A First View of Topics API Usage in the Wild

A First View of Topics API Usage in the Wild

Alberto Verna
Politecnico di Torino
Torino, Italy
alberto.verna@studenti.polito.it

Nikhil Jha
Politecnico di Torino
Torino, Italy
nikhil.jha@polito.it

Martino Trevisan
University of Trieste
Trieste, Italy
martino.trevisan@dia.units.it

Marco Mellia
Politecnico di Torino
Torino, Italy
marco.mellia@polito.it

ABSTRACT

Among several proposals for a privacy-preserving replacement of third-party cookies, Google's new Topics API is widely discussed as a possible solution. Some scepticism still lingers on the new paradigm from researchers and control bodies; however, the industry has started deploying and testing it. This paper measures the current usage of the Topics API in the wild, discovering third parties that started enabling it, the practices they adopt and their interplay with privacy policies and consent acquisition mechanisms.

To do so, we deploy a crawler to record the usage of Topics API on the most popular 50,000 websites worldwide. We observe that this technology starts to get a foothold — with 47 popular ad-related third parties that are testing it to understand the opportunities it offers.

We notably observe typical problems of early deployments: privacy regulation violations, unexpected solutions that allow one to circumvent abuse protection, and deployment errors.

CCS CONCEPTS

• **Information systems** → **Web mining**; • **Security and privacy** → **Privacy protections**.

KEYWORDS

Topics API, Web Privacy, Web measurement

ACM Reference Format:

Alberto Verna, Nikhil Jha, Martino Trevisan, and Marco Mellia. 2024. A First View of Topics API Usage in the Wild. In *Proceedings of the 20th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '24), December 9–12, 2024, Los Angeles, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3680121.3697810

1 INTRODUCTION

Since the birth of online advertising, advertisers and trackers have followed users' browsing habits using cookies, small chunks of text installed in the client's browser, which allow the server to identify the same user on subsequent visits. Third-party cookies — i.e., cookies set by a domain other than the one a user is currently visiting — allow trackers to follow the user on different websites, reconstruct their browsing patterns [21, 25, 31], and build their profiles to ultimately provide targeted ads and personalised content. This approach impacts users' privacy and still sparks debates and countermeasures — i.e. tracker blockers [1, 4, 7], privacy-friendly browsers and search engines (e.g., [2, 5]). Recently, browsers started blocking third-party cookies [6, 11], with Chrome being among the few ones to still admit their usage. Legislators faced unlimited data collection by mandating users to consent before the use of any personal information. GDPR [22], CCPA [19], or LGPD [18] are the most prominent and well-known laws.

This challenges the online ads ecosystem, and industries are looking for alternative paradigms. Google, as one of the largest players in this arena, has proposed new solutions, including the Topics API [13], a key component of its larger Privacy Sandbox framework [10]. The Topics API moves all the tracking activity inside the browser: it observes the sites the user visits, maps them into topics, and, when asked by an enabled third party, shares some of the topics.

The Topics API lets the advertisers access valuable information about topics the user is interested in, without giving access to private information such as the specific website or page the user visits. After some initial setbacks,[1] Google considered introducing the Topics API as part of the Privacy Sandbox for 1% of the Chrome users in the first quarter of 2024 and aims to complete the third-party cookies phase-out by the end of the year,[2] a deadline that was later postponed not earlier than 2025.[3] Users who are not automatically selected can autonomously opt in by activating a flag in Chrome's settings.

Websites and third parties have started to experiment with the Topics API. But what does the actual picture look like? To the best of our knowledge, this work is the first to offer a view of the global usage of the Topics API. We instrument a headless browser to collect the usage of Topics API and run a carefully engineered measurement campaign visiting the top-ranked 50,000 websites. We observe which players use the Topics API, identifying interesting patterns and witnessing unexpected facts as well. We observe even possibly illicit behaviour, such as requests for topics issued before the user gives consent, or the potential usage of Topics API by first and third parties not entitled to do so. In perspective, our work testifies how a new technique for behavioural advertising suddenly gained momentum, complementing the related work on the classical cookie-based approaches [21, 27, 28] and controversial techniques such as device fingerprinting [29, 30].

Some interesting facts emerge from our results:

- Popular ads platforms already adopt the Topics API and appear running live A/B tests to compare their effectiveness with the current cookie technology;

CoNEXT '24, December 9–12, 2024, Los Angeles, CA, USA

ACM ISBN 979-8-4007-1108-4/24/12
https://doi.org/10.1145/3680121.3697810

[1] https://blog.google/products/chrome/update-testing-privacy-sandbox-web/, accessed on October 16, 2024
[2] https://developers.google.com/privacy-sandbox/blog/cookie-countdown-2023oct, accessed on October 16, 2024
[3] https://privacysandbox.com/news/update-on-the-plan-for-phase-out-of-third-party-cookies-on-chrome/, accessed on October 16, 2024.

- A user encounters a party calling the Topics API in one website every two;
- The technology is still in its infancy, with inconsistent deployment, questionable integration with privacy regulations, and even erroneous support in the Chrome browser.

We believe our work, although preliminary, sheds some light on this new technology. We hope this will stimulate other researchers to explore it and monitor its deployment. Moreover, our work seeks to improve practitioners' awareness of the implications of this new technology, as the cases of incorrect use we observe are easily fixable, provided a minimal understanding of the Topics API operation. For this, we offer our tools and dataset to the community.[4]

2 METHODOLOGY AND DATASET

In this section, we describe how the Topics API works, and present how we engineered our custom web crawler and the dataset we gathered.

2.1 The Topics API

The Topics API is a component of the Privacy Sandbox, the new online advertisement ecosystem promoted and designed by Google to look for a new balance between offering valuable information to advertisers and respecting users' privacy.

The functioning of the Topics API is articulated as follows. First, the browser internally monitors the browsing activity of the user. During each *epoch* (currently one week), the browser collects the visited websites and assigns to each of them one or more labels, called *topics*, using a predefined language model. At the end of the epoch, the browser computes the top 5 most-visited topics and stores them in a list. All these processes happen inside the browser, so that no external entity has access to potentially private information. When a user visits a website, any service (e.g., advertisers) on the page can call the Javascript APIs to ask the browser for some topics the user is interested in. The browser returns three topics, one for each of the last three epochs, choosing each randomly from among the epoch's top 5 topics. We exemplify this mechanism with Figure 1.

The Topics API implements specific mechanisms to protect users' privacy: for instance, to add some *plausible deniability*, 5% of the offered topics are replaced by a random topic. This makes it difficult to build the user's profile and gives all topics a minimum exposure probability. To access the Topics API, developers need to complete an enrolment and attestation process. This provides a mechanism to verify which entities can call the API, adds transparency to who is accessing data, and mitigates attempts to misuse the API to gather more data than intended (see below for technical details).

Topics API are supported and implemented in Chromium and Chrome on both their mobile and desktop versions since Chrome version 101 of March 2022. Researchers showed that the Topics API is an improvement with respect to the old "all-allowed" cookie-tracking jungle, and some theoretical [16, 20] and practical [17, 23] results show to various extent that some privacy leak may still happen.

The privacy issues, the setbacks from rival firms' browsers such as Firefox and Safari (which are not implementing Topics API) and worries over the impact that third-party cookies disruption will have on the online advertising ecosystem are slowing down the large-scale introduction of the Topics API on the market.

[4]https://github.com/Novant8/priv-accept-topics

2.2 Data Collection

For our measurement campaign, we rely on a Selenium-based crawler to visit the top-50,000 websites according to the Tranco list [26], as of March 26th, 2024. We employ the Chromium browser version 122.0.6261.128 and manually opt in for the usage of the Topics API. For every visited website, we i) collect the URL of each first- and third-party object downloaded to render the page and ii) record every call to the Topics API by modifying Chromium's `BrowsingTopicsSiteDataManagerImpl` class. Such information includes the domain calling the Topics API — henceforth *Calling Party* (CP) —, the domain of the website on which the call happened, and the timestamp of the last call. We modify the handler to additionally log the API call type [14] (JavaScript, Fetch or IFrame) and record possible multiple calls from the same CP on the same webpage. We show an example of a JavaScript call in Figure 1.

As shown in [24], running reliable crawling campaigns in the wild requires ingenuity. In fact, we expect that the Topics API must follow the same regulatory framework that protects users' privacy — i.e., users have to agree to the privacy policy of a website and explicitly authorise the usage of any personal data. In particular, we run our crawling campaign from Europe, where the GDPR is in force. It clearly mandates any website to collect the user's explicit consent before using any personal information. Thus, during our crawling, we need to mimic the user who grants the usage of personal data by interacting with the Consent Banner shown during the user's first visit. We build on the *Priv-Accept* tool presented in [24] that automatically provides consent by interacting with Privacy Banners, if present. In a nutshell, for each website, we first visit it and record statistics *before* accepting the privacy policy; we then grant consent to personal data usage and, if successful, visit the site *after* acceptance. We delete the browser cache to load again all objects. We call these two visits *Before-Accept* and *After-Accept* as in [24].

Note that if we are not able to find a banner and allow the usage of personal data, we do not proceed with the *After-Accept* visit. This may happen because i) the banner is actually not present, or ii) *Priv-Accept* fails to recognise the "Accept" button.[5]

2.3 Authorised callers

As said, parties interested in the usage of Topics API must complete an onboarding process. To enforce this, the browser checks whether the CP is included in a `allow-list` file stored in the `privacy-sandbox-attestations.dat` file located in the `PrivacySandbox AttestationsPreloaded` folder. If present, the browser allows the call; otherwise, it blocks it. Every time the browser is opened, it updates the `allow-list` file. We call the parties which are present in the list as *Allowed*. We use the file obtained in June 6th, 2024.

[5]Priv-Accept looks for keywords and supports five languages – i.e., English, French, Spanish, German and Italian. The authors show that it is 92—95% accurate with banners in such languages.

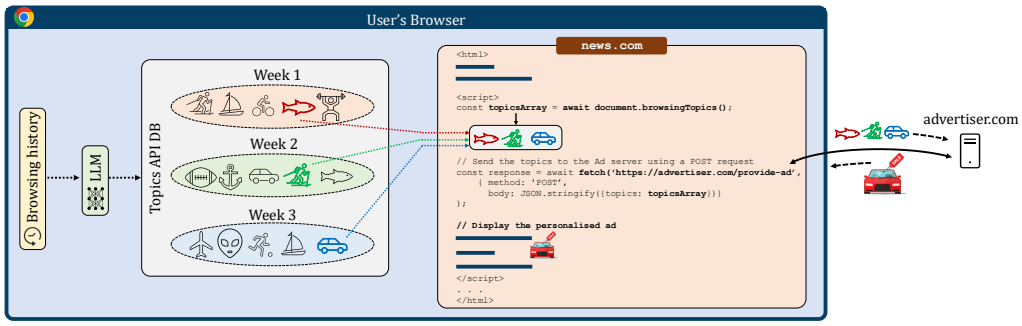

Figure 1: Topics API operation and use in a Javascript.

Table 1: Overall status of Topics API usage. In red, the anomalous usage. In blue, the questionable usage.

	Allowed	193
	Allowed & !Attested	**12**
D_{AA}	***Allowed & Attested***	47
D_{AA}	***!Allowed & Attested***	**1**
D_{AA}	***!Allowed***	**2,614**
D_{BA}	***Allowed & Attested***	**28**
D_{BA}	***!Allowed***	**1,308**

In addition, the Privacy Sandbox policy mandates all the CPs using the Topics API to offer an attestation JSON file in a predetermined URL path, namely <domain>/.well-known/privacy-sandbox-attestations.json. This attestation serves as a declaration by the CP that it will not use the Topics API for re-identification purposes and is considered by Google as part of the verification process for new enrolments [12]. For every first and third party we encounter (i.e., for every domain), we verify whether a valid attestation file is present. If so, we label the party as *Attested*.

During our experiments, we on purpose corrupted the local allow-list of our Chromium browser. Interestingly, we observe the browser allows any first and third parties to call the Topics API in this case. Investigating this, we found that the current implementation permits any Topics API calls as *default* case when the internal database is corrupted or missing. This implementation error would allow any caller to access the Topics API independently whether *Allowed* / *Attested* or not, permitting them to collect users' topics and possibly abuse this information.[6] By removing the `allow-list`, we can thus observe whether not-allowed callers are trying to request topics to the API.

2.4 Dataset and initial findings

We start our crawling on March 30th, 2024 to visit the top-50,000 websites in the Tranco list. The crawl ends after about one day. We successfully visit 43,405 websites[7] for which we obtain a *Before-Accept* visit (without providing any consent). We refer to this dataset as D_{BA}. It includes 19,534 unique third parties in addition to 43,405 first parties.

For 14,719 websites (about 30% of the active sites) *Priv-Accept* accepts the privacy policy and consents to the usage of personal information. For these, we execute an *After-Accept* visit and save the data in a dataset we call D_{AA}.[8]

Table 1 summarises our results:

- 193 domains are *Allowed*. These are the only ones allowed to use the Topics API and include popular advertisers.
- We check all these 193 services to see if they correctly expose the attestation file. 181 do. But 12 do not, erroneously.
- In our crawls, we encounter only 47 CPs that call the Topics API during the *After-Accept* visit (D_{AA} dataset). The 193-47=146 missing potential CPs may not have activated it, or we did not encounter them during our crawling.
- We find that one CP — namely distillery.com — has the attestation file timestamped on November 2023. Yet, it is not included in the `allow-list`. This possibly reflects the attestation process is still ongoing, or that Distillery has no interest in completing it.[9]
- Surprisingly, we observe thousands of websites and CPs that call the API even if they are not among the *Allowed* ones. We investigate this *anomalous* usage in Section 4.
- Considering the D_{BA}, we would expect no usage of the API because the user has yet to consent the privacy policy. However, we find 28 *Allowed* and *Attested* CPs that call the API even if the user did not consent. Astonishing, 1,308 CPs call the Topic API even if they are not *Allowed*, and before collecting the user consent. We will investigate this *questionable* usage in Section 5.

In a nutshell, we observe a very confused deployment, with apparent violations and questionable implementations. In the following, we dig into regular and unexpected cases.

3 LEGITIMATE USAGE

In this section, we offer a characterisation of the penetration of the Topics API inside the Web ecosystem. Here, we take into consideration only legitimate uses of the Topics API: hence, we only include interactions from the 47 CPs which are both in the *Attested* and the

[6]The actual feasibility of an attack goes beyond the scope of this paper. At the moment of writing, we have notified Google and Chromium developers about the error. They recognised the problem and declared to fix it in a future release

[7]The remaining websites fail due to domain name resolution or connection-related errors.

[8]This percentage is in line with [24]. In most of the cases, the website does not implement any banner, or *Priv-Accept* misses language or keyword.

[9]In fact, we observe it using the Topics API on the distillery.com website only, hinting at initial testing.

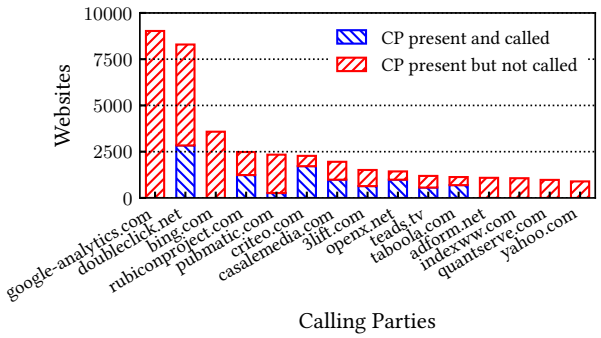

Figure 2: Number of websites where a CP is present and subset where it calls the Topics API. D_{AA} and all *Allowed* and *Attested* parties.

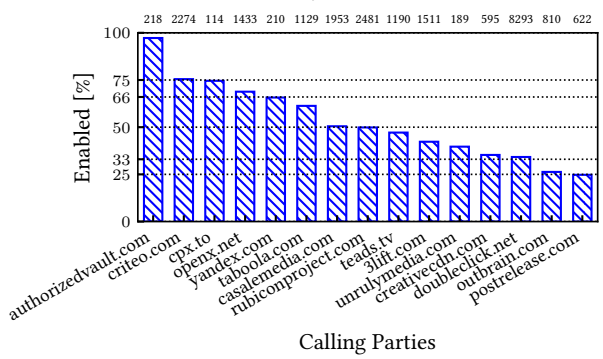

Figure 3: Fraction of times a CP calls the Topics API over the total times it is present on (in the top row). D_{AA} and *Allowed* and *Attested* services.

Allowed sets and that we encounter after successfully accepting the privacy policy (D_{AA} dataset).

Processing each CP attestation file, we observe the onboarding process for the use the Topics API by extracting the attestation certificate issue date. Enrolments kicked off in June 2023, the first attestation being on the 16th. Until May 2024 the enrolment process continues at a low pace: each month, approximately a dozen new services obtain the attestation for the Topics API. Curiously, on October 17th, 2024, many of the enrolled CPs had to update their attestations to include the new `enrollment_site` field and other minor changes.

We now focus on the extent to which popular ad-related platforms adopt the Topics API. In D_{AA}, we observe at least one call to the Topics API in 45% of visited websites. That is, every second website already hosts a CP — not surprising given the pervasiveness of ad services.

Figure 2 details the number of websites on which a given CP is present (red pattern). We show the top-15 most pervasive CPs. No surprise on the players. In blue we highlight the fraction of times in which a CP invokes the Topics API. google-analytics.com is curiously both *Attested* and *Allowed*. Yet, it never calls the Topics API (not being an ad-related service) while Google's doubleclick.net employs the Topics API on about one third of the websites we found it present. Conversely, bing.com (also *Allowed* and *Attested*) does not use the Topic API. criteo.com, rubiconproject.com, and casalemedia.com are leveraging the Topics API the most. Curiously, not enabling it on all websites.

In general, results show that all the major ad-related players started adopting the Topics API. Yet, we seldom observe consistent usage, hinting they are still in a testing phase. We next investigate this aspect.

Given a CP that uses the Topics API, we count the fraction of times it uses them over the total number of times we observe it. We show the CPs with the highest enabled percentage in Figure 3. We highlight some fractions on the y-axis to simplify reading the results. The top of the figure details the number of times we observe such a party. We notice a clustering of behaviours: for instance, authorizedvault.com, present on 218 websites, calls the Topics API almost every time. criteo.com and cpx.to call it 75% of times, yandex.com 66% of times, etc. We impute this to CP implementing some form of A/B tests, with percentages that look predetermined. They test how well the Topics API paradigm behaves compared with the standard third-party cookie solutions for their business metric: even if the Topics API is still below the surface, the most prominent advertising companies in the market are deeply studying its influence. This should convince legislators, privacy advocates, and the general public to closely follow the development process of this framework.

We run repeated tests to observe the policy some CPs use to enable/disable Topics API. We notice consistent alternating periods: for some time, CP, and website, the usage of the API is ON for all visits, followed by some time when it is OFF. This is consistent with some ongoing A/B tests that considers the same population and website but at different times.

4 ANOMALOUS USAGE

We now concentrate on the 2,614 CPs in D_{AA} ($\approx$ 11% of all domains seen) that access the Topics API even if they are not in the *Allowed* set. Recall that we observe them because we removed the `allow-list` in our crawler. With the correct configuration, the browser would not allow such a call.

First, we investigate in which context [3] the call is executed. Surprisingly, the CP is often not a third party but it coincides with the website we are visiting. Out of the 3,450 Topics API anomalous calls, 72% of them come from the website we are visiting — the website and CP second-level domains are the same, e.g., www.foo.com and ad.foo.net. A manual check on the remaining 28% reveals similar situations: i) the same company owns the two domains (e.g. windows.com and microsoft.com); ii) the visited website redirects to a second website which then calls the API — both websites being owned by the same company.

Second, all these bizarre calls use the JavaScript `browsingTopics()` function. This suggests some popular JavaScript libraries could erroneously access the Topics API. If loaded by a website, such a library would execute some calls from the website context (which is not *Allowed*).

To find a possible explanation, we observe the presence of Google Tag Manager's (GTM) [8] JavaScript scripts on 95% of the websites where anomalous calls occur. GTM, in fact, contains a call to the `browsingTopics()` function. The reason of this is unknown to us, being it neither *Allowed* nor *Attested*.

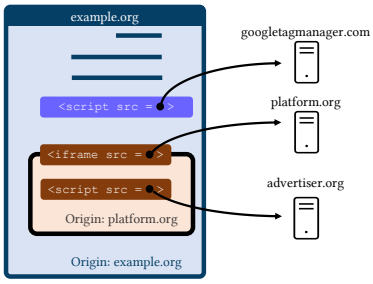

Figure 4: Example of the "Origin" mechanism with scripts and iframes.

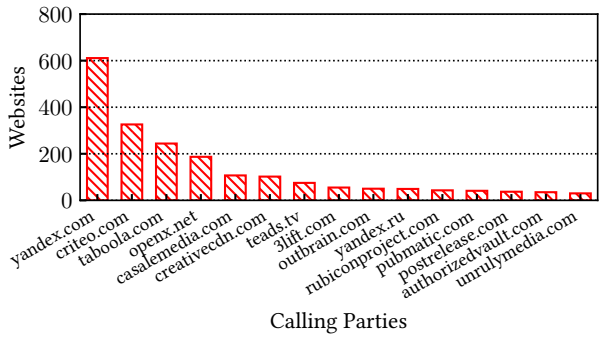

Figure 5: Number of questionable API calls by *Allowed* and *Attested* services. D_{BA} dataset.

Instead, we can explain why the call is executed as coming from the websites and not the GTM context. Indeed, the browser downloads the script from a Google server with a link similar to https://www.googletagmanager.com/gtm.js?id=<ID>. However, the script is executed within the *root* browsing context, resulting in having its context origin [9] set to the website instead of the GTM context. As sketched in Figure 4, this happens because the relevant `<script>` tag is placed directly inside the HTML content of the website's page and not included inside an `<iframe>` with an external source. In our setup with no `allow-list`, our crawler executes the Topic API call that appears as generated by the website itself.

The "wrong context" problem is general and could complicate the deployment of Topics API solutions.[10] This sort of behaviour suggests that websites implementing the Topics API will have to be very careful about the implementation of third parties (like the GTM), as they may cause unexpected and unwanted privacy issues.

5 QUESTIONABLE USAGE

We focus now on those Topics API calls that are performed in the *Before-Accept* visit. Ideally, we expect no API usage since we have not consented to the use of any personal data. However, as reported in Table 1, we observe more than 1,300 CPs. Given we run our crawling campaign from Europe, we appear as a European citizen protected by the GDPR [22]. The above cases are all questionable and can be seen as a violation of said regulations, as one could consider the Topics API usage equivalent to using cookies.[11]

If we restrict to the 47 CPs which are both *Attested* and *Allowed*, 28 of them call the Topics API in the *Before-Accept* visit. Figure 5 shows the number of websites where we observe a violation for a given CP. yandex.com comes first (611 calls in *Before-Accept*), even if it is not among the top callers (1,414 calls in *After-Accept*). In general, we observe little correlation with the service popularity. For instance, doubleclick.net, the top-1 caller, does not perform any call in *Before-Accept* (and more than 2,500 in *After-Accept*). This corroborates the assumption no call shall be issued in the *Before-Accept* visits.

[10]We contacted Google about this issue as well, but at the moment of writing we did not receive any response.

[11]Whether this could be considered an actual violation of the current legislation is outside of the scope of this paper. The fact that some services respect this interpretation reinforces our position.

At least two cases can justify this behaviour: (1) the website does not include any Privacy Banner — and privacy-invasive technologies can be used in every visit. This could be the case with a website outside EU.[12] (2) The website does not correctly implement a privacy banner, e.g., in a shallow-but-in-good-faith behaviour.

We investigate these cases by checking the top-level domain (TLD) of the websites where we observe a violation. We use the TLD as a coarse indication of website country. Here, we focus on the top 4 questionable CPs and break down their API calls by geographic region: `.com`, Japan (`.jp`), Russia (`.ru`), Europe Union (30 TLDs for EU countries where the GDPR is in force) and all the remainder of TLDs. The figure reports the share of websites in which the given CP invokes the Topics API over the number of websites the CP is embedded in. The top x-axis indicates the latter number. We first observe that the presence of CPs strongly varies in different regions. Yandex, a Russian company, is not present in Japan and almost absent in the EU. Conversely, Criteo, based in France, has a worldwide marketplace. Looking at the different bars, we do not identify any clear trend. While the sizeable differences among CPs can be caused by different deployment strategies, we do not identify radical diversity across the geographical regions. We even observe questionable API calls also for websites in the EU, where the GDPR definitively applies.

Next, we check if this questionable behaviour can be due to missing or incomplete configuration of the Privacy Banners by the website administrator. For this, we look for the Consent Management Platform (CMP) a website uses, if any. CMPs are commercial products which simplify the implementation of Privacy Banners. They offer standard libraries that control all the third parties embedded in the websites (such as advertisers or trackers), enabling them only after the user consents to the Privacy Policy. They require minimal configuration by the website administrator. In case this is incomplete, third parties can exhibit non-GDPR-compliant behaviour, i.e., being active in *Before-Accept* [24]. We assume the Topics API should ideally follow the same legislation as any other privacy-intrusive feature. Then, a website that adopts a CMP but allows CPs to call the Topics API on the *Before-Accept* visit (i.e.,

[12]This would still be a GDPR violation which protects Europeans even when accessing international services.

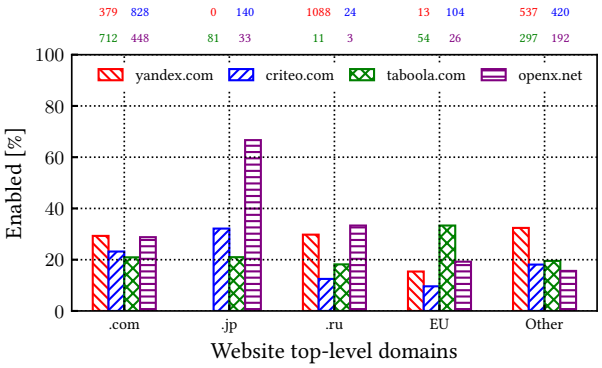

Figure 6: Share of websites where a CP calls the Topics API among all the websites where it appears (D_{BA}).

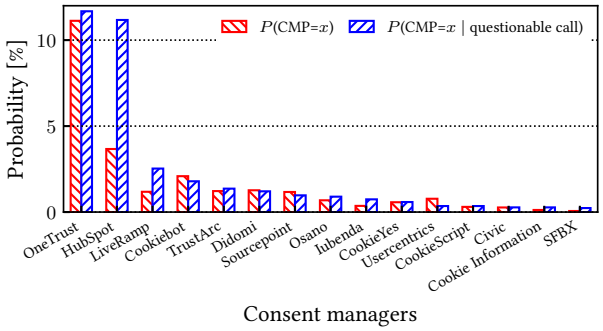

Figure 7: Probability of observing a CMP given (or not) a questionable API call (D_{BA}).

without user consent) is either due to a CMP misconfiguration or bad CMP implementation.

We check which CMP is in use when we visit a website. We rely on the list of the most widespread CMPs (identified by their domain name) offered by Wappalyzer [15]. In Figure 7, we show side-by-side the probability of observing a CMP over all websites ($P(CMP = x)$, red bars) and over websites where we observe a questionable Topics API call ($P(CMP = x|questionable\ call)$, blue bars). We conclude that the popularity of CMPs is generally independent of the presence of questionable calls, being the two probabilities equal: all CMPs suffer from the same problem. Some notable exceptions emerge: Hubspot has a probability of being the CMP in use given a questionable call which is $\approx 3\times$ the probability of observing it. As such, Hubspot does a bad job of properly handling the Topics API: the probability of a questionable call given the CMP is Hubspot is 12%, twice as big as the average probability. The same holds true for Liveramp.

In a nutshell, the complexity of configuring and managing the privacy options has yet to properly integrate the support for the Topics API, allowing possible violations of privacy regulations. This leaves space for more in-depth analysis that we leave for future work.

6 SHORTCOMINGS

Our work represents an initial effort to understand this new technology, but it has several limitations. First, our measurement methodology only detects invocations of the Topics API, and we do not examine how websites and advertisers utilize the retrieved topics (e.g., by providing different ads). This presents an interesting avenue for future research. Second, we provide a snapshot of Topics API usage in early 2024. Given the novelty of the technology, we are measuring early deployments of the Topics API, so our measurements should be conducted continuously to monitor how the technology evolves. Finally, our experiments were conducted from a single location in Europe, and we cannot rule out the possibility that websites may exhibit different behavior based on a user's location. Although these biases persist in the study, we minimised those introduced by the dynamic Web, for instance, taking into consideration the difference between *Before-Accept* and *After-Accept*.

7 CONCLUSION

This paper presented the first study on Topics API deployment. Using a carefully engineered custom-made crawler, we found that the most popular advertising platforms are already deploying and experimenting with the Topics API, in light of the forthcoming phase-out of third-party cookies. We have evidence that they are carrying out forms of A/B tests on controlled subsets of websites and users.

Interestingly, the problem of obtaining user's consent, which led to the proliferation of Privacy Banners and CMPs, recurs with the Topics API: a non-negligible portion of websites and third parties fail in properly handling this new technology, invoking the Topics API even when the user has not explicitly opted in.

We also find evidence of immature or incomplete implementations, which result in erroneous/anomalous Topics API invocations. In turn, we were able to discover such phenomena thanks to an issue in Chromium's Topics API implementation.

To the best of our knowledge, this is the first paper shedding light on the adoption of this new technology. Being proposed by one of the Internet giants, it is likely that the Topics API will become the *de facto* standard for behavioural advertising and one of the pillars of the future web ecosystem. Still, we testify how the deployment is still in an early stage, and the introduction of such a new technology entails shortcomings, bugs, and unexpected behaviours of which all the stakeholders in the system — advertisers, public opinion, privacy advocates, and Google itself — should be aware. Moreover, the commercial approval of this technology is still uncertain. Advertisers ground their business model on fine-grained user profiling, which allows them to track the user's interest in a specific field, brand or even product. The Topics API, which are explicitly designed to pose limits, may not be favourably welcome, thus, making long-term implications of this technology hard to foresee.

ACKNOWLEDGMENTS

This work has been partially supported by the Spoke 1 "FutureHPC & BigData" of ICSC - Centro Nazionale di Ricerca in High-Performance-Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU.

REFERENCES

[1] 2024. AdBlock Plus. https://adblockplus.org, accessed on October 16, 2024.
[2] 2024. Brave. https://brave.com/, accessed on October 16, 2024.
[3] 2024. Browsing context - MDN Web Docs Glossary: Definitions of Web-related terms | MDN. https://developer.mozilla.org/en-US/docs/Glossary/Browsing_context, accessed on October 16, 2024.
[4] 2024. Disconnect. https://disconnect.me, accessed on October 16, 2024.
[5] 2024. DuckDuckGo. https://duckduckgo.com/, accessed on October 16, 2024.
[6] 2024. Firefox. https://www.mozilla.org/en-US/firefox/, accessed on October 16, 2024.
[7] 2024. Ghostery. https://www.ghostery.com, accessed on October 16, 2024.
[8] 2024. Google Tag Manager. https://tagmanager.google.com/, accessed on October 16, 2024.
[9] 2024. Origin - MDN Web Docs Glossary: Definitions of Web-related terms | MDN. https://developer.mozilla.org/en-US/docs/Glossary/Origin, accessed on October 16, 2024.
[10] 2024. Privacy Sandbox. https://privacysandbox.com/, accessed on October 16, 2024.
[11] 2024. Safari. https://www.apple.com/safari/, accessed on October 16, 2024.
[12] 2024. The Privacy Sandbox enrollment attestation model. https://github.com/privacysandbox/attestation, accessed on October 16, 2024.
[13] 2024. Topics API. https://developers.google.com/privacy-sandbox/relevance/topics, accessed on October 16, 2024.
[14] 2024. Topics API integration guide | Privacy Sandbox | Google for Developers. https://developers.google.com/privacy-sandbox/relevance/topics/integration-guide#call_the_topics_api, accessed on October 16, 2024.
[15] 2024. Wappalyzer. https://www.wappalyzer.com/, Accessed on October 16, 2024.).
[16] Mário S. Alvim, Natasha Fernandes, Annabelle McIver, and Gabriel H. Nunes. 2023. A Quantitative Information Flow Analysis of the Topics API *(WPES '23)*. Association for Computing Machinery, New York, NY, USA, 123–127. https://doi.org/10.1145/3603216.3624959
[17] Yohan Beugin and Patrick McDaniel. 2024. Interest-disclosing Mechanisms for Advertising are Privacy-Exposing (not Preserving). In *Proceedings on Privacy Enhancing Technologies*. Vol. 2024. Privacy Enhancing Technologies Symposium, 41–57.
[18] Brazilian President of the Republic. 2018. Lei Geral de Proteção de Dados Pessoais. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm, accessed on October 16, 2024.
[19] California State Legislature. 2018. California Consumer Privacy Act of 2018. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375, accessed on October 16, 2024.
[20] CJ Carey, Travis Dick, Alessandro Epasto, Adel Javanmard, Josh Karlin, Shankar Kumar, Andres Muñoz Medina, Vahab Mirrokni, Gabriel Henrique Nunes, Sergei Vassilvitskii, and Peilin Zhong. 2023. Measuring Re-identification Risk. *Proc. ACM Manag. Data* 1, 2, Article 149 (jun 2023), 26 pages. https://doi.org/10.1145/3589294
[21] Steven Englehardt and Arvind Narayanan. 2016. Online Tracking: A 1-million-site Measurement and Analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* (Vienna, Austria) *(CCS '16)*. Association for Computing Machinery, New York, NY, USA, 1388–1401. https://doi.org/10.1145/2976749.2978313
[22] European Parliament and Council of European Union. 2016. Directive 95/46/EC. General Data Protection Regulation. http://data.consilium.europa.eu/doc/document/ST-5419-2016-INIT/en/pdf, accessed on October 16, 2024.
[23] Nikhil Jha, Martino Trevisan, Emilio Leonardi, and Marco Mellia. 2023. On the Robustness of Topics API to a Re-Identification Attack. In *Proceedings on Privacy Enhancing Technologies*. Vol. 2023. Privacy Enhancing Technologies Symposium, 66–78.
[24] Nikhil Jha, Martino Trevisan, Luca Vassio, and Marco Mellia. 2022. The Internet with Privacy Policies: Measuring The Web Upon Consent. *ACM Trans. Web* 16, 3, Article 15 (sep 2022), 24 pages. https://doi.org/10.1145/3555352
[25] Balachander Krishnamurthy and Craig Wills. 2009. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain) *(WWW '09)*. Association for Computing Machinery, New York, NY, USA, 541–550. https://doi.org/10.1145/1526709.1526782
[26] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Proceedings of the 26th Annual Network and Distributed System Security Symposium (NDSS 2019)*. https://doi.org/10.14722/ndss.2019.23386
[27] Jonathan R Mayer and John C Mitchell. 2012. Third-party web tracking: Policy and technology. In *2012 IEEE symposium on security and privacy*. IEEE, 413–427.
[28] Hassan Metwalley, Stefano Traverso, Marco Mellia, Stanislav Miskovic, and Mario Baldi. 2015. The online tracking horde: a view from passive measurements. In *International Workshop on Traffic Monitoring and Analysis*. Springer, 111–125.
[29] Emmanouil Papadogiannakis, Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos P. Markatos. 2021. *User Tracking in the Post-Cookie Era: How Websites Bypass GDPR Consent to Track Users*. Association for Computing Machinery, New York, NY, USA, 2130–2141. https://doi.org/10.1145/3442381.3450056
[30] Valentino Rizzo, Stefano Traverso, and Marco Mellia. 2021. Unveiling web fingerprinting in the wild via code mining and machine learning. *Proceedings on Privacy Enhancing Technologies* 2021, 1 (2021), 43–63.
[31] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending Against Third-Party Tracking on the Web. In *9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*. USENIX Association, San Jose, CA, 155–168. https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/roesner