

PURL



Safe and Effective Sanitization of Link **Decoration**

Shoor Munir, Patrick Lee, Umar Iqbal,
Zubair Shafiq, Sandra Siby

Agenda

01

Background

What is a link decoration?

02

Measurement and analysis

How prevalent is use of link decoration by trackers?

03

Results and conclusions

Results of countermeasure and deployment statistics

What is a link decoration?

<https://www.theverge.com/2024/7/22/24203893/google-cookie-tracking?fbclid=IY2x...&showComments=1#submit>

What is a link decoration?

<https://> → Scheme

www.theverge.com → FQDN

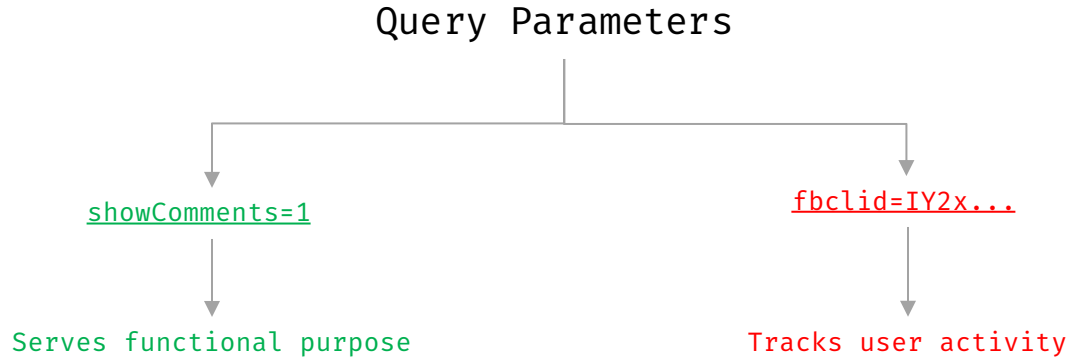
[/2024/7/22/24203893/google-cookie-tracking](https://www.theverge.com/2024/7/22/24203893/google-cookie-tracking) → Resource Path

[?fbclid=IY2x...&showComments=1](https://www.theverge.com/2024/7/22/24203893/google-cookie-tracking?fbclid=IY2x...&showComments=1) → Query Parameters

[#submit](https://www.theverge.com/2024/7/22/24203893/google-cookie-tracking?fbclid=IY2x...&showComments=1#submit) → Fragment

Link decorations

What is a link decoration?



Why are link decorations important?

Why are link decorations important?

Ongoing crackdown against existing mechanisms of online tracking



Third-party cookies have been phased out by major browsers such as Safari and Firefox



Advances towards limiting browser fingerprinting by Google and other stake holders

Has this solved online tracking?

Has this solved online tracking?



**Personal
Information**

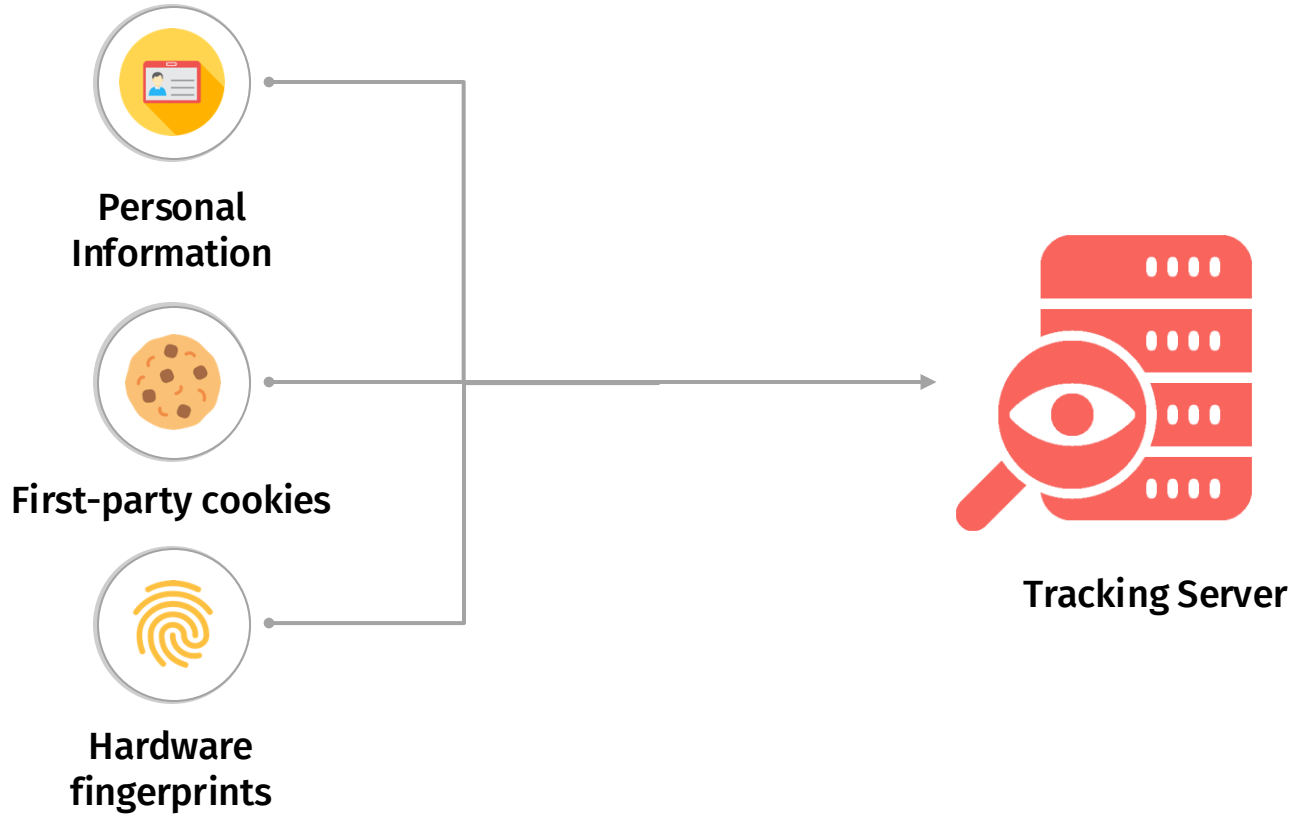


First-party cookies

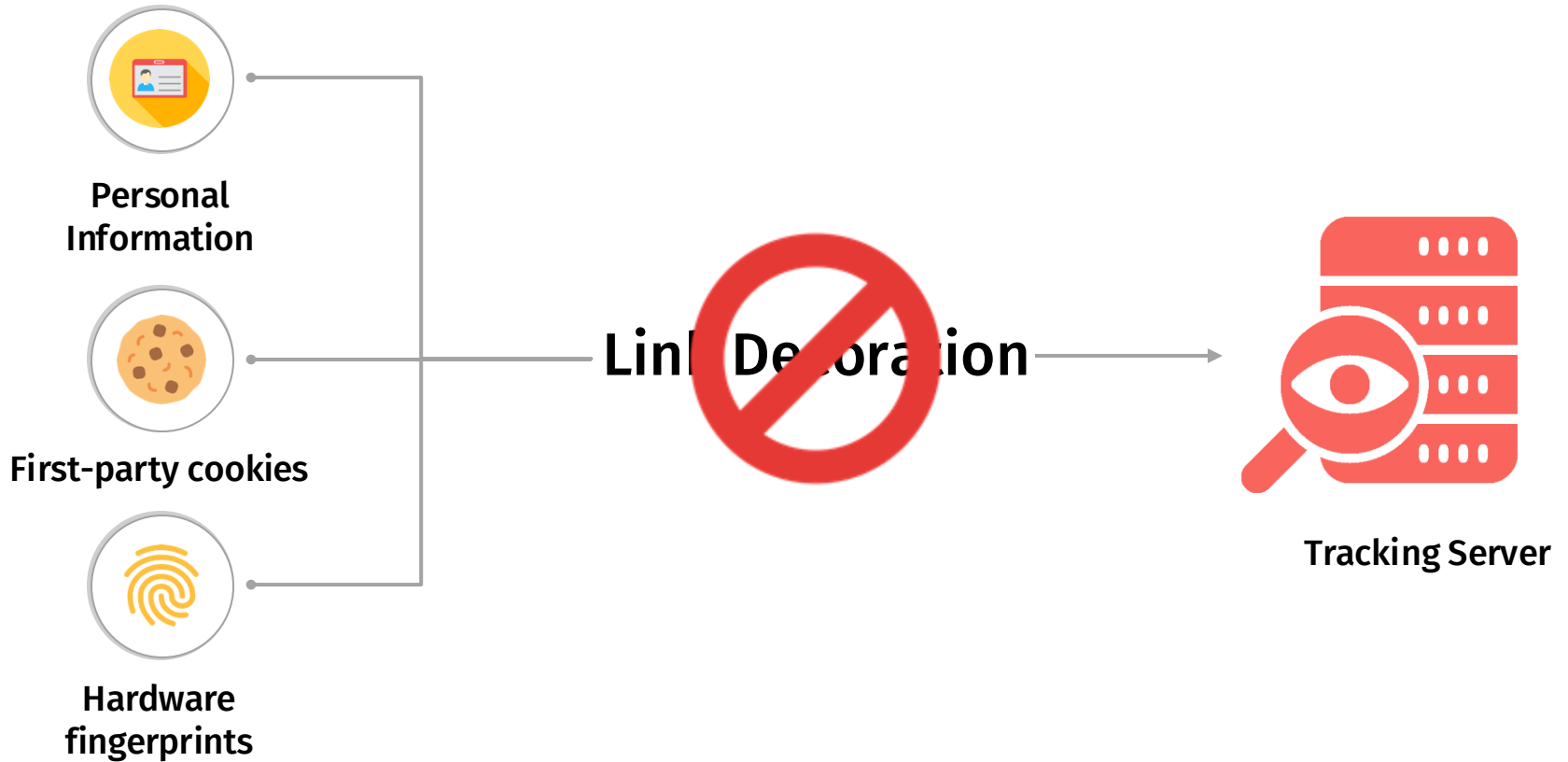


**Hardware
fingerprints**

Has this solved online tracking?



Has this solved online tracking?



Agenda

01

Background

What is link decoration?

02

Measurement and Analysis

How prevalent is use of link decoration by trackers?

03

Results and Conclusions

Results of countermeasure and deployment statistics

Measurement

- Collected data to detect prevalence of tracking link decorations
- Crawled 20,000 websites sampled from top-1M¹ websites using OpenWPM

1. We use tranco list (<https://tranco-list.eu>) for site rankings

Excessive use of link decoration by trackers

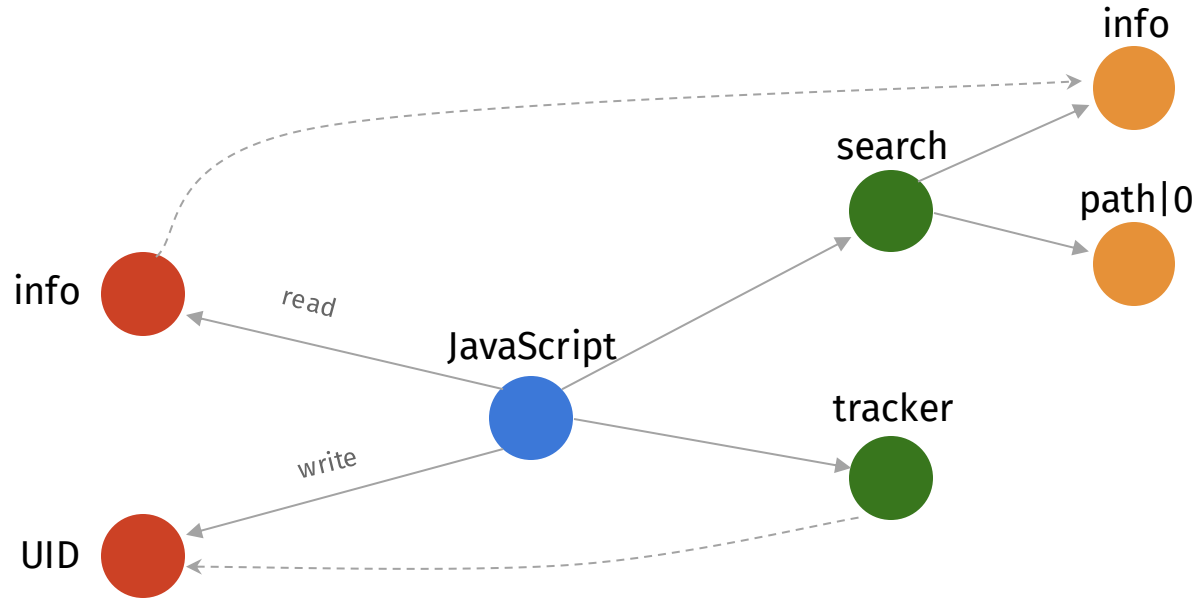
- 45.5% of all link decorations are sent to known advertisers and trackers
- 99.53% of requests containing tracking link decorations are mixed
- Link decorations such as “v” query parameter used by Google were also found to be used for tracking in one context and non-tracking in other
 - In some instances identifies a video
 - In some instances identifies the version

Detecting tracking link decorations

Key behaviors of tracking link decorations are:

- They share values stored in browser storage
- They are sent to endpoints which are involved in infiltration of user identifiers
- They co-exist in URLs with multiple other link decorations
- They are usually strings which can contain more information about user (higher entropy strings)

Graph representation

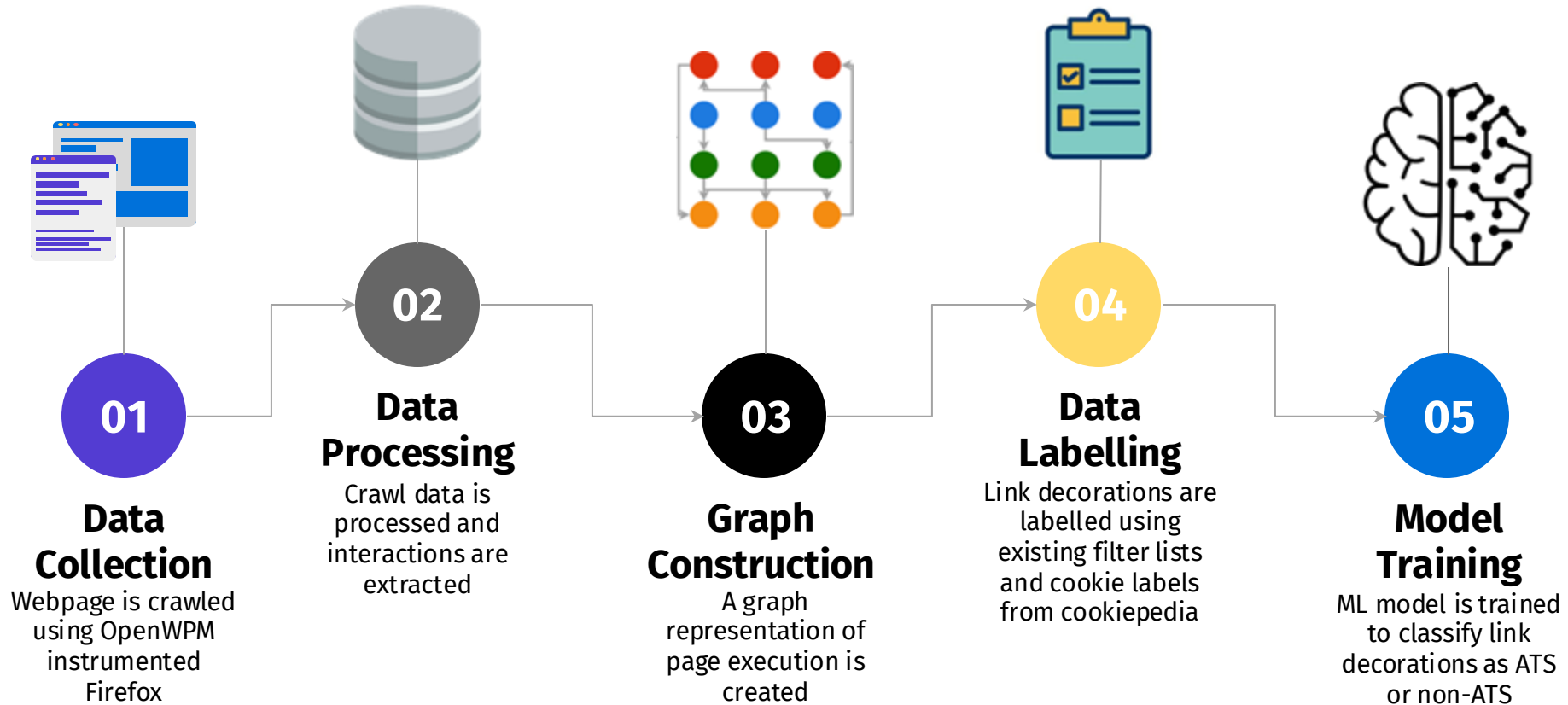


Data Labelling and Model Training

- Data labelled through a combination of filter lists and Cookiepedia
 - **Tracking** labels from
 - Filter lists maintained by Ad Guard (1,000)¹, uBlock Origin (46)², Brave (47)³, Firefox (23)⁴, and Safari (24)⁵
 - Tracking cookies exfiltrated using link decorations
 - **Non-tracking** labels to requests sent to endpoints not in any filter lists
- Model classifies given link decoration as either **tracking** (ATS) or **non-tracking** (non-ATS)

1. https://raw.githubusercontent.com/AdguardTeam/AdguardFilters/master/TrackParamFilter/sections/general_url.txt
2. <https://web.archive.org/web/20230802013435/https://github.com/uBlockOrigin/uAssets/tree/master/filters>
3. https://web.archive.org/web/20230731224815/https://raw.githubusercontent.com/brave/brave-core/f52868910c4d3a1a076644335d9b94916d72f761/browser/net/brave_query_filter.cc
4. <https://web.archive.org/web/20230731221803/https://privacystests.org/>

PURL Pipeline



Agenda

01

Background

What is Link
Decoration?

02

Measurement and Analysis

Are trackers using
link decoration? If
yes, then how?

03

Results and Conclusions

Results of
countermeasure
and deployment
statistics

Performance Comparison

- We compare PURL's performance with four other countermeasures:
 - CrumbCruncher
 - Cookiepedia
 - Request based filter lists (EasyList, EasyPrivacy)
 - Link decoration specific filter lists (Ad Guard, uBlock Origin, Brave, Safari, and Firefox)

Accuracy

Classifier	Accuracy	Precision	Recall
PURL	98.74%	98.62%	98.87%
CrumbCruncher	50.16%	59.09%	10.67%
Filter lists (Requests)	92.31%	90.91%	94.04%
Filter lists (Decorations)	50.50%	100.0%	10.15%

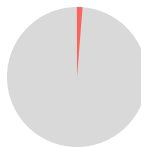
Breakage Comparison

- We compare breakage caused by all countermeasures across 100 different sites
- Two reviewers manually visit each webpage and compare website performance for each countermeasure

PURL

Major Breakage on only 1 site

1%



Filter lists (Requests)

Major breakage on 9 sites

9%



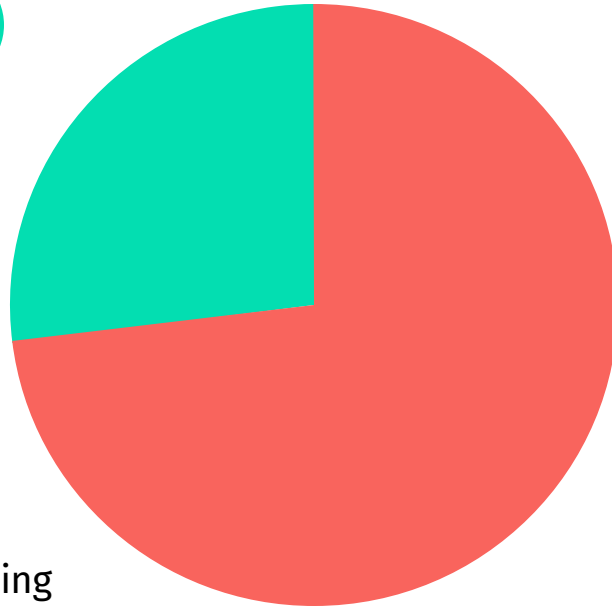
Deploying PURL in the wild

Deploying PURL in the wild

Non-Tracking

These sites send no tracking link decorations

27%



Tracking

These sites send at least one tracking link decoration

73%

Percentage of sites* using tracking link decorations

* Analysis done on 20,000 websites

93.3%

Of all tracking link decorations are query parameters, while 6.6% are resource paths

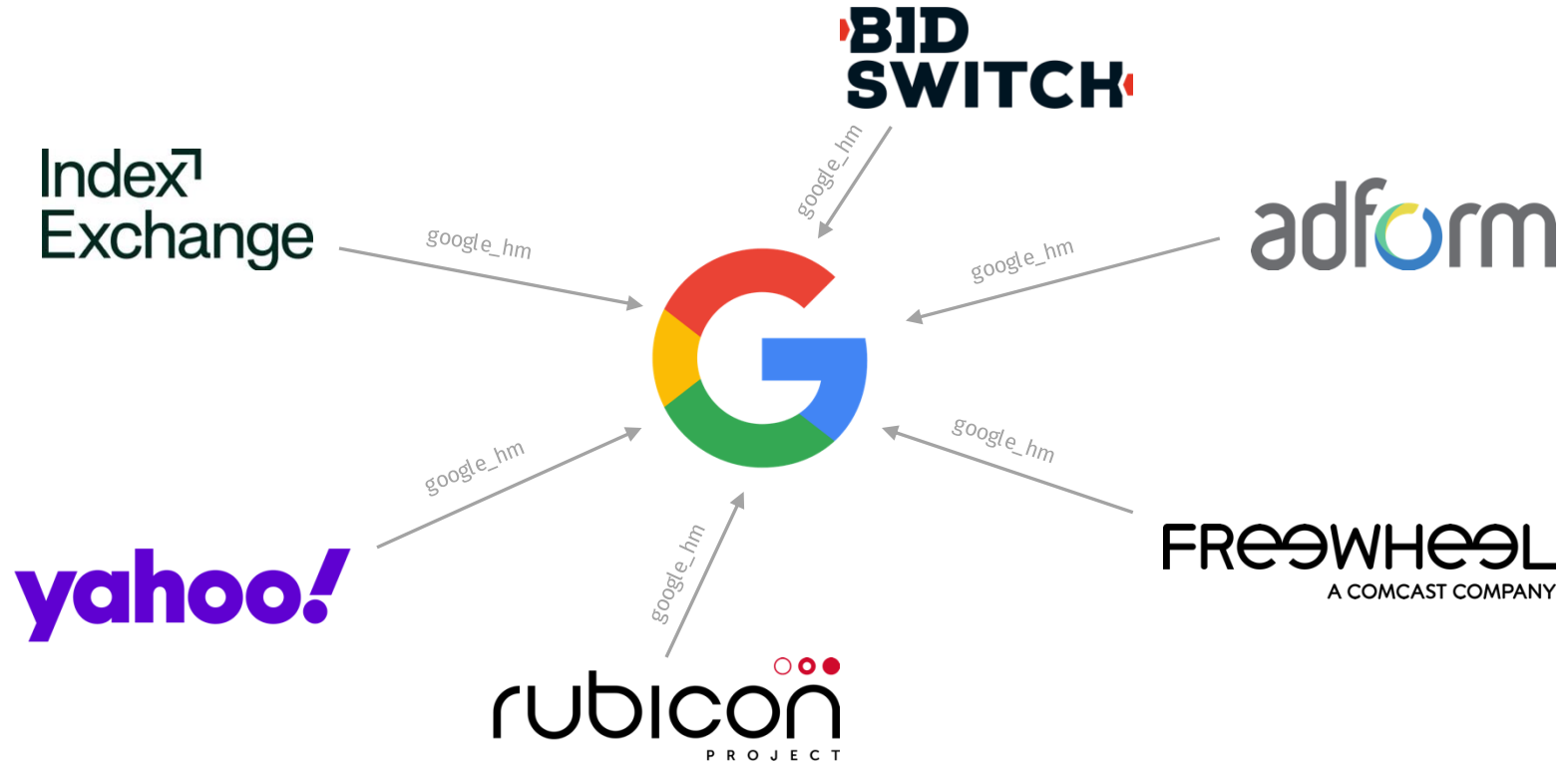
69.4%

Of tested sites contain instances where tracking storage values (first-party cookies and local storage) are shared through link decorations

Distribution of personal data through link decoration

- PURL identified 62 link decorations being used to send out user email addresses to third-parties
 - ttd_id (Feathr)
 - pmUserId (Rich Audience)
- 200 tracking link decorations are sent by scripts involved in fingerprinting
 - ippd (DoubleClick)
 - tid (Rambler)

Distribution and Sharing of tracking link decorations



Conclusion

Conclusion

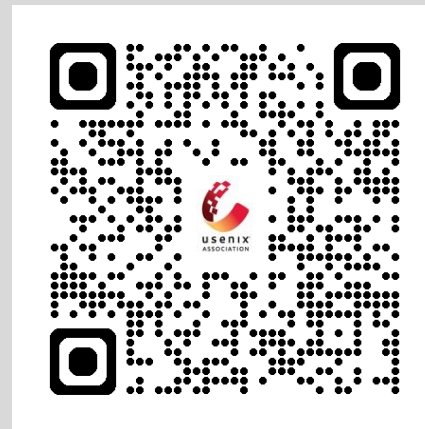
- PURL provides as a safe and effective mechanism to block exfiltration of user information while ensuring website functionality remains intact
- A number of widely used query parameters such as `google_hm`, `gpic`, `ga_sid`, `ttd_puid` are not currently included in existing filter lists
- 1,100 link decorations currently included, PURL identifies 196,890

1. <https://github.com/purl-sanitizer/purl/tree/main/data>

Thank You!

Reach me through email at:

smunir@ucdavis.edu



Scan to read the paper or visit:

<https://www.usenix.org/conference/usenixsecurity24/presentation/munir>