

Bot detection: how to identify and block bot traffic to your websites, mobile apps, and APIs

Bot detection is (or should be) a key security priority for any business with an online presence. About a third of the world's total web traffic is now made up of malicious bots, and bad bots are responsible for many of the most serious security threats that online businesses are facing today.

However, detecting bot traffic is harder than it has ever been. Bot developers are constantly finding new ways to circumvent the bot detection features of standard security solutions. And as they are now starting to make extensive use of artificial intelligence, efficient bot detection will soon be impossible without truly specialized know-how—and without artificial intelligence.

Let's take a look at the current state of malicious bot technologies and distribution, the key requirements for reliable bot detection, and what it will take to secure your online assets against bot threats in the near and mid-term future.

Summary



UNDER ATTACK?

4. [Server-side and client-side bot detection are both required](#)
5. [Bot protection on autopilot](#)

FREE TRIAL

EN

Bot detection is more difficult than ever

Distinguishing bots from human visitors has become a very complex task. Bot developers adopt new technologies faster than ever before, and they deliberately design their bots to bypass bot detection systems.

As such, over the last few years, bots have evolved dramatically. We can summarize this evolution by classifying them in four distinct generations:

Gen 1 bots are simple crawlers (often in-house scripts) that perform basic automated tasks, such as scraping information from a web page. Because they don't maintain a session cookie, they are easy to identify as bots.

Gen 2 bots are web crawlers such as Nutch and Scrapy. Thanks to the absence of JavaScript firing, these are also quite easy to detect.

Gen 3 bots represent a significant change: bots now look like browsers. Examples include PhantomJS and CasperJS. These bots herald the start of low and slow attacks: volume-based thresholding becomes ineffective. Detecting Gen 3 bots requires challenge tests and fingerprinting.

Gen 4 bots mimic human behavior (such as non-linear mouse movements) and can now appear to be a real user or hide inside a user session. They leverage headless browsers like Chrome Headless, and instrumentation frameworks like

**UNDER ATTACK?**

Bots of the most recent generation are **FREE TRIAL** almost **EN** indistinguishable from human visitors, and they are impossible to detect without **truly expert bot detection know-how**. They have prompted the need for tools that are able to determine the visitor's intent, rather than simply analyzing traffic volume and known bot signatures.

Moving forward, we expect the upcoming generation of bots to make massive use of artificial intelligence, which will make it even more challenging to spot them.

IP-centric bot detector solutions are no longer sufficient

In parallel with the technical evolution, bots are also distributed in increasingly elaborate ways in order to eschew detection.

Traditional security software, such as WAFs, tends to rely heavily on IP reputation to manage bots. This is based on the assumption that any form of malicious activity from an IP address means that all activity from that IP address is malicious.

This was efficient enough in the beginning, because most bot operators used to rely on data center proxies. Website owners could simply block data center proxy IP addresses, and their problem was solved. Known data center ISPs are easily identified, and public databases of those IPs are available from organizations like

AbuseIPDB in the US and **IP2LOCATION** in the EU.



UNDER ATTACK?

another exit node and attack from another IP. This made IP-based bot detection and protection much less effective.

Today, bot operators can easily and cheaply rotate through thousands or even millions of different IPs, thanks to [FREE TRIAL](#) as Luminati ^{EN} (the world's largest proxy service). Bots are increasingly distributed [via residential IPs](#), which benefit from excellent reputations, and where the requests they send are indistinguishable from those generated by regular users.

Bot developers also hijack vulnerable IoT devices, which are deployed in the millions, and malicious applications installed on the mobile devices of unsuspecting consumers.

As a result, **bot detector solutions** that rely heavily on IP blacklisting are simply not very effective anymore. Furthermore, as IPs are increasingly shared, IP-based filters carry a high risk of false positives: blocking legitimate users from accessing your website or mobile app.

IP reputation is still an important signal to take into account, but it must be combined with more sophisticated detection techniques.

Behavior-based bot detection is the way of the future


Because bots are now perfectly capable of mimicking human behavior, low and slow attacks that elude volumetric detection have become the norm. Similarly, the massive distribution of bots on residential and IoT IP addresses means that IP-based security systems are no longer of much help in the fight against malicious bot traffic.



UNDER ATTACK?



In order to help CTOs and CISOs protect their websites, mobile apps, and APIs from bad bot traffic, DataDome's bot detection software analyzes 100% of the requests that hit our customers' applications.

For each request, we collect and  more than 250 different events. Across our customer base, this accumulates to 600 billion events per day.

Known bots are detected via technical detection and validation: HTTP fingerprinting, known AI/custom rule pattern matching, and good bot authentication.

New threats represent the real challenge. They are identified via statistical and behavioral detection, using data from server-side fingerprints, a JS rendering engine, SDK inputs and session tracking.

We make extensive use of online machine learning, and our engine detects a new bad bot pattern every 10 milliseconds. That means 1.2 million new bots detected every day, automatically and in real time!

As soon as we detect a new threat, our algorithm is instantly updated and deployed to all our data centers, so that all our customers are protected against the new threat in real time.

Server-side and client-side bot detection are both required

The DataDome solution relies on a combination of client-side and server-side integration. The server-side module collects HTTP requests and fingerprints, analyzes every single request in real time through the DataDome AI detection engine, and makes real-time decisions.



UNDER ATTACK?

exclusively on server-side detection will be completely oblivious to these bots, unless they happen to display suspicious behavior.

Server-side fingerprinting must therefore be combined with client-side signals. The DataDome client-side module [FREE TRIAL](#) analyzes a wide range of browser, app and device features, as well as behavioral signals such as mouse movements and touch events, in order to accurately detect the most sophisticated bots.

Bot protection on autopilot

DataDome analyzes 100% of all requests to our customers' web servers, and uses both server-side and client-side event-related data to distinguish human users from bots. As a result, our customers can run their **bot management on autopilot**.

Whenever an attack is detected, no intervention is needed: the solution automatically applies the most appropriate response for each threat, including new threats.

In this way, bot management is taken off IT teams' hands so that they can spend their time and efforts on more strategic issues. But should they wish to, they can still adapt the response strategy: DataDome lets users fine-tune the configuration of their bot protection via a powerful custom rules engine.

Bot traffic data from the DataDome solution can also be seamlessly integrated into server logs and SIEM/SOC tools, as well as marketing and analytics tools like Google Analytics, Optimizely, Mixpanel, VWO and more.

In conclusion



UNDER ATTACK?



time behavioral analysis are how you can safeguard your digital assets from [DDoS attacks](#), [account takeover](#), [intensive scraping](#) and other bot-related threats.

If you'd like to know more, click [FREE TRIAL](#) to book [“Cost of bots: 13 hidden ways bots hurt your bottom line”](#)

Try DataDome free for 30 days

No credit cards. No commitment.

Install the module corresponding to your architecture and observe the automated traffic in real time for the next 30 days

REQUEST A DEMO

FREE TRIAL

US headquarters
33 Irving Pl
NY 10003
New York
(646) 893-0048

Free bot
12 de



UNDER ATTACK?

RESOURCES

- Bot Management
- Customer Stories
- Engineering
- Events
- Webinars

PRODUCT

- Overview
- New Features
- Integrations
- Documentation
- API server status
- Pricing

COMPANY

- About us
- Careers
- Press
- Blog
- Contact

DATA DOME

[PRODUCT](#)[CUSTOMERS](#)[PRICING](#)[RESOURCES](#)

© Copyright 2020 | DataDome | All Rights Reserved | [Terms of service](#)

[Docs](#)[Login](#)[FREE TRIAL](#)[EN](#) [FR](#)[UNDER ATTACK?](#)