Bringing Content Blocking To The Masses

Dealing with Filter List Development, Maintenance and Compatibility for 50 Million Users

Shivan Kaul Sahib, Anton Lazarev
Brave Software

USENIX PEPR '22

Content Blocking

a.k.a. *adblocking* or *tracker blocking*

TL;DR:

1. Why is ad blocking important for Web privacy?
2. Filter lists and why they're hard to maintain
3. How do you ship a privacy product to millions of users on a platform as diverse as the Web?
4. Productizing research is hard (and some tips!)
5. Supporting an open source community project can be a superpower!

Who uses ad blockers?

37% of the Web

(even in 2016!)

M Malloy, M McNamara, A Cahn, and P Barford. “Ad blockers: Global prevalence and impact”, IMC 2016

Who uses ad blockers, and why?

People who use ad blockers generally fall into the following four groups, represented across the entire demographic spectrum in the U.S. and EU. These groups all respond differently to different types of engagement.

Here are the four groups of people who tend to use ad blockers:

1. **People who install ad blockers to protect their privacy:** This group values choice and control over their browsing experience, and they are more likely to allowlist a site than disable their ad blocker completely.

2. **People who simply don't want to see ads:** This group enjoys the convenience of

NATIONAL SECURITY AGENCY
CYBERSECURITY INFORMATION

BLOCKING UNNECESSARY ADVERTISING WEB CONTENT

Cyber adversaries can leverage malicious advertising ("malvertising") to install malware. Exploit kits in malicious ads can take advantage of unpatched vulnerabilities to silently install malware[1]. Administrators should ensure that software updates are implemented promptly to prevent malware installation. Blocking potentially malicious web advertisements further mitigates malvertising. Additionally, blocking such content can decrease traffic across the network boundary, streamlining incident forensics and enhancing network performance.

BACKGROUND

Web browsers present a major cyber security risk due to their frequent interaction with untrusted, Internet-based content. Due to the vast Internet landscape, it is generally not possible to predict and catalog the "good" websites that a user may visit. Instead, blacklisting approaches (such as Microsoft® SmartScreen®[2] and Google Safe Browsing™[3]) enhance security by blocking known malicious websites. Content that is neither inherently useful nor known to be malicious in nature, such as advertisements, often go unrestricted. Many websites include space for third party advertisers to display content. Despite the benign nature of most advertising content, advertising has been a known malware distribution vector[4] for over a decade[5]. This attack, known as "malvertising," allows a malicious actor to target users based on location, interests, browsing habits, and system specific identifiers, such as software versions[1].

CYBERSCOOP

(Getty Images)

Written by Tim Starks

JAN 14, 2021 | CYBERSCOOP

The U.S. Cybersecurity and Infrastructure Security Agency urged federal agencies on Thursday to deploy ad-blocking software and standardize web browser usage across their workforces in order to fend off advertisements implanted with malware.

“With many agencies greatly expanding telework options, agencies should increase attention on securing federal endpoints, including associated web browsing capabilities,” the Department of Homeland Security’s cyber arm said in a guide for agencies.

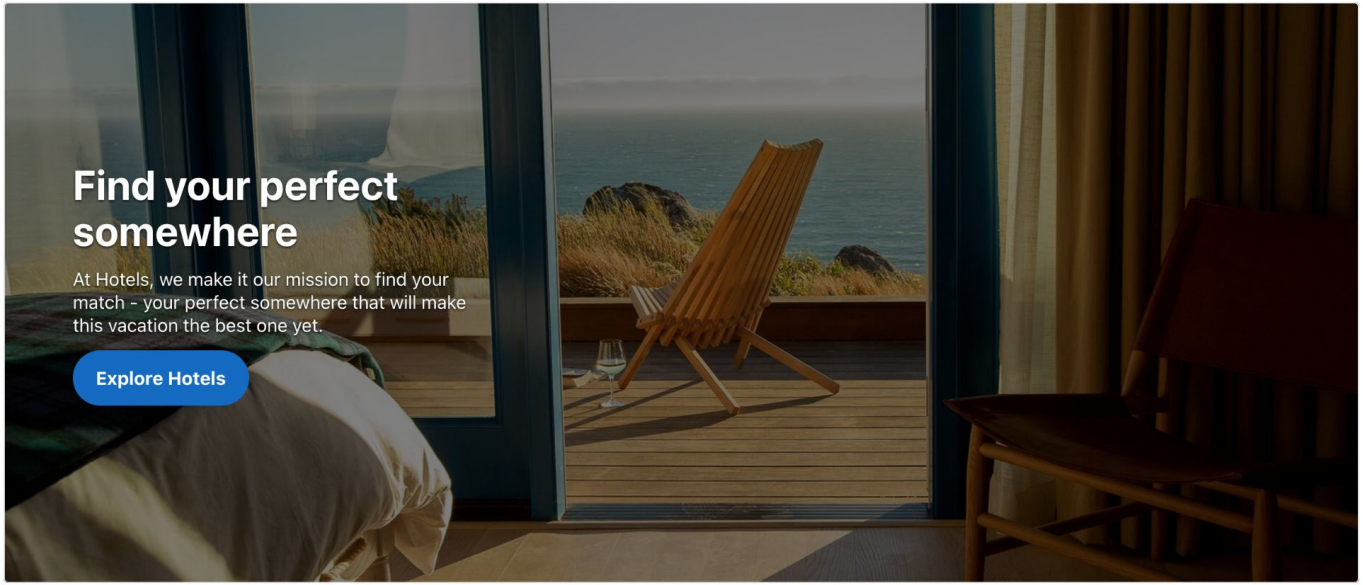
With the alert, CISA joins the National Security Agency, which in 2018 likewise urged agencies to adopt ad blockers in response to the threat from “malvertising” that can spread malware.

Anatomy of a block

Hotels.com™ More travel English List your property Support Trips Account

Where to?

Going to | Dates Jul 6 - Jul 7 | Travellers 2 travellers, 1 room



CANADA

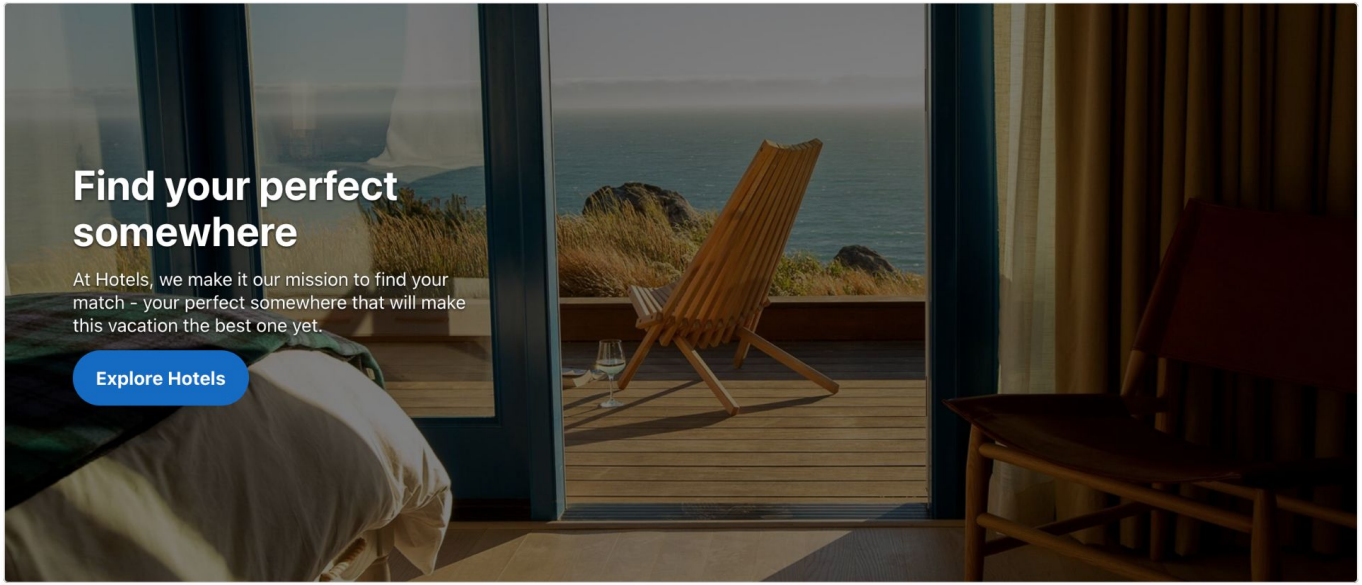
Seek authenticity, seek Canada

Find inspiring Indigenous experiences for your next vacation.

Hotels.com™ More travel English List your property Support Trips Account

Where to?

Going to | Dates Jul 6 - Jul 7 | Travellers 2 travellers, 1 room

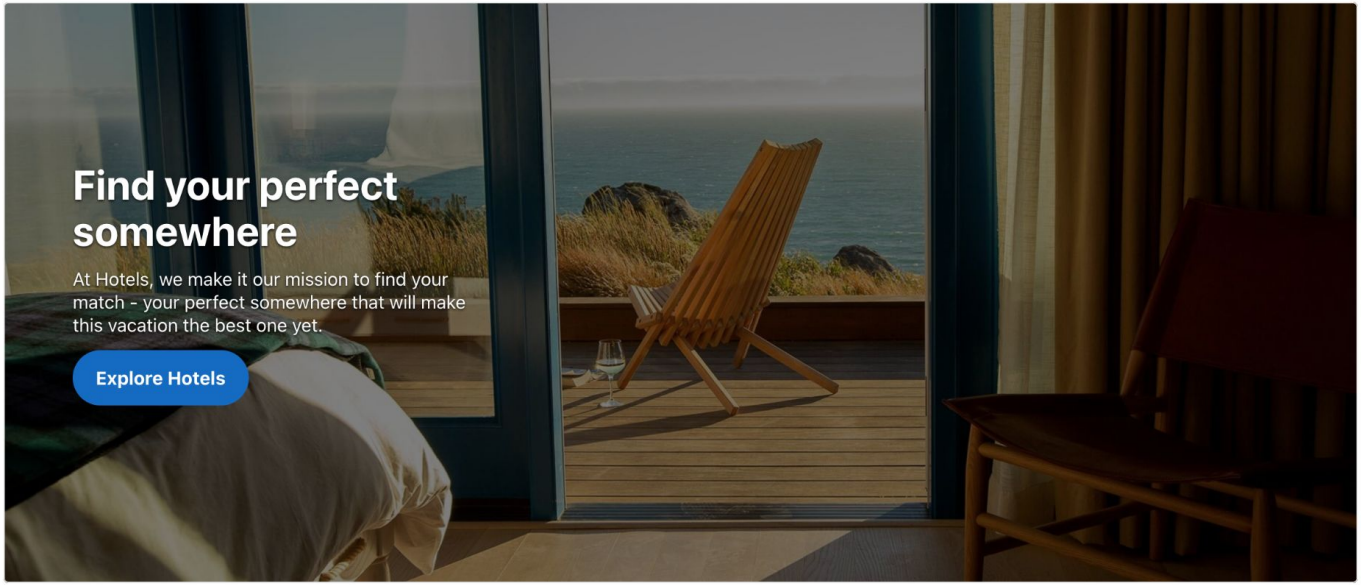


Filter Rule:
/expads-blocked.js

Hotels.com™ More travel English List your property Support Trips Account

Where to?

Going to | Dates Jul 6 - Jul 7 | Travellers 2 travellers, 1 room



Travel with confidence

Many properties have updated us about their enhanced health and safety measures. So, during your search, you may find details like:

Official health standards
Properties adhering to corporate/organizational sanitization guidelines.

Hygiene and sanitization
The use of disinfectant and whether properties enforce a gap period between stays.

Social distancing
Contactless check-in and check-out along with other social distancing measures.

Essentials at the property
Free hand sanitizer for guests and individually wrapped food options.

Ad blocked!

lotels.com™

re to?

oing to

tes
l 6 - Jul 7

vellers
ravellers, 1 room

Search

Elements Console Sources Network Performance Memory

Preserve log Disable cache No throttling

expads Invert Hide data URLs All Fetch/XHR JS CSS Ir

3rd-party requests

200 ms 400 ms 600 ms 800 ms 1000 ms 1200 ms 1400 m

Name	Status	Type
expads-blocked.js	(blocked:other)	script

What are filter lists?

Collection of rules that define what things to block on which websites.

Examples: EasyList, EasyPrivacy

easylist.to/easylist/easylist.txt

```
/eroadvertising.$domain=~eroadvertising.com
/erobanner.
/esi/ads/*
/etology.$domain=~etology.com
/euads/*$domain=~euads.org
/eureka-ads.
/eureka/eureka.js
/eureka_ban/*
/event.ng/*
/exads/*
/excellence/ads/*
/exchange_banner_
/exitpop.
/exitpopunder.
/exitpopunder_
/exitpopup.
/exitsplash.
/exo_bck_
/exoads/*
/exobanner.
/exoclick.$~script,domain=~exoclick.bamboohr.co.uk|~exoclick.com|~exoclick.kayako.com
/exoclick/*$domain=~exoclick.com
/exoclickright.
/exoclickright1.
/exoclickright2.
/exoclickright3.
/exopopunderdesktop.js
/exosrvcode-
/expads-
/expandable_ad.php
/expandable_ad?
/expandingads.
/expandy-ads.
/expop.js
/exports/tour/*$third-party
/exports/tour_20/*
/ext/adform-
/ext/ads/*
/ext_ads.
/extadv/*
/extendedadvert.
/external/ad.
/external/ad/*
/external/ads/*
/external_ad?
```

Which projects use filter lists?

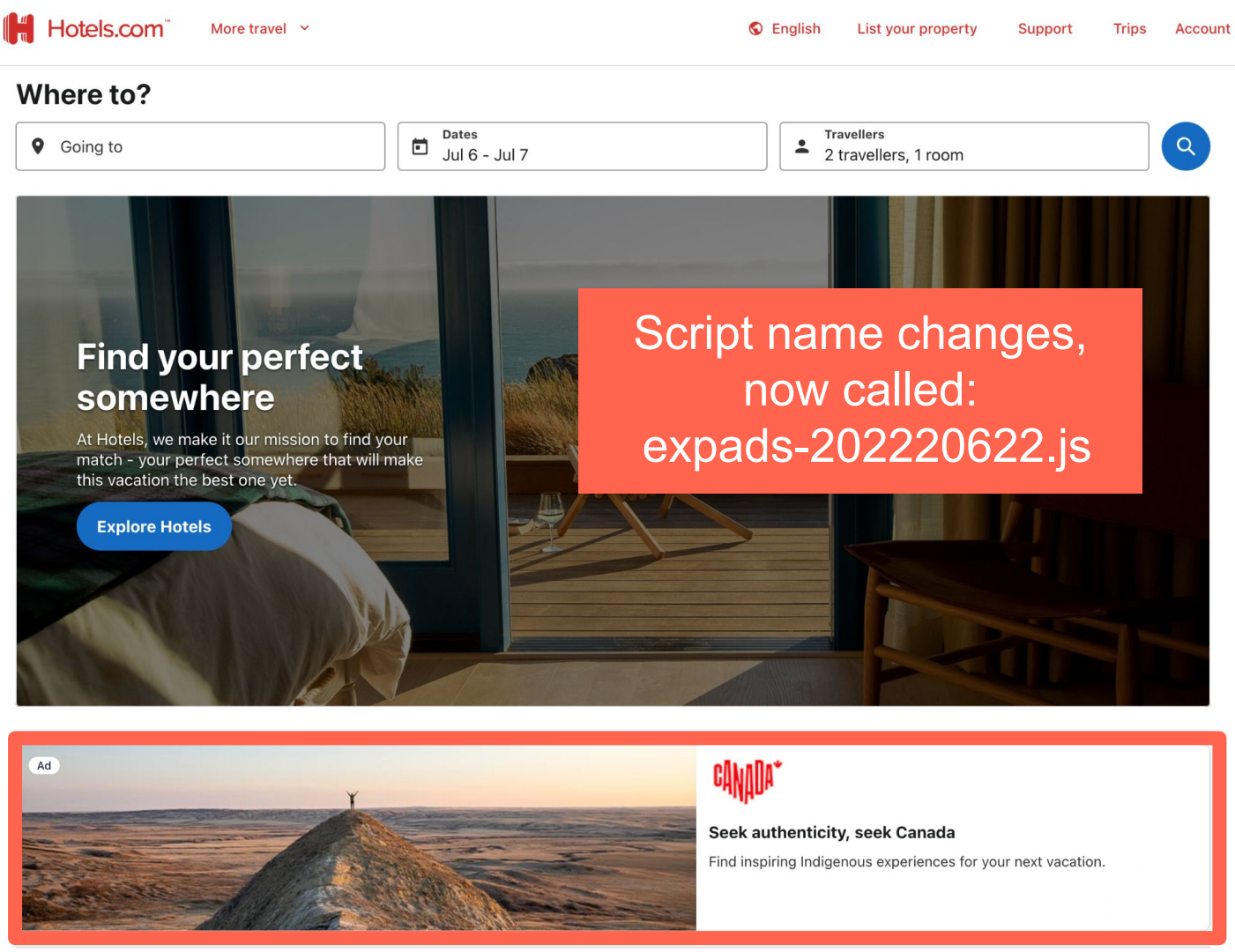
Brave, ABP, uBO, Chromium, AdGuard ...

If you've used an adblocker, you've used a filter list

Why are filter lists hard to maintain?

Why are filter lists hard to maintain?

1. Rules go obsolete



Script name changes, now called: expads-202220622.js

But filter rule is still: /expads-blocked.js

Not blocked :(

Why are filter lists hard to maintain?

1. Rules go obsolete
    a. False sense of security
    b. Can’t just block /ad*.js
    c. Uniquely bad for privacy-focused product

How do we identify adblock evasion?

1. Create signatures of tracker script execution and use that to catch obfuscated, renamed or inlined scripts
2. Detecting Filter List Evasion With Event-Loop-Turn Granularity JavaScript Signatures (IEEE Security & Privacy 2021)
3. Shipped new filter list rules

Why are filter lists hard to maintain?

1. Rules go obsolete
2. Rules can break sites across the Web

```
<html>
…
<-- Defines track_user() -->
<script src="https://evil.com/track.js"></script>
<script>
  // Call tracking function & pass setup function as callback
  track_user(setup_page);
</script>
<body>
  <p id="setup-body"></p> <!-- Empty until setup_page() runs -->
</body>
</html>
```

Let’s just block track.js!

… would break the website :(

Why are filter lists hard to maintain?

1. Rules go obsolete
2. Rules can break sites across the Web
    a. Retaining & acquiring users is crucial.

Why are filter lists hard to maintain?

1. Rules go obsolete
2. Rules can break sites across the Web
    a. Retaining & acquiring users is crucial.
    b. Users should not be expected to be technical.

Why are filter lists hard to maintain?

1. Rules go obsolete
2. Rules can break sites across the Web
    a. Retaining & acquiring users is crucial.
    b. Users should not be expected to be technical.
    c. Lots of users in places where folks might not have latest hardware.

Creating privacy-preserving script replacements

1. Solve the privacy-v.s.-compatibility trade-off by automating the creation of privacy-preserving implementations of tracking libraries.

```
<html>
…
<-- Browser internally redirects to sugarcoated_track.js -->
<script src="https://evil.com/track.js"></script>
<script>
  // Call tracking function but now from sugarcoated_track.js!
  track_user(setup_page);
</script>
<body>
  <p id="setup-body"></p> <!-- Empty until setup_page() runs -->
</body>
</html>
```

Creating privacy-preserving script replacements

1. Solve the privacy-v.s.-compatibility trade-off by automating the creation of privacy-preserving implementations of tracking libraries.
2. SugarCoat: Programmatically Generating Privacy-Preserving, Web-Compatible Resource Replacements for Content Blocking (CCS 2021)

Mitigating website breakage caused by content blocking

1. Blocked or Broken? Automatically Detecting When Privacy Interventions Break Websites (PETS 2022, Issue 4)
2. Predicting when a given filter list rule will break a website

SugarCoat Engineering: Shipping and rolling back

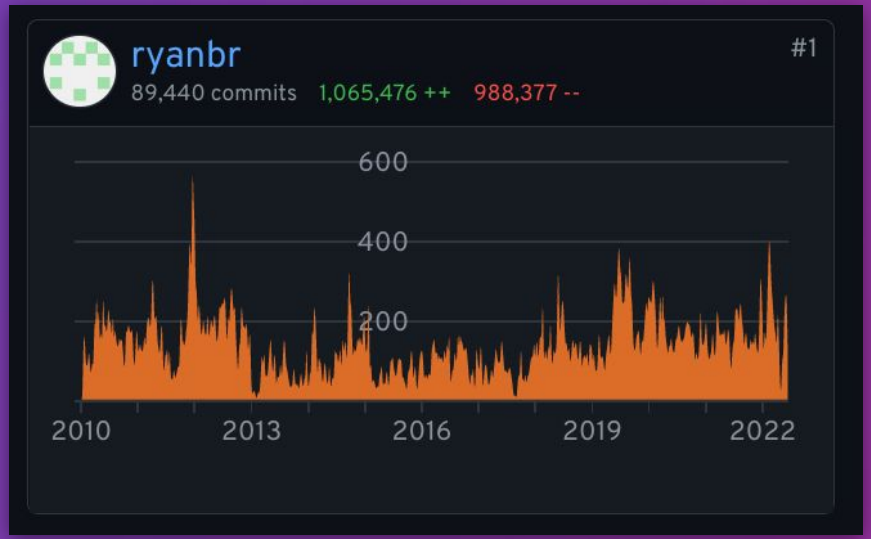
1. Paper -> engineering -> shipped!
2. Reports of browser crashing on devices with very low RAM
3. Rollback

Supporting the filter list community

Supporting Filter List Authorship

Don’t take the community’s work for granted!

Give back whenever possible.

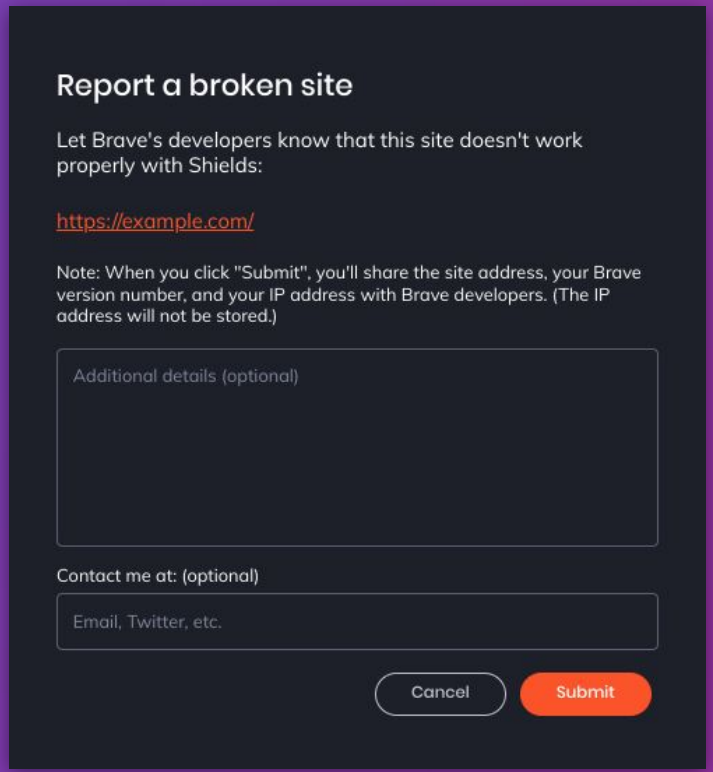


Ryan Brown is the #1 contributor to the EasyList repository on GitHub

Engagement with Active Users

- Feedback channels on:
  - GitHub
  - Twitter
  - Reddit
  - Brave Community Forums
- Nightly and Beta releases for power users

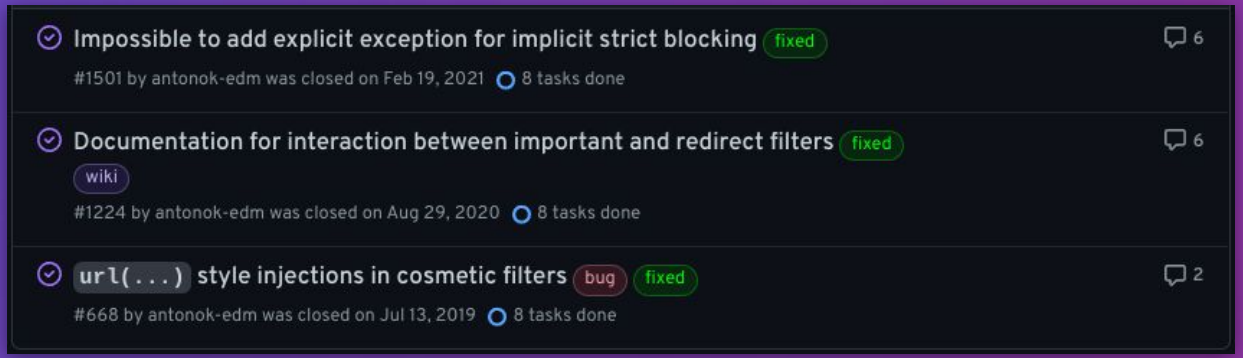
In-Browser Reports



Collaboration with Other Content Blockers

Content blockers share a lot of features and infrastructure.

They're stronger when they work together!



A handful of issues we've opened against uBlock Origin

Open Source

- https://github.com/brave/. . .

  adblock-rust
  adblock-lists
  adblock-resources
  brave-browser
  brave-core

- Adblock engine used in Qutebrowser, Angelfish Browser, others

Takeaways & goodbyes

1. Privacy is meaningless without compatibility.
2. Research ⇒ production is hard. Perf test! Have a rollback plan!
3. Set up feedback channels.
4. Users are delighted by fast fixes!
5. Build trust with the community by hiring & contributing back.


@antonok@fosstodon.org


@shivan_kaul