

# Master Thesis Defense

Sebastian Neef

s.neef@mailbox.tu-berlin.de

Matriculation number: 350692

Supervised by:

Prof. Dr. Jean-Pierre Seifert

Prof. Dr. Florian Tschorsch

M.Sc Julian Fietkau



## Uncovering Fingerprinting Networks. An Analysis of In-Browser Tracking using a Behavior-based Approach



## Agenda

1. Introduction
2. Design & Implementation
3. Studies & Results
4. Conclusion



## Thesis Goal

**Study the fingerprinting networks and actors  
that foster browser fingerprinting.**

The Elephant in the Background: A Quantitative Approach to  
Empower Users Against Web Browser Fingerprinting

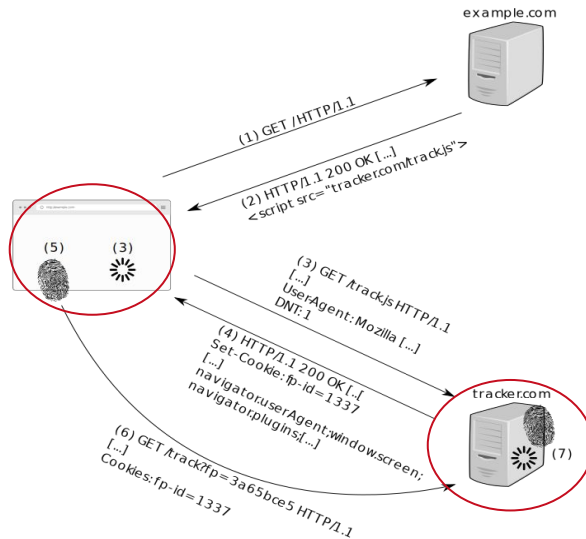
EasyChair Preprint no. 4473

14 pages · Date: October 26, 2020

[Julian Fietkau](#), [Kashjap Thimmaraju](#), [Felix Kybranz](#), [Sebastian Neef](#) and [Jean-Pierre Seifert](#)

# Browser Fingerprinting

- Uniquely (re-)identify browsers based on their properties



- Dual-Use Dilemma: Security vs. Privacy

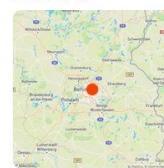
Wir haben einen ungewöhnlichen Anmeldeversuch für Ihr Konto festgestellt.

Wann: Aug 27, 2020 12:36 PM Central European Summer Time

Gerät: Opera Software Opera Linux (Desktop)

In der Nähe: Land Berlin, Germany

- (B) E-Mail notification, including browser details.

Visit History ⓘ	Your Current Visit
<ul style="list-style-type: none"> <li>Current visit</li> <li>March 9, 2021</li> <li>February 18, 2021</li> <li>January 21, 2021</li> <li>January 14, 2021</li> <li>January 14, 2021</li> <li>December 8, 2020</li> <li>December 8, 2020</li> </ul>	<p><b>Your ID:</b> rDY1kT9YHeWAi0NvP3zg ⓘ</p> <p><b>Headless Browser</b> No <b>Location</b> ⓘ</p> <p><b>IP</b> 91.10.152.152</p> <p><b>Incognito</b> No ⓘ</p> <p><b>Browser</b> Firefox on Linux</p> 



## Related Work

### Offensive

- JavaScript engine
- CSS
- Fonts
- Graphics
- History
- Plugins

### Defensive

- FP-Stalker
- Increasing diversity
- Randomness in return-values
- “Paradox of Fingerprintable Privacy Enhancing Technologies”
- Homogeneous fingerprints
- Disabling features

### Large-Scale Studies

- Panoptick
- AmlUnique
- **FPDetective**
- Scans for single offensive or defensive techniques

# The Behavior-based Approach

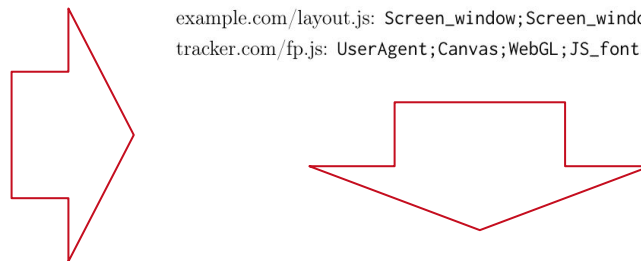
- Observe all JavaScript behavior on a per-script basis

gID	sID	oID	Property or function	Script origin	Feature group
0	0	0	navigator.userAgent	tracker.com/fp.js	UserAgent
1	0	1	screen.height	example.com/main.js	Screen_window
2	1	0	getImageData()	tracker.com/fp.js	Canvas
3	1	1	screen.width	example.com/main.js	Screen_window
4	2	0	drawArrays()	tracker.com/fp.js	WebGL
5	3	0	fillText()	tracker.com/fp.js	JS_fonts
...	...	...	...	...	...

TABLE 4: Example property and function observations including the associated feature group and script origin. The global-ID (gID), script-ID (sID) and origin-ID (oID) are essential to preserve the observation order.

- Compute script signatures and script scores

example.com/layout.js: Screen\_window;Screen\_window;... 1 + ...  
 tracker.com/fp.js: UserAgent;Canvas;WebGL;JS\_fonts;... 4 + ...



- Classify the activity levels based on the script score

Activity	Score	Interpretation
Low	score < 3	The script is likely benign.
Medium	$3 \leq \text{score} \leq 6$	The script exhibits limited fingerprinting activity.
High	score > 6	The script is considered to deliberately fingerprint the user.

TABLE 6: Categorization of scripts into low, medium, high fingerprinting activity based on their score.

# Uncovering Fingerprinting Networks and Actors

- Fingerprinting networks from **similar behavior**

Score	Size	Sig <sub>len</sub>	Files	Typ. Names	Script Domain(s)	Page Domain(s)
8	1,343	30	371	pubads[...].js, ...	doubleclick.net	blackdoctor.org, ebay.com, ...
26	232	97	230	1ad6cd50, 5e4f5e70, ...	dhl.com, dnb.com, ...	dhl.com, dnb.com, ...
36	5	269	1	device.js	maxmind.com	mediafire.com, ...
7	62	25	40	E-v1.js, embed_shepherd- v1.js, ...	wistia.com, wis- tia.net	paychex.com, rochester.edu, privy.com, ...
...	...	...	...	...	...	...

TABLE 7: Example fingerprinting networks with their properties.

- The networks' **script origins identify the actor**

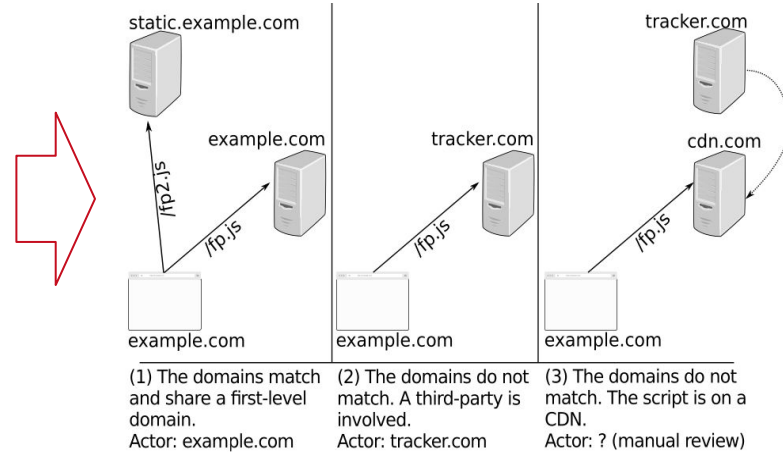


FIGURE 5: Illustration of the three ways this thesis identifies actors.



# FPNET - Fingerprinting Network Scanner

- Framework to collect fingerprinting information for large sets of websites

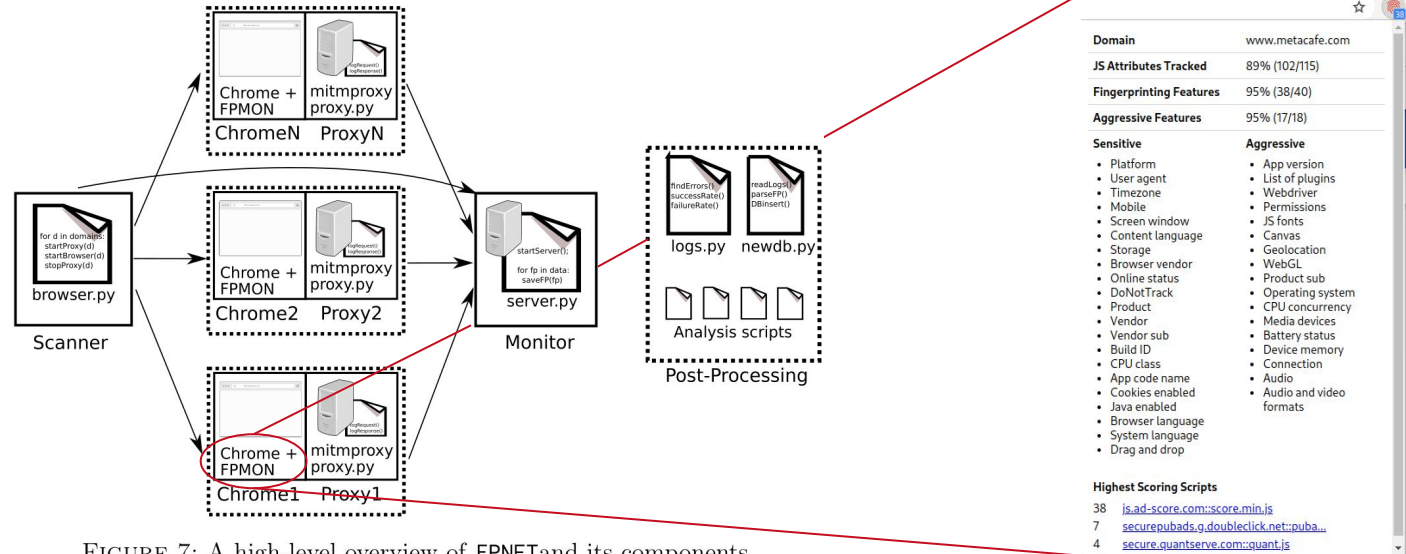


FIGURE 7: A high-level overview of FPNET and its components.





## Study 1: The Internet's Dependency on JavaScript

**Question:** How relevant is JavaScript for modern websites? Is active fingerprinting avoidable?

**Method:** Analyze the event handlers on the Alexa Top 10k websites.

**Results:**

- > 87% pages use event handlers
- Reduction of a website's usability or functionality without JavaScript

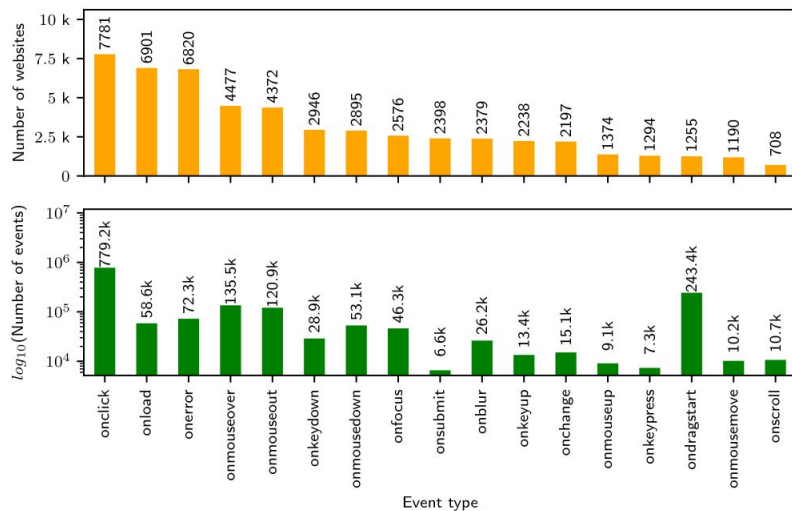


FIGURE 8: Top 15 event handlers by occurrences and website distribution from an analysis of the Alexa Top 10,000.



## Study 2: Identifying Fingerprinting Networks and Actors using Script Signatures

**Question:** What fingerprinting scripts, networks and actors exist on the internet?

**Method:** Scan the Alexa Top 10k with FPNET and analyze the script signatures.

**Results:**

- ~59% low activity
- ~32% medium activity
- ~9% high activity

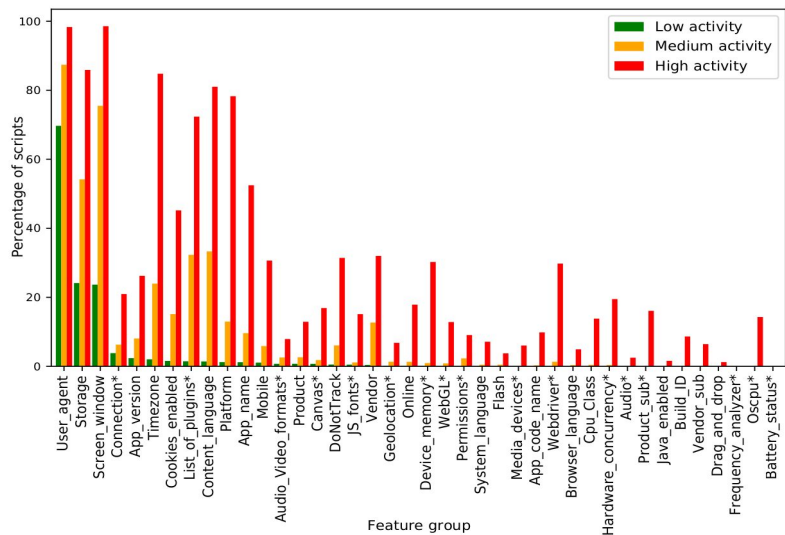


FIGURE 9: Distribution of feature groups for fingerprinting scripts with low, medium and high activity. \* denotes aggressive feature groups.



## Study 2: Identifying Fingerprinting Networks using Script Signatures

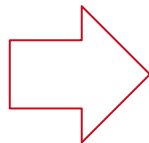
**Question:** What fingerprinting scripts, networks and actors exist on the internet?

**Method:** Scan the Alexa Top 10k with FPNET and analyze the script signatures.

**Results:** - > 375 networks from high activity scripts  
- > 100 actors from ~250 networks

Score	Size	Sig <sub>len</sub>	Files	Typ. Names	Script Domain(s)	Page Domain(s)
8	1,343	30	371	pubads[...].js, ...	doubleclick.net	blackdoctor.org, ebay.com, ...
26	232	97	230	1ad6cd50, 5e4f5e70, ...	dhl.com, dnb.com, ...	dhl.com, dnb.com, ...
36	5	269	1	device.js	maxmind.com	mediafire.com, ...
7	62	25	40	E-v1.js, embed_shepherd-v1.js, ...	wistia.com, wistia.net	paychex.com, rochester.edu, privy.com, ...
...	...	...	...	...	...	...

TABLE 7: Example fingerprinting networks with their properties.



Actor	Category	Networks	Score (Aggr.)	Pages
Google DoubleClick	Web Advertisements	19	10 (2)	1,583
Google AdSense	Search Engines	11	8 (1)	544
Yandex Metrika	Search Engines	52	14 (3)	367
Akamai	Computers	2	28 (10)	292
FingerprintJS	No Category	9	20 (10)	133

TABLE 8: Top 5 actors ranked by the amount of pages they can fingerprint.

Actor	Category	Networks	Score (Aggr.)	Pages
Maxmind	Computers	1	36 (14)	5
Moat	Web Advertisements	5	34 (12)	114
Adscore	Web Advertisements	2	29 (10)	19
Akamai	Computers	2	28 (10)	292
ShieldSquare (Perfdrive)	Vehicles	1	23 (6)	11

TABLE 9: Top 5 actors covering at least 5 pages ranked by the score.



## Study 3: Examining Scripts for Randomization

**Question:** How do fingerprinting scripts behave?

**Method:** Run two consecutive FPNET scans and analyze script signature and script origin changes.

**Results:**

- For **unchanged URLs** (~70k):
  - >86% scripts identical script signature.
  - ~92% of the 14% identical scores  $\Rightarrow$  Changed behavior
- For **changed URLs** (~9.8k):
  - ~ 90% filename changes (e.g., /710de559  $\Rightarrow$  /710de4e3)
  - 12 domain changes (e.g., xqheb9yszyrd.com  $\Rightarrow$  vk77lnizckm6.com)
  - ~ 800 filename and domain changes (e.g., .com\_mssgddsdsl.js  $\Rightarrow$  .co.uk\_jywraijsxptbytq.js)



## Study 4: Comparing Script- and File-Signature Networks

**Question:** Can file-signatures support the behavior-based approach?

**Method:** Compute file-content signatures with ssdeep and compare the resulting networks.

**Results:**

- Similar networks
- Varying properties

Score	# SB	# FB	$SB \cap FB$	$SB \cup FB$	<script domain>:<filename>
32	2	1	1	2	exponential.com:tags.js
27	141	8	7	142	easyjet.com:37f42850
25	16	19	14	21	athome.co.jp:eadjaxlayqcmrfpn.js
22	2	7	2	7	skyscanner.ru:init.js
21	3	2	1	4	px-cloud.net:main.min.js
21	2	7	1	8	zazzle.com:init.js
20	4	4	4	4	sift.com:s.js
20	2	3	2	3	yabidos.com:flimpobj.js?cb=[...]
20	11	13	9	15	adform.net:?pm=1511358&[...]
19	8	2	1	9	dns-shop.ru:log-action.js
19	2	4	1	5	hclips.com:barbar4.12.2.8ba[...].js
19	2	2	2	2	datadome.co:tags.js
18	88	160	82	166	celine.com:1e846d2919242f7e[...]
18	3	3	2	4	detik.net.id:thetracker-cnn-v3.min.js
18	2	160	2	160	hilton.com:2fe9b03fa8242608be[...]
18	18	160	18	160	jetstar.com:02e931a6e189a5dded[...]

TABLE 10: Comparison of signature-based (SB) networks versus file-based (FB) networks, including the number of script files resulting from an intersection or union of both network types, and a script example.



## Study 5: A Web Security Analysis

**Question:** What web security standards are present? Are scripts securely transferred?

**Method:** Analyze the HTTP requests & responses from the previously captured network traffic.

**Results:**

- ~ 98% securely transferred
- ~1k insecurely transferred

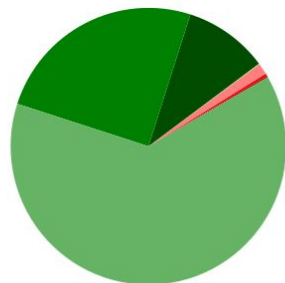
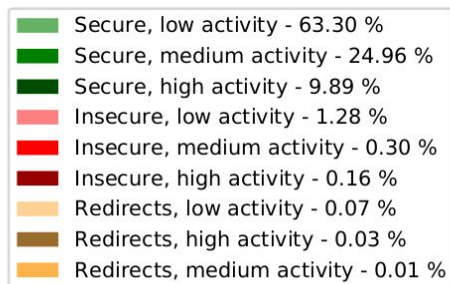


FIGURE 13: Script classification into high, medium and low activity for all fingerprinting script and HTTP request matches per group in study 5.



## Study 5: A Web Security Analysis

**Question:** What web security standards are present? What web security headers are set?

**Method:** Analyze the HTTP requests & responses from the previously captured network traffic.

**Results:** - Room for improvement

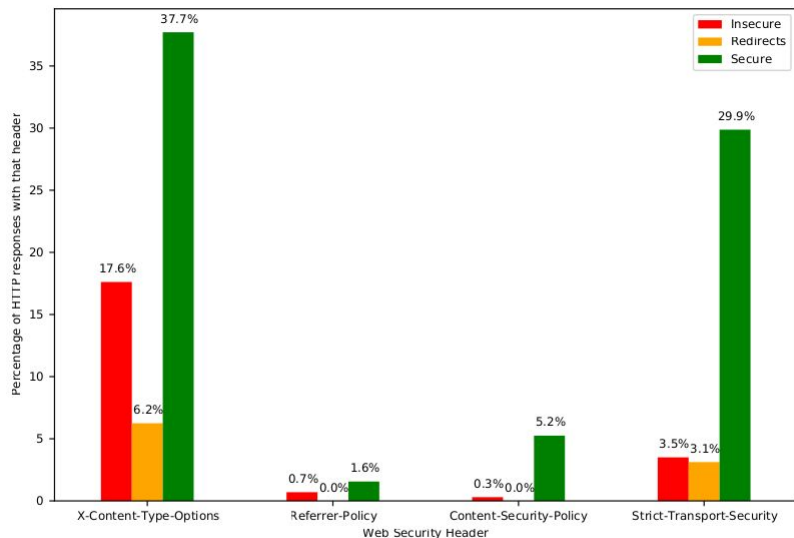


FIGURE 14: Distribution of web security headers for secure, insecure and redirect script URLs





## Comparison to Related Work

- **Enormous increase** of certain fingerprinting techniques
  - 3x the font-based fingerprinting than 8 years ago [2]
  - 2x the canvas-based fingerprinting than 4 years ago [35]
  - 1.6% use audio-based feature groups, instead of “infrequent” use reported by [35]
  - 54% pages running high-activity fingerprinting scripts compared to 69% [8] or 73% [66]
- **Security aspects**
  - Significantly higher adoption of SSL/TLS in third-party scripts with 98% compared to [8,35]
  - Documentation of web security headers
- **Actors change** over time
  - Maxmind, Google, Yandex, Iovation and others also identified by [1, 2, 8]
  - BlueCava, AddThis.com and others not discovered by us
  - New actors join the market



## Future Work

- Improved detection
  - Additional script properties
  - Larger or periodic scans
  - Community-approach
- Advanced analysis
  - Traffic analysis
  - Code analysis
  - Actor identification
- Privacy extensions
  - Blocking scripts based on behavior-/file-signatures
  - Blocking outgoing traffic with fingerprinting information



## Conclusion

- **Analysis**
  - Successfully uncovered fingerprinting networks and their actors
  - Steep increase of fingerprinting and its techniques
- **Tools**
  - FPNET framework for large-scale analysis
  - FPMON with script-level analysis
- **Impact**
  - Browser fingerprinting is widely used
  - Users likely not aware of being fingerprinted
  - Further research and education needed



# Thank you!

Prof. Dr. Jean-Pierre Seifert

Prof. Dr. Florian Tschorsch

M.Sc. Julian Fietkau

For more information, feel free to:

- Reach out to me [s.neef@mailbox.tu-berlin.de](mailto:s.neef@mailbox.tu-berlin.de)
- Read the thesis [github.com/gehaxelt/MasterThesis-FPNET/](https://github.com/gehaxelt/MasterThesis-FPNET/)



## References

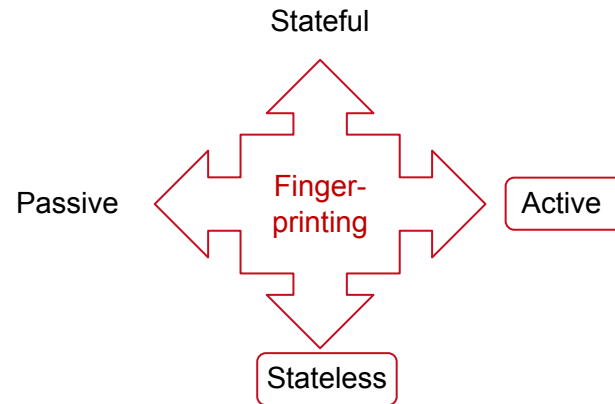
- [1] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz. “The Web Never Forgets”. In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*. ACM Press, 2014. DOI: 10.1145/2660267.2660347.
- [2] G. Acar, M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel. “FPDetective”. In: *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security - CCS '13*. ACM Press, 2013. DOI: 10.1145/2508859.2516674.
- [8] N. M. Al-Fannah, W. Li, and C. J. Mitchell. “Beyond Cookie Monster Amnesia: Real World Persistent Online Tracking”. In: *Developments in Language Theory*. Springer International Publishing, 2018, pp. 481–501. DOI: 10.1007/978-3-319-99136-8\_26.
- [33] P. Eckersley. “How Unique Is Your Web Browser?” In: *Privacy Enhancing Technologies*. Ed. by M. J. Atallah and N. J. Hopper. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 1–18. ISBN: 978-3-642-14527-8. DOI: 10.1007/978-3-642-14527-8\_1.
- [35] S. Englehardt and A. Narayanan. “Online Tracking”. In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, Oct. 2016. DOI: 10.1145/2976749.2978313.
- [66] F. Kybranz. “Uncovering Obfuscated Fingerprinting Techniques. A Large Scale Security and Privacy Analysis of Device Identification.” Master Thesis. Technical University of Berlin, Aug. 13, 2020. URL: <https://github.com/KybranzF/Master-Thesis/blob/master/MasterThesis.pdf>.



## Limitations

- **Connectivity issues**
  - Timeouts, certificate errors, etc.
- **FPNET detection**
  - Headless browsers, missing keyboard/mouse interaction, single IP address, etc.
- **Fingerprinting activity**
  - Limited to FPMON's functions, no GDPR/cookie consent, no sub-pages, etc.
- **Detection accuracy**
  - Randomized behavior, split scripts, asynchronous function calls, etc.
- **Actor attribution**
  - CDNs, multiple script origins, missing information, etc.

# Fingerprinting classifications





## Fingerprinting Traffic

Network	Score	Size	Data Sink	Examples of domain affected
Moatads	80%	58	3rd-party	breitbart.com, wsj.com, westernjournal.com, motor1.com, inquirer.net, nypost.com, ...
Sift	50%	45	3rd-party	udemy.com, scribd.com, patreon.com, kickstarter.com, wayfair.com, flickr.com, ...
Lalaping	88%	17	3rd-party	clipconverter.cc, shahid4u.net, swatchseries.to, o2tvseries.com, maxseries.tv, ...
Datadome	50%	16	3rd-party	nytimes.com, hepsiburada.com, leboncoin.fr, encuentra24.com, fnac.com, oui.sncf, ...
Adform	48%	31	3rd-party	freepik.com, coursehero.com, freepik.es, idnes.cz, tim.it, worldoftanks.eu, ...
Akamai	65%	232	1st party	adobe.com, rakuten.co.jp, foxnews.com, hulu.com, tokopedia.com, ikea.com, ...
fingerprint.js	48%	64	1st party	zhihu.com, agoda.com, olx.com.br, coinmarketcap.com, baixing.com, fmovies.to, ...
Google	20%	1343	3rd-party	reddit.com, okezone.com, twitch.tv, ebay.com, tribunnews.com, nytimes.com, ...

TABLE IV: The most prevalent script distributors with fingerprinting score and network size found with FPMON

[https://easychair.org/publications/preprint\\_open/H7Dc](https://easychair.org/publications/preprint_open/H7Dc)



## Fingerprinting products



### Bot Traffic Detection

Bots make up a consistent chunk of internet traffic. They mean ad fraud, drained advertising budgets and wasted computing power. Your hard-worked ad campaigns might even be stolen by bots of Ad Spy tools.

Come under **ADSCORE's** wing.



### Scraping Protection

You were assuming that having your web content scraped and re-published is a necessary evil. And you can only protect yourself with captchas by compromising your human visitors' experience.

Come under **ADSCORE's** wing.



### Low Quality Human Traffic Detection

Correctly targeting users according to their buying power and conversion rates is essential to running profitable campaigns. You can now accurately detect users with out of date operating systems and devices.

Come under **ADSCORE's** wing.

## Browser Fingerprinting API

Stop fraud, spam, and account takeovers with **99.5% accurate** browser fingerprinting as a service.

<https://www.adscore.com/>

<https://fingerprintjs.com/>

## Script-Signature String Similarity

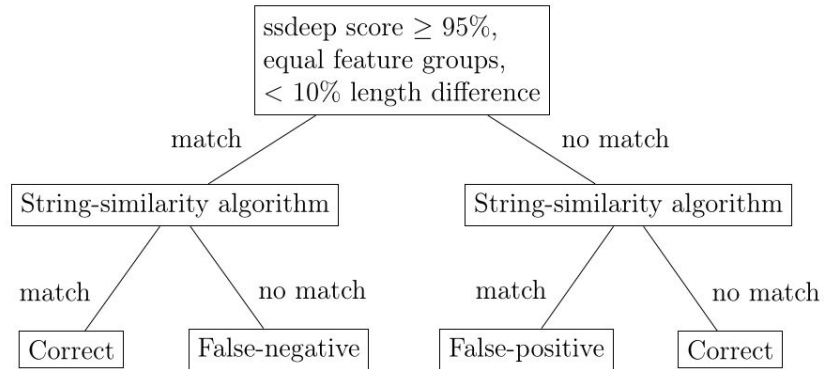


FIGURE 12: Decision tree for the string-similarity algorithm comparison.