REPORT

datacentrix

CLOUDFLARE

Better defenses against modern bots

Bot status quo

Bots target most sites on the Internet with powerful attacks. Enterprises, small businesses, and even individuals deserve equally powerful responses to these attacks. This guide is an overview of Cloudflare's observations, best practices, and advice in bot protection.

But before examining bot attacks, let's quantify just how serious the challenge currently is.

Cloudflare Radar, our team that tracks internet trends, keeps a close eye on automated traffic — and shows that bots represented 43.8 percent of the requests we saw on our network in the first half of 2021. This is up from 39.5 percent in 2020.

Bot Traffic

43.8%
of the traffic flowing through Cloudflare is automated

10.9%
year over year increase

4.5%
of total traffic is verified (good) bots

How bots target us

Bots carry out attacks that are manifold and increasingly problematic for site owners.

Bot attacks

	Attack details	Business impact
Credential stuffing/Account takeover	Bots target login pages with stolen credentials from the dark web to attempt account access.	This leads to account takeover that shakes customer confidence in a company or service.
Inventory hoarding	Bots target checkout/shopping cart pages and grab limited inventory in seconds.	Customers excited for a new product are left frustrated and less likely to return.
Credit card stuffing	Bots target checkout/shopping cart pages with stolen credit card details from the dark web to validate cards to make illicit purchases.	Merchants suffer financial losses when cardholders dispute fraudulent purchases.
Price and content scraping	Valuable, proprietary pricing information or unique content is taken via scraper bots — including content behind forms.	Competitors use scraped data to undercut on price or mount competitive campaigns.
Marketing spam	Bots submit junk data in web forms, damage marketing campaigns, and scrape content from marketing pages.	Sales and go-to-market initiatives are derailed, putting quarterly campaigns at risk while unique content is used by competitors.

Credential stuffing attacks

Cloudflare research shows login endpoints receive a disproportionate amount of automated traffic, likely evidence of sustained, bot-driven credential stuffing attacks. While 44 percent of global traffic is automated, 71 percent of login traffic is automated.

Evasion best practices

Bots have become easy and inexpensive to operate — meaning that site owners must prepare for a variety of attacks from many different sources. Developers online offer inexpensive data scraping, even bragging about the defenses they have been able to evade.

Moreover, many legitimate web testing automation tools like headless browsers unfortunately double as bot automation platforms. Headless browsers are run from the command line and don't use a GUI. You might have heard of Puppeteer, Playwright, Selenium or Phantomjs — these have all been twisted into lethal bot attack tools.

When organizations are targeted by bots, they attempt to fight them off with manual mitigations that can provide immediate, albeit temporary, relief. Organizations frequently start with IP blocking to prevent abuse from known, malicious IPs. IP blocks will often be combined with user-agent filters to detect and filter out clients that are being used by bots. Another approach uses rate limiting to throttle bot activity, delivering a respite to back-end resources.

If organizations are taking steps to slow bots, why are bots so effective?

Any mediocre bot attacker is aware of basic evasions that nullify the above efforts. More sophisticated mitigation approaches are needed to stay ahead of attacks.

Evading defenses

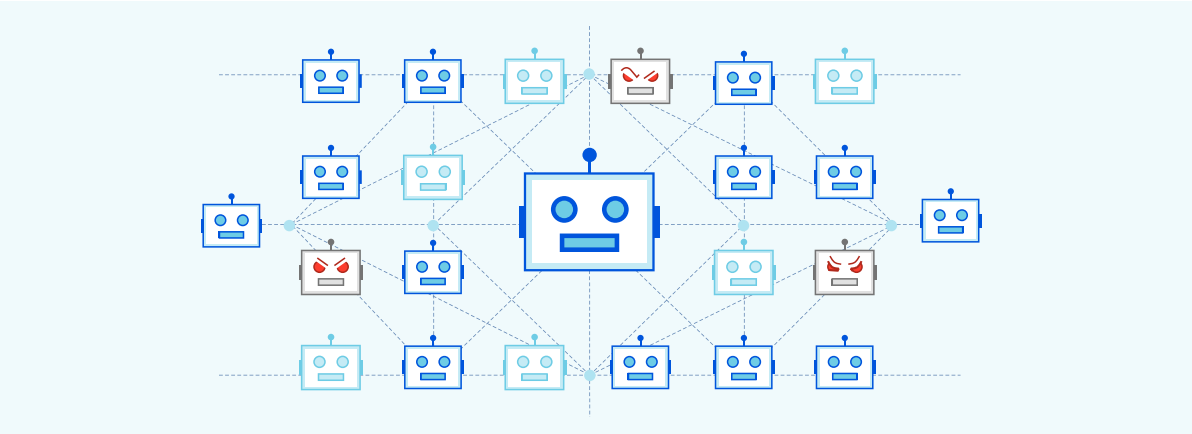
Four habits of highly effective bots

- **IP rotation/proxy use:** Prevent blocking by IP address using rotating residential IP addresses with an aim to circumvent CAPTCHAs.
- **Evasion packages and headless browsers:** Utilize evasion libraries, bypass tools and headless browsers that automate activity while appearing human-like.
- **Emulate people:** Regulate automated crawling speed to emulate humans while inserting random delays, clicks and cursor movements.
- **Fresh user-agents:** Call on multiple user-agents and rotate between them to spoof anti-bot controls. Update the user-agent pool frequently to keep them fresh.

As we see called out in blue above, attackers go to great lengths to avoid detection by obfuscating their device/browser attributes, varying their bot behavior and shifting infrastructure to avoid blocklists.

How to consider bot defenses

Since bots can evade basic defenses, organizations need to consider three important ways to counter the bots: layered detections, detailed analytics and visibility, and adaptive responses and challenges.



Detections

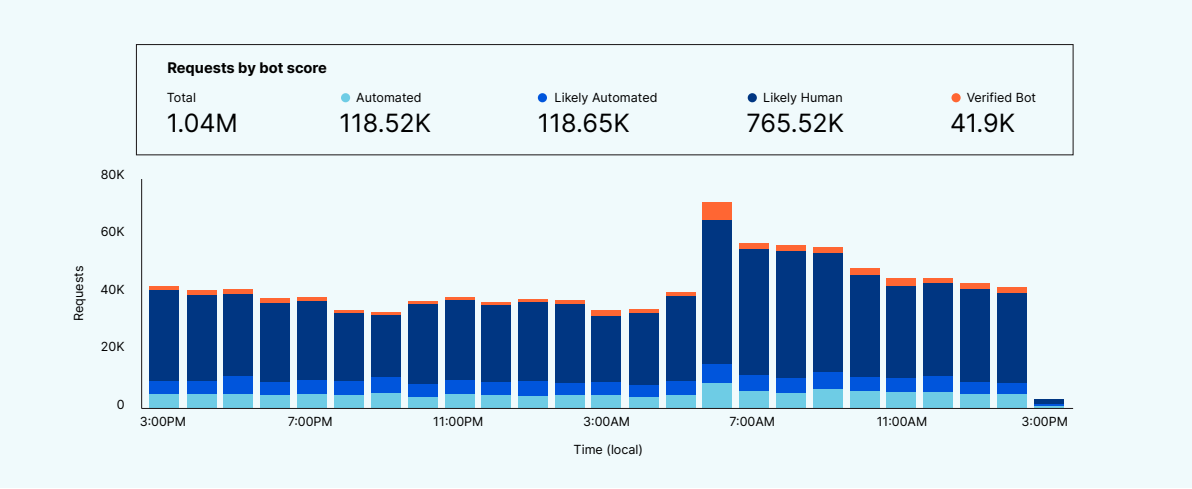
To counter the varied obfuscation approaches of sophisticated bots, a lattice of complementary detections is required. Let's look at how these detections identify bots.

- **Allow the good bots:** We begin by allowlisting the good, verified bots like Google and Bing.
- **Counter obfuscation with heuristics:** With an allowlist in place, we analyze requests with heuristics to scrutinize browser attributes and behavior. This will identify bots by comparing request attributes against multiple heuristic types and hundreds of rules based on hard-to-spoof attributes. Heuristics can detect tiny nuances in bot requests — and then flag these requests as automated.
- **Identify outliers with anomaly detection:** The next detection layer should identify legitimate user activity (specific to a particular site) and detect outlier activity. Anomaly detection should be resilient against metadata spoofing (e.g. fake user agents) and catch new bots without being explicitly trained on them. Each request should be tagged with an outlier score as a measure of certainty.
- **Detect suspicious requests with machine learning:** After identifying anomalies, supervised machine learning should further detect bots. Effective machine learning is trained on vast amounts of data to sharpen its detection effectiveness - the more training data, the better the detections. Machine learning engines should score every request on a scale, evaluating requests as likely automated, likely human - or perhaps in between.
- **Scrutinize clients with JavaScript challenges:** Advanced bot defenses often include JavaScript detection engines that gather more active signals from clients. By injecting client-side JavaScript, it is more difficult for clients to conceal malicious device fingerprints intended to spoof other detection engines. These detections will either flag a request as automated or not.

Visibility and analytics

Given 40-50 percent of Internet traffic comes from bots, bot defenses must be tied to useful analytics and visibility dashboards. This helps organizations determine the scale of bot issues and fine tune bot defenses. Dashboards should provide insights such as:

- **Traffic by type:** Understand if traffic is automated, likely automated, or human.
- **Request scores:** See requests based on bot score.
- **Bot tags:** Know whether a request is from Google or a bot framework.
- **Detection insights:** See which detection engines are most commonly scoring traffic.
- **Request attribute:** View more detailed information on specific IP addresses, user agents and more.



Responses and challenges

With detections and visibility in place, flexible responses and challenges give more granular control to keep business running smoothly with few false positives.

Organizations should have the ability to decide when CAPTCHA or JavaScript challenges are presented to potential bots, based on the bot scoring. Instead of a request being blocked outright, a challenge is served, thereby ensuring that each request is legitimate while avoiding false positives.

Also, the most innovative challenges call on serverless edge compute for custom responses. These can undermine the effectiveness of bot operations by serving fake content or dummy data - or any other desired response.

We have shared our thoughts on the necessary layers of a robust bot defense to better deflect bots attacks that comprise increasingly greater percentages of internet traffic. To learn more about Cloudflare Bot Management, please go to our website.