Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework

Célestin Matte, Nataliia Bielova, Cristiana Santos

Do Cookie Banners Respect my Choice?
Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework

Célestin Matte
Université Côte d'Azur, Inria
France
celestin.matte@inria.fr

Nataliia Bielova
Université Côte d'Azur, Inria
France
nataliia.bielova@inria.fr

Cristiana Santos
Research Centre for Justice and Governance
School of Law, University of Minho
cristianasantos@protonmail.com

***Abstract*—As a result of the GDPR and the ePrivacy Directive, European users encounter cookie banners on almost every website. Many of such banners are implemented by Consent Management Providers (CMPs), who respect IAB Europe's Transparency and Consent Framework (TCF). Via cookie banners, CMPs collect and disseminate user consent to third parties. In this work, we systematically study IAB Europe's TCF and analyze consent stored behind the user interface of TCF cookie banners. We analyze the GDPR and the ePrivacy Directive to identify potential legal violations in implementations of cookie banners based on the storage of consent and detect such suspected violations by crawling 1 426 websites that contains TCF banners, found among 28 257 crawled European websites. With two automatic and semi-automatic crawl campaigns, we detect suspected violations, and we find that: 141 websites register positive consent even if the user has not made their choice; 236 websites nudge the users towards accepting consent by pre-selecting options; and 27 websites store a positive consent even if the user has explicitly opted out. Performing extensive tests on 560 websites, we find at least one suspected violation in 54% of them. Finally, we provide a browser extension to facilitate manual detection of suspected violations for regular users and Data Protection Authorities.**

Keywords*—*Privacy; GDPR; Consent; Web measurement

I. Introduction

Today's web advertising ecosystem heavily relies on continuous data collection and tracking that allows advertising companies as well as data brokers to continuously profit from collecting a vast amount of data associated to the users. Adopted in April 2016 and implemented in May 2018, the General Data Protection Regulation (GDPR) [55] changed the rules on consent, shaking the tracking and advertisement industry in its practices. The ePrivacy Directive, amended in 2009 (ePD, also known as "cookie law") [54] made it mandatory to collect user's consent before any access or storage of non-mandatory data (not strictly necessary for the service requested by the user). In case of websites, the consent is usually presented in the form of *cookie banners*, or cookie notices that inform the user of data collection and should provide a meaningful choice on whether to accept or reject such collection. The website visitors in the European Union observe such banners on many websites they visit today.

Various research studies looked into detection and measurement of web tracking technologies that perform silent data collection without user's explicit consent [51], [46], [12], [1], [44], [39], [48]. Several recent works [41], [9], [57], [53] have been measuring the impact of GDPR on the web tracking and advertising ecosystem. Libert et al. [41] observed a 22% drop in the amount of third-party cookies before and after the GDPR, but only a 2% drop in third-party content. Degeling et al. [9] recently measured the prevalence of cookie banners and showed that the amount of banners increased over time after the GDPR. Legal scholars, authorities and computer science researchers independently noticed that some banners do not allow users to refuse data collection, and raised this in various studies [9], [38], [2], [59]. Several recent works [56], [57], [53] measured the impact of choices set in cookie banners on tracking: upon accepting and rejecting the consent proposed in a cookie banner, researchers evaluated the number of cookies set in the browser and the number of third-party tracking requests across websites. Latest works [58], [45] evaluated whether the design of cookie banners made an impact on how users would interact with them.

Although many research efforts took place after the GDPR to detect and analyze cookie banners and their impact on tracking technologies and on the users, no study has analyzed what actually happens behind the user interface of cookie banners yet. It is unclear how to meaningfully compare the interface of the banners shown to the users to the actual consent that banners store and transmit to the third parties present on the website. Our work is motivated by the following questions:

Do banners actually respect user's choice made in the user interface? Do banners silently register a positive consent even if the user has not made their choice? Do they nudge the user to accept everything by pre-choosing a positive consent?

Answering such questions, ensuring a proper functionality and legal compliance of a cookie banner is usually left to the website publisher and is completely obscure for the website visitor.

In reaction to the GDPR, the European branch of the Interactive Advertising Bureau (IAB Europe), an advertising business organization, produced the Transparency and Consent Framework (TCF) [30] to structure the practices of actors of the tracking and advertisement industry regarding consent collection. Notably, they introduced the notion of *Consent Management Providers (CMPs)* – actors in charge of collecting consent from the end-user, and redistributing this consent to advertisers. Figure 1 shows a typical example of a cookie banner implemented by a CMP that allows the user to agree or disagree with five predefined purposes of data processing.

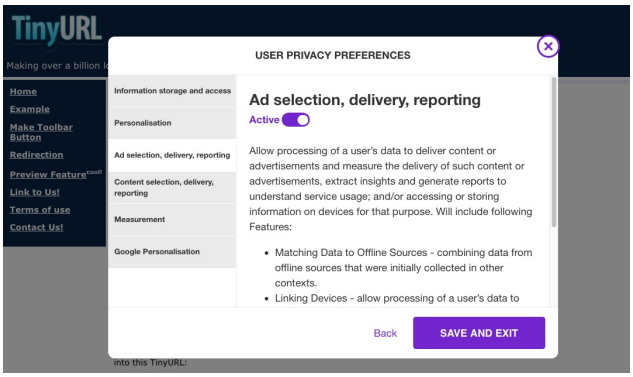


Fig. 1: A cookie banner on `tinyurl.com` of a Consent Management Provider that implements IAB Europe's Transparency & Consent Framework (TCF).

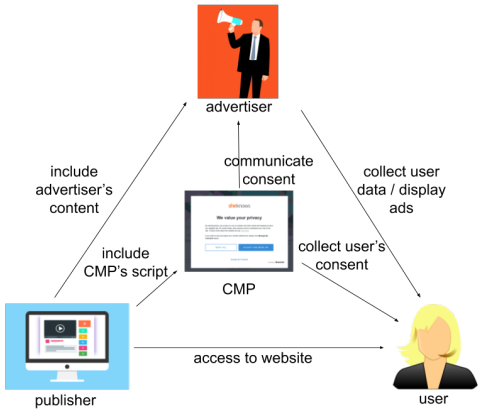


Fig. 2: Consent Management Providers (CMPs) under IAB Europe's TCF.

Contributions. Thanks to the open specification of the TCF, we perform the first systematic comparison of the consent chosen by the users and the consent stored by the CMPs, which is further transmitted to third-party advertisers present on a website. With our analysis of consent, we are able to measure both the GDPR and the ePD compliance of cookie banners implemented in the TCF. We note that the responsibility for the suspected violations are joint between the publishers and the CMPs. Our main contributions are:

1) We design an automatic method to detect the presence of a cookie banner developed by a Consent Management Provider (CMP) (Section IV-B). We automatically detect 1 426 websites with such banners.
2) We develop and use a methodology to intercept the consent stored in the browser (Section IV-C). By analyzing the content of consent, we bring transparency by revealing the companies behind CMPs and publishers.
3) By collaborating with legal scholars (one of the co-authors and external ones), we thoroughly analyze the GDPR, the ePrivacy Directive and other legal texts to identify four potential legal violations specific to cookie banners: *Consent stored before choice*, *No way to opt out*, *Pre-selected choices* and *Non-respect of choice* (Section III).
4) We develop a method to evaluate regulatory compliance of websites (Section IV-E). We quantify the identified suspected violations on 1 426 websites by automatic-, semi-automatic crawls and manual detection (Section VII-A). By analyzing cookie banners' design on a subset of 560 websites (from countries whose language the authors speak), we find that 236 (47%) websites nudge the users towards acceptance by pre-selecting options, while 38 (7%) websites do not provide any means to refuse consent. By analyzing the consent stored in the browser, we automatically detect 141 out of 1 426 (10%) websites that store a positive consent before user has made any choice in the cookie banner, while 27 out of 560 (5%) websites store an all – accepting consent even if the user has explicitly opted out in the cookie banner interface. In total, we find at least one suspected violation in 304 out of 560 websites (54%). We discuss the difficulty to attribute responsibility of these suspected violations in Section XI.
5) We measure the problem of escalation of shared consent between CMPs (Section VII-B). The TCF allows different CMPs and publishers to rely on each other's consent, set in a shared cookie. We observe that 3 websites store a positive consent before user action in the shared cookie, while 20 websites store a positive consent in a shared cookie even if the user has explicitly opted out. Such invalid consent can be reused by any CMP and publisher and therefore escalates non-compliance to other websites.
6) We quantify third-party requests that transmit consent and that belong to known third-party tracking services (Section VIII). We observe that various third parties receive consent with third-party requests, where the origin of consent does not necessarily match the CMP present on the website. Such consents are set before user action on 69 websites and despite user refusal on 38 websites. We observe that the number of third-party tracking requests increases both after positive consent and after refusal.

To measure compliance, we have designed two tools. *Cookinspect* [43] is a Selenium- and Chromium-based crawler which automatically and semi-automatically visits websites, logs stored consent and intercepts transmission of consent to third parties. *Cookie Glasses* [42] is a publicly available browser extension for Google Chrome and Firefox that allows users to detect a CMP that implements a TCF banner and see if their choice is correctly transmitted to advertisers by CMPs.

II. IAB Europe's Transparency and Consent Framework (TCF)

The third-party advertising and tracking ecosystem contains different actors. *Publishers* provide websites to users and include third-party advertising content. *Advertisers* and trackers collect users' data and display ads. Finally, *users* consume content. With the arrival of the GDPR, it became evident that the different actors of this ecosystem were not equipped to properly collect and exchange user's consent.

In April 2018, IAB (Interactive Advertising Bureau) Europe published the Transparency & Consent Framework (TCF) – a technical specification that allows third-parties and publishers to collect and exchange user's consent to data collection and the use of cookies[1]. The TCF was presented as a way

[1]In this work, we study version 1.1 of the TCF. Even though IAB Europe published TCF version 2 on August 21st 2019, we have not observed its application in the wild, and therefore did not address it in this work.

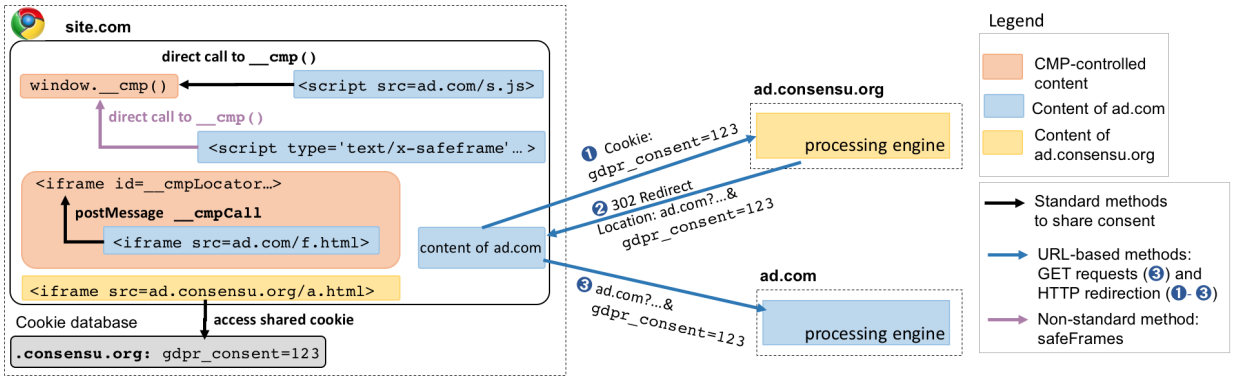


Fig. 3: The methods to share consent in IAB Europe's TCF.

{"created": "2018-11-27 17:24:14",
"cmpId": 139,
"allowedPurposeIds": [1,2,3,4,5],
"allowedVendorIds": [1,2,3 ... ,554,555,556], ...}

Fig. 4: Example of a decoding of a consent string (only fields relevant for this paper are shown).

to help digital advertising industry to "interpret and comply with EU rules on data protection and privacy – notably the GDPR" [30]. IAB Europe introduced new actors, called *Consent Management Providers (CMPs)*, who are responsible for collecting the end user's consent, storing it in the user's browser and implementing methods to respond to advertisers' queries regarding this consent.

Figure 2 summarizes the updated interaction of CMPs with publishers, users and advertisers. To become a part of the TCF, each CMP and advertiser must register with IAB Europe. As a result, IAB Europe maintains: (1) a *public list of CMPs* [27] that participate in the framework, and (2) a *Global Vendor List* (GVL) [31] – a public list of registered advertisers (called "vendors"). As of October 25th 2019, there are 117 CMPs in the CMP list and 550 advertisers in the GVL list. While registering in the GVL, among other information, each advertiser must declare one or more of the *five pre-defined purposes* for which the data is collected and for which of them consent will be used – these are the purposes the user usually sees in the interface of the cookie banner (see Figure 1). Table IX in Appendix A shows the full list and description of all the predefined purposes.

A. Consent String

The TCF defines a standard format for consent [30], called *consent string*. This string contains (1) advertisers for whom the user consented their data to be sent to, (2) purposes of data processing the user consented to, and (3) the CMP identifier, along with other information. This format is a slightly-modified version of base64 of an array of values. We use a script provided by IAB [32] to decode this format.

Figure 4 shows a decoding of the consent string "BOX5uluOX5uluCLAAAENB6-AAAAizAAA", obtained on telerama.fr. The cmpId is an identifier of a CMP from the CMP list [27] responsible for storing a consent string; allowedPurposeIds are the identifiers of the five pre-defined TCF purposes of data processing; and allowedVendorIds are the identifiers of advertisers from the GVL [31]. Note that any third party can identify the CMP that registered the consent string by comparing the cmpId field of the consent string to the public list of CMPs [27].

B. Consent Storage

The TCF does not impose any particular mechanism for storing user's consent in the browser. It only suggests that CMPs use a "first-party service-specific cookie", without further details [29].

As one way to implement consent storage, the TCF proposes CMPs to store a consent string in a cookie, named euconsent in the consensu.org domain owned by IAB, reachable by CMPs through DNS subdomains delegation (we call it "shared cookie" in the rest of the paper)[2]. This mechanism allows CMPs to share consent information despite the Same-Origin-Policy. Since each CMP registered in the TCF has access to its own subdomain (e.g. ad.consensu.org), it can host scripts in it to read and modify the shared cookie.

C. Consent Sharing

Once consent is stored in the user's browser, any advertiser (or, more generally, any third party) present on a page can query a CMP to obtain the consent that was given by the user. In the TCF, consent sharing can be done via: standard APIs, a shared cookie, URL-based methods, and a non-standard method (safeFrames). We present each consent sharing mechanism in more details below. Figure 3 graphically presents how each method obtains a consent string.

Standard APIs. The TCF v1.1 [30] specifies APIs that each CMP must implement – such APIs allow any third-party advertiser present on a publisher website to verify whether a CMP has already stored a consent on a given website. In particular, each CMP must implement:

(i) a javascript function called "__cmp()", that scripts in a first-party position can call directly,
(ii) an iframe named "__cmpLocator", that iframes in a third-party position can communicate with using the postMessage API using a __cmpCall parameter.

[2]TCF mentions that when website-specific cookie and shared cookies are both defined, the website-specific cookie will be used.

Shared cookie. As explained above, CMPs can store consent in a shared cookie named `euconsent` on the `.consensu.org` domain, that any other CMP can read.

URL-based methods. The TCF specifies [34] other methods to send consent to third parties devoid of the possibility to execute JavaScript, such as tracking pixels. First, consent can be passed: through GET requests, in a predefined "`gdpr_consent`" parameter. Second, using an HTTP redirecting mechanism that we show in steps ❶-❸ in Figure 3. Third parties can issue a request to a subdomain of `.consensu.org`, such as `ad.consensu.org`, providing a final destination URL as a parameter of the request, such as `ad.com`. Because browsers automatically attach the cookie of `.consensu.org` to the request, the CMP owning `ad.consensu.org` obtains a shared cookie (step ❶). The CMP then responds with the "302 Redirect" status, indicating `ad.com` as a destination URL and attaching the cookie value in the parameters (step ❷). Finally, `ad.com` receives the shared cookie (step ❸).

Non-standard method: SafeFrames Finally, the TCF proposes an additional non-standard method to share consent: *safeFrames*. A SafeFrame [25] is an API-enabled iframe (implemented via specific first-party scripts) that controls the communication between the webpage content and third-party ads. The SafeFrame proxy obtains consent by calling to the standard `__cmp()` function.

III. GDPR and ePrivacy Suspected Violations

In our work, we focus on the European regulatory framework on data protection and privacy. This section provides a legal analysis of the suspected violations , and the limitations that this analysis entails.

A. Legal Background

In May of 2018, the GDPR enforced the rules regarding the processing of personal data in any environment [20]. In order to lawfully process personal data, companies need to choose a *legal basis of processing* [55, Article 6(1)(a)]. One of the most well-known one is *consent*. Articles 4(11) and 7 of the GDPR have set precise requirements on *valid consent*: it must be freely given, specific, informed, unambiguous, explicit, revocable, given prior to any data collection, and requested in a readable and accessible manner [15].

The ePrivacy Directive ("ePD", also known as "cookie law")[3] [13] provides *supplementary* rules to the GDPR with respect to the processing of personal data in the electronic communication sector, such as websites. While GDPR is a *regulation*, and hence is directly enforceable in every European country, ePD is a *directive*, and hence is left up to each member State to implement in its own national law.

According to the ePD [13, Article 5(3)], and pursuant to the guidance of the European Data Protection Board (EDPB) and Information Commissioner's Office (ICO, the UK's Data Protection Authority), website publishers have to rely on *user consent* when they collect and process personal data using non-mandatory (non strictly necessary for the service requested by the user) cookies or other tracking technologies [19], [35].

[3]The upgrade of the ePD into a regulation is currently under discussion.

An alternative legal basis to process personal data is the *legitimate interest* pursued by the controller or by a third party (Article 6(1)(f) GDPR), e.g. freedom to conduct a business. Nevertheless, the EDPB stated that legitimate interests' grounding is not considered to be an appropriate lawful basis for the processing of personal data in connection with purposes as tracking, profiling and advertising [16],[17].

The following legal analysis is based in the most authoritative legal documents in this specific domain of privacy and data protection law. In particular, we reproduce the arguments already made public by the recent case decisions of the Court of Justice of the EU (CJEU, the highest court in the EU), by the DPAs and the EDPB guidelines, and the legal rules laid down in explicit legal provisions of the GDPR and the ePrivacy Directive, as well as in its recitals (recitals help legal interpretation of provisions in a specific context, but they are not mandatory for compliance). These cited expert generated legal sources are the foundational framework for data protection which we apply in this work to discern whether the declared practices are legally compliant.

B. Legal analysis of potential violations

As a result of a deep legal analysis by a legal scholar (a co-author in this paper), we identify *four potentially legal violations* specific to cookie banners that implement the IAB Europe TCF framework.

Consent stored before choice: *The CMP stores a positive consent before the user has made their choice in the banner. Therefore, when advertisers request for consent, the CMP responds with the consent string even though the user has not clicked on a banner and has not made their choice.*

This practice violates the requirement of *prior consent* which demands that website publishers need to request consent to users (1) prior to any processing activity of personal data [15], and (2) before loading tracking technologies according to Article 5(3) of the ePD [13]. This requirement is further imposed in the guidance from the EDPB [21], the ICO [35] and the CNIL (French Data Protection Authority) [5]. Moreover, the TCF's policy document explicitly states that "a CMP will generate consent signals only on the basis of a clear affirmative action taken by a user" [28].

No way to opt out: *The banner does not offer a way to refuse consent. The most common case is a banner simply informing the users about the site's use of cookies*[4].

This practice configures a violation of the requirement of *unambiguous consent* [55, Art.4(11)] stipulating that in order for the user consent to be valid, the user must give an "unambiguous indication" through a "clear and affirmative action" of his choice [13, Art. 5(3)]. Moreover, Recital 66 of the ePD is quite explicit while directing that *"the methods of offering the right to refuse should be as user-friendly as possible"*. In its guidelines, the EDPB [21] states that *"consent mechanism should present the user with a real and meaningful choice regarding cookies on the entry page"*, and that users *"should have an opportunity to freely choose between the option to accept some or all cookies or to decline all or some*

[4]In previous works, this category has been named as "No option" in Degeling et al. [9] and as "OnlyAccept" in Sanchez-Rola et al.[53]

cookies." In this line, we posit that cookie banner design must offer users an option to either accept or refuse consent.

Pre-selected choices: *The banner gives users a choice between one or more purposes or vendors, however some of the purposes or vendors are pre-selected: pre-ticked boxes or sliders set to "accept".*

Preselected choices consist in a direct violation of the requirement of *unambiguous consent* [55, Article 4(11)]. Recital 32 of GDPR reads further that consent given in the form of a preselected tick in a checkbox does not imply active behaviour of the user and that pre-ticked boxes do not constitute consent. The EDPB [15] indicates that pre-ticked boxes (or opt-out boxes) configure ambiguous behaviours and do not render a valid consent. The ICO guidance [35] and the CNIL [5] observe that *"pre-ticked boxes or any equivalents, cannot be used for non-essential cookies"*. Finally, the Court of Justice of the EU (CJEU) judgment from October 2019 [14] (also known as the "Planet49 GmbH" case), establishes definitely that the consent which a website user must give is not valid if it contains a pre-checked checkbox which the user must deselect to refuse their consent.

Non-respect of choice: *The CMP stores a positive consent in the browser even though the user explicitly refused consent.*

This practice incurs in violation of the *lawfulness principle* established in Articles 5(1)(a) and 6(1) of the GDPR: for the processing to be lawful, it must be based on a legal ground, namely, the user consent to the use of cookies (legal ground demanded by Article 5(3) of the ePD). The EDPB [18] further specified that *"if the individual decided against consenting, any data processing that had already taken place would be unlawful"* due to lacking legal basis for processing.

C. Limitations of the Legal Analysis

Even if our analysis is endorsed in legislation, judicial decisions and expert-generated legal sources, this analysis is yet limited if not sustained judicially. Therefore we deliberately leave space to legal uncertainty on the assessment of the identified questionable practices and emphasize that only a judicial assessment that requires more specific fact finding of each practice could render a final appraisal of such analysis and provide legal certainty.

IV. Methodology

The goal of our study is to detect the suspected violations of the GDPR and the ePrivacy Directive in cookie banners that implement IAB Europe's Transparency & Consent Framework (TCF). We detect all suspected violations defined in Section III on websites that originate from the European Union because a TCF banner is more likely to be observed on EU websites.

In this section, we describe the website selection and data collection processes, followed by our methods to detect TCF banners and intercept the user consent string. We then explain how we detect suspected GDPR and ePD violations with our methodology. To detect suspected violations at scale, we built a crawler, called *Cookinspect* [43], based on a Selenium-instrumented Chromium, that we use to perform large-scale automatic crawls and semi-automatic crawls (with certain human actions) to audit websitesat scale. We describe the two measurement campaigns done with *Cookinspect* in Section V.

A. Websites Selection and Reachability

We used Tranco to build lists [37]. Tranco aggregates results from different lists over a month in order to overcome flaws inherent to these lists' creation: instability, presence of unreachable domains, possible attacks to insert domains, etc[5]. From the top 1 million list of Tranco of September 20th 2019, we extracted the top 1 000 websites of the TLD of 31 European countries and 1 000 websites from three country-independant TLDs: `.com`, `.org` and `.eu`. Altogether, we obtained 28 257 websites (some countries had few websites in Tranco).

Since our study is focused on the respect of consent, we decided to respect publishers' consent regarding bots and crawling on their websites by checking the instructions in each website's `robots.txt` file. For each website d in a list of 28 257 websites, we first visited the address `https://www.`d`/robots.txt` using Python's `urllib` to verify access authorization. If access was denied, we did not crawl the website. As a result, 3 633 (12.86%) websites out of 28 257 refused access to robots in their `robots.txt` file, so we removed them from our further analysis.

While testing authorization, we also verified reachability. If loading the `robots.txt` file failed for a network-related reason, we attempted accessing it through HTTP. If DNS resolution failed for `www.`d, we attempted accessing d instead. We determined if the website was loaded with a timeout of 10 seconds. If loading timed out 3 times, we considered access not successful. We applied the same criteria when visiting the main page with a Selenium-controlled browser. In total, 1 675 (5.9%) websites were not reachable.

As a result, we successfully automatically crawled 22 949 websites originating from up to 1 000 websites from each TLD. The resulting 22 949 websites constitute the basis for the investigations that we describe in Section V.

B. Detecting a TCF Cookie Banner

We first automatically detect websites with cookie banners that implement the TCF. *Cookinspect* detects the presence of a TCF banner by checking whether a `__cmp()` function is defined on a given website (each CMP must implement a `__cmp()` function according to the TCF v1.1 specification [29], see more details in Section II-C).

To validate our detection method in practice, we made a preliminary crawl on 21 000 websites, and analyzed how often `__cmp()` and `__cmpLocator` are used. We observed that all websites (except for one) that implement `__cmpLocator` also implement the `__cmp()` function. We also observed that 20.76% of websites define a `__cmp()` function but do not implement `__cmpLocator`. Thus, we can safely use a hypothesis that if the `__cmp()` function is defined, then a CMP is present on a website. We therefore rely on the presence of a `__cmp()` function to conclude that a CMP is present on a website. When crawling a website, *Cookinspect* waits for the website to finish all loading, waits for additional 3 seconds[6], and then verifies whether the `window` object

[5] See Appendix C for the lists and options we used.

[6] Experimental tests on 200 websites showed that no more than a 2s delay is necessary.

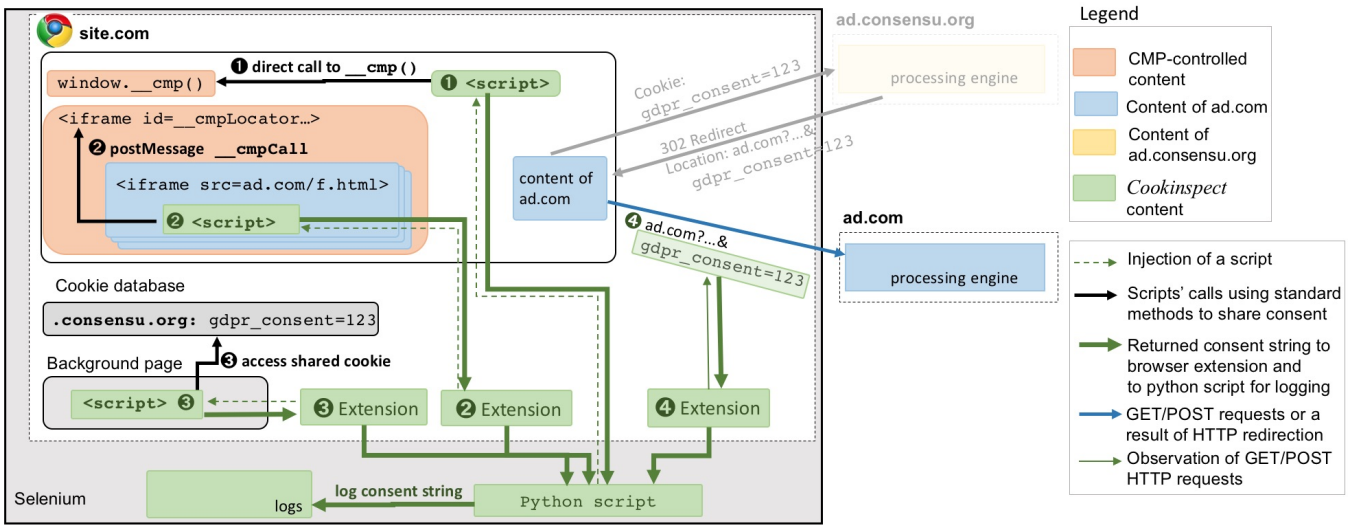


Fig. 5: Detecting stored consent strings with *Cookinspect*.

contains a `__cmp` function. If a `__cmp()` function is not detected, *Cookinspect* does not reload the page.

As a result, we have automatically detected TCF banners on 1 426 websites and we show further results on the prevalence of TCF banners and CMPs that implement them in Section VI. Websites on which we did not find a cookie banner are omitted for the rest of the paper.

C. Intercepting a Consent String

CMPs that implement a TCF banner can use a number of different methods to share a consent string with advertisers present on a website (see Section II-C). In this section, we describe how *Cookinspect* [43] intercepts the consent strings using all available methods, summarized in Figure 5. *Cookinspect* relies on several browser extensions that help injecting scripts and collecting the consent string using different methods. *Cookinspect* contains a Python script that collects all the intercepted consent strings and logs them in a local database.

Standard APIs. First, *Cookinspect* actively pretends to be an advertiser in a first party position: it inserts its script in the context of the crawled website (method ❶ in Fig. 5). The injected script is first-party because it runs in the origin of the crawled website (in the origin of `site.com` in Fig. 5). The injected script makes a direct call to the `__cmp()` function, and if it obtains a consent string in return, then the script transmits it back to the Python script that logs the consent strings.

Second, *Cookinspect* pretends to be an advertiser in a third-party position: *Cookinspect* contains a custom browser extension ❷ that injects a script in all third-party iframes that have the `__cmpLocator` iframe as a parent (only the children of an iframe with the `__cmpLocator` identifier are able to query for the consent string). From each such iframe, the injected script sends a postMessage `__cmpCall` to the `__cmpLocator` iframe to request the consent string (method ❷). The script then transmits it back to the extension and further to the Python script for logging.

Shared cookie. A browser extension ❸ of *Cookinspect* attempts to read the shared cookie, and then transmits it to the extension and back to the Python script.

URL-based methods. To intercept the URL-based methods and obtain consent strings shared with third parties, a custom browser extension ❹ monitors all network requests. According to the TCF, a consent can be transmitted by the URL-based methods inside the GET `gdpr_consent` parameter – we therefore log all the requests containing such parameter (method ❹). Although the TCF only mentions GET requests, we also monitor POST requests parameters. We observed that POST requests transmit a consent string on 399 websites (out of 1 426 TCF websites, i.e. 28%), while GET requests do so on 764 websites (53.6%).

Since the consent redirecting mechanism (detailed in section II-C) always uses HTTP requests to transmit the consent string in the `gdpr_consent` parameter, we already detect it by intercepting all GET and POST requests that contain such parameter.

Non-standard method. According to TCF specification, safeFrames act as a proxy to the `__cmp()` function. *Cookinspect* does not specifically interfere with safeFrames, because they obtain a consent string by making a direct call to the `__cmp()` function, which is already done with method ❶.

D. Identifying the CMP Responsible for a TCF Banner

To identify a CMP behind a banner, we use the consent string that we obtain from the standard APIs and the shared cookie. We decode the consent string with a public script provided by IAB [32]. The decoded array contains the CMP identifier or ID (see `cmpId` in an example of decoded consent string in Fig. 4). We map the CMP ID to the public CMP list [27] to retrieve the CMP company's name.

E. Detecting Suspected GDPR and ePD Violations

In this section, we first explain what information we extract from the consent strings, and then explain how we detect suspected violations. Each consent string contains two arrays: an array of allowed advertisers, and an array of accepted purposes. The TCF indicates [23] that advertisers are supposed

to verify that their identifier is allowed in the advertisers array, and that the purposes they use is allowed in the purposes array. As purposes for data processing have a strong legal meaning (see Section III), we focus on the *purposes* stored in a consent string, and do not analyze the array of allowed advertisers. We do, however, remove websites where a consent string contains an empty array of advertisers and a non-empty array of purposes. We also remove 2 websites where the vendors remaining in the consent string based their processing on legitimate interests only. We leave the discussion of such cases to the Data Protection Authorities (DPAs).

By using *Cookinspect*, we detect the four major suspected GDPR and ePD violations presented in Section III. We explain below what methods we used to detect each violation.

Consent stored before choice: *Cookinspect* logs all the consent strings by using the standard APIs and reading the shared cookie (methods ❶, ❷ and ❸ in Figure 5). *Cookinspect* is able to detect *Consent stored before choice* violation fully automatically while crawling 22 949 websites without performing any user action. If the consent string stored in the user's browser is empty (has no accepted purposes), then we do not label it as a violation. We therefore consider a suspected violation only if the consent string has *one or more accepted purposes* (out of five possible purposes in the TCF, see Appendix A) even though no action was performed on the visited website.

No way to opt out: we detect this suspected violation manually by visiting the websites in a clean Chromium session and using *Cookinspect* to assign the corresponding label. To identify whether there is an option to refuse consent, we click on every button and link to verify whether they lead to a second layer of the banner. In the second layer we refuse consent by deselecting all purposes and advertisers. Notice that the second layer is often hidden behind a misleading terminology (e.g. "learn more"), which does not indicate that refusal is possible.

Pre-selected choices: we detect this suspected violation manually by visiting the websites in a clean Chromium session and using *Cookinspect* to assign the corresponding label. We label a website as violating if it has a "parameters" option, and if in the "parameters" page there is at least one pre-selected checkbox or enabled slider for at least one purpose.

Non-respect of choice: to detect this violation, we perform a semi-automatic crawl based on a human operator and *Cookinspect* on 560 websites from French-, Italian- or English-speaking countries: France, UK, Belgium, Ireland and Italy, and `.com` websites. We only consider banners written in a language that at least one of the authors speak to ensure that we understand the actions we perform. First, in a clean Chromium session, a human operator manually refuses consent on the cookie banner (following the procedure detailed in Appendix D). Then, *Cookinspect* logs all the consent strings by using the standard APIs and reading the shared cookie.

Some TCF banners may display purposes that users cannot refuse, considering that such purposes do not require consent. Such decision can be criticized by legal experts and policy makers, however we exclude such discussions from our work. Instead, we further consider only consent strings that have *all five accepted purposes*, to ensure that a violation indeed took place, even if the user couldn't refuse some of the purposes of data processing in the TCF banner. Thus, we avoid the case where setting a purpose in a consent string may be considered legal because this purpose can be relied on using legitimate interests.

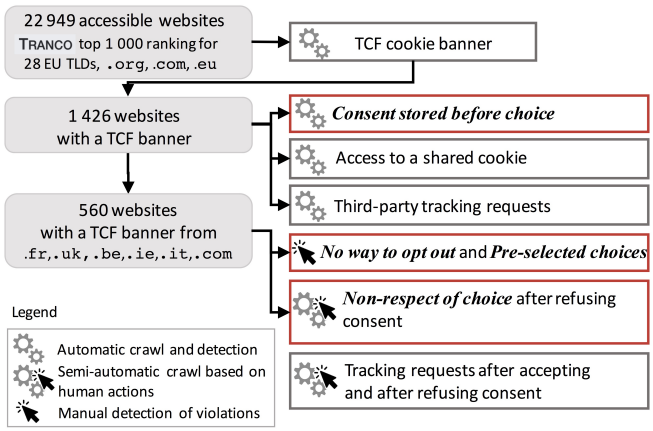


Fig. 6: Overview of the website-auditing process combining automatic crawling, semi-automatic crawling with human interaction and manual analysis to detect suspected GDPR and ePD violations.

TABLE I: Overview of methods to detect the suspected GDPR and the ePD violations in TCF banners.

Short name	Type of crawl	Analyzed component	Number of tested websites
Consent stored before choice	automatic	consent string	1 426
No way to opt out	semi-automatic	banner	560
Pre-selected choices	semi-automatic (when opting out is possible)	banner	508
Non-respect of choice	semi-automatic (when opting out is possible)	consent string	508

We underline that the responsibility for suspected violations is joint between publishers and CMPS and discuss the difficulty to attribute responsibility in Section XI.

V. Two Measurement Campaigns

We perform a large-scale study of websites registered in the EU because the TCF is likely to be observed more often on European websites. We perform two measurement campaigns with *Cookinspect*: a large-scale automatic crawl and a smaller-scale semi-automatic crawl, both conducted in September 2019 from France. The source code of *Cookinspect* and all the extensions is publicly available so that end users and DPAs can test compliance of publishers and CMPs by themselves [43].

Figure 6 provides an overview of the main components of our website auditing process. Table I presents an overview of suspected violations we detect using automatic and semi-automatic crawling campaigns with *Cookinspect*. For each violation, we show the crawl used for its detection and which component of the Web application allowed us to detect it.

A. Automatic Crawl

First, we perform a stateless *automatic crawl* of 22 949 selected websites (see Section IV-A for websites selection) to perform website auditing without human intervention to detect: (1) whether the website contains a TCF banner[7], (2) whether positive consent is stored without any user action (***Consent stored before choice*** violation), (3) whether the website accesses the consent string in a shared cookie that we set in the browser, (4) presence of third-party tracking requests prior to any user consent.

Procedure: In a first browser session, we detect if the website implements the TCF by verifying whether a `__cmp()` function is defined after load and a 3s delay. If so, we attempt to obtain the consent string using *Cookinspect* to detect the ***Consent stored before choice*** violation. *Cookinspect* also logs all requests made to third parties. In a new clean browser session, we place a shared cookie in the browser, and attempt to get it back using a direct `__cmp()` call and a postMessage, to measure whether the CMP on the page uses the shared cookie.

This crawl takes an average of 11.04s per website. Crawling 28 000 websites takes more than 24 hours. This crawl was made from France on September 20th and 21st 2019.

B. Semi-Automatic Crawl

From the resulting 1 426 websites that contain a TCF banner, we select 560 websites of French, Italian or English-speaking countries (the languages that the authors speak fluently) from `.uk`, `.fr`, `.it`, `.be`, `.ie` and `.com` TLDs to perform a *semi-automatic crawl*.

This crawl performs tests requiring interaction with the cookie banner[8]. Upon different browser sessions, we both give a positive consent and refuse consent, to make tests regarding respect of given consent, and setting of the shared cookie. We also note whether banners propose an option to refuse consent, and whether specific parameters are pre-selected in favor of consent-sharing.

Procedure: On a clean browser session, we load the website. If there is no banner or a broken banner, we stop there. We manually label when the ***Pre-selected choices*** or ***No way to opt out*** suspected violation are observed (see the complete procedure describing the labelling process in Section IV-E). We then refuse consent on the banner, and observe with *Cookinspect* what consent is stored by the CMP (including the shared cookie) to detect the ***Non-respect of choice*** violation. After 3 seconds, we refresh the page to log all network requests (to quantify the amount of trackers). We also attempt to obtain the consent string again. Then, on a new session, we repeat this procedure, this time giving a positive consent to the banner. We give the detailed procedure for the human operator to refuse and accept consent on the banner in Appendix D.

Each website takes around 30-40s given a reactive human manipulation. We performed this stateless crawl from France from September 23rd to October 1st 2019.

[7]We do not test for the `__cmpLocator` iframe presence, as we found only 1 website among 21 000 on which we could find a `__cmpLocator` iframe but no `__cmp()` function during a test run.

[8]As the TCF does not specify anything regarding the user's interface, we don't have a way to locate a banner and its different elements, and *a fortiori* to automate banner interaction.

C. Verification Procedure

For the *Non-respect of choice* and the *No way to opt out* violations, we cross-checked choices made in the banner and whether the banner allows to refuse consent by two human operators to limit errors. The three operators are computer scientists working on web tracking. The semi-automatic crawl is first entirely done by one of the operators. For each CMP on which we observe a *Non-respect of choice* on at least one website, we select one of these violating websites. We add an equal number of websites on which the violation is not seen in the pool of sites to be verified. Then, a second human operator refuses consent on all of these websites, unknowingly of which website was considered a violation by the previous tester. We do the same for the *No way to opt out* violation, this time testing for the possibility to refuse consent. Then, the third human operator repeats the procedure on a new pool of websites. In case we obtain different results from different operators, we keep the results of the test returning the least violations on the concerned website, and retest all websites of the considered CMP to take into account the fact that some option of the banner may have been missed.

D. Ethical Considerations

Our data collection process does not involve any human subject. Our study of websites is mostly passive: the only action we perform on websites is clicking on manually-selected cookie acceptance buttons. Hence, we do not tamper with any distant system. Our large-scale measurement does not present any potential harm for websites, while presenting societal benefits. Moreover, we respect publishers' consent regarding bots that they express in a `robots.txt` file.

VI. Prevalence of TCF Banners in Europe

We conducted an automatic crawl of 28 257 websites from 1 000 top Tranco websites for 31 European TLDs and from `.com`, `.org` and `.eu` domains between September 20th and September 23rd 2019. Among reachable and authorized websites, 1 426 (6.2%) had a TCF banner (cookie banner of a CMP implementing the TCF). We show per-TLD details in Table II. The 1 426 websites that have a TCF banner are the target of the following automatic crawls.

We extract information from the consent strings to identify the CMP present on a website. As not all websites were setting up a consent string upon our visit (see our methodology in Section IV), and some consent strings contain an incorrect CMP ID, we have been able to identify the CMP company behind a TCF banner for 298 (20.9%) websites in the automatic crawl, and 511 (92.9%) websites in the semi-automatic crawl. We represent the distribution of identified CMPs in the semi-automatic crawl in Figure 7. The most encountered CMP is Quantcast, far beyond OneTrust, Didomi and Sourcepoint.

We have not found any implementation of the version 2 of TCF that came out in August 20th 2019.

VII. Quantification of Suspected Violations

In this section, we comment on the results regarding the main suspected violations of the GDPR and the ePD described in section III. These suspected violations concern consent

TABLE II: Distribution of websites with a TCF banner across European (and 3 international) TLDs, computed with an automatic crawl.

TLD	Number of domains in the Tranco top-1 million list	Number of reachable and allowed (for bots) domains	Number of domains with a TCF-related cookie banner
.uk	1 000	788 (78.8%)	149 (18.9%)
.fr	1 000	815 (81.5%)	139 (17.1%)
.pl	1 000	858 (85.8%)	129 (15.0%)
.it	1 000	824 (82.4%)	123 (14.9%)
.es	1 000	800 (80.0%)	113 (14.1%)
.nl	1 000	838 (83.8%)	65 (7.8%)
.gr	1 000	720 (72.0%)	53 (7.4%)
.pt	1 000	793 (79.3%)	52 (6.6%)
.de	1 000	881 (88.1%)	56 (6.4%)
.ro	1 000	787 (78.7%)	50 (6.4%)
.bg	547	449 (82.1%)	26 (5.8%)
.fi	1 000	824 (82.4%)	47 (5.7%)
.no	1 000	862 (86.2%)	48 (5.6%)
.dk	1 000	824 (82.4%)	41 (5.0%)
.be	1 000	863 (86.3%)	38 (4.4%)
.at	1 000	873 (87.3%)	33 (3.8%)
.ie	1 000	769 (76.9%)	25 (3.3%)
.cz	1 000	916 (91.6%)	28 (3.1%)
.ch	1 000	849 (84.9%)	26 (3.1%)
.se	1 000	787 (78.7%)	21 (2.7%)
.sk	1 000	879 (87.9%)	14 (1.6%)
.hr	627	513 (81.8%)	8 (1.6%)
.hu	1 000	794 (79.4%)	6 (0.8%)
.lu	186	147 (79.0%)	1 (0.7%)
.lt	745	605 (81.2%)	4 (0.7%)
.lv	537	420 (78.2%)	2 (0.5%)
.si	514	426 (82.9%)	2 (0.5%)
.is	358	248 (69.3%)	1 (0.4%)
.ee	468	373 (79.7%)	1 (0.3%)
.li	105	62 (59.0%)	0 (0.0%)
.cy	108	76 (70.4%)	0 (0.0%)
.mt	62	37 (59.7%)	0 (0.0%)
.com	1 000	711 (71.1%)	97 (13.6%)
.org	1 000	735 (73.5%)	16 (2.2%)
.eu	1 000	803 (80.3%)	12 (1.5%)
all	**28 257**	**22 949 (81.2%)**	**1 426 (6.2%)**

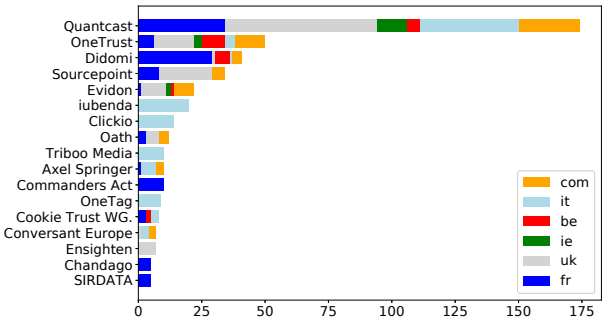


Fig. 7: Distribution of the identified CMPs seen at least 5 times during the semi-automatic crawl.

strings obtained using the standard API and shared cookie (see section IV-C). We note that there is a joint responsibility between publishers and CMPs for these suspected violations. Violations related to consent strings seen in GET and POST requests are shown in section VIII-A, because we cannot attribute CMPs to these cases.

A. Detected GDPR and ePD Suspected Violations

1) Overview of Suspected Violations: We show a summary of the main suspected violations' prevalence, depending on the number of purposes in the consent strings, in Table III. As a reminder, we consider violations of *Consent stored before choice* when we find a consent string with 1 to 5 purposes set, but only when 5 purposes are set for *Non-respect of choice*.

TABLE III: Number of websites seen with the different violations, w.r.t. the maximum number of purposes in observed consent strings. Considered violating cases are shown in bold.

Number of purposes	*Consent stored before choice*	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
1 to 4 purposes	**2.1%** (30/1426)	-	-	6.7% (34/508)
5 purposes	**7.8%** (111/1426)	-	-	**5.3%** (27/508)
Total number of violations	9.9% (141/1426)	6.8% (38/560)	46.5% (236/508)	5.3% (27/508)
Any violation	54.29% 304/560			

We find examples of websites for all considered violations. We find that 38 (6.8%) websites do no provide any way to refuse consent. 236 (46.5%) websites pre-tick the purpose or vendor options. 141 websites (9.9%) set a consent string with 1 to 5 purposes before any user action. 27 websites (5.3%) set a consent string with 5 purposes even though the user refused consent.

2) Quantification per Suspected Violation: Table IV shows the results for each suspected violation, grouped by CMP seen performing a violation at least 3 times in the semi-automatic crawl. For websites for which we found a suspected violation, we provide their global Tranco rank. To compute the results of the *Pre-selected choices* and *Non-respect of choice* violations, we only consider websites having a banner proposing a way to refuse consent (508 websites), i.e. we exclude banners having the *No way to opt out* violations (38 websites), and broken/missing banners (14 websites).

Consent stored before choice: Table V shows results of the automatic crawl per TLD. We observe 141 websites registering a consent string that contains a positive consent even though the user did not perform any action. 111 of them contain all of the TCF's purposes. This is a striking abuse of the framework, happening on more than 1 in 10 websites using it. Interestingly, according to the TCF specification, the APIs we have used to detect consent string *should not return the consent string before the user gives their decision on consent* (or consent is retrieved from existing cookies) [29].

No way to opt out: We observe 38 websites offering no option to refuse consent. These website take part in a framework about user's consent collection, but do not actually offer a way to refuse consent. Collected consent cannot be considered *free*, as required by the GDPR.

Pre-selected choices: Almost half of tested websites (236 out of 508) pre-select choices. In the Planet49 case [14] announced few days after we finished the crawling campaigns, the European court of Justice decided that such pre-selected choices lead to an invalid consent.

Non-respect of choice: 27 websites register a positive consent even though the user refused consent. This strikingly violates user's choice, the framework, and the GDPR.

We observe a variety of suspected violations among the different CMPs. Interestingly, violations are often seen on a partial number of websites. This shows that CMPs offer several versions of their banners that behave differently. We further discuss the shared responsibility of violations between CMPs and publishers in section XI.

TABLE IV: Quantification of violations of the GDPR encountered in the different CMPs for which the considered violations has been seen at least 3 times during the semi-automatic crawl (on `.fr`, `.uk`, `.it`, `.be`, `.ie` and `.com` websites), by CMP. The *Non-respect of choice* and *Pre-selected choices* tables only consider websites on which refusing consent was possible.

(a) *Consent stored before choice*

CMP	total	.uk	.fr	.it	.be	.ie	.com
OneTrust	74.0% (37/50)	81.2% (13/16)	66.7% (4/6)	25.0% (1/4)	100.0% (9/9)	100.0% (3/3)	58.3% (7/12)
Axel Springer	60.0% (6/10)	-	0.0% (0/1)	83.3% (5/6)	-	-	33.3% (1/3)
Quantcast	3.4% (6/174)	0.0% (0/60)	0.0% (0/34)	12.8% (5/39)	20.0% (1/5)	0.0% (0/12)	0.0% (0/24)
Commanders Act	40.0% (4/10)	-	40.0% (4/10)	-	-	-	-
Global Radio Services	100.0% (3/3)	100.0% (3/3)	-	-	-	-	-
Livesport Media	100.0% (3/3)	-	100.0% (1/1)	100.0% (1/1)	-	-	100.0% (1/1)
others	1.9% (6/310)	1.4% (1/70)	1.1% (1/87)	0.0% (0/73)	15.4% (2/13)	0.0% (0/10)	3.5% (2/57)
all	**11.6% (65/560)**	**11.4% (17/149)**	**7.2% (10/139)**	**9.8% (12/123)**	**44.4% (12/27)**	**12.0% (3/25)**	**11.3% (11/97)**

(b) *No way to opt out*

CMP	total	.uk	.fr	.it	.be	.ie	.com
Quantcast	5.2% (9/174)	0.0% (0/60)	0.0% (0/34)	23.1% (9/39)	0.0% (0/5)	0.0% (0/12)	0.0% (0/24)
Axel Springer	70.0% (7/10)	-	100.0% (1/1)	83.3% (5/6)	-	-	33.3% (1/3)
Evidon	22.7% (5/22)	10.0% (1/10)	100.0% (1/1)	-	100.0% (1/1)	50.0% (1/2)	12.5% (1/8)
Global Radio Services	100.0% (3/3)	100.0% (3/3)	-	-	-	-	-
others	4.0% (14/351)	2.6% (2/76)	2.9% (3/103)	2.6% (2/78)	9.5% (2/21)	9.1% (1/11)	6.5% (4/62)
all	**6.8% (38/560)**	**4.0% (6/149)**	**3.6% (5/139)**	**13.0% (16/123)**	**11.1% (3/27)**	**8.0% (2/25)**	**6.2% (6/97)**

(c) *Pre-selected choices*

CMP	total	.uk	.fr	.it	.be	.ie	.com
Quantcast	37.8% (62/164)	55.0% (33/60)	32.4% (11/34)	20.0% (6/30)	60.0% (3/5)	33.3% (4/12)	21.7% (5/23)
OneTrust	83.3% (40/48)	93.3% (14/15)	83.3% (5/6)	100.0% (4/4)	77.8% (7/9)	66.7% (2/3)	72.7% (8/11)
Sourcepoint	64.7% (22/34)	100.0% (21/21)	0.0% (0/8)	-	-	-	20.0% (1/5)
Didomi	39.0% (16/41)	100.0% (1/1)	24.1% (7/29)	0.0% (0/1)	100.0% (6/6)	-	50.0% (2/4)
OneTag	100.0% (9/9)	-	-	100.0% (9/9)	-	-	-
Commanders Act	80.0% (8/10)	-	80.0% (8/10)	-	-	-	-
Conversant Europe	100.0% (7/7)	100.0% (1/1)	-	100.0% (3/3)	-	-	100.0% (3/3)
Ensighten	100.0% (7/7)	100.0% (7/7)	-	-	-	-	-
SFR	100.0% (4/4)	-	100.0% (4/4)	-	-	-	-
Evidon	25.0% (4/16)	11.1% (1/9)	-	-	-	-	42.9% (3/7)
Wikia. (FANDOM)	100.0% (3/3)	-	-	-	-	-	100.0% (3/3)
Cookie Trust WG.	60.0% (3/5)	-	100.0% (2/2)	50.0% (1/2)	0.0% (0/1)	-	-
Axel Springer	100.0% (3/3)	-	-	100.0% (1/1)	-	-	100.0% (2/2)
Snigel Web Services	75.0% (3/4)	-	100.0% (1/1)	-	-	0.0% (0/1)	100.0% (2/2)
Spil Games	100.0% (3/3)	100.0% (1/1)	100.0% (1/1)	100.0% (1/1)	-	-	-
TrustArc	100.0% (3/3)	-	-	-	-	-	100.0% (3/3)
Livesport Media	100.0% (3/3)	-	100.0% (1/1)	100.0% (1/1)	-	-	100.0% (1/1)
others	23.6% (33/140)	25.0% (7/28)	40.0% (14/35)	7.4% (4/54)	100.0% (1/1)	33.3% (1/3)	31.6% (6/19)
all	**46.5% (236/508)**	**60.1% (86/143)**	**42.1% (56/133)**	**28.3% (30/106)**	**77.3% (17/22)**	**36.8% (7/19)**	**47.1% (40/85)**

(d) *Non-respect of choice*

CMP	total	.uk	.fr	.it	.be	.ie	.com
Evidon	25.0% (4/16)	22.2% (2/9)	-	-	-	-	28.6% (2/7)
OneTrust	8.3% (4/48)	6.7% (1/15)	0.0% (0/6)	0.0% (0/4)	11.1% (1/9)	0.0% (0/3)	18.2% (2/11)
WEBEDIA	100.0% (3/3)	-	100.0% (3/3)	-	-	-	-
Livesport Media	100.0% (3/3)	-	100.0% (1/1)	100.0% (1/1)	-	-	100.0% (1/1)
M6 Web	75.0% (3/4)	-	75.0% (3/4)	-	-	-	-
others	2.3% (10/434)	0.0% (0/119)	2.5% (3/119)	2.0% (2/101)	0.0% (0/13)	0.0% (0/16)	7.6% (5/66)
all	**5.3% (27/508)**	**2.1% (3/143)**	**7.5% (10/133)**	**2.8% (3/106)**	**4.5% (1/22)**	**0.0% (0/19)**	**11.8% (10/85)**

TABLE V: Results of the *Consent stored before choice* violation on 1 426 websites via an automatic crawl.

TLD	Number of websites
.uk	11.4% (17/149)
.fr	7.2% (10/139)
.pl	17.8% (23/129)
.it	9.8% (12/123)
.es	7.1% (8/113)
.nl	0.0% (0/65)
.gr	3.8% (2/53)
.pt	1.9% (1/52)
.de	28.6% (16/56)
.ro	2.0% (1/50)
.bg	3.8% (1/26)
.fi	10.6% (5/47)
.no	2.1% (1/48)
.dk	4.9% (2/41)
.be	42.1% (16/38)
.at	12.1% (4/33)
.ie	12.0% (3/25)

TLD	Number of websites
.cz	3.6% (1/28)
.ch	0.0% (0/26)
.se	9.5% (2/21)
.sk	7.1% (1/14)
.hr	0.0% (0/8)
.hu	0.0% (0/6)
.lu	0.0% (0/1)
.lt	25.0% (1/4)
.lv	0.0% (0/2)
.si	0.0% (0/2)
.is	0.0% (0/1)
.ee	0.0% (0/1)
.com	11.3% (11/97)
.org	12.5% (2/16)
.eu	8.3% (1/12)
all	**9.9% (141/1426)**

We give additional presentations of the results (per-country and per-CMPs views) in Appendix E.

3) Quantification per Publisher: We observe suspected violations on a wide range of websites. For each suspected violation, we display the lists of top 10 violating websites, ordered by their rank in the Tranco list in Table VI. `msn.com`, a web portal ranked 48 in the Tranco list, stores a positive consent before any user choice, then offers no way to opt out. `medicalnewstoday.com`, a website about health, does the same, even though medical information is a sensitive category of data. `w3schools.com`, a popular website providing web development tutorials, displays a banner with pre-selected choices, but registers a positive consent even if the user goes to the trouble of deselecting them. `softonic.com`, website of a major software developer, registers a positive consent before user action, then displays a banner with pre-selected choices, and finally does not respect the user's decision.

B. Escalation of Suspected Violations with the Shared Consent Mechanism

Setting a violating consent string in a cookie shared among all TCF websites would constitute an escalation of the problem. We investigate the question: to what extent do websites use the shared cookie? As explained in section IV-C, we try to read it using a browser extension after both giving a positive

TABLE VI: Top 20 websites where we observe each suspected violation, ordered by their rank in the Tranco list. See complete lists of websites for each suspected violation in attachment [3].

Consent stored before choice

Tranco rank	domain
48	msn.com
373	softonic.com
434	merriam-webster.com
453	britannica.com
545	slate.com
551	thesun.co.uk
706	medicalnewstoday.com
821	thetimes.co.uk
841	techtarget.com
1 203	vanityfair.com

No way to opt out

Tranco rank	domain
48	msn.com
322	healthline.com
345	economist.com
545	slate.com
706	medicalnewstoday.com
849	discogs.com
8 017	ilmessaggero.it
8 842	ticketmaster.co.uk
10 972	tomshw.it
12 623	ilgazzettino.it

Pre-selected choices

Tranco rank	domain
58	cnn.com
113	reuters.com
119	tinyurl.com
127	bloomberg.com
133	fandom.com
197	w3schools.com
316	mashable.com
373	softonic.com
425	wikia.com
434	merriam-webster.com

Non-respect of choice

Tranco rank	domain
113	reuters.com
151	telegraph.co.uk
277	sindonews.com
373	softonic.com
562	wowhead.com
841	techtarget.com
1 085	makeuseof.com
1 376	bustle.com
1 420	filehippo.com
1 618	cdiscount.com

consent and refusing consent in the semi-automatic crawl. We observe 126 (22.9%) websites setting the shared cookie.

We then estimate how many websites access and reuse the shared cookie. We place a custom cookie (respecting the specification) in the browser, query the CMP using the standard APIs, and see if the CMP returns the exact same consent string (with no banner interaction). Using this protocol, 62 (4.3%) websites return the same consent strings. This means that CMPs on these websites reuse the shared cookie, even if it has been created by another CMP. This constitutes a lower bound, because CMPs can return another consent string than the one stored in the cookie, and may ignore ours for various reasons (e.g. an unexpected vendor list version).

We also estimate how many websites access the shared cookie by studying how many of them use the HTTP redirect mechanism described in section II-C to do so. We first observe that many consent redirecting domains do not respect the specification. Indeed, during manual inquiry, we find redirecting schemes using different values for the GET parameter specifying the redirection URL. For example, on `mirror.co.uk` we observed a GET request with a `gdpr_consent_string` parameter instead of `gdpr_consent`. As we cannot cover these cases exhaustively, we focus on those respecting the specification. The only domain we observe doing so (`sddan.consensu.org`, owned by the SIRDATA CMP) is used on 53 (9.5%) websites during the semi-automatic crawl. This hints that the practice of reading the shared consent cookie is quite common.

We observe 3 websites setting the shared cookie in the *Consent stored before choice* case, 3 in the *Non-respect of choice* case with 5 purposes, and 20 (3.9%) with 1 to 5 purposes.

Visiting one of the 3 websites on which the cookie is set before any user action on the banner will automatically set a global positive consent cookie. Visiting one of the 20 websites that do not respect user decision will set a global positive consent cookie against the user's decision. This is particularly troubling in terms of privacy: since this consent is reused among different publishers, it constitutes an escalation of the problem. We discuss this further in Section XI.

VIII. Measuring Third-Party Requests: Presence of Consent Strings and Third-Party Trackers

In previous sections, we studied violations in consent strings obtained via the standard API and shared cookies, as described in Section II-C. Responsibility of such violations can be attributed to CMPs and publishers (see the discussion in Section XI). However, when we find a non-compliant consent string via a URL-based method, we have no way to know whether that consent string was legitimately transmitted by the CMP or any other third party present on the page.

In this section, we study third-party requests observed in the two crawls. We first analyse the consent strings transmitted via URL-based methods, and then measure how many third-party trackers are present on the page before user actions, after acceptance and after refusal of consent.

A. Third-Party Requests with Consent Strings

In this section, we detect the four suspected GDPR and ePD violations by analyzing consent strings that we observed in GET and POST requests to third parties.

We observed consent strings with positive consent (1 to 5 allowed purposes) in GET or POST requests before any user action on 151 (10.6%) websites out of 22 949 websites in the automatic crawl – this indicates websites with a *Consent stored before choice* violation. For the *Non-respect of choice* violation, we intercepted consent strings in GET or POST requests with 5 purposes on 63 websites (12.4%). To evaluate whether these results are complementary to our previous findings, we count the number of websites in which we see a violating consent string in GET and POST requests, but do not obtain a violating consent string via intercepting the standard APIs or in the shared cookie.

***Consent stored before choice*:** In addition to 66 websites where we observed this violation while intercepting consent strings using the standard APIs and the shared cookie, we observed it also on additional 69 websites, where GET or POST requests send consent strings with a positive consent (1 to 5 purposes). It means that requests containing violating consent strings are sent while the CMP has not provided a consent string yet.

***Non-respect of choice*:** In addition to 27 websites where we observed this violation while intercepting consent strings with the standard APIs and the shared cookie, we observed it also on additional 26 websites where we obtain consent strings with all 5 purposes in GET and POST requests.

We further investigated whether the identifiers of the responsible CMPs (CMP ID) for each consent string obtained via GET and POST requests match the CMP IDs obtained from consent strings with the standard APIs and the shared cookie. We found CMP IDs in GET and POST requests different from the ones found using the standard APIs on 48 websites. In 37 of them, both CMP IDs found were from valid CMPs, while

TABLE VII: Average number of third-party tracking requests per website before user action, after a positive and after refusing consent.

User action	Number of third party tracking requests	Total number of third party requests
Before user action	22.54	35.04
After refusing consent	28.78	42.50
After a positive consent	39.59	56.75

in the remaining 11 websites, CMP IDs were set to either 0, 1 or 4095, which do not exist in the CMP public list [27]. It seems suspicious that consent strings not created by the website's actual CMP (or even non-existent CMPs) are sent to third parties.

B. *Third-party trackers*

We measure the number of third-party trackers on websites with TCF banners depending on user consent: before any user action, after refusing consent and after a positive consent. To do so, we logged every request to third-party domains with *Cookinspect*. From this, we extract domains which are considered trackers in the Disconnect list [10].

We first measure the number of third-party tracking requests without responding to the cookie banner or doing any other action on the website. Then we count third-party tracking requests after both giving a positive consent and refusing consent to the cookie banner (for websites on which it is possible), and reloading the page. Each measurement of trackers is done in a single browser session, on a single page load. These tests are done on the 508 websites on which refusing consent is possible in the semi-automatic crawl.

Table VII summarizes the results. We observe that giving consent on TCF banners, whether it's a positive or a refusal, has an effect on the number of included third-party trackers. Surprisingly, even refusing consent increases the number of tracking requests. The number of websites having the *Non-respect of choice* violation (and hence setting a positive consent even if the user refused) is not sufficient to explain this increase. We estimate that some scripts, in order to execute and include content, wait for the `__cmp()` function to be defined, which should only happen after the user has given their choice to the banner [29].

Table VIII shows the top 10 companies that own tracking domains present on websites after refusing consent (and a page reload). We matched tracking domains to company names using the Disconnect list [10]. We find whether they are part of the TCF by checking if any company name linked to a tracker domain in WebXRay's database [40] is present in the Global Vendor List (version 168). Some top trackers belong to vendors which are not part of the IAB framework (Google, Facebook or Amazon), but the rest of them are (eg. AppNexus, The Rubicon Project, comScore, etc.).

During our study, we encountered many unusual cases, detailed in appendix F.

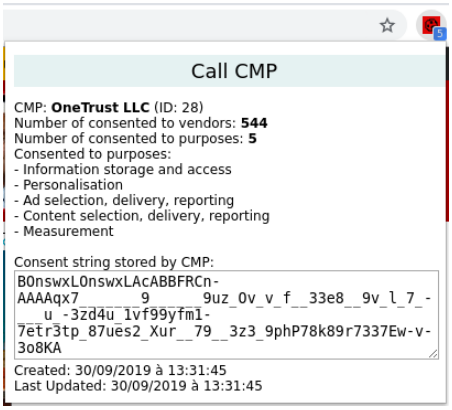
IX. Browser Extension

We publish a browser extension called *Cookie Glasses* [42], to enable users to see if consent stored by CMPs corresponds to their choice. Users can read information stored in the consent string provided by the CMP in a simple interface. The most important pieces of information present in the consent string are decoded and displayed in a readable format (see Figure 8).

TABLE VIII: Top 20 tracking companies observed on 508 websites after refusing consent and a page reload.

Tracking company	Number of websites	TCF Vendor?
Google	491 (96.7%)	
AppNexus	356 (70.1%)	✓
Facebook	337 (66.3%)	
RubiconProject	299 (58.9%)	✓
comScore	280 (55.1%)	✓
Integral Ad Science	258 (50.8%)	✓
Amazon.com	239 (47.0%)	
Casale Media	237 (46.7%)	✓
Criteo	232 (45.7%)	✓
Adform	230 (45.3%)	✓
Yahoo!	221 (43.5%)	✓
OpenX	217 (42.7%)	✓
The Trade Desk	217 (42.7%)	✓
Quantcast	202 (39.8%)	✓
MediaMath	199 (39.2%)	✓
DataXu	192 (37.8%)	✓
Adobe	190 (37.4%)	✓
PubMatic	186 (36.6%)	✓
SmartAdServer	179 (35.2%)	✓
SiteScout	165 (32.5%)	✓

Fig. 8: Interface of our browser extension, *Cookie Glasses*, displaying the consent transmitted by CMPs to advertisers.



Technically, the extension uses postMessages from the standard APIs (see ❷ in Figure 5). It is not possible to use direct calls to the `__cmp()` function because of browser security mechanisms. Our tests show that 79% of TCF websites use the postMessage API. Our extension therefore works on a majority of websites. For the remaining websites, we propose a workaround to manually execute the JS code to obtain the consent string, and decode it with the browser extension.

X. Limitations

Our work has some limitations. First, our scope is limited to banners of IAB Europe's TCF. Since we do not detect other cookie banners, we only observe a subset of all cookie banners. Besides, our results on the prevalence of TCF banners represent a lower bound on the actual usage of TCF banners, due to a variety of implementations of the TCF. For instance, some banners do not define the `__cmp()` function on the first page load. In one of its banners (e.g. on `aol.com`), the Oath CMP redirects the user to another website (of a different

domain) to display a consent wall. On this page, the `__cmp()` function is not displayed. We do not detect such cases in our study. While we detect TCF banners on 17.1% of `.fr` websites, van Eijk et al. [11] found that 40.2% of European websites have a cookie banner.

Secondly, results of our semi-automatic crawl are prone to errors due to dark patterns. Most banners we encountered nudge users towards accepting consent: some of them make it particularly difficult to opt out[9]. As a consequence, results of our semi-automatic crawls are prone to errors due to these dark patterns. To limit such errors, we cross-checked answers to banners by three human operators (see Section V-C). Nonetheless, it is still possible that some banners are designed in such a confusing way that the three persons failed to find the proper way to opt out. We argue that banners that give a technical mean to refuse consent, but make it so difficult that three computer science researchers do not find it, are still in violation with the GDPR.

Finally, we only detect violation in client-side consent strings. Yet, exchanges of consent strings can also happen outside of the browser. IAB provides extension fields [33] for exchanging consent string in its OpenRTB protocol [24]. This protocol is used between ad exchanges and advertisers for Real-Time Bidding. As such communication happens server-to-server, we cannot observe it with a client-side approach.

XI. Discussion

In this section, we reflect on our experiments and our results and comment on open problems that can be addressed by legal professionals or DPAs.

A. Who is responsible for violations in cookie banners?

It is a complex task to attribute the responsibility of a non-compliant cookie banner on a website to either the CMP or the publisher. CMPs often propose different versions of their banners that have different legal implications, and provide a documentation on customizing the banner. For instance, OneTrust, on its webpage presenting its CMP solution [47], proposes publishers to "maximize user opt-in with customizable publisher-specific cookie banners [...] to optimize consent collection". We argue that CMPs providing non-compliant cookie banners cannot exonerate themselves and delegate responsibility to the publishers that include them, especially when they claim to provide GDPR- and ePrivacy-compliant consent collection solutions. Conversely, publishers have a part of responsibility if they choose non-compliant banners. Hence, the responsibility of non-compliant cookie banners is shared between CMPs and publishers. CMPs and publishers might even be considered co-controllers, but we leave this discussion to lawyers.

Moreover, it is possible that publishers customize the banner when they host the CMP script in their website, modifying the original behaviour offered by the CMP. In such a case, the responsibility of a violating banner should be attributed to the publisher. Such cases can be detected with extensive case-by-case manual inspection.

[9]For instance, an uncolored link is hidden in the middle of a 28 000-word-long privacy policy on `liberation.fr`, accessed on October 25th 2019).

B. Problem of shared consent across publishers

The TCF defines a "global" cookie that is writable and readable by all CMPs (see section II-B). We found such an example on `letudiant.fr`: it obtains the consent string set on the website `senscritique.com` previously visited by the user[10]. This behaviour may not be a violation of the GDPR in itself: consent must be specific to a given purpose, not to a publisher. However, it seems suspicious that, even while obtaining the consent string invisibly for the end user, `letudiant.fr` still displays the cookie banner to the user. This may be considered as an excessive request (publishers ask for a response on consent they already have) and a lack of transparency or a deception (user is tricked into thinking `letudiant.fr` does not have their consent)[11]. In fact, in a report about dark patterns, the CNIL already uses the terms "bordering on harassment" to describe the repetitive request for consent on every website, even without any shared consent consideration [7]. The global consent has been criticized by the Privacy International NGO [50], which denounced the lack of users' knowledge that consent is global to websites, and that opting out is near impossible. This concept of global consent requires further analysis by legal experts.

Additionally, such a design in the TCF assumes that all the CMPs who use the global consent mechanism trust each other on setting the consent string accordingly to the choices made by the user. But a TCF-wide problem would arise if one publisher or CMP set this cookie incorrectly, violating user's consent. We found that this was not a hypothetical scenario: we detected 3 websites that set a positive consent in the shared cookie before the user makes any choice in the banner and 20 websites doing so after the user explicitly refused consent (more details in Section VII-B).

C. Unclear purposes in IAB Europe's TCF

The TCF proposes five pre-defined purposes (reproduced in Table IX in the Appendix): we leave for discussion to legal professionals whether defined purposes are explicit and specific. The CNIL has already pronounced that the TCF defined its purposes in a vague, imprecise way, in the decision against the Vectaury company [6].

D. IAB Europe TCF version 2

We have not observed any application of the IAB Europe TCF version 2 that was announced in August 2019. This new version introduces 12 purposes for data processing, and adds more flexibility to choose a legal basis (consent or legitimate interest). Since the implementation of the framework by CMPs and publishers is responsible for these violations, they might still occur on websites with CMPs that implement TCF v2.

E. Consent strings can be created by anyone

As shown in Section VIII-A, we observe on 37 websites requests to third parties containing consent strings that we suspect are being forged by non-CMP scripts running on the page, because they contain a CMP identifier that doesn't

[10]We show a video of this in attachment [3].

[11]However, the appearance of the banner on `letudiant.fr` could be a mistake and not an explicit deception technique.

correspond to the CMP present on the page. Even though the whole purpose of the TCF is to provide a way for actors in the advertisement industry to prove that they received consent from the user, we state that this proof is weak. The consent string's format does not contain any cryptographic proof that it was created by a given CMP, on a given website, in concordance with the user's choice. Consent strings can be forged by anyone, as our observation shows. Such consent strings have been flagged as "consent fraud" by rogue third parties by an actor of the online advertising ecosystem [22].

XII. Related Work

The first lines of research on cookie banners published before the GDPR laid on the legal basis of the ePD and its implementation in various European countries, and were very country-specific. As the GDPR changed behaviour regarding cookies [41], [53], trackers and other third-party content [41] and cookie banners [11], [9], we precise if each work made measurements before or after its enforcement (May 2018).

To the best of our knowledge, the following works measured prevalence of cookie banners after GDPR. Sanchez et al. [53] studied the impact of the GDPR on cookie-setting practices. They found that the GDPR had a global impact, influencing even US-based websites. Similarly to our semi-automatic crawl, they manually refused consent on 2 000 cookie banners to extract statistics such as the number of cookies after consent refusal. Van Eijk et al. [11] studied the impact on user's location on cookie banners. Using a crowd-sourced list, they automatically detected cookie banners on 40.2% of European websites. They found that the provenance of the user has not so much impact as the expected audience of a website regarding the prevalence of banners. They also observed important variations between websites of different top-level domains. Degeling et al. [9] studied characteristics of 31 cookie banner libraries, including several onesprovided by CMPs of the TCF, by installing them locally. They found that 62.2% of European websites displayed cookie banners in October 2018. The authors observed a 16% increase in cookie banners adoption by website pre- and post-GDPR. Nouwens et al. [45] studied dark patterns in 5 popular CMPs of IAB Europe's TCF. They estimate that only 11.8% of banners meet some minimum requirements of European law. They also study how banners design affect users' choice, and notably finds that the absence of a "refuse" button on the first layer of the banner increases positive consent by about 22%.

Similarly to our *No way to opt out* violation, some works measured how many banners offer no way to opt out [9]. In 2015, Leenes and Kosta found this issue on 25% of 100 Dutch websites [38]. The same year, the Article 29 Working Party found it on 54% of top 250 websites of 8 EU countries [2]. Vallina et al. found cookie banners offering no option to refuse consent on 1.36% of porn websites [59].

Several works measured the influence of cookie banners on the number of trackers or cookies. Before the GDPR, Carpineto et al. [4] measured how many websites set cookies without displaying a banner. Traverso et al. [56] measured the number of trackers before and after giving a positive consent on banners on 100 Italians websites. They found an average of 29.5 trackers prior to giving consent. In 2016, Englehardt and Narayanan [12] found 18 third parties per websites prior to any consent. In 2017, Trevisan et al. [57] found that 49% of websites installed profiling cookies before user consent, and that 80.5% of websites installing tracking cookies did not wait for user's consent to do so. After the GDPR came in force, Sanchez et al. [53] measured the number of cookies after refusing consent on banners. Instead, in our work, we measure trackers both before and after both giving a positive consent and refusing consent.

On the legal side, some regulators have already been active. The French DPA (CNIL) sued an advertisement company that used the TCF, invoking a lack of informed, free, specific and unambiguous consent [52]. In early 2020, they published a project document for guidelines on cookie banners [8] and criticized the TCF in a blog post [49]. They also noted that pre-selecting consent checkboxes was not compliant with the Article 32 of the GDPR. The European Court of Justice's decision in the Planet49 [14] case recently settled that pre-selecting options was not GDPR-compliant. Finally, the UK's DPA (ICO) [36] recently published a report on adtech and Real-Time Bidding, studying both IAB Europe's TCF and Google's framework. Among other considerations, they concluded that the TCF lacked transparency and observed a systemic lack of compliance to their data protection requirements.

XIII. Conclusions

In this paper, we have systematically studied cookie banners of IAB Europe's Transparency and Consent Framework (TCF). By collaborating with a researcher in law (one of the co-authors), we have identified four legal violations of both the GDPR and the ePD and we have detected them on 1 426 European websites that use TCF cookie banners. We have detected at least one of these suspected violations in 54% of websites. Finally, to help users and Data Protection Authorities (DPAs) further investigate these violations, we provide a browser extension called *Cookie Glasses*, that is able to detect some of them.

Beyond suspected violations in the implementations of cookies banners, we believe that the TCF suffers from several problems open for discussion and improvement. First, the consent string format has an unclear semantics, which makes it hard to interpret and use by the third parties that rely on such consent. Second, the TCF does not provide guidelines on which actors who obtain user consent (assuming it was obtained in a compliant way) are supposed to respect it: should the publishers, CMPs or some other actors ensure that the third parties respect consent they received? We believe that European regulators should take a more active stand regarding the implementation of cookie banners: either with supportive actions, such as guidance, or regulatory decisions and associated fines.

Acknowledgment

The authors would like to thank Imane Fouad for helping with the verification procedure in cookie banners. This work has been partially supported by the ANR JCJC project PrivaWeb (ANR-18-CE39-0008), the ANSWER project PIA FSN2 No. P159564-2661789/ DOS0060094 between Inria and Qwant, and by the Inria DATA4US Exploratory Action project.

REFERENCES

[1] G. Acar, M. Juárez, N. Nikiforakis, C. Díaz, S. F. Gürses, F. Piessens, and B. Preneel, "Fpdetective: dusting the web for fingerprinters," in *Conference on Computer and Communications Security (CCS'13)*, 2013.

[2] Article 29 Working Party, "Cookie sweep combined analysis - report," https://ec.europa.eu/newsroom/article29/document.cfm?action=display&doc_id=56123, accessed on 2019.11.01, 2015.

[3] "Attachments to the paper (Dropbox repository)," https://www.dropbox.com/sh/fw8ubf23z3ai0ei/AABY4qRO3FKXeGELfiPGFHica.

[4] C. Carpineto, D. Lo Re, and G. Romano, "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*, 2016.

[5] CNIL, "Article 2 of the deliberation n2019-093 of july 4, 2019 adopting guidelines relating to the application of article 82 of the law of january 6, 1978 modified to the operations of reading or writing to a user's terminal (including cookies and other tracers) (corrigendum)," https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000038783337, accessed on 30 October, 2019.

[6] ——, "Décision n MED 2018-042 du 30 octobre 2018 mettant en demeure la société Vectaury," https://www.legifrance.gouv.fr/affichCnil.do?oldAction=rechExpCnil&id=CNILTEXT000037594451&fastReqId=974682228&fastPos=2, accessed on 31 October 2019.

[7] ——, "IP report: Shaping choices in the digital world," https://linc.cnil.fr/fr/ip-report-shaping-choices-digital-world, accessed on 2019.10.30, 2019.

[8] ——, "Projet de recommandation sur les modalités pratiques de recueil du consentement prévu par l'article 82 de la loi du 6janvier 1978 modifiée, concernant les opérations d'accès ou d'inscription d'informations dans le terminal d'un utilisateur (recommandation "cookies et autres traceurs")," https://www.cnil.fr/sites/default/files/atoms/files/projet_de_recommandation_cookies_et_autres_traceurs.pdf, accessed on 17 January 2020., 01 2020.

[9] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy," in *Network and Distributed System Security Symposium (NDSS)*, 2018.

[10] Disconnect, "disconnect-tracking-protection," https://github.com/disconnectme/disconnect-tracking-protection, accessed on 2019.07.16.

[11] R. v. Eijk, H. Asghari, P. Winter, and A. Narayanan, "The impact of user location on cookie notices (inside and outside of the European union)," in *Workshop on Technology and Consumer Protection (ConPro'19)*, 2019.

[12] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *conference on computer and communications security (CCS'13)*, 2016.

[13] "Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009," https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009L0136, accessed on 2019.10.31.

[14] European Court of Justice, "Judgement of the court of justice of the EU, Case c-673/17," http://curia.europa.eu/juris/document/document.jsf;?&docid=218462&doclang=EN&cid=8679428, accessed on 2019.10.31.

[15] European Data Protection Board, "Guidelines on consent under regulation 2016/679" (WP 259 rev.01), adopted on 10 april 2018," https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051.

[16] ——, "Opinion 03/2013 on purpose limitation (WP 203), adopted on 2 april 2013," https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf.

[17] ——, "Opinion 06/2014 on the notion of legitimate interests of the data controller under article 7 of directive 95/46/ec (WP 217)," https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_en.pdf.

[18] ——, "Opinion 15/2011 on the definition of consent (WP 187), adopted on 13 july 2011," https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf.

[19] ——, "Opinion 2/2010 on online behavioural advertising, 22 june 2010, WP 171, p. 10," https://ec.europa.eu/justice/article-29/documentation/opinionrecommendation/files/2007/wp136_en.pdf.

[20] ——, "Opinion 4/2007 on the concept of personal data (WP 136), adopted on 20.06.2007," https://ec.europa.eu/justice/article-29/documentation/opinionrecommendation/files/2007/wp136_en.pdf.

[21] ——, "Working document 02/2013 providing guidance on obtaining consent for cookies, adopted on 2 october 2013," https://www.pdpjournals.com/docs/88135.pdf.

[22] C. Grutchfield, "The adtech truth – don't mess with my consent string!" https://newdigitalage.co/2019/10/09/behind-the-curtain-the-adtech-truth-dont-mess-with-my-consent-string/amp/, accessed on 2020.02.04, 2019.

[23] IAB, "Consent string use cases," https://github.com/InteractiveAdvertisingBureau/Consent-String-SDK-JS/blob/master/consent_string_use_cases.md, accessed on 2020.02.07.

[24] IAB, "Openrtb (real-time bidding)," https://www.iab.com/guidelines/real-time-bidding-rtb-project/, accessed on 2019.09.16.

[25] ——, "Safeframe," https://www.iab.com/guidelines/safeframe/, accessed on 2019.09.16, 2014.

[26] IAB Europe, "CMP ID 1 is not currently assigned to a Consent Management Provider (CMP)," http://advertisingconsent.eu/2019/01/cmp-id-1-is-not-currently-assigned-to-a-consent-management-provider-cmp/, accessed on 2019.09.02.

[27] ——, "CMP list," https://advertisingconsent.eu/cmp-list/, downloaded in 2019.04.

[28] ——, "IAB europe transparency & consent framework policies," https://iabeurope.eu/wp-content/uploads/2019/08/IABEurope_TransparencyConsentFramework_v1-1_policy_FINAL.pdf, accessed on 2019.11.20.

[29] IAB Europe and IAB Tech Lab, "Consent management provider javascript API v1.1: Transparency & consent framework," https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework/blob/master/CMP%20JS%20API%20v1.1%20Final.md#API-provided, 04 2018.

[30] ——, "Transparency and consent framework," https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework, accessed on 2019.05.03, 04 2018.

[31] ——, "Global vendor list (gvl)," https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Frameworkhttps://vendorlist.consensu.org/vendorlist.json, accessed June 2019, 06 2019.

[32] IAB Tech Lab, "Transparency and consent framework: Consent string SDK (javascript)," https://github.com/InteractiveAdvertisingBureau/Consent-String-SDK-JS.

[33] ——, "OpenRTB advisory - GDPR," https://iabtechlab.com/wp-content/uploads/2018/02/OpenRTB_Advisory_GDPR_2018-02.pdf, accessed on 2019.10.16, 02 2018.

[34] IAB Tech Lab and IAB Europe, "GDPR consent passing for URL-based services: Transparency and consent framework," https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework/blob/master/URL-based%20Consent%20Passing_%20Framework%20Guidance.md, 04 2018.

[35] Information Commissioner's Office, "ICO guidance on the rules on use of cookies and similar technologies," https://ico.org.uk/media/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies-1-0.pdf.

[36] ——, "Update report into adtech and real time bidding," https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf, accessed on 2019.07.10, 2019.

[37] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: a research-oriented top sites ranking hardened against manipulation," in *Network and Distributed System Security Symposium (NDSS)*, 2019.

[38] R. Leenes and E. Kosta, "Taming the cookie monster with Dutch law - a tale of regulatory failure," *Computer Law & Security Review*, vol. 31, 2015.

[39] A. Lerner, A. K. Simpson, T. Kohno, and F. Roesner, "Internet Jones and the raiders of the lost trackers: An archaeological study of web tracking

from 1996 to 2016," in *25th USENIX Security Symposium (USENIX Security 16)*, 2016.

[40] T. Libert, "Exposing the hidden web: An analysis of third-party http requests on 1 million websites," *International Journal of Communication*, 2015.

[41] T. Libert, L. Graves, and R. K. Nielsen, "Changes in third-party content on european news websites after GDPR," Reuters Institute for the Study of Journalism, 2018.

[42] C. Matte, "Cookie glasses," https://github.com/Perdu/Cookie-Glasses, 2019.

[43] ——, "Cookinspect," https://github.com/Perdu/Cookinspect, 2020.

[44] N. Nikiforakis, A. Kapravelos, W. Joosen, C. Kruegel, F. Piessens, and G. Vigna, "Cookieless monster: Exploring the ecosystem of web-based device fingerprinting," in *IEEE Symposium on Security and Privacy (SP'13)*, 2013.

[45] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal, "Dark patterns after the gdpr: Scraping consent pop-ups and demonstrating their influence," in *ACM CHI Conference on Human Factors in Computing Systems*, 2020.

[46] L. Olejnik, M. Tran, and C. Castelluccia, "Selling off user privacy at auction," in *Network and Distributed System Security Symposium (NDSS'14)*, 2014.

[47] OneTrust, "Consent management publishers advertisers," https://www.onetrust.com/solutions/consent-management-platform/, accessed on 2019.10.15.

[48] P. Papadopoulos, N. Kourtellis, and E. P. Markatos, "Cookie synchronization: Everything you always wanted to know but were afraid to ask," in *The World Wide Web Conference (WWW'19)*, 2019.

[49] B. Poilvé, "Mécanismes et (r)écueil du consentement," https://linc.cnil.fr/mecanismes-et-recueil-du-consentement, access on 17 January 2020, Laboratoire d'Innovation Numérique de la CNIL, 01 2020.

[50] Privacy International, "Most cookie banners are annoying and deceptive. this is not consent." https://privacyinternational.org/explainer/2975/most-cookie-banners-are-annoying-and-deceptive-not-consent, accessed on 2019.08.12, 2019.

[51] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *USENIX Symposium on Networked Systems Design and Implementation (NSDI'12)*, 2012.

[52] J. Ryan, "French regulator shows deep flaws in IAB's consent framework and RTB," https://brave.com/cnil-consent-rtb/, accessed on 2019.03.28, 2018.

[53] I. Sanchez-Rola, M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Vervier, and I. Santos, "Can I opt out yet?: GDPR and the global illusion of cookie control," in *Asia Conference on Computer and Communications Security (AsiaCCS'19)*, 2019.

[54] The European Parliament and the Council of the European Union, "Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)," 2002.

[55] ——, "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.

[56] S. Traverso, M. Trevisan, L. Giannantoni, M. Mellia, and H. Metwalley, "Benchmark and comparison of tracker-blockers: Should you trust them?" in *Network Traffic Measurement and Analysis Conference (TMA'17)*, 2017.

[57] M. Trevisan, S. Traverso, E. Bassi, and M. Mellia, "4 years of EU cookie law: Results and lessons learned," *Proceedings on Privacy Enhancing Technologies Symposium (PETS'19)*, 2019.

[58] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(un)informed consent: Studying gdpr consent notices in the field," in *Conference on Computer and Communications Security (CCS'19)*, 2019.

[59] P. Vallina, Á. Feal, J. Gamba, N. Vallina-Rodriguez, and A. F. Anta, "Tales from the porn: A comprehensive privacy analysis of the web porn ecosystem," in *Proceedings of the Internet Measurement Conference (ICM'19)*, 2019.

APPENDIX A
PURPOSES DEFINED IN IAB EUROPE'S TCF

We reproduce purposes defined in the TCF in Table IX.

APPENDIX B
ATTACHMENTS

In a public repository [3], we provide files that are relevant to this work: the full list of websites for each suspected violation, and videos showing examples of them.

APPENDIX C
DATA FOR REPRODUCIBLE RESEARCH

For the sake of research reproducibility, we indicate all data relevant to this work in Table X.

For selecting the websites, we use Tranco to build lists [37]. Within Tranco, we select the following options: Alexa and Majestic lists. We don't use The Cisco Umbrella list because it is DNS-based, and may not be representative of web traffic. Likewise, we exclude the Quantcast list because it is based on US traffic only. We also select the option to remove domains flagged as dangerous by Google Safe Browsing

From Tranco's top 1 million list, we extract the first 1 000 websites of the top-level domain (TLD) of each European country, and 1 000 websites from country-independant TLDs: `.com`, `.eu` and `.org` on September 20th 2019.

APPENDIX D
PROCEDURE FOR THE HUMAN OPERATORS

In this section, we give the precise procedure that human operators had to follow to refuse consent and give a positive consent on the banners during the semi-automatic crawl.

First, we attempt to refuse consent. If there is a "refuse" button on the banner, we click it directly; otherwise, we open the banner's "parameters". There, we untick any purpose-related option (checkbox or slider), independently from the kind of option (including e.g. functional cookies). If there is a "refuse all" button, we click it even if options are unticked by default. When banners propose vendors-related options, we ignore them. When the banner does not possess a "parameters" button, but only a link to the privacy policy (such as on `liberation.fr`), we follow this link and attempt to find a way to refuse consent within a reasonable time for a common website user (10 seconds), then come back to the main page. If options to refuse consent are located on another page linked by the banner, we come back to the main page after refusing consent. We always close the banner after refusing consent, by clicking the button whose terminology least indicates that we allow tracking[12]. Once everything is done, we manually label whether we encountered different cases: pre-selected options[13], banner not appearing, non-functional banner, banner proposing

[12] Some banners, such as the one on `rtl.fr`, have both an "accept" button and a small cross to close it. The "accept" button is misleading, because it sets a consent string with all purposes set even if the purpose options are unticked, while the small cross does not.

[13] We consider that there is a *Pre-selected choices* violation according to the visual representation of options (pre-ticked boxes, sliders set to acceptance). Ambiguous cases such as `lefigaro.fr`, where neither "accept" nor "deny" is set by default are not considered a violation.

TABLE IX: Purposes defined in IAB Europe's TCF, accessible at https://register.consensu.org/, accessed on May 3rd, 2019.

Purpose number	Purpose name	Purpose description
1	Information storage and access	The storage of information, or access to information that is already stored, on your device such as advertising identifiers, device identifiers, cookies, and similar technologies.
2	Personalisation	The collection and processing of information about your use of this service to subsequently personalise advertising and/or content for you in other contexts, such as on other websites or apps, over time. Typically, the content of the site or app is used to make inferences about your interests, which inform future selection of advertising and/or content.
3	Ad selection, delivery, reporting	The collection of information, and combination with previously collected information, to select and deliver advertisements for you, and to measure the delivery and effectiveness of such advertisements. This includes using previously collected information about your interests to select ads, processing data about what advertisements were shown, how often they were shown, when and where they were shown, and whether you took any action related to the advertisement, including for example clicking an ad or making a purchase. This does not include personalisation, which is the collection and processing of information about your use of this service to subsequently personalise advertising and/or content for you in other contexts, such as websites or apps, over time.
4	Content selection, delivery, reporting	The collection of information, and combination with previously collected information, to select and deliver content for you, and to measure the delivery and effectiveness of such content. This includes using previously collected information about your interests to select content, processing data about what content was shown, how often or how long it was shown, when and where it was shown, and whether the you took any action related to the content, including for example clicking on content. This does not include personalisation, which is the collection and processing of information about your use of this service to subsequently personalise content and/or advertising for you in other contexts, such as websites or apps, over time.
5	Measurement	The collection of information about your use of the content, and combination with previously collected information, used to measure, understand, and report on your usage of the service. This does not include personalisation, the collection of information about your use of this service to subsequently personalise content and/or advertising for you in other contexts, i.e. on other service, such as websites or apps, over time.

TABLE X: Data for reproducible research

Software - Selenium	python-selenium 3.141.0-1
Software - Chromium	chromium 76.0.3809.100-1
Operating system	Arch Linux
Kernel (result of `uname -a`)	Linux 5.2.5-arch1-1-ARCH #1 SMP PREEMPT Wed Jul 31 08:30:34 UTC 2019 x86_64 GNU/Linux
User-Agent	Mozilla/5.0 (X11; Linux x86_64) AppleWebKit/537.36 (KHTML, like Gecko) HeadlessChrome/73.0.3683.103 Safari/537.36
Location	France
Tranco list	https://tranco-list.eu/list/4NKX/1000000, generated on 2019.09.20
Disconnect list commit	eb817fb1 (2019-12-10)
WebXRay commit	04c3c8e8 (2019-06-18)
Crawling date (automatic crawl)	2019-09-20 - 2019-09-21
Crawling date (semi-automatic crawl)	2019-09-23 - 10-01

no way to refuse consent (considering links inside the banner). In one case (`healthline.com`), the banner proposed a way to refuse consent, but access to the website was then refused. We mark such a case as *No way to opt out*.

Secondly, on a second browser session (or directly if there is no option to refuse consent on the banner), we accept tracking by clicking on the "accept" button, or close the banner when it is the only option (we close the banner in all cases).

If the banner does not appear on first load, we reload the website until the banner appears, up to 3 times.

APPENDIX E
ALTERNATIVE PRESENTATIONS OF THE RESULTS

We display results of violations observed in the semi-automatic crawl organized by country in Table XI, and organized by CMP in Table XII (for CMPs seen at least 5 times). This is interesting for DPAs, who can then see which CMPs to investigate in priority. We do not display results for the automatic crawl because we can only identify CMPs providing consent strings before consent in this case (which would introduce a bias, and only concern 21% of websites).

APPENDIX F
UNUSUAL CASES

We list unusual cases encountered during our whole study.

Multiple banners at once — We observed websites displaying two cookie banners, e.g. psicologiaymente.com or matchendirect.fr. On these two sites, each banner seems to follow different regulation (pre- or post-GDPR). Our guess is that publishers forgot to remove the oldest ones.

Multiple banners on different loads — We encountered one specific website (kayak.fr) displaying 4 different banners under different clean browser sessions. These banners provide different characteristics (consent wall or not, existence of a refuse button, access to more specific configurations). Similarly, public.fr displays 2 different banners when loaded several times with a clean browser: one allowing parameters configuration, and one only providing an accept button.

Specifications not followed — CMPs on some websites do not respect the TCF's specifications at all. On dominos.fr, the `__cmp()` function is defined, but only ever returns an empty JSON object. express.co.uk sets 24 purposes in the consent string, even though only 5 of them are defined in the TCF and mentioned on the banner's text.

Banner not displayed on front page — On some websites, such as gamepedia.com, the banner is not displayed on the front page.

Redirections upon refusal — On some websites, such as `tvguide.co.uk`, users are redirected to the privacy policy page when (and only if) refusing consent. Even more questionably, `mon-programme-tv.be` redirects users automatically to Wikipedia if they refuse consent. On `toro.it`, users are redirected on the privacy policy page of another domain, itself containing a new cookie banner.

`consensu.org`'s page — While the `consensu.org` domain is used for global consent cookie sharing across

TABLE XI: Per-country violation tables. Quantification of suspected violations of the GDPR and the ePD encountered in the different CMPs seen at least 5 times in that country during the semi-automatic crawl (on `.fr`, `.uk`, `.it`, `.be`, `.ie` and `.com` websites), by CMP. The *Non-respect of choice* and *Pre-selected choices* tables only consider websites on which refusing consent was possible.

(a) .uk

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	60	0.0% (0/60)	0.0% (0/60)	55.0% (33/60)	0.0% (0/60)
Sourcepoint	21	0.0% (0/21)	0.0% (0/21)	100.0% (21/21)	0.0% (0/21)
OneTrust	16	81.2% (13/16)	6.2% (1/16)	93.3% (14/15)	6.7% (1/15)
Evidon	10	0.0% (0/10)	10.0% (1/10)	11.1% (1/9)	22.2% (2/9)
Ensighten	7	0.0% (0/7)	0.0% (0/7)	100.0% (7/7)	0.0% (0/7)
Oath	5	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)
others	18	22.2% (4/18)	16.7% (3/18)	33.3% (5/15)	0.0% (0/15)
No consent string found	12	0.0% (0/12)	8.3% (1/12)	45.5% (5/11)	0.0% (0/11)
all	**149**	**11.4% (17/149)**	**4.0% (6/149)**	**60.1% (86/143)**	**2.1% (3/143)**

(b) .fr

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	34	0.0% (0/34)	0.0% (0/34)	32.4% (11/34)	0.0% (0/34)
Didomi	29	0.0% (0/29)	0.0% (0/29)	24.1% (7/29)	0.0% (0/29)
Commanders Act	10	40.0% (4/10)	0.0% (0/10)	80.0% (8/10)	0.0% (0/10)
Sourcepoint	8	0.0% (0/8)	0.0% (0/8)	0.0% (0/8)	0.0% (0/8)
OneTrust	6	66.7% (4/6)	0.0% (0/6)	83.3% (5/6)	0.0% (0/6)
SIRDATA	5	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)
Chandago	5	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)
others	39	5.1% (2/39)	12.8% (5/39)	70.6% (24/34)	29.4% (10/34)
No consent string found	3	0.0% (0/3)	0.0% (0/3)	50.0% (1/2)	0.0% (0/2)
all	**139**	**7.2% (10/139)**	**3.6% (5/139)**	**42.1% (56/133)**	**7.5% (10/133)**

(c) .it

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	39	12.8% (5/39)	23.1% (9/39)	20.0% (6/30)	3.3% (1/30)
iubenda	20	0.0% (0/20)	0.0% (0/20)	0.0% (0/20)	0.0% (0/20)
Clickio	14	0.0% (0/14)	0.0% (0/14)	0.0% (0/14)	0.0% (0/14)
Triboo Media	10	0.0% (0/10)	0.0% (0/10)	0.0% (0/10)	0.0% (0/10)
OneTag	9	0.0% (0/9)	0.0% (0/9)	100.0% (9/9)	0.0% (0/9)
Axel Springer	6	83.3% (5/6)	83.3% (5/6)	100.0% (1/1)	0.0% (0/1)
others	22	9.1% (2/22)	4.5% (1/22)	66.7% (14/21)	9.5% (2/21)
No consent string found	3	0.0% (0/3)	33.3% (1/3)	0.0% (0/1)	0.0% (0/1)
all	**123**	**9.8% (12/123)**	**13.0% (16/123)**	**28.3% (30/106)**	**2.8% (3/106)**

(d) .be

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
OneTrust	9	100.0% (9/9)	0.0% (0/9)	77.8% (7/9)	11.1% (1/9)
Didomi	6	0.0% (0/6)	0.0% (0/6)	100.0% (6/6)	0.0% (0/6)
Quantcast	5	20.0% (1/5)	0.0% (0/5)	60.0% (3/5)	0.0% (0/5)
others	3	66.7% (2/3)	33.3% (1/3)	0.0% (0/1)	0.0% (0/1)
No consent string found	4	0.0% (0/4)	50.0% (2/4)	100.0% (1/1)	0.0% (0/1)
all	**27**	**44.4% (12/27)**	**11.1% (3/27)**	**77.3% (17/22)**	**4.5% (1/22)**

(e) .ie

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	12	0.0% (0/12)	0.0% (0/12)	33.3% (4/12)	0.0% (0/12)
others	7	42.9% (3/7)	14.3% (1/7)	40.0% (2/5)	0.0% (0/5)
No consent string found	6	0.0% (0/6)	16.7% (1/6)	50.0% (1/2)	0.0% (0/2)
all	**25**	**12.0% (3/25)**	**8.0% (2/25)**	**36.8% (7/19)**	**0.0% (0/19)**

(f) .com

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	24	0.0% (0/24)	0.0% (0/24)	21.7% (5/23)	0.0% (0/23)
OneTrust	12	58.3% (7/12)	8.3% (1/12)	72.7% (8/11)	18.2% (2/11)
Evidon	8	0.0% (0/8)	12.5% (1/8)	42.9% (3/7)	28.6% (2/7)
Sourcepoint	5	20.0% (1/5)	0.0% (0/5)	20.0% (1/5)	20.0% (1/5)
others	36	8.3% (3/36)	5.6% (2/36)	58.8% (20/34)	14.7% (5/34)
No consent string found	12	0.0% (0/12)	16.7% (2/12)	60.0% (3/5)	0.0% (0/5)
all	**97**	**11.3% (11/97)**	**6.2% (6/97)**	**47.1% (40/85)**	**11.8% (10/85)**

TABLE XII: Quantification of suspected violations of the GDPR and the ePD encountered in the different CMPs seen at least 5 times during the semi-automatic crawl (on `.fr`, `.uk`, `.it`, `.be`, `.ie` and `.com` websites), by CMP. The *Non-respect of choice* and *Pre-selected choices* columns display results w.r.t. the number of websites on which refusing consent was possible.

CMP	Number of websites	Violations			
		Consent stored before choice	*No way to opt out*	*Pre-selected choices*	*Non-respect of choice*
Quantcast	174	3.4% (6/174)	5.2% (9/174)	37.8% (62/164)	0.6% (1/164)
OneTrust	50	74.0% (37/50)	4.0% (2/50)	83.3% (40/48)	8.3% (4/48)
Didomi	41	0.0% (0/41)	0.0% (0/41)	39.0% (16/41)	0.0% (0/41)
Sourcepoint	34	2.9% (1/34)	0.0% (0/34)	64.7% (22/34)	2.9% (1/34)
Evidon	22	0.0% (0/22)	22.7% (5/22)	25.0% (4/16)	25.0% (4/16)
iubenda	20	0.0% (0/20)	0.0% (0/20)	0.0% (0/20)	0.0% (0/20)
Clickio	14	0.0% (0/14)	0.0% (0/14)	0.0% (0/14)	0.0% (0/14)
Oath	12	0.0% (0/12)	0.0% (0/12)	16.7% (2/12)	0.0% (0/12)
Triboo Media	10	0.0% (0/10)	0.0% (0/10)	0.0% (0/10)	0.0% (0/10)
Commanders Act	10	40.0% (4/10)	0.0% (0/10)	80.0% (8/10)	0.0% (0/10)
Axel Springer	10	60.0% (6/10)	70.0% (7/10)	100.0% (3/3)	33.3% (1/3)
OneTag	9	0.0% (0/9)	0.0% (0/9)	100.0% (9/9)	0.0% (0/9)
Cookie Trust WG.	8	25.0% (2/8)	25.0% (2/8)	60.0% (3/5)	0.0% (0/5)
Conversant Europe	7	0.0% (0/7)	0.0% (0/7)	100.0% (7/7)	0.0% (0/7)
Ensighten	7	0.0% (0/7)	0.0% (0/7)	100.0% (7/7)	0.0% (0/7)
SIRDATA	5	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)
Chandago	5	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)	0.0% (0/5)
incorrect CMP ID	9	11.1% (1/9)	11.1% (1/9)	62.5% (5/8)	12.5% (1/8)
others	73	11.0% (8/73)	6.8% (5/73)	54.4% (37/68)	22.1% (15/68)
No consent string found	40	0.0% (0/40)	17.5% (7/40)	50.0% (11/22)	0.0% (0/22)
all	**560**	**11.6% (65/560)**	**6.8% (38/560)**	**46.5% (236/508)**	**5.3% (27/508)**

publishers and for consent redirection through its subdomains, its main web pages https://consensu.org and https://www.consensu.org display a generic park page.

Claiming GDPR does not apply — The URL-based consent passing method specification [34] includes a parameter called `gdpr`, used to indicate whether GDPR applies. We observe many queries setting this parameter to 0, claiming that GDPR does not apply. As there are many reasons for the GDPR not to apply to a given script, we cannot decide whether such claims are founded.

Extremely tiresome cases — During our semi-automatic crawl, some banners were extremely hard to configure. For instance, the one on `rtl.fr` will display 8 purposes separated by hundreds of vendors, making it hard to disable each purpose. Furthermore, each vendor in each list is preticked, making it extremely tiresome to disable each of them.

Unticked options ambiguity — Some banners, e.g. Quantcast's banner on `sciencesetavenir.fr`, show unticked options when parameters are opened. However, a consent refusal is set upon saving, while a positive consent is set if user accepts without opening the parameters. This can lure users into thinking they have nothing to do to refuse consent, while they actually have to open the parameters to do so.

No choice before acceptance — Some banners, e.g. Evidon's banner on `ticketweb.co.uk`, only give the option to define consent preferences *after* user has accepted tracking: the banner only displays an "accept" button, and reveals the parameters button once this accept button has been clicked.

Hidden parameters — On some websites, parameters to refuse consent are hidden into a long cookie policy document linked by the cookie banner. For instance, on `liberation.fr`, the link to open these parameters is hidden in the middle of a 12 000-word-long policy document and is visually indistinct from the rest of the text.

No implementation — Some websites display a banner of one of the TCF-affiliated CMPs, but do not implement elements from the specification. For instance, `dominos.fr` displays a classical OneTrust banner, but does not provide a `__cmp()` function nor a `__cmpLocator` iframe. We cannot detect these cases in our automatic crawl.

Wrong CMP id — We observe the following incorrect CMP IDs in consent strings: 1, 0 and 4095 (resp. 155, 45 and 3 websites). As of September 2nd 2019, identifiers in IAB Europe's public CMP list [27] range from 2 to 265. IAB Europe stated that CMP ID 1 is incorrect and should not be used [26], which indicates that this is clearly a violation of the TCF. While some CMPs always return a consent string containing an invalid CMP ID, some CMPs only do so before users give their consent, e.g. Conversant Europe on `inc.com`.

Broken banner – We observe banners on which either refusing or accepting consent is not possible due to a bug on 6 websites. Ex: `olympia.ie`

Consent to nonexistent vendors — Some CMPs set consent for nonexistent vendors in the consent string. For instance, the CMP on `mycanal.fr` sets vendor IDs from 1 to 2000, even though vendor identifiers go up to a maximum of 670 in the GVL (as of September 2019). We observe this issue on 114 (20%) websites in the semi-automatic crawl.

HTTP only — 95 (7%) TCF-websites only provide an HTTP access. It is worrisome that websites using tracking technologies do so on an unencrypted connection.

Unusual consent verification — While monitoring consent verification made by third parties (using browser extensions to override the `__cmp()` function to catch direct calls, monitor postMessages, GET and POST requests), we observe third parties unregistered in the TCF doing so. We detect if third parties are trackers using the Disconnect list. We observe at least one tracker unregistered in the TCF querying the CMP to obtain consent in 44% of websites, and at least one third-party unregistered in the TCF querying the CMP to obtain consent in 55% of websites. It is unclear why vendors would verify consent if they're not registered to the framework.