Unveiling the Impact of User-Agent Reduction and Client Hints: A Measurement Study

Asuman Senol

imec-COSIC, KU Leuven

asuman.senol@esat.kuleuven.be

www.asumansenol.com

Gunes Acar

Radboud University

g.acar@cs.ru.nl

gunesacar.net

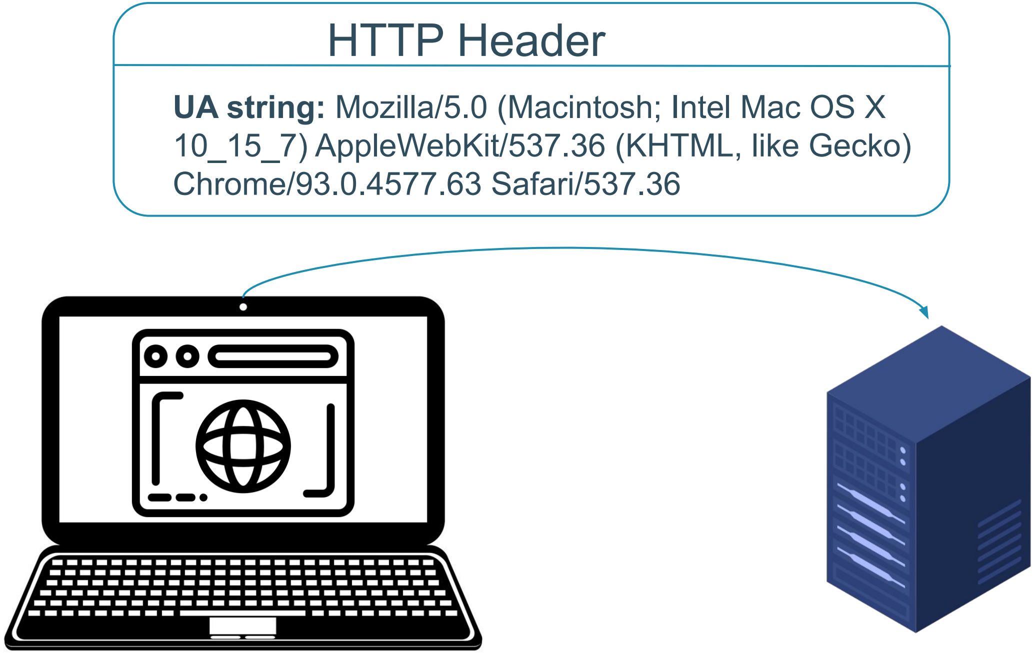
Background

What is user-agent string?

- Contains the details of a user's **device, platform and browser.**

Why does browser send this?

- Analytics
- Debugging
- Content adaptation
- Detecting incompatible, outdated or vulnerable browsers

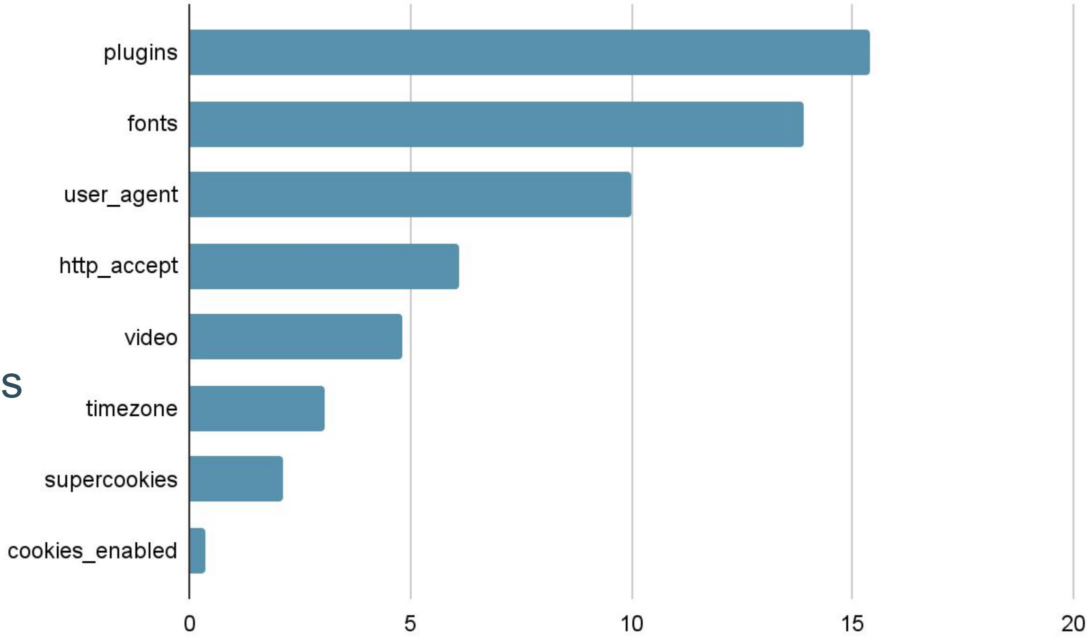


HTTP Header

UA string: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_7) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.4577.63 Safari/537.36

Motivation

- It enables *passive fingerprinting.*
  - Can be used for **cross-site tracking** by combining with
    - Screen dimensions, installed fonts, or graphics capabilities.
  - Affects **the uniqueness** of a user’s fingerprint.

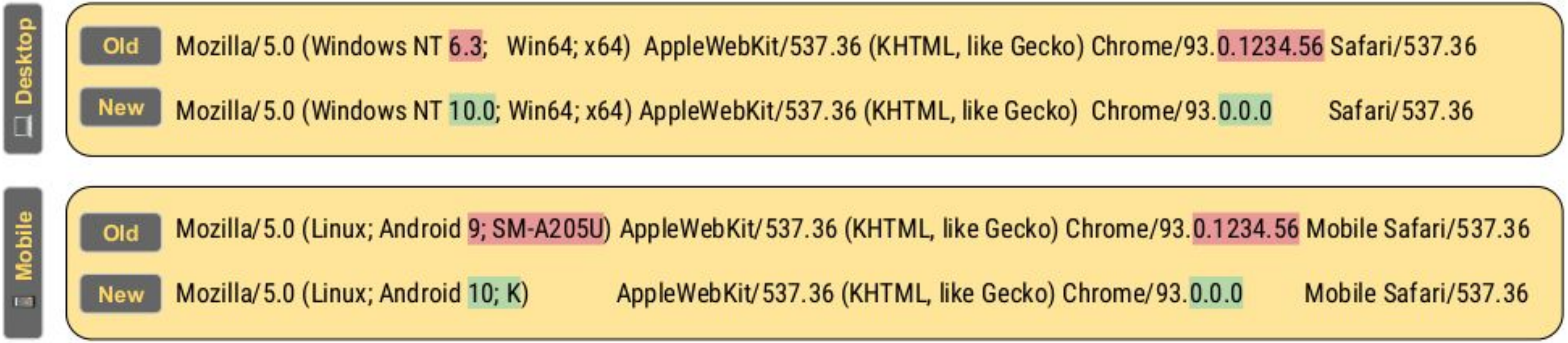


The most distinguishing browser features by entropy values [1].

[1] Peter Eckersley. 2010. How unique is your web browser? Privacy Enhancing Technologies (2010), 1–18. https://doi.org/10.1007/978-3-642-14527-8_1

Motivation

- Browsers reduced the identifying information in UA strings to enhance user privacy
- To access reduced details, Chrome introduced:
    - High-entropy user-agent client hints (UA-CH)
    - A new JavaScript API: *`navigator.userAgentData.getHighEntropyValues`*



Desktop

Old Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.1234.56 Safari/537.36

New Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.0.0 Safari/537.36

Mobile

Old Mozilla/5.0 (Linux; Android 9; SM-A205U) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.1234.56 Mobile Safari/537.36

New Mozilla/5.0 (Linux; Android 10; K) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0.0.0 Mobile Safari/537.36

Study Objectives

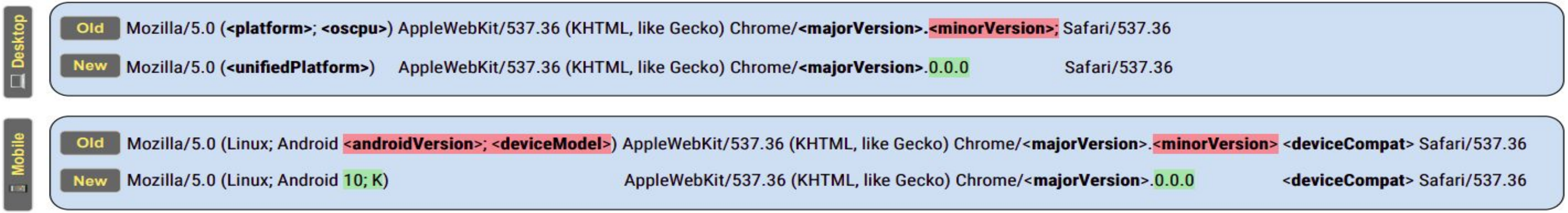
- Characterizing the effects of these major changes on the top 100K websites.
- Quantifying access to high-entropy browser features through
  - UA-CH HTTP headers
  - the JavaScript API
- Measuring access delegation to third parties such as trackers, advertisers, etc.

What changed and how?

1. Reduction of the UA string

For instance:

- Chrome 101 (June, 2022), minor version numbers were replaced with zeros
- Chrome 107 (Feb, 2023), CPU and platform-related details were simplified


Desktop

Old Mozilla/5.0 (**<platform>**; **<oscpu>**) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/**<majorVersion>**.**<minorVersion>**; Safari/537.36

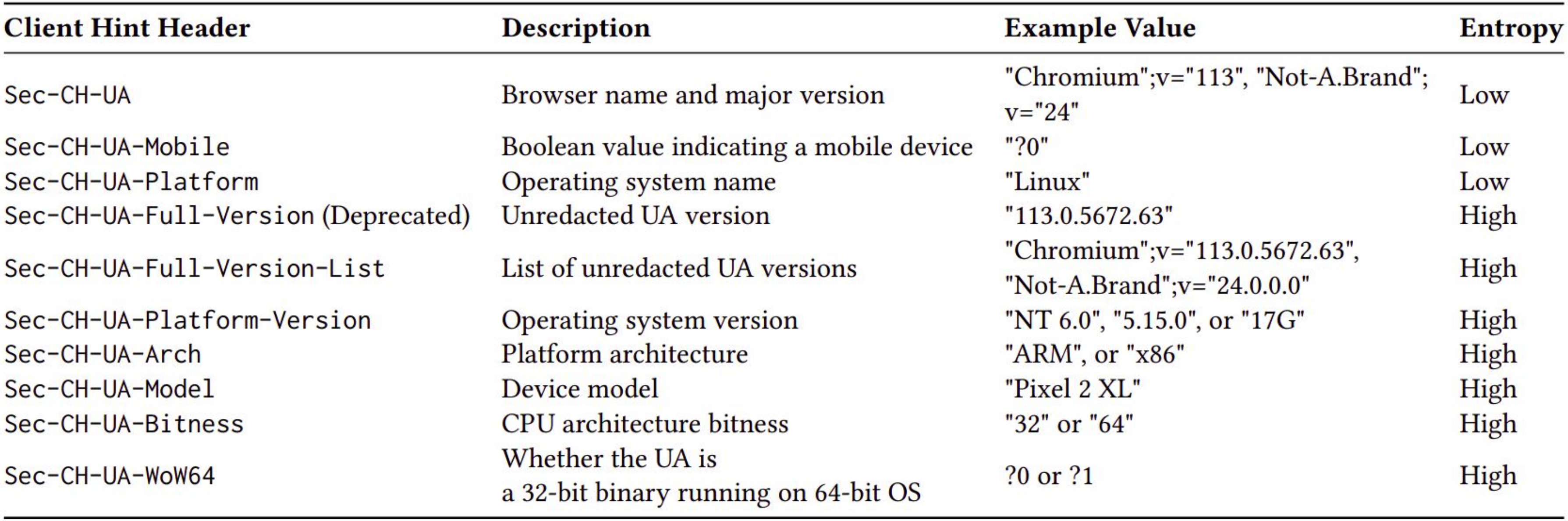
New Mozilla/5.0 (**<unifiedPlatform>**) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/**<majorVersion>**.0.0.0 Safari/537.36

Mobile

Old Mozilla/5.0 (Linux; Android **<androidVersion>**; **<deviceModel>**) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/<**majorVersion**>.<**minorVersion**> <**deviceCompat**> Safari/537.36

New Mozilla/5.0 (Linux; Android 10; K) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/<**majorVersion**>.0.0.0 <**deviceCompat**> Safari/537.36

2. User-agent client hint (UA-CH) HTTP headers



Client Hint Header	Description	Example Value	Entropy

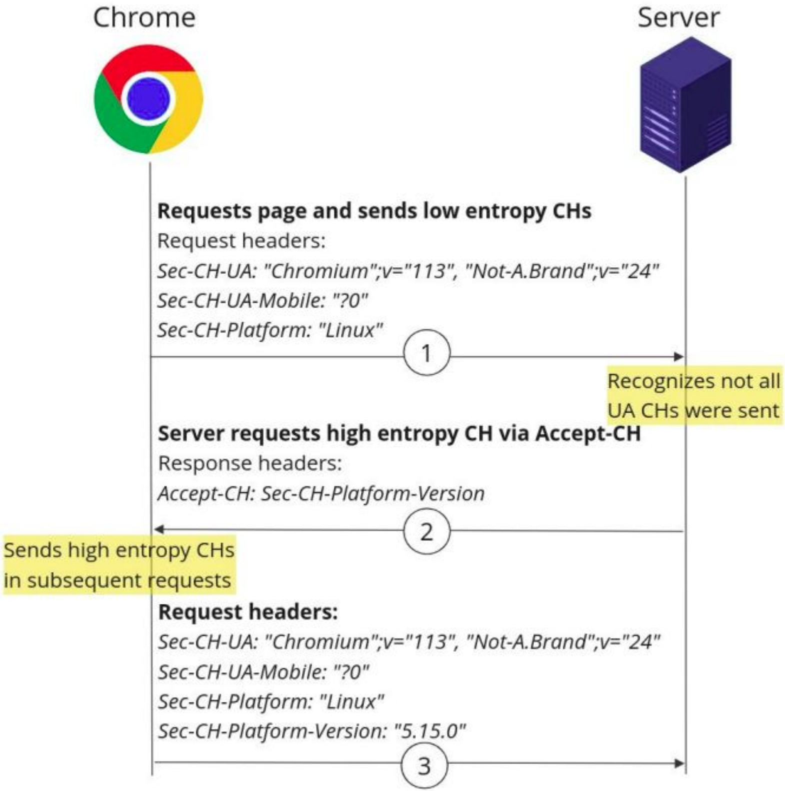
`Sec-CH-UA`	Browser name and major version	"Chromium";v="113", "Not-A.Brand";
		v="24"
			Low

`Sec-CH-UA-Mobile`	Boolean value indicating a mobile device	"?0"	Low
`Sec-CH-UA-Platform`	Operating system name	"Linux"	Low
`Sec-CH-UA-Full-Version` (Deprecated)	Unredacted UA version	"113.0.5672.63"	High
`Sec-CH-UA-Full-Version-List`	List of unredacted UA versions	"Chromium";v="113.0.5672.63", "Not-A.Brand";v="24.0.0.0"	High
`Sec-CH-UA-Platform-Version`	Operating system version	"NT 6.0", "5.15.0", or "17G"	High
`Sec-CH-UA-Arch`	Platform architecture	"ARM", or "x86"	High
`Sec-CH-UA-Model`	Device model	"Pixel 2 XL"	High
`Sec-CH-UA-Bitness`	CPU architecture bitness	"32" or "64"	High
`Sec-CH-UA-WoW64`	Whether the UA is a 32-bit binary running on 64-bit OS	?0 or ?1	High

Access to UA-CHs via HTTP

- Three low-entropy CHs is sent by Chrome by default in each request
  - platform name
  - major browser version
  - mobileness
- High-entropy CHs require
  - Explicit opt-in for 1st parties
  - Delegation to 3rd parties

Opt-in to high-entropy CHs via Accept-CH header



Requests page and sends low entropy CHs

Request headers:

Sec-CH-UA: "Chromium";v="113", "Not-A.Brand";v="24"
Sec-CH-UA-Mobile: "?0"
Sec-CH-Platform: "Linux"

Recognizes not all UA CHs were sent

Server requests high entropy CH via Accept-CH

Response headers:

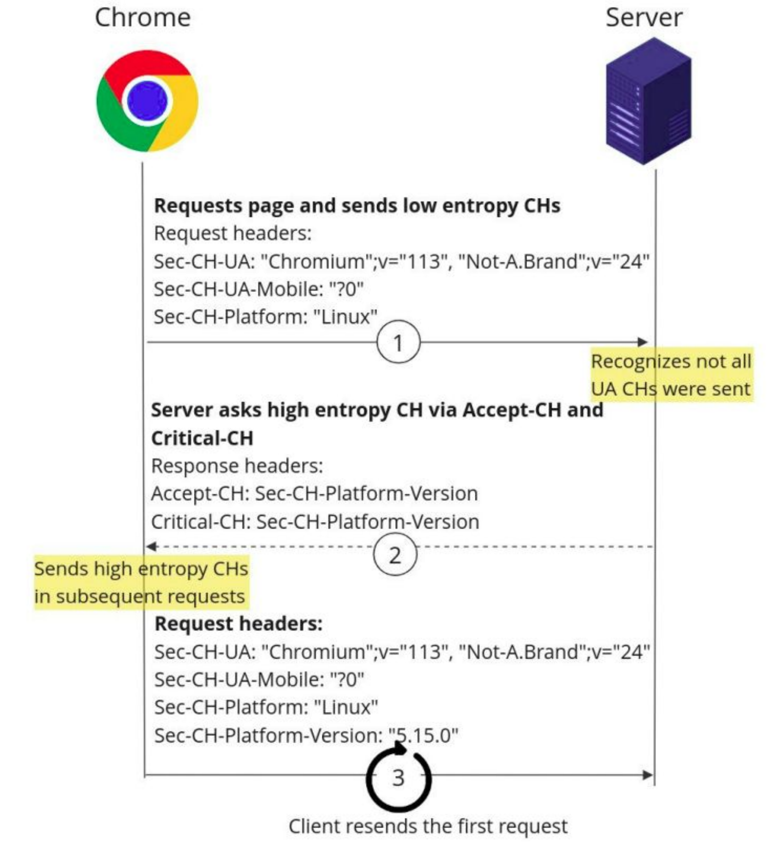
Accept-CH: Sec-CH-Platform-Version

Sends high entropy CHs in subsequent requests

Request headers:

Sec-CH-UA: "Chromium";v="113", "Not-A.Brand";v="24"
Sec-CH-UA-Mobile: "?0"
Sec-CH-Platform: "Linux"
Sec-CH-Platform-Version: "5.15.0"

High-entropy CHs in initial request via Critical-CH header



Requests page and sends low entropy CHs

Request headers:
Sec-CH-UA: "Chromium";v="113", "Not-A.Brand";v="24"
Sec-CH-UA-Mobile: "?0"
Sec-CH-Platform: "Linux"

Recognizes not all UA CHs were sent

Server asks high entropy CH via Accept-CH and Critical-CH

Response headers:
Accept-CH: Sec-CH-Platform-Version
Critical-CH: Sec-CH-Platform-Version

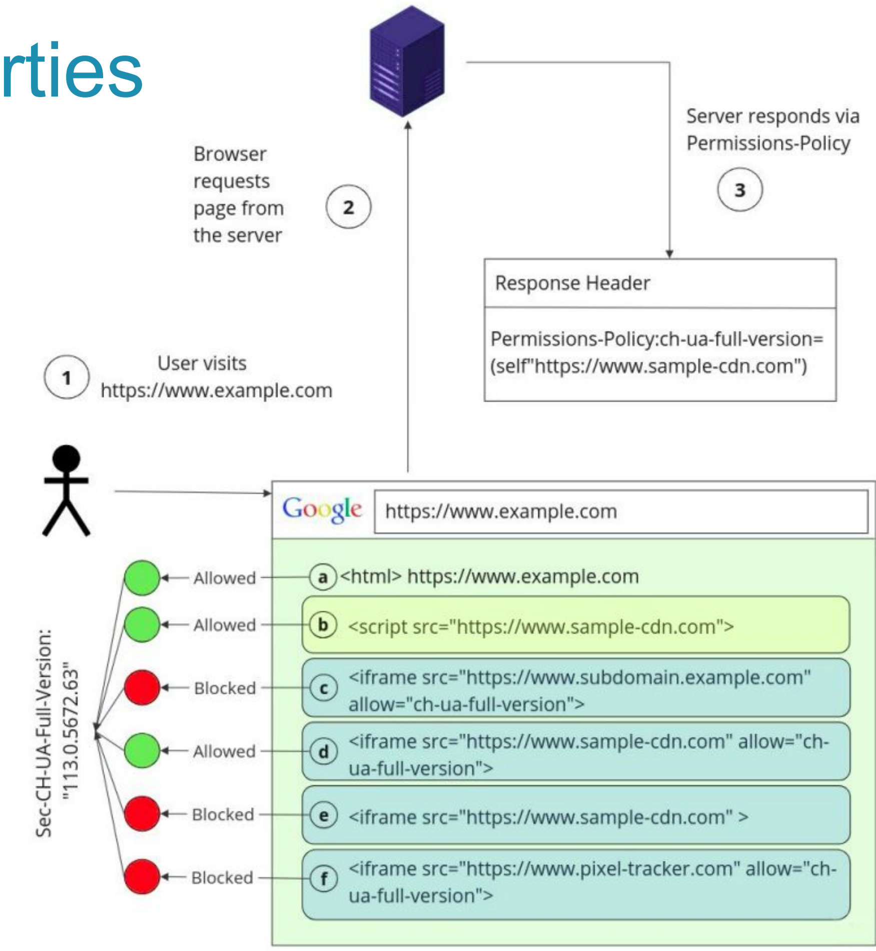
Sends high entropy CHs in subsequent requests

Request headers:
Sec-CH-UA: "Chromium";v="113", "Not-A.Brand";v="24"
Sec-CH-UA-Mobile: "?0"
Sec-CH-Platform: "Linux"
Sec-CH-Platform-Version: "5.15.0"

Client resends the first request

Delegating hints to third-parties

- First-party server must send a Permissions Policy header



Browser requests page from the server

2

Server responds via Permissions-Policy

3

Response Header

Permissions-Policy:ch-ua-full-version=(self"https://www.sample-cdn.com")

1

User visits https://www.example.com

Google

https://www.example.com

Sec-CH-UA-Full-Version: "113.0.5672.63"

Allowed

a <html> https://www.example.com

Allowed

b <script src="https://www.sample-cdn.com">

Blocked

c <iframe src="https://www.subdomain.example.com" allow="ch-ua-full-version">

Allowed

d <iframe src="https://www.sample-cdn.com" allow="ch-ua-full-version">

Blocked

e <iframe src="https://www.sample-cdn.com" >

Blocked

f <iframe src="https://www.pixel-tracker.com" allow="ch-ua-full-version">

Delegating hints to third-parties

- Via HTML (For publishers who cannot modify their website’s Permissions Policy HTTP header)
  - HTML <meta> tag
    - http-equiv="accept-ch" with content attribute
    - http-equiv="delegate-ch" with content attribute

3. New JavaScript interface: NavigatorUAData

Properties

1. NavigatorUAData.brands
2. NavigatorUAData.mobile
3. NavigatorUAData.platform

Methods

1. NavigatorUAData.getHighEntropyValues()
2. NavigatorUAData.toJSON()

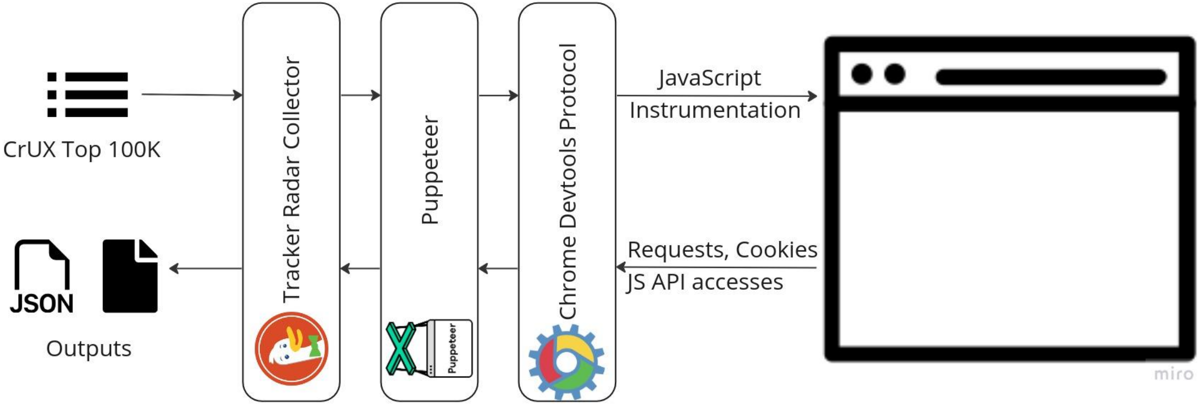
Statuses and positions of other browser vendors

moz://a

- Froze the rendering engine version,
- Reduced the information exposed in the UA string over time,
- Labeled UA-CHs as neutral in their web standard positions, but no work has been done as of today.

- Froze UA string in 2017 but later unfroze the major OS version,
- Negative stance against UA-CHs
- All browsers running on iOS have to use the WebKit rendering engine.

Method – Extending Tracker Radar Collector



Modifications:

1. Added 10 UA-CH HTTP headers and also Accept-CH and Critical-CH.
2. Intercepted JavaScript calls to `navigator.userAgentData.getHighEntropyValues` and save the arguments and the call stack.
3. Parsed the meta and iframe elements' attributes.
4. Instrumented fingerprinting-related method calls and property accesses.
5. Accepted personal data processing by porting Priv-Accept (Jha et al.)

Detection of high-entropy value exfiltrations

- Inspected HTTP request payloads and URLs to detect high-entropy CH exfiltrations.
  - Can be encoded, hashed or obfuscated.
  - Followed Englehardt et al.'s approach [2]
    - Searching for multi-layered encodings and hashes

[2] Steven Englehardt, Jeffrey Han, and Arvind Narayanan. 2018. I never signed up for this! Privacy implications of email tracking. Proc. Priv. Enhancing Technol. 2018, 1 (2018), 109–126

Identifying tracking-related requests

- Used uBlock Origin npm package
  - Includes filter lists such as EasyList, EasyPrivacy

Crawl

- Homepages of the top 100K sites (CrUX- April’23)
- In June’23
- On a cloud-based (DigitalOcean) server located in the United States

Results

getHighEntropyValues Calls and Exfiltrations

- 98.6% of the calls are due to third-party and tracking-related scripts

	All	Third party	Tracking related
getHighEntropyValues calls	53,148	52,392	51,630
Hi-ent. UA-CH exfiltration	48,355	47,691	47,285

getHighEntropyValues Calls and Exfiltrations

High Entropy API calls		High Entropy API exfiltrations	
Tracker domain	Num. Sites	Tracker domain	Num. Sites
googletagmanager.com	28,929	google-analytics.com	22,517
googlesyndication.com	6,843	google.com	9,325
doubleclick.net	3,633	doubleclick.net	8,853
googletagservices.com	1,414	googlesyndication.com	2,018
googleadservices.com	673	crwdcntrl.net	985

Top tracker domains calling getHighEntropyValues and exfiltrate high-entropy values

getHighEntropyValues Calls and Exfiltrations

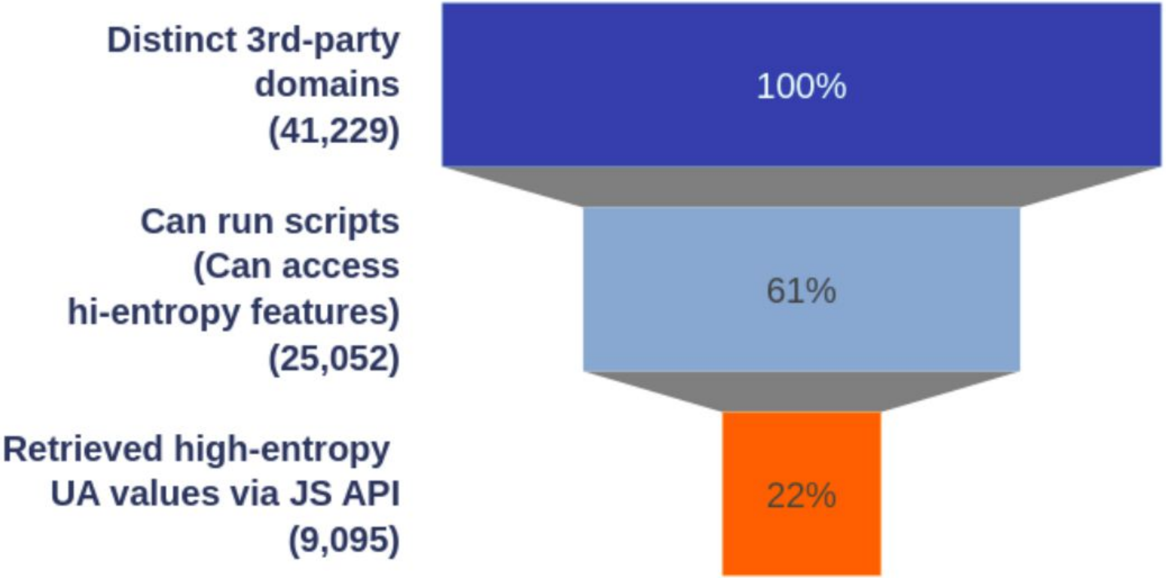
Script Category	Num. Sites.
Ad Motivated Tracking	44,084
Advertising	43,976
Audience Measurement	40,901
Third-Party Analytics Marketing	40,491
Analytics	40,347
Action Pixels	13,224
Embedded Content	4,523
CDN	4,342
Social - Share	2,338

Most common categories of third-party scripts calling getHighEntropyValues method.

getHighEntropyValues Calls and Exfiltrations

- The most frequently requested UA client hints via the JavaScript
  - `model` ⇒ on 52,270 sites
  - `platformVersion` ⇒ on 52,214 sites
- Call with mistyped argument: `uaFulVersion`
- Called with the argument None, only returns low entropy hints.

Reduction in high-entropy User-Agent exposure



The collection of User-Agent Client Hint HTTP headers

Ent.	UA-CH Header	All	Third Party	Tracking Related
High	Sec-CH-UA-Platform-Version	886	331	134
	Sec-CH-UA-Model	886	329	132
	Sec-CH-UA-Full-Version-List	696	261	67
	Sec-CH-UA-Arch	667	257	63
	Sec-CH-UA-Full-Version	581	217	25
	Sec-CH-UA-Bitness	491	217	25
	Sec-CH-UA-Wow64	401	210	21
Low	Sec-CH-UA	89,141	78,476	67,560
	Sec-CH-UA-Mobile	89,141	78,476	67,560
	Sec-CH-UA-Platform	89,141	78,476	67,560

Opt-in via Accept-CH header

Ent.	UA-CH Header	Num. Sites
High	Sec-CH-UA-Model	1,046
	Sec-CH-UA-Platform-Version	870
	Sec-CH-UA-Full-Version-List	824
	Sec-CH-UA-Arch	667
	Sec-CH-UA-Full-Version	799
	Sec-CH-UA-Bitness	443
	Sec-CH-UA-Wow64	354
Low	Sec-CH-UA-Platform	818
	Sec-CH-UA	434
	Sec-CH-UA-Mobile	403

Delegation via Permissions Policy

Ent.	UA-CH Header	Num. Sites
High	Sec-CH-UA-Platform-Version	338
	Sec-CH-UA-Model	337
	Sec-CH-UA-Full-Version-List	266
	Sec-CH-UA-Arch	266
	Sec-CH-UA-Bitness	225
	Sec-CH-UA-Full-Version	225
	Sec-CH-UA-Wow64	222
Low	Sec-CH-UA-Platform	225
	Sec-CH-UA	6
	Sec-CH-UA-Mobile	6

User-Agent Client Hint opt-in and delegation via HTML

Delegation	Num. Sites
http-equiv='accept-ch'	117
iframe-allow	32
http-equiv='delegate-ch'	11

Discussion

✓ UA reduction efforts **achieved to limit** the potentially **identifying information** in the UA HTTP **header**.

⚠ High-entropy client hints are **accessible** to scripts **without any control**. We believe browser vendors should consider imposing **stricter controls**.

Summary

- The first empirical study of impact of user-agent string reduction
- High-entropy UA CHs are accessed by third-party scripts on nearly 60% of the sites
- Over 90% of the websites, the obtained hints were exfiltrated to remote servers by tracker scripts
- Found the use of high-entropy UA-CH headers to be very limited

Project’s Website

https://homes.esat.kuleuven.be/~asenol/ua-reduction

Source Code

https://github.com/ua-reduction/ua-client-hints-crawler

Thank you!

Any Question?

asuman.senol@esat.kuleuven.be

www.asumansenol.com