imperva

CUSTOMER SUCCESS STORY

Leading Job Site Blocks Bots

Leading job site blocks bots with Imperva Appliance.

Overview

This leading jobs website is an innovator in the way it helps people find jobs and the way it helps companies recruit talent., and allows visitors to research which employers are hiring and gain a deeper understanding about the culture of the company. Thousands of employers also advertise open positions on this platform.

% 99.9%

real human traffic

Challenges

Unwanted bots were crawling the site, muddying web metrics

A leading job site had a suspicion that unwanted bots were crawling their site. Though they welcomed friendly search engine bots like Google and Bing, they were concerned about bad bot activity.

One of the primary drivers for the concern was web metrics. The site wanted to be absolutely sure their metrics were accurate so they could share them with customers.

Preventing

bad bots from crawling

Advertising platform traffic lacked integrity, impacting customers

As an advertising platform for employers, the leading job site's traffic metrics are important. The SVP, Engineering and CTO at the site, was adamant about taking a proactive approach to bot detection and mitigation in order to keep the metrics clean.

"Web metrics are important to our business model. As an ad platform, we host job advertising on our site. One of the key ways we show value to our customers is by demonstrating how much activity we're driving to their sites for jobs. If we have a lot of unwanted bots crawling our site and driving traffic to them, that would be bad. That wouldn't be doing right by our customers," he explained.

"We wanted to be proactive, to make sure that we had a high level of integrity, and be doing our best for our customers," he said. "This was something we wanted to get in front of."

Lower

infrastructure costs

Homegrown solution could not proactively block bad bots

The site's CDN had some utilities to block bots, and the Engineering team built a homegrown solution to supplement those efforts. These quick fixes could block a specific IP address, IP range or user agent. But the homegrown solution was always reactive, instead of proactive.

The team was unable to identify and block bots before they became a problem.

"The identification process was very manual. We were crawling through server logs and using third parties to monitor performance. The monitoring tools would show us, 'Hey, this part of our site is going slow.' We'd dig in and try to root cause it." said the SVP, Engineering and CTO.

"On a handful of occasions, the team identified a really badbehaving bot; that bad-behaving bot would typically drive a lot of activity to one part of the site, causing that part of the site to slow. It wasn't great for the user experience.

"We wanted to catch those bots ahead of time, as opposed to reacting to a slowdown. It wasn't proactive; we were only stopping the bad bot activity after it happened," he noted.

"We wanted to catch those bots ahead of time, as opposed to reacting to a slowdown. It wasn't proactive; we were only stopping the bad bot activity after it happened,"

Malicious traffic consumed resources, drove up infrastructure costs and hurt performance

The SVP, Engineering and CTO was also concerned about his infrastructure costs. He wanted to be sure that unwanted bot traffic was not driving a large percentage of traffic, inflating the amount of resources he would need.

"This is also something I wanted to be proactive about. As the site became more popular, we began to see more bad bot activity. Though it wasn't at a problematic level yet, it was growing. It was important to know if metrics were skewed and to be proactive in managing the associated infrastructure costs," he said.

Furthermore, he was concerned about performance of the site. "We take performance very, very seriously. We want to make sure that our user experience is always great, and a big part of the user experience is having a fast site. Any time that we have a performance issue pop up, it's a problem. So I was very focused on finding a way to proactively prevent performance issues caused by bot traffic," he said.

Engineering lacked control over the data set

The leading job site's Engineering team was also worried about their data set; there was some indication that bots could be stealing content and using it in ways that were not authorized. Though they had not yet seen hard evidence to this effect, the team wanted to be proactive to make sure they had complete control of the valuable data.

Requirments

In searching for a bot protection system, the site looked for an appliance which offered redundancy, but did not increase latency. With a CloudFlare CDN and F5 load balancers already in place, the Engineering team wanted a solution that would fit well with their existing infrastructure.

"We were looking very specifically for bad bot protection. We chose to use an appliance, rather than a public reverse proxy, because we didn't want to add yet another layer. We already had a CDN. We didn't want to add another service that was another proxy, out in the wild, to add latency to our user experience," said the SVP, Engineering and CTO.

The leading job site went with the Imperva Bot Management (formerly Distil Networks) Appliance in a high availability configuration (2 boxes; redundant servers). "Traffic comes from our CDN into our F5 load balancer. It then proxies that information, and sends a request over to the Imperva appliance. Assuming Imperva gives it the thumbs up and allows the traffic through, it sends that back to our load balancer, and the load balancer then distributes it to our application servers.

"Configuration and testing were a bit time consuming, but once we got it working, it's been great," he concluded.

Why Imperva?

The site evaluated many solutions, but none worked as well as Imperva

The site looked at internal solutions, third parties and various CDNs. As they evaluated the options, they realized that while many offered basic bad bot protection, none of them seemed to do as good a job as Imperva.

"The big value that Imperva leverages effectively is utilizing the data gathered from the entire Imperva network. Rather than just using a simple rule-based web application or firewalls, Imperva captures information about the malicious bots from all of its customers and shares that learning, to everyone's benefit. All of us are able to leverage what is learned every day. So there is a strong network effect, which is really powerful. There were very few other companies that can offer this benefit,"

"The big value that Imperva leverages effectively is utilizing the data gathered from the entire Imperva network. Rather than just using a simple rule-based web application or firewalls, Imperva captures information about the malicious bots from all of its customers and shares that learning, to everyone's benefit."

Results

Bad bots are blocked from crawling, and web metrics are accurate

Implementing the Imperva appliance was a proactive move for the site. Since Imperva Bot Management was installed, they have seen a reduction in bad bot traffic, and they have not experienced any site performance issues due to bots. The implementation was, in the SVP, Engineering and CTO's words, "a huge success."

"Ultimately, the biggest win is we have a lot more confidence in our site metrics. We feel much more comfortable talking to our advertising customers and saying, "We are confident that these numbers are accurate, because we know that we've done the right things to reduce any noise."

Advertising platform traffic is 99.9% real human traffic

After the Imperva appliance was installed and optimized, traffic from the leading job site's advertising platform was significantly cleaned up. The engineering team no longer worries that customers are getting bad bot traffic on their profile pages or referral pages.

Bad bots are blocked proactively, not after-the-fact

With the Imperva appliance in place, the site is now in front of the problem and doesn't have to worry about going through logs and reacting after the fact. They have outsourced their bot problem, and bad bots are stopped before the traffic is distributed to the application servers.

Infrastructure costs are not driven up artificially by bots

The SVP, Engineering and CTO noted, "We expect the Imperva appliance will save us costs in the long run, as we restrict this extra bad bot traffic to our site."

Engineering has restored confidence and control over the data set

"We have more confidence now that our data is being used in the way that it was intended to be used. Without the Imperva appliance, we wouldn't know, and in all likelihood, our data could be misused.

"We use the Imperva Bot Management portal extensively. It's been important for us to see that the appliance is doing what we want it to be doing, and to confirm it is not doing things we don't want it to be doing. We can see that it is blocking a percentage of our traffic, so we're getting value with it. We can also make sure that too many people

"We expect the Imperva appliance will save us costs in the long run, as we restrict this extra bad bot traffic to our site."

aren't getting the Captcha screen; that would be an indication the settings were too aggressive. Those are really important metrics," he explained.

"One of the most exciting features from Imperva Bot Management is their ability to leverage their network. If they detect a bad bot from one of their other customers, The leading job site can be the beneficiary of that, and block that suspicious activity. Likewise, we're a part of the network, and we're happy to share with others to make sure that they're blocking the bad activity. I think leveraging that network effect is really powerful and something few other vendors do," he concluded.