



HAL
open science

Client-Side GTM Tags: hidden data leaks and potential violations of GDPR

Gilles Mertens, Nataliia Bielova, Vincent Roca, Cristiana Santos

► To cite this version:

Gilles Mertens, Nataliia Bielova, Vincent Roca, Cristiana Santos. Client-Side GTM Tags: hidden data leaks and potential violations of GDPR. PETS 2024 - 24th Privacy Enhancing Technologies Symposium, Jul 2024, Bristol, United Kingdom. pp.1-1, 2024. hal-04699754

HAL Id: hal-04699754

<https://hal.science/hal-04699754v1>

Submitted on 17 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Client-Side GTM Tags: hidden data leaks and potential violations of GDPR

Background

Websites collect their visitors' data for various purposes, such as targeted advertising, and rely on third parties for such collection. One option for Website Publishers to **share visitor data** with these 3rd parties is to install Tags they provide and use **Google Tag Manager (GTM)** to manage them. Until now research considers GTM as a single tracking script, our work casts light on the internals of GTM.

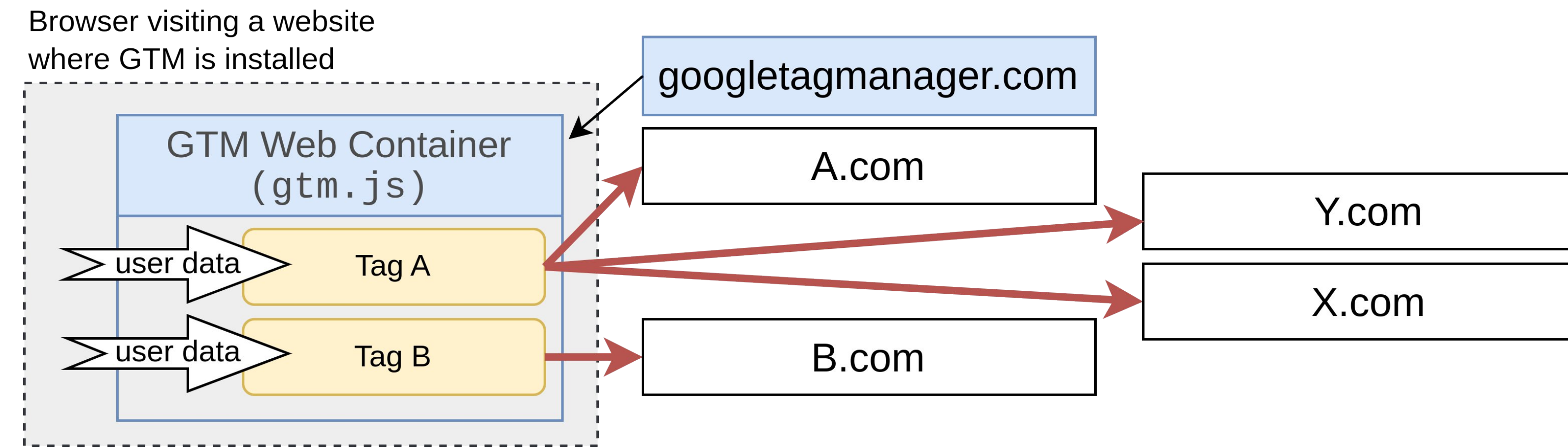
Methods

We study Tags from the **point of view of the Publisher** on an empty website. This approach enables us to install tags in **an environment we fully control**.

- **In-depth analysis.** We analyze 6 tags in-depth by hand to list types of data collected from: technical and legal documentation, tag dashboard and network traffic.
- **Large-scale study.** We develop a program that automatically installs 719 Tags and collects their traffic on our website. For the 352 Tags making network requests, we study both the data collected and with whom it is shared.
- **Policy analysis.** With a legal scholar (Cristiana Santos), we study the privacy policies of the 3rd party providing the tag. We especially focus on the clarity of the reference to the tag, the disclosure of third parties collecting data and use of cookies.
- **Consent mode.** Finally we study Google Consent mode, a feature that enables tags to be aware of user consent: we install the Google Tag on our website and visit it twice, refusing and accepting consent, to assess potential differences in data collection practices.

Findings

Tags collect information about the website visited, device features and use first- and third-party cookies. We found that **49 tags share data with multiple companies**, with one tag contacting 26 of them! Analysis of privacy policies shows that **Tag providers often omit to declare with whom the data is shared** and the cookies used. Moreover, privacy policies studied never mention tags explicitly, leaving Publisher to guess whether declarations apply to the tag they installed. Finally, we discovered that the Google Tag through its use of Google Consent Mode **collects personal data when users decline consent**, potentially violating EU data protection laws.



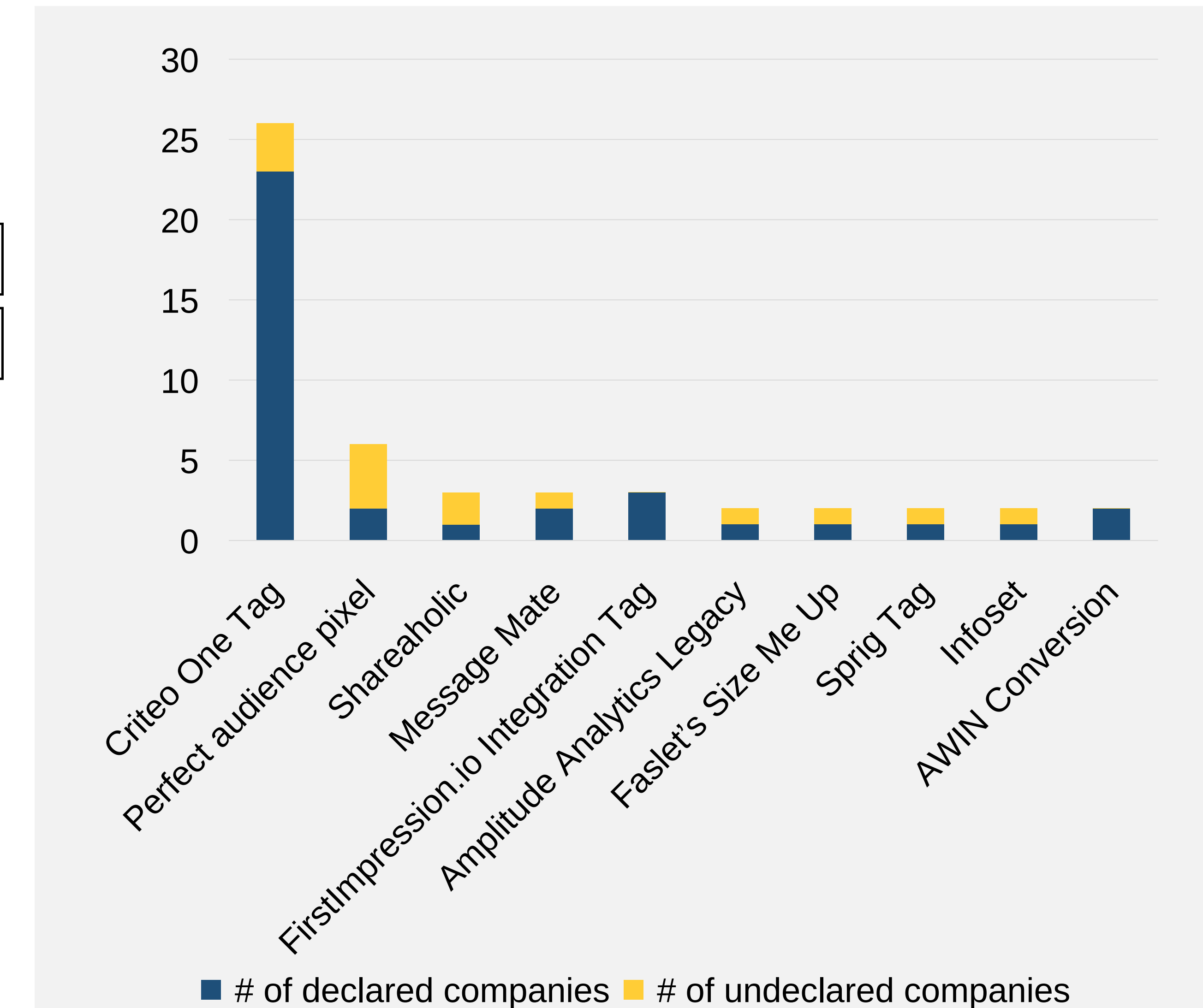
In-depth analysis : Architecture of a GTM installation. Tags run in the browser, collect data and send it to third parties.

URL param. name	Value of url parameter when refusing all	Value of url parameter when accepting all
v	2	2
tid	G-XSKNFL9CWM	G-XSKNFL9CWM
gtm	45je43p0v9181346581za200	45je43p0v9181346581za200
_p	1711640211524	1711640309156
gcs	G100	G111
gcd	13p3p3p2p5	13t3t3t2t5
npa	1	0
dma_cps	-	sypham
dma	1	1
cid	1337464789.17116	1994365208.17116
ul	en-us	en-us
sr	2560x1440	2560x1440
uaa	x86	x86
uab	64	64
uaflv	Chromium%3B122.0.6261.94%7CNot(...	Chromium%3B122.0.6261.94%7CNot(...
uamb	0	0
uap	Linux	Linux
uapv	6.6.17	6.6.17
uaw	0	0
pscdl	denied	noapi
_s	2	2
sid	1711640211	1711640280
sct	1	1
seg	0	1
dl	https%3A%2F%2Fga4consent.webexp...	https%3A%2F%2Fga4consent.webexp...
dr	https%3A%2F%2Fga4consent.webexp...	https%3A%2F%2Fga4consent.webexp...
dt	My%20experiment%20page%20title	My%20experiment%20page%20title
en	view_search_results	view_search_results
tfd	5620	5613
_eu	AEA	AEA
_et	2	
ep.search_term	my%20search	my%20search

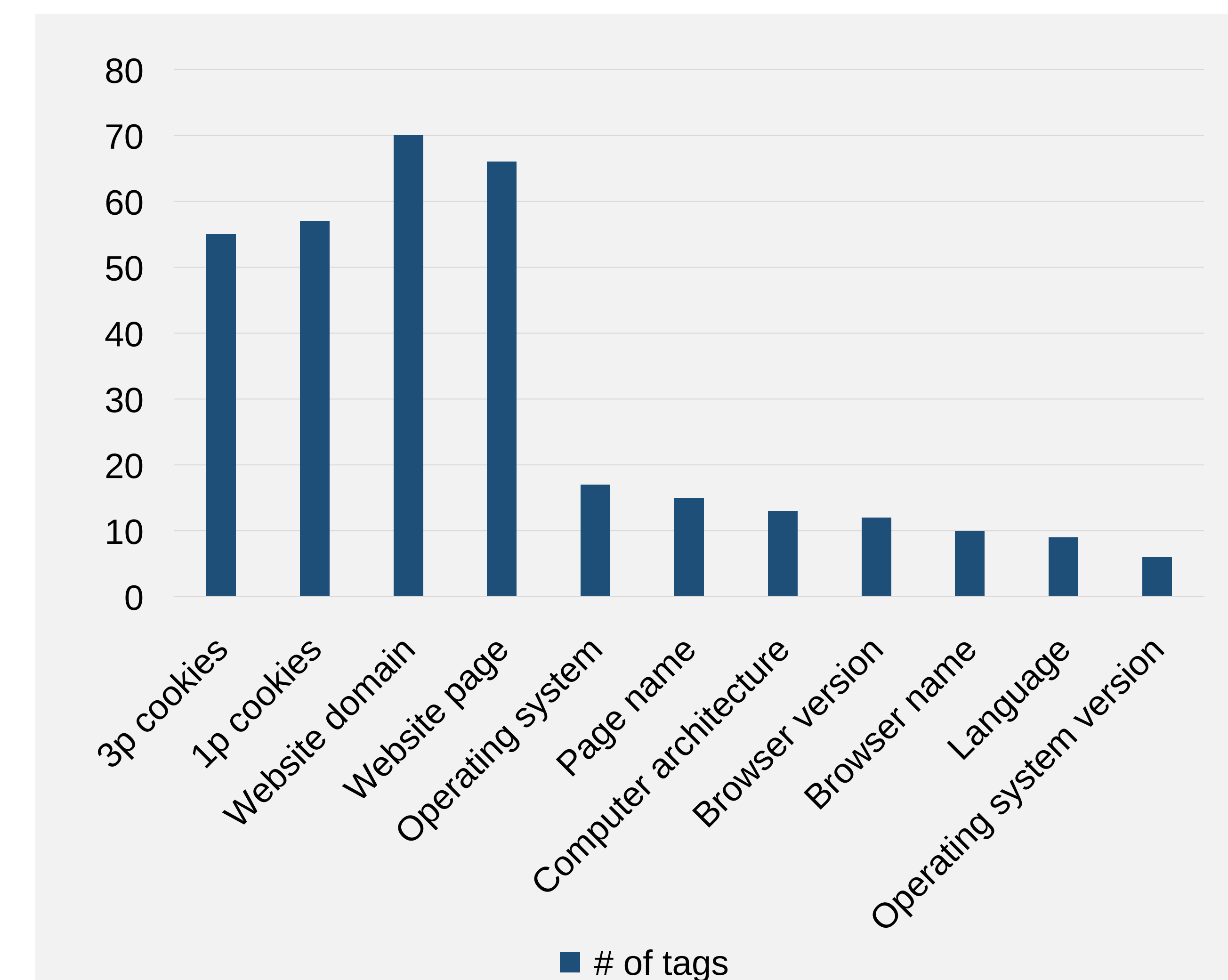
Consent mode: Data actively sent by the Google Tag when accepting and refusing consent. The values in red are sent independently of consent.



Previous submission
<https://arxiv.org/abs/2312.08806>



Large-scale & Policy analysis: Tags send data to multiple companies and don't declare it.



Large-scale study: A broad set of personal data types is actively collected by the tags

