Crumbl: An awareness enhancing tool for cookie collection

Yi-Shyuan Chiang*
University of Illinois at Urbana-Champaign

Ho Shan Lam*
University of Illinois at Urbana-Champaign

Xingjian Zhang
University of Illinois at Urbana-Champaign

Eshwar Chandrasekharan
University of Illinois at Urbana-Champaign

**These authors contributed equally.*

Abstract

Websites use cookies to track and record users' behaviors. In the post General Data Protection Regulation era, data collection consent from end users is widely mandatory, resulting in a plethora of cookie decisions for users everyday. Some might install cookie management extensions to take care of these decisions, but such extensions tend to create a false sense of security when not every unwanted cookie is blocked. We present *Crumbl*, which documents cookie collection and displays the information in ways that are accessible to novice users. In this paper, we focused on the implementation and design details and made our source code available online. Going forward, *Crumbl* will allow individuals to better understand online privacy as they learn more about HTTP cookie collection through their newly gained insights.

1 Motivation

Data collection has become ubiquitous with the advancement of the Internet. Comprehensive data protection laws have been proposed around the world since GDPR granted users control of their shared data online, such as mandating the implementation of cookie consent dialogs. However, researchers have previously argued that it is impossible for users to make true informed consent with a simple consent tickbox [17].

To bridge this gap, studies have shown that browser extensions can help raise privacy awareness [21] and such awareness can lead to adoption of new privacy technology [20]. Yet, plugins on the market, e.g. Ghostery, Adblock, Privacy Badger, etc. can be overwhelming and potentially misunderstood by beginners who are unaware of the logic and the limits of such plugins, leading to a false sense of protection [26]. Therefore, we developed *Crumbl* to offer a beginner-friendly browser extension to surface real-time cookie collection, to help raise privacy awareness among users. Key functionalities that set *Crumbl* apart from other privacy browser extensions are: beginner-friendly interface, real-time cookie collection alerts, and browser cookie inspection.

2 Related Work

2.1 Transparency enhancing tools

Transparency enhancing tools (TETs) provide insights into how users' data is being collected, stored, processed and disclosed in an accurate and comprehensible way [13] . Hedom believed that TETs could be systems that give related information, provide accesses to the stored data, or provide counter profiling capabilities [10]. Janic, Wijbenga, and Veugen came up with a slightly different set of characteristics inspired by Hedom. TETs can provide user with information on how service providers "claim" to handle users' personal information, how service providers "actually" handle user's personal information, or the aforementioned information altogether [13]. In this article, we adopt Hedom's taxonomy to be more comprehensive, as Janic, Wijbenga, and Veugen's version is more of an elaboration of the first characteristic of Hedom's.

Privacy bird[1], the PRIME project[2], and primelife[3] are TETs that have been studied most frequently in the past [6, 29]. Unfortunately, older applications are either offline or ill-maintained. Therefore, in this paper we focus on taking inspiration from up-to-date TETs. Browser extensions and browser with native blocking mechanisms are two of the most common accessible TETs. Merzdovnik et al. surveyed the most common blocker plugins available: AdBlock, Ghostery, and

Copyright is held by the author/owner. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee.

USENIX Symposium on Usable Privacy and Security (SOUPS) 2022.
August 7–9, 2022, Boston, MA, United States.

[1] http://www.privacybird.org

[2] https://www.w3.org/2005/02/17-prime-pr/all.htm

[3] http://www.primelife.eu

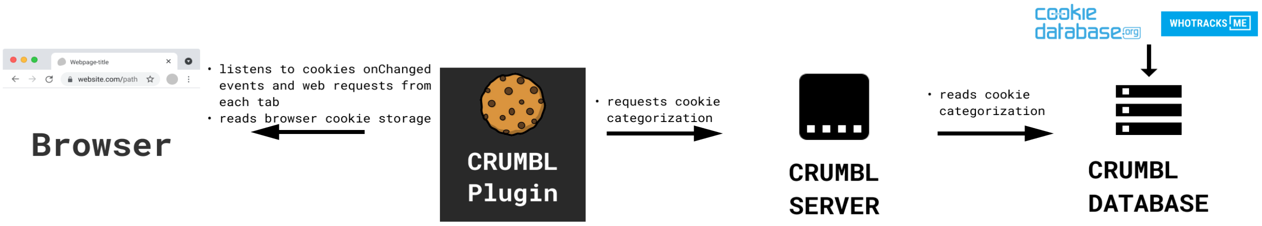

Webpage-title
website.com/path
Browser
• listens to cookies onChanged events and web requests from each tab
• reads browser cookie storage
CRUMBL Plugin
• requests cookie categorization
CRUMBL SERVER
• reads cookie categorization
cookie database.org
WHOTRACKS ME
CRUMBL DATABASE

Figure 1: Overview of Crumbl's system architecture. The extension module listens to cookies onChanged events and web requests originating from all tabs, aggregates real-time cookie information to the extension interface while querying the server for cookie categorization.

uBlock were three of the most downloaded plugins [22]. Kontaxis, Georgios and Chew, Monic analyzed the efficiency of Firefox's native blocker and reported that there is a 65% reduction in the cookies collected [16].

2.2 Impact of General Data Protection Rule

GDPR has brought changes to the privacy sphere as it grants data subjects multiple rights, such as the right of access, the right to data portability, and the right to explanation [2, 4, 28]. Privacy research has taken new directions after GDPR has taken effect. Momen, Matamian and Fritsh examined the effects of GDPR by comparing the mobile phone manifest data usage [23] while Schufrin et al. looked into the data made possible to access by GDPR with a visualization dashboard [27]. Scholars have also continued their endeavor to enhance the transparency level by visualizing personal data [1, 7]. Besides dashboards geared towards data subjects, some have proposed a GDPR-aligned information language and toolkit for developers [9], while others focused on the design of GDPR compliance tools and systems [19, 25]. On the contrary, a plethora of anti-GDPR browser extensions were launched to remove cookie consent dialogs automatically [15, 24]. In particular, "I don't care about cookies" has over 0.6 million downloads, this shows that GDPR still has loopholes in protecting privacy of the public and raising their awareness to data collection [15].

2.3 Analysis of privacy policy changes

To study the changes of privacy policy in large scales, most studies adopt website crawling and automated examination. Dabrowski et al. used the crawled data to study the changes in privacy policies in different jurisdiction by comparing the cookies requested by websites. They gathered websites from the European Union and American IP addresses and compared the policies before and after GDPR came into effect [3]. Merzdocnik et al. also leverage crawled data to categorize popular trackers on websites [22]. Others adopted a hybrid method. Besides web crawlers, the researchers inspected websites that systems can not get information on manually. Degeling et al. observed an increase in privacy policy statement and cookie consent notices on websites [5]. Meanwhile, Momen et al. approached the policy changes differently, and focused more on contrasting the information required mobile applications before and after GDPR took effect [23].

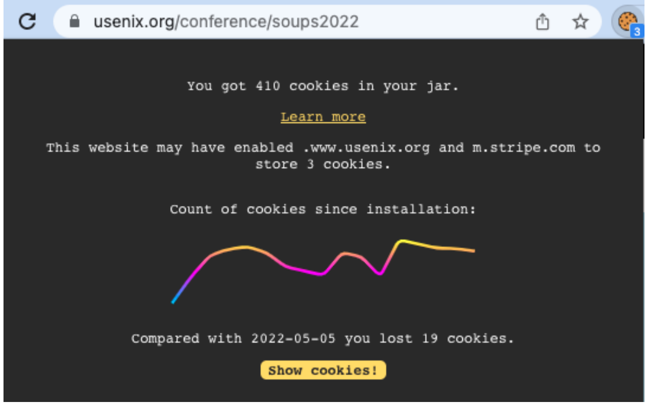
3 System: *Crumbl*

Existing cookie or tracker related plugins do not adequately capture how cookies are being injected into browsers, and mostly function as anti-trackers or automatic cookie clearers. Though users can view cookies related to a tab via inspector mode on Chromium-based browsers, still it is a developer feature that many are not aware of. Therefore, we propose a TET solution to bridge this gap—*Crumbl*.

3.1 Design

When Crumbl is pinned on the browser top bar, a user can see a number on the icon indicating the number of potential cookies being used or placed by the current tab. Overview page is the default page of Crumbl, it provides a sense of how full a browser cookie store is and what kind of cookies are being used in current tab, as shown in figure 2. It pops up whenever the user clicks on the extension icon on the browser top bar. When it is not clicked, the user can see the total number cookies being used by the current tab via the number on the top right corner of the icon. Moreover, Crumbl keeps track of daily changes in the number of cookies since its installation, allowing users to be more mindful of their Internet footprint.

Top sites page is a visualizer of browser cookie storage as shown in figure 4. On the top sites page, Crumbl shows the top domains that store the most number of cookies on a browser and their counts. We hope users can become well



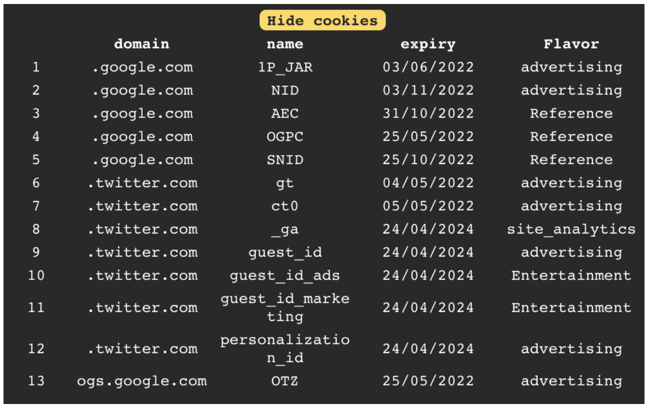
You got 410 cookies in your jar.

This website may have enabled .www.usenix.org and m.stripe.com to store 3 cookies.

Count of cookies since installation:

Compared with 2022-05-05 you lost 19 cookies.

Figure 2: Overview page of Crumbl. The layover number on the icon shows the amount of cookies on the current tab. The plugin page shows (from top to bottom): (1) total number of cookies on browser, (2) link to view analytics of stored cookies, (3) growth chart of cookies since plugin installation, (4) a breakdown of the types of cookies.



	domain	name	expiry	Flavor
1	.google.com	1P_JAR	03/06/2022	advertising
2	.google.com	NID	03/11/2022	advertising
3	.google.com	AEC	31/10/2022	Reference
4	.google.com	OGPC	25/05/2022	Reference
5	.google.com	SNID	25/10/2022	Reference
6	.twitter.com	gt	04/05/2022	advertising
7	.twitter.com	ct0	05/05/2022	advertising
8	.twitter.com	_ga	24/04/2024	site_analytics
9	.twitter.com	guest_id	24/04/2024	advertising
10	.twitter.com	guest_id_ads	24/04/2024	Entertainment
11	.twitter.com	guest_id_marketing	24/04/2024	Entertainment
12	.twitter.com	personalization_id	24/04/2024	advertising
13	ogs.google.com	OTZ	25/05/2022	advertising

Figure 3: Expanded view of cookie purposes. It shows the name, domain, expiry date, and "flavor" of each cookie when visiting Twitter.com.

acknowledged with common tracking domains that they encounter with this page. As a call-to-action, we recommend users install a general-purpose blocker to enhance privacy and suggest two non-profit blockers, namely Privacy Badger [8] and uBlock Origin [11].

Expiration page is a summary of the current expiration status of cookies living on a browser. In the first part, it counts the number of cookies per expiration window, in order to give users a sense of the expiration nature of their cookies. They might find it surprising to see that most cookies has no expiration date. Therefore, the second part highlights 10 of the non-expiring cookies belonging to a browser for users to investigate further.

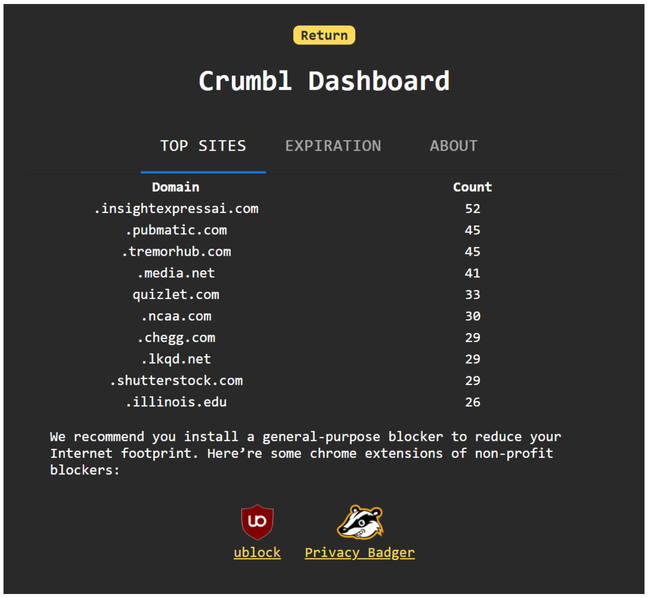
Besides the exploration page and analytics page, we also have an about page that explains the meanings of the number shown on the *Crumbl* icon. We also linked a brief introduction on HTTP cookies as supplement materials (figure 6).

3.2 Implementation

Architecture of *Crumbl* consists of a Chrome extension and a cookie categorization server. The extension reads web requests and browser cookie storage to learn about cookies on every tab, and queries the cookie categorization server when analyzing "flavors" of cookies for displaying cookie categorizations when it shows the popup. It also records daily cookie growth since its installation. See figure 1.

Cookie categorization server consists of an API endpoint that is exposed to the *Crumbl* extension and a database cookie categorization data. We use the Open Cookie Database dataset [18] that contains 709 entries, and another tracker categorization data from WhoTracksMe [14] that has 3195 entries. The latter is an open-source project by Ghostery GmbH dedicated to capturing the landscape of tracking cookies on the Internet. A cookie is composed of a name, a value, and a domain that the cookie belongs to. To ensure privacy, our extension only sends cookie name and domain to the server that does not log any web requests. Upon receipt of a request, the server returns the relevant category by looking up the database. As the data from [18] and [14] is not exhaustive, if no cookie record is found, the server returns either the category of the associated domain using data from [14] or *unknown*. The domain categorization data has 2289 entries. Note that category of a domain does not necessarily equate to that of a cookie, but we hope to offer users at least an idea of where their cookies originate from. The challenge is that cookies often have obfuscated name which are difficult to conclude their purposes without manually tracing source code.

The Extension is implemented using React.js and Chrome API for extension [12]. Once our extension is enabled, it listens to all completed web requests for each tab to capture all tab-related domains, then queries the browser cookie storage to determine how many cookies a website is using. It is important to understand that a cookie is only associated with a domain but not a tab, so it is possible that the number of cookies used by a website on each load is over-counted. As for recording the number of cookies since installation, *Crumbl* updates the total number of cookies of a day in browser every time a *cookies.onChanged* event is fired, so that the growth since installation graph in figure 2 can display the changes of the number of cookies in each day. Lastly, the reason we did not count cookies by reading the *Set-Cookie* header from web request is because it is stripped by browser for security reasons. Therefore, we count cookies based on domains called by the current tab to best estimate the count.



Return

Crumbl Dashboard

TOP SITES EXPIRATION ABOUT

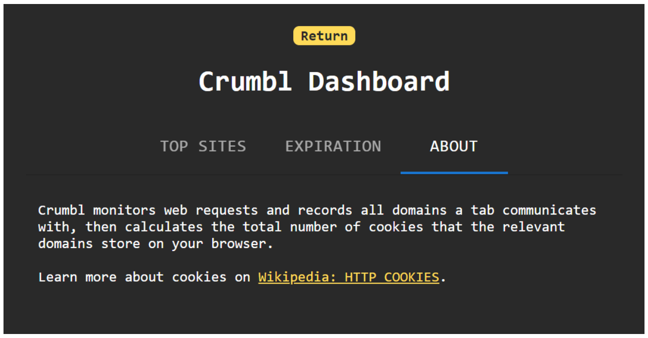
Domain	Count
.insightexpressai.com	52
.pubmatic.com	45
.tremorhub.com	45
.media.net	41
quizlet.com	33
.ncaa.com	30
.chegg.com	29
.lkqd.net	29
.shutterstock.com	29
.illinois.edu	26

We recommend you install a general-purpose blocker to reduce your Internet footprint. Here're some chrome extensions of non-profit blockers:

ublock Privacy Badger

Figure 4: Top sites page. It shows the top domains that stores the most amount of cookies in a browser and summary of the categories of the 50 most recent cookies.

The source code is hosted on GitHub[4].



Return

Crumbl Dashboard

TOP SITES EXPIRATION ABOUT

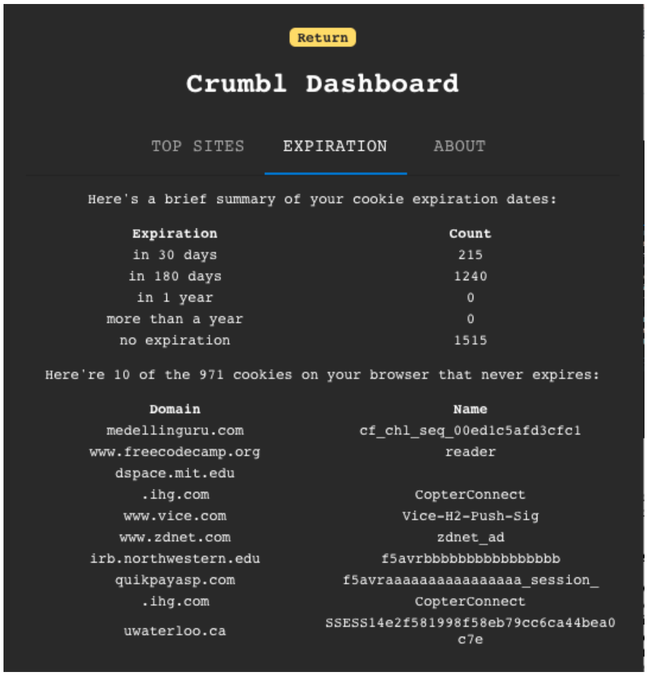
Crumbl monitors web requests and records all domains a tab communicates with, then calculates the total number of cookies that the relevant domains store on your browser.

Learn more about cookies on Wikipedia: HTTP COOKIES.

Figure 6: About page in analytics page. It explains how *Crumbl* counts the number of cookies used per tab and provides a link to learn more about HTTP cookies.

4 Discussion and future work

Past research has found that existing TETs could create false sense of protection as users did not understand the tools sufficiently [26]. We designed *Crumbl* as the first step towards nurturing privacy-focused users by providing a straightforward interface that encourages users to learn more about the cookie collection process.



Return

Crumbl Dashboard

TOP SITES EXPIRATION ABOUT

Here's a brief summary of your cookie expiration dates:

Expiration	Count
in 30 days	215
in 180 days	1240
in 1 year	0
more than a year	0
no expiration	1515

Here're 10 of the 971 cookies on your browser that never expires:

Domain	Name
medellinguru.com	cf_chl_seq_00ed1c5afd3cfc1
www.freecodecamp.org	reader
dspace.mit.edu	
.ihg.com	CopterConnect
www.vice.com	Vice-H2-Push-Sig
www.zdnet.com	zdnet_ad
irb.northwestern.edu	f5avrbbbbbbbbbbbbbbbb
quikpayasp.com	f5avraaaaaaaaaaaaaaaa_session_
.ihg.com	CopterConnect
uwaterloo.ca	SSESS14e2f581998f58eb79cc6ca44bea0c7e

Figure 5: Expiration page. It shows the expiration status of current cookies categorized by expiration window to show "stickiness" of cookies.

We conducted a pilot study with 6 participants as a preliminary step to investigate the potential impact of *Crumbl*. Our pilot study suggested that *Crumbl* may be especially helpful for users with less privacy-related knowledge. In the future, we plan to obtain IRB approval to conduct a user study to evaluate how participants with varying-levels of privacy and security knowledge perceive *Crumbl*. We plan to enhance the usability if *Crumbl* and aim to release the *Crumbl* on browser extension platform after the final revision.

5 Conclusion

In this paper, we developed *Crumbl*, a Chromium browser extension to help increase user awareness about cookie collection. With an easy-to-use interface and information presented in accessible ways to lay users, we hope that Crumbl can successfully raise awareness about data collection among users who are curious but not savvy about online privacy.

[4] https://github.com/sharonhsl/Crumbl

References

[1] Julio Angulo, Simone Fischer-Hübner, Tobias Pulls, and Erik Wästlund. Usable transparency with the data track: a tool for visualizing data disclosures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1803–1808, 2015.

[2] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. Gdpr: when the right to access personal data becomes a threat. In *2020 IEEE International Conference on Web Services (ICWS)*, pages 75–83. IEEE, 2020.

[3] Adrian Dabrowski, Georg Merzdovnik, Johanna Ullrich, Gerald Sendera, and Edgar Weippl. Measuring cookies and web privacy in a post-gdpr world. In *International Conference on Passive and Active Network Measurement*, pages 258–270. Springer, 2019.

[4] Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. The right to data portability in the gdpr: Towards user-centric interoperability of digital services. *Computer law & security review*, 34(2):193–203, 2018.

[5] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy. *arXiv preprint arXiv:1808.05096*, 2018.

[6] Ana Ferreira and Gabriele Lenzini. Can transparency enhancing tools support patient's accessing electronic health records? In *New Contributions in Information Systems and Technologies*, pages 1121–1132. Springer, 2015.

[7] Simone Fischer-Hübner, Julio Angulo, Farzaneh Karegar, and Tobias Pulls. Transparency, privacy and trust–technology for tracking and controlling my data disclosures: Does this work? In *IFIP International Conference on Trust Management*, pages 3–14. Springer, 2016.

[8] Electronic Frontier Foundation. Privacy badger, 2021.

[9] Elias Grünewald and Frank Pallas. Tilt: A gdpr-aligned transparency information language and toolkit for practical privacy engineering. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 636–646, 2021.

[10] Hans Hedbom. A survey on transparency tools for enhancing privacy. In *IFIP Summer School on the Future of Identity in the Information Society*, pages 67–82. Springer, 2008.

[11] Raymond Hill. ublock. https://github.com/gorhill/uBlock, 2022.

[12] Google Inc. Api reference, 2022.

[13] Milena Janic, Jan Pieter Wijbenga, and Thijs Veugen. Transparency enhancing tools (tets): an overview. In *2013 Third Workshop on Socio-Technical Aspects in Security and Trust*, pages 18–25. IEEE, 2013.

[14] Arjaldo Karaj, Sam Macbeth, Rémi Berson, and Josep M Pujol. Whotracks. me: Shedding light on the opaque world of online tracking. *arXiv preprint arXiv:1804.08959*, 2018.

[15] Daniel Kladnik. I don't care about cookie, 2022.

[16] Georgios Kontaxis and Monica Chew. Tracking protection in firefox for privacy and performance. *arXiv preprint arXiv:1506.04104*, 2015.

[17] Bert-Jaap Koops. The trouble with european data protection law. *International data privacy law*, 4(4):250–261, 2014.

[18] Jack Kwakman. Open cookie database. https://github.com/jkwakman/Open-Cookie-Database, 2022.

[19] Connor Luckett, Andrew Crotty, Alex Galakatos, and Ugur Cetintemel. Odlaw: A tool for retroactive gdpr compliance. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 2709–2712. IEEE, 2021.

[20] Delfina Malandrino, Vittorio Scarano, and Raffaele Spinelli. How increased awareness can impact attitudes and behaviors toward online privacy protection. In *2013 International Conference on Social Computing*, pages 57–62. IEEE, 2013.

[21] Aditya Marella, Chao Pan, Ziwei Hu, Florian Schaub, Blase Ur, and Lorrie Faith Cranor. Assessing privacy awareness from browser plugins. In *Proceedings of the Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[22] Georg Merzdovnik, Markus Huber, Damjan Buhov, Nick Nikiforakis, Sebastian Neuner, Martin Schmiedecker, and Edgar Weippl. Block me if you can: A large-scale study of tracker-blocking tools. In *2017 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 319–333. IEEE, 2017.

[23] Nurul Momen, Majid Hatamian, and Lothar Fritsch. Did app privacy improve after the gdpr? *IEEE Security & Privacy*, 17(6):10–20, 2019.

[24] nasir. Cookie popup blocker, 2022.

[25] Philip Raschke, Axel Küpper, Olha Drozd, and Sabrina Kirrane. Designing a gdpr-compliant and usable privacy dashboard. In *IFIP international summer school on privacy and identity management*, pages 221–236. Springer, 2017.

[26] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. Watching them watching me: Browser extensions' impact on user privacy awareness and concern. In *NDSS workshop on usable security*, pages 1–10, 2016.

[27] Marija Schufrin, Steven Lamarr Reynolds, Arjan Kuijper, and Jörn Kohlhammer. A visualization interface to improve the transparency of collected personal data on the internet. In *2020 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pages 1–10. IEEE, 2020.

[28] Andrew Selbst and Julia Powles. "meaningful information" and the right to explanation. In *Conference on Fairness, Accountability and Transparency*, pages 48–48. PMLR, 2018.

[29] Kim-Phuong L Vu, Vanessa Chambers, Beth Creekmur, Dongbin Cho, and Robert W Proctor. Influence of the privacy bird® user agent on user trust of different web sites. *Computers in industry*, 61(4):311–317, 2010.