Understanding the Privacy Risks of Popular Search Engine Advertising Systems

Salim Chouaki, Oana Goga, Hamed Haddadi, Peter Snyder

ACM Internet Measurement Conference 2023

# Introduction

Popular traditional and private search engines.

	Traditional search engines		Private search engines		
	Bing	Google	DuckDuckGo	StartPage	Qwant
Advertising platform	Microsoft Advertising	Google Ads	Microsoft Advertising	Google Ads	Microsoft Advertising

DuckDuckGo, StartPage, and Qwant claim to not track users and respect their privacy.

However, they all rely on Microsoft Advertising or Google Ads to deliver ads to users!
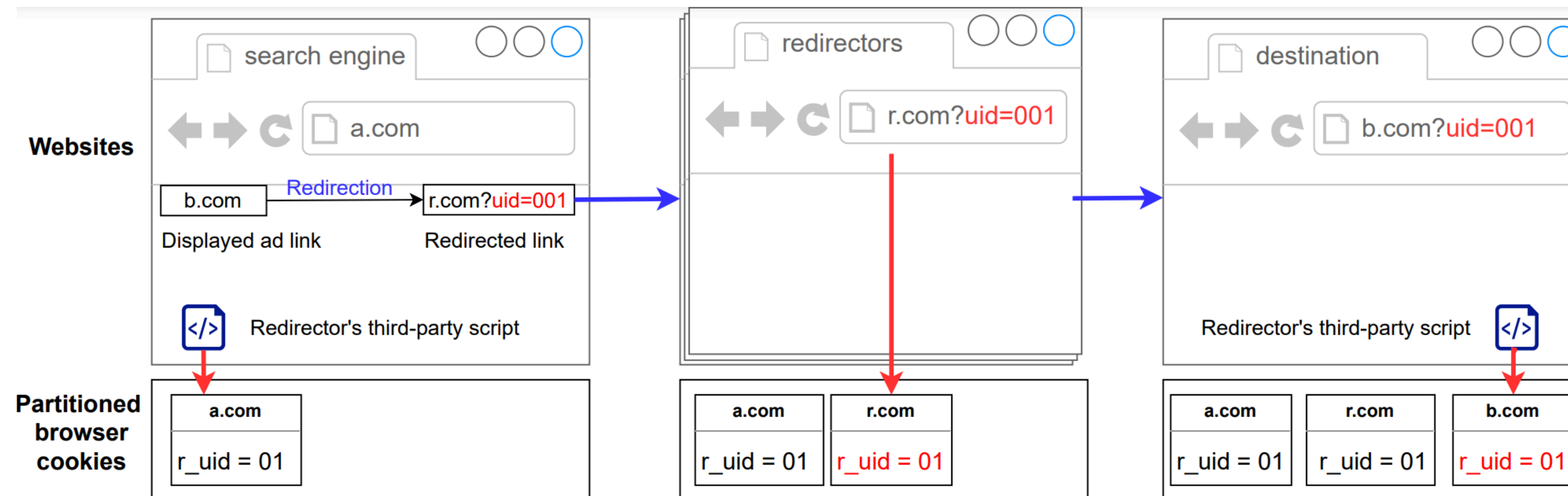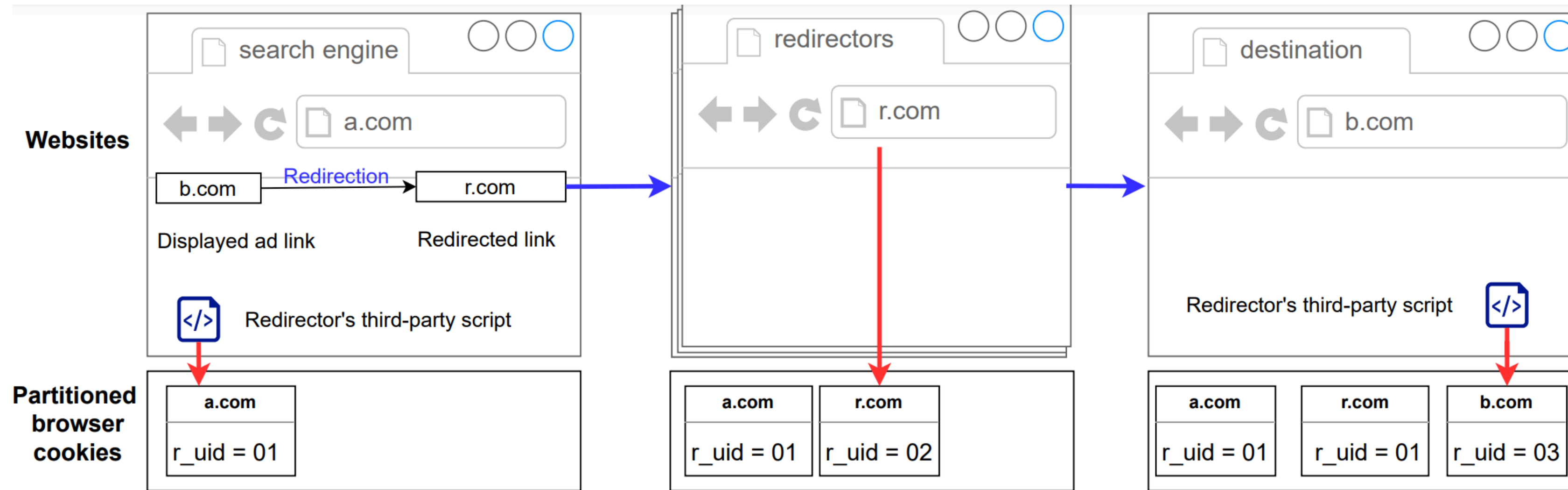
Motivation

Observation upon clicking an Ad on DuckDuckGo and manually inspecting the requests.
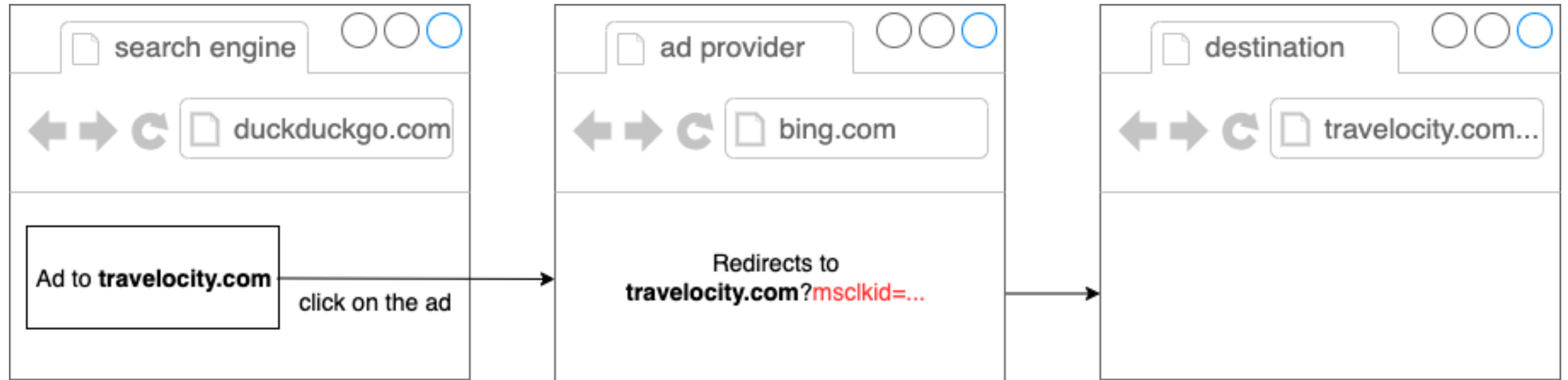
- **Two navigational tracking techniques:**

  - **Bounce tracking:** Redirection through bing when clicking the ad.

  - **User identifiers smuggling:** Adding the *msclkid* identifier as a query parameter.

Navigational tracking

Bounce tracking

UID smuggling

Objectives

- **Assess the privacy risks of private search engines advertising systems and how they compare to traditional search engines:**

  - Do search engines attempt to re-identify users across visits?

  - How often does bounce tracking happen when clicking on ads?

  - How often does users identifiers smuggling happen when clicking on ads?

  - Do advertisers' websites include known privacy-harming resources?

Crawling methodology

- **Puppeteer-based crawling system:**

  1. Access a search engine and accept cookies (if prompted).

  2. Type and submit a search query and identify search ads.

  3. Click on a search ad to access the destination page and wait for 15 seconds.

- **Challenges:**

  - Record first-party storage (cookies and local storage) and web requests at each step.

  - Identify web requests to online trackers.

  - Identify occurrences of bounce tracking (redirections) and user identifiers smuggling.

Identify web requests to online trackers

- **Use of URL filtering:**

  - Two open-source filter lists: EasyList and EasyPrivacy.

  - Combined and parsed using adblock-rs, resulting in 86,488 filtering rules.

- **Use of the Disconnect Entity List:**

  - Identify tracker domains and their associated entities.

  - Contains 1,449 entities and 3,371 related web domains.

Identify instances of bounce-tracking

- **Bounce-tracking:** When ads destination links are altered to go through redirectors.

- **Redirection chain:** Web domains navigated through after clicking an ad.

- **Examining HTTP response headers:**

  - Location header: contains the new redirection URL.

  - Status code header: 301, 302, 307, and 308.

Request URL:	https://www.bing.com/aclick?ld=e859KwYo_52ywJ2yx-UUQe2jVUCUzWugl6klVSntGC6moRhnJYoEhryILTolOWjRQbsW796Mjwfv-NA_uAA7NSQ6Fmib-mMOTLjH2XcsPNgMsps9wr1JTGr2G2YtW0KLmt4bvdEyGbIrCVXVw-TcGVw-zOnDDyylRYicXx0Q5mkCi-k7uZZTxjvmqYQI2XlVaKs3dPoQ&u=aHR0cHMlM2ElMmYlMmZmci52cG5tZW50b3IuY29tJTJmYmVzdC12cG4tZm9yLWZyYW5jZSUzZmtleXdvcmQlM2R2cG4lMjZnZW8lM2QxMjcxMjYlMjZkZXZpY2UlM2QlMjZ1dG1fc291cmNlJTNkYmluZyUyNmFkaWQlM2Q3NjQxNjE5NTcwNjMyMiUyNm1zY2xraWQlM2QyZWRkMTY1ZWNjZjcxMGMwOTVkMjczNmVjZjdmOGU1YQ&rlid=2edd165eccf710c095d2736ecf7f8e5a
Request Method:	GET
Status Code:	🟠 302 Found
Location:	https://fr.vpnmentor.com/best-vpn-for-france?keyword=vpn&geo=127126&device=&utm_source=bing&adid=76416195706322&msclkid=2edd165eccf710c095d2736ecf7f8e5a

Identify instances of UID smuggling

1. **Each iteration is executed in a new browser instance:**

   - We discard tokens with the same value across all or a subset of browser instances.

2. **An extra iteration is executed for each crawler, one day later:**

   - We discard tokens with different values across the two iterations.

3. **Multiple advertisements are rendered for each iteration:**

   - We discard tokens with different values across advertisers of the same iteration.

4. **Programmatic heuristics and manual filtering:m**

   - We discard tokens that appear to be timestamps, URLs, English words, coordinates.

   - We discard tokens that are seven characters or less.

Audrey Randall, Peter Snyder, Alisha Ukani, Alex C. Snoeren, Geoffrey M. Voelker, Stefan Savage, and Aaron Schulman. 2022. Measuring UID Smuggling in the Wild. In ACM Internet Measurement Conference (IMC '22).

Before clicking on an ad - on search engines

	Bing	Google	DuckDuckGo	StartPage	Qwant
First-party reidentification	✔	✔	✘	✘	✘
Requests to trackers	✘	✘	✘	✘	✘

Private search engines do not attempt to reidentify users across visits

When clicking on an ad - redirection chains

Top 3 most frequent redirection chains observed on each search engine

Search Engine	Redirection Chain	%
Bing	bing.com - destination	96%
	bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	3%
	bing.com - t23.intelliad.de - 1045.netrk.net - destination	1%
Google	google.com - googleadservices.com - destination	69%
	google.com - googleadservices.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	17%
	google.com - googleadservices.com - pixel.everesttech.net - ad.doubleclick.net - destination	4%
DuckDuckGo	duckduckgo.com - bing.com - destination	82%
	duckduckgo.com - bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	14%
	duckduckgo.com - bing.com - 6102.xg4ken.com - destination	2%
StartPage	startpage.com - google.com - googleadservices.com - destination	73%
	startpage.com - google.com - googleadservices.com - clickserve.dartsearch.net - ad.doubleclick.net - dest	17%
	startpage.com - google.com - destination	6%
Qwant	qwant.com - bing.com - destination	66%
	qwant.com - destination	14%
	qwant.com - bing.com - clickserve.dartsearch.net - ad.doubleclick.net - destination	10%

After clicking on an ad - on advertisers' sites

Top entities reached from advertisers websites

Bing	Google	DuckDuckGo	StartPage	Qwant
Unknown (32%)	Unknown (34%)	Unknown (30%)	Google (36%)	Google (26%)
Google (24%)	Google (29%)	Google (22%)	Unknown (28%)	Amazon (23%)
Microsoft (14%)	Microsoft (10%)	Amazon (16%)	Microsoft (4%)	Unknown (22%)
Facebook (4%)	Amazon (3%)	Facebook (3%)	Facebook (3%)	Microsoft (4%)
Criteo (2%)	Criteo (3%)	Criteo (2%)	Criteo (3%)	Criteo (4%)

Private search engines do not require advertisers to abide by privacy-respecting practices.

After clicking on an ad - on advertisers' sites

Fraction of iterations where advertiser websites received user identifiers as query parameters

	Bing	Google	DuckDuckGo	StartPage	Qwant
MSCLKID	79%	0%	66%	0%	51%
GCLID	12%	92%	12%	92%	8%
Other UID parameters	3%	8%	6%	12%	7%

Private search engines' advertising systems collude with advertisers and aid them in profiling visitors

Conclusion

- Private search engines' reliance on traditional advertising systems (Google and Microsoft) exposes their users to navigational tracking.

    - Bounce tracking observed in most ad-click occurrences on private search engines.

    - UID smuggling observed on all private search engines.

- There is a need for increased attention to privacy protections within the advertising systems of private search engines.

    - Build their own advertising systems.

    - Collaborate with Google or Microsoft to address privacy concerns (e.g., DuckDuckGo).

Questions?