

Shy Guys: A Light-Weight Approach to Detecting Robots on Websites

Rémi Van Boxem, Tom Barbette, Cristel Pelsser, Ramin Sadre

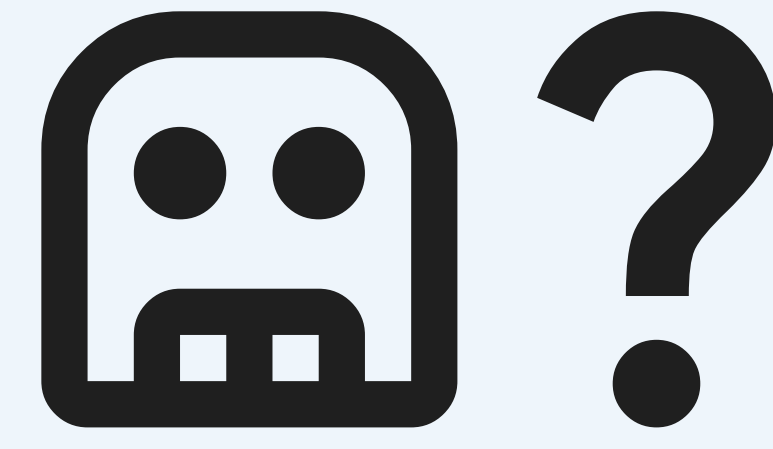
Motivation

In 2025:⁽¹⁾
50% x15

Requests are made by bots Expansion in AI-crawler requests

Detecting bots is important for multiple reasons

- ⚠️ Avoiding outage and (unintentional) DDoS attacks
- 💰 Reduce traffic cost by removing unwanted load
- 🚫 Block AI crawler to steal content without proper authorization
- 🔒 Block automated vulnerability scanning and exploitation



worse result, better UX, easy to set up

good result, bad UX, maybe resource hungry

Passive

- 🔍 IP Reputation
- 👤 Behaviour analysis
- 🌐 User-Agent string

Active
Proof of Work

reCAPTCHA (3) easy!

Anubis (2) Compute hash in browser

"Use the arrows to rotate the animal to face the direction of the hand"

huh?!

We need a passive method with results similar to an active one!

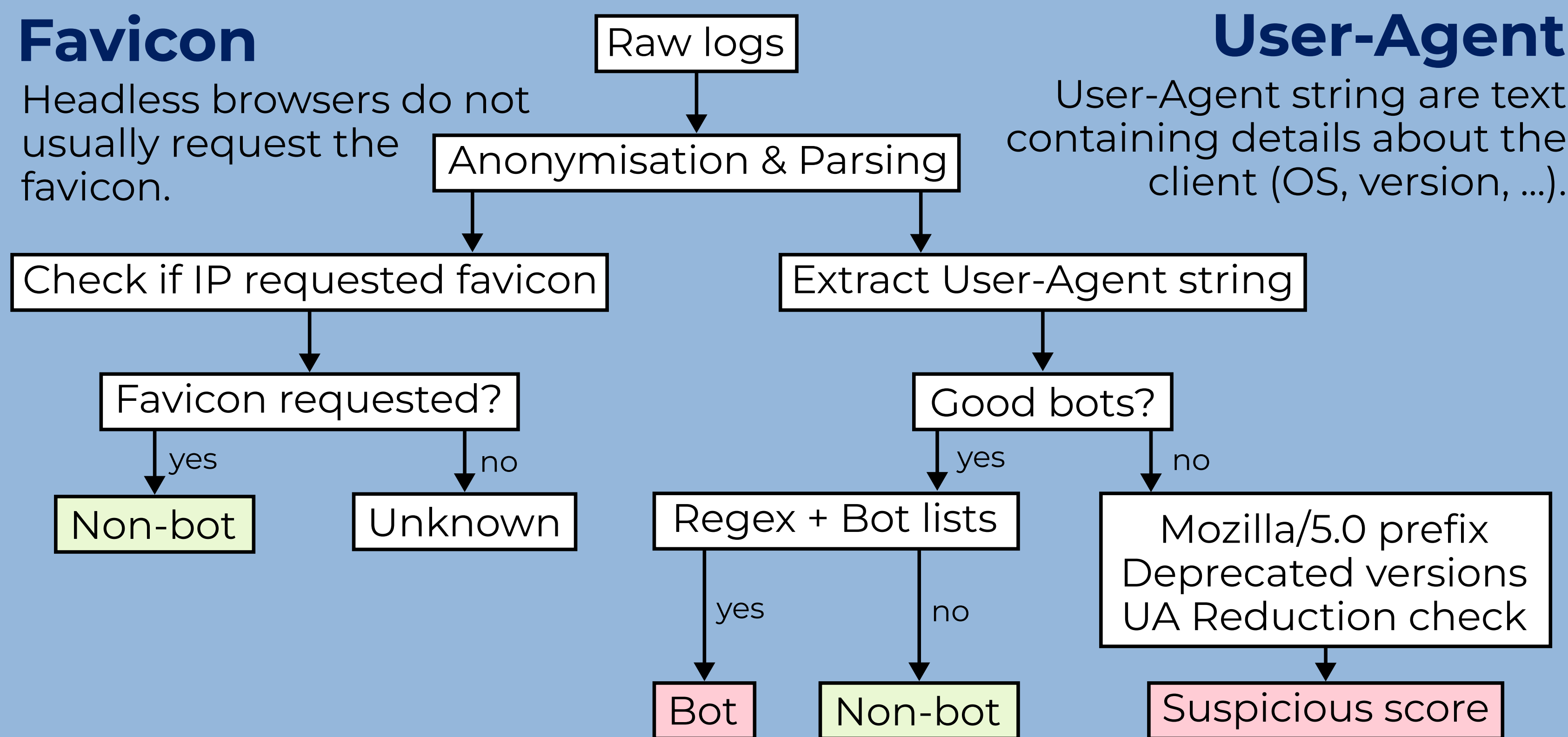
Methodology

Favicon

Headless browsers do not usually request the favicon.

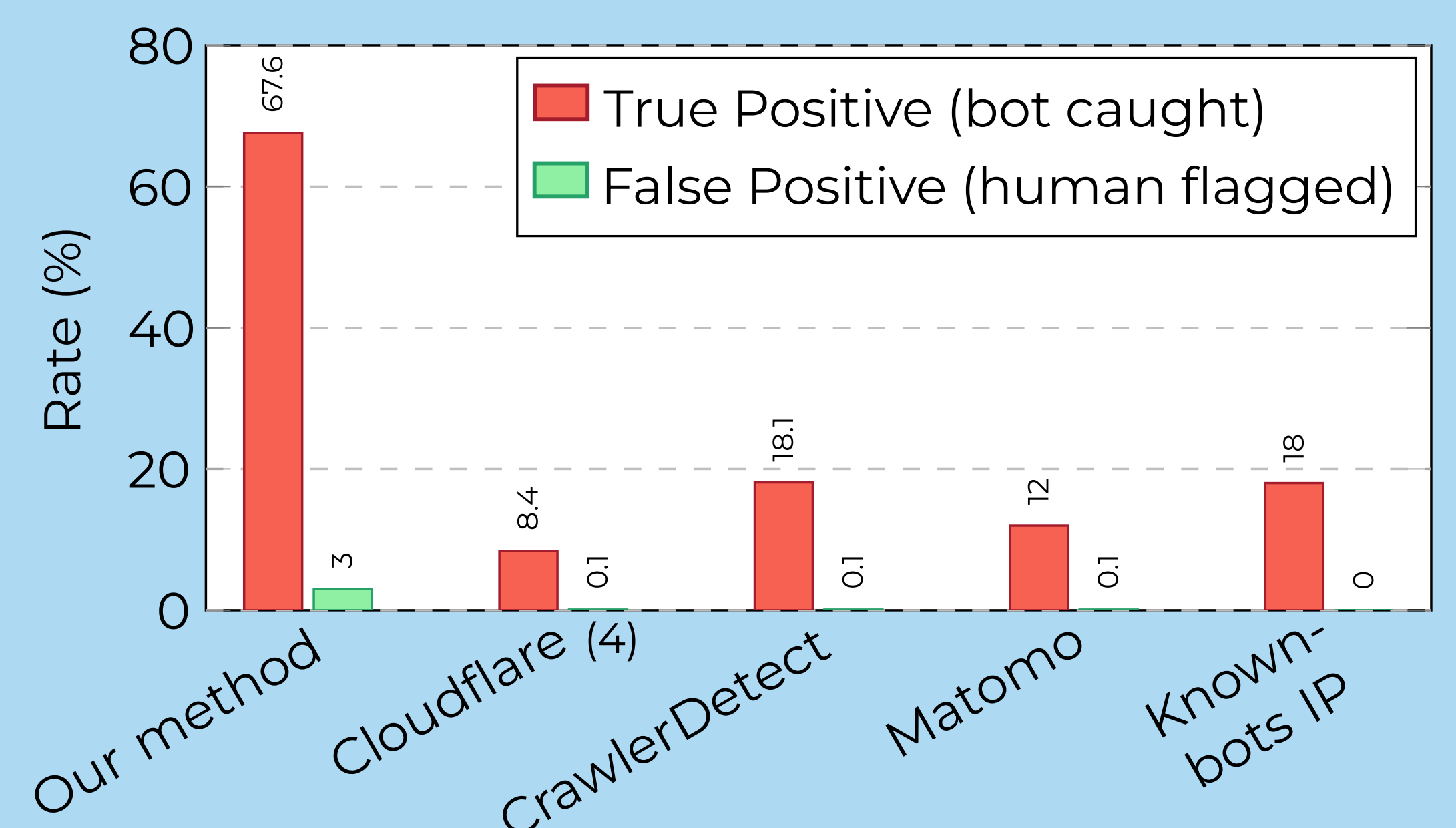
User-Agent

User-Agent string are text containing details about the client (OS, version, ...).



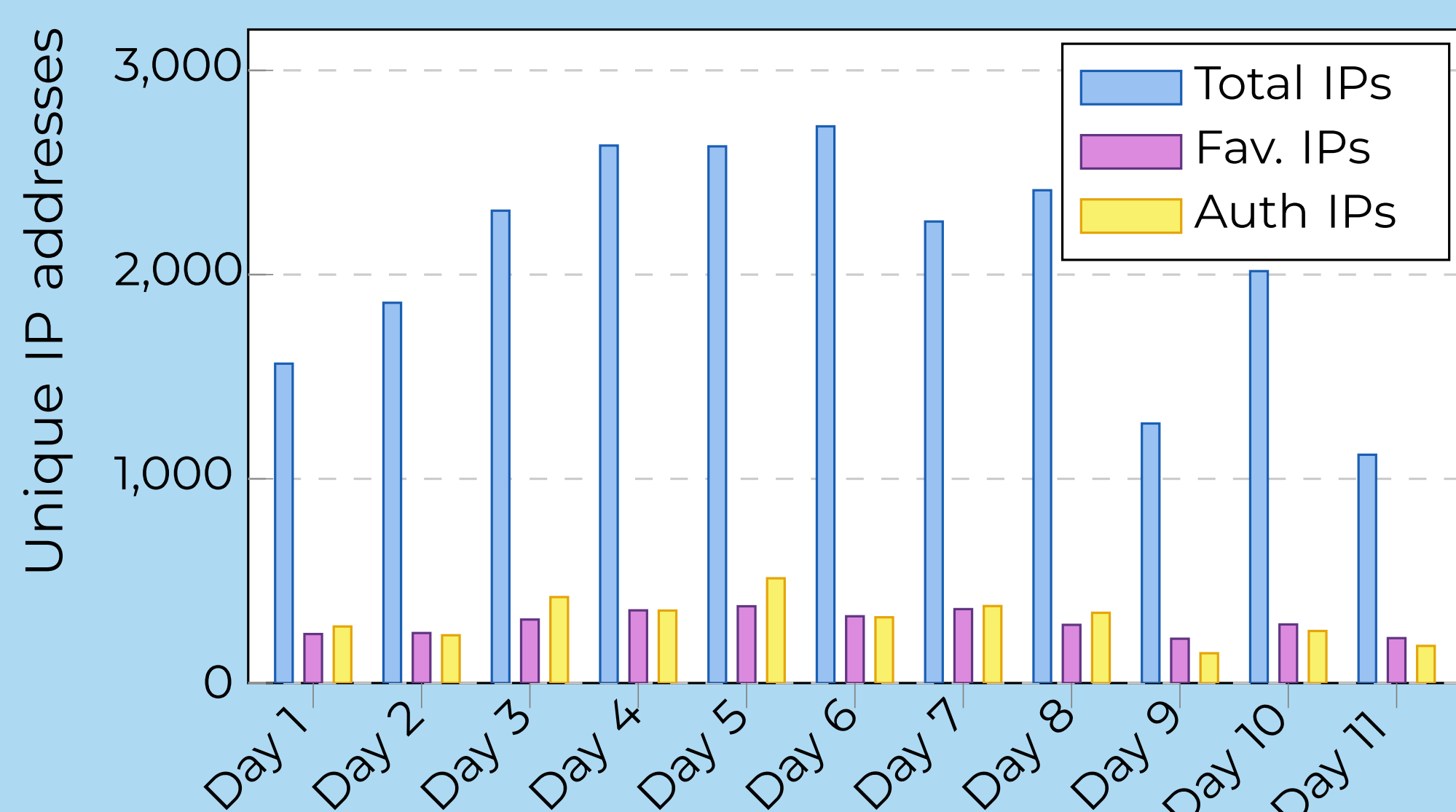
User-Agent evaluation

491 unique bot user-agents flagged (56.2%) vs 58 by Cloudflare (6.6%), an 8× increase in coverage.

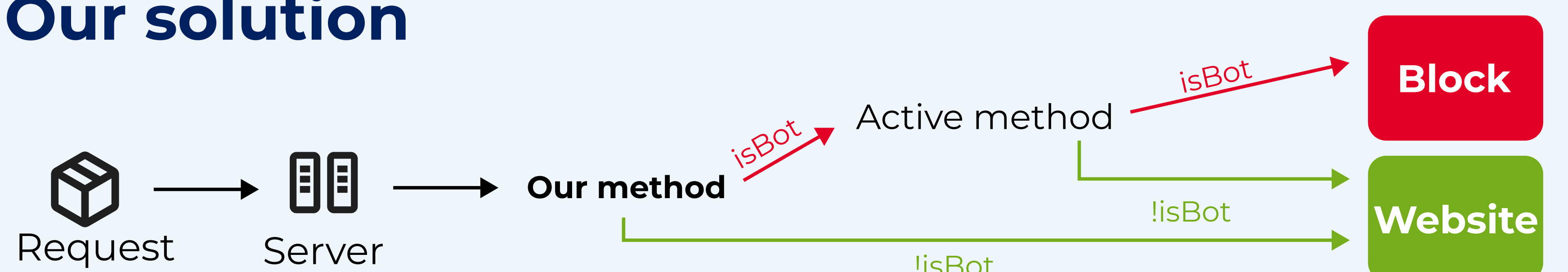


Favicon validation

Daily unique IP addresses issuing favicon requests and auth IP over the observation period



Our solution



Ground truth built from two reliable sources: humans from authenticated LMS traffic, bots from an unindexed honeypot (71% of traffic from Azure, Cloudflare, AWS, DigitalOcean, GCP).

Bots are half of the web. Existing tools catch under 20% of them. **We catch 67.6% passively, from logs alone without compromising user experience.**

References:

- (1) David Belson, "The 2025 Cloudflare Radar Year in Review- the Rise of AI, Post-Quantum, and Record-Breaking DDoS Attacks," The Cloudflare Blog, December 15, 2025
- (2) Luis Von Ahn et al., "CAPTCHA: Using Hard AI Problems for Security," in Advances in Cryptology — EUROCRYPT 2003, ed. Gerhard Goos, Juris Hartmanis, and Jan Van Leeuwen, vol. 2656, ed. Eli Biham, Lecture Notes in Computer Science (Berlin, Heidelberg: Springer Berlin Heidelberg, 2003), 294–311, https://doi.org/10.1007/3-540-39200-9_18.
- (3) TecharoHQ/Anubis, Go, March 17, 2025; Techaro, released February 16, 2026, <https://github.com/TecharoHQ/anubis>.
- (4) "Cloudflare Bot Management Solution," accessed March 25, 2026, <https://www.cloudflare.com/en-gb/application-services/products/bot-management/>.