OpenStreetMap Ops Team
@osm_tech@en.osm.town

If you write about the messy reality behind "free" internet services: we're seeing #OpenStreetMap hammered by scrapers hiding behind residential proxy/embedded-SDK networks. We're a volunteer-run service and the costs are real. We'd love to talk to a journalist about what we're seeing + how we're responding. #AI #Bots #Abuse

Jan 27, 2026, 07:45 PM · 🌐

1.7K boosts · 12 quotes · 1.1K favorites

↩ ⇄ ☆ 🔖 •••

Derick Rethans 🌐 22h
@derickr@phpc.social

@osm_tech Maybe one for you @ricmac ?

↩ 0 ⇄ ☆ 🔖 •••

Baloo Uriza ☾ 19h
@BalooUriza@social.tulsa.ok.us

@osm_tech I wonder if there's a way to fail2ban requests coming in faster than typically found in human requests.

↩ 3 ⇄ ☆ 🔖 •••

OpenStreetMap Ops Team ☾ 18h
@osm_tech

@BalooUriza We use fail2ban to handle some of this with custom rules, but eventually fail2ban becomes a bottleneck after 100,000 IP addresses.

↩ 2 ⇄ ☆ 🔖 •••

Cassandrich ☾ 16h
@dalias@hachyderm.io

@osm_tech @BalooUriza For IPv4, a bitmask of the entire address space is a viable "efficient" implementation of blocking. I wonder if

there are tools that can do it that way rather than needing a gigantic list.

↩ 3   ⇄   ☆   🔖   •••

Cassandrich ☾ 16h
@dalias@hachyderm.io

@osm_tech @BalooUriza Like, a bitmask of IPv4 space is several times smaller than a Chrome instance. 🙃 🤡

↩ 0   ⇄   ☆   🔖   •••

MAgetröte ☾ 1h
@magezwitscher@det.social

@dalias @BalooUriza But that is one of the points @osm_tech are making in their post. These crawlers resort to using massive amounts of "scrapers hiding behind residential proxy/embedded-SDK networks" - meaning they are using Adware-infested phones all over the world for their scraping attaks. So banning IP ranges won't help much. Playing cat-and-mouse with these scrapers is resource intensive, which is increasingly hard for FOSS projects and is also driving up cost for commercial offerings.

↩ 1   ⇄   ☆   🔖   •••

Cassandrich ☾ 1h
@dalias@hachyderm.io

@magezwitscher @BalooUriza @osm_tech Not ranges. Just the single IP, and a short-lived ban. All you need to do is get them down from thousands of requests per minute to one request per hour (because they get banned for an hour each time they start again).

↩ 0   ⇄   ☆   🔖   •••

Nils Goroll 🕊 ☾ 9h
@slink@fosstodon.org

@dalias @osm_tech @BalooUriza we have a very efficient implementation in #vinylcache (formerly #varnishcache)

↩ 0   ⇄   ☆   🔖   •••

Kees de Kooter 🍉 🌐 19h
@kdekooter@mastodon.social

@osm_tech @schellevis @laurensvhg ?

↩ 0   ⇄   ☆   🔖   •••

fooxl   🌐 19h
@fooxl@fosstodon.org

@osm_tech @suka_hiroaki

↩ 0   ⇄   ☆   🔖   •••

Luciano Mieldazis   🌐 20h
@LMieldazis@mastodon.social

@osm_tech @geerlingguy I would 100% watch a video about that. Just throwing it out there :)

↩ 1   ⇄   ☆   🔖   •••

OpenStreetMap Ops Team   🌐 18h
@osm_tech

@LMieldazis @geerlingguy oooh do we get to show him our out-of-band (remote access) Raspberry Pi with dual power feeds, 4G modem and loads of serial connections? Saved our skin a good few times.

↩ 1   ⇄   ☆   🔖   •••

Jeff Geerling   🌐 15h
@geerlingguy@mastodon.social

@osm_tech @LMieldazis would love to talk maps ops! I've seen many projects wrapping in map data and adding scripts to dl entire regions

↩ 0   ⇄   ☆   🔖   •••

MAgetröte   ☾ 20h
@magezwitscher@det.social

@osm_tech
@leah just recently discussed the same topic from the perspective of a web hosting provider in the @chaosradio podcast.

↩ 0   ⇄   ☆   🔖   •••

🔍

📈 Trending

🌐 Live feeds

ⓘ About

Mastodon is the best way to keep up with what's happening.

Follow anyone across the fediverse and see it all in chronological order. No algorithms, ads, or clickbait in sight.

Create account

Login

Jorge Sanz
@jorgesanz@mapstodon.space
🌙 20h

@osm_tech maybe @civio @dcabo can be interested or help finding someone

↩ 1   ⇄   ☆   🔖   •••

David Cabo
@dcabo@mastodon.social
🌙 7h

@jorgesanz @osm_tech @civio hmm, it doesn't fit in Civio's scope I'm afraid. But it's definitely an issue I'm aware of, it's worse now with all the AI scrapers and I wonder if we should block them all, they flood my apps too :/ Maybe the 404 Media guys would be interested in this? 404media.co/ai-scraping-bots-a…

> 404 Media · Jun 17, 2025
>
> **AI Scraping Bots Are Breaking O…**
>
> "This is a moment where that community feel…

↩ 0   ⇄   ☆   🔖   •••

Claire
@ClaireH@blahaj.zone
🌐 20h

@osm_tech@en.osm.town 404 media might be interested in this - they've been doing a lot of pieces about the impact of AI

↩ 1   ⇄   ☆   🔖   •••

Fluffy Kitty Cat
@fluffykittycat@furry.engineer
🌐 18h

@ClaireH @osm_tech @404mediaco here's a good story idea, OSM is having issues with AI scrapers

↩ 0   ⇄   ☆   🔖   •••

Codeberg
@Codeberg@social.anoxinon.de
🌙 20h

@osm_tech You're absolutely not alone, we wish you good luck! ~n

↩ 0   ⇄   ☆   🔖   •••

Mastodon is the best way to keep up with what's happening.

Follow anyone across the fediverse and see it all in chronological order. No algorithms, ads, or clickbait in sight.

Create account

Login

Fooker
@Fooker@infosec.exchange
🌐 20h

@osm_tech @dangoodin maybe this might be up your alley? If not you may know someone appriopriate

↩ 0   🔁   ☆   🔖   •••

---

draNgNon
@draNgNon@hachyderm.io
🌐 20h

@osm_tech so I recently read a couple writeups from @briankrebs about malware and residential proxies

↩ 0   🔁   ☆   🔖   •••

◈
@utf_7@mastodon.social
🌐 21h

@osm_tech what is a embedded-Sdk network?

↩ 2   🔁   ☆   🔖   •••

AliveDevil
@AliveDevil@tauri.earth
🌐 20h

@utf_7 @osm_tech

App developers can embed some "Sdk" into their apps or games.
The developer receives money.
The "Sdk"-provider proxies requests through these apps and games, to gain residential IPs.
And scrapers can buy these services, to tunnel their requests from residential IPs.

↩ 3   🔁   ☆   🔖   •••

AliveDevil
@AliveDevil@tauri.earth
🌙 20h

> Viewer Discretion, example of these Sdks
> **Show more**

↩ 1   🔁   ☆   🔖   •••

Cassandrich
@dalias@hachyderm.io

⌄ 16h

@AliveDevil @utf_7 @osm_tech So ridiculous that Google and Apple won't just permaban any developer embedding one of these "SDKs".

↩ 1   ⇄   ☆   🔖   ⋯

AliveDevil
@AliveDevil@tauri.earth

⌄ 16h.*

@dalias I'd wish for them to enforce policies, but they get Ad- and IAP-revenue, so why bother.

Also, these "Sdks" probably have kill-switches (or rather, delayed activation) built-in, to not immediately contact their C&C servers.

↩ 1   ⇄   ☆   🔖   ⋯

Cassandrich
@dalias@hachyderm.io

⌄ 16h

@AliveDevil Yes but they could still be banned when caught. A few devs getting banned would be a big deterrent for others to ship this malware.

The right *technical* defense, however, is not to allow apps arbitrary network access unless they're declared in the manifest as a "browser" or other "client software" that the user can use with any service they want (like IRC clients, mail clients, Mastodon clients, etc.).

Instead, the manifest should declare a single domain the app can contact, or multiple if the developer is willing to pay for more intensive vetting of them, and only allow network access to the declared domain(s).

↩ 1   ⇄   ☆   🔖   ⋯

�
@utf_7@mastodon.social

⌄ 11h

@dalias @AliveDevil dafuq? if so, "software development kit sounds" wrong in that contedt. this is plain malware.

imagine using an app and someone downloads child porn or classical torrent over your connection. how will you proof you're

innocent

↩ 0   ⇄   ☆   🔖   •••

inamruzui   🌐 12h
@Kuniti_shino@pounced-on.me

@AliveDevil @utf_7 @osm_tech basically botnet/malware

↩ 0   ⇄   ☆   🔖   •••

luksfarris   🌐 20h *
@luksfarris@ruby.social

@AliveDevil @utf_7 @osm_tech how do I know if any of my installed apps is doing this crap? Could that be just buried in terms of use? Or would I need to give it explicit consent?

↩ 1   ⇄   ☆   🔖   •••

AliveDevil   🌐 20h
@AliveDevil@tauri.earth

@luksfarris @utf_7

Probably terms of use, but this is so shady, that I doubt anyone would even bother disclosing this.
Best you can do: Monitor network traffic, and use DNS block lists for these known proxy services.

They definitely won't ask you for consent.
The only way to know an app _doesn't_ use these services is checking for the "requires internet access"-flag in AppStores, but that is basically futile, as most apps require internet access for … something.

↩ 0   ⇄   ☆   🔖   •••

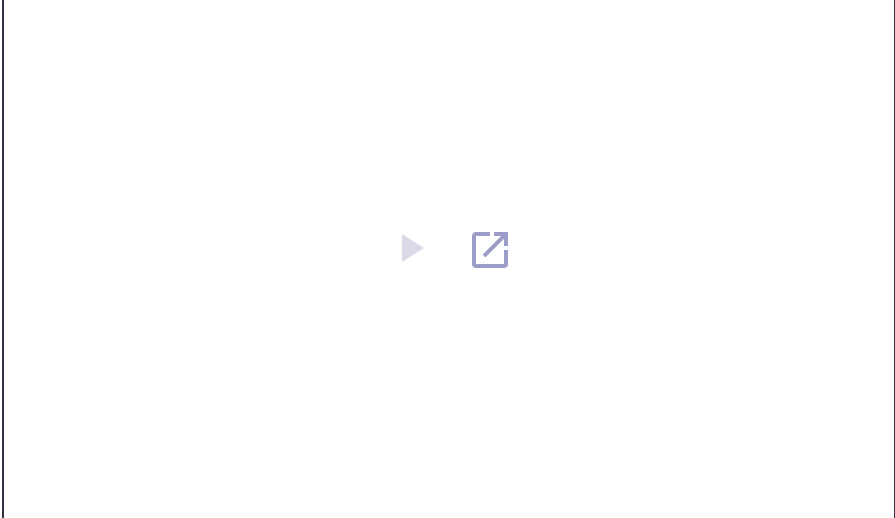OpenStreetMap Ops Team   🌐 18h
@osm_tech

@utf_7 See youtube.com/watch?v=qhie14YKa-…

YouTube

2025: Servers on Fire: Keeping OpenStreetMap Online

By **State of the Map Europe**

↩ 0   ⇄   ☆   🔖   •••

Jonathan Corbet   🌐 21h
@corbet@social.kernel.org

@osm_tech You are definitely not alone: https://lwn.net/Articles/1008897/ The situation is not sustainable but I'm not sure what we do about it beyond waiting for the AI bubble to burst.

LWN.net

Fighting the AI scraperbot scour...

There are many challenges involved with runn...

↩ 0   ⇄   ☆   🔖   •••

zuggamasta   🌐 21h
@zuggamasta@merveilles.town

@osm_tech hey @emanuelmaiberg this sounds interesting and might add yet another part to your reporting on museums and others being scraped.

↩ 0   ⇄   ☆   🔖   •••

Internet Rando   ☾ 21h
@mousey@seattlematrix.org

Mastodon is the best way to keep up with what's happening.

Follow anyone across the fediverse and see it all in chronological order. No algorithms, ads, or clickbait in sight.

Create account

Login

@osm_tech You guys are heroes, nonetheless!! ✊

⤺ 0 ⇄ ★ 🔖 •••

Raineer 🦆 🌐 21h
@raineer@frontrange.co

@osm_tech cc @404mediaco as well

⤺ 0 ⇄ ★ 🔖 •••

zuggamasta 🌐 21h
@zuggamasta@merveilles.town

@osm_tech this sounds right up 404 medias alley.

They all have contacts and have reported on museums and Wikipedia having similar issues.

404media.co/about/

404 Media · Jul 24, 2023

Who We Are

Who is 404 Media?

⤺ 0 ⇄ ★ 🔖 •••
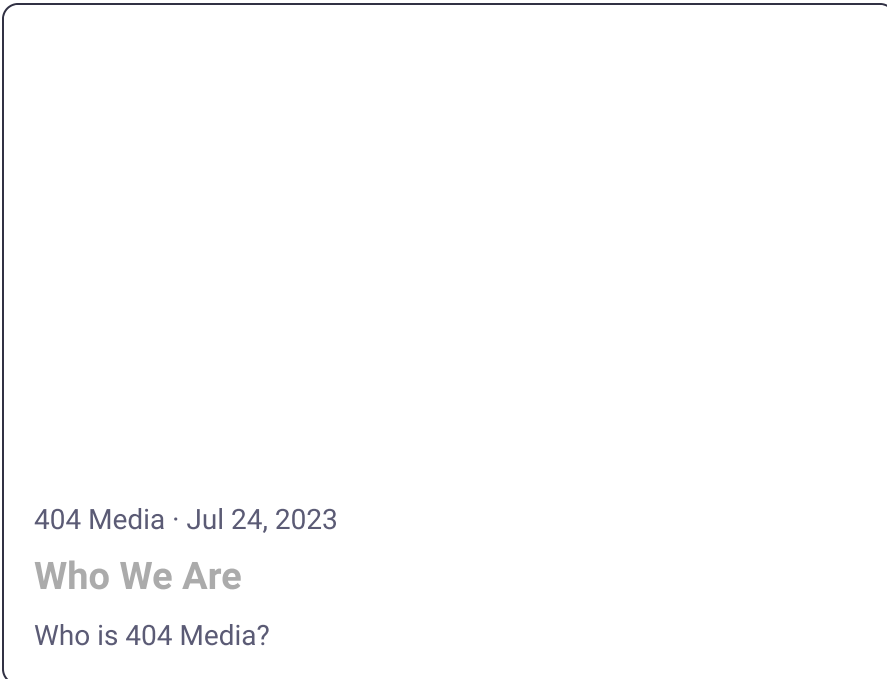
Ray McCarthy 🌐 21h
@raymaccarthy@mastodon.ie

@osm_tech
Decades ago when I read Dune I thought the Butlerian Jihad against computers was the silliest thing in it.

Suddenly it makes sense. The sooner the LLM AI bubble bursts the better!

↩ 0   ⇄   ★   🔖   •••

🔍

Kierkethumbs up convincingly   🌐 21h
@Kierkegaanks@beige.party

@osm_tech definitely not google's doing

↩ 0   ⇄   ★   🔖   •••

Nick Kerker   🌐 21h
@thekerker@mstdn.social

@osm_tech @josephcox

↩ 0   ⇄   ★   🔖   •••

Dr. Christopher Kunz   🌐 21h
@christopherkunz@chaos.social

@osm_tech Oh, that sounds interesting! If you want, I'd be interested in talking about this.

↩ 0   ⇄   ★   🔖   •••

Jernej Simončič �   🌐 21h
@jernej__s@infosec.exchange

@osm_tech I'm administering a web server for a client that has about 50 web sites. Every few days they get hammered by residential proxy IPs for a few hours, so I finally installed Anubis.

↩ 0   ⇄   ★   🔖   •••

BrianKrebs   🌐 21h
@briankrebs@infosec.exchange

@osm_tech Hey. Sorry to hear about that. Drop me a line on Signal? username: briankrebs.07. Thanks!

↩ 0   ⇄   ★   🔖   •••

João Tiago Rebelo (NAFO J-121)   🌙 21h
@jt_rebelo@ciberlandia.pt

@osm_tech I think @briankrebs could be interested.

↩ 0   🔁   ☆   🔖   ⋯

InsertUser   🌐 21h
@InsertUser

@osm_tech The proxy SDK providers need to be treated like the DDOS providers they are and prosecuted.

↩ 1   🔁   ☆   🔖   ⋯

Andrew Zonenberg   🌐 16h .*
@azonenberg@ioc.exchange

@InsertUser @osm_tech Pulling them from app stores and banning developers of the SDKs would be a good start. Save the criminal charges for after the damage control is done.

↩ 0   🔁   ☆   🔖   ⋯

Isaac   🌐 21h
@isaac@hachyderm.io

@osm_tech @josephcox Open source maps project dealing with AI scrapers, requesting journalists who might be interested ☝️

↩ 0   🔁   ☆   🔖   ⋯

Sven222   🌐 21h
@sven222@soc.hardwarepunk.de

@osm_tech muss schon wieder @evawolfangel hier taggen.

↩ 0   🔁   ☆   🔖   ⋯

Vincent Bergeot   🌙 21h
@vinber@mstdn.fr

@osm_tech @mathildesaliou perhaps ?

↩ 0   🔁   ☆   🔖   ⋯

blub   🌙 22h
@blub@norden.social

@osm_tech Or @heiseonline ?

↩ 1   🔁   ☆   🔖   ⋯

Dr. Christopher Kunz
@christopherkunz@chaos.social ☾ 9h

@blub @osm_tech @heiseonline Yeah I already replied.

↩ 0   ⇄   ☆   🔖   …

Jan Kröger
@ankrjoe@social.tchncs.de ☾ 22h

@osm_tech
Could that be a topic for @evawolfangel ?

↩ 0   ⇄   ☆   🔖   …

Floris
@floris@freiburg.social 🌐 22h

Vielleicht ist das ein Thema für die @lagedernation?

↩ 1   ⇄   ☆   🔖   …

Lage der Nation
@lagedernation@chaos.social 🌐 4h

@floris
@osm_tech Hi, please get in touch, we've covered OSM many times before and would love to learn more: team a lagedernation org

↩ 0   ⇄   ☆   🔖   …

Pietervdvn
@pietervdvn 🌐 22h

@osm_tech ugh.... why don't they use the exports...

↩ 1   ⇄   ☆   🔖   …

InsertUser
@InsertUser 🌐 21h

@pietervdvn Because that would involve a human using their brains or having a shred of conscience and those both go against the basic principles of the companies doing this.

@osm_tech

1

Cassandrich 🌐 16h

Mastodon is the best way to keep up with what's happening.

Follow anyone across the fediverse and see it all in chronological order. No algorithms, ads, or clickbait in sight.

Create account

Login