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

NDSS 2021

Tongwei Ren, Alexander Wittman, Lorenzo De Carli, Drew Davidson

Worcester Polytechnic Institute, University of Kansas

2022-08-17

JaeHyun Lee (jhlee2021@mmlab.snu.ac.kr)

Contents

Motivation

❖ Recently advertisers / trackers get into misuse of CNAMEs to bypass blocklists and privacy policies

✓ Supplement the lack of analysis of the effect of CNAME cloaking on browser cookie policies

✓ Understand the actual effect of existing mitigations against CNAME cloaking on cookie exfiltration

Main Contributions

- Perform a large-scale analysis of the impact of advertising-related CNAME redirections on cookie propagation (Alexa Top-10000 sites)

- Find that in a number of sites, the deployment of 1st party redirections cause sensitive cookies to leak to 3rd party advertising domains

CNAME Redirection

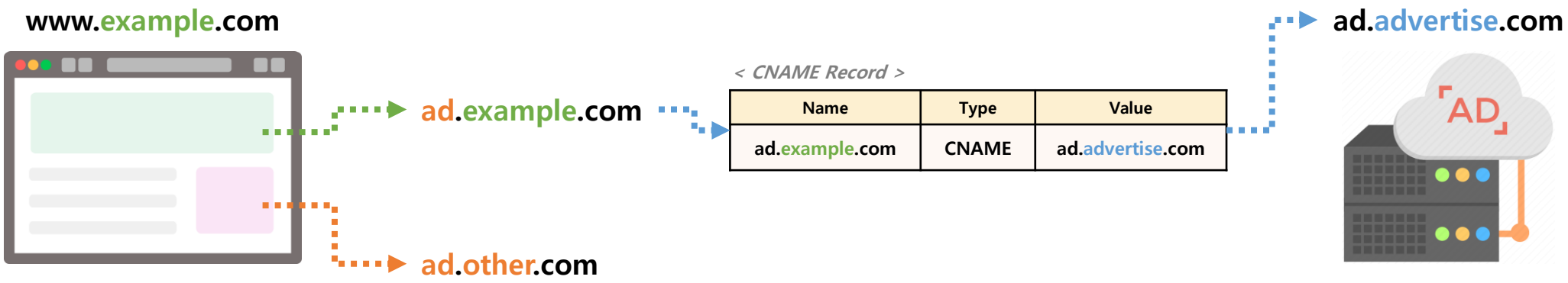
- A **CNAME** (Canonical Name) record is a type of resource record in DNS that maps one domain name to another
  - ✓ CNAME redirection is also used for content delivery network (CDN)
    - Browsers identify and trust CDN content as coming from host domain itself
  - ✓ However, CNAME records can also be used for some **malicious purposes**

< DNS Records >

Name	Type	Value
blog.example.com	CNAME	example.com
www.example.com	CNAME	example.com
example.com	A	128.0.4.80

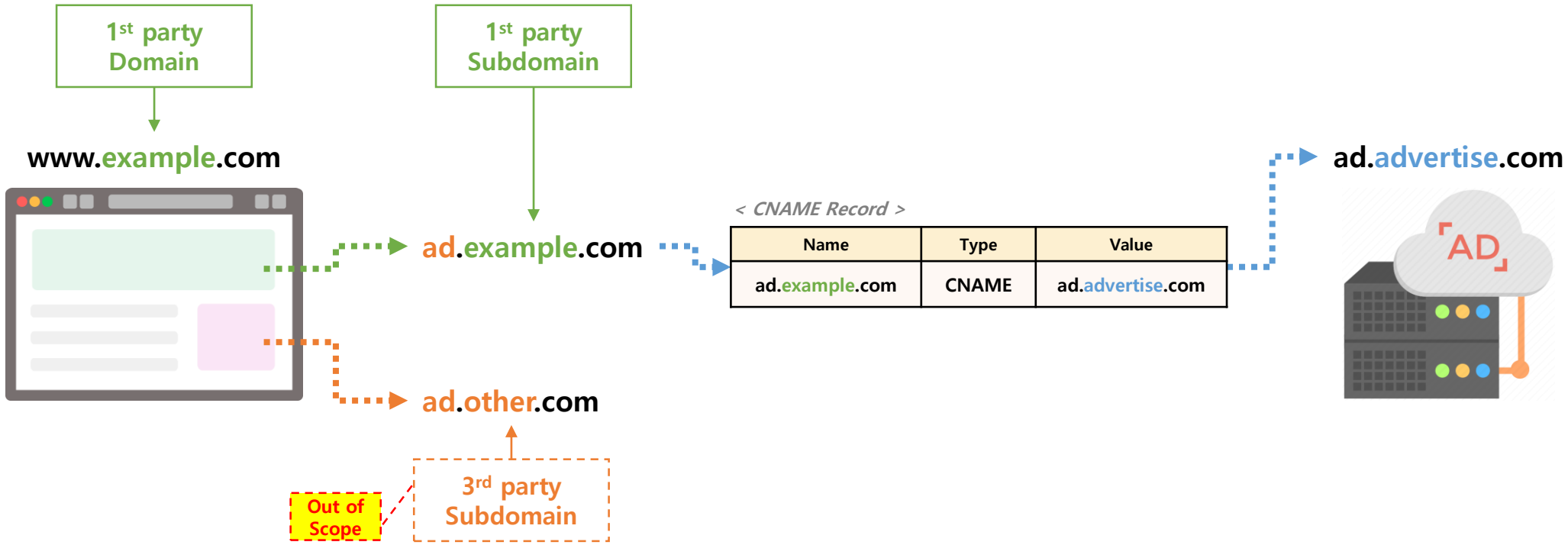
What is a CNAME Cloaking?

- By using ***CNAME redirections***, a 3rd party domain gets cloaked as a subdomain of a 1st party or trusted 3rd party
  - ➢ Same powers as the true 1st party or trusted 3rd party



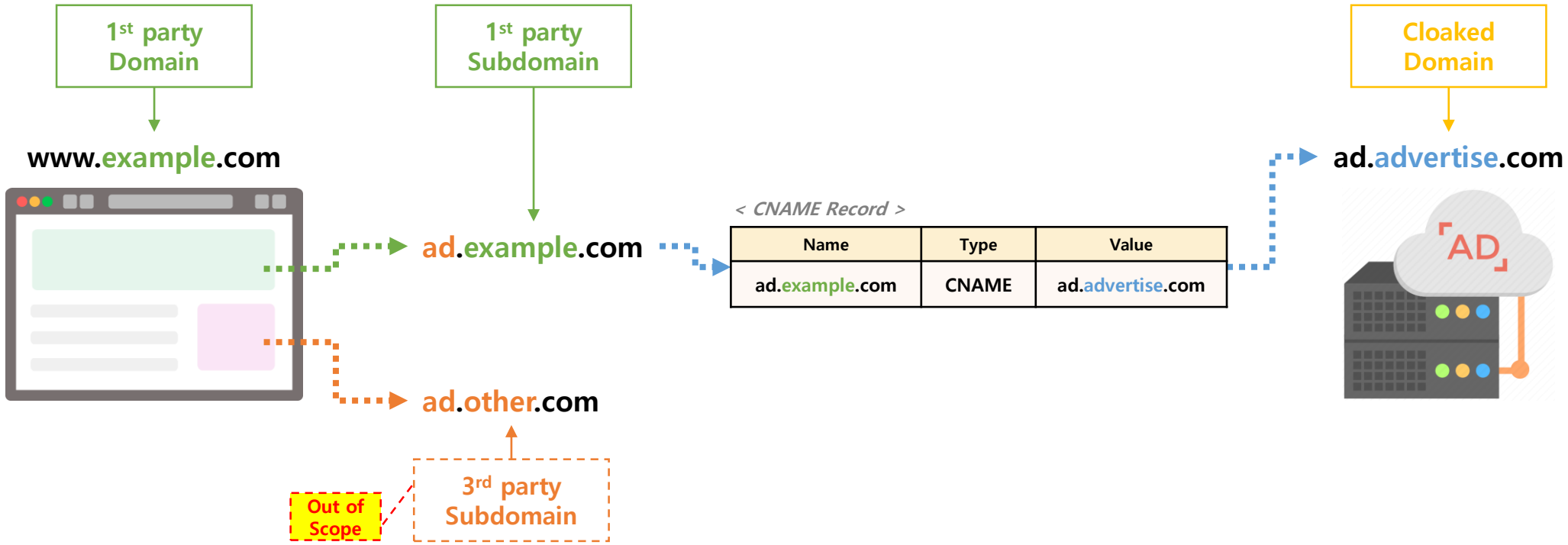
What is a CNAME Cloaking?

- By using ***CNAME redirections***, a 3rd party domain gets cloaked as a subdomain of a 1st party or trusted 3rd party
  - ➢ Same powers as the true 1st party or trusted 3rd party



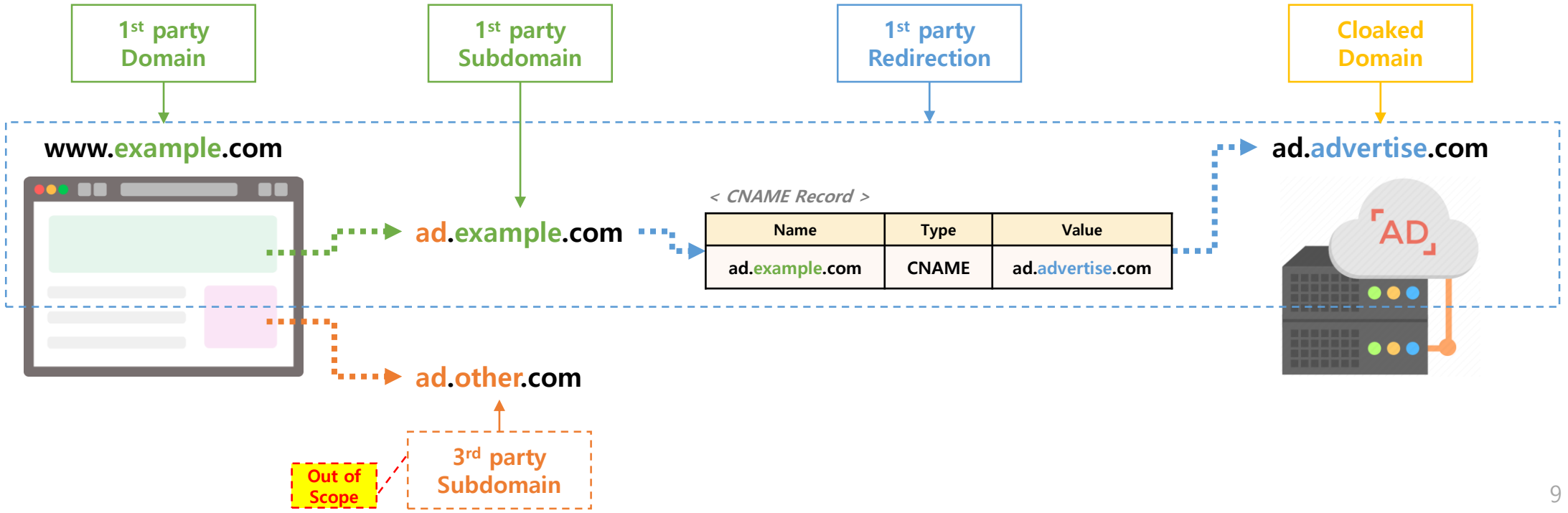
What is a CNAME Cloaking?

- By using ***CNAME redirections***, a 3rd party domain gets cloaked as a subdomain of a 1st party or trusted 3rd party
  - ➢ Same powers as the true 1st party or trusted 3rd party



What is a CNAME Cloaking?

- By using ***CNAME redirections***, a 3rd party domain gets cloaked as a subdomain of a 1st party or trusted 3rd party
  - ➢ Same powers as the true 1st party or trusted 3rd party



Browser Cookie Policy

- Cookie ?
  - One of the most straightforward ways to maintain user identities on the web
  - Text-based key-value pairs that are managed by the browser
  - Popular for authentication and/or user-tracking

- **Same-Origin Policy** (SOP)
  - To ensure that data is not leaked through cookies
  - Browser can prohibit access to a cookie from other origins

➢ However, **CNAME Cloaking** may obscure the true origin of a web request for a resource such as cookie

Existing Mitigations

- AdBlockers
  - Rely on manually-curated blocklist

- Browser protection
  - Brave / Safari: Proposed solution (recently on 2020)
    - ✓ But need re-implementation (to update blocklist)
  - Chrome / Firefox / Edge: No CNAME defenses at all

Methodology

- Analysis steps
  1) Data Collection
  2) Domain Classification
  3) Cookie Lifecycle Analysis
  4) Manual Cookie Analysis
  5) Browser Blocklist Evaluation

- Test environments
  - Custom crawler/logger on Firefox browser
  - Selenium[1] / mitmproxy[2] / dnspython[3]
  - Singularity container with mongoDB
  - Tested on June and December 2020

[1] https://www.selenium.dev/ Web testing framework,
[2] https://mitmproxy.org/ HTTPS proxy
[3] https://www.dnspython.org/ DNS toolkit for Python

Data Collection

❖ *Gather main dataset*

- Logging Redirections
  - All the requests, responses and DNS resolution chain for each request
  - Target: Alexa Top-10000 list
- Isolating Candidate Redirections
  - Redirections that are likely to be ad/tracking related
  - Two approaches
    - ✓ Domain-based: based on the popular blocklists used by commercial ad-blockers
    - ✓ URL-based: based on string pattern on URL such as "ad", "track", etc.

Data Collection - Result

	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

Around 400 sites

< **Summary of main dataset** >

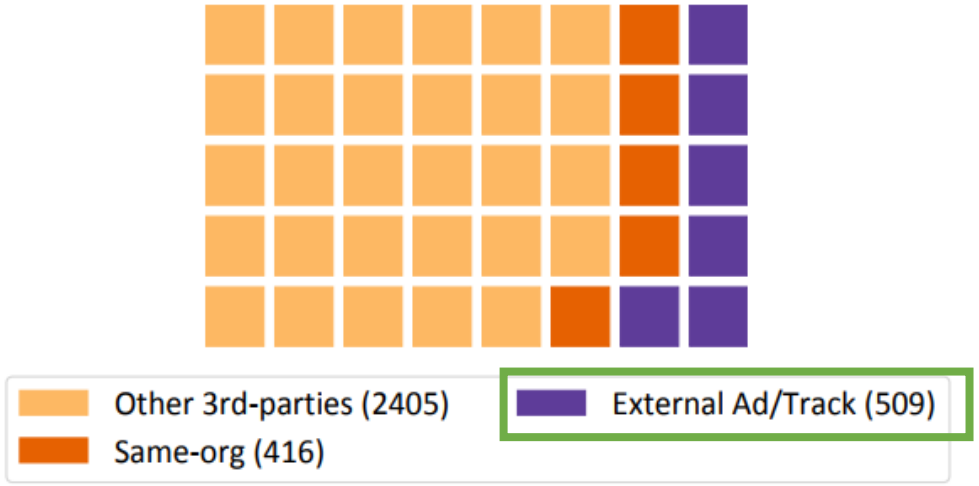
- 1st party redirection
  - More than 4% of the total websites
- Ad/Tracking related candidates
  - Approximately **19%** of all 1st party redirections

Domain Classification

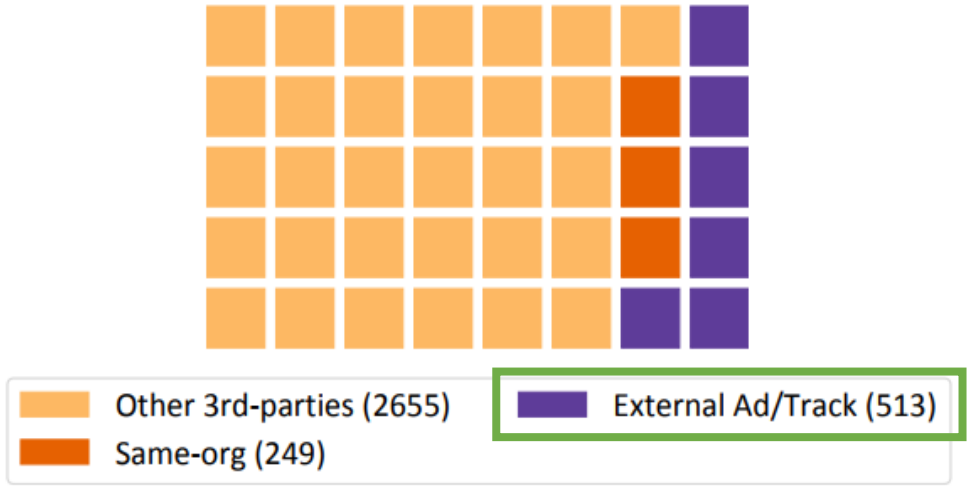
❖ *Obtain accurate and comprehensive information about destination domains of filtered redirections*

- Through manual investigation of source and destination domains of each redirection, divide all 1st party redirections into three categories
  - Same-organization
    - ✓ Source and destination domain belong to the same organization (ex. msn and Microsoft)
  - **External ad/tracking**
    - ✓ Destination domain belongs to an ad/tracker. (ex. ads.google.com)
  - Other 3rd parties
    - ✓ Not under either Same-organization or External ad/tracking.

Domain Classification - Result



(a) June 2020



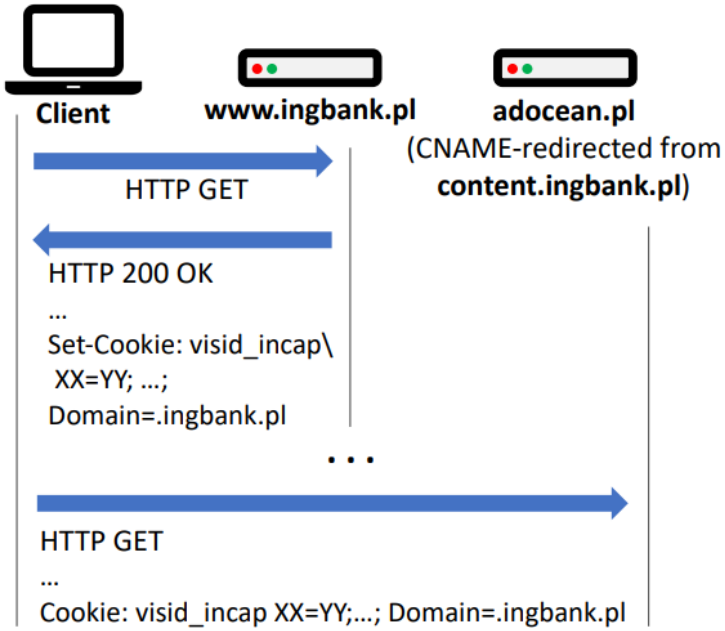
(b) December 2020

- External Ad/Tracking
  - Approximately **15%** of all filtered redirections

Cookie Lifecycle Analysis

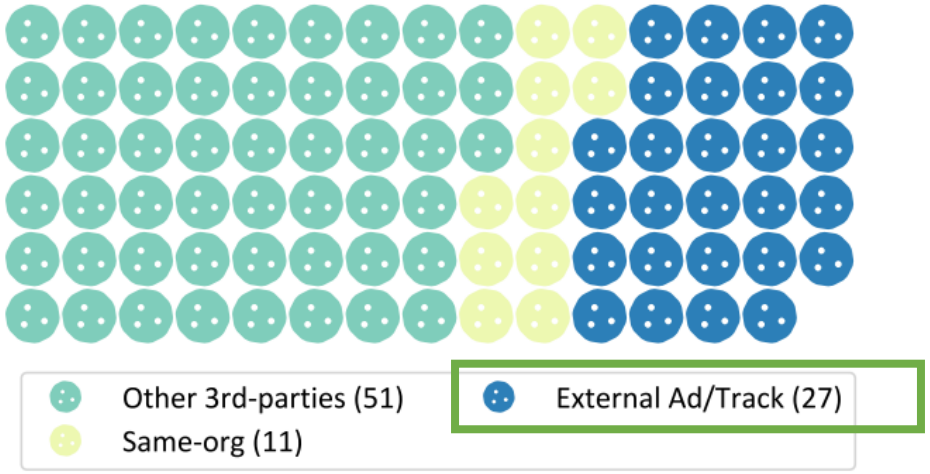
❖ *Learn more about how frequently cookies are transmitted to External Ad/Tracking 3rd party domains*

- **Cross-domain Cookie Transmission**
  - Observed many times on this experiment (unclear)
  - Cookies may be set by the real 1st party domain, but sent to the cloaked 3rd party domain

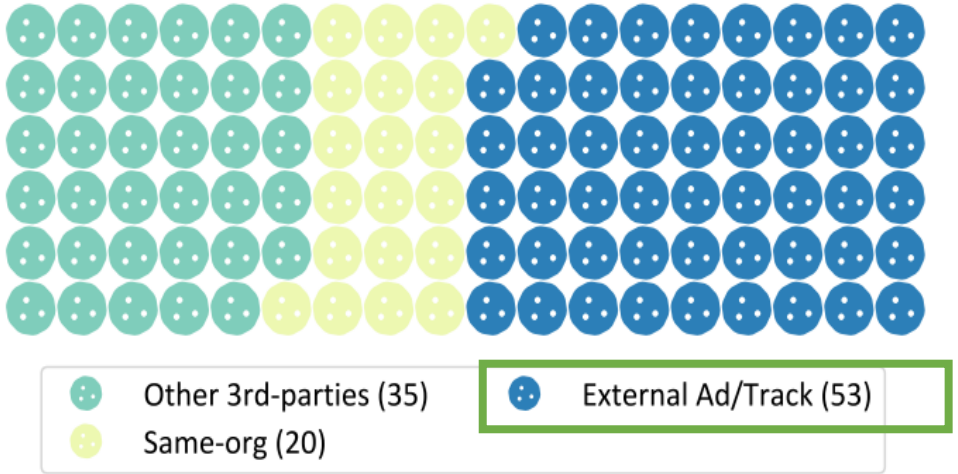


< extracted example of cookie transmission>

Cookie Lifecycle Analysis - Result



(a) June 2020



(b) December 2020

- In total,
  - **89** (June) and **108** (December) cookies are identified as 'cross-domain transmission'
- Among them,
  - **27** (June) and **53** (December) cookies belong to 'External Ad/Track' category

Manual Cookie Analysis

❖ *Gain insight on inter-domain cookie transmission on actual website*

- Methodology
  - Randomly select 62 websites (External Ad/Tracking)
  - Create user accounts and record authenticated browsing sessions
  - Analyze each cookie to determine whether it is a sensitive cookie

  - **Sensitive cookies**
    - ✓ Information cookie – contains one or more of the user data such as name, email, etc.
    - ✓ Authentication cookie – causes the website to ask users to re-auth after deleting it
    - ✓ Identity cookie – causes the user to log back in without re-entering PW after deleting it

Manual Cookie Analysis - Result

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

14

46

< **A**: Authentication cookie; **I**: Information cookie; **D**: iDentity cookie >

- Sensitive cookies that exfiltrated to 3rd parties
  - ➢ **Actual data (46 cookies in 14 sites) were found in the wild**

Browser Blocklist Evaluation and Result

❖ *Evaluate 'Safari' and 'Brave' browsers since they have explicitly announced their ability to prevent CNAME cloaking*

- Visit and log in to the 7 websites having cross-domain authentication cookie transmission from the previous experiment

- Results
  - ✓Safari: 2 out of 7 instances of exfiltration were blocked
  - ✓Brave: 6 out of 7 instances of exfiltration were blocked

  ➢ **Impressive but not perfect**

Summary

- A non-negligible fraction of the Alexa Top-10000 websites perform CNAME redirection (more than 4%)
- Many sites exfiltrate cookies to 3rd party ad/tracking domain on their homepage
- Sensitive cookies are exfiltrated to 3rd parties beyond the homepage (totally 46 cookies in 14 of 62 websites)
- The ability of blocking these exfiltration vary between browsers

Conclusion – Security Implications

- CNAME cloaking has undesirable implications for user security and privacy
- CNAME cloaking appears to be a feasible means for advertisers to evade blocklists when they have the cooperation of 1st parties
- 3rd parties and 1st parties are willing to collaborate in ways that blur origin-based security
- The exfiltration of authentication cookies may open the door to impersonation and account takeover, extend the 1st party attack surface

Thank you

Appendix