Analysis of Third-Party Cookies on Top Japanese Websites

JIAYING FENG[1], YUKI YADA[1], TSUNEO MATSUMOTO[1,2],
NAO FUKUSHIMA[3], FUKUYO KIDO[4,1],
HAYATO YAMANA[1]

Abstract: Third-party cookies, set by a website other than the website we are accessing, may become an invasion of privacy. From the enforcement of the amended act on the "Act on the Protection of Personal Information" in Japan in April 2022, if the third-party cookies are used with personal information, they shall take users' consent of the collected data usage. Besides, in the EU, GDPR has already been initiated to standardize the usage of cookies. In this paper, we gathered the third-party cookies from top-accessed Japanese websites to analyze. Then, the characteristics of the third-party cookies are reported.

Keywords: Third-party Cookies, Dark Patterns, Online Privacy

1. Introduction

Cookies, stored in a browser, are text files containing such as a user id and session id, which are typically used for continuous access to a specific website without additional logins after the first login. However, cookies can be used for malicious purposes once the cookies are shared among third parties. In a typical case, users' web access information is gathered by using cookies to tune the personalized advertisements. Thus, if the cookies are shared among multiple organizations and used with the user's personal information, they should be protected. Therefore, regulations have been adopted to protect users' privacy [5][6][8].

From the enforcement of the amended act on the "Act on the Protection of Personal Information" in Japan in April 2022, if the third-party cookies are used with personal information, they shall take users' consent of the collected data usage.

In this report, we investigate the usage of third-party cookies used in the top-accessed websites in Japan to investigate the current situation. Our contributions are as follows.

(1) **This report is the first attempt to analyze how third-party cookies are used in Japanese websites.**
(2) **We analyzed cookies' lifespans over different categories of websites.**
(3) **Dark patterns related to the cookies are analyzed.**

The rest of the paper introduces the background of cookie usage and related work in Section 2. Next, third-party cookie analysis is shown in Section 3. Then in Section 4, we analyze dark patterns for cookie consent. Finally, section 5 concludes this paper.

2. Background and related work

This section reviews the classifications of cookies, regulations adopted towards uses of cookies and definitions, and taxonomies of dark patterns.

2.1 First-party cookies and third-party cookies

Cookies are text files stored in browsers with small pieces of data. When a user visits a website, the webserver requests the user's browser to save the files onto the users' devices. Those files are called *cookies*, in which session id, user id, checked items, and other user-related information are stored to enable continuous access to the same website or track the user's browsing history. The information in cookie files depends on what the webserver places inside the cookie files.

Cookies have different taxonomies according to different properties. Cookies are classified into two categories by their provenance; first-party cookies and third-party cookies. First-party cookies are placed by the websites visited, and third-party cookies are placed by the other websites the user does not visit. For example, assume that a user visits a website called abc.com. The cookies placed by abc.com are called first-party cookies. On the other hand, the cookies placed by the other websites are called third-party cookies. For example, the third-party cookie called "IDE" is placed by doubleclick.net, serving targeted advertisements relevant to users across the websites [7].

Cookies can also be classified by their lifespan, i.e., how long the cookies stay in the user's device. For example, session cookies will stay while the browser is open, i.e., the cookie files will be discarded after closing the browser. This is because session cookies are usually used to keep or track the user's session while the user walks around the same website. On the other side, persistent cookies remain in operation even though the browser is closed.

2.2 Regulation towards uses of cookies

There have been a few regulations adopted to restrict the use of cookies. The EU adopted the General Data Protection Regulation (GDPR) to protect the online privacy of its residents on May 25,

1 Waseda University, Shinjuku-ku, Tokyo 169-8555, Japan
2 National Consumer Affairs Center of Japan, Sagamihara, Kanagawa 252-0229, Japan
3 LINE Corporation, Shinjuku, Tokyo 160-0004, Japan
4 National Institute of Informatics, Chiyoda-ku, Tokyo 101-8430, Japan

Figure 1: An example of cookie banners (http:/gdpr.eu/)

2018 [2]. Consent for cookies that can identify a user uniquely is explicitly mentioned in Recital [6].

The California Consumer Privacy Act (CCPA) also introduced regulations to inform and provide consumers of their right to opt-out of the sale of their personal information in Civil Code section 1798.120 [8]. Consent for cookies that can identify a user uniquely is explicitly mentioned in Civil Code section 1798.140 [11].

To comply with the above regulations, websites employ various methods to notify users of the use of cookies and provide opt-out options. One standard method is cookie banner, as shown in Figure 1; it usually offers two options: accept or customize the cookies' use.

In Japan, from the enforcement of the amended act on the "Act on the Protection of Personal Information" in April 2022, if the third-party cookies are used with personal information, they shall take users' consent of the collected data usage.

2.3 Dark Patterns

Dark patterns are defined as slick user interfaces, leading users to intentionally do specific things, such as buying something or subscribing to some services. The term "dark patterns" was first introduced by Harry Brignull in 2010 [4]. In 2018, Colin [3] classified dark patterns into five categories: nagging, obstruction, sneaking, interface interference, and forced action (s), as described in parentheses below.

Nagging:
Redirection of expected functionality that persists beyond one or more interactions.
Obstruction:
Making a process more complicated than it needs to be, with the intent of dissuading specific action (s).
Sneaking:
Attempting to hide, disguise, or delay the divulging of relevant information to the user.
Interface Interference:
Manipulation of the user interface that privileges specific actions over others.
Forced Action:
Requiring the user to perform a certain action to access (or continue to access) certain functionality.

Even though some websites have cookie banners that enable users to opt-out of their personal information, it may be dark patterns depending on how the cookie banners are configured. Figure 2 is a screenshot of a website's cookie banner, which falls under interface interference because the button "Accept All Cookies" is more conspicuous than "Manage Choices." An analysis of 300 consent notices from online news outlets shows that 297 of 300 websites use dark patterns when eliciting consent from users [9].

3. Cookie analysis for 400 Japanese sites

To understand current cookie usage, we analyzed the cookies in top-accessed Japanese sites in four categories. We analyzed the following four points:

1. The number of websites adopting third-party cookies
2. The average number of first-party cookies and third-party cookies on a website
3. Duration of third-party cookies in each website category
4. Commonly used third-party cookies

3.1 Datasets

The datasets are prepared based on the traffic data provided by Similar Web (https://www.similarweb.com/), a company that provides web analytics services. We constructed the following four datasets based on the statistics as of September 2021:

1) top-accessed 100 websites in Japan (all categories)
e.g.) www.google.com, www.yahoo.co.jp, www.youtube.com
2) top-accessed 100 websites in business and consumer
e.g.) www.mynavi.jp, impress.co.jp, suumo.jp
3) top-accessed 100 websites in marketing
e.g.) nend.net,research-panel.jp,moshimo.co.jp
4) top-accessed 100 websites in publishing
e.g.) tkj.jp,machicomi.jp, ansin-anzen.jp

3.2 Results – usage of third-party cookies

Figure 3 shows how many websites use third-party cookies as of January 2022. In our experiment, nearly 77% of websites use third-party cookies. Especially, 84 websites out of the top-accessed 100 websites, i.e., 84%, use third-party cookies. On the other hand, the websites in the publishing category tend not to use third-party cookies; however, 65% of them still use third-party cookies.

Among the websites with third-party cookies, few websites have cookie banners to inform and allow users to opt-out. In order to

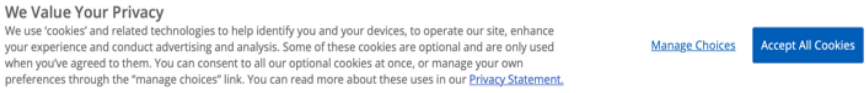

Figure 1: A dark pattern of cookie banner (https://home.kpmg/xx/en/home.html)

figure out the use of third-party cookies, we then focus on websites that use third-party cookies. Table 1 shows the average number of first-party cookies and third-party cookies in each category. It is notable that in the "All" categories, i.e., in the top-accessed 100 websites, there are three times as many third-party cookies as first-party cookies. For comparison, websites from "Marketing" have similar numbers of first-party and third-party cookies. Websites related to users closer have more third-party cookies compared to others.

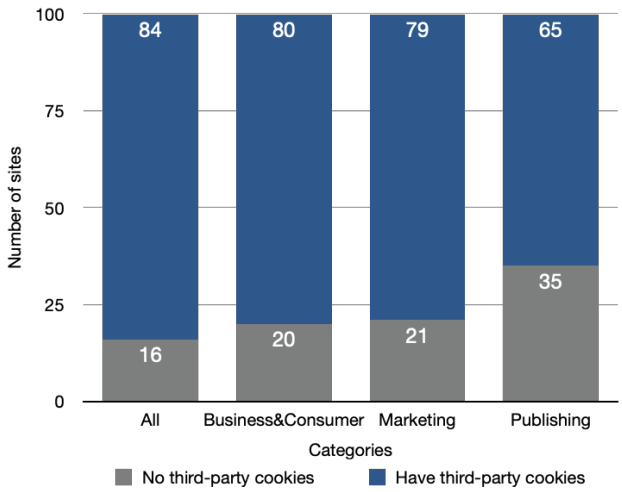

Figure 2: Use of third-party cookies in each top 100 websites as of January 2022

Table 1: The average number of cookies per website

Category	All (top 100)	Business & Consumer	Marketing	Publishing
first-party cookies	14.25	14.83	12.05	8.94
third-party cookies	54.63	34.45	14.83	15.97
Total	68.88	49.28	26.88	24.91

3.3 Result – lifespan of cookies

Apart from the number of third-party cookies, lifespan is also essential. As we mentioned before, different cookies have different lifespans. According to the EU's "ePrivacy Directive," cookies should not last longer than 12 months; however, they could remain on users' devices much longer in practice if an action to manage them is not taken [5]. A long lifespan also means cookies can keep more personal information and thus be considered a severe invasion of privacy.

Our experiment divides lifespan time into five ranges: 1) in a session (session cookies), 2) less than one day, 3) one day to one month, 4) one month to one year, and 5) longer than one year. Figure 4 shows the number of cookies on average in different lifespans.

The websites in the "All" and "Business & Consumer" categories have more cookies whose lifespan is longer than one year, as shown in Figure 4 and Table 2. Even though keeping cookies longer than one year is not illegal in Japan right now, we should pay attention to these long-lifespan cookies.

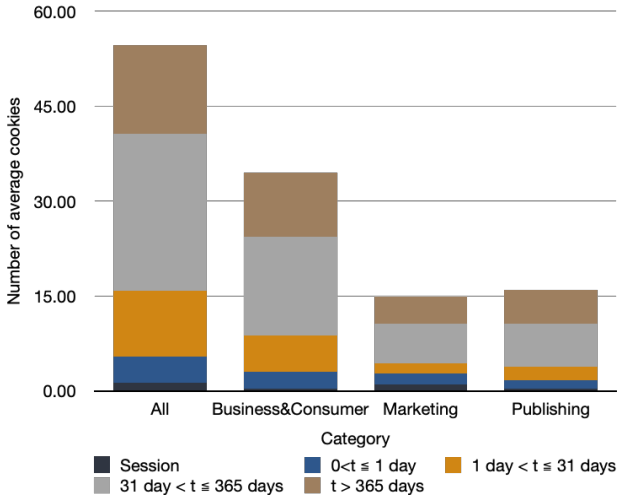

Figure 3: Lifespan of cookies in each category

Table 2: Percentage of cookies' lifespan in each category

Category / lifespan	All (top 100)	Business& Consumer	Marketing	Publishing
in a session	2.44%	0.99%	6.75%	2.38%
0<t ≤ 1 day	7.45%	7.72%	11.47%	8.45%
1 day < t ≤ 31 days	18.99%	16.52%	10.86%	13.09%
31 day < t ≤ 365 days	45.59%	45.46%	42.78%	42.20%
t > 365 days	25.54%	29.32%	28.14%	33.88%

4. Dark patterns for cookie consent

Attention should be paid to the consent of cookies since the data in cookie files may be considered personal information. However, websites usually do not actively inform users about cookies and seek their consent. A few websites in our experiment ask for the consent of cookies. Even though a few websites ask for the consent of cookies, they use dark patterns to trick users. Note that whether this trick is on purpose or not is unknown.

There have been studies conducted to investigate the situation outside Japan. For example, in 2019, Utz et al. [13] conducted a study of consent notices on e-commerce websites. They show that small UI design substantially impacts how users interact with cookie consents. Nouwens et al. [12] also researched the impact of different designs of consent notices. They focused on the presence of three features: whether the consent is explicit; whether the acceptance button is the same as the rejection button; whether pre-ticked boxes are present. They found that only 80 out of the 680 notices satisfy all three conditions.

In this section, our experiment focused on Japanese top-accessed websites. First, if a website uses third-party cookies, we will check whether a cookie banner exists or other notifications related to cookies. Second, we also analyzed the relationship between dark patterns and these cookie banners and notifications. Then, we investigated dark patterns to classify. Note that the use of cookies and dark patterns depend on which device, i.e., from smartphones or PCs, is used to access. Therefore, cookies and dark patterns when accessing PCs are investigated in this experiment.

4.1 Datasets

The datasets are prepared based on the traffic data provided by Similar Web (https://www.similarweb.com/), a company that provides web analytics services. We constructed the following two datasets based on the statistics as of September 2021:

1) top-accessed 256 websites in Japan (all categories)
   e.g.) www.youtube.com, www.amazon.co.jp, www.google.com
2) top-accessed 721 websites in business and consumer
   e.g.) www.chatwork.com, www.cybozu.com, www.askul.co.jp

Since we focused on the PC version's web pages, we chose the websites that meet the following three conditions: 1) desktop share is more than 50%, i.e., frequently accessed from PCs compared to mobile devices (smartphones); 2) the website's traffic should be high, i.e., the volume of users visiting a websites is high; 3) the websites use third-party cookies. After applying the above conditions, the total number of websites becomes 821.

4.2 Experiment

Dark patterns may manifest in several locations inside the websites, relying heavily upon interface manipulation, such as changing the hierarchy of interface elements or prioritizing specific options over others using different colors [9]. Because of the above characteristics of dark patterns, it is tough to automatically capture web pages' dark patterns. Therefore, we accessed the 821 websites manually and recorded items related to cookies' dark patterns with the help of invited 26 students.

4.3 Results – analysis of dark patterns

Only 61 out of the 821 websites have cookie banners. Even though these websites have cookie banners, dark patterns were used to trick users as follows:

Sneaking:

Even if a cookie banner exists, 32 out of the 61 websites do not have any decline or customize button. This can be considered sneaking because websites attempt to hide opt-out options.

Interface Interference:

Interface interference is commonly used in cookie banners. In this experiment, we focus on the two features to analyze the cookie banners: 1) the location of the cookie banner; 2) whether the acceptance button is the same as the decline/customize button.

Only 14 out of the 61 websites put the cookie banners in the upper 1/3 of web pages. Previous research showed higher interaction rates for desktop screens for the notices displayed at the bottom and left sides of the screen [13]. Therefore, the users are less likely to pay attention to the cookie banners placed at the lower 2/3 of web pages, i.e., the remaining 47 websites put the cookie banners at the bottom 2/3 of the web pages.

The difference also appears in the design of the acceptance and decline and customize button. The 32 websites have cookie banners, and 18 out of 32 have differences in accepting buttons and the other buttons. The differences include 1) color, 2) font size, and 3) display format. Websites tend to use conspicuous colors and larger font in the acceptance button. Moreover, eight websites use links instead of customizing/declining buttons.

Obstruction:

We also checked the explanation of cookies if there is an explanation page. 13 out of 29 websites use more than 1000 words to explain the cookies, which is hard to understand by common users. Therefore, this can be considered "obstruction" since handling cookies is difficult to understand.

5. Summary and discussion

In this paper, we presented two analyses of the use of third-party cookies on Japanese websites. The first analysis focused on top-accessed Japanese sites in four different categories. The analysis shows that most websites use third-party cookies but do not inform users of their usage. We also checked the lifespan of third-party cookies and found that about 21% of third-party cookies last longer than one year. The second analysis focused on the relationship between cookies and dark patterns. Our experiment shows that even among websites that inform users of the uses of cookies, they use dark patterns to lead users to give out their personal information.

Unlike previous research [1][2][9], our study focuses on Japanese websites which have not been investigated yet. We believe that with the implementation of the "Act on the Protection of Personal Information" in April 2022, the websites will make changes to the use of cookies. The use of third-party cookies dropped by 10% after GDPR [10]. For future work, checking the use of cookies regularly to monitor the effects of this regulation is necessary.

Acknowledgment

This study was supported by LINE Corporation.

Reference

[1] California of Department of Justice. "Civil Code section 1798.120". https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.120.&nodeTreePath=8.4.47&lawCode=CIV, (accessed on 2022-01-26).
[2] California of Department of Justice. "Civil Code section 1798.140". https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?sectionNum=1798.140.&nodeTreePath=8.4.47&lawCode=CIV, (accessed on 2022-01-29).
[3] "Dark Patterns". https://www.darkpatterns.org/, (accessed on

2021-05-30).
[4] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. The Dark (Patterns) Side of UX Design. Proc. of the 2018 CHI Conf. on Human Factors in Computing Systems. Paper 534, pp.1–14, 2018.
[5] GDPR.EU. “Cookies, the GDPR and the ePrivacy Directive.”. https://gdpr.eu/cookies/, (accessed on 2022-01-29)
[6] GDPR Recital 30. “Online identifiers for profiling and identification”. https: //gdpr-info.eu/recitals/no-30/, (accessed on 2022-01-28).
[7] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. Circumvention by design - dark patterns in cookie consent for online news outlets. Proc. of the 11th Nordic Conf. on Human-Computer Interaction: Shaping Experiences, Article 19, pp.1–12, 2020.
[8] Xuehui Hu and Nishanth Sastry. Characterising Third Party Cookie Usage in the EU after GDPR. Proc. of the 10th ACM Conf. on Web Science, pp. 137–141, 2019.
[9] “IDE Summary”. https://cookiedatabase.org/cookie/google-doubleclick/ide/, (accessed on 2022-01-26).
[10] Rasmus Kleis Nielsen, Nic Newman, Richard Fletcher, and Antonis Kalogeropoulos. “Reuters Institute Digital News Report 2019”. Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2019-06/DNR_2019_FINAL_1.pdf, (accessed on 2022-01-29).
[11] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. Proc. of ACM Hum.-Comput. Interact. CSCW, Article 81, pp.1-32, 2019.
[12] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence. Proc. of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-13, 2020.
[13] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed Consent: Studying GDPR Consent Notices in the Field. Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security, pp. 973–990, 2019.