



CUSTOMER SUCCESS STORY

# Whitepages Protects Online Data and Blocks Form Spam

whitepages™

# Whitepages Protects Online Data and Blocks Form Spam Using Imperva Bot Management (formerly Distil Networks)

## Overview

Founded in 1997 by Alex Algard out of his Stanford dorm room, Whitepages has become the leading provider of contact information in North America. Whitepages helps people find, understand and verify personal and business identities, and has more than 50 million monthly Users.

## Challenges

Whitepages was a prime target for scrapers. Because of the volume of valuable identity data stored there, Whitepages.com was a prime target for data scrapers. Web scrapers would either use the data for themselves, or resell it for profit. Protecting premium content was an additional challenge. “When scrapers get into the premium content, they trigger third-party API calls to our vendors, who are billing us,” said Michael Bradshaw, IT Operations Manager at Whitepages. “So we were paying for third-party vendors to feed data to unauthorized web scrapers.”

### Homegrown solution was time-consuming and error-prone

Although Bradshaw’s IT Operations team had a solution for bot detection and prevention, it lacked accuracy. In addition, it was prone to errors because it relied on identifying and blocking IP addresses, not specific users.

“Our homegrown bot detection solution performed Splunk queries roughly every 15 minutes and applied logic to determine the legitimacy of an IP address,” Bradshaw said. “We’d find a heavy user and ban the address, which resulted in a lot of false positives. Actual scrapers, especially the more sophisticated ones who performed distributed searches across a wide number of IPs, would still slip through the cracks.”

### Lack of insight into user variability and increased query load was degrading user experience

The existing IP model was unable to capture the variability of users. “We could be blocking an office building where people were trying to look up a good lunch spot, or a guy running a script to scrape data in the background—we couldn’t tell the difference,” he lamented.

**20%**

**DevOps time  
saving**

**Prevented**

**scrapers from  
stealing data**

**Stopped**

**false-positive  
detection**

To make matters worse, infrastructure costs were rising due to increased query load from malicious bots. “We noticed that the better we made the user experience, the more bots we let through and the more data we gave away,” said Bradshaw. “We simply didn’t have the sophistication in our detection methods to identify which bots were bad and which weren’t, and to block the bad traffic while ensuring a good user experience.”

### Bot detection and prevention took a lot of care and feeding and required constant adjustment

“Our customer service department received a lot of complaints about blocked pages,” said Bradshaw. “We had to respond to those complaints daily, as they were reported. It was taking one of our Operations staff the equivalent of 20% of his time.” Bradshaw said the team received up to 50 inbound queries a day, and would have to tweak the algorithm and respond to actual users, which took a lot of time. “We’d easily spend at least one day each week trying to figure out what was going on.”

## Requirements

Bradshaw’s team began looking for purpose-built bot detection and mitigation solution that would enable them to identify threats on a very granular, per-user level. “We wanted the capability to just log activity, use CAPTCHA, or serve up a blocked page at various levels of activity,” he said.

“We’re not experts in bot detection; that’s not our core competency,” he said. “Data is the lifeblood of our business, and when it’s leaked out, its value decreases. Getting a solution that worked was essential to our business.”

## WHY Imperva?

### Advanced detection algorithms enable per-user, per-request filtering

“One of the things that appealed to us most about Imperva was its ability to block or ‘CAPTCHA’ bots on a per-user, per-request basis, rather than on a per-IP basis,” he said. “Previously, we used a hammer-like approach—either we blocked the whole IP address or not. The ability to ‘CAPTCHA’ bots and perform rate-limiting was definitely appealing.”

### Straightforward deployment makes blocking bad bots easy

**“We’re not experts in bot detection; that’s not our core competency,” he said. “Data is the lifeblood of our business, and when it’s leaked out, its value decreases. Getting a solution that worked was essential to our business.”**

Michael Bradshaw, IT Operations Manager at Whitepages

“We set up Imperva Bot Management to have access, and they did all of the configuration,” said Bradshaw. “Imperva passes legitimate traffic back to our load balancer, at which point we continue processing pools and send them to the appropriate web servers,” he said. “It’s an innovative solution that works well and was easy to deploy.”

### Responsive support goes a long way

The Imperva dashboard provides many levers which allow Bradshaw to adjust and fine tune his bot blocking parameters. “One time we noticed we were blocking more pages than we should have been,” he said. “We made an adjustment that fixed the problem in minutes.”

## The Result

### Imperva stops comment spam and prevents scrapers from stealing data

Since implementing Imperva, Bradshaw’s team has reduced time spent worrying about bot detection, while increasing protection. Bradshaw is also using Imperva to filter comment spam displayed on the Whitepages “reverse phone” comments page. “All the ‘reverse phone’ pages are built through Imperva and are therefore protected,” he said.

### Imperva Bot Management reduces time spent detecting bad bots and scrapers by 20 percent

Before Imperva, the organizational burden of detecting and blocking bots was heavy, taking at least 20% of a DevOps employee’s time. “We were spending at least one day out of the week trying to figure out what needed to be fixed and how to fix it. We’d have to update the list of IP addresses to block, and re-generate a file every 15 minutes,” said Bradshaw. “With Imperva, we don’t have to do any of that—it’s all Automatic.”

### Imperva’s accurate, self-optimizing solution helps stay ahead of the bad Bots

Because of the solution’s ability to fingerprint bots based on over 40 variables, Bradshaw’s team only blocks the bad bots; false positives have become a thing of the past. “Our previous IP-based model didn’t give us the granular insight we have with Imperva,” said Bradshaw. “The various methods Imperva applies to every request -- along with its self-optimizing capability -- enables us to stay ahead of the curve. It constantly evolves in the background, learning and identifying new malicious behaviors as they’re introduced. That is a huge plus.”

**“With Imperva, we don’t have to do any of that—it’s all Automatic.”**

Michael Bradshaw, IT Operations  
Manager at Whitepages

## Reduces need and expense of taking legal action against malicious scrapers

When Bradshaw's team suspected content theft or IP infringement, they sometimes initiated legal proceedings. Depending on the particular scraper, Bradshaw contacted the ISP and had lawyers draft "cease-and-desist" orders. Imperva helps with this process by identifying the ISP from which the bad actor originates. "We definitely wouldn't be able to identify as many of these bad actors without Imperva," he said. "In addition, we're now able to block the bad bots, eliminating the need and the expense of taking legal action."

## Eliminating bot traffic decreases query load on production servers, reducing infrastructure costs and improving user experience

With fewer bots getting past the Imperva servers, the query load on Whitepages' production servers has been reduced substantially, improving site performance and the user experience. "Using Imperva Bot Management, we're protecting our data and preserving a good user experience," said Bradshaw. "We've essentially outsourced our bot problem."

**"We definitely wouldn't be able to identify as many of these bad actors without Imperva,"**

Michael Bradshaw, IT Operations  
Manager at Whitepages