An Analysis of First-Party Cookie Exfiltration due to CNAME Redirections

Tongwei Ren*, Alexander Wittman†, Lorenzo De Carli*, Drew Davidson†
*Worcester Polytechnic Institute, †University of Kansas

***Abstract*—DNS CNAME redirections, which can "steer" browser requests towards a domain different than the one in the request's URI, are a simple and oftentimes effective means to obscure the source of a web object behind an alias. These redirections can be used to make third-party content appear as first-party content. The practice of evading browser security mechanisms through misuse of CNAMEs, referred to as CNAME cloaking, has been recently growing in popularity among advertisers/trackers to bypass blocklists and privacy policies.**

While CNAME cloaking has been reported in past measurement studies, its impact on browser cookie policies has not been analyzed. We close this gap by presenting an in-depth characterization of how CNAME redirections affect cookie propagation. Our analysis uses two distinct data collection samples (June and December 2020). Beyond confirming that CNAME cloaking continues to be popular, our analysis identifies a number of websites transmitting sensitive cookies to cloaked third-parties, thus breaking browser cookie policies. Manual review of such cases identifies exfiltration of authentication cookies to advertising/tracking domains, which raises serious security concerns.

I. Introduction

The domain name system (DNS) is a fundamental part of the web infrastructure, allowing domain names to be used as identifiers throughout much of the web browser stack. The most obvious use of domain names is to serve as aliases to IP addresses, but more complex and subtle aliasing relationships can be constructed through DNS records. The entity that owns the domain controls the DNS records, and may either map a (sub)domain to an IP address via an A/AAAA record, or to another domain name—which the DNS will recursively look up—via a CNAME record. While CNAME redirections are invisible to higher layers of abstraction in the browser, they play a crucial role in informing identities and trust. A website may, for instance, legitimately map one of its subdomains to a domain name belonging to a content delivery network (CDN) so that scripts, images, and rich media served from the CDN can interact with content served from the first-party server without crossing a security boundary. By specifying, through DNS, that the CDN and first party content come from the same entity, a visiting browser will *identify and trust* the CDN content as coming from the website domain itself. However, CNAME records can also be used for less benign purposes.

Misuse of CNAME records to hide the source of advertising/tracking content is a recurring issue [30], [29] which has seen a resurgence [21]. Recently, several websites have been found to deploy redirections for this purpose [31], [19], [36]. This utilization of CNAME records allows advertisers to serve content that appears to come from the first-party, complicating the task of ad-blockers. CNAME use is by no means the only strategy to integrate advertising services into first-party code [29], but it has a notably problematic effect of weakening a core mechanism of web security, the same origin policy (SOP)—by causing browsers to treat third-party subdomains as part of the first party origin. Forgoing origin-based isolation gives scripts imported from the CNAME domain permissions that are potentially contrary to the intent of the website administrator, such as the capability to read and manipulate first-party cookies and the DOM. Due to their negative connotation, these CNAME redirections of advertising/tracking resources from a first- to third-party domain have been termed *CNAME cloaking.* We furthermore term a third-party domain which is used as a target of CNAME cloaking as a *cloaked domain.*

In this paper we focus on the implications of CNAME cloaking on browser cookie policies. For web developers, correctly configuring a cookie policy is a delicate balancing act between security and usability. Cookies set with incorrect configurations can have significant consequences, from major functionality breakdown to allowing malicious code to impersonate an authenticated user. CNAME redirections throw in an additional layer of complexity, as they entangle otherwise-distinct origins. In our work, we perform an analysis of cookie behavior across Alexa top-10000 websites. We identify numerous instances of CNAME cloaking and a significant number of cases where first-party cookies are sent to third-parties due to cloaking. Furthermore, we carry a manual, in-depth analysis on select websites, showing first-party authentication cookies being propagated to third-party advertisers. Propagation of authentication cookies to (non-authentication-related) third-parties serve no functional purpose, and is likely an unintended consequence of cloaking. At the same time, such events represent serious security lapses.

Mitigations against CNAME cloaking are still nascent, as the problem is not well-studied. Recent additions to ad-blocking tools including UBlock Origin and AdBlocker were introduced as recently as November 2019 [16]. Unfortunately, these tools rely on manually-curated blocklists, which have no guarantee of being exhaustive. 95 of the 101 distinct entities that we found to be the targets of CNAME cloaking were not covered in these lists. Despite previous anecdotal evidence of sensitive cookies being forwarded to cloaked domains [19], we believe our work to be the first to investigate the security/privacy impact of CNAME cloaking with respect to cookies. Integrated browser protections are also limited, and

Workshop on Measurements, Attacks, and Defenses for the Web (MADWeb) 2021
25 February 2021
ISBN 1-891562-67-3
https://dx.doi.org/10.14722/madweb.2021.23xxx
www.ndss-symposium.org

those protections that do exist are quite new: Brave's CNAME cloaking mitigations for content blocking were announced as recently as October 27th, 2020 [17], and Apple's defenses were only released to Safari alongside the recent Big Sur update [14]. Furthermore, the popular Chrome, Firefox, and Edge browsers have no CNAME defenses.

Studying the threat of CNAME cloaking in practice is essential to inform the debate surrounding the security and privacy implications of online tracking and advertising. However, conducting a large-scale study of the impact of CNAME cloaking on cookie policies is challenging, due to the scale and complexity of the analysis. Identifying cloaking and cookie exfiltration requires sifting through all objects served by the large number of websites that can be conceivably construed as popular. To handle the scalability challenge, we leverage a custom-built analysis platform that allows us to analyze millions of HTTPS requests towards 95% of Alexa Top-10000 websites, identifying and reporting all instance of CNAME redirections. We augmented this automation platform with extensive manual analysis of exfiltrated cookies and the domains to which they are propagated. **Our main contributions are:**

- We track cookie creation and exfiltration for Alexa top-10000 websites, identifying a number of instances where first-party cookies are sent to unrelated advertising third-parties due to CNAME cloaking.
- We perform an in-depth manual analysis for a select sample of websites, which shows that sensitive content is exposed in practice to third-parties through CNAME cloaking, including authentication tokens that may allow for impersonation attacks.
- We perform our data collection and analysis in two experiments six months apart (June and December 2020), identifying trends that underscore the prevalence of the phenomenon.
- We make our dataset available for other researchers to investigate CNAME cloaking[1].

To the best of our knowledge, our work is the first to shed light on the impact of CNAME cloaking on cookie exfiltration.

II. Background

A. CNAME Redirections

A CNAME (Canonical Name) redirection is used to inform a DNS cache performing a resolution that the provided hostname is an alias for a different hostname. Upon receiving a CNAME answer, a DNS cache issues a second query for the canonical name, and subsequently packages (i) the original CNAME record, and (ii) the A (or AAAA) record for the canonical hostname, in a single response to the client. For example, consider a hostname $www.domainA.com$ whose resolution encounters a CNAME record pointing to $srv1.cloudprovider.com$. The answer received by the client will include both this $CNAME$ record, and an A record expressing the IP address of $srv1.cloudprovider.com$. In some cases, multiple CNAME records may be encountered on the path to an A/AAAA record.

CNAME redirections offer an additional layer of indirection on top of the one naturally provided by DNS, and simplify a number of benign use cases (e.g., redirecting requests towards the main hostname to an external cloud provider). Importantly, since DNS resolutions are low-level operations unconstrained by browser policy, a CNAME redirection may be to a completely distinct domain, as in the previous example wherein a request for $www.domainA.com$ was CNAME-redirected to $srv1.cloudprovider.com$.

B. Definitions

First, we define the **first-party domain** to be the domain of the website the user is visiting in a browsing session, i.e., the domain shown in the browser URL bar. Note that although *website* and *domain* are distinct concepts, for the purpose of our analysis a website corresponds to exactly one domain. A **first-party subdomain** is a subdomain of the first-party domain. A **third-party domain** is a domain, different from the first-party domain, from which content is fetched while browsing the first-party domain. For example, if the user is visiting *www.domainA.com* and the website includes an iframe with content from *domainB.com*, *domainA.com* is a first-party domain and *domainB.com* a third-party domain. A **third-party subdomain** is a subdomain of a third-party domain.

*A **first-party redirection** consists of a CNAME redirection from a first-party subdomain to a different domain.* As an example, consider a website at *www.domainA.com*, which includes some web resources which are fetched from *sd1.domainA.com*. In turn, this subdomain CNAME-redirects to *sd2.domainB.com*.

*A **third-party redirection** consists of a CNAME redirection from a third-party subdomain to a different domain.* For example, consider the situation where the website at *www.domainA.com* includes an iframe pulling content from *sd1.domainB.com*, and this subdomain CNAME-redirects to *sd2.domainC.com*.

In this study, we only consider first-party redirections. We consider this type of redirection to be particularly important, since it allows interactions between first- and third-party content. Therefore, we ignore third-party redirections and redirections that remain within the same domain, be it first- or third-party. As a CNAME redirection makes it difficult for the browser logic to identify the true source which serves an HTTP request, we refer to the domains at the end of first-party redirection chains as **cloaked domains**. We also refer to first-party redirections as **cloaking**.

C. Browser Cookie Policies

One of the most straightforward ways to maintain user identities on the web is to associate a stateful token to each browser in the form of a cookie. Cookies are text-based key-value pairs that can be set by a first- or third-party, and are managed by the browser. Cookies are popular for authentication, since an authentication token can be stored as a key-value pair. Cookies are also popular for user tracking, since they offer a means to re-identify a user across multiple web requests. In its most basic form, a first-party can track a user with a key-value pair representing a unique ID in a cookie, and then request that cookie from every visiting browser. On the

[1] https://www.dropbox.com/sh/nd1b4bzjhlp1fr9/AABTMDV9OeeOoL4gu_TzATKga

first visit to that site, the cookie will not yet exist, so a unique identifier will be generated and sent to be managed by the browser. On subsequent visits, the first-party can retrieve the tracking ID from the cookie, and re-identify the user for whom the cookie was generated. Similarly, third-party entities can generate and serve cookies to users for a variety of purposes, including tracking across the web. Due to the security- and privacy-sensitive content of cookies, browser policies limiting propagation of cookies use a notion of origin-based isolation as part of the cookie policy: a cookie may specify the subdomain of servers that are allowed to see the cookie (by default the domain of the URL serving the cookie will be used if no subdomain is set). The cookie policy works in cooperation with the well-known same-origin policy (SOP) [33] to ensure that data is not leaked through cookies: the browser can prohibit access to a cookie from other origins even if it within the cookie's domain. These mechanisms are important security measures, since cookies can contain authentication data and personally-identifiable information.

Indeed, a use of CNAME cloaking is to allow a third-party to set tracking cookies that appear to come from the first-party domain. The effect of CNAME Cloaking is to obscure the true origin of a web request for a resource like a cookie. In the process, the cloaked domain can subvert the cookie policy for the purpose of more invasive interactions, such as user tracking. Since cookies can be used for authentication, weakening the distinction may also lead to information leakage to third parties, such as credential stealing.

D. Browser-based Countermeasures

In recognition of the threat of CNAME cloaking, several client-side solutions have been proposed or implemented to mitigate the threat. Since the CNAME record is a part of DNS, it is sufficient for a browser to recursively check CNAME records that are encountered as part of the request. In effect, this approach uncovers the true provenance of a domain obscured by the CNAME record. This approach, which is implemented natively in the Brave browser and in UBlock Origin on Firefox, is particularly effective in allowing block-lists to transcend the threat of CNAME cloaking. However, without a re-implementation of the browser cookie policy that is CNAME-cloaking aware, cloaked third parties that do not appear on a blocklist can still access first-party resources.

E. Object of Study

The goal of our work is to characterize first-party cloaking which causes cookies to propagate beyond the first-party domain which sets them. However, not all such propagations are significant. In particular, we focus on instances where cookies propagate to advertising- and tracking-related third-party domains which bear no direct relation to the first party setting the cookie. Such redirections are problematic because they reveal potentially sensitive data belonging to the first party. In some cases, evidence strongly suggests that cookies are exfiltrated by mistake, i.e., as an unintentional consequence of cloaking (ref. Section IX-B).

III. Methodology

In this section, we outline our data collection infrastructure and methodology.

Our goal is to measure the extent to which CNAME cloaking affects the propagation of first-party cookies to third-parties. We limit the scope of our measurements to the Alexa top-10,000 websites. This is necessary as our analysis (detailed below) involves complex, time-consuming manual steps. This decision is further discussed in Section IX-D.

Our analysis is structured in four steps: (a) bulk data collection; (b) domain classification; (c) cookie lifecycle analysis; and (d) manual website analysis. We further performed a high-level evaluation of browser defenses (e). In the following, we detail each step.

A. Data Collection

Cookies are set and transmitted by HTTP(s) responses and requests; therefore, studying the lifecycle of cookies for a website requires logging all requests/responses generated while visiting a website. Furthermore, the DNS resolution chain associated with the hostname in each request must also be logged. Due to the volume of data, we limit bulk data collection to the homepage of each website in the Alexa Top-10,000 list (we discuss why we consider this an acceptable limitation in Section IX-D).

To collect bulk website data, we use a custom crawler/logger based on a current version of the Firefox browser, the Selenium web testing framework [9], and the mitmproxy HTTPS proxy [7]. As neither Selenium nor mitmproxy provide insight into DNS resolutions, immediately after each request we issue a DNS resolution for the respective subdomain using dnspython (unless we previously encountered the subdomain during the same visit). Changes in DNS records between resource fetching and DNS resolution are unlikely: the interval between the two operations remains below a second, while DNS updating frequencies tend to be minutes to hours.

We encapsulate our tooling in a Singularity container and spawn multiple instances on an HPC cluster at the University of Kansas. A separate container running MongoDB server is used as log storage. During data collection, our Singularity cluster is configured to spawn 20 concurrent instances. Each container is short-lived, and receives a list of 500 websites to visit on startup. It visits each website's homepage for 30 seconds, to ensure complete loading of dynamic and delayed content. The container logs the content of all HTTP requests and responses, including header and request/response body. Browser caches are cleared between each visit.

In order to obtain an historical perspective on cloaking we repeated the collection in June and December 2020, in both cases using the infrastructure described above. The raw data upon which we perform our analysis consist of HTTP requests and responses, and DNS responses associated to the URI in each request.

1) Isolating Candidate Redirections: Our object of study concerns redirections that cause cookies to be sent to third-party advertisers/trackers. Before further analyzing our dataset, we begin by isolating a set of candidate redirections that are likely to be advertising-/tracking-related, and we restrict further analysis to those. We use two approaches:

Domain-based: We extract all domains in the EasyList [4] and EasyPrivacy [6] lists, which are popular blocklists used by browser-based ad-blockers. We only extracted domains that explicitly appear in blocking rules, e.g., the rule `||myfinance.com^$domain=cnn.com|marketwatch.com` would cause us to extract the domain `myfinance.com`. We then select all first-party redirections ending in one of such domains.

URL-based: The lists introduced above also include rules that identify advertising and tracking content based on the presence of specific byte patterns in the URL (regardless of domain). We extract all such patterns and we select all first-party redirections affecting requests that match those patterns.

Together, the redirections collected via the methods above constitute 19% of all first-party redirections. Note that the above methods are prone to false positives, as not all identified domains are solely used to distribute advertising/tracking content. In practice this does not constitute a problem, as identified domains are further manually analyzed, as discussed below. We characterize our dataset (pre- and post-filtering) in Section IV.

B. Domain Classification

Philosophically, the issue under study is propagation of cookies to domains whose ownership has been misrepresented due to CNAME redirections. However, redirections can be deployed in a variety of ways, not all of which misrepresent the entity to which cookies are transmitted.

For example, redirecting a first-party domain to an ad provider which is also owned by the first-party, is arguably not a misrepresentation. Such redirections are deployed often and as a matter of traffic optimization, and data exchanged across the two domains remain within control of the user and the first-party. On the opposite end of the spectrum, a first-party redirecting requests to an external tracking company is concerning, since the user may not wish to expose its private information to such a third party. In between these extremes there are a number of ambiguous cases, which include redirections towards a number external CDNs and PaaS which are sometimes used for advertising/tracking (and therefore appear in blocklists). These entities host content upon which the first party retains varying degrees of control.

Given the considerations above, we implemented a high-level taxonomy based on manual investigation of each *(source domain, destination domain)* pair in the dataset, and we make sources for each labeling decision publicly available in our dataset (ref. Section I). We include the following categories:

Same-organization: includes redirections in which source and target domain are broadly part of the same organization. An example is *msn.com* redirecting to *microsoft.com*. We put a redirection in this category if two domains can be determined to belong to the same organization either via publicly available data, or by using the "adns" methodology by Krishnamurthy and Wills [29]. The latter puts two domains under the same ownership if they declare the same authoritative DNS server (we manually reviewed the output to filter out false positives).

External ad/tracking: This category includes redirections which do not fall under **Same-organization**, and whose destination domain belongs to an organization whose primary business is online advertising and/or tracking. In order to make this determination we: (i) reviewed whether the destination domain advertises itself as a provider of advertising services; and (ii) reviewed company information aggregators such as *crunchbase.com*. Conservatively, we only put a domain in this category if the information sources described above unambiguously identify the domain as belonging to an advertiser/tracker.

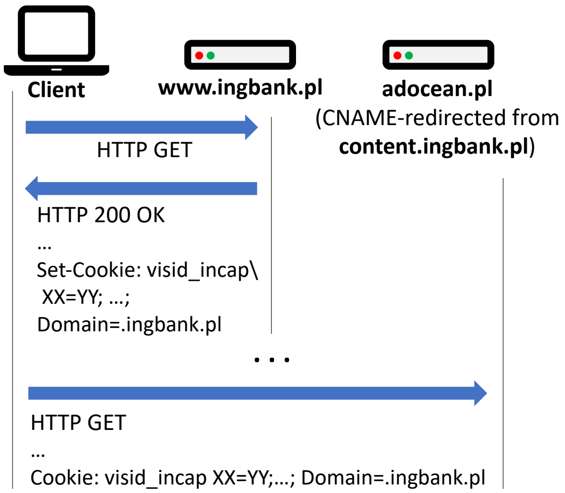


Fig. 1: Cross-domain cookie transmission

Other 3rd-Parties: This category includes all redirections which do not fall under either **Same-organization** or **External ad/tracking**. Most first-party redirections not meeting this condition are directed towards CDNs and cloud providers.

We report the results of our domain analysis in Section V.

C. Cookie Lifecycle Analysis

After having performed domain classification, we specifically focus our analysis on the lifecycle of cookies set by first parties. In both the June and December datasets, we isolate browsing sessions, among those in our dataset, in which the first-party domain issues a *Set-Cookie* for a cookie which is later transmitted to a third-party via a first-party redirection. During the process, we filter out cookies that are set by first-party domains which consist of cloaked third-party ones, as those cookies are not truly set by a first-party. For example, Figure 1 presents a situation (extracted from our June 2020 dataset) in which what appears to be a first-party tracking cookie set during a visit to *www.ingbank.pl* is later transmitted to *adocean.pl*, despite browser design forbidding explicit cross-domain cookie transmissions.

In practice, the first-party may issue multiple *Set-Cookie* directives across several responses; however, the browser coalesces these cookie key/value pairs into a single string when they are sent back in HTTP request headers via a *Cookie:* field. To avoid ambiguities, here we consider the content of a single *Cookie:* HTTP request header field as one cookie. This analysis is automated and at the end of it we map each cookie to the label (*Same organization*, *External ad/tracking*, *Other 3rd-parties*) assigned to the third-party domain to which it gets transmitted. Results are reported in Section VI.

D. Manual Cookie Analysis

In order to gain insight on inter-domain cookie transmission beyond each website's homepage, we performed a second experiment focused on exfiltration of sensitive cookies. Our definition of *sensitive cookie* is provided below; intuitively, those are cookies which store information pertaining the user identity and/or authentication. We further define the implication of these cookies being exfiltrated in Section IX.

For this experiment, we selected 62 websites, among those found perform redirections towards domains in the External Ad/Tracking category (roughly 14% of all sites in this group). We selected these sites at random, excluding those which require payment for creating an account. We then created user accounts on each of them and recorded authenticated browsing sessions using our *mitmproxy* setup. Finally, we repeated this analysis for the same websites in December 2020 (we used the same set of sites in both sessions).

Collecting candidate cookies: in each experiment, during the login process we manually identified candidate cookies via the Firefox network console. Using the console UI, we identified first-party cookies set in the HTTP traffic generated in response to logging into the website (we also verified that each of these cookies is set by the actual first-party, and not by a cloaked third-party). We then searched HTTP session logs to determine whether each candidate had been transmitted to a cloaked domain falling in the "External Ad/Tracking" category. We retained only candidates for which at least one such transmission was observed.

In the next step, we analyzed each candidate cookie to determine if it is to be considered *sensitive*. Note that here a candidate cookie is intended as a key/value pair, and not the whole *Cookie* field in the HTTP header. We define a cookie sensitive if it matches one of the following conditions (or both):

a) Cookie stores sensitive content: We consider this condition matched if the cookie visibly contains one or more of the following data: user name, user email, user website ID, geographical location, profile data, account status, and timestamp. We manually analyzed each candidate to attempt to determine its content. We found this to be a complex process, due to the fact that cookie content is not bound to follow any standardized format. Many cookies appear to be textual serializations of internal backend state; without knowledge of the backend logic, it is impossible to make sense of them. Luckily, however, some cookies are formatted according to standards (e.g., JWT, OpenAM), and/or have easily recognizable structure. Overall, decoding was an empirical, heuristic process where we used our own domain knowledge to determine the appropriate decoding technique. For example, apply base64 decoding if the content appears to be base64-encoded data; apply an OpenAM decoder if the cookie is in OpenAM format, etc. (OpenAM [8] is an access management platform). As a result, the analysis is conservative: it may exhibit false negatives, but not false positives. We term cookies belonging to this category as "information cookies".

b) Cookie is necessary for user identity/authentication: We verified each candidate by visiting the website again after manually deleting the candidate from the browser cookie storage. A website requiring the user to re-authenticate themselves after deleting a cookie strongly suggests the cookie plays a role in the authentication process (we also verified that the website recognizes the user if the cookie is not deleted). We termed such cookies "authentication cookies". We also noted that in some cases, after deleting a cookie it was possible to log back in by confirm the user identity without entering a password. A brief analysis suggests that server-side logic may be recovering the session via data in other cookies and/or local storage. We term cookies leading to this behavior as "identity cookies".

	June 2020	**December 2020**
Websites	9,578	9,683
HTTP Requests	1,576,505	1,554,789
HTTP Responses	1,552,791	1,533,379
Avg Req size [B]	1,364	1,428
Avg Resp size [B]	104,535	102,566
First-party redirections	188,300	203,957
Redirections after filtering	28,250	46,745

TABLE I: Summary of main dataset

Note that information cookies and identity/authentication cookies are not disjoint sets, since some cookies may match both conditions above. The results of our manual analysis are reported in Section VII.

E. Browser Blocklist Evaluation

As described in Section II-D, some client-side security mechanisms exist to prevent CNAME cloaking from evading blocklists. These mechanisms, embedded in the browser or loaded via extensions, are implemented by recursively dereferencing CNAME records to check for entities on a blocklist. However, cookie leaks through CNAMEs (inadvertent or otherwise) may still occur when the cloaked domain is not blocked.

To verify the behavior of cookie sharing between sites that are not blocklisted, we arranged a simple experiment: we created two sites with distinct domains, to simulate a third party and a collaborating first party. We added a DNS record containing a CNAME record for a subdomain of the first party to the third party site, and embedded the third party page into a first party page. This basic experiment is sufficient to observe the behavior of browsers' cookie policies when neither site is on a blocklist. We verified that no browser prevents cookie sharing through CNAME records, including those browsers that implement recursive CNAME lookups for blocklists: in every major browser, the simulated third party had access to all cookies not otherwise prevented by the cookie policy.

This preliminary analysis encouraged us to perform a further experiment. We selected the Safari and Brave browsers as these have explicitly advertised their ability to prevent CNAME cloaking [3], [10]. We then evaluated whether the instances of authentication cookie exfiltration identified in our analysis of the December 2020 dataset (ref. Section III-D) happen under these browsers. Results are provided in Section VIII.

IV. Data Collection Results

Our Dataset is characterized in Table I, rows 1-5. In the June and December sessions, our data collection infrastructure successfully processed **9578** (June) and **9683** (December) websites among those in the Alexa Top-10,000 list. The remaining

	June 2020	December 2020
First-party redirections	3,330	3,417
Same-organization:	416	249
External Ad/Tracking:	509	513
Other 3rd-Parties:	2,405	2,655
Source domains	1,509	1,597
Same-organization:	176	154
External Ad/Tracking:	453	459
Other 3rd-Parties:	998	1,207
Destination domains	135	133
Same-organization:	29	24
External Ad/Tracking:	31	33
Other 3rd-Parties:	78	77

TABLE II: First-party redirections and domains in each category

websites failed to return any content, or returned various types of HTTP error codes. This data was then processed to isolate all requests for which the DNS answer included a first-party CNAME redirection (row 6 in Table I). We ignored redirections starting and ending within the same domain, which are extremely common and irrelevant to this study.

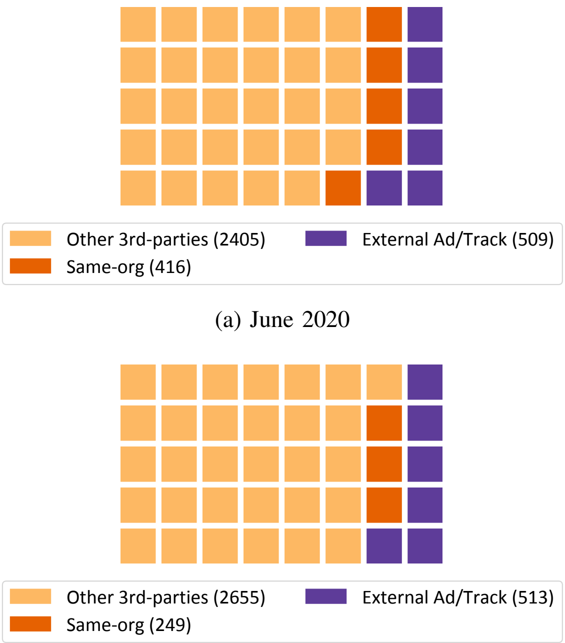
The set of first-party redirections was further filtered according to the methodology described in Section III-A1. The remaining redirections (detailed in row 7 of Table I) constitute approximately **19%** of all first-party redirections.

V. Domain Classification Results

We isolated all the *unique* first party redirections within the set described in Section IV (i.e. we collapsed redirections appearing more than once into one). For each unique redirection, we classified the target domain according to the categories of Section III-B. Figure 2 breaks down first-party redirections into the three categories above, for June and December data. Each square represents 2.5% of the total redirections.

Table II shows details of the set of redirections. In particular, it presents breakdowns by category for both the redirections themselves, and domains that appear as sources and target of such redirections. Note that some domains appear in multiple categories. The number of unique source domains deploying first-party redirections, across our two data collection sessions, is **2818** (There are **1527** overlapping domains between the sets of June and December domains). The number of unique domains appearing as target of redirections is **101** (There are **31** overlapping domains between the sets of June and December domains). Interestingly, **95 of 101** such domains did not appear in public CNAME cloaking blocklists [36] at the time of writing.

The set of domains appearing as source of advertising-related first-party redirections is far larger than that of destinations, and there are no domains belonging to both sets. Analysis of the graph defined by source and destination domains suggests that there exist multiple communities of websites using CNAME cloaking, each clustered around a highly popular content provider. At the same time, most websites which redirect their users to third-parties via CNAME cloaking depend on a single provider of cloaked content. The average outdegree for redirection sources in June (December) is **1.04** (**1.05**) and the average indegree for targets is **15.22** (**14.58**). For example, the most popular provider, *omtrdc.net*, served requests from **325** sources in the June dataset and **361** sources in the December dataset.



(a) June 2020

(b) December 2020

Fig. 2: Categorization of first-party redirections by category of destination domain

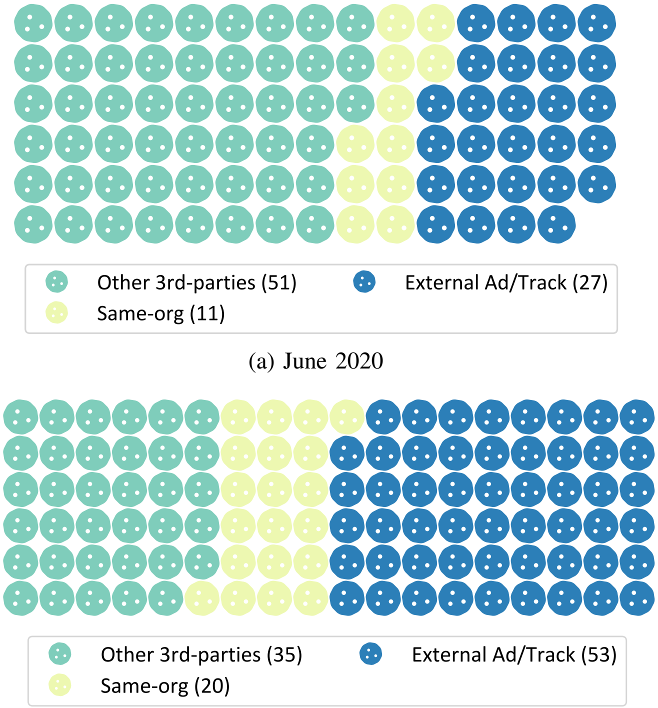
A. CNAME Redirections Chains

An iterative DNS resolution may include multiple CNAME redirections; we encountered **49** (**508**) redirections of length greater than 2 in the June (December) dataset. To categorizing these, we use the following labeling heuristic: first, if the first and last domain share ownership, we apply the "same-organization" label. Otherwise, if $>= 1$ domains in the chain are labeled as "External Ad/Tracking", we label the entire redirection as such. Otherwise, we label the redirection as "Other 3rd-parties". We also identified **85** (**318**) cases where the source domain redirects to an external one, which then redirects back to the original domain. These appear to be related to DNS-based load-balancing, and we ignored them.

VI. Cookie Lifecycle Analysis Results

In this section we demonstrate that CNAME-based cross-domain cookie transmission happens in the wild. In particular, we isolate all instances of first-party cookies transmitted to cloaked third-parties from website homepages, according to the methodology described in Section III-C.

In the June dataset, the analysis results in **89** cookies that appear in cross-domain transmissions (for a total of **521** key/value pairs). In December, we identify **108** cookies representing **914** key/value pairs. Next, we classify the third-party domains receiving first-party cookies according to the



(a) June 2020

(b) December 2020

Fig. 3: Categorization of cross-domain cookie transmission instances by category of destination domain

categories of Section III-B; results are presented in Figure 3. Table III further breaks down the number of source/destination domains in each category.

VII. Manual Cookie Analysis Results

As discussed in Section III-D, this experiment involves manually creating accounts in a select sample of 62 websites, and isolating cookies which are (1) generated in response to authentication; (2) sent to a cloaked third-party in the "External Ad/Tracking" category; and (3) contain user information and/or are necessary to conduct an authenticated session.

Table IV summarizes the results of the analysis. The table lists all domain found to leak sensitive cookies. For each domain, we specify if exfiltration happened in the June (column 2) and December (column 3) analysis. We furthermore list the type and number of leaked key/value cookie pairs ("cookies" for brevity in the following) (column 4). Cookies listed under the **I** label are Information cookies; cookies listed under the **A** label are Authentication cookies; and cookies listed under the **D** label are iDentity cookies (refer to Section III-D for a detailed description of each category). Some cookies are both information cookies and authentication (or identity) cookies. Finally, for each domain we provide a few relevant details in regards to exfiltrated cookie content (column 5).

While determining if a cookie belongs to the authentication/identity category is simple (deleting the cookie results in a user logout), determining whether a cookie contains personal information is more complex. Despite our efforts, some cookies had no obvious plaintext meaning. Other used open, reversible standards that are amenable to decoding. Figure 4 summarizes the analysis of two such cookies. *realestate.com.au* transmits three cookies storing JSON data. One contains, among other things, the semi-obfuscated user email. Another stores a JWT token [26]; such tokens are used to encode website and user identity in single sign-on systems (contextual clues suggests that this website uses Amazon Cognito). *sas.com* transmits an OpenAM session cookie. OpenAM [8] is an access management platform; the purpose of an OpenAM cookie is to facilitate single-sign on [5]. Indeed, Figure 4 shows that it contains internal server-side session identifiers. Other authentication-related cookies contain opaque data which is probably only meaningful in the context of the server-side logic. For example, *carsales.com.au* sends an ".XdmAuth" cookie, which contains 644 bytes of binary data, to advertisers. Data appears to be a randomly generated token according to entropy analysis. Removing this cookie fully logs out the user.

	June 2020	Dec. 2020
First-party (source) domains	28	26
Sends to Same-organization:	5	5
Sends to External Ad/Tracking:	6	9
Sends to Other 3rd-Parties:	18	12
Third-party (destination) domains	10	12
Belongs to Same-organization:	1	2
Belongs to External Ad/Tracking:	4	4
Belongs to Other 3rd-Parties:	5	6

TABLE III: Number of domains involved in cross-domain cookie transmission

VIII. Browser Blocklist Evaluation Results

In this section, we used the Brave and Safari browsers to visit all websites found to exfiltrate authentication cookies in our December 2020 dataset (ref. Table IV). When using Safari, **2 out of 7** instances of exfiltration were blocked (cookies were blocked on *autotrader.com* and *carsales.com.au*). When using Brave, **6 out of 7** instances of exfiltration were blocked (cookies were exfiltrated on *mathworks.com*).

IX. Discussion

A. Summary of Results

Our analysis shows that despite recent press coverage of CNAME cloaking, the practice is still observable, even on popular sites. Overall, data in Section V shows that the practice is uncommon, but not rare: more than **4%** of Alexa-10000 websites perform first-party redirections towards destination domains which can be unambiguously identified as third-party advertisers/trackers. Analysis of the historical data show a negligible decrease (**1%**), between June and December 2020, in the number of domains performing first-party redirections to destinations that are unambiguously classified as third-party advertisers/trackers. Interestingly, the decrease in the number of domains acting as destination of such redirections is more significant (**7%**). We speculate that some advertisers may be renouncing the practice due to community pressure; however, there is no conclusive evidence.

Domain	June 2020	Dec. 2020	#Key/Value Pairs	Content found in cookies
autotrader.com	✗	✓	**A/I:1**	HEX data; user email address
carsales.com.au	✓	✓	**A:1**	Opaque HEX data
cheaptickets.com	✓	✓	**A:1**; I:1	Opaque encoded data; username
childrensplace.com	✓	✓	**A:5**; I:9	Base64 data; user's name, location, ZIP, account n., reg. date
denik.cz	✓	✓	D:2; D/I:1	User email address
everydayhealth.com	✗	✓	**A:3**; I:3	Opaque HEX data; user email, username, name, birthday, ZIP
intel.com	✗	✓	**A:1**	Opaque Base64 data
mathworks.com	✗	✓	**A:1**; I:1	HEX data; username and profile-picture filename
realestate.com.au	✓	✗	D/I:1	JWT token (see Figure 4); user email address
royalcaribbeans.com	✓	✗	**A:1**	OpenAM authentication cookie
sas.com	✓	✗	D:1; I:1	OpenAM-formatted cookie (see Figure 4); username
startribune.com	✓	✓	D:5; D/I:5	JWT token; user email address, registration date and ZIP code
travelzoo.com	✓	✗	**A:1**	Opaque HEX data
vagaro.com	✗	✓	I:1	City-level user location and ZIP code

TABLE IV: Categorization of sensitive key/value cookie pairs exfiltrated to third parties (**I**: information cookie; **A**: Authentication cookie; **D**: iDentity cookie; **A/I**: Authentication and Information cookie; **D/I**: iDentity and Information cookie). In column 4, items in bold represent authentication cookies exfiltrated to third parties.

The analysis of Section VI shows that these redirections have an impact on cookie policies. A number of websites were found exfiltrating cookies to advertising/tracking third-parties on their homepage. The number is small, but increasing (**9** in the December dataset, vs **6** in the June dataset). The number of first-party cookies transmitted in such events also increased slightly (from **89** in June to **108** in December 2020). Likewise, the percentage of such cookies which is sent to advertisers/trackers increased from **30%** to **49%**. It should be pointed out, however, that these exfiltrations involve cookies generated simply in response to a non-authenticated users browsing a website homepage. Upon close analysis, most cookies appear to be related to session and/or user identification and geolocation (based the presence of keys such as "visid", "sid", "Country", and similar). A few (containing the key "ak_bsmc") appear to be related to Akamai's bot detection platform. Unfortunately, without knowledge of each specific website design rationale, it is impossible to determine whether such cookies were intended to be exposed to third-parties. It is also worth noting that the practice of having first-parties set tracking cookies on behalf of advertisers is well-documented [29]. At least some cases identified above may be instances of this practice.

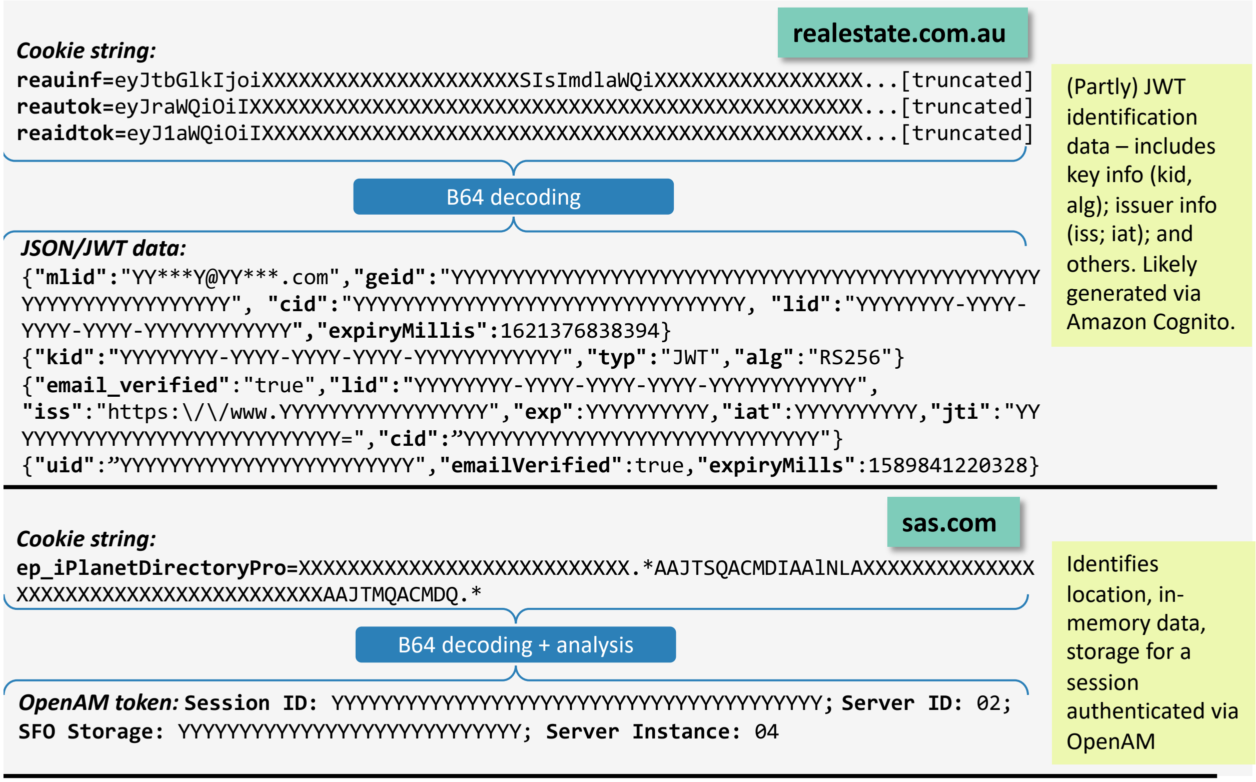
In order to evaluate the extent of cookie exfiltration beyond each site's homepage, we performed the manual analysis of Section VII. The analysis focused on a sample of websites known to perform first-party redirections, and found that roughly one every seven websites analyzed exfiltrates cookies which are generated in response to user authentication, and contain sensitive information. Through this process, we identified **9** domains exfiltrating authentication/identity cookies to third-parties in June, and **10** in December. Only **5** domains appears in both sets, suggesting churn in the websites suffering from this type of issues. Overall, the identified domains exfiltrated **46** cookies. Among those, we identified **15** authentication-related cookies which are particularly interesting. We further discuss those in the following section.

Our final experiment (Section VIII) looked at the effectiveness of browser-based blocklist in preventing sensitive cookie exfiltration. At high level, the experiment suggests that neither Safari nor Brave prevents cookie sharing through CNAME records, unless the redirection target is explicitly included in a blocklist.

B. Security Implications of our Findings

The impact of any individual instance of CNAME cloaking is difficult to predict; although a first party site may expose cookies to a given third party, as discussed above in some cases that behavior may be intentional. Without visibility into the behavior of the third party, it is likewise impossible to determine how the cookie is being used (if at all). Nevertheless, CNAME Cloaking as an aggregate phenomenon has undesirable implications for user security and privacy. In a broad, theoretical sense, origin-based isolation is one of the most recognized security principles of web browsers, which is muddied by CNAME cloaking. As our work shows, third parties and first parties are willing to collaborate in ways that blur origin-based security. Regardless of the parties' intent, browser protections that would prevent incidental data collection (i.e. cookies) are rendered less effective.

Identity and authentication-related cookies are particularly interesting for studying the impact of weakening browser protections. These cookies are generated by a website in response to a user login, and may contain authentication tokens (or similar values) whose purpose is to recognize the user as authenticated, and/or various classes of user- and session-related data. The exfiltration of authentication cookies as those detailed in Table IV carries direct practical implications, as it may open the door to impersonation and account takeover. The most direct mean to impersonate the user involve an attacker directing requests to the first-party website, presenting the obtained authentication cookies in an attempt to pose as the victim user. Note that access to authentication cookies alone may not be sufficient. First, authenticated sessions may be operated with mechanisms beyond cookies (e.g., header-based tokens). Also, some frameworks deploy countermeasures, like CSRF tokens, which are designed to ensure authentication cookies are used within a legitimate browsing session. We also point out that the domains on the receiving end of exfiltrated sensitive cookies receive them as a matter of standard browser



Cookie string:

```
reauinf=eyJtbGlkIjoiXXXXXXXXXXXXXXXXXXXXXXSIsImdlaWQiXXXXXXXXXXXXXXXXX...[truncated]
reautok=eyJraWQiOiIXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...[truncated]
reaidtok=eyJ1aWQiOiIXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX...[truncated]
```

JSON/JWT data:

```
{"mlid":"YY***Y@YY***.com","geid":"YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY
YYYYYYYYYYYYYYYYY", "cid":"YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY, "lid":"YYYYYYYY-YYYY-
YYYY-YYYY-YYYYYYYYYYYY","expiryMillis":1621376838394}
{"kid":"YYYYYYYY-YYYY-YYYY-YYYY-YYYYYYYYYYYY","typ":"JWT","alg":"RS256"}
{"email_verified":"true","lid":"YYYYYYYY-YYYY-YYYY-YYYY-YYYYYYYYYYYY",
"iss":"https:\/\/www.YYYYYYYYYYYYYYYYY","exp":YYYYYYYYYY,"iat":YYYYYYYYYY,"jti":"YY
YYYYYYYYYYYYYYYYYYYYYYYYYY=","cid":”YYYYYYYYYYYYYYYYYYYYYYYYYYYYYY"}
{"uid":”YYYYYYYYYYYYYYYYYYYYYYYYY","emailVerified":true,"expiryMills":1589841220328}
```

(Partly) JWT identification data – includes key info (kid, alg); issuer info (iss; iat); and others. Likely generated via Amazon Cognito.

Cookie string:

```
ep_iPlanetDirectoryPro=XXXXXXXXXXXXXXXXXXXXXXXXXXX.*AAJTSQACMDIAAlNLAXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXAAJTMQACMDQ.*
```

OpenAM token: `Session ID: YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY; Server ID: 02;`
`SFO Storage: YYYYYYYYYYYYYYYYYYYYYYYYYYYY; Server Instance: 04`

Identifies location, in-memory data, storage for a session authenticated via OpenAM

Fig. 4: Examples of sensitive cookies transmitted across domains

cookie policies; most likely, the exfiltration we identified are not performed maliciously, and the third parties receiving those simply ignore them. However, exfiltration of authentication material such as tokens beyond the boundary of the first-party website still constitutes a significant relaxation of security, and can simplify attacks. In this case, exfiltrated information is directed to third-parties which are also, at the same time, injecting content into the first-party website. Thus exfiltration can be construed as extending the first-party attack surface, and therefore we believe it constitutes a significant finding.

From the point of view of our study, such exfiltrations are also relevant because they may not be intentional. A number of websites set authentication/identity cookies with either the “Secure” or “HttpOnly” flag, or both (**7** out of **9** websites in both dataset). These flags respectively instruct the browser to only transmit the cookie over encrypted connection, and to negate access to the cookies from JavaScript code; their presence further suggest that these cookies are considered at least somewhat security sensitive by the website maintainers. Fine-grained data pertaining to authenticated sessions is not necessary for user tracking; as such, we suspect these cookies are broadcasted to third-parties as an unintended consequence of CNAME cloaking.

In terms of incentives, CNAME cloaking appears to be a feasible means for advertisers to evade blocklists when they have the cooperation of first parties. Ours and previous analyses [31], [19] suggests that evading blocklists is the main incentive for employing advertising-related CNAME cloaking. Furthermore, there is no apparent technical disincentive from employing CNAME cloaking.

C. *Possible Remediations*

As we showed in Sections III-E and VIII, no browser or security mechanism attempts to detect or prevent illegitimate use of CNAME records, nor prevent cookie sharing through CNAME records, including those browsers that implement recursive CNAME lookups for blocklists. Rather, tools simply appear to explore CNAMEs to find blocklist hits. This design decision is unsurprising, since preventing cookie sharing across CNAME subdomains would break functionality where the subdomain is a legitimate part of the main domain. This is a fundamental limitation of blocklists, which are agnostic to the *intent* of the website designer, and therefore cannot discriminate between legitimate and illegitimate cookie propagation. For the same reason, it is unclear that detecting misuse of CNAMEs is technically achievable without additional information from the first-party, though it may be an area for possible area for future work. However, enhancing any DNS-based content blocking mechanism with recursive CNAME lookups may be sufficient to end the widespread use of CNAME cloaking, since it will no longer be useful for cloaking advertising.

We also emphasize that a first-party that employs CNAME cloaking can still prevent cookie sharing with third-parties by creating a cookie policy that explicitly names the subdomains from which a cookie can be accessed. The cookie policy can thus allow the cloaked domain to be excluded from the cookie’s domains, without breaking other functionality. Unfortunately, explicitly naming subdomains is more onerous than using an implicit wildcard to allow access on all subdomains. An administrator that would agree to enable CNAME cloaking may be unlikely to expend extra effort in locking down cookie policies. In such cases, clients rely entirely on blocklists to prevent CNAME cloaking as a vector for (unintentional) cookie leaks.

D. Threats to Validity

A threat to *external validity* stems from only including the Alexa top-10,000 websites. As pageview distribution is understood to follow a heavy-tailed distribution [12], analysis of the top websites is enough to capture a significant portion of daily internet activity. Therefore, we do not believe including additional websites would return a clearer or more useful picture of the impact of cloaking on cookie policies.

In terms of *internal validity*, limiting bulk data collection to each website's homepage may cause false negatives, i.e. a conclusion that CNAME cloaking is not performed while in fact it is deployed on parts of the website other than the homepage. We consider this an acceptable limitation, which is compensated by the in-depth analysis of a select number of websites in Section VII. Furthermore, while this implies that our results may represent a lower bound for the prevalence of cookie exfiltration, the number of instances that we uncovered is sufficient to conclude that the phenomenon continues to be a problem.

E. Ethical Disclosure

Prior to submitting this paper, we contacted maintainers for the seven websites that were found to exfiltrate authentication cookies in our December 2020 dataset (ref. Table IV), and informed them of the issue.

X. Related Work

a) CNAME Cloaking: The risk of DNS aliases as an origin-obfuscation tool has been previously highlighted. Krishnamurthy and Wills identify *hidden third-parties* (CNAME-cloaked domains) as one of multiple means advertisers use to avoid blocklists [30], [29]. These studies are more than 10 years old and they do not characterize DNS redirections in-depth, as they focus chiefly on user privacy loss. Olejnik and Castelluccia [37] also identify a specific instance of cloaking, but do not investigate the phenomenon on a larger scale. More recently, Dao et al. characterize the landscape of advertising-related cloaking on the web [21]. Their work, which constitutes the inspiration for ours, investigate the prevalence of redirections but does not analyze their impact on cookie policies and exfiltration. Our work was also informed by a number of expert blog posts [31], [19] describing a recent rise in first-party redirections. Members of the community and commercial providers were quick to propose solutions for blocking suspicious cloaked domains [39], [2], [1], [3], [10] and build public blocklists [36]. However, we are concerned that without a comprehensive view of the phenomenon, those who deploy countermeasures risk to be "flying blind" and failing to identify all ramifications of the problem (a concern supported by our results).

b) Online Advertising and Tracking: Characterizing the impact of online advertising [15], [25] and tracking [18], [22], [29], [11], [43] has received much attention from the security and measurement communities. Most works find a continued and concerning impact of advertising on user privacy, elaborate and constantly evolving tracking technology, and rampant user data sharing among advertisers. Advertising and tracking also extend beyond web browsing. Razaghpanah et al. [41] identify thousands of advertising and tracking services. Andreou et al. [13] show similar widespread and invasive tracking and targeting on Facebook. Other work investigates the use of browser-based blocklists specifically to block online ads. Pujol et al. [40] find that use of ad-blocking technology is widespread, although it seems to be rooted more in annoyance towards ads than privacy concerns. Wills and Uzunoglu [45] investigate differences among ad-blockers. Gomer et al. [24] analyze the exposure of users to tracking cookies specifically for search. Englehardt et al. [23] show the potential for cookies to be used to violate user privacy in the context of tracking and mass surveillance. Despite this significant amount of prior work on online advertising, tracking, and ad-blockers, to the best of our knowledge the impact of advertising-related CNAME cloaking on cookie policies has not been previously investigated in detail.

c) DNS Security: DNS is an aging protocol, and efforts to improve its security have been slow and marred by deployment mistakes [32], [20], enabling various attacks [28], [27] and misuse by ISPs [44]. New technologies like DNS-over-HTTPS [34] have been proposed for remediation, but deployment has only recently begun. As CNAME redirections are part of the DNS infrastructure (and CNAME cloaking is not clearly distinguishable *prima facie*), none of the existing countermeasures affect it. Pearce et al. [38] investigate the role of DNS manipulations in internet censorship; while relevant, this work is orthogonal to advertising-related manipulations.

d) Website Security: Several large-scale analyses of website security focus on the presence of potentially vulnerable code (e.g., [42], [35]). Cavalier CNAME redirections (as the one uncovered in this paper) can enable unwanted access from third-party code to sensitive resources without the need of exploiting pre-existing vulnerabilities.

XI. Conclusions

In this paper, we presented the results of a large-scale analysis of the impact of advertising-related CNAME redirections on cookie propagation. Our data confirm that a non-negligible fraction of the Alexa-10,000 websites perform such redirections, consistent with prior studies. Furthermore, we show that in a number of sites, the deployment of first-party redirections causes sensitive cookies—which include personal information and/or authentication data—to leak to third-party advertising domains. While exfiltration of these cookies is likely unintentional, it is also problematic as it facilitates impersonation and takeover attacks. We also find that the ability of blocking these exfiltrations vary between browsers.

Overall, these results show that CNAME cloaking has serious ramifications, affecting cookie propagation in unexpected ways. A combination of more sophisticated browser-based blocking strategies, and more fine-grained website policies, may be necessary to tackle the problem.

References

[1] "Address 1st-party tracker blocking · Issue #780 · uBlockOrigin/uBlock-issues," 2019. [Online]. Available: https://github.com/uBlockOrigin/uBlock-issues/issues/780

[2] “Apply Pi-Hole blocking to CNAMEs,” Nov. 2019. [Online]. Available: https://discourse.pi-hole.net/t/apply-pi-hole-blocking-to-cnames/25445/95

[3] “Cname cloaking and bounce tracking defense — webkit,” 2020. [Online]. Available: https://webkit.org/blog/11338/cname-cloaking-and-bounce-tracking-defense/

[4] “easylist/easylist/ at master,” May 2020. [Online]. Available: https://github.com/easylist/easylist/tree/master/easylist

[5] “Faq: Cookies in am/openam,” 2020. [Online]. Available: https://backstage.forgerock.com/knowledge/kb/article/a19829000

[6] “https://easylist.to/easylist/easyprivacy.txt,” May 2020. [Online]. Available: https://easylist.to/easylist/easyprivacy.txt

[7] “mitmproxy - an interactive https proxy,” May 2020. [Online]. Available: https://mitmproxy.org

[8] “Openam - open access manager,” 2020. [Online]. Available: https://www.openidentityplatform.org/openam

[9] “Seleniumhq browser automation,” May 2020. [Online]. Available: https://www.selenium.dev

[10] “What’s brave done for my privacy lately? episode #6: Fighting cname trickery,” 2020. [Online]. Available: https://brave.com/privacy-updates-6/

[11] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, “The web never forgets: Persistent tracking mechanisms in the wild,” in *CCS*, 2014.

[12] L. A. Adamic and B. A. Huberman, “Zipf’s law and the internet.” *Glottometrics*, vol. 3, no. 1, pp. 143–150, 2002.

[13] A. Andreou, M. Silva, F. Benevenuto, O. Goga, P. Loiseau, and A. Mislove, “Measuring the facebook advertising ecosystem,” in *NDSS*, 2019.

[14] Apple, “Tracking prevention in webkit,” Nov. 2020. [Online]. Available: https://webkit.org/tracking-prevention/#intelligent-tracking-prevention-itp

[15] P. Barford, I. Canadi, D. Krushevskaja, Q. Ma, and S. Muthukrishnan, “Adscape: Harvesting and analyzing online display ads,” in *WWW*, 2014.

[16] D. Bradbury, “Ad-blocking companies block ’unblockable’ tracker,” Nov. 2019. [Online]. Available: https://nakedsecurity.sophos.com/2019/11/25/ad-blocking-companies-block-unblockable-tracker/

[17] Brave, “What’s brave done for my privacy lately? episode #6: Fighting cname trickery,” Oct. 2020. [Online]. Available: https://brave.com/privacy-updates-6/

[18] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris, “I always feel like somebody’s watching me: measuring online behavioural advertising,” in *CoNext*, 2015.

[19] R. Cointepas, “CNAME Cloaking, the dangerous disguise of third-party trackers,” Dec. 2019. [Online]. Available: https://medium.com/nextdns/cname-cloaking-the-dangerous-disguise-of-third-party-trackers-\ \195205dc522a

[20] T. Dai, H. Shulman, and M. Waidner, “Dnssec misconfigurations in popular domains,” in *International Conference on Cryptology and Network Security*. Springer, 2016, pp. 651–660.

[21] H. Dao, J. Mazel, and K. Fukuda, “Characterizing CNAME Cloaking-Based Tracking on the Web,” in *TMA*, 2020.

[22] S. Englehardt and A. Narayanan, “Online tracking: A 1-million-site measurement and analysis,” in *CCS*, 2016.

[23] S. Englehardt, D. Reisman, C. Eubank, P. Zimmerman, J. Mayer, A. Narayanan, and E. W. Felten, “Cookies that give you away: The surveillance implications of web tracking,” in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 289–299.

[24] R. Gomer, E. M. Rodrigues, N. Milic-Frayling, and M. Schraefel, “Network analysis of third party tracking: User exposure to tracking cookies through search,” in *2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, vol. 1. IEEE, 2013, pp. 549–556.

[25] S. Guha, B. Cheng, and P. Francis, “Challenges in measuring online advertising systems,” in *IMC*, 2010.

[26] M. Jones, J. Bradley, and N. Sakimura, “RFC 7519 - JSON Web Token (JWT),” May 2015. [Online]. Available: https://tools.ietf.org/html/rfc7519

[27] A. Klein and B. Pinkas, “Dns cache-based user tracking,” in *NDSS*, 2019.

[28] A. Klein, H. Shulman, and M. Waidner, “Internet-Wide Study of DNS Cache Injections,” in *INFOCOM*, 2017.

[29] B. Krishnamurthy and C. Wills, “Privacy diffusion on the web: a longitudinal perspective,” in *WWW*, 2009.

[30] B. Krishnamurthy and C. E. Wills, “Generating a privacy footprint on the internet,” in *IMC*, 2006.

[31] J. Leyden, “Web trackers using CNAME Cloaking to bypass browsers’ ad blockers,” Nov. 2019. [Online]. Available: https://portswigger.net/daily-swig/web-trackers-using-cname-cloaking-to-bypass-browsers-ad-blockers

[32] W. Lian, E. Rescorla, H. Shacham, and S. Savage, “Measuring the practical impact of DNSSEC deployment,” in *USENIX Security Symposium*, 2013.

[33] J. R. Mayer and J. C. Mitchell, “Third-party web tracking: Policy and technology,” in *2012 IEEE Symposium on Security and Privacy*. IEEE, 2012, pp. 413–427.

[34] P. McManus and P. Hoffman, “RFC 8484 - DNS Queries over HTTPS (DoH),” Oct. 2018. [Online]. Available: https://tools.ietf.org/html/rfc8484

[35] M. Musch, C. Wressnegger, M. Johns, and K. Rieck, “New kid on the web: A study on the prevalence of webassembly in the wild,” in *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*. Springer, 2019, pp. 23–42.

[36] nextdns, “nextdns/cname-cloaking-blocklist,” May 2020. [Online]. Available: https://github.com/nextdns/cname-cloaking-blocklist

[37] L. Olejnik and C. Castelluccia, “Analysis of openx-publishers cooperation,” in *HotPETs*, 2014.

[38] P. Pearce, B. Jones, F. Li, R. Ensafi, N. Feamster, N. Weaver, and V. Paxson, “Global Measurement of DNS Manipulation,” in *USENIX Security Symposium*, 2017.

[39] O. Poitrey, “NextDNS added CNAME Uncloaking support, becomes the first cross-platform solution to the problem,” Nov. 2019. [Online]. Available: https://medium.com/nextdns/nextdns-added-cname-uncloaking-support-becomes-the-first-cross-\ \platform-solution-to-the-problem-e3f437f84342

[40] E. Pujol, O. Hohlfeld, and A. Feldmann, “Annoyed users: Ads and ad-block usage in the wild,” in *Proceedings of the 2015 Internet Measurement Conference*, 2015, pp. 93–106.

[41] A. Razaghpanah, R. Nithyanand, N. Vallina-Rodriguez, S. Sundaresan, M. Allman, C. Kreibich, and P. Gill, “Apps, trackers, privacy, and regulators,” in *NDSS*, 2018.

[42] G. Richards, C. Hammer, B. Burg, and J. Vitek, “The eval that men do,” in *ECOOP*, 2011.

[43] F. Roesner, T. Kohno, and D. Wetherall, “Detecting and defending against third-party tracking on the web,” in *NSDI*, 2012.

[44] N. Weaver, C. Kreibich, and V. Paxson, “Redirecting dns for ads and profit.” in *FOCI*, 2011.

[45] C. E. Wills and D. C. Uzunoglu, “What ad blockers are (and are not) doing,” in *IEEE HotWeb*, 2016.