imperva

phase pacific

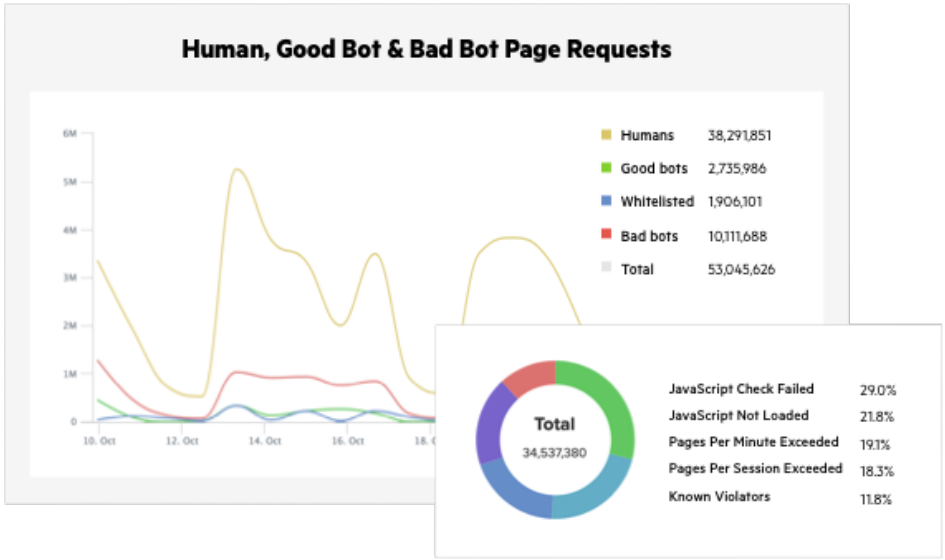
Thirteen Questions You Should Ask Your Bot Mitigation Vendor

Use these 13 questions to evaluate bot mitigation solutions, identify key differentiators, and make an informed decision.

TABLE OF CONTENTS

Thirteen Questions You Should Ask Your Bot Mitigation Vendor

The volume of automated threats on the internet today is staggering—in fact over 20 percent of all internet traffic comes from bad bots. As the sheer volume, sophistication, and business damage caused by bad bots grows, many vendors are beginning to claim they have the ability to identify and mitigate bots. You'll want to evaluate the claims of these vendors. Use these 13 questions to help differentiate between solutions and make an informed decision.[2]



1. How does your solution deal with advanced persistent bots?

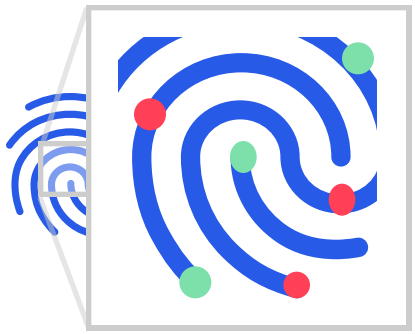
Today's bots are more sophisticated than ever. In fact, these bots are so different from the primitive bots of yore, we've created a new term to describe them: advanced persistent bots (APBs).

An APB can dynamically rotate IP addresses and to distribute their attacks over hundreds or thousands of IP addresses. Rotating IP addresses helps a bot avoid detection by circumventing rate-limiting and blacklisting.

Likewise, distributing attacks across many IP addresses reduces requests per IP, allowing these attacks to fly under the radar. After all, 1,000 IP addresses making one request each achieves the same goal as 1 IP address making 1,000 requests.

To combat these tactics, your bot protection solution should include device fingerprinting to track bad bots across IP addresses. This will greatly enhance the solution's accuracy and allow it to detect much more sophisticated bots.

When evaluating vendors, ask if their solution includes device fingerprinting. If so, ask what criteria go into creating each fingerprint? Solid fingerprinting technology will stick to the bot even if it attempts to cycle through random IP addresses or hide behind an anonymous proxy. In contrast, IP-centric bot detection will suffer from high false positives and false negatives.



2. Are you able to detect browser automation tools like Selenium and PhantomJS? If so, how?

On the forefront of bad bot innovation are a series of tools such as PhantomJS and Selenium that use browser automation to perform their attacks.

2.

Many of these tools are actually browser plug-ins that possess all of the capabilities of a legitimate browser including a JavaScript engine with the ability to launch, fill, and submit forms, click on specific areas of the screen, follow links, and even take screenshots.

Because these tools process JavaScript and run a browser in nearly the same fashion as a human would, they are difficult to identify and block.

Talk to your anti-bot vendor and ask how they handle these threats. Make sure the answer isn't, "with JavaScript checks or IP reputation lists," because these techniques are no match for browser automation tools.

Instead, you'll want to find a solution that can perform a detailed analysis of the webhooks used in these browsers and find the tell-tale signs that a browser automation bot is behind the wheel.

3. Does your solution make use of modern bot detection techniques?

Legacy bot mitigation solutions lean heavily on IP-based rate limiting, IP blocking, user agent validation, and JavaScript checks. Ask your vendor what modern, proactive security mechanisms they utilize.

Does their solution inject active challenges and honeypots into HTTP traffic to trap bad bots? Are they making use of behavioral modeling and machine learning to understand what is normal behavior for your website, and then looking for deviations from the norm? Besides being more accurate, modern bot detection techniques are self-optimizing, minimizing their administrative burden.

Ask your vendor what modern, proactive security mechanisms they utilize.

4. Does your product make use of machine learning? If so, how does it work?

Machine learning works by establishing and constantly updating a baseline of behavior for your website, then identifying anomalies to that baseline in real time. It is going to learn things like browser popularity by country and be able to correlate that data against your site.

Machine learning can quickly add in new classifications such as:

- Time of day
- Mouse movements
- Navigation paths

For example, is a visitor from a given country more likely to use a mobile or desktop browser during a specific time of day? This type of automated analysis greatly bolsters the power of the bot detection and mitigation solution. Keep in mind that running machine learning algorithms on big data is great, but intelligence requires a feedback mechanism. The system needs a way to understand the impact of its decisions.

One example could be a Turing test presented when machine learning intercepts a malicious user. The system can challenge that user and update its algorithms based on whether its decision was right or wrong.

Does the vendor's machine learning have this level of sophistication?

Also, machine learning shouldn't be a black box. You should have some degree of control over its settings. Can you dial up the settings on critical websites or URLs, or turn it off for specific applications? Make sure you fully understand the product's machine learning capabilities so you can select a solution that's more than just a marketing buzzword.

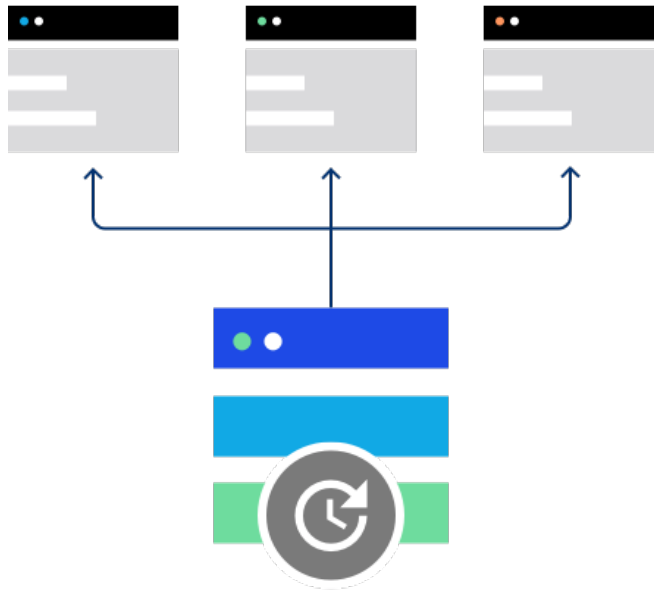
5. Does your solution incorporate external and community sourced intelligence feeds?

Bots are a tricky beast. They can rotate IPs, mimic human behavior, and distribute their attacks, all of which makes them hard to pin down with IP addresses.

Simply getting a list of malicious IP addresses and piping them into your defense strategy may not be an effective strategy against this type of adversary. Why not? Because not all security solutions can detect advanced persistent bots.

If the source of your intelligence feeds doesn't make use of techniques proven to detect bots, then the data they provide may add little value. One type of threat feed that is extremely valuable for bot mitigation is a threat feed derived from bad bot device fingerprints from other companies using the same anti-bot solution you have implemented.

Does your solution include community sourcing that enables you to benefit from fingerprints obtained from other customers using the same solution when they are attacked by bots? How and when are these fingerprints updated?



6. Are you able to protect my APIs from automated threats?

When you start blocking bots from accessing your website, you're cutting off a revenue stream for a would-be hacker. In response, many determined attackers will shift their focus to your APIs in the hopes of regaining access to their revenue stream. That's why it is also important to protect your APIs against these threats, as well as developer errors, integration bugs, and general abuse.

Talk with your potential vendors and ask what their capabilities are for handling automated threats launched at your application programming interfaces.

7. What deployment options do you offer?

Generally, bot mitigation vendors offer their solutions in a few deployment types: appliances, cloud services, and API calls. Each of these deployment types has its pros and cons, so make sure to find one that fits your web environment and operational style. Below is a brief description of each common deployment type that will you understand their strengths and weaknesses:

Appliances

Appliances are an option for many customers. They work well behind content delivery networks (CDNs). They are in your data center, and thus subject to your security protocols, and they can seamlessly integrate with your existing infrastructure. When evaluating a vendor with an appliance, you should determine if they offer both physical and virtual appliances and how long it takes to get the solution up and running.

Cloud-Based Network

Typically, cloud-based network deployments refer to operating a CDN that has bot mitigation as one of its features. To use this type of solution, you'll route traffic to the vendor's network, they will provide their service, and route the clean traffic back to you. This deployment type is easy to implement and may provide other useful services such as content caching.

What many vendors won't tell you is that if you already have a CDN, this may not be a good option for you as you will need to daisy-chain the CDNs together which increases latency. Additionally, many CDN-based bot mitigation solutions do not function at full capacity when deployed behind another CDN.

API Calls

Finally, there are solutions on the market that function via API calls. These solutions may add latency to websites because requests must be sent to the vendor over their API for inspection before the traffic is allowed to reach the protected website. Also, consider which technologies the solution integrates with to make sure they are appropriate for your environment.

8. Can your solution be implemented and configured on a per-domain basis?

Depending on how an anti-bot solution works, it may be possible to provision and configure the service on a per-website basis, or even in bulk by managing hundreds or thousands of domains at once. A solution architectured in this manner will greatly reduce the management overhead your team will experience with the tool.

This is in stark contrast to solutions that can only be set up and configured on a per-URL basis. If you're managing many URLs, websites, or both, then you need to ask how websites and specific web pages are provisioned, configured, and managed. Implementing a solution with a bulk management interface will be advantageous.

The flip side of this coin is granularity. Once you've established your initial deployment, you may have some especially sensitive pages you'd like to ratchet up protection on. Some examples of these types of pages might be account login pages or account registration pages. Look for a solution that offers per-URL customization and security controls that will help you fine-tune your protection.

9. How much effort is required to manage your solution? What features do you have that will reduce my management overhead?

The fact is, some solutions will require more hands-on attention to be effective than others.

Talk with your potential vendors about:

- How much attention their identification and enforcement mechanisms will take
- How reactive or proactive their solutions are
- Whether or not they include self-tuning capabilities

Usually, there is a correlation between maintenance and reactive security approaches, meaning that if your solution requires you to take action, such as blacklisting or whitelisting based on newly identified threats, you're going to spend a lot of time doing that.

On the other hand, solutions that include proactive strategies such as machine learning , behavioral modeling, self-tuning, honey potting, or browser validation will greatly reduce the amount of time you spend in the user interface turning proverbial knobs and dials.

10. What parameters does your access control list include and does it perform self-maintenance?

Access control lists (ACLs) are a necessary part of any security strategy. However, they are a perfect example of a reactive security mechanism. You need to update them constantly based on new information. If you discover a new threat, you blacklist it. If you get a new partner, you adjust your whitelist. It never ends.

Don't worry, there are tactics that can help minimize your labor. Ask your anti-bot vendor if they can block by country or organization. Instead of blacklisting specific IP addresses, can their solution blacklist or whitelist entire countries or organizations and dynamically obtain IPs for these entities? This approach ensures that the IP addresses on your ACLs are accurate.

Another feature of a modern ACL that will cut down on maintenance is self-depreciation. Instead of blocking an IP into perpetuity, a timed approach where specific violations are blacklisted for specific lengths of time can be used. Repeat offenders are blocked for longer periods of time. Using this approach, your ACL is more likely to be accurate and to have current data.

11. What options does your solution provide for rate-limiting controls?

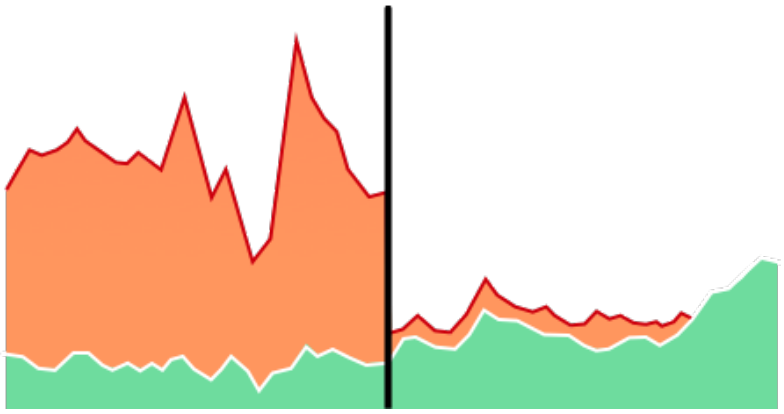
Ask your vendor if they can rate limit based on:

- Client
- Device
- Authentication tokens in addition to simple IP addresses

IP addresses are easy to obtain, spoof, and rotate, so you'll want to be able to rate-limit based on additional characteristics of the visitor.

Does the vendor offer graduated controls for rate-limiting? For example, when traffic from a specific client or IP address surpasses a predefined threshold such as 1,000 requests per minute, can it send an alert? If traffic continues to rise and surpasses a second predefined threshold such as 5,000 requests per minute, can it be blocked?

Finally, does the vendor offer predictive analytics that can help you forecast the impact of any changes you make to your rate-limiting settings before you go live with them? By showing you how your new controls will play out on your traffic, you'll be able to make informed decisions and avoid accidentally blocking legitimate users while adjusting security settings.



12. What kind of visibility do you provide in terms of bot traffic, trends and motives?

We've all used tools that generate great data, but in an unabsorbable way such as massive tables or log files. Perhaps they include some filters to help you make sense of it all. But you need more insightful information and context. For example, a solution may be able to identify that it has blocked 100,000 bots, but can it tell you what those bots were doing and why?

Ask your potential vendors what pre-made dashboards and reports are available to help you understand your bot problem. Look for solutions that provide more information than just the number of blocked bots or the standard traffic distribution. Ask what insights the solution can provide and how.

To truly understand your bot problem, you'll want to be able to answer the following questions:

- What were the top pages the bots targeted?
- Whose bots were they?
- What mechanisms caught the bots?

You'll also want reports and dashboards that detail the activity of good bots visiting your website. Having a detailed analysis of helper bots and search engine bots gives you insights that may aid you in optimizing your web applications.

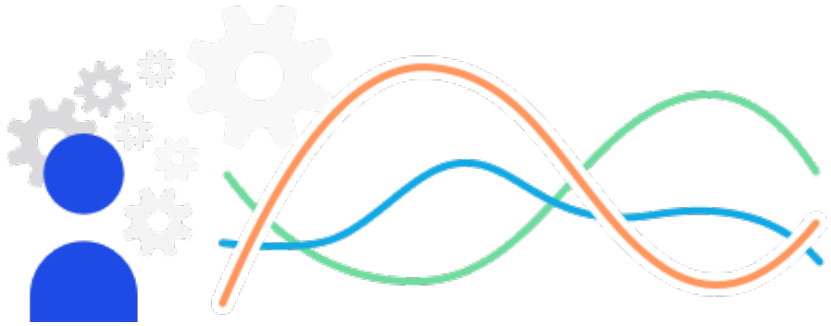
For example, does Google visit you at the same time every day? Perhaps you can release new content before its arrival. Did your security team scan your website for vulnerabilities? That might explain the sudden jump in page views you saw in your analytics tool. Ask your vendor if their solution includes pre-made reports for good bots visiting your web property.

13. What options do I have for enforcement?

Let's assume that whatever solution you implement finds some bad bots, because it will. Once you've identified these bots, you'll need to do something about them.

Ask your potential vendor what mitigation options you have. Can the solution be run in a monitor or alert-only mode? If you want to block traffic, what enforcement mechanisms do they offer? Can you customize the blocking pages to include your own messaging and branding?

You may even want to fight back. Imagine that your biggest competitor is constantly scraping your website for information to use against you. You might want to turn the tables on them and show them some fake information. Some vendors can provide you with a mechanism to accomplish this.



imperva

phase pacific

About Imperva

Imperva Bot Management protects your websites, mobile applications, and APIs from automated threats without affecting the flow of business-critical traffic. As part of the Imperva Application Security Solution, Bot Management protects all your access points and provides you with the choice of multiple response options for incoming bots. Imperva collects and analyzes your bot traffic to pinpoint anomalies.

Our machine learning models identify real-time bad bot behavior across our network and feed it through our known violators database. Biometric data validation, such as mouse movements, mobile swipe, and accelerometer data, catches malicious botnets. Rate limits based on device fingerprints — not IPs — provide further protection.