Websites' Global Privacy Control Compliance at Scale and over Time

Katherine Hausladen[1], Oliver Wang[1], Sophie Eng[1], Jocelyn Wang[2], Francisca Wijaya[1], Matthew May[1], and Sebastian Zimmeck[1] *

[1]*Department of Mathematics and Computer Science, Wesleyan University*
[2]*Department of Computer Science, Princeton University*

Abstract

The California Consumer Privacy Act (CCPA) gives California residents the right to opt out of the sale or sharing of their personal information via Global Privacy Control (GPC). In this study we show how to evaluate websites' compliance with GPC. Using longitudinal data collected by crawling a set of 11,708 sites, we show the extent to which sites are respecting California residents' opt out rights expressed via GPC. We do so by examining the values of four privacy strings that indicate a web user's opt out status: the US Privacy String, the Global Privacy Platform String, the OptanonConsent cookie, and the `.wellknown/gpc.json`. We find that about a third of sites that have evidence of selling or sharing personal information per the CCPA implement at least one of the four privacy strings. In December 2023, 44% (1,411/3,226) of such sites opted users out via all implemented privacy strings. In February 2024, this percentage decreased to 43% (1,473/3,402) before increasing to 45% (1,620/3,566) in April 2024. Despite the slight uptick between December 2023 and April 2024, compliance rates remained at a low level overall, indicating widespread disregard for California residents' right to opt out. Our findings highlight the importance of effective enforcement of the CCPA, in particular, with a focus on big web publishers.

1 Introduction

In 2018 the California State Legislature passed the California Consumer Privacy Act (CCPA) giving California residents the right to opt out of the sale or sharing of their personal information [74]. They can exercise their opt out right via Global Privacy Control (GPC) [82, 83, 84, 85, 86]. This binary privacy preference signal can be implemented via an HTTP header or a JavaScript DOM property. The California Privacy Protection Agency (CPPA) requires that businesses subject to the CCPA honor "user-enabled global privacy control, like the GPC" as a valid opt out request, and the Office of the California Attorney General (OAG) enforces this requirement [75]. A growing number of browsers, including Brave, DuckDuckGo, and Firefox, now natively support GPC.

When a first party website receives a GPC signal, it must propagate this preference to its integrated third parties, if any. This propagation is done using privacy strings. The first party updates any implemented privacy strings based on the user's preferences, and all integrated third parties are required to modify their behavior accordingly. For instance, site operators can propagate a user's preferences to third parties via the US Privacy String (USPS) [40], a privacy string provided by the Interactive Advertising Bureau (IAB), an ad industry organization, that specifically supports CCPA requirements [39]. As of January 31, 2024, the IAB deprecated the USPS and replaced it with the Global Privacy Platform (GPP) String, which supports requirements for multiple US state laws and laws in other jurisdictions [33]. Sites that integrate OneTrust [60], a Consent Management Platform (CMP) that provides web and app compliance libraries, can set `isGpcEnabled` in the OptanonConsent cookie. Also, site operators can include a `.well-known/gpc.json` resource on their site to indicate to the public that it supports GPC [85]. We will collectively refer to the set of USPS, GPP String, OptanonConsent cookie, and `.well-known/gpc.json` as privacy strings.[1]

In this study, we show how to evaluate websites' compliance with the CCPA opt out right expressed via GPC and based on the propagation of privacy strings to third parties. In particular, we are addressing the following research questions:

RQ1. CCPA Opt Out Right Applicability: How do we identify sites subject to the CCPA opt out right?

RQ2. GPC Compliance Implementation: How do we identify sites' compliance with the CCPA opt out right when receiving GPC signals at scale and over time?

*Katherine Hausladen, Oliver Wang, and Jocelyn Wang worked on this study while they were at Wesleyan University. They graduated in Spring 2024. This study is based on the master's thesis of Katherine Hausladen [29]. Corresponding author: Sebastian Zimmeck (szimmeck@wesleyan.edu).

[1] See Appendix A.1 for background information on GPC and privacy strings. See Appendix A.3 for relevant CCPA definitions in the context of this study.

RQ3. GPC Compliance Evaluation: To which extent do sites indicate that they respect GPC? How does GPC compliance on the web change over time?

RQ4. Transition from USPS to GPP: Are sites that implement the IAB's USPS transitioning to GPP?

RQ5. Recommendation for Regulators: How can regulators improve GPC compliance?

To answer these questions we implemented a web crawler based on the Selenium WebDriver [73], which we used to visit 11,708 sites and which instrumented our browser extension to determine if a site is subject to the CCPA and, if so, respects GPC signals. We based our browser extension on the OptMeowt extension in analysis mode [65], as described and implemented by Zimmeck et al. [86]. We collected three longitudinal snapshots — in December 2023, February 2024, and April 2024 — a time period which coincides with the IAB's transition from the USPS to the GPP String. After discussing the background and related work of our study (§2) and how we constructed our crawl set of 11,708 sites (§3.1), we make the following contributions:

- We develop a methodology for identifying CCPA opt out right applicability to websites based on their amount of web traffic and integration of third party libraries that collect, buy, sell, or share personal information. (**RQ1**: §3.2.)
- We design and implement a web crawler and browser extension to identify sites' GPC compliance by sending GPC signals and detecting the presence of privacy strings, their values, and changes of their values, each with high accuracy. (**RQ2**: §3.3 to §3.6.)
- Despite a slight uptick, the overall compliance level remained low for our observation period. In December 2023, 44% (1,411/3,226) of sites that sell or share personal information per the CCPA and that implement at least one of the four privacy strings opted users out via all implemented privacy strings. In February 2024, this percentage decreased to 43% (1,473/3,402) before increasing to 45% (1,620/3,566) in April 2024. (**RQ3**: §4.1 and §4.2.)
- GPP adoption rose markedly at the time of USPS deprecation and then slowed down substantially. (**RQ4**: §4.2.3.)
- Our main recommendation for regulators is to focus enforcement on big web publishers. For example, instances of inconsistent opt outs resulting from multiple implemented privacy strings are often related to a few big publishers. Further, as big publishers tend to roll out the same privacy strings to all of their sites, improved compliance will result in broad impact. (**RQ5**: §4.3, §4.4, and §5.)

2 Background and Related Work

Our study is motivated by the lack of compliance with the CCPA (§2.1) for opting out from web tracking (§2.2). GPC and other privacy preference signals can help users to exercise their opt out rights (§2.3). Thus, we seek to analyze websites' compliance with GPC at scale and over time (§2.4).

2.1 The CCPA and its Evolution

The CCPA grants California residents, among others, the rights to (1) know what personal information businesses collect about them, (2) have deleted the personal information that was collected, (3) opt out of the sale or sharing of personal information, and (4) not be discriminated against for exercising CCPA rights [74]. Originally, the CCPA applied to businesses that (A) have an annual revenue of at least twenty-five million dollars, (B) collect the personal information of 50,000 or more consumers, or (C) derive more than half of their annual revenue from selling personal information [76]. Any business that sold or shared personal information was required to "[p]rovide a clear and conspicuous link on [its] Internet homepage, titled 'Do Not Sell My Personal Information"' (DNSL) where the consumer can exercise the opt out right [76]. Under the CCPA, "sale" is defined as transferring personal information "for monetary or other valuable consideration" (CCPA, §1798.140(ad)(1)). "Sharing" is defined as transferring personal information "for cross-context behavioral advertising, whether or not for monetary or other valuable consideration" (CCPA, §1798.140(ah)(1)).[2]

In 2020, Californians voted to amend the CCPA through the California Privacy Rights Act (CPRA). While this amendment modified the CCPA in a variety of ways, there are two changes that are of particular interest for our purposes. First, the CPRA reduced the scope of businesses to which the CCPA applies by increasing the threshold in part (B) of the original "business" definition to 100,000 or more consumers (CCPA, §1798.140(d)(1)(B)). Second, the CPRA exempted businesses from the DNSL requirement if they respect automated opt out signals (CCPA, §1798.135(b)(1)). Additionally, this amendment created the CPPA to enforce the CCPA. The CPRA became operative January 1, 2023, and enforcement began on March 29, 2024.

2.2 Opting Out from Web Tracking

Despite the CCPA's rights opting out from web tracking remains a challenge for users and is difficult to enforce for regulators. Many websites use tracking cookies first and ask for consent later [79]. Invisible tracking pixels are commonly used by third parties [23, 69]. The ads that appear on sites can be related to various personal attributes and potentially even lead to discrimination [17]. While more recent approaches of ad personalization are more privacy-preserving, e.g., the Topics API by Google has better privacy compared to tracking cookies [1], interest-disclosing mechanisms are generally privacy-exposing and not -preserving [5]. For users it is difficult to gain control over their data. Privacy controls and

[2]Full legal definitions are included in Appendix §A.3.

opt out choices can vary significantly and confuse users even when they want to have more privacy [28].

User-friendly design for privacy notice and choice interactions has been studied to help improve the usage of choice mechanisms [26, 27, 71, 84]. Browser extensions can help users to opt out by automatically clicking on opt out links [3] or cookie opt out buttons [42]. From the sites' perspective, CMPs can help streamline privacy controls [70]. To propagate users' CCPA opt out choices from first parties to third parties and among third parties the IAB developed GPP as part of the IAB CCPA Compliance Framework [36]. However, various adoption and implementation issues limit users' opt outs from being honored by third parties [2]. While laws and regulations were found to have some effect, e.g., the introduction of the GDPR lowered the amount of third party cookies and tracking [45], regulators currently do not have any mechanism to audit ad networks' compliance with user consent, i.e., to determine if ad networks indeed do not collect, process, and share user data when users opt out [46].

2.3 GPC and other Privacy Preference Signals

Privacy preference signals hold the promise of helping users to efficiently and effectively exercise their opt out rights. However, their adoption represents a coordination problem dating back to the 1990s [31]. The Platform for Privacy Preferences Project (P3P) was an early effort to help people understand and make privacy choices automatically based on machine-readable privacy policies [14, 15]. Its complexity and the lack of adoption of machine-readable privacy policies lead to the development of Do Not Track (DNT) [21], a simple binary signal for people to express their opt out of tracking per the California Online Privacy Protection Act [9]. However, DNT adoption remained low as there was no legal obligation for signal recipients to comply with DNT; only to *say* whether they comply [9]. Learning from previous experiences GPC aims to be a technologically practical and legally enforceable opt out mechanism [85]. GPC is not limited to the CCPA but could also be applied in the EU [4]. Competent Data Protection Authorities or the Court of Justice of the EU could clarify how GPC should be interpreted under EU law [16].

Our study is based on various previous GPC studies. One study has shown that most users understand what they declare by turning on GPC and that they would do so in larger numbers if GPC were supported by their browser [86]. In addition, websites and their integrated third parties would also need to adopt GPC functionality. At the time of the study [86], August 2022, only 12% of sites respected GPC signals, which substantially increased per our findings in this study. At that time, the applicability of the CCPA was determined by the presence of a DNSL, an approach that no longer works as DNSLs are not mandated anymore under the CCPA as amended by the CPRA (CCPA, §1798.135(b)). While nearly a quarter of the top 25,000 Tranco sites [64] that had a DNSL before the CPRA took effect no longer support any opt out mechanism, the number of sites that respect GPC signals increased from November 2022 to November 2023 [11]. By measuring CCPA compliance based on DNSLs and determining CCPA applicability via a combination of unique visitor count estimates and detection of embedded ad networks using resource inclusion trees another study showed that DNSLs were much more prevalent in the top 25,000 Tranco sites than in lower-ranked sites. In general, as the CCPA is evolving there is a positive spillover effect of the CCPA opt out requirements to states with weaker or no privacy laws [78]. In this sense, GPC may reach beyond jurisdictions that formally adopt it.

2.4 Website Compliance Analysis

In this study we evaluate websites' compliance with GPC. Generally, website compliance with legal requirements can be evaluated based on various browser automation tools [20, 25, 51]. We implemented a web crawler using a Selenium WebDriver [73] and Firefox Nightly instance [55]. We identify trackers with Firefox's built-in Enhanced Tracking Protection [54], which identifies tracker URLs and exposes an API that categorizes the trackers based on the Disconnect list [19]. In addition to websites' compliance with GPC, previous work evaluated compliance with EU cookie laws, focusing on consent requests, and found various infractions [10]. A number of sites registered positive consent without user choice, nudged users with pre-selected options, or ignored explicit opt outs [47]. While a significant number of sites adjusted their privacy policies as the GDPR became effective, the functionality and usability of opt out mechanisms saw less improvement [18]. Thus, compliance problems under both the GDPR and CCPA continue to persist [81].

3 Methodology and Application

We constructed a crawl set of 11,708 sites (§3.1) and evaluated the applicability of the CCPA opt out right for each site (§3.2). Our evaluation methodology is based on detecting changes in privacy strings that sites send to integrated third parties after receiving a GPC signal (§3.3). We implemented this methodology in a browser extension that runs on a Selenium web crawler (§3.4) and evaluated its accuracy for detecting sites' compliance with GPC (§3.5). This approach is subject to various limitations (§3.6).

3.1 Constructing the Crawl Set

We started the construction of our set of sites to crawl (the *crawl set*) using BuiltWith [8], a lead generation service that identifies sites integrating certain technologies, such as GPC, USPS, or a certain ad network.[3] We identified sites potentially

[3]While BuiltWith identifies sites with GPC or USPS, its detection mechanism is shallow, essentially, only detecting whether a site contains any GPC-

subject to the CCPA due to their integration of technologies that are likely involved in the selling or sharing of personal information (§3.2.1). We used the Selenium WebDriver [73] to scrape BuiltWith's free preview for United States sites with GPC [6], USPS [7], or a popular ad network from the Disconnect list [19]. We ran the scrape between October 11 and November 3, 2023. We scraped sites from BuiltWith that:

1. Have GPC code or
2. Have USPS code or
3. Use Facebook, Twitter, Amazon, Google, Automattic, TikTok, Microsoft, MailChimp, Akamai, OptinMonster, or Criteo. We chose the ad networks in the following way:
   (a) Begin with all ad network URLs in the Advertising section of the Disconnect list, filtered for .com, .net, .org, .info, .biz, .mobi, .us, .services, .xyz, .tech, .co, .tv, .works, .pro, .online, .media, .space, .io, and .html.
   (b) Sort the ad network URLs by how many server domain names they have on the Disconnect list as a proxy for prevalence to maximize crawl set coverage.
   (c) If the ad network has more than 5 server domain names, manually search BuiltWith for its name.
   (d) If the ad network has up to 5 server domain names, automate the search on BuiltWith by changing spaces in the ad network name to "-", concatenate the new name to `https://trends.builtwith.com/ads/`, and load each resulting URL using Selenium.
   (e) For each search per (c) and (d) collect the USA Live sites number from their BuiltWith site.
   (f) Sort the ad networks by how many USA Live sites each has. Pick the 11 highest-ranked ad networks.

To identify all sites on BuiltWith using GPC, USPS, or ad network technologies, we constructed URLs by concatenating each of the following base URLs and paths for all available US states and cities resulting in a base set of 42,312 sites (the *base set*). We loaded each constructed URL with Selenium and scraped the site metadata from BuiltWith, including sites' Traffic categories ("Very High," "High," "Medium," "–").

1. Base URLs:
   GPC: `https://trends.builtwith.com/websitelist/Global-Privacy-Control`
   USPS: `https://trends.builtwith.com/websitelist/US-Privacy-User-Signal-Mechanism`
   Ad Network: `https://trends.builtwith.com/ads/<ad_network_name>`
2. Paths:
   (a) `/United-States`
   (b) `/United-States/<state>/` for each US state
   (c) `/United-States/<state>/<city>` for each city listed on a `/United-States/<state>/` path

To select sites that are more likely to be a business due to having at least 100,000 annual California visitors (§3.2.2) and, thus, being subject to the CCPA, we filtered the base set to only include sites that had a value of at least "Medium" in BuiltWith's Traffic category and then selected the top 11,708 sites based on BuiltWith's reported rank of a site on the Tranco list [64]. These 11,708 sites form our crawl set. We selected this number of sites because we estimated that sites with less traffic and lower Tranco rank would not meet the threshold of 100,000 California residents to qualify as a business. To evaluate whether a site meets the threshold, we used web traffic statistics from Similarweb [72], a web analytics service. We correlate the web traffic with the Tranco rank as we consider the Tranco list to be the canonical web ranking.[4]

3.2 CCPA Opt Out Right Applicability

For **CCPA opt out right applicability (RQ1)** a site must (1) sell or share personal information (§3.2.1) and (2) be part of a business (§3.2.2).

3.2.1 Selling or Sharing Personal Information

A site "sells" or "shares" personal information per the CCPA if at least one integrated third party buys or collects such from the site. To make this determination we leverage Firefox's Disconnect list integration [19, 54], which identifies trackers, classifies them into different categories, and exposes the classification result via an API in Firefox's `urlClassification` object [49]. We want to identify those categories in the Disconnect list that have the highest chance of containing services that collect or buy personal information. To construct a set of services with this property we performed a manual analysis of the privacy policies of 115/1,315 (9%) unique services in the Advertising, Analytics, Fingerprinting General, Fingerprinting Invasive, and Social categories of the Disconnect list. We determined for each policy whether it allows a service to buy or collect personal information (or sell or share such downstream).[5] We estimated that a number of around 100 analyzed policies would allow us to sufficiently calculate the statistical significance of a site in our crawl set buying or collecting personal information.

or USPS-related code. As such, it cannot be used to determine whether a site has a functional GPC or USPS implementation.

[4] Note that while we could not find BuiltWith's Traffic category definitions, we got an estimate via the sites' Tranco ranks. Appendix A.2, Figure 15 shows the ranges of Tranco ranks of sites included in each of BuiltWith's Traffic categories for the base set. After excluding the "–" Traffic category, which presumably means that a site has no or only nominal traffic, and sorting by Tranco rank, all sites in our crawl set had a Tranco rank of 163,503 or higher.

[5] Our analysis is based on the Disconnect list of September 15, 2023. While summing the services of each of the aforementioned categories yields 1,506 total services, some of these services are included in multiple categories. Excluding duplicates, there are 1,315 services.

Disconnect Category	Total Services	# of Services Analyzed	% that Buy or Collect
Social (S)	19	10	**100%**
Advertising (A)	1,083	68	**93%**
Fingerprinting General (FG)	44	13	**100%**
Analytics	243	15	53%
Fingerprinting Invasive	123	9	78%
S ∪ A ∪ FG (SAFG)	1,097	97	**95%**

Table 1: Privacy policy analysis results for 115 services on the Disconnect list. For individual categories, Total Services is the number of services on the Disconnect list for that category. For the SAFG services, Total Services is the number of unique services in the union of categories. Note that the number of SAFG services is not equal to the sum of individual services in the Social, Advertising, and Fingerprinting General categories as the SAFG union includes cross-listed services.

Initially, we randomly selected and analyzed policies of 70/1,083 (6%) Advertising, 15/243 (6%) Analytics, 9/44 (20%) Fingerprinting General, 9/123 (7%) Fingerprinting Invasive, and 9/19 (47%) Social services. We omitted 2 Advertising services that did not have a privacy policy, resulting in 68/1,083 (6%). We also decided to omit the Analytics and Fingerprinting Invasive categories as they self-declared CCPA applicability in their privacy policies at a lower rate. We took the union of the remaining categories — Social, Advertising, and Fingerprinting General (the *SAFG* categories). To increase our sample size in the SAFG categories, we analyzed policies of 4 additional randomly selected services in Fingerprinting General for 13/44 (30%) and 1 additional randomly selected service in Social for 10/19 (53%). We also included policies of 6 services that we had originally analyzed as part of the Analytics or Fingerprinting Invasive categories but were cross-listed in the Advertising category, resulting in a total of 97 SAFG services. As shown in Table 1, 95% of policies we analyzed in the SAFG union declare to buy or collect personal information, that is, sites on which they are integrated sell or share personal information. Calculating a confidence interval for proportion using z-scores, with 95% confidence the true proportion of SAFG services buying or collecting personal information is 91-99%. This interval is a lower bound since the determination is based on self-identification in services' privacy policies.

3.2.2 Business

To evaluate whether a website is part of a business we used web traffic statistics from Similarweb [72]. We want to estimate which sites in our crawl set have at least 100,000 annual California visitors, which would qualify the site as part of a business per the CCPA. Similarweb provides a site's monthly traffic and percentage of traffic by country available for three-month periods. We estimate the annual California visitors of a site based on the following equation:

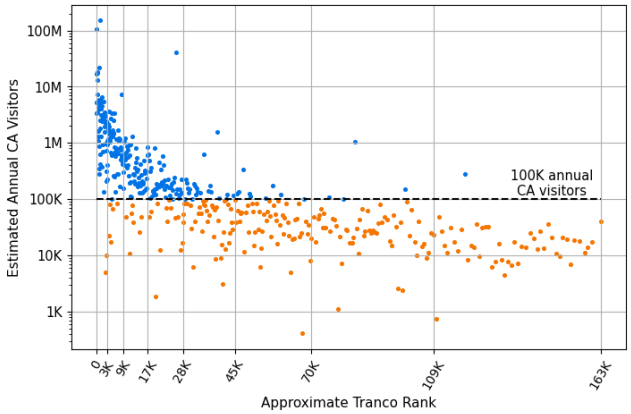

Figure 1: The annual California traffic for every 25th site in our crawl set ($n = 468$) estimated per Equation 1. As the trend was the same when we only included half of the data, i.e., every 50th site, we believe the inclusion of every 25th site to be sufficient.

$$12 * \text{Average Monthly Traffic} * \left(\frac{\text{CA Population}}{\text{US Population}}\right) \quad (1)$$

where $12 *$ Average Monthly Traffic is an upper bound for annual unique visitors as some will be repeat visitors.[6] We applied this equation to manually collected Similarweb October–December 2023 data, from which we calculated the Average Monthly Traffic, for every 25th site in our crawl set, as ordered by the Tranco list. As indicated in Figure 1, most sites below a Tranco Rank of 28K would be considered part of a business under the CCPA based on their web traffic. We cannot draw conclusions about the sites that have fewer than 100,000 annual California visitors, as they may satisfy a different threshold of the CCPA's business definition. For instance, more than half of their revenue could come from selling personal information, which is not an unreasonable assumption given that data for content is the prevalent business model on the web.[7]

We correlated ranges of Tranco ranks with the probability that a site has at least 100,000 California visitors. As shown in Figure 2, a lower Tranco rank means that a site is more likely to have California traffic that exceeds 100,000 annual visitors. With 95% confidence, a site in the first 1,500 sites of our

[6] While it would be ideal to have a lower bound, we did not find available statistics or a way to calculate such. Also, while the number of visitors of a site from California will not be proportional to the CA Population for every site, on average, we believe this estimate to be reasonable.

[7] Generally, operationalizing the thresholds of the CCPA business definition from the perspective of an outside observer is challenging. The "translation" of the law into technological criteria can thwart its effectiveness. This challenge extends beyond the CCPA and applies to other US state laws with similar definitions as well. It is rooted in the limits of geographical jurisdictions and the attempt of not overburdening small businesses.

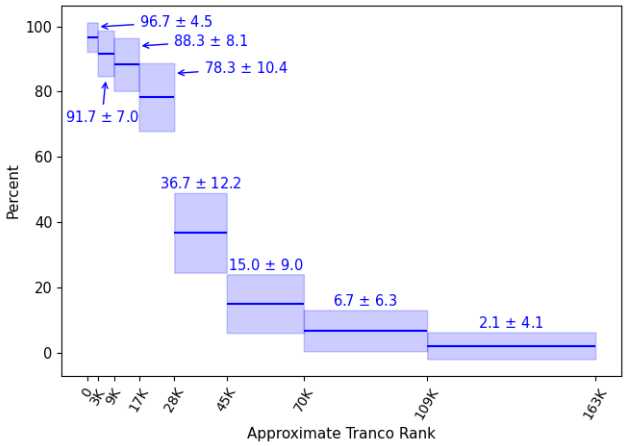

Figure 2: The 95% confidence intervals for ranges of sites having at least 100,000 annual California Visitors. Each range contains 60 Similarweb data points and represents 1,500 crawl set sites (1,500/25 = 60), except the rightmost, which contains 48 Similarweb data points representing 1,208 crawl set sites.

crawl set (Tranco rank 0–3K) has a 96.7 ± 4.5% probability of being part of a business per the CCPA based on its web traffic. However, a site in the set of the last 1,208 sites of our crawl set (Tranco rank 109K–163K) has only a 2.1 ± 4.1% probability. Taken together, the leftmost four boxes in Figure 2 indicate that there is about an 89% probability that a site in the first 6K sites of our crawl set is part of a business per the CCPA based on web traffic.

3.2.3 Selling or Sharing & Business

Combining the eight confidence intervals shown in Figure 2 with the identified SAFG requests from §3.2.1, we can estimate the number of sites in our crawl set that are subject to the CCPA. To determine the upper bound of sites subject to the CCPA based on web traffic, we use the following equation:

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 \end{bmatrix} * \begin{bmatrix} min(0.967+0.045,1) \\ min(0.917+0.07,1) \\ min(0.883+0.081,1) \\ min(0.783+0.104,1) \\ min(0.367+0.122,1) \\ min(0.15+0.09,1) \\ min(0.067+0.063,1) \\ min(0.021+0.041,1) \end{bmatrix} \quad (2)$$

where x_n is the number of sites with an SAFG request between index $1{,}500n-1{,}500$ and $min(1{,}500n-1{,}11{,}708)$ in our crawl set. We illustrate the **CCPA opt out right applicability (RQ1)** using $n=1$ in December 2023 as an example: since $n=1$, we need to know x_1, the number of sites with an SAFG request between index $1{,}500*1-1{,}500=0$ and $min(1{,}500*1-1,\ 11{,}708)=1{,}499$, i.e., $0\ to\ 1{,}499$. In our December 2023 crawl data, x_1 = 1,261. Therefore, we arrive at $x_1*min(0.967+0.045,1)=1{,}261*1=1{,}261$ as an upper bound for number of sites subject to the CCPA in the first 1,500 sites of our crawl set. To determine the lower bound, we subtract the error and take the maximum with 0. Continuing our $n=1$ example in December 2023, we arrive at $x_1*max(0.967-0.045,0)=1{,}261*0.922=1{,}163$ of the first 1,500 sites in our crawl set are subject to the CCPA.[8]

3.3 Detecting Privacy String Value Changes

Given the probability of a site being subject to the CCPA opt out right (§3.2), we can run our **GPC compliance implementation (RQ2)** as follows:

1. Check whether the site sells or shares personal information via Firefox's `urlClassification` object (§3.2.1).
2. Check the values of the US Privacy String, OneTrust's OptanonConsent cookie, and GPP String, if any.
3. Send a GPC signal to the site.
4. Recheck the values per step 2.

In order for a site to be GPC-compliant, the following must be true after the GPC signal was sent for each mentioned privacy string that the site implements:

1. The third character of the US Privacy String is a `Y`.
2. The value of the OptanonConsent cookie contains `isGpcEnabled=1`.
3. The opt out columns in the GPP String's relevant US sections (i.e., `SaleOptOut` and `SharingOptOut`) have a value of `1`. In California, the relevant sections are `uscav1` or `usnatv1`, and the relevant opt out columns are `SaleOptOut` and `SharingOptOut`.

Appendix A.1 contains further background information on the mentioned privacy strings. For the `.well-known/gpc.json`, if any, we detect its `gpc` value.

3.4 Implementation and Procedure

To perform the 4-step analysis outlined in §3.3 we implemented a Selenium web crawler and Firefox browser extension, which is based on OptMeowt's analysis mode [65]. We crawled our crawl set of 11,708 sites three times: in December 2023, February 2024, and April 2024. All crawls were run on a 2018 MacBook Pro with an Intel I7 processor and 16 GB RAM set to a Los Angeles IP address using Mullvad VPN [56]. Figure 3 shows an overview of our setup.

[8] These bounds do not apply to other thresholds of the CCPA business definition, such as the percentage of revenue from selling or sharing personal information.

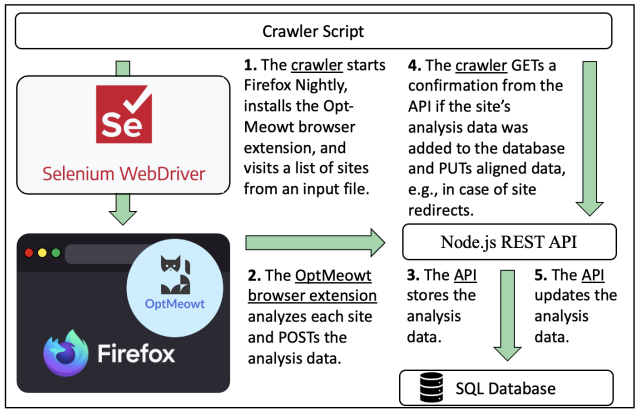

Figure 3: The high-level architecture of our setup, which includes a Selenium web crawler, SQL database, and our OptMeowt analysis extension, all run locally.

3.4.1 Browser Extension

Our extension begins the analysis of a site 7 seconds after DOM content has been loaded. Requests to SAFG category sites are identified via Firefox's `urlClassification` object [49] and stored in a SQL database for later analysis. Our extension uses the cookies API [48] to check for US Privacy and OptanonConsent cookies and injects scripts to call the USPAPI and GPP's Consent Management Platform API (CMPAPI). We allotted a 2.5-second timeout for the collection of any privacy string values. Then, the extension sends a GPC signal, reloads the site, collects the USPS, GPP String, and OptanonConsent cookie values again, and stores the results in the database. After a 3-second timeout the crawler loads the next site for our extension to analyze. We selected all timeouts to minimize idle time while ensuring accurate data collection. The timeouts are based on preliminary experiments, particularly, ensuring that all site resources are fully loaded.

3.4.2 Selenium Web Crawler

We implemented our crawler using Selenium WebDriver for Firefox Nightly [73]. The crawler's primary purpose is to automate the loading of sites for analysis by the extension. We ran each crawl in eight batches of around 1,500 sites. Larger batch sizes lead to more Selenium crashes. After finishing all eight batches, we identified each site that had a Selenium error and a subdomain, removed the subdomain from the site's URL, and recrawled all such sites since an invalid subdomain was one of the main error sources we encountered.[9] Our crawler uses non-headless mode to appear more similar to a human user. Each site is allotted 35 seconds to load and 22 seconds to be analyzed. We found that a 35-second loading period gives slow-loading sites sufficient time to load. The 22-second analysis period is the sum of the timeouts in the analysis extension (§3.4.1).

The crawler logs and attempts to catch any errors that occur while a site loads. Generally, errors happen because a site failed to load in 35 seconds, had an insecure certificate, or led to an error page. Before moving to the next site, the crawler checks if analysis data for the current site was added to the database. If that is not case and there is no Selenium error, a second attempt at loading the site may be successful. Thus, the crawler loads the site one more time. This strategy greatly improved the success rate for sites that failed to load in 35 seconds on the first attempt. Over all crawls, nearly 40% of sites that failed to load on the first attempt were loaded and successfully analyzed on the second attempt. In each of our three crawls, about 2% of sites failed to load on both attempts. Considering all errors, about 7–9% of sites could not be analyzed in each crawl (§4).

Another purpose of the crawler is to identify sites with human checks loading an intermediate page that either indicates that access to a site has been denied or prompts users to perform some challenge to prove that they are human. Since the text at the `/html/head/title` XPath on a site is often indicative of the presence of a human check, we use a list of regular expressions to match this text and determine whether the site has a human check. We rely on this approach as opposed to a more complex approach because many sites use a service such as Cloudflare [13] to identify bot visitors and, therefore, have the same intermediate page, which is identifiable by regular expressions.

Finally, the crawler automates identification of sites that redirect to another domain. If the crawler loads a URL that redirects to another domain, the analysis extension logs the analysis data in our database under the new domain. There were 644 sites in our crawl set that redirected to a different domain in at least one of our three crawls.

3.4.3 `.well-known/gpc.json`

To determine which sites have a `.well-known/gpc.json` we use the Python requests library resource [68]. This crawl is performed separately and is not part of our Selenium web crawler.

3.5 Accuracy of Non-Compliance Detection

We tested the accuracy of our extension using a 100-site test set (the *test set*). After our first crawl, we randomly selected 60 test set sites from the second batch, which we had randomly selected as well. To ensure that there were sufficient sites with GPP Strings in the test set we randomly selected 40 sites from the 109 sites for which our first crawl had detected a GPP String. In our test procedure we simultaneously compared analysis values, generated by our extension, and ground truth values, manually observed by us. As our crawler analyzes data

[9] See Appendix A.4 for details on the errors we encountered.

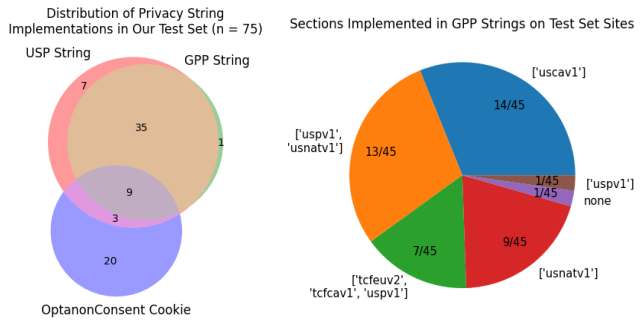

Figure 4: Privacy strings (left) and GPP String sections (right) found on the sites in the test set after a GPC signal was sent.

Analysis Item	TP, FP, TN, FN	#	P, R, F1
USPS Found Before GPC Sent	54, 0, 46, 0	100	1, 1, 1
USPS Found After GPC Sent	54, 0, 46, 0	100	1, 1, 1
USPS Opt Out after GPC Sent	25, 0, 29, 0	54	1, 1, 1
USPS Change to Opt Out after GPC Sent	25, 0, 0, 0	25	1, 1, 1
OAC Found Before GPC Sent	32, 0, 68, 0	100	1, 1, 1
OAC Found After GPC Sent	32, 0, 68, 0	100	1, 1, 1
OAC Opt Out after GPC Sent	21, 0, 11, 0	32	1, 1, 1
OAC Changes to Opt Out After GPC Sent	21, 0, 0, 0	21	1, 1, 1
GPP String Found Before GPC Sent	44, 0, 56, 0	100	1, 1, 1
GPP String Found After GPC Sent	45, 0, 55, 0	100	1, 1, 1
Sale Opt Out After GPC Sent	3, 0, 97, 0	100	1, 1, 1
Sharing Opt Out After GPC Sent	3, 0, 97, 0	100	1, 1, 1
Sale Change to Opt Out After GPC Sent	3, 0, 28, 0	31	1, 1, 1
Sharing Change to Opt Out After GPC Sent	3, 0, 28, 0	31	1, 1, 1

Table 2: Our extension's performance of finding privacy strings in the test set ($n = 100$). OAC Found indicates that an OptanonConsent cookie with an `isGpcEnabled` value was found. Sale and Sharing refer to the `SaleOptOut` and `SharingOptOut` fields, respectively, in the `usnatv1` or `uscav1` sections of the GPP String. (USPS = US Privacy String, OAC = OptanonConsent cookie, TP = True Positives, FP = False Positives, TN = True Negatives, FN = False Negatives, # = Total, P = Precision, R = Recall, F1 = F1 score.)

on the same site load as we observe ground truth values, we can be sure that differences are not due to varying site loads. We declared a USPS present if the USPAPI returned a non-null USPS value or a `usprivacy` or similarly named cookie was found. We declared a GPP String present if the GPP CMPAPI returned a non-null GPP String value. We declared an OptanonConsent cookie present if we found a cookie of that name. If it did not have an `isGpcEnabled` field, we gave it a value of `no_gpc`. We performed our tests with Mullvad VPN connected to a Los Angeles IP address.

Our extension correctly identified the presence and value of all privacy strings in the test set (Figure 4). When identifying the presence and value of GPP Strings, our extension had an accuracy, precision, recall, and F1 score of 1. Of the 45/100 (45%) test set sites with a GPP String, 36/45 (80%) had a relevant California section (`usnatv1` or `uscav1`). 3/36 (8%) sites changed the `SaleOptOut` and `SharingOptOut` fields to opt out after receiving a GPC signal, both of which were correctly identified by our extension. It is noticeable that the number of sites changing their values to opt out after receiving a GPC signal is generally lower compared to sites already having an opt out value set before receiving a GPC signal.[10]

In total, a USPS was implemented on 54/100 (54%) sites. With an accuracy, precision, recall, and F1 score of 1, our extension reliably identified the existence and value of the USPS. Our extension also correctly identified all 25/54 (46%) sites that changed the third character of the USPS to a `Y` after receiving a GPC signal, i.e., opted us out. 26/54 (48%) test set sites implemented a USPS using the USPAPI only, and the other 28/54 (52%) used both the USPAPI and cookies highlighting the importance of correctly identifying a USPS via the USPAPI. Our extension also had an accuracy, precision, recall, and F1 score of 1 when identifying the presence and value of OptanonConsent cookies. An OptanonConsent cookie with an `isGpcEnabled` field was found on 32/100 (32%) sites, 21/32 (66%) of which opted us out after receiving a GPC signal. Our extension also correctly identified the 1 site that had an OptanonConsent cookie without an `isGpcEnabled` field. Table 2 shows our complete results.

3.6 Limitations

Our approach is subject to various limitations. First, we determine CCPA applicability based on an estimate of a site's California web traffic. For our estimate we only analyzed web traffic data for 1/25th of our crawl set. We also do not consider other CCPA applicability thresholds, e.g., the annual revenues from selling or sharing personal information, which, however, could only lead to an increase of sites subject to the CCPA. For determining whether a site is selling or sharing personal information we rely on the classification of an integrated third party as buying or collecting such as described in its privacy policy, which is subject to legal interpretation, may not reflect their actual practice, and can change over time. We also assume that Firefox's webRequest API always returns the correct `urlClassification` for a site. Also, our extension cannot analyze sites that block script injection as we use such to call the USPAPI and GPP CMPAPI. Previous work found that up to 14% of websites detect automated browsers [43]. We identified that 1.6–2.6% of sites in our crawl set had a

[10] One site only had a GPP String after we sent the GPC signal, which appears to be a result of AdRoll's GPP implementation. The GPP CMPAPI responded both before and after we sent the signal. First, it returned an undefined GPP String. Then, it returned an empty GPP String. We found the AdRoll script that implemented GPP to respond in this way by searching for `_set_global("__gpp"` in the Firefox Debugger. There were 85 sites in our crawl set that had this behavior in at least one of our crawls. We did not manually verify that each of these sites used AdRoll's GPP implementation. As of January 2025 it appears that the AdRoll script has been modified.

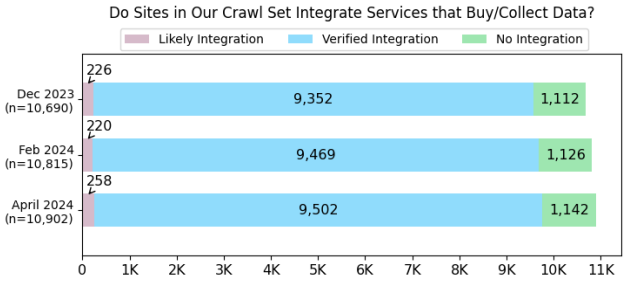

Figure 5: SAFG services integrated in the crawl set sites.

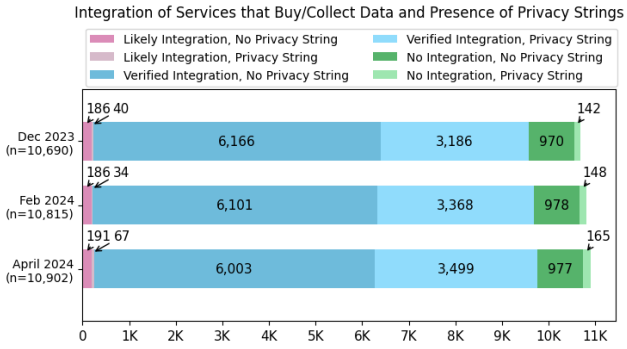

Figure 6: SAFG services integrated in the crawl set sites and their privacy strings.

human check. While we used a California VPN, some sites may have detected that we are not actually in California, treating us differently. However, if sites were detecting our true location, we would expect to see the section of the GPP String of the state in which we are truly located, which we did not see using a VPN but did see without it. Our crawl set as a whole is certainly not comprehensive and may also not be representative for the overall GPC compliance on the web or in regard to our other findings as we constructed it from BuiltWith and relied on its pre-selection of sites.

4 Results

Our crawler successfully analyzed 10,690/11,708 (91%) sites in December 2023, 10,815/11,708 (92%) sites in February 2024, and 10,902/11,708 (93%) sites in April 2024. The remaining sites' analyses failed due to errors.[11] The success rate of our crawler is similar to the performance reported in previous work [78]. The average analysis time per site was 30.7 seconds in December 2023, 30.5 seconds in February 2024, and 30.4 seconds in April 2024. We now evaluate the applicability of the CCPA opt out right to the sites in our crawl set (§4.1), determine their opt out rates (§4.2), evaluate inconsistencies in sites' opt out behavior due to the implementation of multiple privacy strings (§4.3), and illustrate the impact big publishers' privacy string configurations can have on opt out rates (§4.4).

4.1 CCPA Opt Out Right Applicability

About 90% of sites in our crawl set integrated at least one SAFG category service in all three crawls.

4.1.1 How Many Sites Sell or Share?

As shown in Figure 5, we further divided this subset of sites into Verified and Likely Integration categories based on whether a site integrated a service we manually verified (§3.2.1). For Verified Integration sites, which account for about 87% of sites in all three crawls, we are certain that they sell or share. For Likely Integration sites, which account for about 250 sites, we have 95% confidence that 91–99% sell or share (§3.2.1). The remaining 10% of sites, shown in the No Integration category, do not integrate any SAFG service.[12]

Figure 6 shows that sites in the Verified Integration category are most likely to have a privacy string, i.e., at least one of USPS, GPP String, OptanonConsent cookie, or `.wellknown/gpc.json`. More specifically, 3,186/9,352 (34%) of Verified Integration sites, 40/226 (18%) of Likely Integration sites, and 142/1,112 (13%) of No Integration sites had a privacy string in December 2023. In February 2024, 3,368/9,469 (36%) of Verified Integration sites, 34/220 (15%) of Likely Integration sites, and 148/1,126 (13%) of No Integration sites had a privacy string. These percentages increased to 3,499/9,502 (37%), 67/258 (26%), and 165/1,142 (14%) in April 2024, respectively.

4.1.2 Which Sites Are Subject to the CCPA Opt Out?

Given that nearly 90% of the sites we successfully crawled in December 2023 have evidence of selling or sharing personal information, we determine if these sites are part of a business per the CCPA. We use Equation 2 to calculate the upper bound and the analogous equation to calculate the lower bound. Based on this calculation we estimate that 4,465–5,929 of the 9,578 sites, 47–62% according to the 95% confidence interval, that had a request from an SAFG service in December 2023 also meet the traffic requirement per §3.2.2 and, thus, are subject to the CCPA opt out right.[13]

Overall, as Figure 7 illustrates, sites higher ranked in the Tranco list are more likely to be subject to the CCPA as they

[11] See Appendix A.4 for details on the errors we encountered.

[12] It could also be that a site in this category had an uncaught error that did not crash our crawler and, thus, went unnoticed. The same consideration applies to sites in the No Integration, No Privacy String category in Figure 6 and the No Ad Network category in Figure 7.

[13] In addition, just the inclusion of a privacy string, as such, given that it relates to California, indicates that a site is subject to the CCPA.

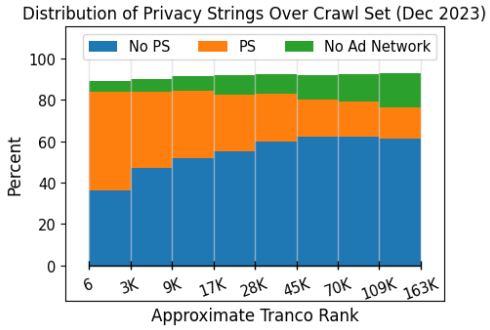

Figure 7: The distribution of privacy strings over our crawl set in December 2023. The trend is similar for the two subsequent crawls. Each bar represents 1,500 sites, except the rightmost, which represents 1,208 sites. The remaining percentage of sites were sites that could not be analyzed due to errors. No Ad Network means that the site did not integrate any SAFG category service, which includes both Likely and Verified Integration SAFG categories. (PS = Privacy String.)

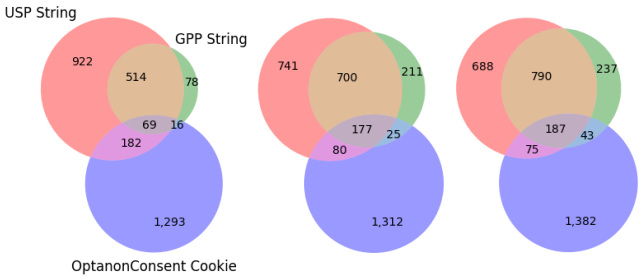

Figure 8: The distribution of sites' privacy string implementations in December 2023 ($n = 3,074$), February 2024 ($n = 3,246$), and April 2024 ($n = 3,402$), from left to right. We define a privacy string to be implemented if it ever has a non-null value (i.e., before GPC, after GPC, or both).

include a privacy string and are also more likely to sell or share as they more often include an SAFG service. Smaller sites seem to monetize their content to a lesser degree. While Figure 2 suggests a sharp drop in CCPA applicability and, thus, implemented privacy strings after a Tranco rank of about 28K, we observe a steadier decrease here. It could be that sites ranked lower on the Tranco list may still be subject to the CCPA opt out right due to their revenues. These trends stay consistent for February and April 2024.

4.2 Privacy String Adoption and Opt Outs

For evaluating sites' privacy string adoption and resulting opt out behavior we only consider sites with evidence of selling or sharing as these could be subject to the CCPA. Specifically,

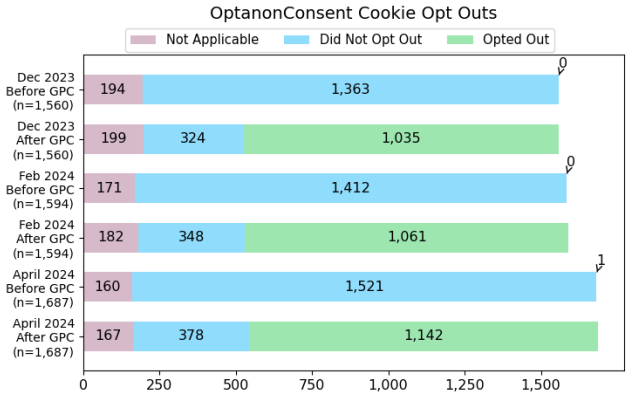

Figure 9: OptanonConsent cookie opt outs. The total counts, n, refer to the total number of sites that implemented an OptanonConsent cookie per our definition of an implemented privacy string (Figure 8). Some sites had a non-null privacy string value either only before or only after sending a GPC signal. Thus, the values on the bars do not all sum to n. All counts were slightly lower before sending a GPC signal.

we only consider sites in the Likely and Verified Integration categories (Figure 5). From Figure 6, we see that the percentage of sites with evidence of selling or sharing and implementation of at least one privacy string increased slowly from 34% (3,226/9,578) in December 2023 to 35% (3,402/9,689) in February 2024 and 37% (3,566/9,760) in April 2024. GPP String implementations significantly increased between December 2023 and February 2024 from 7% (677/9,578) to 12% (1,113/9,689). This time period coincided with the IAB's deprecation of the USPS. Between February and April 2024 GPP String implementation continued at a slower rate. Despite its deprecation, the percentage of sites with a USPS remained largely the same between December 2023 and April 2024. Overall, for our **GPC compliance evaluation (RQ3)** we find that in December 2023, 44% (1,411/3,226) of sites opted out via all privacy strings they implemented (i.e., one or more of the USPS, GPP String, OptanonConsent cookie, and `.wellknown/gpc.json`). In February 2024, this percentage decreased to 43% (1,473/3,402) before increasing to 45% (1,620/3,566) in April 2024. The percentage of sites that implemented at least one privacy string and opted out via none was 46% (1,477/3,226) in December 2023, 45% (1,519/3,402) in February 2024, and 45% (1,598/3,566) in April 2024.

4.2.1 OptanonConsent Cookie Adoption and Opt Outs

The percentage of sites opting out via the OptanonConsent cookie increased slightly each crawl (Figure 9). Compared to the other privacy strings, it has the highest percentages of being in an opted out state after a GPC signal; 66% (1,035/1,560), 67% (1,061/1,594), and 68% (1,142/1,687) of

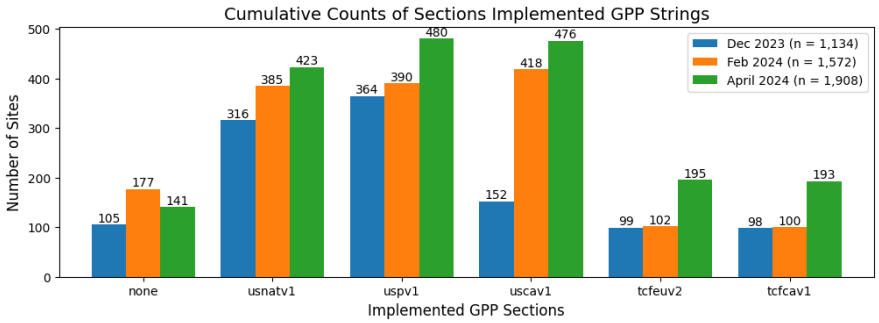

Figure 10: Cumulative counts of GPP String sections. For example, if we observed a `[usnatv1,uspv1]` string, i.e., sections for the deprecated USPS and US national laws, as described in Appendix A.1.2, such string adds one count to `usnvatv1` and one count to `uspv1`. The `tcfeuv2` and `tcfcav1` GPP sections are the EU and Canadian sections, respectively.

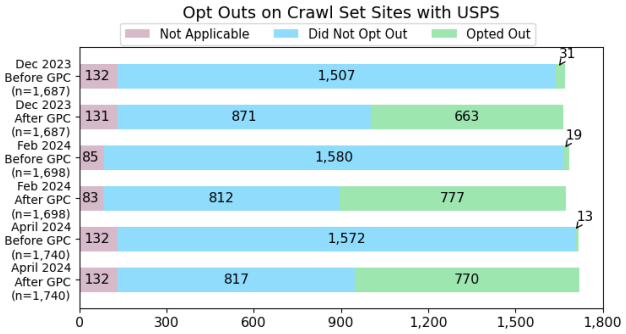

Figure 11: USPS opt outs. Some sites had a non-null privacy string value either only before or only after sending a GPC signal. Also, 3 (December 2023), 2 (February 2024), and 3 (April 2024) sites had an invalid USPS value (e.g., `1---N`).

OptanonConsent cookies opted out in December 2023, February 2024, and April 2024, respectively. One site, in April 2024, was in an opted out state before we sent a GPC signal. Some sites had an OptanonConsent cookie without `isGpcEnabled` field, shown as Not Applicable in Figure 9. Across crawls about 10% of sites fell into this category. A few sites removed the `isGpcEnabled` field after receiving a GPC signal causing the Not Applicable category to increase after GPC.

4.2.2 USPS Adoption and Opt Outs

Despite its deprecation at the end of January 2024, the USPS remained the most prevalent privacy string in our crawl set (Figure 11). The percentage of sites that implemented a USPS remained relatively constant each crawl indicating the slow pace of change in the industry. The percentage of sites that opt out via USPS after a GPC signal increased from 39% (663/1,687) in December 2023 to 46% (777/1,698) in February 2024 before decreasing to 44% (770/1,740) in April 2024.

4.2.3 GPP Adoption and Opt Outs

We observed a substantial increase in GPP String implementations reflecting the IAB's **transition from USPS to GPP (RQ4)**. In December 2023, `uspv1`, the deprecated USPS section within the GPP String, was the most common section, followed by `usnatv1` (Figure 10). While the number of sites with a `usnatv1` or `uspv1` section increased slightly in February 2024, the frequency of the `uscav1` section, which is replacing the deprecated USPS, nearly tripled and became the most common section. Most of the new instances came from sites that did not have a GPP String in December 2023 and implemented a GPP String that only includes the `uscav1` section in February 2024. It appears these sites use Google Funding Choices to implement GPP indicating the impact a single service can have.[14] In April 2024, the `uspv1` section became the most common section, followed by `uscav1`. Interestingly, in all three crawls, the `uspv1` section rarely occurred alone in a GPP String but usually paired with at least one other section indicating a period of transition and spillover effects.

Considering only sites with a GPP String in each crawl, the rate of GPP adoption between December 2023 and February 2024 was 218 sites per month and decreased to 72 sites per month between February and April 2024. Even if the initial rate of 218 sites per month resumes, it would take about four years for the remaining sites in our crawl set to implement a GPP String, and only about 72% of sites would have a section relevant to California. In fact, most sites do not make use of the GPC option that is already available to them. The

[14] Funding Choices was Google's CMP created in 2017 with the intent of helping publishers recover ad revenue lost to ad blockers [12]. In 2020, Funding Choices added functionality to allow publishers to communicate with site visitors regarding CCPA and GDPR consent [41]. While Funding Choices has since been integrated into other Google advertising platforms [24], the analyzed sites have a file, `m=kernel_loader,loader_js_executable`, which serves as the GPP CMPAPI for those sites and also references Funding Choices objects. This implementation may not explain the increase for all sites. However, we manually spot-checked and inspected this file for 20 sites. These results are from the time of our crawls and may have changed since.

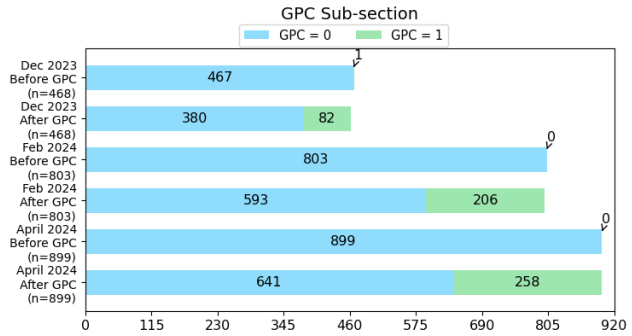

Figure 12: The value of the GPC subsection found in the `usnatv1` or `uscav1` sections before and after sites receive a GPC signal. A value of `1` indicates that the site has received a GPC signal, and a value of `0` indicates that the site has not.

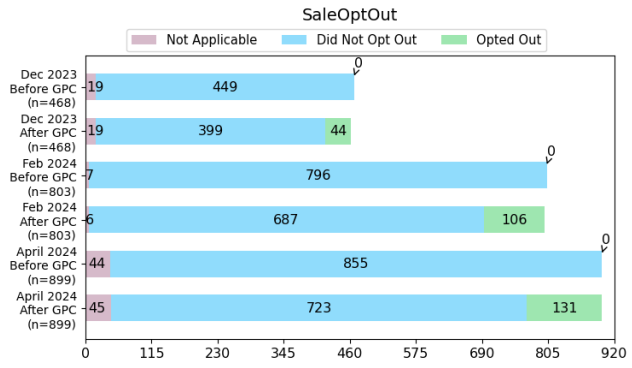

Figure 13: Opt outs via the `SaleOptOut` field in the `usnatv1` and `uscav1` sections of the GPP String. For opting out via the `SharingOptOut` field see Appendix A.6, Figure 16.

`usnatv1` and `uscav1` sections include a GPC subsection that stores whether the site has received a GPC signal.[15] Per the IAB, sites "should" check for a GPC signal and store its value in this subsection [35]. However, only 18% (82/468), 26% (206/803), and 29% (258/899) of sites did so in December 2023, February 2024, and April 2024, respectively (Figure 12). Thus, a large percentage of sites are either not looking for a GPC signal or omit logging it in this subsection.

For evaluating GPP opt outs in California, where GPC compliance is required and enforced [75], we consider sites with a `uscav1` or `usnatv1` GPP section. GPC selling opt out rates via the GPP String are low (Figure 13). In December 2023, after a GPC signal was sent, 9% (44/468) of sites opted the user out of selling. These rates increased to 13% (106/803) in February 2024 and 15% (131/899) in April 2024. These rates remain much lower than those of the OptanonConsent cookie

[15] Colorado's `uscov1` and Connecticut's `usctv1` sections also have the GPC subsection as both states recognize GPC as a valid opt out mechanism.

Privacy Strings	Number Occurrences		
	Jan 2024	Feb 2024	April 2024
`.well-known/gpc.json` only	152	156	164
`.well-known/gpc.json`, USPS	7	5	4
`.well-known/gpc.json`, GPP	4	2	2
`.well-known/gpc.json`, USPS, GPP	1	3	4
`.well-known/gpc.json`, OAC	17	16	18
`.well-known/gpc.json`, USPS, GPP, OAC	0	0	1
Total	181	182	193

Table 3: Overlap between `.well-known/gpc.json` implementation and other privacy strings. (USPS = US Privacy String, GPP = GPP String, OAC = OptanonConsent cookie.)

(§4.2.1) and the USPS (§4.2.2). One reason could be that the OptanonConsent cookie and USPS are easy to understand; OneTrust handles the implementation of the OptanonConsent cookie and the USPS is only four characters. There is evidence that implementing a GPP String is challenging as it requires understanding how to encode, decode, and construct various sections of the string [67]. The IAB extended the deprecation date of the USPS twice to give publishers more time to implement GPP (Appendix A.1.2). Thus, lower opt out rates may be, partly, due to the complexity of GPP. The online ad ecosystem's general slow pace may also play a role.

4.2.4 `.well-known/gpc.json` Adoption

We ran three crawls to examine `.well-known/gpc.json` adoption. The optional `.well-known/gpc.json` allows sites to declare to the world that they are respecting GPC by posting a public JSON file [85]. However, as shown in Table 3, fewer than 200 sites made use of this option in each crawl. 3 (Jan 2024), 2 (Feb 2024), and 2 (April 2024) sites in the `.well-known/gpc.json` only category of Table 3 set the value of `gpc` to `False`. The remaining sites set it to `True`. There is not much overlap of sites with a `.well-known/gpc.json` and other privacy strings. This finding could mean that a number of sites use custom implementations for GPC compliance as they do not implement the USPS, GPP String, or OptanonConsent cookie but still claim compliance via the `.well-known/gpc.json`. However, sites could also implement another privacy string we do not cover here. It may also be the case that compliance rates are actually higher than we found via privacy strings. Generally, this uncertainty shows the difficulty of determining GPC compliance from the outside, especially, for custom implementations.

4.3 Inconsistent Privacy String Opt Outs

Because of the lack of overlap between other privacy strings and `.well-known/gpc.json`, we set aside implementations of the `.well-known/gpc.json` in this section, considering only implementations of the other three privacy strings. A number of sites implemented more than one of these privacy strings. We consider a privacy string to be implemented if

Opt Out	Dec 2023			Feb 2024			April 2024		
	USPS + GPP	USPS + OAC	GPP + OAC	USPS + GPP	USPS + OAC	GPP + OAC	USPS + GPP	USPS + OAC	GPP + OAC
Both	8	79	10	14	48	15	23	42	24
Neither	221	78	3	317	14	7	463	11	16
USPS only	**285**	20	-	**368**	14	-	**300**	15	-
GPP only	0	-	0	1	-	0	4	-	0
OAC only	-	5	3	-	4	3	-	7	3
Total	514	182	16	700	80	25	790	75	43

Table 4: Opt out counts of crawl set sites with exactly two privacy strings. The shaded rows indicate counts of inconsistent opt outs. Note that for GPP the opt out must be in the `uscav1` or `usnatv1` sections of the string to be counted. Sites in this table are mutually exclusive from sites in Table 5 and Appendix, Table 7.

Opt Out	Dec 2023	Feb 2024	April 2024
All Opt Out	24	69	72
None Opt Out	33	97	102
OAC Opt Out Only	1	1	1
OAC + USPS Opt Out Only	**8**	2	2
OAC + GPP Opt Out Only	0	**4**	5
GPP Opt Out Only	0	0	0
USPS Opt Out Only	1	2	3
USPS + GPP Opt Out Only	2	2	2
Total	69	177	187

Table 5: Opt out counts of crawl set sites with exactly three privacy strings. The shaded rows indicate inconsistent opt outs. Sites in this table are mutually exclusive from sites in Table 4 and Appendix, Table 7.

it has a non-null value. We further consider it to opt out if it satisfies the criteria for opting out per §3.3. In December 2023, 11% (325/3,074) of sites had inconsistent opt outs. This percentage increased to 12% (400/3,246) in February 2024 before decreasing to 10% (338/3,402) in April 2024.[16]

4.3.1 Sites with Exactly Two Privacy Strings

Considering sites with exactly two privacy strings, Table 4 shows that sites with a USPS and GPP String predominantly either opt out via USPS only or not at all. For instance, per the first column, there were 514 total sites in December 2023 with both a USPS and GPP String. 285 of these sites opted out via USPS only after receiving a GPC signal. This lack of GPP opt outs is partially due to 30% of sites in this category lacking the relevant California sections, `usnatv1` or `uscav1`, and thus, they cannot meet our definition of opting out. For instance, a group of 51 MediaNews Group [50] sites opted out via the USPS in all three crawls but did not have the `usnatv1` or `uscav1` section in their GPP Strings. The GPP's fairly recent introduction and its complexity could be reasons for the divergence between USPS and GPP opt outs. There were also differences in the number of sites that opted out via USPS only. Rather than being a result of many sites individually changing their behavior, these differences were primarily caused by big publishers. For example, a group of 41 Lee Enterprises [44] sites did not opt out in December 2023, opted out in February 2024, and then did not opt out again in April 2024. Also, a group of at least 75 Townsquare Media [77] sites opted out in February 2024 but not in April 2024.[17]

4.3.2 Sites with Exactly Three Privacy Strings

Considering sites with exactly three privacy strings, Table 5 shows that sites with such implementation are likely to opt out either via all privacy strings or none. All sites with three privacy strings implemented either a `uscav1` or `usnatv1` section in their GPP String. The biggest inconsistency was 8 sites in the OAC + USPS Opt Out Only category in December 2023. 7/8 (88%) sites were Penske Media Corporation (PMC) [63] sites. Their GPP implementation was corrected for the next two crawls. However, there were additional PMC sites that did not opt out via any privacy string in December 2023 but opted out via all privacy strings in February and April 2024. This result suggests that deployment of privacy string implementations to a group of sites may not be a trivial task as there may be edge cases when automating this process. The 4 sites in the OAC + GPP Opt Out Only category in February 2024 had USPS values of `1---` indicating that the sites deemed the CCPA not applicable. These sites had the same behavior in April 2024. The None Opt Out category includes sites from Nexstar Media Group, Inc. [57], which had only a USPS and OptanonConsent cookie in December 2023 and implemented a GPP String sometime before February 2024. While they did not opt out via any string in any crawl, they updated the GPP GPC subsection after receiving a GPC signal.

4.4 Identifying Big Publishers that Handle GPP Implementation for Multiple Sites

As discussed (§4.3), a substantial number of sites is impacted by GPP String settings that are handled by big publishers for multiple sites. Such publishers may have tens, hundreds, or even thousands of sites. They may implement the same GPP String on each, possibly, relying on a library to set its values. To understand the impact of multi-site rollouts of GPP Strings for opt out compliance we set out to identify big publishers that handle GPP implementation.[18]

[16] In the following we discuss opt out counts of crawl set sites with exactly two or three privacy strings. Appendix A.7, Table 7 contains further details on the opt out counts of crawl set sites with exactly one privacy string.

[17] Our extension failed to analyze most of the Townsquare Media sites in December 2023. Of the Townsquare Media sites that were analyzed in December 2023, most did not opt out. Appendix, Table 6 shows details of the analysis errors we encountered during our crawls. Note that, generally, any change in behavior of a site may not be due to a change in implementation but rather a result of fluctuation during different site loads.

[18] Appendix A.5 contains our protocol for this identification task.

When constructing their GPP String, publishers select sections as well as field values within each section, which are then encoded to form the GPP String. To determine common choices for combinations of sections and field values in GPP Strings, we compared the encoded strings (i.e., the string characters) and found 46 unique strings in December 2023, 55 unique strings in February 2024, and 57 unique strings in April 2024. In each crawl, the five most frequent GPP Strings accounted for over half of all GPP String instances. We looked at a few different prevalent GPP Strings and found that a small number of big publishers has outsize impact.

For instance, there are 188 sites in our crawl set with a GPP String value of "DBABBg~BUoAAAKA.QA" before receiving a GPC signal in February 2024.[19] Most of these sites are published by the USA Today Network [80], Nexstar Media Group, Inc. [57], and PMC [63]. Similarly, in February 2024, Raptive [66] sites all had a GPP String value of "DBABzw~~BVQqAAAAAgA" before receiving a GPC signal.[20] We noticed that when we reloaded a Raptive site twice after sending a GPC signal, the GPP String would opt out of sale and sharing.

We informed Raptive staff of this finding. They responded that the code to check for GPC signals was built into the wrong JavaScript file and ran too late after the first page load, however, only affected the first page load of the first session of a new user. They informed us that after we reported the bug they fixed it, and GPC now works correctly on all page loads, including the first. Since our extension only loaded a site once before and once after sending a GPC signal, we counted all Raptive sites as not opting out in our crawls. There are 192 Raptive sites in our crawl set. Given the correction, the April 2024 opt out rates for the selling of personal information increase from 15% (131/899) (§4.2.3) to 36% (323/899) showing the impact of a single privacy string implementation of a big publisher on the overall opt out compliance rates.

5 Discussion

Effective enforcement is critical for making the opt out right meaningful on the web and beyond. However, bridging the gap between legal requirements and technological implementations requires significant effort and technical expertise on part of the regulators, e.g., for determining whether a site is subject to the CCPA opt out right (§3.2). Our main **recommendation for regulators (RQ5)** is to focus on big publishers to alert them of their obligations and fix incorrect GPC opt out implementations (§4.3 and §4.4). Many small and medium-sized businesses rely on big publishers for their compliance implementations. By addressing the source of these issues, regulators can efficiently drive widespread improvements. This strategy can be complemented by enforcement actions against high-profile publishers of individual sites to broadcast to the industry as a whole to take compliance seriously. Furthermore, it is important that regulators provide education and guidance to businesses and users.

While some businesses may be delaying their GPC implementations deliberately, especially, as a big part of the industry is still non-compliant (§4.2), others may simply be unaware of their obligations or face technical hurdles. The latter point is illustrated by the many inconsistent privacy string implementations we found (§4.3). We think it would be useful if regulators reach out to the IAB and other industry organizations as well as CMPs requesting that those remind their members and customers of their obligations and where they can find guidance. There are already a number of useful resources available, e.g., on `https://globalprivacycontrol.org/` or published by the Office of the Connecticut Attorney General [59]. To help in this effort we make our software and data publicly available (§8). For example, using our crawler, publishers with multiple sites can efficiently evaluate the compliance status of their entire portfolio. In addition to educating businesses, raising awareness among users and the tools they can use to exercise their rights is equally important.

The effectiveness of opt out compliance depends on the further evolution of the online ad ecosystem and its integration into the web platform. To evaluate whether a site respects a user's opt out choices we relied on privacy strings. However, it is not guaranteed that every site implements those as required. Further, even if that were the case, sites receiving personal information may simply disregard the privacy strings they receive. The only way for an outside observer to determine compliance in the current nontransparent environment is through experimentation, for example, by adding products to shopping carts on retailer sites with GPC enabled and checking whether those products later show up in ads on other sites. However, such experiments do not solve the core problem. To fundamentally improve data transparency and privacy enforcement effectiveness on the web platform a comprehensive reform of the online ad ecosystem is necessary.

6 Conclusions

Our results show that GPC adoption is gradually increasing and corresponding to the evolving privacy law landscape. Adapting to new privacy laws requires expertise and resources. Especially, smaller businesses may not be aware of their obligations. Some businesses may also perceive the risk of noncompliance to be low. Others may face technical challenges, especially, if they need to roll out changes to a large number of sites. Our findings highlight the importance of effective enforcement. Ultimately, opt out compliance is reliant on the evolution of the online ad ecosystem overall. Big improvements require a systems solution.

[19] This GPP String contains the `uscav1` section and does not opt out of sale or sharing. The string can be decoded at `https://iabgpp.com/`.

[20] This GPP String contains the `uspv1` and `usnatv1` sections. The `uspv1` section is empty, and the `usnatv1` does not opt out of sale or sharing.

Acknowledgments

We thank our anonymous reviewers and our shepherd for their improvement suggestions. We also thank Wesleyan University students Ebuka Akubilo and Samir Cerrato for refining our crawler and Don Marti from Raptive for sharing his insights from a publisher's perspective. We are grateful to the National Science Foundation for their support of this research (Award #2055196). We also thank Wesleyan University, its Department of Mathematics and Computer Science, and the Anil Fernando Endowment for their additional support. Conclusions reached or positions taken are our own and not necessarily those of our supporters, its trustees, officers, or staff.

7 Ethics Considerations

We consider the ethical risks of our study to be low. We only visited each site a few times and did not extensively probe or penetrate sites' security or privacy protections. We only collected publicly available data. The data we shared with the sites and third parties was our own and did not involve data of any external research participants. We notified each publisher named in our study and gave them an opportunity to comment. We only received a substantive reply from Raptive, as described. All legal analysis was performed by the last author who is admitted to the California bar association (inactive status) and has expertise in privacy law. We are currently working with three regulators identifying non-compliant sites based on the methodology and results of this study. For confidentiality reasons we are unable to share further details.

8 Open Science

The code of our web crawler and browser extension, Python script for decoding GPP Strings, crawl set, and privacy policy analysis results as well as the policies themselves are publicly available under the MIT license [30].

References

[1] M. S. Alvim, N. Fernandes, A. McIver, and G. H. Nunes. A Quantitative Information Flow Analysis of the Topics API. In *WPES*, pages 123–127, New York, NY, USA, 2023. ACM.

[2] M. A. B. Aziz and C. Wilson. Johnny Still Can't Opt-out: Assessing the IAB CCPA Compliance Framework. In *PETS*, volume 4, pages 349–363, Bristol, UK, July 2024. PETS.

[3] V. Bannihatti Kumar, R. Iyengar, N. Nisal, Y. Feng, H. Habib, P. Story, S. Cherivirala, M. Hagan, L. Cranor, S. Wilson, F. Schaub, and N. Sadeh. Finding a Choice in a Haystack: Automatic Extraction of Opt-Out Statements from Privacy Policy Text. In *The Web Conference*, pages 1943–1954, New York, NY, USA, 2020. ACM.

[4] R. Berjon. GPC under the GDPR. https://berjon.com/gpc-under-the-gdpr/, 2021. Accessed: January 30, 2025.

[5] Y. Beugin and P. McDaniel. Interest-disclosing Mechanisms for Advertising are Privacy-Exposing (not Preserving). In *PETS*, volume 1, pages 41–57, Bristol, UK, July 2024. PETS.

[6] BuiltWith. Websites Using Global Privacy Control. https://trends.builtwith.com/websitelist/Global-Privacy-Control/. Accessed: January 30, 2025.

[7] BuiltWith. Websites Using US Privacy User Signal Mechanism. https://trends.builtwith.com/websitelist/US-Privacy-User-Signal-Mechanism/. Accessed: January 30, 2025.

[8] BuiltWith. Builtwith. https://builtwith.com/, 2025. Accessed: January 30, 2025.

[9] California Legislative Information. California Online Privacy Protection Act. https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=8.&chapter=22.&lawCode=BPC. Accessed: January 30, 2025.

[10] C. Carpineto, D. Lo Re, and G. Romano. Automatic Assessment of Website Compliance to the European Cookie Law with CooLCheck. In *WPES*, pages 135–138, New York, NY, USA, 2016. ACM.

[11] J. Charatan and E. Birrell. Two Steps Forward and One Step Back: The Right to Opt-out of Sale under CPRA. In *PETS*, volume 2, pages 91–105, Bristol, UK, July 2024. PETS.

[12] V. Chirravuri. Helping publishers recover lost revenue from ad blocking. https://blog.google/technology/ads/helping-publishers-recover-lost-revenue-ad-blocking/, 2018. Accessed: January 30, 2025.

[13] Cloudflare. Cloudflare. https://cloudflare.com/. Accessed: January 30, 2025.

[14] L. F. Cranor, B. Dobbs, S. Egelman, G. Hogben, J. Humphrey, M. Langheinrich, M. Marchiori, M. Presler-Marshall, J. M. Reagle, M. Schunter, D. A. Stampley, and R. Wenning. The Platform for Privacy Preferences 1.1 (P3P1.1) specification. https://www.w3.org/TR/P3P11/, November 2006.

[15] L. F. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. M. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) specification. https://www.w3.org/TR/P3P/, April 2002.

[16] T. Crepax. Global Privacy Control and Portability of Privacy Preferences Through Browser Settings: A Comparative Study of Techno-Legal Challenges Under the

CCPA/CPRA and the GDPR. http://dx.doi.org/10.2139/ssrn.4710372, 2024. Accessed: January 30, 2025.

[17] A. Datta, M. C. Tschantz, and A. Datta. Automated Experiments on Ad Privacy Settings. In *PETS*, volume 1, pages 92–112, Philadelphia, PA, USA, June 2015. De Gruyter Open/Sciendo.

[18] 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 *NDSS*. Internet Society, 2019.

[19] Disconnect. Tracking protection lists. https://disconnect.me/trackerprotection. Accessed: January 30, 2025.

[20] S. Englehardt and A. Narayanan. Online tracking: A 1-million-site measurement and analysis. In *CCS*, pages 1388–1401, New York, NY, USA, 2016. ACM.

[21] R. T. Fielding and D. Singer. Tracking Preference Expression (DNT). https://www.w3.org/TR/tracking-dnt/, 2019. Accessed: January 30, 2025.

[22] A. Folks. US State Privacy Legislation Tracker. https://iapp.org/resources/article/us-state-privacy-legislation-tracker/. Accessed: January 30, 2025.

[23] I. Fouad, N. Bielova, A. Legout, and N. Sarafijanovic-Djukic. Missed by Filter Lists: Detecting Unknown Third-Party Trackers with Invisible Pixels. In *PETS*, volume 2, pages 499–518, online, July 2020. De Gruyter Open/Sciendo.

[24] Google. Funding Choices has moved. https://support.google.com/fundingchoices/answer/9010669?hl=en. Accessed: January 30, 2025.

[25] Google. Puppeteer. https://pptr.dev/. Accessed: January 30, 2025.

[26] H. Habib and L. F. Cranor. Evaluating the Usability of Privacy Choice Mechanisms. In *SOUPS*, pages 273–289, Boston, MA, USA, August 2022. USENIX Association.

[27] H. Habib, M. Li, E. Young, and L. Cranor. "Okay, whatever": An Evaluation of Cookie Consent Interfaces. In *CHI*, New Orleans, LA, USA, 2022. ACM.

[28] H. Habib, Y. Zou, A. Jannu, N. Sridhar, C. Swoopes, A. Acquisti, L. F. Cranor, N. Sadeh, and F. Schaub. An Empirical Analysis of Data Deletion and Opt-Out Choices on 150 Websites. In *SOUPS*, pages 387–406, Santa Clara, CA, USA, August 2019. USENIX Association.

[29] K. Hausladen. Investigating the Current State of CCPA Compliance on the Internet. MA thesis, Wesleyan University, Middletown, Connecticut, May 2024. https://doi.org/10.14418/wes01.2.451. Accessed: January 30, 2025.

[30] K. Hausladen, E. Akubilo, M. May, F. Wijaya, J. Wang, O. Wang, S. Eng, and S. Zimmeck. GPC Web Crawler. https://doi.org/10.5281/zenodo.14729170. Accessed: January 30, 2025.

[31] M. Hils, D. W. Woods, and R. Böhme. Privacy Preference Signals: Past, Present and Future. In *PETS*, volume 4, pages 249–269, Online, July 2021. De Gruyter Open/Sciendo.

[32] IAB. IAB Second Amended and Restated Multi-State Privacy Agreement. https://iabtechlab.com/wp-content/uploads/2024/01/IAB-Second-Amended-and-Restated-Multi-State-Privacy-Agreement-MSPA.pdf, 2024. Accessed: January 30, 2025.

[33] IAB Tech Lab. Global-Privacy-Platform. https://github.com/InteractiveAdvertisingBureau/Global-Privacy-Platform. Accessed: January 30, 2025.

[34] IAB Tech Lab. Global Privacy Platform String. https://github.com/InteractiveAdvertisingBureau/Global-Privacy-Platform/blob/main/Core/Consent%20String%20Specification.md. Accessed: January 30, 2025.

[35] IAB Tech Lab. GPP Extension: California Privacy Technical Specification. https://github.com/InteractiveAdvertisingBureau/Global-Privacy-Platform/blob/main/Sections/US-States/CA/GPP%20Extension%3A%20California%20Privacy%20Technical%20Specification.md. Accessed: January 30, 2025.

[36] IAB Tech Lab. IAB CCPA Compliance Framework for Publishers & Technology Companies. https://iabtechlab.com/standards/ccpa/. Accessed: January 30, 2025.

[37] IAB Tech Lab. IAB GPP Encoder/Decoder. https://iabgpp.com/. Accessed: January 30, 2025.

[38] IAB Tech Lab. iabgpp-es. https://www.npmjs.com/package/@iabgpp/cmpapi. Accessed: January 30, 2025.

[39] IAB Tech Lab. US Privacy String. https://github.com/InteractiveAdvertisingBureau/USPrivacy/blob/master/CCPA/US%20Privacy%20String.md. Accessed: January 30, 2025.

[40] IAB Tech Lab. USPrivacy. https://github.com/InteractiveAdvertisingBureau/USPrivacy. Accessed: January 30, 2025.

[41] V. Johnsen. Helping publishers manage consent with funding choices. https://blog.google/products/admanager/helping-publishers-manage-consent-funding-choices/, 2020. Accessed: January 30, 2025.

[42] R. Khandelwal, A. Nayak, H. Harkous, and K. Fawaz. Automated Cookie Notice Analysis and Enforcement. In *USENIX Security*, pages 1109–1126, Anaheim, CA, August 2023. USENIX Association.

[43] B. Krumnow, H. Jonker, and S. Karsch. How gullible are web measurement tools? A case study analysing and strengthening OpenWPM's reliability. In *CoNEXT*, pages 171–186, Roma, Italy, 2022. ACM.

[44] Lee Enterprises. Lee Enterprises. https://lee.net. Accessed: January 30, 2025.

[45] V. Lefrere, L. Warberg, C. Cheyre, V. Marotta, and A. Acquisti. The impact of the GDPR on content providers. https://ideas.repec.org/p/hal/journl/hal-03111801.html, December 2020. Accessed: January 30, 2025.

[46] Z. Liu, U. Iqbal, and N. Saxena. Opted Out, Yet Tracked: Are Regulations Enough to Protect Your Privacy? In *PETS*, volume 1, pages 280–299, Bristol, UK, July 2024. PETS.

[47] C. Matte, N. Bielova, and C. Santos. Do Cookie Banners Respect my Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *S&P*, pages 791–809. IEEE, 2020.

[48] MDN Web Docs. cookies. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API/cookies. Accessed: January 30, 2025.

[49] MDN Web Docs. urlclassification. https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/API/webRequest/onHeadersReceived#urlclassification. Accessed: January 30, 2025.

[50] MediaNews Group. MediaNews Group. https://www.medianewsgroup.com. Accessed: January 30, 2025.

[51] Microsoft. Playwright. https://playwright.dev/. Accessed: January 30, 2025.

[52] J. Moscow. US Privacy Signal Deprecation Deadline Extended to January 31, 2024. https://iabtechlab.com/us-privacy-signal-deprecation-deadline-extended-to-january-31-2024/. Accessed: January 30, 2025.

[53] J. Moscow. US Privacy Signal Deprecation Deadline Extended to September 30, 2023. https://iabtechlab.com/us-privacy-signal-deprecation-deadline-extended-to-september-30-2023/. Accessed: January 30, 2025.

[54] Mozilla. Enhanced Tracking Protection in Firefox for Desktop. https://support.mozilla.org/en-US/kb/enhanced-tracking-protection-firefox-desktop. Accessed: January 30, 2025.

[55] Mozilla. Firefox. https://www.mozilla.org/en-US/firefox/channel/desktop/. Accessed: January 30, 2025.

[56] Mullvad VPN. Mullvad VPN. https://mullvad.net/en. Accessed: January 30, 2025.

[57] Nexstar Media Group, Inc. Nexstar Media Group, Inc. https://www.nexstar.tv. Accessed: January 30, 2025.

[58] Office of the Colorado Attorney General. Universal Opt-Out Shortlist. https://coag.gov/uoom/. Accessed: January 30, 2025.

[59] Office of the Connecticut Attorney General. The Connecticut Data Privacy Act. https://portal.ct.gov/ag/sections/privacy/the-connecticut-data-privacy-act. Accessed: January 30, 2025.

[60] OneTrust. OneTrust. https://www.onetrust.com/. Accessed: January 30, 2025.

[61] OneTrust. OneTrust Cookies. https://my.onetrust.com/articles/en_US/Knowledge/UUID-2dc719a8-4be5-8d16-1dc8-c7b4147b88e0. Accessed: January 30, 2025.

[62] OneTrust. Cookie Consent Integration with Google Tag Manager. https://my.onetrust.com/s/article/UUID-301b21c8-a73a-05e8-175a-36c9036728dc, July 2024. Accessed: January 30, 2025.

[63] Penske Media Corporation. Penske Media Corporation. https://www.pmc.com. Accessed: January 30, 2025.

[64] V. L. Pochat, T. V. Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen. Tranco. https://tranco-list.eu/. Accessed: January 30, 2025.

[65] privacy-tech-lab. OptMeowt. https://github.com/privacy-tech-lab/gpc-optmeowt/tree/v3.0.0-paper. Accessed: January 30, 2025.

[66] Raptive. Raptive. https://raptive.com. Accessed: January 30, 2025.

[67] Reddit. IAB GPP Consent Strings. https://www.reddit.com/r/adops/comments/1384saa/iab_gpp_consent_strings/. Accessed: January 30, 2025.

[68] K. Reitz. Requests. https://pypi.org/project/requests/. Accessed: January 30, 2025.

[69] J. Ruohonen and V. Leppänen. Invisible Pixels Are Dead, Long Live Invisible Pixels! In *WPES*, pages 28–32, Toronto, Canada, January 2018. ACM.

[70] C. Santos, M. Nouwens, M. Toth, N. Bielova, and V. Roca. Consent Management Platforms Under the GDPR: Processors and/or Controllers? In *Privacy Technologies and Policy*, pages 47–69. Springer, 2021.

[71] F. Schaub, R. Balebako, A. L. Durity, and L. F. Cranor. A Design Space for Effective Privacy Notices. In *SOUPS*, pages 1–17, Ottawa, Canada, July 2015. USENIX Association.

[72] Similarweb. Similarweb. https://similarweb.com/. Accessed: January 30, 2025.

[73] Software Freedom Conservancy. WebDriver. https://www.selenium.dev/documentation/webdriver/. Accessed: January 30, 2025.

[74] State of California Department of Justice. California Consumer Privacy Act (CCPA). https://oag.ca.gov/privacy/ccpa. Accessed: January 30, 2025.

[75] State of California Department of Justice. Attorney General Bonta Announces Settlement with Sephora as Part of Ongoing Enforcement of California Consumer Privacy Act. https://oag.ca.gov/news/press-releases/attorney-general-bonta-announces-settlement-sephora-part-ongoing-enforcement, 2022. Accessed: January 30, 2025.

[76] State of California Legislative Information. Assembly Bill No. 375. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375, 2018. Accessed: January 30, 2025.

[77] Townsquare Media. Townsquare Media. https://www.townsquaremedia.com. Accessed: January 30, 2025.

[78] V. H. Tran, A. Mehrotra, M. Chetty, N. Feamster, J. Frankenreiter, and L. Strahilevitz. Measuring Compliance with the California Consumer Privacy Act Over Space and Time. In *CHI*, Honolulu, HI, USA, 2024. ACM.

[79] M. Trevisan, S. Traverso, H. Metwalley, and M. Mellia. 4 Years of EU Cookie Law: Results and Lessons Learned. In *PETS*, volume 2, pages 126–145, Stockholm, Sweden, 2019. De Gruyter Open/Sciendo.

[80] USA Today Network. USA Today Network. https://cm.usatodaynetwork.com/myusatodaynetwork. Accessed: January 30, 2025.

[81] M. Zhang, W. Meng, Y. Zhou, and K. Ren. CSChecker: Revisiting GDPR and CCPA Compliance of Cookie Banners on the Web. In *ICSE*, pages 1–12, Lisbon, Portugal, 2024. ACM.

[82] S. Zimmeck. Standardizing Global Privacy Control (GPC). https://github.com/privacycg/proposals/issues/10, 2020. Accessed: January 30, 2025.

[83] S. Zimmeck and K. Alicki. Standardizing and Implementing Do Not Sell. In *WPES*, pages 15–20, Online, November 2020. ACM.

[84] S. Zimmeck, E. Kuller, C. Ma, B. Tassone, and J. Champeau. Generalizable Active Privacy Choice: Designing a Graphical User Interface for Global Privacy Control. In *PETS*, pages 258–279, Bristol, UK, July 2024. PETS.

[85] S. Zimmeck, P. Snyder, J. Brookman, and A. Zucker-Scharff. Global Privacy Control (GPC). https://w3c.github.io/gpc/. Accessed: January 30, 2025.

[86] S. Zimmeck, O. Wang, K. Alicki, J. Wang, and S. Eng. Usability and Enforceability of Global Privacy Control. In *PETS*, volume 2, pages 1–17, Lausanne, Switzerland, July 2023. PETS.

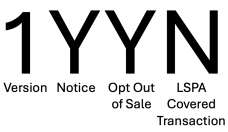

Figure 14: An example USPS indicating (1) version 1 is used, (2) the site gave the user notice of the CCPA and the opportunity to opt out, (3) the user opted out of sale, and (4) the site is not operating under the IAB's Limited Service Provider Agreement. Our approach focuses on the third character to detect whether a site respects GPC (§3.3).

A Appendix

A.1 GPC and Privacy String Background

A.1.1 GPC Signals

GPC was developed by a coalition of privacy-minded academics, browser vendors, web publishers, CMPs, and non-governmental organizations to help people exercise their opt out rights on the web and other platforms [85]. In contrast to site-by-site DNSLs, this binary signal is intended to provide a "comprehensive option that broadly signals [a consumer's] opt-out request" [74]. The OAG states that GPC signals "must be honored by covered businesses as a valid consumer request to stop the sale or sharing of personal information" and enforces this requirement [74, 75]. The first enforcement of GPC by the OAG happened in August 2022 against Sephora resulting in a settlement in which Sephora was ordered to pay a $1.2 million penalty and reform its privacy practices to comply with the CCPA [75]. GPC is also adopted as a universal opt out mechanism in Colorado [58] and Connecticut [59]. To date, 19 US states have passed comprehensive privacy laws [22]. 12 of these laws provide for opt out rights to be exercised via universal opt out mechanisms, such as GPC.[21]

A.1.2 Privacy Strings

The USPS and GPP String In November 2019, the IAB provided the USPS (Figure 14), a four-character string that communicates a user's opt out status and other CCPA requirements [39]. The third character of this string indicates the user's opt out status; a value of `Y` means the user was opted out, a value of `N` means the user was not opted out, and a value of "–" means the CCPA does not apply.

The USPS can be implemented with either the IAB's client-side JavaScript USPAPI or an HTTP cookie [86]. The IAB released the USPS's successor, the GPP String, on September 28, 2022 in an effort to support additional privacy laws [33]. Sites using the USPS were expected to transition to the GPP

[21] The states are: CA, CO, CT, NJ, OR, MT, NE, TX, MN, MD, DE, NH.

String by the USPS's deprecation date. Originally, the deprecation date was July 1, 2023, but the IAB pushed the date back twice, for a final deprecation date of January 31, 2024 [52, 53]. The GPP String has sections for privacy laws in Canada, the EU, and multiple US states.

The `uscav1` section specifically supports CCPA requirements and replaces the deprecated USPS. Site operators also have the option to simultaneously comply with all US state privacy laws by following the US National Approach and implementing the `usnatv1` section.[22] Since we are concerned with the opt out right in California, we focus on the `SaleOptOut` and `SharingOptOut` fields of the `uscav1` and `usnatv1` sections. For both of these fields, a value of `1` means the user was opted out, a value of `2` means the user was not opted out, and a value of `0` means this field is not applicable [35].

GPP String Decoding Because site operators can include multiple regional sections in a GPP String, it is possible that GPP Strings can be "too long for certain applications" [34]. To manage string length, GPP Strings are encoded using Fibonacci encoding [34]. We collected encoded GPP Strings during our crawls, as described in §3.3, and decoded them to determine user opt out status. The IAB provides a JavaScript library that websites can use to handle encoding and decoding of GPP Strings [38]. The IAB also maintains a website with a form that will encode and decode single GPP Strings [37]. Since we used Python to perform our data analysis and needed to decode hundreds of GPP Strings at once, neither of the IAB's options suited our needs. Thus, we converted the IAB's JavaScript library to a Python script for decoding GPP Strings during our data analysis.

The OptanonConsent Cookie Some site operators outsource privacy law compliance to CMPs or similar platforms that provide a library to capture people's privacy preference signals and obtain consent. One such CMP is OneTrust [60]. OneTrust's OptanonConsent cookie is a first-party cookie that indicates the consent status of a site visitor. It can be found on sites that integrate the OneTrust Banner Content Delivery Network, OneTrust's cookie consent banner solution [61]. The OptanonConsent cookie has an `isGpcEnabled` field, which has a value of `1` when (1) GPC is enabled, meaning the site responds to GPC signals, and (2) the site has received a GPC signal from a user's browser. If GPC is not enabled or a GPC signal was not received from the browser, the value of the `isGpcEnabled` field will be `0` [61]. Some OptanonConsent cookies lack this field, which seems to be the result of combining OneTrust's `dataLayer` object with Google Tag Manager [62].

[22] The US National Approach is defined in section 1.81 of the IAB's Multi-State Privacy Agreement (MSPA) [32]. Under the US National Approach, sites must treat users generally as if each privacy law jurisdiction is applicable to them. In practice, this approach means that the site is complying with the most stringent set of opt out requirements.

A.2 Tranco Rank of Sites by BuiltWith Traffic Category

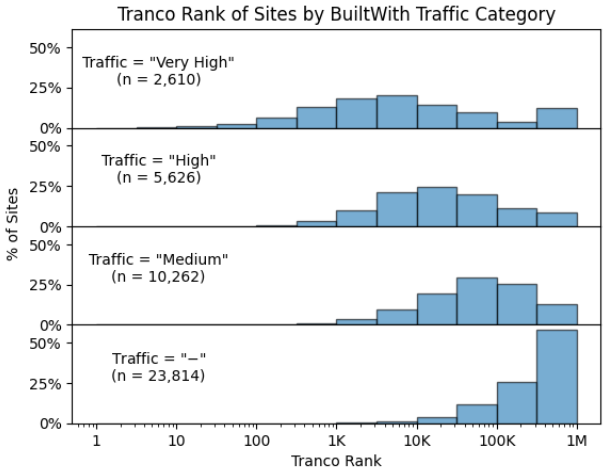

Figure 15: A histogram of Tranco rank ranges for sites in each of the four BuiltWith Traffic categories: "Very High," "High," "Medium," and "–." Sites with a rank of "1,000,000+" were given a rank of 1M. The x-axis is on log scale.

A.3 CCPA Definitions

- **Selling:** "selling, renting, releasing, disclosing, disseminating, making available, transferring, or otherwise communicating orally, in writing, or by electronic or other means, a consumer's personal information by the business to a third party for monetary or other valuable consideration" (CCPA, §1798.140(ad)(1))
- **Sharing:** "sharing, renting, releasing, disclosing, disseminating, making available, transferring, or otherwise communicating orally, in writing, or by electronic or other means, a consumer's personal information by the business to a third party for cross-context behavioral advertising, whether or not for monetary or other valuable consideration, including transactions between a business and a third party for cross-context behavioral advertising for the benefit of a business in which no money is exchanged" (CCPA, §1798.140(ah)(1))
- **Business:** "A sole proprietorship, partnership, limited liability company, corporation, association, or other legal entity that is organized or operated for the profit or financial benefit of its shareholders or other owners, that collects consumers' personal information, or on the behalf of which such information is collected and that alone, or jointly with others, determines the purposes and means of the processing of consumers' personal information, that does business in the State of California, and that satisfies one or more of the following thresholds:

Error	Dec 2023	Feb 2024	April 2024
Human Check	**307**	**307**	**186**
Reached Error Page	185	198	196
Timeout (on both loads)	184	171	170
Insecure Certificate	172	173	170
Unexpected Alert Open	1	2	4
WebDriver	1	0	1
Other	**168**	**42**	**79**
Total	1,018	893	806

Table 6: A detailed breakdown of the errors logged during crawls. Our crawler is able to catch these errors and continue its analysis. Human Check and Timeout errors are described in §3.4.2. A WebDriver error indicates that the WebDriver has failed to execute part of the script. An Unexpected Alert Open error indicates that a popup on the site disrupted Selenium's ability to analyze the site (such as a mandatory login).

(A) As of January 1 of the calendar year, had annual gross revenues in excess of twenty-five million dollars ($25,000,000) in the preceding calendar year, as adjusted pursuant to paragraph (5) of subdivision (a) of Section 1798.185.
(B) Alone or in combination, annually buys, sells, or shares the personal information of 100,000 or more consumers or households.
(C) Derives 50 percent or more of its annual revenues from selling or sharing consumers' personal information." (CCPA, §1798.140(d)(1))

A.4 Errors Caught During Crawls

Errors in the "Other" category of Table 6 are errors that we could not categorize otherwise. The decrease in this category appears to be largely due to the successful analysis of a group of Townsquare Media sites in February and April 2024 that failed in December 2023. We are unsure of why the analysis failed in December 2023. The primary reasons for other unidentified errors were that sites blocked script injection or redirected between multiple domains. There was a large decrease in human checks in April 2024. We are unsure of the cause. However, some variation in the particular sites that have a human check in each crawl is expected. For instance, while there were 307 sites with a human check in each of our first two crawls, only 270 sites had a human check in both of these crawls, and in total there were 372 sites that had a human check in at least one of our crawls. The remaining error frequencies stayed relatively consistent for all crawls.

A.5 Protocol for Identifying Big Publishers that Handle GPP Implementation for Multiple Sites

1. Filter the crawl set for sites with a specific GPP String.
2. Qualitatively assess whether these sites could be part of the same organization:
   (a) Look at all the site names and try to group them into categories (i.e., news, pop culture, blog).
   (b) Visit some sites in a category and look for evidence that the site is part of a larger organization. Usually, the page footer or privacy policy is indicative of this.
3. Determine whether sites are using the same files to set the GPP String.
   (a) Search in the Firefox Debugger for terms unique to GPP that could relate to setting a GPP String. For example, "usnat," "setgpp," or "SaleOptOut" could be productive search terms, while searching for "gpp" alone usually brings up instances of services looking for GPP String values.
   (b) Determine if any of the files found in the previous step actually set the GPP String. If so, visit more sites and see if the GPP String is set by the same code on those sites.

A.6 Sharing Opt Outs via GPP Strings

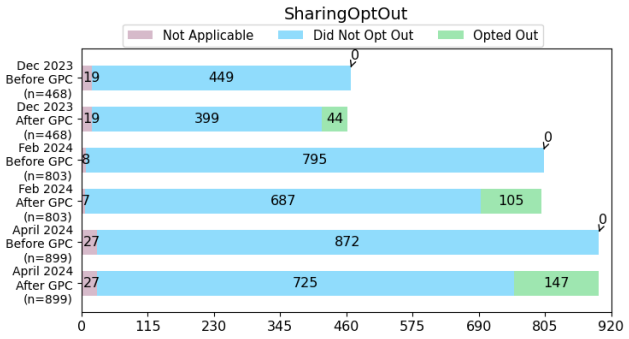

Figure 16: Opt outs via the `SharingOptOut` field in the `usnatv1` and `uscav1` sections of the GPP String.

A.7 Sites with Exactly One Privacy String

Opt Out	Dec 2023			Feb 2024			April 2024		
	USPS	GPP	OAC	USPS	GPP	OAC	USPS	GPP	OAC
Opt Out	236	0	905	258	0	915	311	0	986
No Opt Out	686	78	388	483	211	397	377	237	396
Total	922	78	1293	741	211	1312	688	237	1382

Table 7: Opt out counts of crawl set sites with exactly one privacy string. Sites in this table are mutually exclusive from sites with exactly two privacy strings in Table 4 and from sites with exactly three privacy strings in Table 5.