Can I Opt Out Yet?: GDPR and the Global Illusion of Cookie Control

Conference Paper · July 2019

DOI: 10.1145/3321705.3329806

7 authors, including:

Matteo Dell'Amico
EURECOM
56 PUBLICATIONS **864** CITATIONS
SEE PROFILE

Davide Balzarotti
EURECOM
110 PUBLICATIONS **5,276** CITATIONS
SEE PROFILE

Igor Santos
University of Deusto
114 PUBLICATIONS **2,152** CITATIONS
SEE PROFILE

Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control

Iskander Sanchez-Rola
University of Deusto
Symantec Research Labs

Matteo Dell'Amico
Symantec Research Labs

Platon Kotzias
IMDEA Software Institute
Univ. Politécnica de Madrid

Davide Balzarotti
EURECOM

Leyla Bilge
Symantec Research Labs

Pierre-Antoine Vervier
Symantec Research Labs

Igor Santos
University of Deusto

ABSTRACT

The European Union's (EU) General Data Protection Regulation (GDPR), in effect since May 2018, enforces strict limitations on handling users' personal data, hence impacting their activity tracking on the Web. In this study, we perform an evaluation of the tracking performed in 2,000 high-traffic websites, hosted both inside and outside of the EU.

We evaluate both the information presented to users and the actual tracking implemented through cookies; we find that the GDPR has impacted website behavior in a truly global way, both directly and indirectly: USA-based websites behave similarly to EU-based ones, while third-party opt-out services reduce the amount of tracking even for websites which do not put any effort in respecting the new law. On the other hand, we find that tracking remains ubiquitous. In particular, we found cookies that can identify users when visiting more than 90% of the websites in our dataset—and we also encountered a large number of websites that present deceiving information, making it it very difficult, if at all possible, for users to avoid being tracked.

KEYWORDS

user privacy; browser cookies; GDPR

ACM Reference Format:
Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. 2019. Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control. In *ACM Asia Conference on Computer and Communications Security (AsiaCCS '19), July 9–12, 2019, Auckland, New Zealand.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3321705.3329806

1 INTRODUCTION

On May 25th, 2018 the European Union's (EU) General Data Protection Regulation (GDPR) entered into effect. This resulted in users worldwide being flooded with emails about updated privacy terms, and in many websites starting to ask for explicit consent to collect and share user information; some even redirected users to a text-only version of their site or simply refused to serve content to anyone connecting from the European Union [22]. The reason for this is that the GDPR introduced sterner regulations on processing personal data, impacting the practice of user tracking which was performed in more than 90% of the highest traffic websites [13, 45].

Much of the economy behind the Web is moved by advertising: a huge, fast-growing industry estimated to be worth around 227 billion dollars in 2018 [35]. A cornerstone of the digital advertisement industry is personalization in the form of targeted ads, which increase the likelihood that users will follow them. Unfortunately, collecting the very data which is needed to personalize advertisements carries important privacy risks. In particular, tracking done through web cookies results in advertisers having access to a large part of user's browsing history—data that could reveal sensitive information, because it can lead to disclosing many kinds of confidential information such as medical conditions or political opinions.

To protect its residents' privacy, the European Union introduced a set of laws. In 2009, the ePrivacy Directive [15] required user consent before websites could track them through cookies, unless those cookies were strictly necessary to perform the required services. However, the implementation of this directive often resulted in simple pop-up messages that did not provide any real option to users. The definition of consent in the ePrivacy Directive referenced previous legislation [14] which was implemented differently among EU countries. The GDPR [16], which was approved in April 2016 and entered in force on May 25, 2018 (we discuss in more details the content of the GDPR in Section 2) raised the bar for consent, and in turn we have seen big changes in how companies manage cookie consent practices. Just a month after that, California also introduced a similar law, which will take effect in 2020 [30].

Researchers already started to investigate the effect of the GDPR in other independent studies. *Whotracks.me* [53] collected data on the tracking performed when users visit websites, and observed that the advertising market is recently shifting towards fewer larger companies. Degeling et al. [8] looked at how European websites have changed after the GDPR and noted an increment in the amount of information and control available to users in the EU. Dabrowski et al. [6] evaluated the persistent cookies set while accessing the same websites from the USA and from the EU, showing that visitors from the EU are less likely to receive persistent cookies, and that the number of persistent cookies set for visitors from the USA appears to have diminished as well after the introduction of the GDPR. While these studies show that the GDPR is already changing the tracking panorama, it is still unclear what the final effect is for the

AsiaCCS '19, July 9–12, 2019, Auckland, New Zealand

ACM ISBN 978-1-4503-6752-3/19/07…$15.00
https://doi.org/10.1145/3321705.3329806

end users and whether, and to which extent, it is now possible for European citizens to effectively avoid being tracked. Moreover, it is also interesting to understand the actual "boundaries" of the effects of this regulation and whether the required privacy controls are different between the EU and the rest of the world.

To answer the previous questions for cookie-based tracking, we performed an extensive manual analysis of 2,000 popular websites across the world, belonging to several categories. We visited each site and tried to refuse tracking every time we could, while a custom browser plugin collected all cookies set by those websites (both before and after our opt-out attempts), as well as the information presented to the user and the privacy policies and the privacy controls available for expressing an informed consent. Note that our goal here is not to assess whether these websites comply with legislation, but rather to understand the influence of the GDPR on the privacy of Internet users in the context of web tracking. In particular, we aim at measuring how easy it is to opt-out from web tracking if the user desires to do so and assessing whether it is possible at all.

Our results show that tracking is prevalent, happens mostly without user's consent, and opt-out is difficult. Most websites perform some form of tracking, and 92% of them do it before providing any notice to the user. Only 4% of the websites provide a clear opt-out option in their cookie notice, but even when they do opt-out is mostly ineffective—only 2.5% of the websites erase some cookies.

While tracking is still ubiquitous, we also find that the impact of the new regulation is largely perceivable globally, in terms of privacy controls and information presented to users: in particular, USA-based websites have a behavior which is on aggregate similar to the one of EU-based ones. While sites in other countries appear to be less influenced by the regulation, they are still often impacted indirectly: opting out of tracking through third parties reduces the amount of tracking. Within each website category we observe a similar degree of tracking, with a few exceptions that deviate from the typical behavior of commercial websites.

In summary, our main contributions are the following:

- We perform a global analysis of how websites handle the new European data protection rules. We check both how the user is informed about cookie tracking, and the actual behavior in terms of tracking cookies installed.
- We discover that the GDPR has a global effect on a very large number of websites, either directly or indirectly. In particular, we found similar behavior in the EU and the USA.
- We discover that, often, opting out is not properly implemented and most websites end up tracking users with long-lasting cookies despite them having opted out.

2 BACKGROUND

In this section we provide a short introduction to HTTP cookies and we discuss their relation with the GDPR. For a more in-depth discussion, we refer the reader to the document by Cookiebot [5].

2.1 HTTP Cookies

Cookies [36] were first introduced by Netscape in 1994 to enable stateful browsing over the stateless HTTP protocol. They allow small pieces of data created by a web application to be stored on web browsers. For example, they can include information on user authentication, preferences, and keep track of session identifiers used to enable complex workflows (e.g., "shopping carts"). Each cookie is characterized by a name, a value, a URL it is associated with, and an expiration date. Whenever a browser issues an HTTP(S) request, all cookies mapped to a configurable prefix of the target URL are sent to the corresponding web server.

On top of their original goal of providing stateful navigation, nowadays cookies are routinely used to track users *across* different websites, most often for advertising and analytics [13, 45]. This is possible because websites often include resources provided by third-party domains such as advertisement companies. Upon connection to the website, these third-parties can set uniquely identifiable cookies on the visitor's computer; they will then be able to track these identifiers across multiple associated websites, thus reconstructing part of the user's online activity.

Cookies are the most widespread way of tracking web visitors, but they are not the only one. Other kinds of tracking, which are outside the scope of this paper, are discussed in Section 8.

2.2 The GDPR

EU legislation regarding privacy consists of several documents which interacts in complex ways between them and with national laws. With the goal of describing the context rather than aiming for exhaustivity or giving legal guidance, in the following we outline some of the content that is most closely related to cookie tracking.

The GDPR [16] is an EU legislation (which came into force on May 25, 2018) that regulates the handling of personal data with the goal to *"strengthen individuals' fundamental rights in the digital age and facilitate business by clarifying rules for companies and public bodies in the digital single market"* [4]. It is widely regarded as having a major impact on the world's corporations, with provisions for very large fines (up to 20 million Euros or 4% of a company's annual global turnover, whichever is the largest) and costs to comply that are estimated in the order of billions of dollars [25]. While the GDPR was introduced in the EU, its impact can extend to the entire world as the legislation has extra-territorial scope, meaning it applies to entities outside the EU processing the personal data of EU residents (*data subjects)* while offering them goods and services or monitoring their behavior *"regardless of whether the processing takes place in the Union or not"* (Article 3(1); for more in-depth discussion on the GDPR's jurisdiction see the document by Wiley Rein [54]). It is worth noting that this legislation will still apply in the United Kingdom if it exits from the EU, as an equivalent legislation has been adopted into its national law [31].

The GDPR applies to personal data defined as *"any information concerning an identified or identifiable natural person,"* including pseudonymous data *"which could be attributed to a person by the use of additional information,"* i.e., by taking into consideration methods that could be used to de-anonymize pseudonymous data [39, 55]. Hence, the legislation applies when identification is *possible*, not only when it is actually performed. Cookies are expicitly mentioned: *"natural persons may be associated with online identifiers provided by their devices, applications, tools and protocols, such as internet protocol addresses, cookie identifiers or other identifiers [...]. This may leave traces which, in particular when combined with unique*

identifiers and other information received by the servers, may be used to create profiles of the natural persons and identify them" (Recital 30). Hence, the GDPR does not differentiate between first- and third-party cookies, and it is not hard to see how the unique user identifiers that are present in many cookies—even session cookies, which only last for a single browser session—can be used to match personal information with the natural persons owning them.

Personal data in identifiable form should be retained for no longer than necessary (for the purpose for which they were processed): controllers have the duty of *"ensuring that the period for which the personal data are stored is limited to a strict minimum. [...]. In order to ensure that the personal data are not kept longer than necessary, time limits should be established by the controller for erasure or for a periodic review."* For the purpose of processing (and any associated legal or statutory requirements), various practitioners [5, 11, 43] use a threshold of 12 months.

User Consent. The ePrivacy Directive [15] required user consent to cookie tracking, with exceptions for cases required by law or strictly necessary ones (e.g., memorizing a user's log-in credentials or shopping cart). To comply with this, websites often included small banners containing just an `OK` button to accept all cookies, or notes such as *"if you continue browsing, we will consider that you accept cookies."* Conversely, the GDPR clearly states that consent cannot be implicit as it "*should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her [...]. Silence, pre-ticked boxes or inactivity should not therefore constitute consent.*" EU law also contemplates browser settings as a way to handle user consent, but this is still not generally implemented in practice.

Consent should be given *prior* to data processing: websites cannot create non-strictly necessary cookies and then delete them if the user rejects them, because cookies may have already been sent to the server if any request was performed after cookie creation; this could be implemented with a form of script blocking until users give consent.

Finally, in line with previous legislation, the GDPR states that websites cannot refuse serving users who do not agree to tracking for purposes that are not essential to the functioning of the website: *"Consent should not be regarded as freely given if the data subject has no genuine or free choice or is unable to refuse or withdraw consent without detriment."*

Policies. Privacy and cookie policies or notices are the typical way of disclosing information on data processing, and they should be transparent and expressed in a clear language: *"It should be transparent to natural persons that personal data concerning them are collected, used, consulted or otherwise processed and to what extent the personal data are or will be processed. The principle of transparency requires that information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used."* Privacy policies, moreover, should be tailored to the specific website they apply to rather than being just taken from a template or from automated generators: *"the specific purposes for which personal data are processed should be explicit and legitimate and determined at the time of the collection of the personal data."*

	EU	USA	China	Others
Business	18%	52%	8%	21%
Entertainment	24%	53%	5%	18%
Finance	30%	27%	10%	33%
Food	27%	49%	10%	14%
Gaming	28%	53%	8%	11%
Government	28%	32%	–	40%
Health	18%	57%	16%	9%
Hobby	21%	56%	7%	16%
Kids	35%	57%	1%	7%
News	25%	35%	11%	29%
Politics	15%	65%	–	20%
Pornography	31%	52%	1%	16%
Real Estate	26%	37%	12%	25%
Religion	12%	63%	–	25%
Science	35%	47%	7%	11%
Shopping	35%	30%	10%	25%
Sports	48%	25%	1%	26%
Technology	11%	66%	9%	14%
Travel	37%	36%	9%	18%
Weapons	15%	72%	1%	12%

Table 1: Websites analyzed, by region and category.

3 METHODOLOGY

The goal of our study is to verify to what extent, after the GDPR went in force, users in the EU are able to control cookie tracking: evaluating whether a website's behavior complies with the law is outside the scope of our work. It is important to note that since the effects of the GDPR extend beyond the EU borders, we perform a *global* measurement to capture the ability of users to opt-out from tracking on a worldwide scale. If the top-level domain (TLD) of a website corresponds to a country, we attribute that website to the country; otherwise, for international TLDs, we use IP geolocation via `ip-api.com` to determine the country. We are aware that IP geolocation sometimes returns incorrect results [51]. We consider efforts to improve this method's precision as outside the scope of our work.

For each website we verify how much tracking is performed when users *never* consented to it, and rejected it when given the option. We also analyze how difficult it is for users to opt out and what kind of interface and choices websites offer in this respect.

By design, our study cannot verify whether cookies are actually used to profile users; however, as discussed in Section 2.2, the GDPR applies as soon as cookies can *potentially* be used to identify users, even in the—arguably unlikely—case of websites that set cookie identifiers without using them later.

3.1 Domain Selection

We used the list of Alexa top-1M websites [47], divided in fine-grained categories returned by Symantec Rulespace [48]. We then chose 20 of the most highly accessed categories; from each, we

collected the top 100 entries in the Alexa list. All together, this resulted in a list of 2,000 websites: Table 1 on the preceding page contains the list of categories and a breakdown by region. Of these 2,000 websites, 474 belong to the top 1K websites for worldwide traffic according to the Alexa ranking, 1,228 are ranked in the top 5K, and the remaining rank outside these boundaries.

3.2 Data Collection

We performed our data collection using a manual approach as opposed to an automated one, since we do not know of any automated approach that would reliably navigate websites rejecting tracking, and find and categorize cookie and privacy information.

Between July 6th and 30th 2018, we browsed the selected websites from EU IP addresses located in France, Ireland and Spain, recording which options they provided for their privacy and which cookies they created on the user's browser. We collected this information by filling a multi-step questionnaire that we implemented as a Chrome browser extension. The dataset was split evenly and each site was randomly assigned to one of the authors only after its exact interpretation—described in the following—was agreed and clarified in multiple in-person meetings. With our extension in place, we manually visited each website described above and, for each of them, recorded a number of features associated to the following five phases:

(1) **Loading**. Before visiting each website, the extension deletes all cookies stored in the browser. Once the website is loaded, and before any interaction with it, the extension automatically saves the lists of newly created cookies and the list of fetched resources.

(2) **Cookie Notice**. We manually classify the size and content of cookie notices. For size, we distinguish whether they are banners that allow navigation while they are present or "blocking" UI elements that preclude browsing until they are dismissed. The categories for content are: a) `Anyway` for notices that just inform users that they will be tracked; b) `AutoAccept` for notices informing that, by continuing the visit to the website, users accept any cookies that it will set; c) `OnlyAccept` for notices having an accept button, but no quick way to reject tracking (a way to reject cookies is sometimes present through a link at the cookie settings dialogues); d) `AcceptReject` for notices that provide users with both an accept and a reject button; e) `JustSettings` for cases that lead users straight in a more complex settings dialog.

(3) **Cookie Settings and Policy**. We manually browse to the cookie settings and privacy policy; we collect the full HTML of both and save it for later analysis. We also classify the kind of options available in the settings (e.g., to allow to reject by types of tracking purposes or to individually control tracking from each company).

(4) **Rejection**. We reject all tracking whenever possible, to detect if websites actually comply with user choice. In this step, we did not visit third-party services for opting out, which we analyzed as described in the following.

(5) **Reload**. We reload the website to see the difference, if any, after the user explicitly rejected all tracking; our plugin saves again the list of cookies created and resources fetched. Finally, we check whether a cookie notes appears again and, if so, we note its type as described above.

Cookie value	$\log_{10}$(strength)	Ent
`__test`	4.29	1.91
`12198692`	6.06	2.25
`W1TMYg**`	8.00	3.46
`6RqV3mB2nac`	**10.17**	3.45
`drZshFnT6g1M670owknOIA`	**22.00**	4.27
`2ee92b9f-2781-43a2-a326-af9cc922a942`	**35.67**	3.45

Table 2: Example cookie values and their "strength" estimated by `zxcvbn`. We conservatively consider identifiers those with a strength whose base-10 logarithm is at least 9.

As discussed in Section 2.2, EU legislation gives exceptions to the tracking requirement for cases where this is strictly necessary. We remark that these cases generally covers tracking information such as user logins of shopping carts—none of which is needed for the use case discussed above. Hence, we consider that the number of strictly necessary tracking cookies we encounter in our study is likely to be negligible.

3.3 Third-Party Services

By analyzing cookie-related banners, settings, and policies, we repeatedly observed the presence of four popular third-party services used to opt-out from multiple tracking companies at once: *aboutads* [58], *youronlinechoices* [57], *networkadvertising* [38], and *TRUSTe* [49]. These services handle the whole process of opting in/out from third-party trackers and save the decision on cookies. We analyzed which companies are involved with each of them opting out from all the services and saving all the opt-out cookies created in the browser. In Section 4, we analyze how these services affect the real deployment of tracking cookies and how many websites link to them in order to opt out from tracking.

3.4 Recognizing Identifiers

The GDPR applies to data that can identify users, whether or not they are meant or used to actually track them. While many cookies are personal identifiers, not all are: for example, some record preferences that are too coarse to identify users. While several cookies carrying little information can sometimes be combined to uniquely identify users, we take a conservative approach to avoid overestimating tracking. Other information in the cookie, such as for example the cookie name, could in principle be used to identify the user; however, in practice cookie names tend to be the same for all users. For this reason, again following a conservative approach, we ignore this information.

A traditional approach to determine whether cookie values carry enough information to identify users would be to measure their entropy—an information-theoretical approach that considers longer strings with several different characters as carrying more information. In our case, though, the entropy calculation may be misleading because it would attribute an excessive weight to common strings, such as dictionary words. Hence, we take inspiration from the literature on passwords [9, 10, 34, 52], which measures a password's strength—i.e., the amount of information it carries—as the number

Region	Tracking	Long-lasting ids	3rd-party opt-out link	Browser instructions	Cookie settings	Third-party notice
EU	92.3%	81.7%	32.0%	19.6%	17.7%	3.8%
USA	91.7%	80.1%	23.6%	6.6%	11.5%	4.7%
China	90.5%	82.5%	–	–	–	–
Others	92.6%	79.5%	8.5%	3.3%	3.3%	0.5%

Table 3: Statistics about tracking, cookie user interface and information, broken down by region.

of guesses it would take an attacker to guess it. We use `zxcvbn` [52]—an approach that conservatively estimates a password's strength by evaluating the worst-case scenario of an attacker choosing the best approach from a set of possible ones—as our estimator; in Table 2 we show a few examples of cookie values and their strength output by `zcxvbn`. The Table also shows the entropy values even though, as expected, they do not provide a meaningful indication of the randomness of those cookies. We conservatively consider as identifiers cookies whose strength as a password is considered to be greater than 10^9 by `zxcvbn`, considering that they could be used to distinguish among at least 1 billion users. We remark that this approach behaves reasonably even if cookies store plain-text personal information: for example, a name that is common enough to be shared by several people and hence not useful to identify a single individual (e.g., "Bob Smith") falls below the 10^9 strength threshold, while one that is less likely to have omonyms (e.g., "Robert J. Smith-Johnson") has a score that makes us consider it a unique identifier. As timestamps are often stored in cookies, we ignore all parts of a cookie that contain plain-text Unix timestamps in the period of our experiment. Again conservatively, we do not consider as identifiers cookie values that have been observed more than once in our whole dataset. After applying these rules, we found that 84% of the unique cookies we collect in our experiments were complex enough to identify users, resulting in 54,803 unique cookies that we consider identifiers.

For each identifier cookie we checked if its domain is a known third-party tracker (e.g., analytics and advertisement), using the list that Firefox uses for its tracking filter [37]. Our measurements in Section 4 show that this database is quite effective in categorizing third-parties in the USA and EU, but does not include many trackers we found in Chinese websites.

4 DATA ANALYSIS

We first discuss our results through general statistics using the whole data; we then breakdown the results by region, and finally by website categories.

4.1 General Findings

We begin our discussion by highlighting a number of important points that apply to the entire dataset, independently from the geographic location or website category.

Most web pages track users even if they do not give their consent. Figure 1 shows that around 92% of the websites we considered perform some form of tracking—i.e., they set at least one identifier cookie (as discussed later, more than 80% of websites set cookies that last a year or more). This happens even before

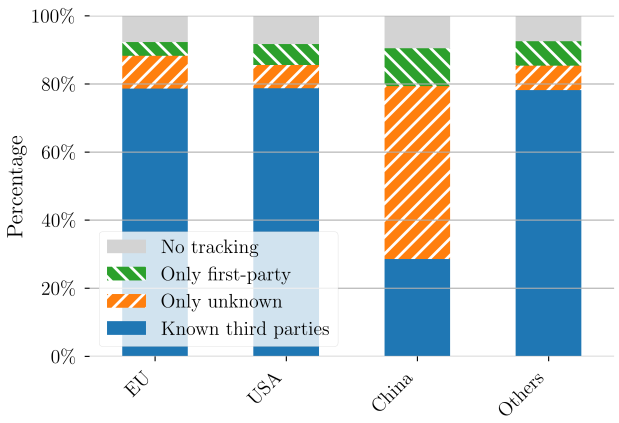
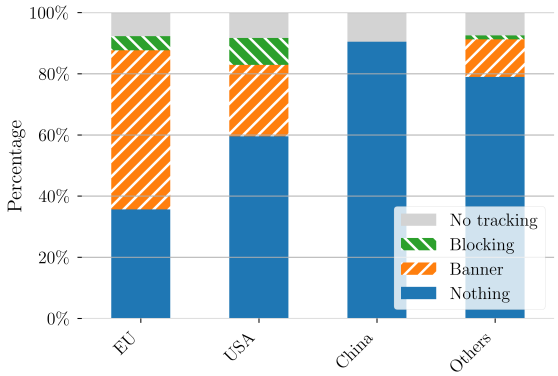


Figure 1: Kind of tracking observed.

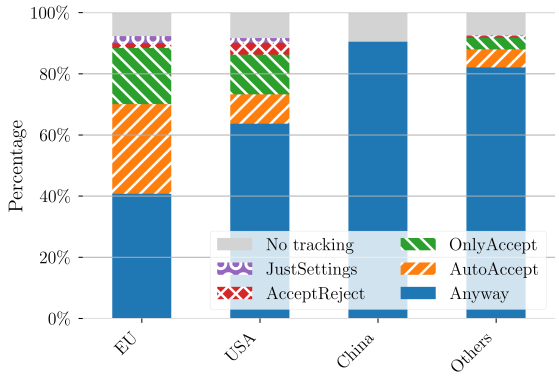
showing any banner about cookie policies, and even if the user chooses to opt-out from being tracked. Additionally, we observe that the vast majority of websites perform tracking using known third-parties services (generally for advertisements and analytics), and only 12% of domains rely uniquely on first-party tracking. Note that, as discussed in Section 2, the GDPR does not differentiate between first-party and third-party tracking; hence, users should have the right to reject both types of tracking. We also see, as discussed in Section 3.4, that the tracker database we use does not include many third parties observed in Chinese websites (i.e., "only unknown" third-party domains).

Few websites provide an easy way to opt out from tracking. In Figure 2b on the following page we break down the content of the cookie notice, if present. In some cases, users have no choice (i.e., the notice just mentions that cookies will be set); in others, they are informed that continuing navigation on the website will result in setting cookies (`AutoAccept`); some websites just have an "accept" button without a "reject" one (`OnlyAccept`). Unfortunately, the cases where users have a clear "reject" option (`AcceptReject`) or are presented right away with a cookie settings dialog (`JustSettings`) are, together, less than 4%.

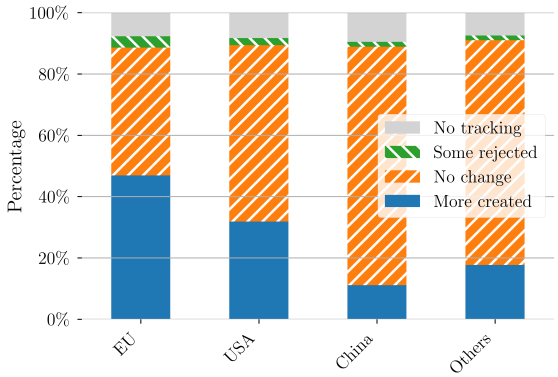
Rejecting tracking is often ineffective. In Figure 2c we compare the number of cookies before and after the user rejected them (if the interface gave an option to do so) and reloaded the website. In most cases, the number of cookies set by the server remains the same or even increases. The cases where some cookies are erased after refusing them are, on average, 2.5% of the whole set of



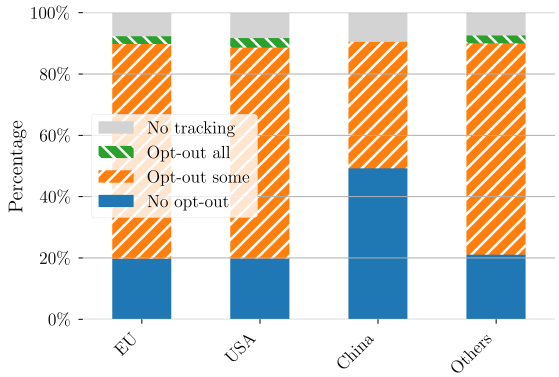
(a) Cookie notice size. While a relevant number of websites in EU and USA show cookie notices, no website we analyze hosted in China does.



(b) Type of cookie notice. Most websites leave users no choice, inform that cookies will be set if navigation continues ("AutoAccept") or give only the possibility to accept tracking ("OnlyAccept.")



(c) Number of cookie identifiers after rejecting them, if possible, and reloading the website.



(d) Number of cookie identifiers cookies after opting out from external websites.

Figure 2: Statistics about cookie user interfaces and tracking, broken down by region.

websites. The fact that *more* cookies can be created after rejecting them may appear counter-intuitive at first. However, this can be due to additional elements (e.g., ads) fetched when reloading the website, or because the entire page content was not loaded in the first place (e.g., because a blocking notice prevented to fully load the underlying webpage). In Section 6, we show the details of a case where more cookies are loaded after rejecting cookies.

Few websites have cookie settings. In Table 3 on the previous page, we observe that only 16% of EU websites and 12% of USA websites have a cookie settings interface. Numbers are even smaller for the other regions; for instance, none of the Chinese websites in our dataset had a cookie settings interface. We will discuss regional differences in detail in Section 4.2.

9 websites out of 10 create long-lasting identifiers. Even if the GDPR requires storing personal data for a minimum amount of time, this seems to be rarely taken into account. Table 3 shows that around 90% of websites create identifiers lasting more than 12 months—the threshold we discussed in Section 2.

Third-party cookie notices are not very common. There are some third-party services that handle the cookie notice and, if present, the cookie settings. Therefore, we checked how many websites used services provided by the main third parties: OneTrust [42], TRUSTe [49], Quantcast [44], Cookiebot [5], and Evidon [17]. Table 3 shows that only 5% of USA websites use these services and the number is even lower in the European Union (3%). These services are practically not used at all in the rest of the world.

Opting out through external services does not prevent all tracking. In Figure 2d we show that the number of identifying cookies varies when the user previously opts out through one of the third-party opt-out services included in Section 3.3. However, while the number of identifier cookies often decreases ("opt-out some"), the case in which all identifiers are blocked ("opt-out all") is lower than 3%.

The general picture we can draw from these results is that the overwhelming majority of the websites we visited track users even if, during our manual crawling, we never consented to cookies. This is not to say that GDPR had no effect: in fact, information about

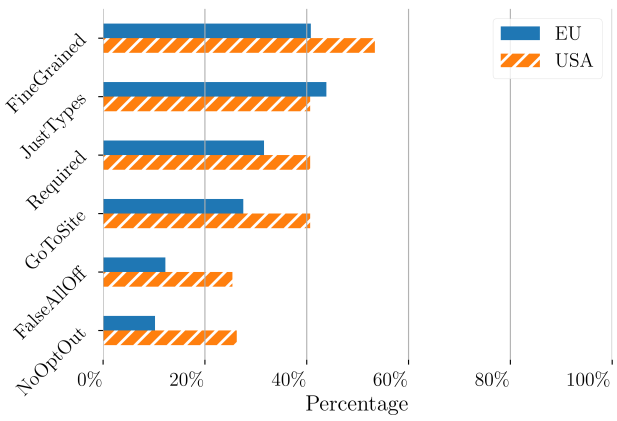


Figure 3: Characteristics of the settings dialogs.

cookies is present in 33% of the websites, and it is often possible for the user to reject at least *some* of the tracking. However, from our tests it is hard to conclude if this is done on purpose or, as we will discuss in more details in Section 6, if this is a consequence of the challenge websites encounter when they want to prevent third-party tracking.

4.2 Regional Differences

We observe that the EU- and USA-based websites appear more influenced by the GDPR regulation; EU citizens that visit websites located in other parts of the world, on the other hand, may expect less in terms of privacy. Most of the websites we analyze are located in the European Union, USA, or China; we therefore group them in these three areas, plus one group referring to all other countries.

US-based websites look almost as impacted by the GDPR as the EU websites do. Throughout all of the information we present, it is apparent that websites in the USA appear to approach cookie regulations similarly to the EU. From Table 3 we see that the number of websites providing third-party opt-out links and cookie settings dialogs are similar between EU and USA; Figure 2a shows that cookie notices appear in 32% of US-based websites against the 57% of EU while the more intrusive "blocking" dialog is even more frequent in the USA than in the EU. Also Figure 2b shows that, even if cookie notices give more often control to the users in EU websites, the percentage of US websites that provide cookie control is still around 30%.

China-based websites do not seem to have GDPR-induced modifications, but they are still impacted by it. From Table 3 and Figures 2a and 2b we see that Chinese websites, unlike many US and EU-based ones, provide very little cookie control and information; yet, the result of Figure 2d is interesting because it proves that, if one opts-out from the global third-party services described in Section 3.3, they will be tracked less also when browsing 41% of Chinese websites.

Cookie settings in the USA make opting out more difficult. In Figure 3 we show the types of and prevalence of cookie settings

Category	1st-party	Unknown	Analytics	Ads	Content
All	76%	70%	70%	62%	41%
Pornography	69%	72%	**30%**	**39%**	**11%**
Weapons	79%	64%	62%	**45%**	38%
Religion	69%	50%	71%	58%	37%
Government	60%	**46%**	**41%**	**29%**	**17%**

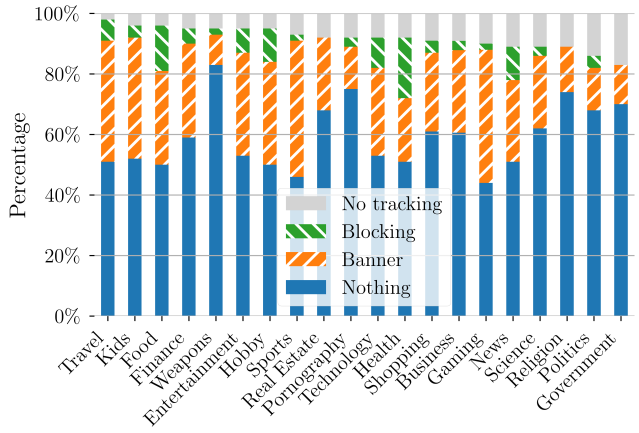
Table 4: Kind of tracking for particular categories.

dialogs we observed. When users are allowed to opt-out from each of the ad providers or third-party trackers individually, we flag the cookies settings to be "FineGrained." In some cases ("JustTypes"), to simplify the user's job while opting out, trackers are grouped by categories (e.g., advertisement, analytics, etc.). In other more desirable cases, the cookie settings interface gives both options. Often, there are so-called essential cookies that are required by the websites for functioning properly ("Required"). When opting out requires going to an external website, we set the "GoToSite" flag. Sometimes, there are shortcuts for opting out from "all" tracking, however, when we expand the list, we observe that a significant number of third party services listed under the category does not allow the opt-out: we flag this as "FalseAllOff". We also flag cases when opt out must be performed by going to external websites ("GoToSite"), and cookies for which a way of opting out is just not provided ("NoOptOut"). Our results of Figure 3 show that cookie settings dialogs in the US are more complex, and that they generally require more work to opt out from all cookies.

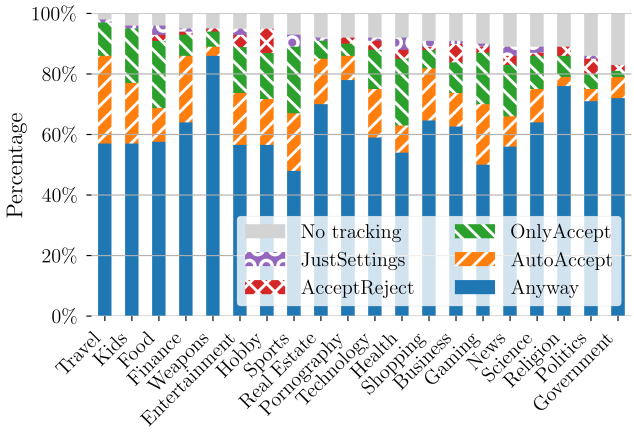
Other Remarks. Cookie policies in the EU tend to give instructions on how to delete cookies from a user's browser more often than other websites ("browser instructions" in Table 3 corresponds to this); the results from the "other countries" aggregated category are, unsurprisingly, generally in the middle between the results of the USA and those of China. We do not breakdown these results into more fine grained regions because the number of websites would be too small to be statistically significant and representative. However, we observed that countries that are economically and/or geographically closer to EU/USA/China have similar characteristics to the neighboring region. For example, for European Economic Area countries like Switzerland or Norway we observe characteristics similar to the EU.

4.3 Categories

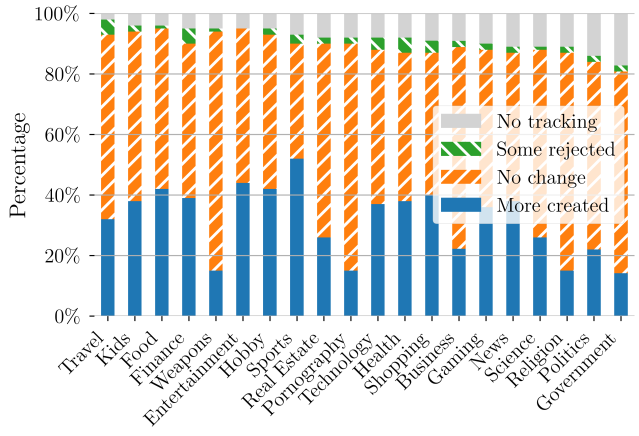
We now compare how GPDR is adopted by different website categories. Figure 4 on the next page depicts details about type of cookie notice, cookie notice size, what happens after the user rejects tracking from the website and external websites for the 20 site categories of our analysis. Categories are sorted by the percentage performing tracking–i.e., those on the left have the highest percentage of websites that do tracking, while those on the right have the least. Most of the categories do not deviate much from the average behavior of all websites; we still observe that categories catering to more general audiences tend to give more information and control to the users (in particular, see Sports and Gaming in Figure 4a and 4b). We remark that we avoid breaking down our results by location *and*



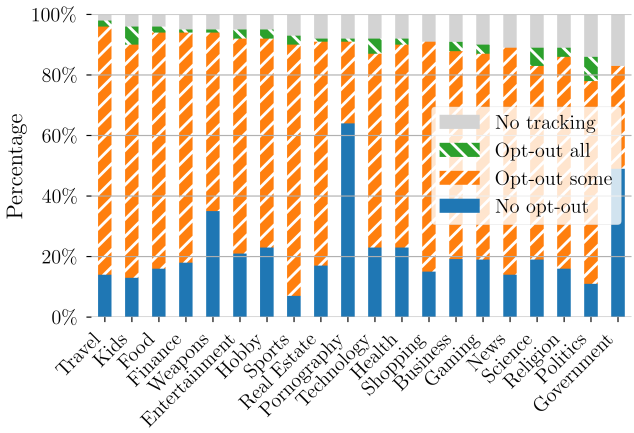
(a) Cookie notice size.



(b) Type of cookie notice.



(c) Cookie identifiers after rejecting tracking (if possible) and reloading. "No tracking" refers to websites that performed no tracking even *before* rejecting cookies.



(d) Cookie identifiers after opting out from external websites.

Figure 4: Statistics about cookie user interfaces and tracking, broken down by categories.

category together because numbers would become low enough to impact the statistical significance of our results. In the following, we highlight some categories whose results significantly deviate from the mean; a breakdown of the kind of tracking we encounter on them is provided in Table 4 on the preceding page.

Pornography. Adult websites appear to behave very differently from others. While there is a considerable amount of websites in the category that perform some type of tracking, only less than 20% of them displays cookie notices (Figures 4a and 4b). Also, opting out from the services described in Section 3.3 reduces tracking only in less than 30% of the websites (Figure 4d). We think this happens because this category has its own ecosystem and advertising networks [56], which appears to be quite segregated from the rest of the Internet. In this industry, the adoption of the GDPR rules appears to be progressing more slowly than on the rest of the Internet.

Weapons. This category has the most websites that perform tracking without informing users about it (over 80%). This might be because these sites are mostly hosted in the USA, and—unlike other categories—appears to be directed mostly to a local audience rather than EU citizens; hence, they may be outside the scope of the GDPR (see also the discussion on extra-territorial scope in Section 2.2).

Religion. Over 70% of the websites in this category perform tracking but do not inform users. A possible interpretation is that some of these websites may have been designed and implemented by volunteers rather than professionals; this could explain a slower adoption of the GDPR rules.

Government. Government pages are also largely not commercial, and this may explain why this is the category for which we observe the least tracking. Cookies could result from technical choices (e.g., including libraries, scripts or videos) rather than commercial ones; we hypothesize that in some cases website administrators might be

Region	Policies	FRES	FKRL	Length (words)
EU	190	57.1	10.5	2,261
USA	617	54.5	11.2	2,698
All	849	54.1	11.1	2,250

Table 5: Average readability scores and length of privacy policies per region. High FRES and low FKRL scores indicate easy texts.

unaware of some of the tracking done by their systems or of the relative consent requirements: this could explain the relative lack of information and/or control on cookie tracking.

5 READABILITY OF PRIVACY POLICIES

Users often need to consult the cookie and privacy policies of the websites they visit to understand how their data is processed, to opt-out from tracking, or even to get information on how to change their browser configuration (e.g., disabling third party cookies). In fact, 59% of the websites in our study do not prompt their visitors with any privacy banner and they present the available options only in the policy page. Thus, it's critical that policies are readable and understandable, a requirement included in the GDPR (see Section 2.2).

We measure the length and the readability of privacy policies using two simple metrics, the Flesch Reading Ease Score (FRES) [18] and the Flesch-Kincaid Reading Level (FKRL) [27]. Both metrics estimate the text complexity using the average sentence length and number of syllables per word but they differ on the weighting factors they use. FRES emphasizes the word length and outputs a score up to 121.22, that is the easiest to understand possible text; it does not have a theoretical lower bound. For example, *Time* magazine has an FRES score around 52, and *Harvard Law Review* scores in the low 30s. On the other hand, the FKRL index puts more emphasis on the sentence length, it has a lower bound of -1.3 and no theoretical upper bound. The output of this test can be interpreted either as a US grade level or as the years of formal education required to understand a text (for scores higher than 10). Despite the simplicity of these formulas and the known caveats [23], both metrics are standards used by legislators and government agencies. For example, US law requires insurance policies to have an FRES score above 45 [19]. Prior work used these tests to measure the complexity of privacy policies [24] and End User License Agreements (EULAs) [21].

We perform the readability analysis for the privacy policies of 849 websites written in English; 617 are from USA, 190 from the EU and 42 from the rest of the world. Table 5 shows, for each region, the average FRES and FKRL scores as well as the average policy length. The FRES scores average at 54.1 with a standard deviation of 9.3—minimum and maximum scores corresponding to 14,7 and 104.3. The easiest and hardest to read policies are both from websites in the USA: the easiest belongs to a government website, and the two most difficult, which belong to the same IT company, are for online communities. Documents for the general public should ideally score between 60 and 70 [21].

In prior work, Jensen and Potts [24] obtained a an average FRES score of 34.2 for 64 privacy policies in 2004, while in 2007 Grossklags and Good [21] obtained an average FRES score of 35.7 for 50 EULAs. Both used datasets that are orders of magnitude smaller than ours and performed their measurements more than a decade ago. Our results show an improvement in readability over the years.

FKRL scores exhibit a similar improvement. We obtain an average score of 11.1; therefore, a reader with at least 11 years of formal education should be able to understand these documents. Since the average score of privacy policies a decade ago was 14.2 [24], the improvement is obvious. Despite the improvement on both metrics, the complexity of privacy policies still remains high and outside the ideal range (FRES: 60–70; FKRL: 8–9) to be understandable by the general public. A comparison among regions shows that today EU and US privacy policies have a similar complexity, with the European ones being slightly more readable.

We also measure the length of privacy policies: on average, they consist of 2,250 words or 10 pages of double-spaced text. By considering the amount of websites a user may visit every day, we believe they are overly lengthy. Their size is very similar to the size of EULAs [21], which unfortunately users tend to regularly ignore [20].

Finally, we searched the content of privacy policies for mentions of the GDPR. Over 84% do not mention it, and 58% of those websites do not prompt users with any privacy banner; while neither is required to comply with regulation, both are possible indicators of websites that haven't made GDPR-related changes. From the 16% (136) of policies that mention the GDPR, 94 belong to US websites (11% of all policies from the US), 39 to EU (21% of the EU ones), and three from the rest of the world (7%). From the three policies outside US and EU, two belong to media conglomerates in India, and one to an Australian bank.

6 CASE STUDIES

In this section, we first (Section 6.1) discuss a couple of interesting cases, to highlight some particular and perhaps surprising results that we observed in our evaluation. Then (Section 6.2) we present in more detail various third-party opt-out services. Finally, (Section 6.3) we discuss some concrete examples, highlighting those that we consider good, bad, and misleading.

6.1 False Rejections

As shown in Section 4, most websites do not let visitors avoid tracking. In fact, more than 90% create tracking cookies immediately after loading the website, even before users can take any decision on tracking. We present the cases of two large websites implementing cookie control incorrectly, either on purpose or as a consequence of a bug in their code.

Rejecting cookies does not impact tracking. The website of a major company in the food and beverages business has a banner on the bottom of the home page stating *"[COMPANY NAME], 'we', envisages to use certain categories of cookies for several reasons, but need your agreement to do so"* and *"We will not use any categories of cookies other than those that are strictly necessary for the website to function if you do not elect to opt-in to those categories of cookies."* While visitors would think that they could opt out from tracking, this is not the case. In reality, all the tracking cookies are created before any consent is given, and the answer to the banner (i.e.,

accept or reject) does not affect the already present cookies. In either case, the website just creates a new cookie with the name/key `OptanonAlertBoxClosed` and the current date as value. In addition to this new cookie, the website continues to keep the tracking cookies from companies such as Gigya, Google or Adobe. For instance, the website includes the `ucid` cookie from Gigya, described by them as a *"Unique computer identifier used for generating reports, and used by the Web SDK to get saved response."* This cookie has a time-to-live of over a year.

More cookies created after rejecting tracking. A major website directed to software developers includes a blocking banner that only gives access to the website after accepting or rejecting tracking cookies. The banner indicates *"Click 'I Accept' below to consent to the use of this technology across the web,"* after explaining that cookies are used for advertisement and analytics. Before any decision is taken, a cookie named `VISITOR` with a unique user identifier is created. If the user rejects cookies, the `VISITOR` cookie won't be deleted, and several new tracking cookies from third-party advertisers are created—one of them described in their website as *"one of the main advertising cookies"* of the company. In fact, the website implements a script blocker: some cookies were never created before any choice was made, but not all the tracking scripts were suspended.

6.2 Third-party Opt out

In Section 4, we discussed four third-party opt-out services (*aboutads*, *youronlinechoices*, *networkadvertising*, and *TRUSTe*) designed to help websites and their clients opt-out from third-party tracking. We tested them for three days in a row and measured the time it took to reject cookies, and the number of companies from which we could successfully opt out from the list they offered.

We observed that these websites were rather slow, making external requests that require between 2 and 5 minutes for each service—i.e., opting out from all the four requires 8–20 minutes. Each of these services claims to let users opt out from 90–244 third party trackers; however, several of them returned errors, and on average we were only able to opt out from 85% of the "supported" services.

6.3 UIs: the Good, the Bad and the Ugly

Here we present a detailed, manual analysis of three popular websites and their user consent notices.

The Good. When accessing an important technology website, a banner that blocks the interaction with the website until a choice is made appears. This notice includes three options: "I ACCEPT," "I DO NOT ACCEPT" and "More Options." The user can choose whether to accept or reject all cookies with just one click, which we consider definitely useful and straightforward. There is also an option letting users define which specific cookies they accept and which they do not. The option pane further provides the user with two different groups: first-party and third-party cookies. Each of these groups is further broken down in five sub-groups: "Information storage and access," "Personalisation," "Ad selection, delivery, reporting," "Content selection, delivery, reporting," and "Measurement." Finally, users can even "See full vendor list" and opt out from each independently. We consider this an example of how websites can and should give users control over the tracking cookies created on their browsers.

The Bad. A large news website does not show any banner to the user indicating cookie usage related to tracking. There is a clickable text at the bottom of the website called "Privacy"—reachable only after scrolling through more than 25 pages of content. After finding the link and following it, there is no quick way to opt out from the tracking cookies, as the website just informs the reader about a third-party opt-out service [57]. In this case users can stop tracking cookies somehow, but they do not have a simple way of doing it.

The Ugly. An important gaming website provides a UI that we found reminiscent of "dark patterns," i.e., *"tricks used in websites and apps that make you buy or sign up for things that you didn't mean to"* [2]. This website presents a blocking notice with two options: a big "I ACCEPT" button and a smaller "Show Purposes" link. When clicking the second option, a list of categories appears; if the user marks any of them and clicks "Save and continue," a big "Accept all" button appears followed by a long list of companies. If the user avoids the button, scrolls down to the end of the list, and hits the small "Reject all" link, an "Are you sure?" dialog appears, warning that rejecting tracking could result in a "limited experience." A "Go back" button brings back to the previous dialogs where users can reconsider their opinions; the only way to actually enter the website without giving consent to tracking is instead clicking the smaller "Leave" link. If any of the accept buttons is ever pressed during the process, instead, the user is directed straight away to the website without any additional formalities.

7 DISCUSSION

Our results show that many websites, even outside the EU, attempt at least to some extent to comply to the GDPR regulations by including various kinds of cookie controls and privacy notices. Even websites that do not, such as most Chinese websites, are affected by its impact, because opting out from third-party services can also reduce the amount of tracking on those websites.

Nevertheless, we still found that a large majority of websites track users through cookies, and we think that the spirit of the law is often still not applied. This can be the consequence of legal, economic, and technical reasons. With respect to legislation, it currently deals with personal data in general, without much detail about how it should be interpreted in particular use cases (the word "cookies" appears exactly once in the 88-page document of the GDPR). Moreover, specific technical guidelines are still missing. From an economic point of view, stakes are big: on the one hand, fines imposed because of non-compliance to the GDPR can be very large; on the other hand, though, the main source of income of many websites is advertising, and we speculate that—even though some recent developments may suggest otherwise [7]—letting users easily opt out from tracking could negate them a part of their income that could potentially be even larger than the fines they could face. Uncertainty with respect to the scope of the legislation and the likelihood of it being enforced may also be involved. Finally, from a technical point of view, enforcing a user-specified opt-out can be very difficult, in particular when a website includes external content (say, a video or a script). In fact, the current HTML specification

does not let a website specify that third party resources should not track users, according to the preferences they expressed. The result is that, to avoid tracking from thid parties, some websites we visited chose to provide only a static, text-only page to users who opted out from tracking.

We conclude that, as of now, despite undeniable improvements cookie tracking is still far from what privacy advocates envision.

8 RELATED WORK

Tracking. Web tracking has been the focus of several pieces of work. Krishnamurthy and Wills [28] were among the first to analyze tracking related to HTML cookies in 2009. One year later, Eckersley [12] realized that users could be identified using certain information from the browser, such as the `user-agent` or the size of the screen.

In the following years, new techniques appeared to fingerprint the user, such as using WebGL rendering to obtain a precise identifier [3] or exploiting differences in the computer internal clock signals [46]. Recent studies compare the effectiveness of existing fingerprinting techniques [29, 50].

Other research analyzes the prevalence of fingerprinting, showing that it is not as widespread as cookie tracking, but is still commonly used [1, 40]. Recent works show that a large and increasing fraction of websites have some kind of tracking either via cookies or fingerprinting [13, 45].

Three recent pieces of work evaluate the changes in the tracking panorama induced by the GDPR. Degeling et al. [8] performed a longitudinal study comparing the information presented to users of EU websites before and after GDPR, focusing on the changes in privacy policies and information presented to users. We complement these results by observing websites *outside* the EU and by studying the tracking that websites actually perform, and the chasm with what they communicate to users. The *whotracks.me* database [26], based on results from real users running privacy-oriented browsers and plugins, has been updated with results remarking that the GDPR induced consolidation in the advertising market, increasing the market share of the biggest players while reducing that of smaller competition [53]. Our results differ from these latter because: (1) that dataset is based on a list of known third-party trackers [32], which by design does not include first-party tracking and may miss some third parties as we discuss in Section 3.4; (2) we evaluate tracking that happens when users *do not* consent to it, while information on user consent is absent from the *whotracks.me* data. Dabrowski et al. [6] crawled a set of websites both from the EU and the USA, and discovered that (1) EU-based visitors are less likely to receive persistent cookies; (2) the number of persistent cookies set for USA-based visitors appears to have diminished as well after the GDPR was introduced. These results are complementary to ours, since we analyze distinctions based on the location of web servers rather than clients; moreover, unlike Dabrowski et al., we distinguish cookies that have enough information to actually identify a user from those that cannot.

Policies. Prior work tried to increase privacy policy transparency by annotating and extracting the most essential parts of privacy policies. Privee [59] is an architecture for automatically extracting essential policy terms and presenting them to a user using a combination of crowdsourcing and machine learning classification techniques. Similarly, Oltramari et al. [41] proposed a framework for annotating policies using a combination of crowdsourcing, machine learning and natural language processing.

Other works measured the complexity of privacy policies and their perception by end users. McDonald et al. [33] compared different privacy policy formats and evaluated the understanding of privacy policies by 749 Internet users; they reported that users were unable to reliably understand companies' privacy policies and all formats were simlarly disliked by users. Jensen and Potts [24] analyzed the content and the complexity of 64 privacy policies from popular websites; thanks to their measurement we can observe that privacy policies were more complex a decade ago than now. Grossklags and Good [21] evaluated the readability and usability of 50 EULAs from popular programs; their results show that complexity of EULAs is comparable to privacy policies, and that many users do not have a clear understanding of the terms they accept.

9 CONCLUSION

We presented an evaluation of web user tracking after the new European General Data Protection Regulation (GDPR) entered in effect. Rather than judging compliance to the law, our goal was investigating to what extent this law helps users better control their privacy.

We found that this regulation has a global reach, giving users information and means to control cookie tracking in both EU-based and USA-based websites; in addition, third party opt-out services also affect tracking on websites that did not explicitly attempt to comply with the new law. On the other hand, we find that despite this, tracking is still ubiquitous and present in more than 90% of websites, even those in the EU. We think that the panorama of tracking in the Web is in flux, and the future results will depend on the technical and standards-based solutions that will make it easier to block tracking, on the enforcement of current regulation, and on the new laws that will be drafted in Europe and in the rest of the world.

ACKNOWLEDGMENT

The authors would like to thank Petros Efstathopoulos and Sunny Athwal for their help in clarifying the legal issues surrounding the GDPR, which have been useful in improving this paper. This work is partially supported by the Basque Government under a pre-doctoral grant given to Iskander Sanchez-Rola.

REFERENCES

[1] Gunes Acar, Marc Juarez, Nick Nikiforakis, Claudia Diaz, Seda Gürses, Frank Piessens, and Bart Preneel. 2013. FPDetective: dusting the web for fingerprinters. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*.
[2] Harry Brignull. 2018. Dark Patterns. https://darkpatterns.org/
[3] Yinzhi Cao, Song Li, and Erik Wijmans. 2017. (Cross-)Browser Fingerprinting via OS and Hardware Level Features. In *Proceedings of the Network and Distributed System Symposium (NDSS)*.
[4] European Commission. 2018. Data protection in the EU. https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en
[5] Cookiebot.com. 2018. GDPR and cookies. https://www.cookiebot.com/en/gdpr-cookies/
[6] Adrian Dabrowski, Georg Merzdovnik, Johanna Ullrich, Gerald Sendera, and Edgar Weippl. 2019. Measuring Cookies and Web Privacy in a Post-GDPR World. In *International Conference on Passive and Active Network Measurement (PAM)*.

[7] Jessica Davies. 2019. After GDPR, The New York Times cut off ad exchanges in Europe — and kept growing ad revenue. Digiday UK. https://digiday.com/media/gumgumtest-new-york-times-gdpr-cut-off-ad-exchanges-europe-ad-revenue/

[8] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. 2019. We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR's Impact on Web Privacy. In *Proceedings of the Network and Distributed System Security Symposium Symposium (NDSS)*.

[9] Matteo Dell'Amico and Maurizio Filippone. 2015. Monte Carlo strength evaluation: Fast and reliable password checking. In *Proceedings of ACM SIGSAC Conference on Computer and Communications Security (CCS)*.

[10] Matteo Dell'Amico, Pietro Michiardi, and Yves Roudier. 2010. Password strength: An empirical analysis. In *Proceedings of IEEE INFOCOM*.

[11] DMA Italia, FedoWEB, Iab Italia, Netcomm, UPA, and Iubenda. 2018. Cookies Instructions Kit. https://help.iubenda.com/wp-content/uploads/2018/04/Cookie-Law-Official-Kit-en.pdf.

[12] Peter Eckersley. 2010. How unique is your web browser?. In *Proceedings of the Privacy Enhancing Technologies (PETS)*.

[13] Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*.

[14] 1995. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Union* (1995). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31995L0046

[15] 2009. Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009. *Official Journal of the European Union* (2009). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32009L0136

[16] 2016. 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). *Official Journal of the European Union* (2016). http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[17] Evidon. 2018. Digital Governance, Privacy Compliance, Website Monitoring. https://www.evidon.com/.

[18] Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology* 32, 3 (1948), 221.

[19] Florida Statutes. 2016. Florida Statutes Section 627.4145 - Readable Language In Insurance Policies. https://law.onecle.com/florida/title-xxxvii/627.4145.html

[20] Nathaniel Good, Rachna Dhamija, Jens Grossklags, David Thaw, Steven Aronowitz, Deirdre Mulligan, and Joseph Konstan. 2005. Stopping spyware at the gate: a user study of privacy, notice and spyware. In *Proceedings of the Symposium on Usable privacy and security (SOUPS)*.

[21] Jens Grossklags and Nathan Good. 2007. Empirical studies on software notices to inform policy makers and usability designers. In *International Conference on Financial Cryptography and Data Security*.

[22] Alex Hern and Jim Waterson. 2018. Sites block users, shut down activities and flood inboxes as GDPR rules loom. The Guardian. https://www.theguardian.com/technology/2018/may/24/sites-block-eu-users-before-gdpr-takes-effect

[23] Dahlia Janan and David Wray. 2012. Readability: The limitations of an approach through formulae. http://www.leeds.ac.uk/educol/documents/213296.pdf

[24] Carlos Jensen and Colin Potts. 2004. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI)*.

[25] Jeremy Kahn, Stephanie Bodoni, and Stefan Nicola. 2018. It'll Cost Billions for Companies to Comply With Europe's New Data Law. https://www.bloomberg.com/news/articles/2018-03-22/it-ll-cost-billions-for-companies-to-comply-with-europe-s-new-data-law

[26] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M. Pujol. 2018. WhoTracks.Me: Monitoring the online tracking landscape at scale.

[27] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. (1975).

[28] Balachander Krishnamurthy and Craig Wills. 2009. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the International Conference on World Wide Web (WWW)*.

[29] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. 2016. Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints. In *Proceedings of the IEEE Symposium on Security and Privacy (Oakland)*.

[30] Issie Lapowsky. 2018. California Unanimously Passes Historic Privacy Bill. Wired. https://www.wired.com/story/california-unanimously-passes-historic-privacy-bill

[31] Legislation.gov.uk. 2018. Data Protection Act 2018. http://www.legislation.gov.uk/ukpga/2018/12/contents/enacted

[32] Sam Macbeth. 2017. Tracking the Trackers: Analysing the global tracking landscape with GhostRank. (2017). https://www.ghostery.com/wp-content/themes/ghostery/images/campaigns/tracker-study/Ghostery_Study_-_Tracking_the_Trackers.pdf

[33] Aleecia M Mcdonald, Robert W Reeder, Patrick Gage Kelley, and Lorrie Faith Cranor. 2009. A comparative study of online privacy policies and formats. In *International Symposium on Privacy Enhancing Technologies Symposium (PETS)*.

[34] William Melicher, Blase Ur, Sean M Segreti, Saranga Komanduri, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2016. Fast, Lean, and Accurate: Modeling Password Guessability Using Neural Networks.. In *Proceedings of the USENIX Security Symposium (Sec)*.

[35] Rani Molla. 2018. Advertisers will spend $40 billion more on internet ads than on TV ads this year. Recide. https://www.recode.net/2018/3/26/17163852/online-internet-advertisers-outspend-tv-ads-advertisers-social-video-mobile-40-billion-2018

[36] Lou Montulli and David M. Kristol. 2000. HTTP State Management Mechanism. RFC 2965. https://rfc-editor.org/rfc/rfc2965.txt

[37] Mozilla. 2018. Security/Tracking protection. https://wiki.mozilla.org/Security/Tracking_protection.

[38] NAI Consumer. 2018. Opt Out of interest-based advertisement. http://optout.networkadvertising.org.

[39] Arvind Narayanan and Vitaly Shmatikov. 2009. De-anonymizing social networks. In *Proceedings of IEEE Symposium on Security and Privacy (Oakland)*.

[40] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *Proceedings of IEEE Symposium on Security and Privacy (Oakland)*.

[41] Alessandro Oltramari, Dhivya Piraviperumal, Florian Schaub, Shomir Wilson, Sushain Cherivirala, Thomas B Norton, N Cameron Russell, Peter Story, Joel Reidenberg, and Norman Sadeh. 2017. PrivOnto: A semantic framework for the analysis of privacy policies. *Semantic Web* (2017).

[42] OneTrust. 2018. Privacy Management Software. https://www.onetrust.com/.

[43] Piwik. 2018. Turn on/off GDPR compliance on the website. https://help.piwik.pro/consent-manager/setting-consent-manager/.

[44] Quantcast. 2018. AI-driven Audience Insights, Targeting & Measurement. https://www.quantcast.com/.

[45] Iskander Sanchez-Rola and Igor Santos. 2018. Knockin' on Trackers' Door: Large-Scale Automatic Analysis of Web Tracking. In *Proceedings of the International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment (DIMVA)*.

[46] Iskander Sanchez-Rola, Igor Santos, and Davide Balzarotti. 2018. Clock Around the Clock: Time-Based Device Fingerprinting. In *Proceedings of the ACM SIGSAC conference on Computer & communications security (CCS)*.

[47] Amazon Web Services. 2018. Alexa Top Sites. https://aws.amazon.com/es/alexa-top-sites/.

[48] Symantec. 2018. Symantec RuleSpace: OEM URL Categorization Database and Real-Time Web Categorization Technology. https://www.symantec.com/products/rulespace

[49] TRUSTe. 2018. Your advertising choices. http://preferences-mgr.truste.com/.

[50] Antoine Vastel, Pierre Laperdrix, Walter Rudametkin, and Romain Rouvoy. 2018. FP-STALKER: Tracking Browser Fingerprint Evolutions. In *Proceedings of IEEE Symposium on Security and Privacy (Oakland)*.

[51] Zachary Weinberg, Shinyoung Cho, Nicolas Christin, Vyas Sekar, and Phillipa Gill. 2018. How to Catch when Proxies Lie: Verifying the Physical Locations of Network Proxies with Active Geolocation. In *Proceedings of the Internet Measurement Conference (IMC)*.

[52] Daniel Lowe Wheeler. 2016. zxcvbn: Low-Budget Password Strength Estimation.. In *Proceedings of the USENIX Security Symposium (Sec)*.

[53] Whotracks.me. 2018. GDPR—What Happened? https://whotracks.me/blog/gdpr-what-happened.html

[54] Wiley Rein. 2017. The GDPRs Reach: Material and Territorial Scope Under Articles 2 and 3. https://www.wileyrein.com/newsroom-newsletters-item-May_2017_PIF-The_GDPRs_Reach-Material_and_Territorial_Scope_Under_Articles_2_and_3.html

[55] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. 2010. A practical attack to de-anonymize social network users. In *Proceedings of IEEE Symposium on Security and Privacy (Oakland)*.

[56] Gilbert Wondracek, Thorsten Holz, Christian Platzer, Engin Kirda, and Christopher Kruegel. 2010. Is the Internet for Porn? An Insight Into the Online Adult Industry.. In *Proceedings of the Workshop on the Economics of Information Security (WEIS)*.

[57] Your Online Choice. 2018. A guide to online behavioural advertisement. http://www.youronlinechoices.com/es/preferencias/.

[58] YourAdChoices. 2018. WebChoices: Digital Advertising Alliance's Consumer Choice Tool. http://optout.aboutads.info.

[59] Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *Proceedings of the USENIX Security Symposium (Sec)*.