SOTI

V10 ISSUE 03

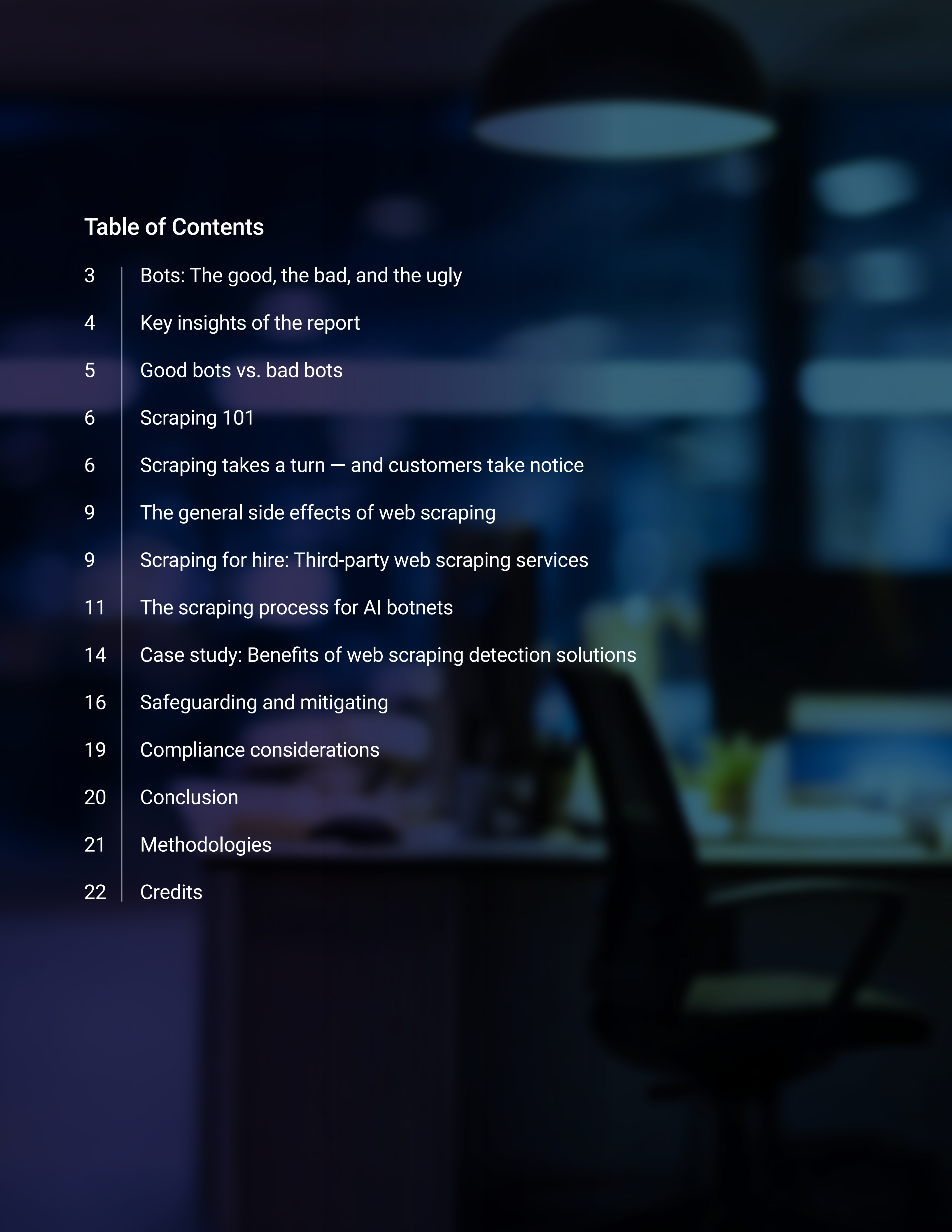
10 YEARS
OF SECURITY INSIGHT

# Scraping Away Your Bottom Line

## How Web Scrapers Impact Ecommerce

Akamai

State of the Internet/Security



## Table of Contents

Did you know that bots generate more than half of all web traffic? The commerce vertical, in particular — with its reliance on revenue-generating web applications and assets — has been most affected by high-risk bot traffic (Figure 1). And while we often hear that bots are evolving, web scraper bots are the type that's grabbing the attention of ecommerce-driven organizations today because their economic impacts — oftentimes hidden below the surface — differ from those of other types of bots. The detection of scraper bots has also become much more difficult due to the rise of artificial intelligence (AI) botnets and headless browser technologies, which make them extremely evasive. For example, one of Akamai's ecommerce customers had 99% of high-risk traffic stopped that they didn't even know was from scraper bots.

**Monthly Bot Requests: Top 3 Verticals**

January 1, 2023 – March 31, 2024

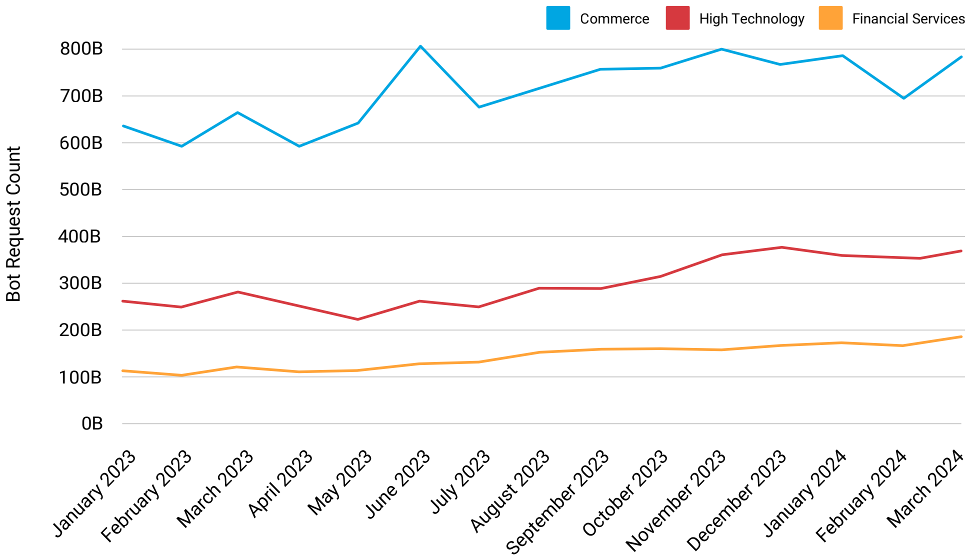


*Fig. 1: Commerce is the top vertical for bot requests, and a rise in global bot traffic in the commerce vertical can be observed from the beginning of 2023 through Q1 2024*

Therefore, in this State of the Internet (SOTI) report, we focus on the evolution and specialization of these bots — and their operators. Although bots have been around for some time, we continue to see their application across a variety of groups to enable criminal attacks, fraud schemes, and competitive intel. Recently, we have seen a trend toward the increased use of all bots and a rise in the negative business impacts of scraper bots. This report is designed to share both technical insights and attack methodology to raise awareness of this growing problem across the commerce industry.

## Bots: The good, the bad, and the ugly

Every major ecommerce-focused organization suffers from bots that are continuously evolving and becoming more specialized depending on what they aim to accomplish. Within the commerce vertical, there’s a large variety of bot types that perform many different tasks. An easy way to think of them is to break them into three groups: good bots, bad bots, and gray bots. Good bots help customers find your site. Bad bots scrape your site for malicious purposes. Gray bots tend to be noisy even though they are still legitimate; they are really a subcategory of the good bots (e.g., partner bots that are constantly pinging and other program APIs that make frequent calls).

So, as we think about helpful chatbots and search engine bots that can have beneficial impacts — such as answering users’ basic questions and providing website content that returns more accurate search results — we want to optimize those types of bots while containing IT costs. For harmful ones, like credential stuffing bots that try to gain unauthorized access to a customer’s account leading to account takeover, we want to take preventive measures without impacting the overall customer experience. One type of bot that hit the scene recently is becoming especially problematic by reducing revenue, diminishing loyalty, and increasing costs — web scraper bots.

Scraper bots, a botnet used to directly extract data and content from websites on the internet, are unique. They are demanding attention because of how they operate differently and how their business impacts and detections vary from those of other bots. Web scrapers are also multifaceted in that their use cases vary depending on how organizations and operators monetize the information these bots collect. Regardless of the particular goal, scrapers are costing revenue, increasing IT costs, and lowering overall customer experiences.

In this SOTI report, we examine the impacts of scraping across ecommerce and examine why business owners (think digital, marketing, brand, finance, risk, and security) should have a shared interest in stopping abusive scrapers. To better understand these impacts, it is crucial to view the full picture of why web scraper bots have evolved, what they are being used for, how they operate, what their impacts are, and what commerce organizations can do about them.

## Key insights of the report

Web scraping is not just a fraud or security problem, it is also a business problem. Scraper bots have a negative effect on many facets of the organization, including revenue, competitive edge, brand identity, customer experience, infrastructure costs, and digital experience, just to name a few.

According to an Akamai research case study, 42.1% of overall traffic activity was from bots, with 65.3% of that bot traffic from malicious bots. And a total of 63.1% of the bad bots traffic used advanced techniques.

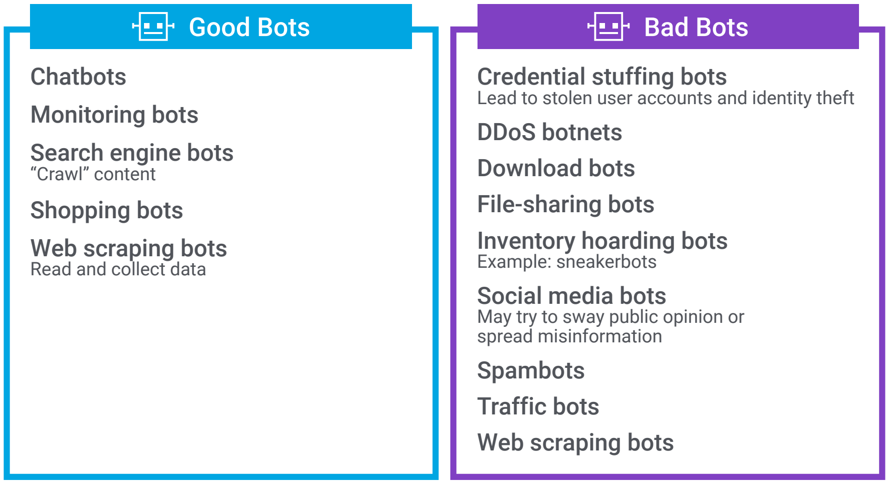
Headless browser technology has changed the scraper landscape, requiring an approach to managing this type of bot activity that is more sophisticated than other JavaScript-based mitigations.

Technical impacts that organizations face as a result of being scraped, whether the scraping was done with malicious or beneficial intentions, include website performance degradation, site metric pollution, compromised credentials attacks from phishing sites, increased compute costs, and more.

It is important to observe and understand the different traffic patterns to identify whether a website is incurring human, basic bot, or sophisticated bot traffic. These patterns can range from circadian to intermittent to continuous.

## Good bots vs. bad bots

Let's start with the basics: A bot, short for "robot," is a computer program that can perform automated tasks faster and more accurately than a human can. The various roles and types of bots fall into two main categories: good bots and bad bots (Figure 2). Gray bots are a subcategory of good bots, but we'll merge them with the good bots for now to simplify the comparison.



**Good Bots**

- Chatbots
- Monitoring bots
- Search engine bots
  "Crawl" content
- Shopping bots
- Web scraping bots
  Read and collect data

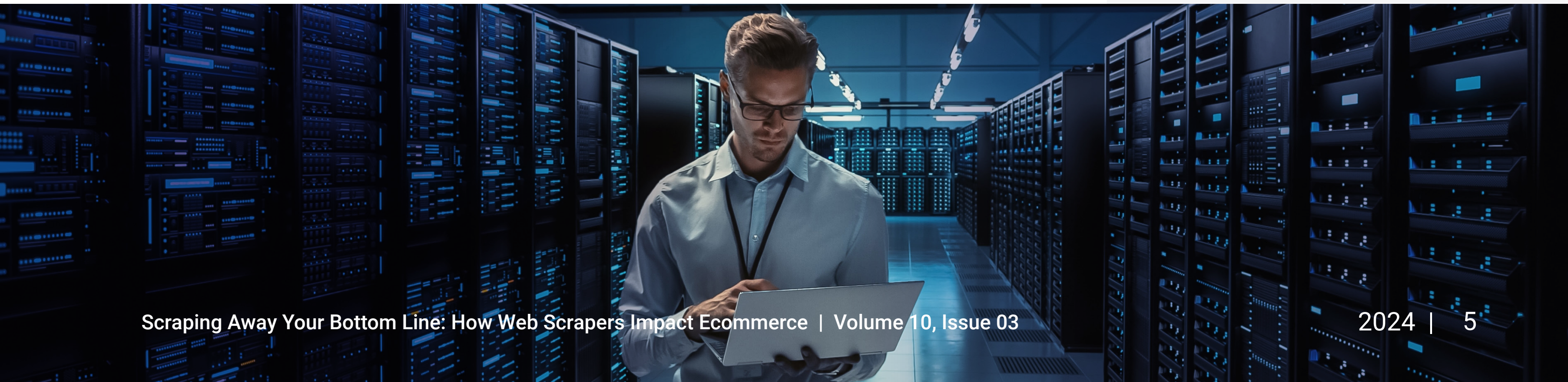
**Bad Bots**

- Credential stuffing bots
  Lead to stolen user accounts and identity theft
- DDoS botnets
- Download bots
- File-sharing bots
- Inventory hoarding bots
  Example: sneakerbots
- Social media bots
  May try to sway public opinion or spread misinformation
- Spambots
- Traffic bots
- Web scraping bots

*Fig. 2: A side-by-side comparison, with examples, of good bots and bad bots*

Good bots are useful bots that help provide tools and services, while bad bots are often used with malicious intent by cybercriminals and fraudsters. An example of this type of malice is a traffic bot that mimics human behavior online to increase clicks and traffic on a website (i.e., commit ad fraud).

Web scraping bots appear in both the good bot and bad bot categories. The distinction has to do with how organizations use the information that these bots collect. We will now focus more closely on various use cases associated with both the good and the bad effects of scraper bots that are faced by some of the largest retailers and ecommerce brands in the world.



Akamai
Data Collection
Data Extraction
Data Proces
Competitive analysi (product positionin pricing, etc...)
Adaptive scraping and anti-ban decision tree
Retail site 1
Retail site 2
Social media
Data Corp / Business Intelligence
Customer sentiment analysis

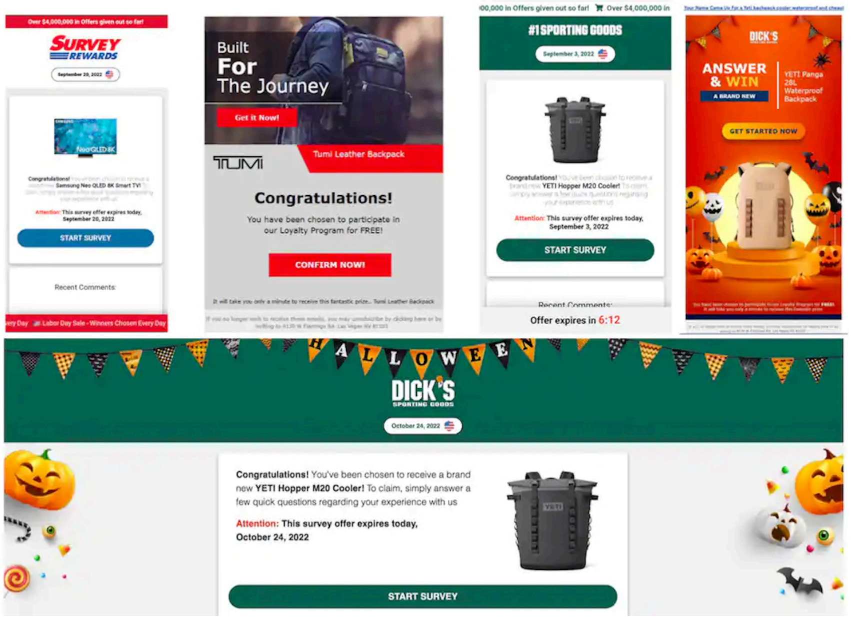
## Scraping 101

Web scraping is commonly used by ecommerce businesses. In the travel and hospitality sectors, for example, travel aggregators scrape dynamic content from their hotel and airline partners to stay up to date on availability and pricing. This type of scraping is expected, and businesses use common bot controls to throttle scrapers during times of day when real users are looking to make a reservation. Organizations also use data extraction services providers to gather leads and other related information from competitors. Additionally, scraping bots can be used for analyzing data and identifying trends. Scraping may also be beneficial for site review to improve online offerings and services, and to allow potential consumers to more easily find company products, such as via a search engine. All these actions can help businesses achieve a competitive edge. However, there is no denying that many entities are using scrapers for less commendable reasons.

## Scraping takes a turn — and customers take notice

Unfortunately, we often hear about consumers who have fallen victim to phishing scams. In this case, scraper bots may have been used to grab product images, descriptions, and pricing information to create counterfeit storefronts or phishing sites aimed at stealing credentials or credit card information. These phishing/counterfeit sites are a form of brand impersonation, in which the intellectual property of victim organizations is being used to establish trust with potential customers.

Some of the largest ecommerce brands in the world have been impacted by counterfeit sites, phishing campaigns, and the theft of company web data as part of brand impersonation campaigns (Figure 3). Unfortunately, when phishing sites are successful, the legitimate brands are left with the fallout from lost customer trust and loyalty.



Congratulations! Samsung Neo QLED 8K Smart TV!

Attention: This survey offer expires today, September 20, 2022

Congratulations!

You have been chosen to participate in our Loyalty Program for FREE!

It will take you only a minute to receive this fantastic prize... Tumi Leather Backpack

Congratulations! YETI Hopper M20 Cooler!

Attention: This survey offer expires today, September 3, 2022

Offer expires in 6:12

Congratulations! You've been chosen to receive a brand new YETI Hopper M20 Cooler! To claim, simply answer a few quick questions regarding your experience with us

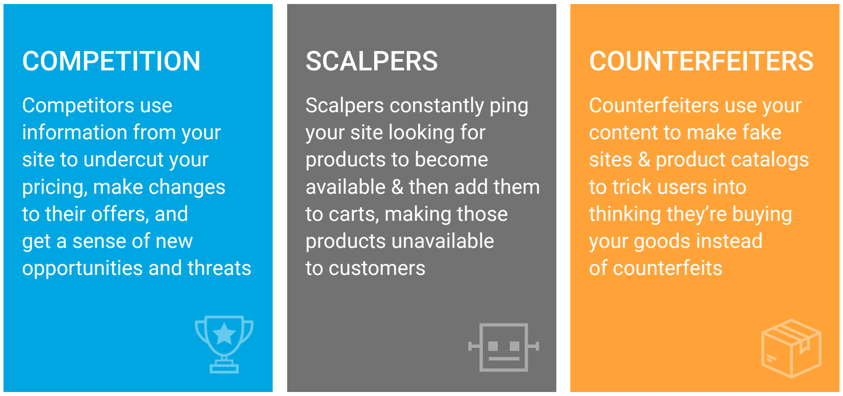
Attention: This survey offer expires today, October 24, 2022

*Fig. 3: An example of some major ecommerce companies that have fallen victim to brand impersonation*

Scalping may also be attributed to web scraping as scalpers can scrape a site for available products and purchase them before legitimate customers have a chance (Figure 4).

**Scraper Use Cases**

There is money to be made by scraping your content



COMPETITION

Competitors use information from your site to undercut your pricing, make changes to their offers, and get a sense of new opportunities and threats

SCALPERS

Scalpers constantly ping your site looking for products to become available & then add them to carts, making those products unavailable to customers

COUNTERFEITERS

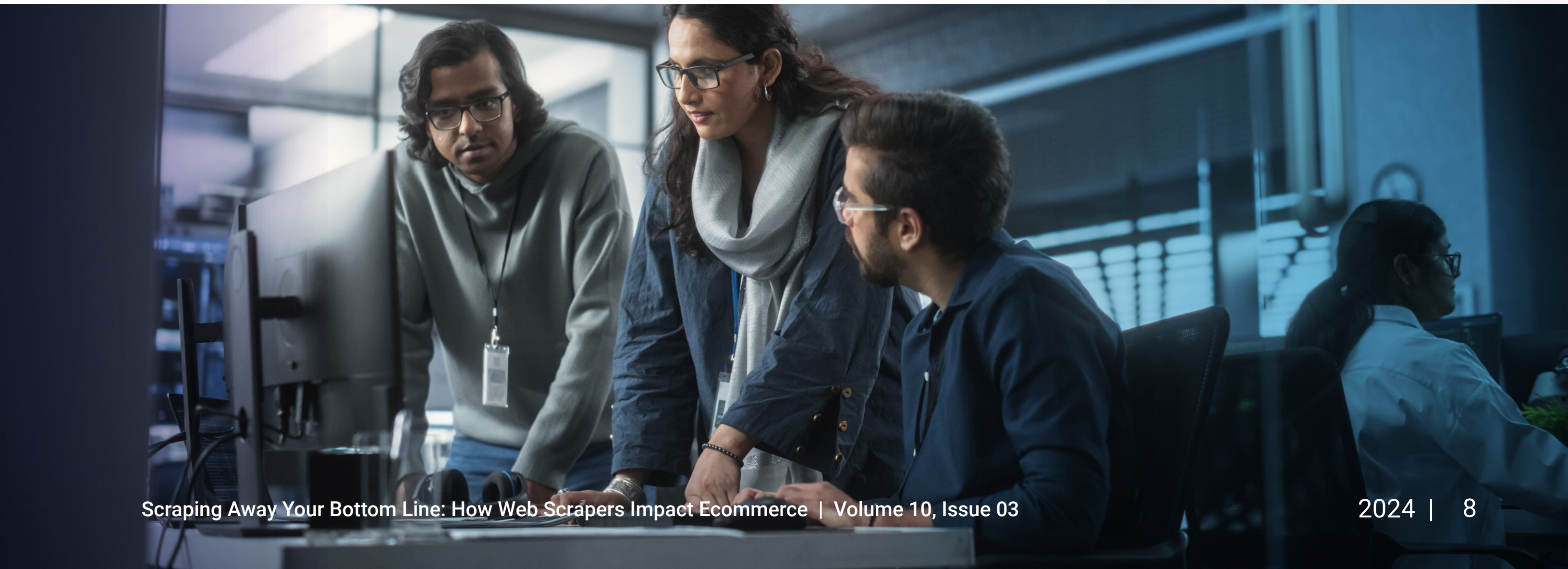
Counterfeiters use your content to make fake sites & product catalogs to trick users into thinking they're buying your goods instead of counterfeits

*Fig. 4: Scraper use cases*

Threat actors who conduct these kinds of harmful scraping activities are aware of the effects that their malicious objectives have on victims. This includes the negative impacts of competitive intelligence/espionage, inventory hoarding/scraping, counterfeiting and imposter sites/goods, and media site scraping and reposting (Table 1). And there are no existing laws that explicitly prohibit the use of scraper bots.

| Impact | Description |
|---|---|
| Competitive intelligence/espionage | Competitors use information from an organization's site to undercut pricing, make changes to their offers, and get a sense of new opportunities and threats. |
| Inventory hoarding/scraping | Scalpers ping targeted sites constantly looking for products to become available and then add them to carts, making those products unavailable for real customers. |
| Counterfeiting and imposter sites/goods | Counterfeiters use scraped content to make fake sites and product catalogs to trick users into thinking they're buying legitimate goods instead of counterfeits. |
| Media site scraping and reposting | Attackers can scrape news articles, blogs, and other content and place it on their own sites, causing the original organization to lose visitors and potential advertising revenue.<br>Advertising rates are often based on site visitor numbers/audience, so fewer visitors means the media site loses the revenue they'd have gotten from higher ad rates. |

*Table 1: Intentional negative impacts caused by web scrapers*



## The general side effects of web scraping

Regardless of the intent of the web scraping, organizations have to deal with expenses from its side effects. Some companies pay for beneficial scraping services, but the companies that are being scraped are incurring costs of their own. These include expenses for anti-bot solutions, and the negative economic impacts of site performance degradation and key metrics pollution (Table 2).

| Impact | Description |
|---|---|
| Increased server, CDN, and cloud costs to serve bot traffic | This impacts revenue and causes reputational loss from competitors', attackers', and counterfeiters' use of content. |
| Site performance degradation | Since scraper bots run continuously until stopped, these bots increase server and delivery costs as organizations serve unwanted bot traffic and suffer from impaired user experiences, such as slower site and app performance. |
| Key metrics pollution | Undetected bot activity severely skews key metrics like site conversion that business teams rely on to make investment decisions, such as product positioning strategies and marketing campaigns. |

*Table 2: Unintentional negative impacts caused by web scrapers*

## Scraping for hire: Third-party web scraping services

As we've mentioned, web scraper bots may be used for good or bad. Unlike the bots used for credential stuffing attacks, which are known malicious bots and therefore justifiably blocked, there are companies that offer legitimate web scraping bots. Many organizations use these third-party web scraping services to extract and provide data to their own organization, which can be beneficial, especially in the world of competitive marketing.

There are dozens of these companies that provide different types of web scraping/data extraction services; there are even conferences that promote them. For example, Bright Data hosts a conference called ScrapeCon that brings together experts on evading bot detections so that companies can learn how to scrape data. Table 3 includes examples of the levels of services that may be provided by third-party web scraping companies.

| | |
|---|---|
| **Service Level 1** | Proxy services may be a part of the scraping and offer infrastructure that could include data centers' mobile IP and residential addresses. |
| **Service Level 2** | This second level may also include automated data extraction that cleans and structures the data for easier use by the customer's data science team members, who extract the valuable intelligence to steer business decisions. |
| **Service Level 3** | The highest level may add the extraction of actual business intelligence itself, which can further enhance the decision-making process for businesses. These are referred to as "AI botnets." |

*Table 3: Various levels of services provided by third-party web scraping companies*

Customers may choose any of these service levels, from the most basic to the most advanced, as well as the frequency of the data collection, and they can specify their targets. Often, the level of service provided, or botnet chosen, depends on the level of protection they need to overcome. A more basic botnet can collect data via an advanced script with a few thousand proxy servers located in data centers that balance the traffic load. If the protection is rudimentary enough, then the botnet could use this technique to pass through the bot management defenses and web application firewall of the security infrastructure.

If, however, the protection is more advanced, then a more sophisticated approach to scraping, such as a headless browser attack, may be necessary. This is true whether the scraping is conducted by an actor with good or bad intent. And it's not cheap, as companies are going to incur costs that are generally much higher for the more sophisticated infrastructure than for the basic service level. An advanced defense may include challenge technologies (like CAPTCHA or proof of work), several detection layers designed for client-side fingerprint assessment, and an analysis of the Hypertext Transfer Protocol (HTTP) and Transport Layer Security (TLS) characteristics.

# The scraping process for AI botnets

Although basic web scrapers may be more consistent in their scraping techniques, AI botnets have the ability to discover and scrape unstructured data and content that is in a less consistent format or location. Additionally, AI botnets can use actual business intelligence to enhance the decision-making process. The sophisticated AI botnets, mentioned in Table 3, service level 3, have a three-step process to scraping data. They operate by collecting, extracting, and then processing data (Figure 5).

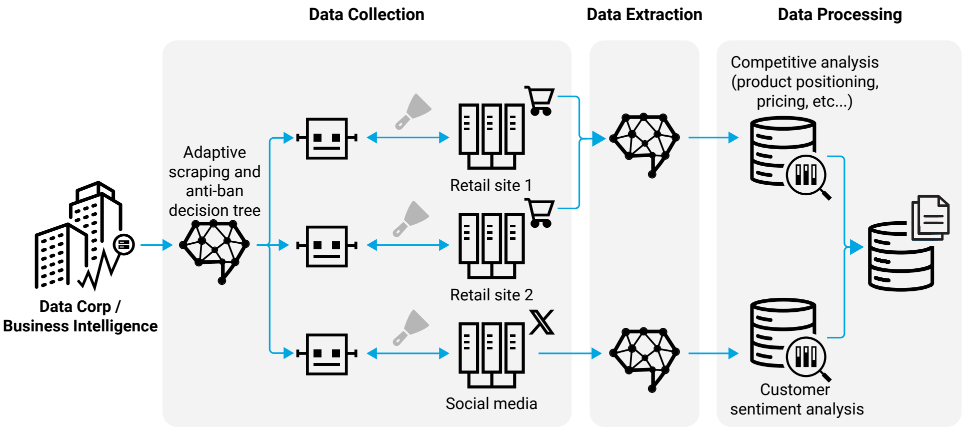

Data Collection
Data Extraction
Data Processing
Adaptive scraping and anti-ban decision tree
Data Corp / Business Intelligence
Retail site 1
Retail site 2
Social media
Competitive analysis (product positioning, pricing, etc...)
Customer sentiment analysis

*Fig. 5: A representation of an AI botnet and its three-step process*

Let's examine these three steps in more depth to better understand what they entail.

## Data collection

Web scraping entails organizing data that's been extracted from a website, or websites, so that organizations may produce new datasets that can be applied and analyzed as they see fit. And it begins with gathering the data.

It is necessary for data collection to consist of adaptive scraping combined with "anti-ban" or "anti-bot-detection" technologies to operate quickly and smoothly. These technologies are set up as decision trees to detect various aspects of any protections that may be in place. Resiliency is the name of the game here. Bot protection may include JavaScript fingerprinting, HTTP and TLS fingerprinting (assessing the HTTP headers and TLS handshake), and Internet Protocol (IP) reputation detection (Figure 6). Some of these workflows may include machine learning (ML), especially when gathering statistics on the success rate; adjusting to the cookie strategy, HTTP header, and TLS parameters; and evaluating the JavaScript fingerprinting code. This is also where a headless browser may come into play.

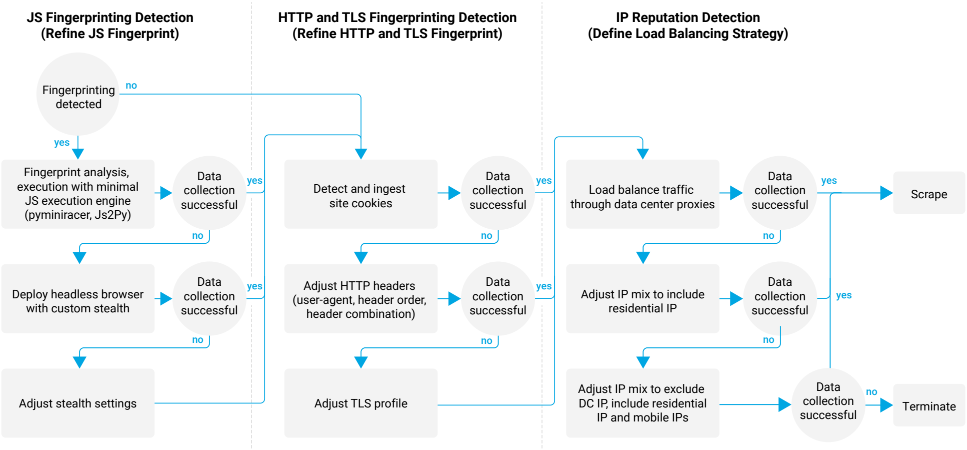

JS Fingerprinting Detection
(Refine JS Fingerprint)
HTTP and TLS Fingerprinting Detection
(Refine HTTP and TLS Fingerprint)
IP Reputation Detection
(Define Load Balancing Strategy)
Fingerprinting detected
no
yes
Fingerprint analysis, execution with minimal JS execution engine (pyminiracer, Js2Py)
Data collection successful
yes
no
Deploy headless browser with custom stealth
Data collection successful
yes
no
Adjust stealth settings
Detect and ingest site cookies
Data collection successful
yes
no
Adjust HTTP headers (user-agent, header order, header combination)
Data collection successful
yes
no
Adjust TLS profile
Load balance traffic through data center proxies
Data collection successful
yes
no
Scrape
Adjust IP mix to include residential IP
Data collection successful
yes
no
Adjust IP mix to exclude DC IP, include residential IP and mobile IPs
Data collection successful
no
Terminate

*Fig. 6: When attempting to collect data, this anti-bot-detection decision tree tries to avoid JavaScript fingerprinting, HTTP and TLS fingerprinting, and IP reputation detection*

## The browser with no head

A headless browser is a web browser that doesn't have a graphical user interface (GUI). This means that humans cannot interact directly with the web page on which the headless browser appears, and the browser is instead executed via a command-line interface (CLI) or by a network communication. In the case of Selenium, a popular open-source headless browser, the browser is automated and widely used for web scraping. This can be very helpful for data seekers who are attempting to scrape dynamic content.

Headless browsers can also allow for screenshots and website code to be copied efficiently, and for the chosen data to be extracted without rendering the entire page. However, headless browser attacks are expensive to conduct and can sometimes still be detected by the fingerprints they leave behind. Expenses for other sophisticated infrastructure, however, are similar to those of headless browsers; that is, generally high.

## Data extraction and data processing

The extracted information generally consists of HTML and JSON content. Of all the data extracted, only a fraction of it may be helpful for the analysis. For example, competitive analysis usually includes prices, discounts, inventory, and product SKU numbers, categories, and descriptions. Essential pieces of information may be extracted automatically by ML models that can be trained with multiple structures and data formats to recognize it. This helps avoid all the extra processing work that must be done to manually extract the data and helps avoid the requirement to study the HTML and JSON content code structure. Furthermore, the content code structure may change as the design of the site evolves. Additional ML logic is also necessary for processing if multiple websites are involved as part of the analysis scope.

## Case study: Benefits of web scraping detection solutions

Akamai researchers observed a subset of ecommerce customers who were protected by a web scraping solution that was detecting scraping activities, and looked at the traffic activity breakdown for one week. This amounted to a sample size of approximately 6.9 billion requests. The analysis only took into account HTML and AJAX requests. The static content (images, JavaScript, style sheets) was not included in the analysis since most bots do not request static content; this omission also helped avoid unnecessarily inflating the data.

The overall activity was classified by Akamai Content Protector and consisted of 49.3% low-risk human traffic, 42.1% bot traffic (27.5% high-risk bad bots and 14.6% good bots), and 8.7% medium-risk unclassified traffic (Figure 7).

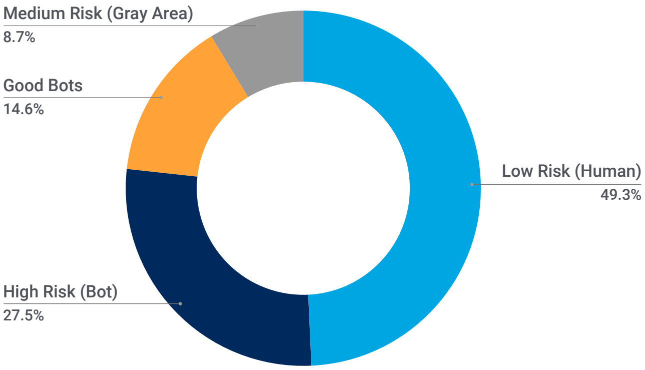


*Fig. 7: Traffic activity classification breakdown*

Figure 8 shows that of the 42.1% of traffic that was from bots, 65.3% originated from scrapers that are considered to be bad bots, and the remaining 34.7% were from scrapers that are classified as good bots (e.g., web search engines, SEO, social media, and online advertising).

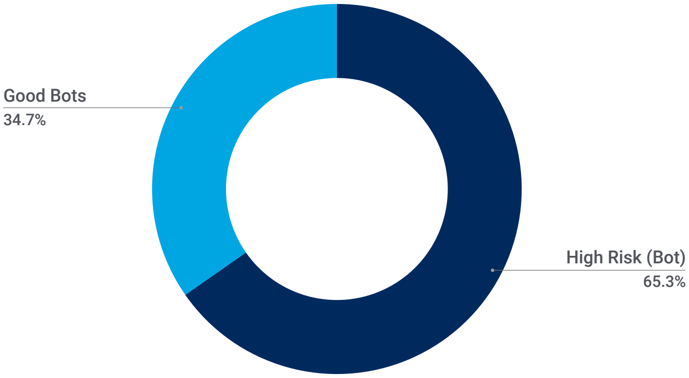


*Fig. 8: Good bot traffic vs. bad bot traffic*

The levels of sophistication were also measured for the high-risk bad bots that contributed to 65.3% of overall bot traffic. Thirty-seven percent of that traffic came from basic scripted botnets that are easy to detect through simple stateless methods, 47.6% came from more advanced scripted botnets that require more advanced stateful detection methods using ML, and 15.5% came from headless browsers that require advanced JavaScript fingerprinting and stateful detection methods (Figure 9).

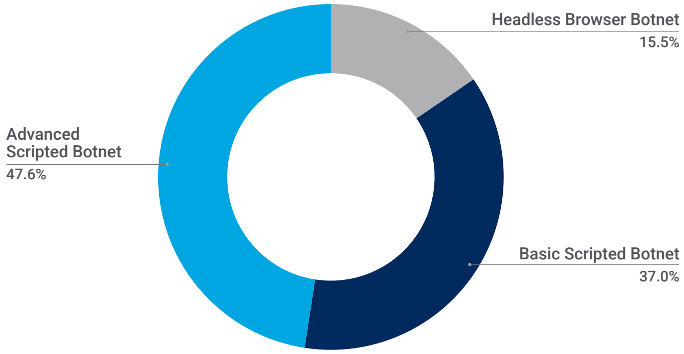


*Fig. 9: Bad bots traffic distribution based on their sophistication (totals do not sum to 100% due to rounding)*

So, from this data, we can see that bad bot scrapers are significantly more numerous than good bot scrapers, and that close to half the overall traffic consisted of bots, with the advanced scripted botnets producing the most bad bot traffic (47.6%).

Website activity will run much faster and more efficiently, and site metrics will be cleaner to read, once defenses against these bots are in place and scrapers are removed. And these outcomes will result in better user/customer experiences. As shown in Figure 10, the number of high-risk bot requests decreased substantially once the mitigation was activated.

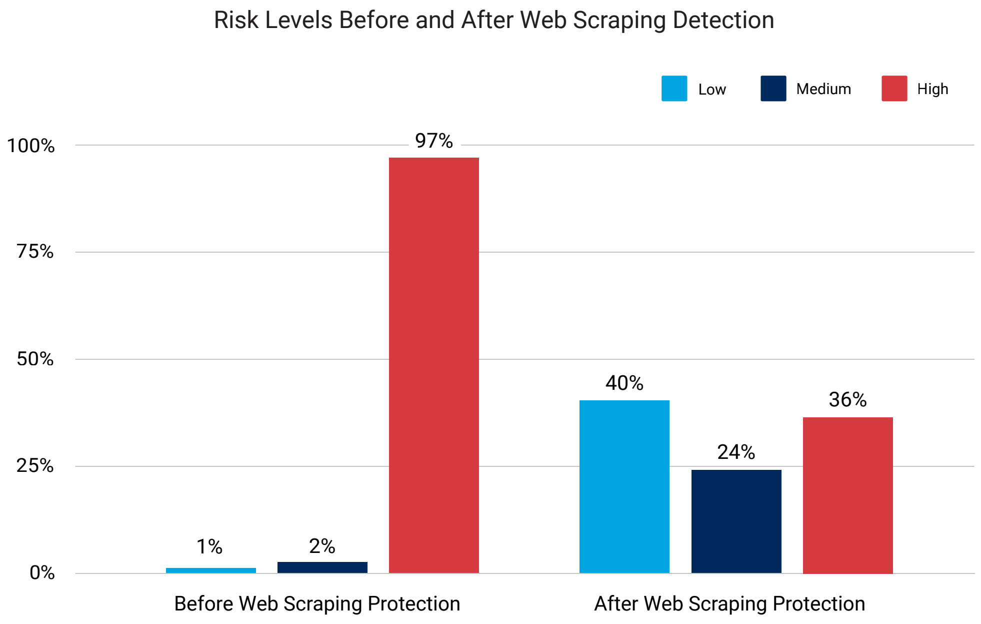


*Fig. 10: Risk levels before and after mitigation with Content Protector*

## Safeguarding and mitigating

This section provides some crucial indicators in detecting web scrapers, and information about tools that can provide defensive measures against them.

### Detecting basic scrapers

Although sophisticated scrapers may be difficult to detect, bot management solutions can defend against data being collected by all kinds of intrusive scrapers and can especially look out for the following characteristics to detect simpler web scraper bots:

- Requests that advertise older browsers and OS versions
- Anomalies in the HTTP header signature
- The use of old versions of HTTP (e.g., v1.1) instead of the more common HTTP v2 or the emerging HTTP v3
- Requests that come from thousands of cloud services/data centers

## Detecting more advanced scrapers

None of the characteristics in the list above will be observable for the more advanced scrapers. So, here are some characteristics for the more sophisticated scapers:

- Requests that come from the latest browser and OS version
- The HTTP header set looks identical to the legitimate browser
- The use of HTTP v2
- Requests that come from hundreds of thousands of residential and mobile IP addresses

## Identifying traffic patterns

There are some key indicators that can identify whether the type of traffic a website is incurring is human (Figure 11), basic bot (Figure 12), or sophisticated bot (Figure 13).

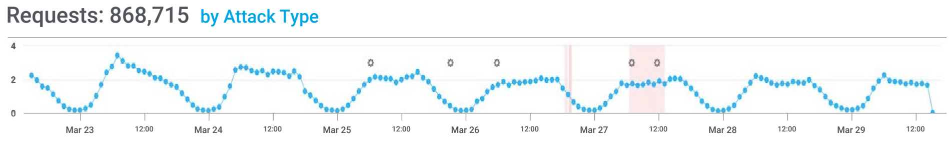


*Fig. 11: Legitimate user traffic generally shows a circadian cycle of activity*

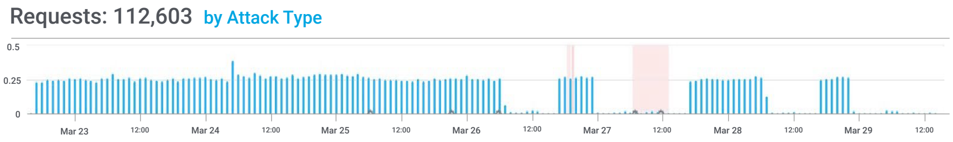


*Fig. 12: Typical bot traffic displays regular activity with occasional breaks*

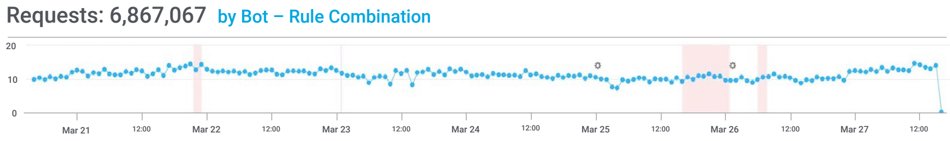


*Fig. 13: More sophisticated bots show traffic continuously day and night*

We often also see botnets that are somewhere in the middle, with a weak load balancing strategy but a sophisticated fingerprint strategy (or vice versa). However, more advanced botnets may be so sophisticated that they can pass as having a perfect fingerprint or even reproduce a legitimate user traffic pattern.

In addition to being on the lookout for these scraper bots, tools that protect against web scraping, such as a content protector, may allow for special benefits and smoother sailing among the choppy scraper-infested waters. Benefits may include:

- Higher conversion rates and reduced IT costs
- More accurate metrics, which can lead to better investment decisions and drive revenue increase
- Reduced pricing pressure, which can translate to sales saved from competitor undercutting
- Happy customers who can access desired goods, and increased revenue from upsell opportunities when customers add additional products to cart once they've secured the hot item
- Preserved brand reputation as customers are protected from poor quality fakes that they think are legitimate goods from the original seller
- Retained product revenue and maintained customer loyalty
- Increased/protected ad revenue
- Retained audience and site visitors

## Compliance considerations

Payment Card Industry Data Security Standard (PCI DSS) v4.0 is now in effect, and many of the changes were driven by threat trends that are still having an impact on companies. Visibility is key to addressing these attacks. Whether they are in your historical JavaScript environment or APIs used to facilitate transformation, it is critical to rapidly detect and remediate these attacks.

We also see emerging compliance trends in the new NIST Cybersecurity Framework version 2.0, which has added a governance function. NIST tends to be a foundation for a number of government regulations and bleeds into many commercial cybersecurity frameworks. So, now is a great time to review the new guidance and either use it to update your policies or map your current documentation to see where you have gaps.

For publicly traded companies and those using generally accepted accounting principles (GAAP), another area of compliance is cybersecurity materiality. The need to define material risks and threats requires collaboration across the leadership team. Once you identify material threats (like ransomware), you need to map mitigations (such as microsegmentation). Make sure your crisis management plans address disclosure timelines and be sure to have a playbook for the worst case scenario for which you would need to file an Security and Exchange Commission Cyber Incident Form 8-K.

## Conclusion

We hope this report gives you insight into an area that could be having a negative economic impact on your organization. Bots are affecting your sites in ever-increasing volume, and it is important to optimize beneficial bots, mitigate malicious bots, and ensure low friction in the overall customer experience. This is a security issue with business impacts. As with all security issues, the first step is gaining visibility, the second step is analyzing impact, and the final step is determining ROI for risk and revenue so you can implement appropriate security controls.

You can't protect what you can't see, so now is the time to determine where you have gaps in visibility. To do this, you must determine the level of web scraping activity on your sites — and its intent. Both good bots and bad bots compose the bot landscape, and scraper bots are in both categories, depending on their use. Although the line between beneficial and harmful scraper bots can be blurry, the evolution of bot sophistication (e.g., web scrapers conducting headless browser attacks) continues. This all comes with the immense impact web scraper bots have among ecommerce entities on both IT costs and customer experience. It is key to make sure you have the tools in place to analyze the bot activity and impacts on your site.

What you don't want is attackers who execute their criminal business model on your sites and commit a variety of malicious activities, like cashing out loyalty points, placing fraudulent orders, or even conducting return fraud. You also don't want ticket bots to buy out limited events or shopping bots to buy hot products. Bots can be used to facilitate new account opening abuse by taking advantage of special offerings, which impacts campaign analysis and costs. Large Distributed Denial-of-Service (DDoS) botnets can overwhelm web-facing applications and cause a poor user experience or the inability to place orders or make reservations, resulting in lost revenue and customer friction. Bots can even mimic human behavior online to increase clicks and traffic on a website, skewing both the marketing and performance analytics of carefully crafted digital experiences. You definitely don't want any of that.

As we noted earlier, more than half of the global commerce web traffic is made up of bots, and the bot traffic levels continue to rise. Akamai has based the insights and advice in this report on our security platform, which includes content protection with defense against web scraping. We partner with many ecommerce leaders, so we wanted to share safeguards and mitigations that companies can use to best protect their customers. We anticipate an increase in the use, service level choices, and types of available web scraper bots. Therefore, it is necessary to continuously evaluate your company's risk posture and determine if your current security controls are meeting your leadership's risk appetite.

Stay plugged into our latest research by checking out our security research hub.

## Methodologies

### Content Protector data

This data sample describes the risk level classifications our Content Protector tool assigns to the traffic it monitors. These classifications are used to detect both bot scraping activities and to determine whether we are dealing with a good or bad bot. Since most bots do not request static content, this analysis only took into account HTML and AJAX requests to avoid unnecessarily inflating the data.

*This data sample covered the one-week period from April 12 through April 19, 2024. Our total sample size consisted of more than 6.5 billion requests.*

### Bot attacks

This data describes application-layer alerts on traffic seen through our web application firewall (WAF) and bot management tool. The bot alerts are triggered when we detect a bot payload within a request to a protected website, application, or API. These bot alerts can be triggered by both malicious and benign bots. The alerts do not indicate the successfulness of an attack. Although these products allow a high level of customization, we collected the data presented here in a manner that does not consider custom configurations of the protected properties. The data was drawn from an internal tool for analysis of security events detected on Akamai Connected Cloud, a network of approximately 340,000 servers in more than 4,000 locations on nearly 1,300 networks in 130+ countries. Our security teams use this data, measured in petabytes per month, to research attacks, flag malicious behavior, and feed additional intelligence into Akamai's solutions.

*This data covered the 15-month period from January 1, 2023, through March 31, 2024.*

## Credits

### Editor in chief

Lance Rhodes

### Editorial and writing

David Senecal
Maria Vlasak

### Review and subject matter contribution

Mitch Mayne
Susan McReynolds
Christine Ross
Badette Tribbey
Steve Winterfeld

### Data analysis

Chelsea Tuttle

### Promotional materials

Annie Brunholzl

### Marketing and publishing

Georgina Morales
Emily Spinks

## More State of the Internet/Security

Read back issues and watch for upcoming releases of Akamai's acclaimed State of the Internet/Security reports. **akamai.com/soti**

## More Akamai threat research

Stay updated with the latest threat intelligence analyses, security reports, and cybersecurity research. **akamai.com/security-research**

## Access data from this report

View high-quality versions of the graphs and charts referenced in this report. These images are free to use and reference, provided that Akamai is duly credited as a source and the Akamai logo is retained. **akamai.com/sotidata**

## More on Akamai solutions

To learn more information on Akamai solutions for detecting and protecting against web scrapers, visit our **Content Protector page**.