex scenario offered by tracking technology advertising industries' business models, users' rights to privacy and confidentiality, and regulatory bodies. Our study does not focus on the legal reasons of the failure, but quantifies it by showing technical evidences.

7 Background

In this section we first describe the cookie technology in the context of online tracking. Then, we introduce the European regulatory framework that governs the usage of cookies. For the latter, we involved experts in regulations to detail the possible interpretations of the law.

7.1 HTTP cookies as tracking technology

HTTP cookies are used to preserve state of HTTP transactions in web browsers, and have been standardized in IETF's RFC 6265 [16]. A cookie is defined as a set of `{name=value}` pairs. It may have attributes such as `expires=date`, `path=path`, `domain=domain_name`, `secure`. To set a cookie, a server can either use the `Set-Cookie` header in HTTP responses, or use Javascript to write entries. Once a cookie is stored, the browser includes it in the header of each HTTP request headed to servers belonging exclusively to domain `domain_name` with path `path` until `date` expires.

When a user visits some web page W_1, hosted by domain D_1, the browser typically initiates HTTP transactions to fetch URLs $U_2, ..., U_n$ contained in the page, to, e.g., get images, `css`, `js` files, etc. Some of these objects may be served by other domains $D_2, ..., D_n$. We refer to the cookie installed by D_1, i.e., the domain the user intentionally visited, as a *first-party* cookie. Any of the cookies installed by $D_2, ..., D_n$ are *third-party*. The cookies employed to track users belong to this second family as typically the tracker domain D_i is different from D_1.

Also expiration time makes cookies different. *Session cookies* are temporary, i.e., deleted when the user closes the browser or after a short period of time. *Persistent cookies* contain an explicit expiration date and may stay stored in the browser for long time.

Third-party persistent cookies are the most prominent means used by trackers to reconstruct users' habits trajectories and compute per-user profiles which are employed to, e.g., deliver personalized advertisement [31]. As these threaten users' privacy, policymakers undertook initiatives to regulate their usage.

In this paper we focus on cookies installed by well-known trackers that use persistent profiling cookies.

7.2 The current European regulatory framework for web privacy

At the international level, no comprehensive regulatory framework exists concerning web privacy, and each country eventually provides its own [52]. In this paper we focus on the European scenario, which is considered to have one of the most comprehensive frameworks (alongside with the Canadian and the Swiss ones) for the regulation of personal data collection and usage in web communications. Three motivations move our choice: i) it provides strict rules on tracking; ii) it has broad territorial application; iii) it is currently being updated.

The first European regulation on web tracking was introduced in 2002 by Dir-2002/58/EC on privacy and electronic communications [21]. Article 5(3) prescribed two conditions for "the use of electronic communications networks to store information or to gain access to information stored in the terminal equipment of a subscriber or user": i) "the subscriber or user concerned is provided with clear and comprehensive information" pursuing the Directive 95/46/EC principles; and ii) "is offered the right to refuse such processing by the data controller".

This Directive has been amended in 2009 [22]. According to its last version, Article 5(3) explicitly disciplines the conditions for the use of any tracking "devices" (e.g., cookies, supercookies, fingerprinting, snippets, etc.), moving from the necessity of the user's consent. It "is only allowed on condition that the subscriber or user concerned has given consent, having been provided with clear and comprehensive information, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing". The European legislator rejected the opt-out mechanism for the opt-in one, prescribing the necessity of the user's prior informed consent before using any tracking technology.

The Opinion 15/2011 on the definition of consent [11] of the Article 29 Data Protection Working Party (Art. 29 WP) clarifies how to manage the prescription of the user's informed consent.[7] It details the consent procedure, stating that "consent based on the lack of individuals' action [...] *does not meet* the requirements of valid consent under the Directive 95/46/EC". Thus, "browser settings which would accept by default the targeting of the user (through the use of cookies)" are not possible. It fails "to meet in particular the requirements for an unambiguous indication of wishes. It is essential that the data subject is given the opportunity to *make* a decision and to *express* it, for instance by ticking the box himself, in view of the purpose of the data processing". In 2013 [13], Art. 29 WP went further providing guidance on obtaining consent for cookies and lists four requirements to be jointly fulfilled: i) specific information must be provided; ii) consent has to be obtained *before* the cookies' setting; iii) the consent should express an *active* choice; iv) the consent should be freely given exercising a real choice.

It follows that the websites have to offer i) a short resume of the privacy policy to the user, ii) a link to the page containing all details, and iii) an interactive element to ask users' consent to install tracking devices. All this information is typically provided in an "Accept Cookie Bar" which is offered to the user the first time she visits the website.

7.2.1 Cookies subject to prior consent

Dir-2002/58/EC defines two exceptions to the informed consent principle: technical storage or access to information is allowed "for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as *strictly necessary* in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service". In the Opinion 04/2012 on Cookie Consent Exemption [12] the Art. 29 WP explicitly clarifies which kinds of cookies are exempted from the requirement of informed consent: session cookies, and cookies that are *essential* to provide the service are exempted (e.g., to handle a cart in an e-commerce website). Informed consent is *required* for all the others. The document explicitly observes that third-party persistent cookies are not

7 Art. 29 WP is an EU advisory body set out in Art. 29 of Dir-1995/46/EC (Data Protection Directive). Opinions are documents which help practitioners and players interpret, understand and apply the content of directives.

Interpretation	Subject to consent			
A				
B				✓
C			✓	✓
D		✓	✓	✓
E	✓	✓	✓	✓
Type of cookie	**First-party**	**Technical third-party**	**Non-technical third-party**	**Profiling**
Examples	Authentication, user-input memory, security	Multimedia player settings, load balancing	Social plug-ins, performance analytics	Behavioral advertising, tracking services

Table 4. Cookies subject to consent in different interpretations of the ePrivacy Directive. We verify violations according to (B).

essential and, thus, not exempted. This is the case of profiling cookies.

However, such guidelines are not part of the ePrivacy Directive. Hence, the definition of a "necessary element" to offer a service is subject to interpretation by National Data Protection Authorities. We illustrate different approaches in Table 4, based on different definitions of necessary cookie. Helped by Opinion 04/2012, we identify four types of cookies, represented as columns in the table. Starting from the left, they are sorted from the most to the least likely to be necessary. Indeed, first-party cookies are very likely to be necessary (e.g., for authentication), while profiling cookies, installed by e.g., behavioral advertising platforms, are very far from being essential. We then formalize five different interpretations of the ePrivacy Directive that differ in the type of cookies subject to consent. The most permissive approach (A) represents the extreme scenario where no cookie is subject to consent, that clearly breaks the spirit of the Directive. On the opposite side, the approach (E) represents the strictest interpretation, which mandates that all cookies need prior consent. In this work, we verify the compliance to the interpretation (B), which is the most tolerant approach that mandates consent for some kind of unnecessary cookies. According to (B), only profiling cookies need prior users's consent, while all others do not. Thus, our definition of violations becomes very conservative – i.e., we prefer to neglect some violations, than to incur false positives.

Fig. 12. ePrivacy Directive transposition into EU MSs legislation.

7.2.2 Transposition into national legislation

The European regulatory landscape is not homogeneous. All EU MSs must transpose EU Directives into their legislation as a minimum level of harmonization. As such, EU MSs have transposed the ePrivacy Directive in different ways [44, 56]. For the sake of simplification, we group such transpositions into three categories. A graphical representation of ePrivacy Directive interpretations (and EU MSs of application) is provided in Figure 12.

(i) The strictest category prescribes the opt-in mechanism – i.e., the user must provide consent. The consent must be obtained *before* cookies are used. This interpretation is in force in 11 EU MSs.

(ii) More flexible interpretations allow the consent to be *implied* – i.e., inferred from the behavior of the user. In other words, consent can be conveyed by clicking a link or scrolling through a website. In 9 MSs, regulations allow implied consent.

(iii) Tolerant interpretations of the ePrivacy Directive do not require consent. The website must show a Cookie Bar warning users about the use of cookies, that can be installed without any prior consent. Websites must

provide an opt-out mechanism to let users get rid of cookies. Legislations of this kind are in force in 7 MSs.

In this work, we embrace the first two interpretations. The ePrivacy Directive and Art. 29 WP Opinions contain explicit references to the opt-in mechanism, and we believe categories (i) and (ii) adhere more properly to the spirit of the Directive. In other words, websites installing cookies without prior consent (even if *implied*) are considered violating the ePrivacy Directive in the paper. Experiments in this paper are performed from Italy, that implements the opt-in mechanism.

Finally, the ePrivacy Directive does not sketch procedures to guide the enforcement of its principles, nor provides guidelines to perform proper audits. Despite Recital 66 of Dir-2009/136/EC amending Dir-2002/58/EC disposes that "The enforcement of these requirements should be made more effective by way of enhanced powers granted to the relevant National Authorities", this point has been largely ignored.

8 Discussion

In this section we present the reasons we identify behind the problematic issues of the ePrivacy Directive, and add some considerations on the new e Privacy Regulation proposals.

8.1 Reasons behind the failure of the ePrivacy Directive

Our results demonstrate that a large fraction of websites ignores the ePrivacy Directive. We identify three main reasons behind this:

(*i*) The Directive does not offer guidelines to perform systematic auditing procedures with a coordinated supervision. Two official reports from independent third parties confirm this [26, 57]. In particular, Deloitte [26] emphasizes that enforcement of rules is "insufficient and inconsistent", as currently in charge of each EU MS's Authority which "tend to audit in cases where there is a specific risk or complaint by an individual. Ex-officio audits remain a minority."

(*ii*) The lack of automatic tools which can verify whether a website violates the Directive makes it possibly complicated for the deputed agencies to plan systematic audits.

(*iii*) Users' consciousness of their online privacy is still too low. EU promulgated the ePrivacy Directive without accompanying proper awareness campaigns aiming at educating users about how their privacy is threaten under the surface of the web. Regarding this, the EU body Regulatory Fitness and Performance (REFIT), which is in charge of verifying effectiveness of directives, states the current rules end being counter-productive as "the constant stream of cookie pop-up-boxes that users are faced with completely eclipses the general goal of privacy protection as the result is that users blindly accept cookies" [53].

8.2 The new Proposal for the ePrivacy Regulation

In January 2017 a Proposal for the ePrivacy Regulation which repeals the ePrivacy Directive [32] was published by the European Commission. Hopefully, it will change positively the current scenario. After a long review process that involved stakeholders such as providers, regulatory bodies, citizens' associations [14, 33, 35, 38, 48], the Proposal entered the EU approval procedure. In October 2017 and in September 2018, the EU Parliament [37] and the EU Council [24] proposed two new revised texts, respectively.

The proposed Regulation, as announced by the EU Digital Single Market Strategy [9], will integrate the privacy regulatory framework and shall be read in close connection with the new General Data Protection Regulation which entered into force in May 2018 in all EU MSs [23], as part of the same EU strategy. Among its objectives, it will draw the new rules on cookies and a more effective enforcement mechanism for such rules. Moreover, the proposed Regulation will also extend the territorial scope of the safeguards for the protection of the confidentiality to electronic communication services from outside the European Union to end-users inside.

Both the EU Commission and the Parliament stressed the importance of the privacy-by-default principle in designing websites, giving users a free choice on the use of tracking devices and techniques. This point was emphasised by the European Data Protection Supervisor (EDPS) who invites the legislators to ensure that consent will be genuinely freely given with no prejudice to the access and fruition of web services [34, 35], and by the European Data Protection Board (EDPB) in its Statement on the ePrivacy Regulation [36] published on May 2018. The EDPB also recommended that "In order for consent to be freely given as required by the GDPR, access to services and functionalities must not be made conditional on the consent of a user to the

processing of personal data or the processing of information related to or processed by the terminal equipment of end-users, meaning that cookie walls should be explicitly prohibited".

Notwithstanding, the EU Council proposed to revise again how to design websites privacy settings on cookies [25]. This said, as far as November 2018, the revision process is far from being closed.

8.3 Unsolved issues

The new Proposal for the ePrivacy Regulation will surely improve current scenario. However, it will have to face the same difficult problem addressed by the ePrivacy Directive, i.e., regulating the troubled marriage between ad-fueled web services willing to monetize their content and users desirous of preserving their privacy. In such a complex ecosystem, dominated by the interests of advertisers, we believe issues (i) and (ii) of Section 8.1 will be key to avoid another failure. Indeed, we advocate EU will subsidize the development of automatic and scalable auditing instruments and require Data Protection Authorities to perform systematic and regular audit campaigns to verify the enforcement of the new Regulation. In this context, we believe tools like CookieCheck are extremely valuable. Moreover, they may help in fostering the principle of transparency of the data processing.

9 Conclusion

In this paper, we run large-scale measurement campaigns and testify that a wide fraction of websites does not respect the the Cookie Law set up by the ePrivacy Directive, with a few popular third parties causing such violations. To this end, we designed CookieCheck, a simple tool to audit whether a website violates the ePrivacy Directive. It is available online [3], along with its source code [4]. We also share the datasets collected for this study [2].

Apart from being helpful for increasing users' online privacy awareness, the results, as well as the tools and discussion provided with this study, are useful for researchers, policymakers and players of the web industry in the debate on how to monitor and enforce privacy policies on the web.

Considering the research community, we are among the first to face the verification of privacy regulations, and we hope the effort in producing tools for auditing privacy violations will increase and get momentum.

Finally, we are improving CookieCheck to include further checks, and we are contacting local Data Protection Authority to present our findings, hoping this will lead to a first step to find a remedy to the shady scenario depicted by our results.

Acknowledgments

The research leading to these results has been funded by the Vienna Science and Technology Fund (WWTF) through project ICT15-129 (BigDAMA) and the SmartData@PoliTO center for Big Data technologies. We are grateful to the Nexa Center for Internet & Society of Politecnico di Torino for assisting us in understanding all caveats of regulations and directives. We are grateful to the *scans.io* team for hosting our (large) dataset on their servers, and make it available for download to the community. Finally, the author Stefano Traverso acknowledges a grant for the Lagrange Project - CRT Foundation / ISI Foundation.

References

[1] Chrome DevTools Protocol. https://chromedevtools.github.io/devtools-protocol/.

[2] CookieCheck Dataset. https://scans.io/study/polito-har-crawl.

[3] CookieCheck online tool. http://cookiecheck.polito.it/.

[4] CookieChecker Code Repository. https://github.com/CookieChecker/CookieCheckSourceCode.

[5] HttpArchive. https://httparchive.org/.

[6] HttpArchive dataset. https://console.cloud.google.com/storage/browser/httparchive.

[7] Similarweb. https://www.similarweb.com/.

[8] Webpagetest. https://www.webpagetest.org/.

[9] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A Digital Single Market Strategy for Europe, COM(2015) 192 final), 2015.

[10] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A., and Diaz, C. The web never forgets: Persistent tracking mechanisms in the wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2014), CCS '14, ACM, pp. 674–689.

[11] Art. 29 Data Protection Working Party. Opinion 15/2011 on the definition of consent, 2011. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp187_en.pdf.

[12] Art. 29 Data Protection Working Party. Opinion 04/2012. on Cookie Consent Exemption, 2012. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2012/wp194_en.pdf.

[13] Art. 29 Data Protection Working Party. Working Document 02/2013 providing guidance on obtaining consent for cookies, 2013. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp208_en.pdf.

[14] Art. 29 Data Protection Working Party. Opinion 03/2016 on the Evaluation and Review of the ePrivacy Directive (2002/58/EC), 2016. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2016/wp240_en.pdf.

[15] Barford, P., Canadi, I., Krushevskaja, D., Ma, Q., and Muthukrishnan, S. Adscape: Harvesting and analyzing online display ads. In *Proceedings of the 23rd International Conference on World Wide Web* (New York, NY, USA, 2014), WWW '14, ACM, pp. 597–608.

[16] Barth, A. HTTP State Management Mechanism. RFC 6265, Apr. 2011.

[17] Borghi, M., Ferretti, F., and Karapapa, S. Online data processing consent under EU law: a theoretical framework and empirical evidence from the UK. *International Journal of Law and Information Technology 21*, 2 (2013), 109–153.

[18] Carpineto, C., Lo Re, D., and Romano, G. Automatic assessment of website compliance to the european cookie law with coolcheck. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society* (New York, NY, USA, 2016), WPES '16, ACM, pp. 135–138.

[19] Castelluccia, C., Kaafar, M.-A., and Tran, M.-D. Betrayed by your ads! In *Privacy Enhancing Technologies* (Berlin, Heidelberg, 2012), S. Fischer-Hübner and M. Wright, Eds., Springer Berlin Heidelberg, pp. 1–17.

[20] Cofone, I. N. The way the cookie crumbles: online tracking meets behavioural economics. *International Journal of Law and Information Technology 25*, 1 (2017), 38–62.

[21] Council of European Union. Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on Privacy and Electronic Communications), 2002. http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32002L0058.

[22] Council of European Union. Directive 2009/136/EC amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws, 2009. http://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32009L0136.

[23] Council of European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016.

[24] Council of the European Union. Proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_12336_2018_INIT&from=EN.

[25] Council of the European Union. Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications)- Examination of the Presidency text, Document nr. ST 10975 2018 INIT, 2018. https://www.consilium.europa.eu/register/en/content/out?&typ=ENTRY&i=ADV&DOC_ID=ST-10975-2018-INIT.

[26] Deloitte. Evaluation and review of Directive 2002/58 on privacy and the electronic communication sector, 2016. https://ec.europa.eu/digital-single-market/en/news/evaluation-and-review-directive-200258-privacy-and-electronic-communication-sector.

[27] Disconnect. https://disconnect.me.

[28] Eckersley, P. How unique is your web browser? In *Privacy Enhancing Technologies* (2010), vol. 6205, Springer, pp. 1–18.

[29] Englehardt, S., Han, J., and Narayanan, A. I never signed up for this! privacy implications of email tracking. *Proceedings on Privacy Enhancing Technologies 2018*, 1 (2018), 109 – 126.

[30] Englehardt, S., and Narayanan, A. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2016), CCS '16, ACM, pp. 1388–1401.

[31] Englehardt, S., Reisman, D., Eubank, C., Zimmerman, P., Mayer, J., Narayanan, A., and Felten, E. W. Cookies that give you away: The surveillance implications of web tracking. In *Proceedings of the 24th International Conference on World Wide Web* (Republic and Canton of Geneva, Switzerland, 2015), WWW '15, International WWW Conferences Steering Committee, pp. 289–299.

[32] European Commission. Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in Electronic Communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communcations), COM(2017) 10 final., 2017. https://ec.europa.eu/digital-single-market/en/news/proposal-regulation-privacy-and-electronic-communications.

[33] European Data Protection Supervisor. Opinion 5/2016. Preliminary EPDS Opinion on the review of the ePrivacy Directive (2002/58/EC), 2016. https://edps.europa.eu/data-protection/our-work/publications/opinions/eprivacy_en.pdf.

[34] European Data Protection Supervisor. EPDS recommendations on specific aspects of the proposed ePrivacy Regulation, 2017. https://edps.europa.eu/sites/edp/files/publication/17-10-05_edps_recommendations_on_ep_amendments_en.pdf.

[35] European Data Protection Supervisor. Opinion 6/2017. EPDS Opinion on the Proposal for a Regulation on Privacy and Electronic Communications (ePrivacy Regulation), 2017. https://edps.europa.eu/sites/edp/files/publication/17-04-24_eprivacy_en.pdf.

[36] European Data Protection Supervisor. Statement of the EDPB on the revision of the ePrivacy Regulation and its impact on the protection of individuals with regard to the privacy and confidentiality of their communications, 25 May 2018. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_statement_on_eprivacy_en.pdf.

[37] European Parliament. Report on the proposal for a regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications), 2018. http://www.europarl.europa.eu/sides/getDoc.do?type=REPORT&reference=A8-2017-0324&language=EN.

[38] European Parliament, EPRS. Reform of the e-Privacy Directive, Briefing legislation in progress, 2017. https://www.europarl.europa.eu/RegData/etudes/BRIE/2017/608661/EPRS_BRI(2017)608661_EN.pdf.

[39] Garante per la Protezione dei Dati Personali. Informativa e Consenso per l'uso dei Cookie, 2014. http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3585077.

[40] Garante per la Protezione dei Dati Personali. Simplified Arrangements to Provide Information and Obtain Consent Regarding Cookies, 2014. http://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/3167654.

[41] Ghostery. https://www.ghostery.com.

[42] Gonzalez, R., Jiang, L., Ahmed, M., Marciel, M., Cuevas, R., Metwalley, H., and Niccolini, S. The cookie recipe: Untangling the use of cookies in the wild. In *2017 Network Traffic Measurement and Analysis Conference (TMA)* (June 2017), pp. 1–9.

[43] Hoofnagle, C., Urban, J., and Li, S. Privacy and modern advertising. In *Proceedings of Amsterdam Privacy Conference* (2012).

[44] Interactive Advertising Bureau. Europe's cookie laws: e-Privacy Directive Implementation Center . https://www.iabeurope.eu/eucookielaws/.

[45] Koops, B.-J. The trouble with european data protection law. *International Data Privacy Law 4*, 4 (2014), 250–261.

[46] Leenes, R., and Kosta, E. Taming the cookie monster with dutch law–a tale of regulatory failure. *Computer Law & Security Review 31*, 3 (2015), 317–335.

[47] Leenes, R., van Lieshout, M., and Hoepman, J.-H. The Cookiewars: From regulatory failure to user empowerment? *M. van Lieshout, & J-H. Hoepman (Eds.), The Privacy & Identity Lab: 4 years later* (2015), 31–49.

[48] Legislative Train Schedule. Connected digital single market. Proposal for a regulation on privacy and electronic communications, 2017. https://www.europarl.europa.eu/legislative-train/theme-connected-digital-single-market/file-e-privacy-reform.

[49] Li, T.-C., Hang, H., Faloutsos, M., and Efstathopoulos, P. *TrackAdvisor: Taking Back Browsing Privacy from Third-Party Trackers*. Springer International Publishing, Cham, 2015, pp. 277–289.

[50] Marciel, M., Cuevas, R., Banchs, A., González, R., Traverso, S., Ahmed, M., and Azcorra, A. Understanding the detection of view fraud in video content portals. In *Proceedings of the 25th International Conference on World Wide Web* (Republic and Canton of Geneva, Switzerland, 2016), WWW '16, International World Wide Web Conferences Steering Committee, pp. 357–368.

[51] Markou, C. Behavioural Advertising and the New "EU Cookie Law" as a Victim of Business Resistance and a Lack of Official Determination. *Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection* (2016), 213–247.

[52] Pratical Law. Data protection global guide. https://uk.practicallaw.thomsonreuters.com/Browse/Home/International/DataProtectionGlobalGuide.

[53] REFIT Platform. REFIT Platform Opinion on the submission by the Danish Business Forum on the E Privacy

Directive and the current rules related to "cookies", 2016. http://ec.europa.eu/info/sites/info/files/opinion_comm_net.pdf.

[54] Roesner, F., Kohno, T., and Wetherall, D. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation* (Berkeley, CA, USA, 2012), NSDI'12, USENIX Association, pp. 12–12.

[55] Software is hard. Har 1.2 spec. http://www.softwareishard.com/blog/har-12-spec.

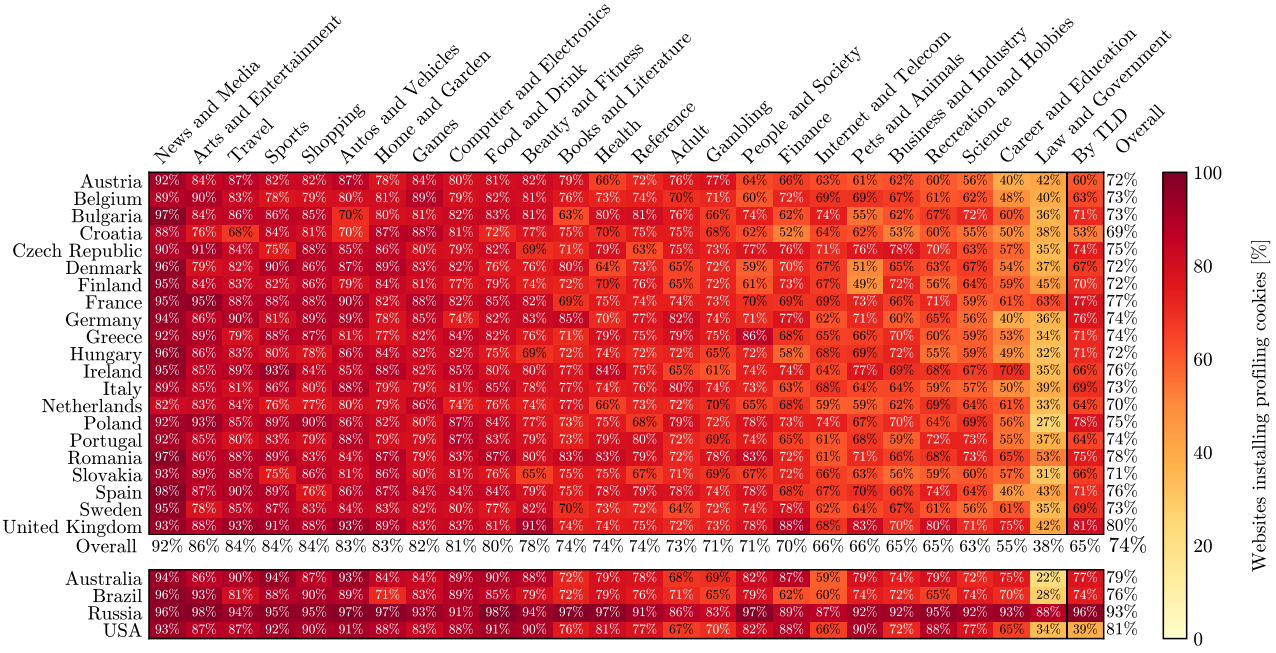
[56] The International Association of Privacy Professionals. EU law on cookies. https://iapp.org/media/pdf/resource_center/DLA_EU_cookie_implementation_9-14.pdf.

[57] Time.lex and Spark. ePrivacy Directive: assessment of transposition, effectiveness and compatibility with proposed Data Protection Regulation, 2013. https://ec.europa.eu/digital-single-market/en/news/eprivacy-directive-assessment-transposition-effectiveness-and-compatibility-proposed-data.

[58] Traverso, S., Trevisan, M., Giannantoni, L., Mellia, M., and Metwalley, H. Benchmark and comparison of tracker-blockers: Should you trust them? In *2017 Network Traffic Measurement and Analysis Conference (TMA)* (June 2017), pp. 1–9.

[59] Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., and Hennessy, M. Americans reject tailored advertising and three activities that enable it. Available at SSRN 1478214.

[60] Yen, T.-F., Xie, Y., Yu, F., Yu, R. P., and Abadi, M. Host Fingerprinting and Tracking on the Web: Privacy and Security Implications. In *Proceedings of the 2012 Network and Distributed System Security Symposium* (2012).



Appendix - Results with a less conservative definition of profiling cookie

In this paper, we employ a strict definition of profiling cookies. They are those installed by behavioral advertising platforms, and with a lifetime higher than one month. If results presented in subsection 5.4 provide a lower bound for ePrivacy Directive violations, here we want to provide a higher bound, relaxing our definition of profiling cookie.

Here, we consider violating the ePrivacy Directive those websites that install at least one third-party cookie when accessed the first time, regardless the lifetime and nature. The above figure shows violations across countries and website categories, and is identical to Figure 5 except for the more relaxed definition of profiling cookie. Overall violations grow from 46% to 74%, and the increase is rather uniform across rows and columns.