

How Adblockers Work

And Why Communities Make it all Possible.

Rémi Berson

October 23th, 2019

Cliqz, Ghostery

Adblockers, adblockers...



What's a Website?

The screenshot shows the homepage of The Economist website. At the top, there is a banner for PATEK PHILIPPE GENEVE featuring a gold watch with a black dial. Below the banner, the site's navigation bar includes links for 'The Economist', 'Topics', 'Current edition', 'More', 'Subscribe', 'Log in or sign up', 'Manage subscription', and a search bar.

A main headline reads: "Not so super Saturday: MPs kick the Brexit can down the road. Again". A subtext below it says: "Boris Johnson is forced to ask for an extension, but vows Britain will leave this month".

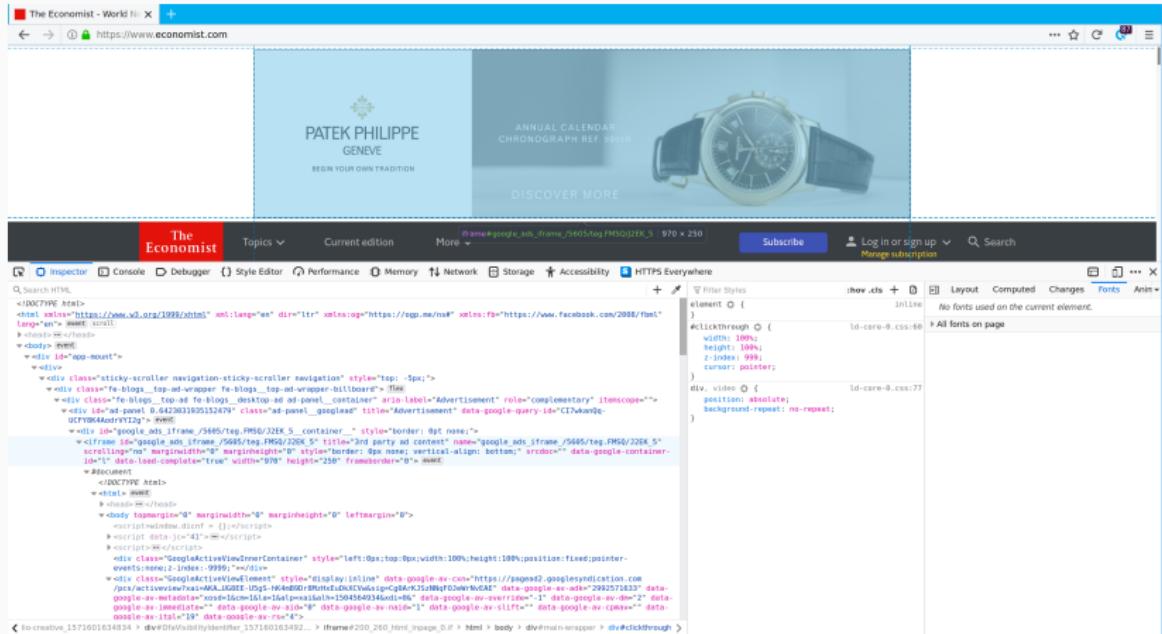
The central image is a photograph of Boris Johnson sitting in what appears to be the House of Commons. To his left is a small image of a protest with flags, and to his right is a small image of a plane.

On the right side, there is a sidebar for the "Current edition" dated Oct 19th 2019, which includes links to "Cyril Ramaphosa: Reforming South Africa", "Flight change: Skies darken for BA", and "Nordic noir: Europe's moody, boudoir-ing problem".

At the bottom, there are two calls-to-action: "Subscribe: 12 weeks for €20" and "Sign up: \$ articles per month".

More than meets the eye.

What's an Ad?



The *DOM*.

Blocking: What Can Extensions Do?

Adblockers are **extensions**:

- They are JavaScript **programs** running inside the browser.
- They can *monitor*, *modify* or **block** network requests.
- They can **hide** elements from pages.

Blocking: Network Filtering

Status	Method	Domain	File	Cause	Type	Transferred	Size	Time	
200	GET	px.moatads.com	pix.gif?r=21&q=06yc0=0x156_00814_00x0_00a180_00q=005yt=0a..._img	gif	419 B	43 B		152 ms	
404	GET	securepubsub.p.doubleclick.net	ad?gdfp_req=25795523244692196&correlator=28456036414..._xhr	plain	6.06 KB	16.39 KB		409 ms	
500	POST	es-sec.easymedia.com	headersstats?=>186960504+https://www.economist.com/w/e=3	xhr	plain	1.23 KB	0 B	78 ms	
200	GET	api.googleadsyndication.com	17229899439128411365	img	566 B	0 B		163 ms	
200	GET	ad.doubleclick.net	823196911_256982726_dc_trk_id=453149783_dc_trk_id=121701422..._img	gif	763 B	42 B		100 ms	
500	GET	securepubsub.p.doubleclick.net	ad?gdfp_req=25795523244692196&correlator=28456036414..._xhr	plain	716 B	0 B		96 ms	
200	GET	px.moatads.com	..._img	gif	419 B	43 B		62 ms	
500	POST	api.amebitive.com	..._xhr	json	557 B	79 B		120 ms	
200	GET	cdn.amebitive.org	script	js	cached	19.04 KB			
200	GET	cdn.amebitive.org	script	js	cached	0 B			
200	GET	cdn.amebitive.org	script	js	cached	0 B			
200	GET	api.googleadsyndication.com	17229899439128411365	img	63.06 KB	63.31 KB		102 ms	
200	GET	z.mzatads.com	iframe.html	https://tpc.googlesyndication.com/simged/17229899439128411365	html	4.77 KB	4.37 KB		59 ms
200	GET	z.mzatads.com	msatad.js	script	js	8.95 KB	34.34 KB		54 ms
200	GET	z.mzatads.com	msatadmp.js	script	js	6.39 KB	16.18 KB		117 ms

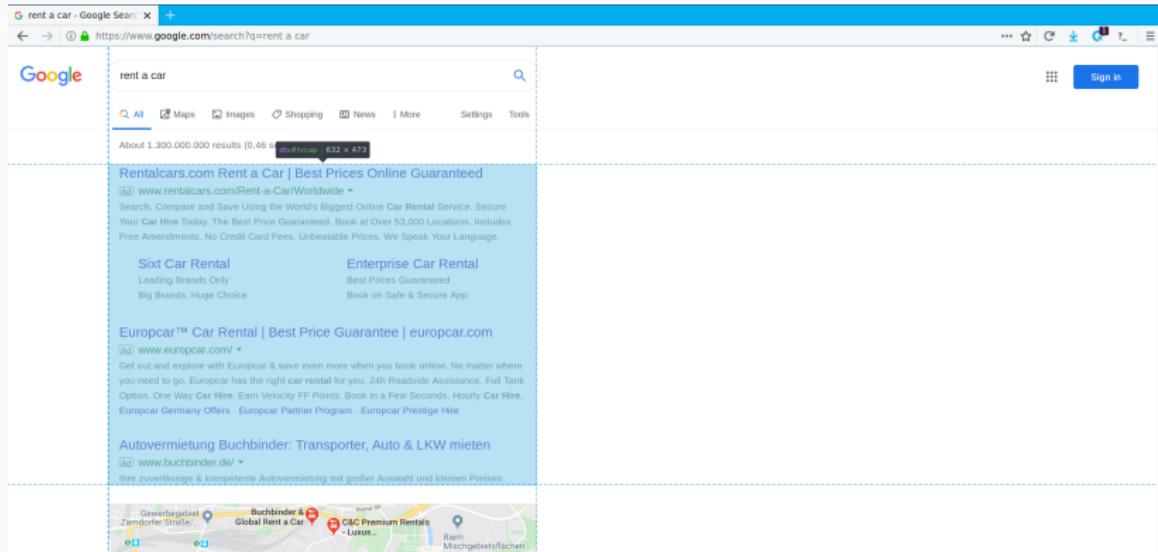
The ads have to be fetched somehow.

Hiding: Cosmetics Filtering

A screenshot of a Google search results page. The search query "rent a car" is entered in the search bar. The results show several links, including "Rentalcars.com Rent a Car | Best Prices Online Guaranteed" and "Europcar™ Car Rental | Best Price Guarantee | europcar.com". Each result includes a snippet of text and a link to the website. The interface shows standard Google search controls like "All", "Maps", "Images", "Shopping", "News", and "Settings". The overall layout is clean and typical of a web browser's search results view.

When blocking network requests is not enough.

Hiding: Cosmetics Filtering



We can remove or hide parts of the *DOM*.

Filters: Network

How to decide which network requests to block?

```
||example.com$script,third-party  
/ads/$image
```

~100k by default.

Filters: Cosmetic

How to decide **what to hide** in the page?

```
##.ad-element  
example.com##^script:has-text(loadAds)
```

~50k by default.

What's an Adblocker?

- Big rule engine
- Decide what to block in real-time
- Store rules in optimized format
- Needs to be *really fast*

- *fast* decision time (~ 0.01 ms per request)
- *low memory* usage (5.6 MB to 7.7 MB for **170k** filters)
- zero-cost initialization (0.1-20 ms)

whotracks.me/blog/adblockers_performance_study.html

Cliqz: Stack

 github.com/cliqz-oss/adblocker

-  — Hosted on *GitHub*
-  — Written in *TypeScript*
-  — Published on *NPM*
-  — Tested with *Jest*
-  — Managed with *Lerna*
-  — Bundled with *Rollup*
-  — Packaged with *Yarn*
-  — Formatted with *Prettier*

Run **everywhere** —     

Open Source?

Sharing **knowledge**: performance, designs, data, benchmarks:

-  influential work on efficient designs and features.
-  performance study and open benchmarks.
-  proposed currently most efficient approach to adblocking.
-  translated Cliqz' adblocker into Rust.
-  Adblocker Dev Summit.
-  shared mobile trackers database with whotracks.me.

Open Source!

Adblocking is not possible without a *strong community*!

-  — github.com/easylist/easylist
-  — github.com/uBlockOrigin/uAssets 
-  — github.com/AdguardTeam/AdguardFilters 

filterlists.com counts more than **1700** subscriptions!

Contributing?

So many ways to do *better together!*

- *report* site breakage or ads not blocked
- help *Maintain* up-to-date rules
- open *issues*, ask *questions*, start *discussions*
- *talk* and *write* about projects
- *share* your knowledge!

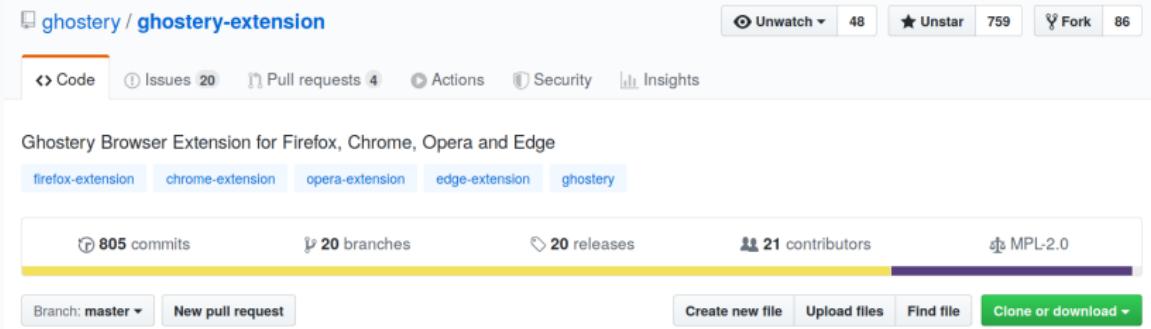
Contributing: GitHub!

— github.com/cliqz-oss/adblocker

The screenshot shows the GitHub repository page for `cliqz-oss/adblocker`. The page includes the repository name, a star icon, and a fork icon. It features a navigation bar with links for Code, Issues (6), Pull requests (0), Projects (0), Wiki, Security, Insights, and Settings. Below the navigation bar, a description states: "A fast, pure-JavaScript content-blocking library <https://cliqz.com>". There are buttons for content-blocking, javascript, adblocker, easylist, and Manage topics. Key statistics are displayed: 322 commits, 1 branch, 53 releases, 9 contributors, and a license of MPL-2.0. A progress bar indicates the repository's status. At the bottom, there are buttons for Branch: master, New pull request, Create new file, Upload files, Find file, and Clone or download.

Contributing: GitHub!

ghostery — github.com/ghostery/ghostery-extension

A screenshot of a GitHub repository page. At the top, there's a header with a profile icon, the repository name, and navigation links for Code, Issues (20), Pull requests (4), Actions, Security, and Insights. Below the header, a brief description reads "Ghostery Browser Extension for Firefox, Chrome, Opera and Edge". A horizontal bar displays metrics: 805 commits, 20 branches, 20 releases, 21 contributors, and the license MPL-2.0. At the bottom, there are buttons for Branch: master, New pull request, Create new file, Upload files, Find file, and Clone or download.

Contributing: Twitter!



— twitter.com/@pythux or twitter.com/@cliqz

Take a screenshot of any website

Visit URL Wait for Overlay Background

into a snapshot



Microlink screenshot - Take a screenshot of any website | Product Hunt producthunt.com

Released using ghostery behind the hood 😊



Jul 18, 2019, 9:45 AM

Questions?

- *jobs@cliqz.com* is **hiring!**
- Ping me: `remi@cliqz.com`
- Twitter: `@pythux`

⌚ — `github.com/cliqz-oss/adblocker`

- `npm` `@cliqz/adblocker-webextension`
- `npm` `@cliqz/adblocker-electron`
- `npm` `@cliqz/adblocker-puppeteer`