Online Tracking and Privacy Task 2

Bram Pulles – S1015194
Sabah Serhir – s1124043
Robin – s1031986

April 2024

	Error type	Accept	Block
0	Page load timeout	8	7
1	Consent click failure	28	30

Table 1: (Exercise 1) Number of failures encountered during each crawl. Note the page load timeout is set 10 seconds.

	Metric	A, min	A, median	A, max	B, min	B, median	B, max
0	Page load time (s)	0.156	0.656	5.369	0.125	0.608	3.186
1	Number of requests	1.000	43.000	165.000	1.000	37.000	158.000
2	Distinct third party domains	0.000	3.000	10.000	0.000	2.000	6.000
3	Distinct tracker domains	0.000	1.000	8.000	0.000	0.000	1.000
4	Distinct third party domains with cookie	0.000	0.000	2.000	0.000	0.000	1.000

Table 2: (Exercise 3) A stands for Allow and B stands for Block.
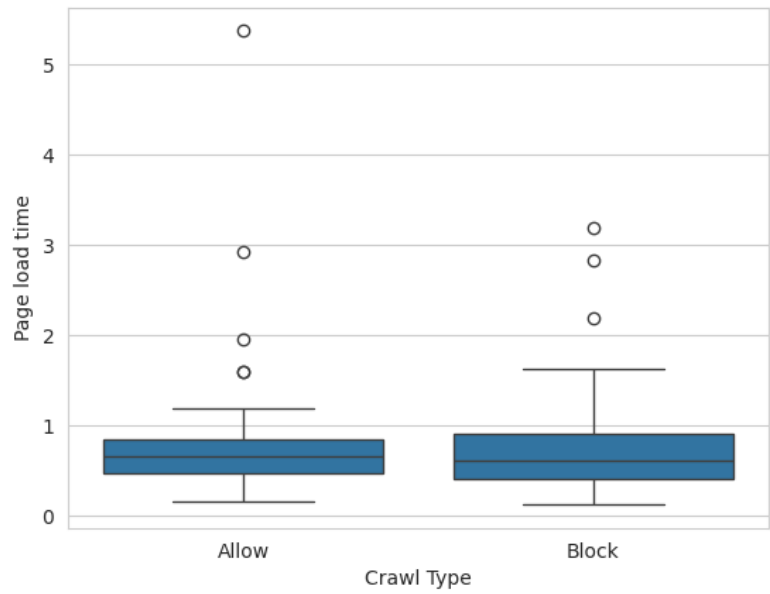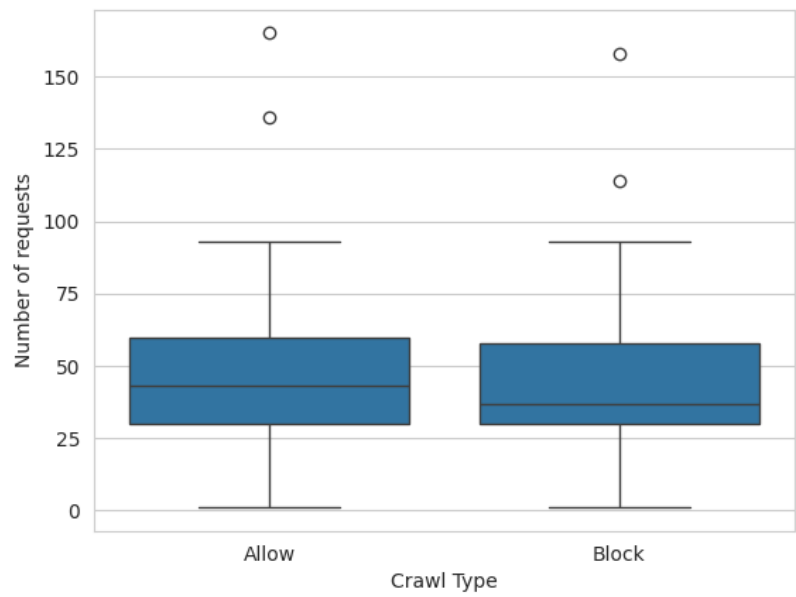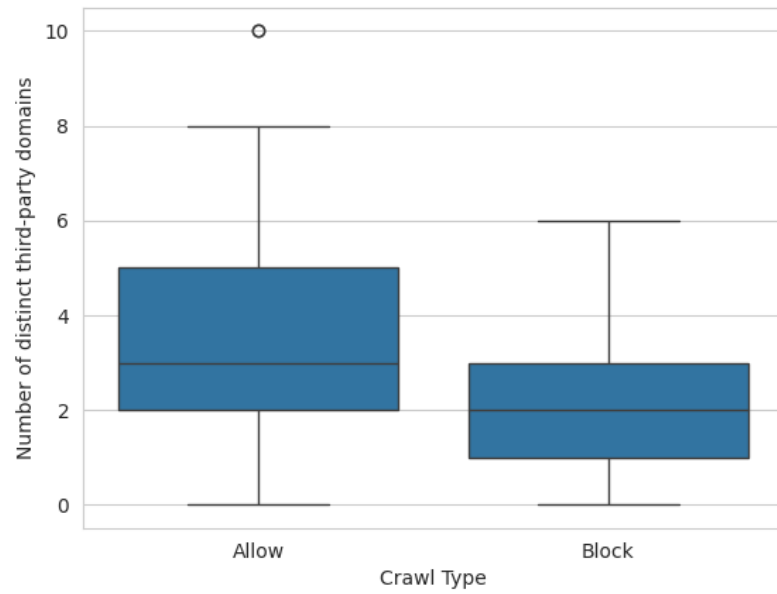
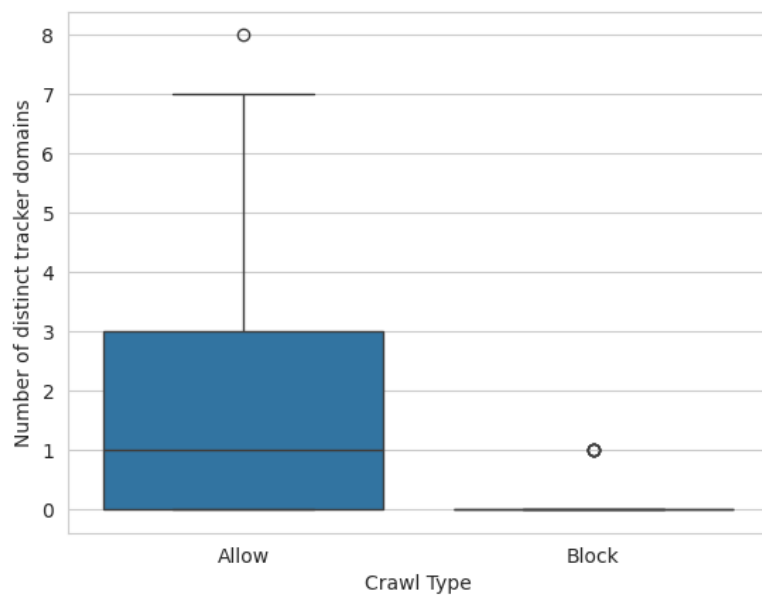Figure 1: (Exercise 2)

Figure 2: (Exercise 2)
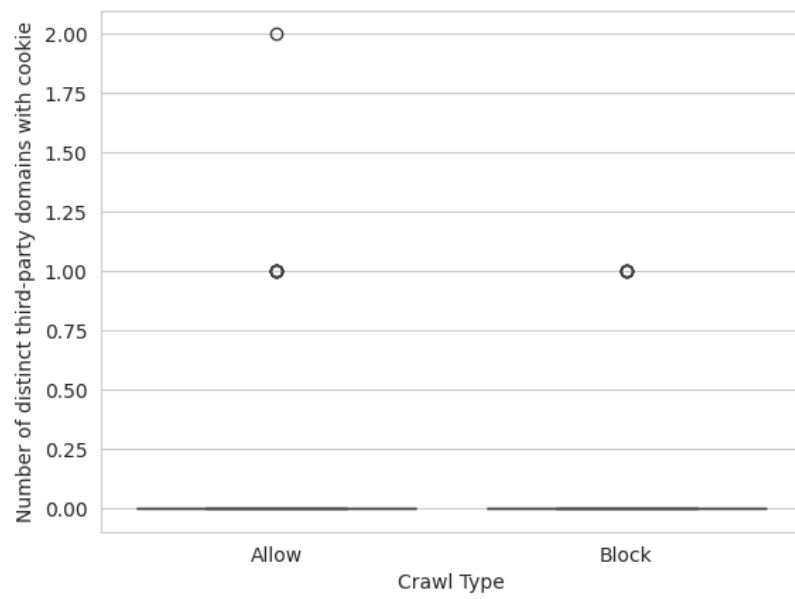
Figure 3: (Exercise 2)

Figure 4: (Exercise 2)

Figure 5: (Exercise 2)

	Third-party domain	Number of distinct websites (Allow)	is Tracker?
0	piwik.pro	44	No
1	google-analytics.com	40	Yes
2	sim-cdn.nl	38	No
3	rijksoverheid.nl	35	No
4	siteimproveanalytics.io	34	Yes
5	googletagmanager.com	26	No
6	readspeaker.com	21	No
7	abtasty.com	21	Yes
8	typekit.net	21	No
9	report-uri.com	16	No

Table 3: (Exercise 4) Ten most prevalent third-party domains during crawl-allow.

	Third-party domain	Number of distinct websites (Block)	is Tracker?
0	piwik.pro	44	No
1	sim-cdn.nl	43	No
2	rijksoverheid.nl	34	No
3	googletagmanager.com	25	No
4	google-analytics.com	23	Yes
5	readspeaker.com	21	No
6	typekit.net	21	No
7	report-uri.com	16	No
8	mopinion.com	15	No
9	citynavigator.nl	14	No

Table 4: (Exercise 4) Ten most prevalent third-party domains during crawl-block.

	HTTP Method	Number of requests (Allow)	Number of requests (Block)
0	GET	2282	2197
1	POST	122	105
2	HEAD	1	1

Table 5: (Exercise 5)

(Exercise 6) Websites that disable access for all parties (Allow).

- geolocation: nwo.nl

- camera: nwo.nl

- microphone: nwo.nl

(Exercise 6) Websites that disable access for all parties (Block).

- geolocation: nwo.nl

- camera: nwo.nl

- microphone: nwo.nl

(Exercise 7) Referrer-Policy headers that use `no-referrer` or `unsafe-url`.

- No referrer sites (Allow): noord-holland.nl, provinciegroningen.nl, gelderland.nl, ind.nl, werk.nl, denhaag.nl, digid.nl, studyinnl.org, gov.nl

- No referrer sites (Block): noord-holland.nl, provinciegroningen.nl, gelderland.nl, ind.nl, werk.nl, denhaag.nl, digid.nl, studyinnl.org, gov.nl

- Unsafe url sites (Allow): utrecht.nl

- Unsafe url sites (Block): utrecht.nl

	Client hint	Number of websites (Allow)
0	Sec-CH-UA-Full-Version-List	1
1	Sec-CH-UA-Platform-Version	1
2	Sec-CH-UA-Arch	1
3	Sec-CH-UA-Model	1
4	Sec-CH-UA-Bitness	1

Table 6: (Exercise 8)

	Client hint	Number of websites (Block)
0	Sec-CH-UA-Full-Version-List	1
1	Sec-CH-UA-Platform-Version	1
2	Sec-CH-UA-Arch	1
3	Sec-CH-UA-Model	1
4	Sec-CH-UA-Bitness	1

Table 7: (Exercise 8)

(Exercise 9) Both the Allow and Block variant only had one redirection from citynavigator.nl to plaece.nl at visitdrenthe.com.

(Exercise 10)

a)

One of the most challenging parts of the assignment was to actually click the cookies accept button. In a lot of cases the button on the website was not an actual button but only a http link. We have tried to also search for links that include words that could indicate an accepting cookies link or button, sadly there were too many websites that contained links with these accepting words as sub-words such that the crawler clicked on links that it was not supposed to.

It was also challenging to ensure that the analyses represent accurately the behavior and performance of the websites as we had to check manually that the results were correct, so it required careful attention to detail.

b)

The number of distinct third-party domains is relatively low, suggesting a focused approach to third-party integration or limited use of external resources. There is also a low number of distinct third-party domains with cookies set to SameSite = None and without the Partitioned attribute which aims to protect users privacy and security. Additionally, while some websites interact with a high number of tracking domains, others have none. There is a high variability in tracking practices among these websites.

We further found very few high entropy client hints, which might indicate not a lot of government websites use fingerprinting.