Recommended Paper

Risk Analysis of Cookie Sharing by Link Decoration and CNAME Cloaking

Yuta Takata[1,a)] Daiki Ito[1] Hiroshi Kumagai[1] Masaki Kamizono[1]

Received: December 24, 2020, Accepted: July 7, 2021

Abstract: Privacy issues due to web tracking are a continuously evolving problem. One tracking method utilizes third-party cookies. This method analyzes user behavior and interest on the Web by sharing cookies with third-party vendors such as analytics and advertising brokers. Several regulations on third-party cookies have been considered by countries and browser vendors to address privacy issues due to such excessive web tracking. However, third-party vendors continue to track users with new technologies such as link decoration that embeds cookies in URLs and CNAME cloaking which tricks browsers into treating third-party cookies as a first-party. In this paper, we analyze cookie sharing by link decoration and CNAME cloaking and reveal their privacy issues. In addition, we reveal new security risks emerging from these technologies.

Keywords: Tracking, Cookie, Link Decoration, CNAME Cloaking

1. Introduction

Web tracking also simply called "tracking" that can track user behavior on the Internet is well known for example for being used in measuring website access and calculating conversion rates for web advertising. Measuring website performance and optimizing web content through tracking is an important element in improving marketing strategies [11]. Cookies are one of the major tracking methods. A cookie is data with small size, stored in a browser, and used for stateful web applications. Third-party vendors, such as analytics and advertising brokers, use cookies for tracking. However, excessive tracking using third-party cookies has become an important privacy issue [17], [20].

Various regulations and limitations on tracking using third-party cookies are being discussed. For example, the General Data Protection Regulation (GDPR) and the ePrivacy Regulation require clear notice and explanation to users about the use and sharing of cookies [21]. Browser vendors such as Apple and Google also limit the use of third-party cookies [5], [27].

On the other hand, third-party vendors are deploying new technologies such as link decoration and CNAME cloaking, to share cookies with external websites without using third-party cookies in order to evade the above limitations [13], [28]. In this paper, we analyze these technologies for sharing cookies with third-party vendors and investigate the risks these technologies create.

In summary, we make the following contributions.

- We identify websites that share cookies using only link decoration or only CNAME cloaking.
- We show that link decorations and CNAME cloaking share 147 (5.59%) and 149 (35.28%) `Session` cookies, respectively.
- Within the sampled websites' `SameSite` cookies, we find that 66 (1.86%) cookies are shared by bypassing the cross-site request limitations.
- We find that *Strictly Necessary* cookies on first-party websites are shared by CNAME cloaking.

The rest of this paper is structured as follows. In Section 2, we review related work. We provide background on tracking and cookie sharing and explain risks they cause in Section 3. We explain a method for detecting and analyzing cookie sharing in Section 4 and we present the results of our analysis in Section 5. In Section 6, we discuss newly found risks of cookie sharing. Finally, we conclude this paper in Section 7.

2. Related Work

2.1 Cookie Analysis

Many researchers have conducted measurement studies and risk analysis regarding cookies. There are several studies on privacy risk analysis based on cookie features [12], tracking using cookies [17], and cookie measurement on the Internet [20]. Other researchers have studied methods for keeping cookies by evercookies [11] and cookie synchronization [18]. However, all of the above studies focused on third-party cookies. As the most relevant study, Dao et al. investigated CNAME cloaking usage and evaluated the effect of privacy protection techniques by filters, browsers, and extensions [15]. Although they reported the characteristics and pervasiveness of CNAME cloaking, this study was unclear about the actual security and privacy risks there are because they did not analyze cookies that were themselves shared by CNAME cloaking. On the other hand, we analyze technolo-

[1] Deloitte Tohmatsu Cyber LLC, Chiyoda, Tokyo 100–0005, Japan
[a)] yuta.takata@tohmatsu.co.jp

The preliminary version of this paper was published at IPSJ SIG on CSEC in July 2020. The paper was recommended to be submitted to Journal of Information Processing (JIP) by the chief examiner of SIGCSEC.

gies for sharing cookies including not only CNAME cloaking but also link decoration, and cookies shared by them for identifying security and privacy risks.

2.2 Cookie Protection

Cookie protection features only consist of `HttpOnly`, `Secure`, and `SameSite` attributes [6]. An `HttpOnly` attribute can limit the use of cookies by JavaScript (`document.cookie`) and prevent cookie leakage through cross-site scripting (XSS) attacks. A `Secure` attribute can limit the use of cookies to HTTPS connections and prevent cookie leakage on the network. A `SameSite` attribute can prevent cross-site request forgery (CSRF) attacks by setting either the "`Strict`" or "`Lax`" value. Although other browser extensions can protect cookies, Franken et al. have reported that some of these security features can be bypassed [16]. We show that link decoration and CNAME cloaking also bypass the `SameSite` attribute in Section 5.4.

3. Background

3.1 Cookie

A Cookie is text data stored in a browser and set by an HTTP response header "`Set-Cookie: name=value`" from a web server or an HTTP request header "`Cookie: name=value`" from a client, i.e., a browser [6]. There are two types of cookies depending on the expiration period: `Persistent` and `Session` cookies. A `Persistent` cookie can keep data in a browser until a date specified by an `Expires` attribute, or with a period of time specified by a `Max-Age` attribute. On the other hand, a `Session` cookie can keep data without these attributes, and the `Session` cookie is deleted when the session ends. Therefore, cookies are a feature used by stateful websites rather than traditional stateless websites and are primarily used for session management, personalization, and tracking.

3.2 Use of Cookie

When a browser uses cookies for identifying users or authenticating sessions, sensitive data might be set in the cookies. If these cookies can be sent and/or received without any limitations, then session hijacking and CSRF attacks become possible. Therefore, URLs that can use cookies are limited to a scope defined by the `Domain` and `Path` attributes. A `Domain` attribute defines a domain name (including the subdomains) that can use the cookie. A `Path` attribute defines a URL path that can use the cookie. Note that if domain names set in the `Domain` attribute match the domain name in the browser's address bar, we call the affected cookies first-party cookies, and the other cookies third-party cookies [1].

3.3 Tracking Using Third-party Cookie

Third-party vendors maximize the effectiveness of web advertisements by tracking user behavior on the Web and publishing advertising content based on user interests [19]. We explain the process of tracking by third-party cookies using **Fig. 1**. (1) When a browser visits website A, an HTTP request is sent requesting third-party content possessed by website A. (2) Third-party servers receive the requests, set an identifier in the response content using a "`Set-Cookie`" header, and record the access to website A. (3) The browser visits website B that contains the same third-party content. (4) This time, the browser automatically sends the aforementioned identifier through a "`Cookie`" header to the third-party server. The third-party server can thereafter track visits by the browser across websites, i.e., from website A to website B. Since many first-party websites profit by selling advertising space to third-party vendors, they can gather these cross-site requests and analyze user behavior on the Web.

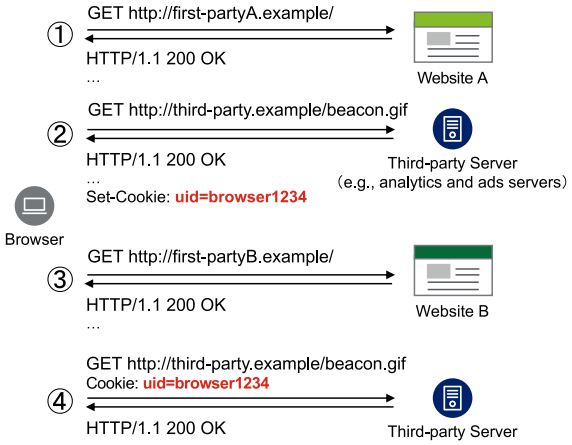

Fig. 1 Tracking with third-party cookie.

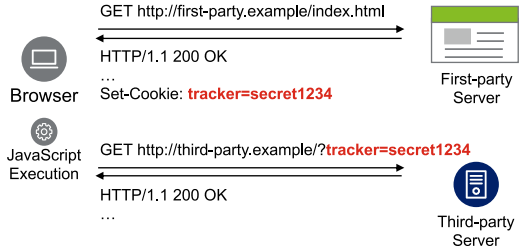

Fig. 2 First-party cookie shared by link decoration.

3.4 Trends of Privacy Regulations and Limitations

Privacy issues caused by tracking using third-party cookies are growing, and regulations and limitations have been considered by countries and browser vendors. The EU and United States have considered and published regulations such as GDPR, ePrivacy regulation, and CCPA, which state that clear explanations to users about cookies are required [21]. Browser vendors are also considering functions to limit third-party cookies themselves. For example, Apple announced that third-party cookies with tracking functions must be immediately removed and first-party cookies are limited to storage periods of up to 24 hours starting from April 2019 [28]. Google also announced that browsers will block third-party cookies without `SameSite=None` and `Secure` attributes from October 2019 [5].

3.5 Cookie Sharing

Third-party vendors try to evade the regulations and limitations described above by **link decoration** that embeds cookie data in URLs and **CNAME cloaking** which makes third-party cookies behave like a first-party.

3.5.1 Link Decoration

Link decoration is a technique to embed first-party cookies in

```
first-party.example.              IN A      192.168.0.1.
ipsj2020.first-party.example.     IN CNAME  user1.third-party.example.
```

Fig. 3 DNS resource records registered by first-party webmaster.

```
user1.third-party.example.   IN A   172.16.0.1.
user2.third-party.example.   IN A   172.16.0.1.
user3.third-party.example.   IN A   172.16.0.1.
...
```

Fig. 4 DNS resource records registered by third-party vendor.

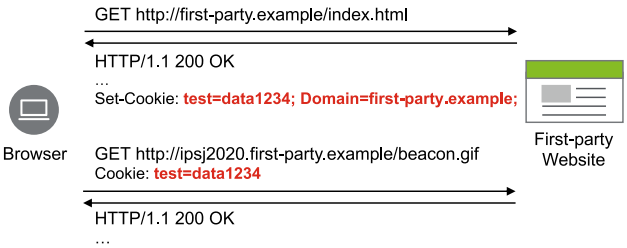

Fig. 5 First-party cookie shared by CNAME cloaking.

third-party URLs and share them [28]. We show an example of sharing first-party cookies to third-party servers using link decoration in **Fig. 2**. When a browser visits a first-party website, cookies are set in the browser by a `Set-Cookie` header in the server response. The cookie data is read and embedded in third-party URLs through JavaScript's `document.cookie` and shared with the third-party servers. In addition to the dynamic methods used by JavaScript, there is a static method for creating links embedded with cookie data in advance. Link decoration is often used for third-party content contained in first-party websites.

3.5.2 CNAME Cloaking

CNAME cloaking is a new technique to use domain names (IP addresses) set by third-party vendors as canonical names of first-party subdomains [13]. We explain the process using **Figs. 3**, **4**, and **5**. First, a first-party webmaster and a third-party vendor set DNS resource records shown in Figs. 3 and 4, respectively. The webmaster sets the domain name "`user1.third-party.example`" provided by the third-party vendor as a CNAME record of "`ipsj2020.first-party.example`". Next, the webmaster uses contents of a URL hosted on the subdomain for the first-party website. Then, as shown in **Fig. 5**, a request for the content with a first-party `Cookie` header occurs when a browser visits the first-party website. The request destination is "`user1.third-party.example`" which has the canonical name (CNAME) "`ipsj2020.first-party.example`" pointing to the third-party IP address "`172.16.0.1`". Therefore, the first-party cookie is shared with the third-party vendor. This method takes advantage of the fact that a cookie with the `Domain` attribute is automatically shared with subdomains as well.

3.6 Risks Caused by Cookie Sharing

Technologies of sharing cookies described in the previous section can bypass cookie protection features such as `Session`, `Secure`, and `SameSite` attributes. If cookies are embedded in URLs and shared by link decoration, all these features are disabled. CNAME cloaking can also share first-party cookies to third-party vendors while bypassing the `SameSite` attribute. If these protected cookies are used for identifying users or authenticating sessions, then users are exposed to security and privacy risks as described in Sections 2.2, 3.2, and 3.3. In our study, we analyze and identify such protected cookies shared by link decoration and CNAME cloaking while bypassing cookie protection features.

4. Detection and Analysis of Cookie Sharing

We detect first-party cookies shared by link decoration and CNAME cloaking, and analyze them to identify risks exposed by cookie sharing.

4.1 Detection of Cookie Sharing

Detection of Link Decoration. To detect link decoration, we analyze whether first-party cookie values are embedded in third-party URLs during website crawling. If embedded, we detect them as first-party cookies shared with third parties by link decoration.

Detection of CNAME Cloaking. To detect CNAME cloaking, we analyze DNS resource records of domain names set in the `Domain` attributes of first-party cookies. If all of the following conditions match, we detect them as first-party cookies shared with third parties by CNAME cloaking.

- A domain name set in a `Domain` attribute does not have any A records.
- A domain name set in a `Domain` attribute has CNAME record(s).
- The domain name set in the above CNAME record is not a first-party domain name.
- The IP address of the domain name set in the above CNAME record is not an IP address of the input URL.

4.2 Analysis of Cookie Sharing

First, we count the total number of collected cookies, categorize `Persistent` or `Session` cookies, analyze the `Secure`, `HttpOnly`, `SameSite` attributes, and calculate the uniqueness of cookie values. The uniqueness of cookie values indicates the possibility of uniquely identifying users on the Web. Cookie values with higher uniqueness have higher privacy and tracking risks. To calculate the uniqueness, we used `zxcvbn` [26] which can calculate password strength as well as the approach taken by Sanchez-Rola et al. [21].

Next, we investigate the purpose and use of cookies with high uniqueness shared by link decoration and CNAME cloaking. We used search results of cookie names with Cookiepedia [1] to analyze the purpose.

Finally, we analyze whether `Secure` cookies are leaked through HTTP links by link decoration. We also analyze whether `SameSite` cookies bypass cross-site request limitations by CNAME cloaking.

4.3 Source and Destination Analysis of Cookie Sharing

We analyze features of *source* that share cookies (i.e., first-party websites) and *destination* that receive these cookies (i.e., third-party vendors).

First, we analyze what kind of services are provided by third-

Table 2 Attributes and uniqueness of collected cookies (Top5K).

Domain	Total	Persistent	Session	HttpOnly	Secure	SameSite	zxcvbn_log10≥10
1st-party	19,481	14,211 (72.95%)	5,270 (27.05%)	4,012 (20.59%)	3,385 (17.38%)	2,296 (11.79%)	12,873 (66.08%)
3rd-party	66,609	61,714 (92.65%)	4,895 (7.35%)	9,587 (14.39%)	51,504 (77.32%)	50,931 (76.46%)	48,428 (72.70%)

Table 3 Attributes and uniqueness of collected cookies (Rand5K).

Domain	Total	Persistent	Session	HttpOnly	Secure	SameSite	zxcvbn_log10≥10
1st-party	13,118	10,160 (77.45%)	2,958 (22.55%)	2,689 (20.50%)	1,089 (8.30%)	1,247 (9.51%)	8,592 (65.50%)
3rd-party	28,994	25,838 (89.11%)	3,156 (10.89%)	5,553 (19.15%)	21,007 (72.45%)	20,443 (70.51%)	21,119 (72.84%)

Table 1 Results of website crawling.

Dataset	HTTP 200	HTTP error	Crawl error
Top5K	4,297	173	530
Rand5K	4,598	90	312
Total	8,895	263	842

Table 4 Values of SameSite attributes.

Domain	None	Lax	Strict
1st-party	1,002 (28.28%)	2,436 (68.76%)	105 (2.96%)
3rd-party	71,266 (99.85%)	103 (0.14%)	5 (0.01%)

party vendors that often appear as cookie sharing destinations. We use domain category data provided by Cisco Talos [2] to categorize these domain names to their service names.

Next, we analyze whether first-party websites that use link decoration and CNAME cloaking are aware of the cookie sharing itself. If they collect and use cookies and/or share and sell them on other websites, the process must be listed in their privacy policy [1]. Therefore, for our analysis, we manually investigated whether a privacy policy is provided on these source websites and whether descriptions of cookie processes are listed in said privacy policy.

4.4 Experimental Environment

To crawl websites and collect cookies, we used a Chromium browser 79.0 [10] installed on Ubuntu. Since JavaScript executions and asynchronous communications will occur after loading web content, we forced the browser to wait for network idle time while crawling. Note that if content loading did not complete within three minutes, we timed out the access. In addition, we minimized the effect of browsing histories by using the browser in Incognito mode [8].

For website crawling, we built two types of datasets: Top5K and Rand5K. The former includes top 5,000 domain names listed on AlexaTopSites [7]. The latter includes 5,000 domain names randomly extracted from AlexaTopSites. Note that we crafted URLs with "http://" and these domain names.

5. Analysis Results

5.1 Data Collection of Cookie

We crawled a total of 10,000 URLs in February 2020. As a result, 8,895 websites successfully responded with some type of content and the others were HTTP 400 s or 500 s errors, DNS, or timeout errors as shown in **Table 1**. We collected 32,599 first-party cookies and 95,603 third-party cookies by loading the content of 405,601 first-party URLs and 551,791 third-party URLs while crawling. We found that the maximum number of third-party cookies were 218 on a website in Top5K and 239 in Rand5K. In the following sections we analyze these collected cookies.

5.2 Attributes and Uniqueness of Cookie

First, we show attribute distributions of collected cookies used in Top5K and Rand5K websites in **Tables 2** and **3**, respectively. The total number of third-party cookies of Top5K was double that of Rand5K and we can infer that Top5K websites deployed more advertisement and analytics content because of their popularity. There were no significant differences between Top5K and Rand5K in the other attributes.

Next, we focus on differences between first-party and third-party rather than Top5K and Rand5K. There were more third-party cookies than first-party ones and almost all cookies were Persistent instead of Session, or namely these cookies have a set expiration date. Moreover, the Secure and SameSite attributes were more distributed in third-party cookies although there were no significant difference in the distribution rate of the HttpOnly attribute. It signifies that cookies with safer settings were used with external domain names. We can infer that this finding was affected by Google's announcement in February 2020, that Chrome version 80 or newer will block third-party cookies without SameSite=None and Secure attributes [5]. Indeed, the SameSite attribute value was overwhelmingly set to "None". Also, "Lax" was used in a few percent of the requests, and use of "Strict" was negligible as shown in **Table 4**. In other words, many third-party vendors set the "None" as a tentative response rather than "Lax" and "Strict" that can limit cross-site requests. On the other hand, "Lax" values were set most in the SameSite attribute of first-party cookies. Note that cookies without the SameSite attribute are processed as SameSite=Lax by Google Chrome from February 2020 [9].

Lastly, we calculated the uniqueness of cookies for analyzing the tracking capability. More precisely, we calculated it using guesses_log10 values of zxcvbn [4]. The guesses_log10 means the number of identifiable users. 10^9 = 1,000,000,000 users can be identified when the value is 9. Since the number of Internet users all over the world was reported as 45.7 billion in December 2019 [*1], our analysis collected cookies with over 10 guesses_log10 values [3]. As a result, over 60% of first-party cookies and over 70% of third-party cookies used values with high uniqueness as shown in Tables 2 and 3, respectively. In the following sections, we analyze cookie sharing with zxcvbn_log10 $\geq$ 10 to target more meaningful cookie sharing.

[*1] Internet users identify different devices or browsers used by the same person as different users. The number therefore exceeds the number of humans alive at this point in time.

Table 6 Attributes of first-party cookies shared by link decoration.

Data	# of Sharing	Persistent	Session	HttpOnly	Secure	SameSite
Top5K	1,480	1,379 (93.18%)	101 (6.82%)	22 (1.49%)	110 (7.43%)	97 (6.55%)
Rand5K	1,057	1,011 (95.65%)	46 (4.35%)	8 (0.76%)	23 (2.18%)	27 (2.55%)

Table 5 Methods of cookie sharing.

Third-party Cookie	Link Decoration	CNAME Cloaking	Top5K	Rand5K
✗	✗	✗	832	1,605
✔	✗	✗	2,491	2,394
✗	✔	✗	54	58
✔	✔	✗	760	509
✗	✗	✔	7	1
✔	✗	✔	109	24
✗	✔	✔	0	0
✔	✔	✔	44	7

5.3 Detection Results of Cookie Sharing

We detected and counted websites using link decoration and CNAME cloaking as well as third-party cookies. **Table 5** shows the results. More than half of the websites used only third-party cookies and none of the evasion techniques covered in this paper. Approximately 16.10% and 2.16% of websites used link decoration and CNAME cloaking, respectively. In addition, there were 112 and 8 websites using only link decoration or CNAME cloaking, respectively. Webmasters of these websites must confirm whether first-party cookie sharing with third-party vendors was intended.

5.4 Analysis Results of Cookie Sharing

5.4.1 Features of Link Decoration

We counted the attributes of first-party cookies shared by link decoration **Table 6** shows the results. Almost all of the shared cookies were `Persistent` and only 147 (5.79%) were `Session`. Since `Session` cookies are used for identifying users and authentication sessions, leaking these cookies increases the security risks of session hijacking and CSRF attacks [6]. Although 133 `Secure` cookies were shared, there were no `Secure` cookie leaks via HTTP links and mixed contents. Note that the browser update in February 2020 also blocked mixed content, reducing the security risks of `Secure` cookie leakage [24]. On the other hand, 54 `SameSite` cookies were shared despite setting the attribute to "`Lax`" or "`Strict`". We can consider cookie sharing as cross-site requests that evaded the `SameSite` limitation and increase privacy risks in addition to the above security risks [16].

Next, we investigated the purpose and use of cookies shared by link decoration. We show the top 10 search results of cookie names with Cookiepedia on Top5K and Rank5K websites in **Table 7** and **8**, respectively. Top5K had the most cookie sharing for "Targeting/Advertising" purposes, and Rank5K for "Performance" measurement purposes. As mentioned in Section 5.2, these shared cookie names also suggest that advertisements and analytics are often used in Top5K websites. Although almost of cookies were categorized as "Targeting/Advertising" or "Performance", if other "Unknown" cookies, e.g., `UM_distinctid` and `_ym_uid` in Tables 7 and 8, are session cookies used for authentications, the attackers may be able to conduct session hijacking and CSRF attacks as mentioned above.

Table 7 Top 10 shared cookie names by link decoration (Top5K).

Name	Count	Purpose
`__asc`	198	Targeting/Advertising
`__auc`	198	Targeting/Advertising
`__utma`	100	Performance
`_fbp`	83	Targeting/Advertising
`__atuvs`	67	Functionality
`UM_distinctid`	59	Unknown
`_ym_uid`	24	Unknown
`cX_P`	20	Unknown
`cX_S`	15	Unknown
`_cb`	14	Unknown

Table 8 Top 10 shared cookie names by link decoration (Rand5K).

Name	Count	Purpose
`__utma`	123	Performance
`_fbp`	103	Targeting/Advertising
`__atuvs`	100	Functionality
`_shopify_y`	59	Performance
`_shopify_fs`	59	Performance
`_y`	59	Performance
`_s`	59	Performance
`_shopify_s`	58	Performance
`_shopify_sa_t`	47	Performance
`_ym_uid`	46	Unknown

5.4.2 Features of CNAME Cloaking

We counted attributes of first-party cookies shared by CNAME cloaking. As shown in **Table 9**, over 45% of `Session` cookies and 35% of `SameSite` cookies were shared by CNAME cloaking. Although there were 12 cookies with "`Lax`" or "`Strict`", these `SameSite` cookies were shared regardless of the attribute values since cookies shared by CNAME cloaking behave as first-party cookies. This result signifies that users are exposed to security and privacy risks similar to link decoration in the previous section.

Next, we investigated the purpose and use, and show these results in **Tables 10** and **11** as well as in the previous section. Although the total number was small, we observed cookie sharing with "Functionality" and "Strictly Necessary" cookies. We assume that this represents a feature of CNAME cloaking that directly shares first-party cookies.

5.5 Source and Destination Analysis Results of Cookie Sharing

5.5.1 Domain Name for Cookie Sharing

We investigated service categories of third-party domain names set as cookie sharing destination. We show the top 10 domain categories provided by Cisco Talos on Top5K and Rank5K websites in **Tables 12**, **13**, **14**, and **15**, respectively. In link decoration, domain names of big tech companies, such as Google, Facebook, and Amazon, were frequently used. On the other hand, many domain names used in CNANE cloaking provided "Infrastructure and CDN" and "SaaS and B2B" services. We can infer that the objective of these canonical domain names (i.e., CNAME records) was load balancing with combinations of multiple domain names and IP addresses, not CNAME cloaking. These do-

Table 9 Attributes of first-party cookies shared by CNAME Cloaking.

Data	# of Sharing	Persistent	Session	HttpOnly	Secure	SameSite
Top5K	352	221 (62.78%)	131 (37.22%)	156 (44.32%)	185 (52.56%)	136 (38.64%)
Rand5K	54	36 (66.67%)	18 (33.33%)	13 (24.07%)	25 (46.30%)	20 (37.04%)

Table 10 Top 10 shared cookie names by CNAME cloaking (Top5K).

Name	Count	Purpose
JSESSIONID	23	Strictly Necessary
AWSALB	14	Unknown
AWSALBCORS	14	Unknown
pardot	14	Targeting/Advertising
gmid	13	Unknown
ucid	13	Unknown
_sp_v1_uid	11	Functionality
_sp_v1_data	11	Functionality
_sp_v1_ss	11	Functionality
acw_tc	7	Unknown

Table 11 Top 10 shared cookie names by CNAME cloaking (Rand5K).

Name	Count	Purpose
anon_id	7	Targeting/Advertising
pardot	5	Targeting/Advertising
JSESSIONID	2	Strictly Necessary
__cfduid	2	Strictly Necessary
_sp_v1_uid	2	Functionality
_sp_v1_data	2	Functionality
_sp_v1_ss	2	Functionality
gmid	2	Unknown
ucid	2	Unknown
_shopify_y	1	Performance

Table 12 Top 10 destination domain names of link decoration (Top5K).

Domain Name	Count	Category
certify.alexametrics.com	399	SaaS and B2B
www.facebook.com	97	Social Networking
www.google-analytics.com	85	Computers and Internet
m.addthis.com	69	Business and Industry
ssl.google-analytics.com	63	Computers and Internet
v.shopify.com	57	Business and Industry
securepubads.g.doubleclick.net	46	Advertisements
scomcluster.cxense.com	38	Computers and Internet
googleads.g.doubleclick.net	33	Advertisements
www.google.com	33	Computers and Internet

Table 13 Top 10 destination domain names of link decoration (Rand5K).

Domain Name	Count	Category
v.shopify.com	342	Business and Industry
www.facebook.com	106	Social Networking
m.addthis.com	100	Business and Industry
www.google-analytics.com	92	Computers and Internet
ssl.google-analytics.com	61	Computers and Internet
mc.yandex.ru	53	Search Engines and Portals
certify.alexametrics.com	43	SaaS and B2B
kraken.rambler.ru	25	Search Engines and Portals
bam.nr-data.net	23	Infrastructure and CDN
www.google.com	19	Computers and Internet

main names may not be malicious at this time. However, risks of cookie leakage will increase if these domain names are hijacked or become malicious. For example, a subdomain takeover is known as one method to gain control of a domain name, and a security vendor reported that websites in the United States have been hacked by subdomain takeovers via misconfigured CNAME records [14]. If domain names used in CNAME cloaking are taken over by this attack, then users are exposed to risks due to cookie leakage.

5.5.2 Privacy Policy and Cookie Description

We manually investigated 139 first-party websites in English

Table 14 Top 10 destination domain names of CNAME cloaking (Top5K).

Domain Name	Count	Category
mms.fra.sp-prod.net	15	Infrastructure and CDN
cluster3.technolutions.net	9	Computers and Internet
Frontier-Airlines-lb -2074229919.us-east-2.elb .amazonaws.com	8	SaaS and B2B
pi-ue1-lba1.pardot.com	8	SaaS and B2B
lb.eu1.gigya.com	8	Business and Industry
pi-ue1-lba2.pardot.com	8	SaaS and B2B
message-fra.sp-prod.net	7	Infrastructure and CDN
elb-multiapps-570371819.us -east-1.elb.amazonaws.com	7	SaaS and B2B
message200-fra.sp-prod.net	6	Infrastructure and CDN
pi-ue1-lba3.pardot.com	6	SaaS and B2B

Table 15 Top 10 destination domain names of CNAME cloaking (Rand5K).

Domain Name	Count	Category
cs1143.wpc.chicdn.net	7	Infrastructure and CDN
pi-ue1-lba6.pardot.com	4	SaaS and B2B
pi-ue1-lba2.pardot.com	4	SaaS and B2B
mms.iad.sp-prod.net	4	Infrastructure and CDN
message200-iad.sp-prod.net	3	Infrastructure and CDN
cluster3.technolutions.net	3	Computers and Internet
a.api.permutive.app	2	Not Actionable
vbest-elb001-1086081390 .ap-northeast-1.elb .amazonaws.com	2	SaaS and B2B
d1mp2mpfkjrjed.cloudfront.net	2	Infrastructure and CDN
domains.shoplineapp.com	2	Shopping

that share cookies to the destination domain names listed in the previous section to determine whether they define a privacy policy. As a result, 99 (71.22%) websites defined privacy policies. On the other hand, we found that some of the other websites without privacy policy definitions have only "Privacy Policy" strings without links or only links to privacy policies without content.

Next, we investigated cookie descriptions in these privacy policies. As a result, 85 (61.15%) privacy policies had descriptions about cookies. However, many policies were only general statements and only 35 (25.18%) had more specific descriptions such as third-party vendor names.

6. Discussion

6.1 Security Risks of Cookie Sharing and Suggestion

Security risks such as session hijacking and CSRF attacks are exposed by unnecessary cookie sharing [16], [25].

In the case of link decoration, we can decrease the risks by limiting access through JavaScript's `document.cookie` with the `HttpOnly` attribute and being careful not to embed sensitive data in URLs. On the other hand, CNAME cloaking shares first-party cookies using the cookie feature itself, not URLs. Therefore, we cannot limit cookie sharing of CNAME cloaking using the `HttpOnly` attribute as mentioned above. Moreover, even if we set `SameSite` attributes that can limit cross-site requests, cookies are still shared regardless of the attribute values as described in Section 5.4.2 since cookies shared by CNAME cloaking behave as first-party cookies. Therefore, we need to use third-party vendors

without CNAME cloaking techniques and/or deploy DNS-level blocking to decrease these risks [13].

6.2 Privacy Risks of Cookie Sharing and Suggestion

Unnecessary cookie sharing exposes user behavior and interest to privacy risks due to tracking [17], [23]. Webmasters must clearly describe the process of cookie collection and analysis in privacy policies. Moreover, they need to clearly explain the purpose, use, and content of third-party cookies, and shared cookies. In addition to the above, they should provide opt-in/opt-out functions of cookies for website visits from foreign countries such as the United States and EU and must assume that cookie-related regulations will apply.

Third-party vendors must use approved methods such as the privacy sandbox [22] rather than sneaky approaches. It is important to seek a harmonious balance between user privacy and their benefits while gaining user consensus.

6.3 Limitation

Our analysis collected cookies observed while crawling websites from Japan. Therefore, we could not analyze cookies observed while crawling websites from other regions such as the United States and EU. Since we expect that these analysis results will differ from the results in our paper, we will perform such collection and analysis in future work.

In our analysis of link decoration and CNAME cloaking, we filtered out cookies with low uniqueness. Moreover, we analyzed simple link decoration only when cookie values were directly embedded in URLs and did not analyze obfuscated/encoded link decorations. We expect all the above limitations to have some impact on our results. However, we consider the effect to be small because it is equivalent to underestimating the amount of cookie sharing.

7. Conclusion

Third-party vendors are collecting cookies shared by link decoration and CNAME cloaking in addition to third-party cookies. In this paper, we analyze cookie sharing and evaluated security and privacy risks to users. We find that some cookie sharing evades security features such as `SameSite` attributes although there seem to be no malicious third-party vendors. We hope that our evaluation results will accelerate countermeasures by countries, vendors, and webmasters and improve web security and privacy for users.

References

[1] How We Classify Cookies, available from ⟨https://cookiepedia.co.uk/classify-cookies⟩.
[2] Intelligence Categories - Cisco Talos Intelligence Group - Comprehensive Threat Intelligence, available from ⟨https://talosintelligence.com/categories⟩.
[3] World Internet Users Statistics and 2020 World Population Stats, available from ⟨https://www.internetworldstats.com/stats.htm⟩.
[4] zxcvbn, available from ⟨https://github.com/dropbox/zxcvbn⟩.
[5] Developers: Get Ready for New SameSite=None; Secure Cookie Settings (2019), available from ⟨https://blog.chromium.org/2019/10/developers-get-ready-for-new.html⟩.
[6] HTTP Cookie (2019), available from ⟨https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies⟩.
[7] Alexa Top Sites (2020), available from ⟨https://www.alexa.com/topsites⟩.
[8] How private browsing works in Chrome (2020), available from ⟨https://support.google.com/chrome/answer/7440301⟩.
[9] SameSite Cookie Changes in February 2020: What You Need to Know (2020), available from ⟨https://blog.chromium.org/2020/02/samesite-cookie-changes-in-february.html⟩.
[10] The Chromium Projects (2020), available from ⟨https://www.chromium.org/Home⟩.
[11] Acar, G., Eubank, C., Englehardt, S., Juarez, M., Narayanan, A. and Diaz, C.: The Web Never Forgets: Persistent Tracking Mechanisms in the Wild Categories and Subject Descriptors, *ACM SIGSAC Conference on Computer and Communications Security* (*CCS*) (2014).
[12] Cahn, A., Alfeld, S., Barford, P. and Muthukrishnan, S.: An Empirical Study of Web Cookies, *World Wide Web Conference* (*WWW*) (2016).
[13] Cointepas, R.: CNAME Cloaking, the dangerous disguise of third-part trackers (2019), available from ⟨https://medium.com/nextdns/cname-cloaking-the-dangerous-disguise-of-third-party-trackers-195205dc522a⟩.
[14] Curran, P.: Trump Website Hacked: Subdomain Takeover Defaces Fundraising Site (2017), available from ⟨https://www.checkmarx.com/blog/trump-website-hacked-subdomain-takeover-defaces-fundraising-subdomain/⟩.
[15] Dao, H., Mazel, J. and Fukuda, K.: Characterizing CNAME Cloaking-Based Tracking on the Web, *IEEE/IFIP Network Traffic Measurement and Analysis Conference* (*TMA*) (2020).
[16] Franken, G., Van Goethem, T., Leuven, K.U. and Joosen, W.: Who Left Open the Cookie Jar? A Comprehensive Evaluation of Third-Party Cookie Policies, *USENIX Security Symposium* (2018).
[17] Mayer, J.R. and Mitchell, J.C.: Third-party web tracking: Policy and technology, *IEEE Symposium on Security and Privacy* (*S&P*), pp.413–427 (2012).
[18] Papadopoulos, P., Kourtellis, N. and Markatos, E.P.: Cookie synchronization: Everything you always wanted to know but were afraid to ask, *The Web Conference* (*WWW*), pp.1432–1442 (2019).
[19] Papadopoulos, P., Kourtellis, N., Rodriguez, P.R. and Laoutaris, N.: If you are not paying for it, you are the product: How much do advertisers pay to reach you?, *ACM SIGCOMM Conference on Internet Measurement Conference* (*IMC*) (2017).
[20] Reisman, D., Englehardt, S., Eubank, C., Zimmerman, P. and Naranyanan, A.: Cookies that give you away: Evaluating the surveillance implications of web tracking, *World Wide Web Conference* (*WWW*) (2015).
[21] Sanchez-Rola, I., Amico, M.D., Balzarotti, D. and Santos, I.: Can I Opt Out Yet? GDPR and the Global Illusion of Cookie Control, *ACM Symposium on Information, Computer and Communications Security* (*AsiaCCS*) (2019).
[22] Schuh, J.: Building a more private web (2019), available from ⟨https://www.blog.google/products/chrome/building-a-more-private-web/⟩.
[23] Staicu, C.A. and Pradel, M.: Leaky images: Targeted privacy attacks in the web, *USENIX Security Symposium*, pp.923–939 (2019).
[24] Stark, E. and Lopez, C.J.R.I.: No More Mixed Messages About HTTPS (2019), available from ⟨https://security.googleblog.com/2019/10/no-more-mixed-messages-about-https_3.html⟩.
[25] Watanabe, T., Shioji, E., Akiyama, M., Sasaoka, K., Yagi, T. and Mori, T.: User Blocking Considered Harmful? An Attacker-Controllable Side Channel to Identify Social Accounts, *IEEE European Symposium on Security and Privacy*, pp.323–337 (2018).
[26] Wheeler, D.L.: zxcvbn: Low-budget password strength estimation, *USENIX Security Symposium*, pp.157–173 (2016).
[27] Wilander, J.: Preventing Tracking Prevention Tracking (2017), available from ⟨https://webkit.org/blog/7675/intelligent-tracking-prevention/⟩.
[28] Wilander, J.: Intelligent Tracking Prevention 2.2 (2019), available from ⟨https://webkit.org/blog/8828/intelligent-tracking-prevention-2-2/⟩.

Editor's Recommendation

This paper analyzes new security risks emerging from link decoration that embeds cookies in URLs and CNAME cloaking which tricks browsers into treating third-party cookies as a first-party. Authors present a method of analyzing cookie sharing and evaluated security and privacy risks to users. In addition, this paper reported the analysis results of cookie sharing for two types of datasets: Top5K and Rand5K from AlexaTopSites. The paper gives insights to readers in this research field and thus is selected

as a recommended paper.

(Chief examiner of SIGSCEC Toshihiro Yamauchi)

Yuta Takata received his B.E., M.E., and Ph.D. degrees in computer science and engineering from Waseda University, Japan in 2011, 2013, and 2018. He was a researcher at NTT from 2013 to 2018. He is currently a senior researcher and a manager at Deloitte Tohmatsu Cyber LLC, Tokyo, Japan. Since joining Deloitte in 2019, he has been engaged in R&D of technologies and solutions related to cyber security, web security and privacy while working on utilizing the research results in business.

Daiki Ito received his B.E. and M.E. degrees in electrical and electronic engineering from Kobe University, Japan in 2015 and 2017. He joined Deloitte Tohmatsu Cyber LLC, Tokyo, Japan in 2019. He is currently a researcher interested in OSINT (Open Source INTelligence) and security threats related to Internet-reachable devices, domain names, and phishing. He has been engaged in R&D of technologies and solutions based on these interests.

Hiroshi Kumagai worked as a lead analyst in JPCERT/CC from 2011 to 2015. He was a researcher at PwC from 2015 to 2019. In 2019, he joined Deloitte Tohmatsu Cyber LCC, Tokyo, Japan where he is currently a principal researcher. His research interests include threat intelligence, dark web, cryptocurrency, fake news, and he has been engaged in R&D of technologies and solutions based on these interests.

Masaki Kamizono led the Cyber Security Laboratory at PwC and worked as a senior researcher at NICT from 2015 to 2019. In 2019, he joined Deloitte Tohmatsu Cyber LCC, Tokyo, Japan as CTO to launch the Advanced Cyber Security Laboratory. He leads the R&D team, and consistently develops new solutions and new businesses based on R&D. He has also been engaged in human resource development.