NETACEA

Technical Showcase

How Netacea Delivers Rapid Results with Advanced Bot Detection

Introduction

Why Your Business Needs Advanced Bot Management

Business logic exploits are leaving businesses wide open to myriad attacks, including scraping, scalping, credential stuffing, carding, and account takeover, to name just a few.

The best defense against advanced bot attacks is advanced bot detection technology. By analyzing all traffic coming into your website, mobile apps and APIs, and applying powerful anomaly detection algorithms, these tools can classify traffic as 'good bots', 'bad bots' or human – and then prevent bad actors from carrying out their attacks.

Alongside protection layers like web application firewalls (WAFs) and distributed denial of service (DDoS) protection, bot management is an essential part of any web security stack.

Why You Must Prove The Value Of Bot Management Quickly

It's crucial to prove ROI for security solutions because they often cost a great deal, require board buy-in, and are notoriously difficult to measure success.

The ultimate value of a threat mitigation tool like Netacea Bot Management is in stopping attacks, but there's also value in describing how this is done, and by outlining the details of the mitigated attacks:

- What was the intent of the attack?
- How big of an attack was it?
- How much damage might it have caused?
- How sophisticated was the attack?

Many businesses assume that because no attacks affected a business after investing in defenses, they don't need to rely on their protection solutions. In fact, the opposite is often true. By giving a deeper level of detail, even in snapshots, security teams can quickly gain buy-in for the solution across the business.

The first phase of engagement with any third-party solution, especially one responsible for protecting web traffic and potentially even customer data, is running a POC/POV (proof of concept/proof of value).

How to Run a Successful POC

To run detailed and accurate analysis of bots in web traffic, first we need raw data to process. We use web logs so we can investigate every single request made to clients' web servers and gain a full picture of web traffic.

The first step is to gain access to the customer's web logs, either as a live feed via a simple integration, or by looking at historic data from a previous period. Netacea Bot Management can detect bots with as little as a day's worth of data, but to provide detailed analysis, we investigate a minimum of one week's worth of data. We recommend four weeks' worth of data for a full POC.

Data is then cleansed, parsed and input into our Intent Analytics® engine, as well as being investigated by our team of bot experts, as it would be in a full-scale ongoing engagement with any customer.

After each week of the POC, we present a report to the client revealing the bot activity discovered. Following the fourth and final week of analysis, we deliver a summary report answering any queries from the customer.

While Netacea's proof of concept cycle typically involves four weeks of analysis and reporting, we sometimes find customers wish to move from POC into full on-boarding after just one week, based on our findings and actions within that timeframe.

Breaking Down the POC Process

Cleaning Up Web Log Data For Analysis

Business logic attacks are defined as exploits using the intended functions of a system for malicious purposes. For example, many websites legitimately require users to access accounts via a login page, but bots can exploit this intended functionality to flood login pages with credential stuffing attacks, and takeover accounts for malicious purposes.

To carry out business logic attacks, every bot must make HTTP requests to web servers. Every HTTP request contains information about the method of the request (path, parameters, protocol etc.) and the traffic host, user-agent, language etc. within the request headers.

At Netacea, we use these requests to build a profile of all traffic interacting with the client's systems. Unlike client-side detection, web log data can't be manipulated by malicious actors; if they wish to carry out their intended purpose, they must make requests, even if they attempt to spoof their origin or identity.

These requests are recorded by the client's servers as web logs, which businesses store in different formats (the most popular formats being JSON or text files). As a result, Netacea's ingest process involves parsing non-standard web log data into our standard format to ensure it is clean and ready to be analyzed.

```
>>> data.head(10)
                   client              timestamp status response_size method                                      path  protocol
0           in24.inetnebr.com  01/Aug/1995:00:00:01    200          1839    GET  /shuttle/missions/sts-68/news/sts-68-mcc-05.txt  HTTP/1.0
1             uplherc.upl.com  01/Aug/1995:00:00:07    304             0    GET                                         /  HTTP/1.0
2             uplherc.upl.com  01/Aug/1995:00:00:08    304             0    GET                  /images/ksclogo-medium.gif  HTTP/1.0
3             uplherc.upl.com  01/Aug/1995:00:00:08    304             0    GET                /images/MOSAIC-logosmall.gif  HTTP/1.0
4             uplherc.upl.com  01/Aug/1995:00:00:08    304             0    GET                  /images/USA-logosmall.gif  HTTP/1.0
5  ix-esc-ca2-07.ix.netcom.com  01/Aug/1995:00:00:09    200          1713    GET                  /images/launch-logo.gif  HTTP/1.0
6             uplherc.upl.com  01/Aug/1995:00:00:10    304             0    GET                /images/WORLD-logosmall.gif  HTTP/1.0
7         slppp6.intermind.net  01/Aug/1995:00:00:10    200          1687    GET              /history/skylab/skylab.html  HTTP/1.0
8        piweba4y.prodigy.com  01/Aug/1995:00:00:10    200         11853    GET                  /images/launchmedium.gif  HTTP/1.0
9         slppp6.intermind.net  01/Aug/1995:00:00:11    200          9202    GET            /history/skylab/skylab-small.gif  HTTP/1.0
>>>
```

Examples Of Bot Traffic Analysis

With web logs parsed and ready to analyze, we can start looking at patterns and trends in the data to spot anomalous behavior. Because some business logic attacks can be very short lived (yet very damaging), they need to be blocked quickly. By using a combination of detection techniques and analytical approaches, Netacea Bot Management can detect and block attacks within seconds.

After our bot experts have investigated the anomalies spotted by our various detection models, we present our findings in reports to provide the POC customer with a full view of all anomalous traffic identified within the given timeframe.

Here are just a few examples of analysis methods we use:

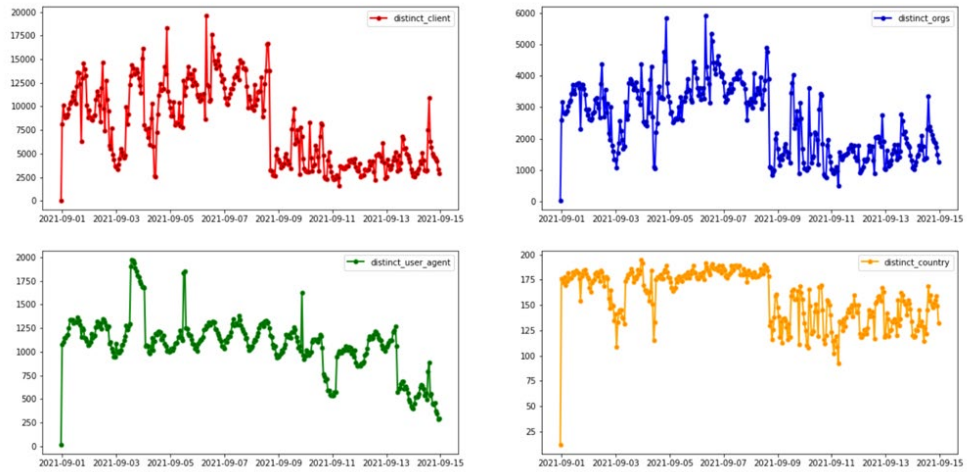
Investigating Peaks

Using a time plot of the volume of traffic, we can spot volumetric attacks in a blunt way. These attacks are usually the least sophisticated but can nevertheless cause significant damage quickly if unmitigated.

It is important to note that spikes in web traffic could also be the result of activity by real visitors, for example interest in a new product release, the customer's own marketing efforts or any other non-malicious cause for increased traffic. We discuss these peaks with the client to ensure there are no false positives and question whether any dramatic increases in traffic are in line with expectations.

Distinct Counts of Values Over Time

In more sophisticated attacks, bots hide in human traffic not only by interacting with pages and sites like humans, but also by distributing their origins, for example, by using multiple IP addresses or spoofing a range of user agents.



By cross referencing these data points we can see patterns and distinguish distributed traffic acting in the same way, despite coming from different countries, devices, data centers or IP addresses.

Login Analysis

To detect credential stuffing attempts, we investigate activity on login pages. This is done by analyzing login attempts, successful logins, failed logins and the login failure rate over time.

To quickly pinpoint attacks, we set a failure rate threshold based on a distribution estimation. If a user exceeds the threshold (for example, within an hour they make 100 login attempts and 90% are unsuccessful) we can use this to identify anomalous traffic.

The successful login rate is highly effective in uncovering account takeover (ATO) attacks. These usually occur after a successful credential stuffing attack, during which user details are tested in high volume to narrow down a list of credentials proven to work. The successful login rate, especially when paired with a prior login failure rate, can be an indicator of ATO activity as well as fake account creation attacks.

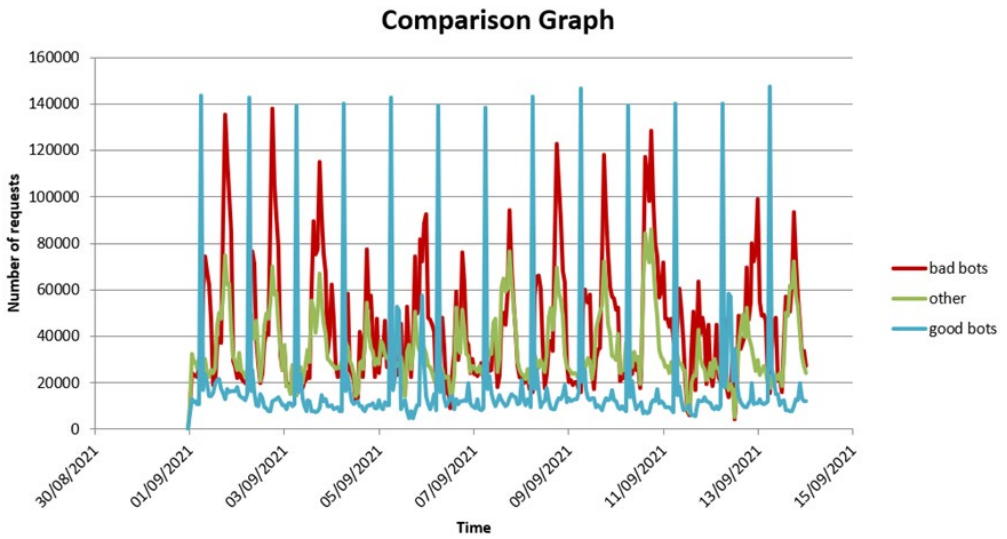
Good Bots Vs Bad Bots Vs Other

Another resource available to all Netacea customers is our Active Threat Database. This dynamic list is constantly kept up to date with all known traffic classifications identified across all Netacea customers, to keep our whole client base in line with all known threats.

Not only does the Active Threat Database quickly identify bad bots (those with malicious intent), but 'good bots' can also be identified.

These could be search engine crawlers such as Googlebot, or social media scrapers looking to syndicate content. However, as it's possible for the bad bots to spoof the known good ones to bypass defenses, we investigate suspicious good bots by finding out where the traffic is really hosted and if this matches their declared identity.

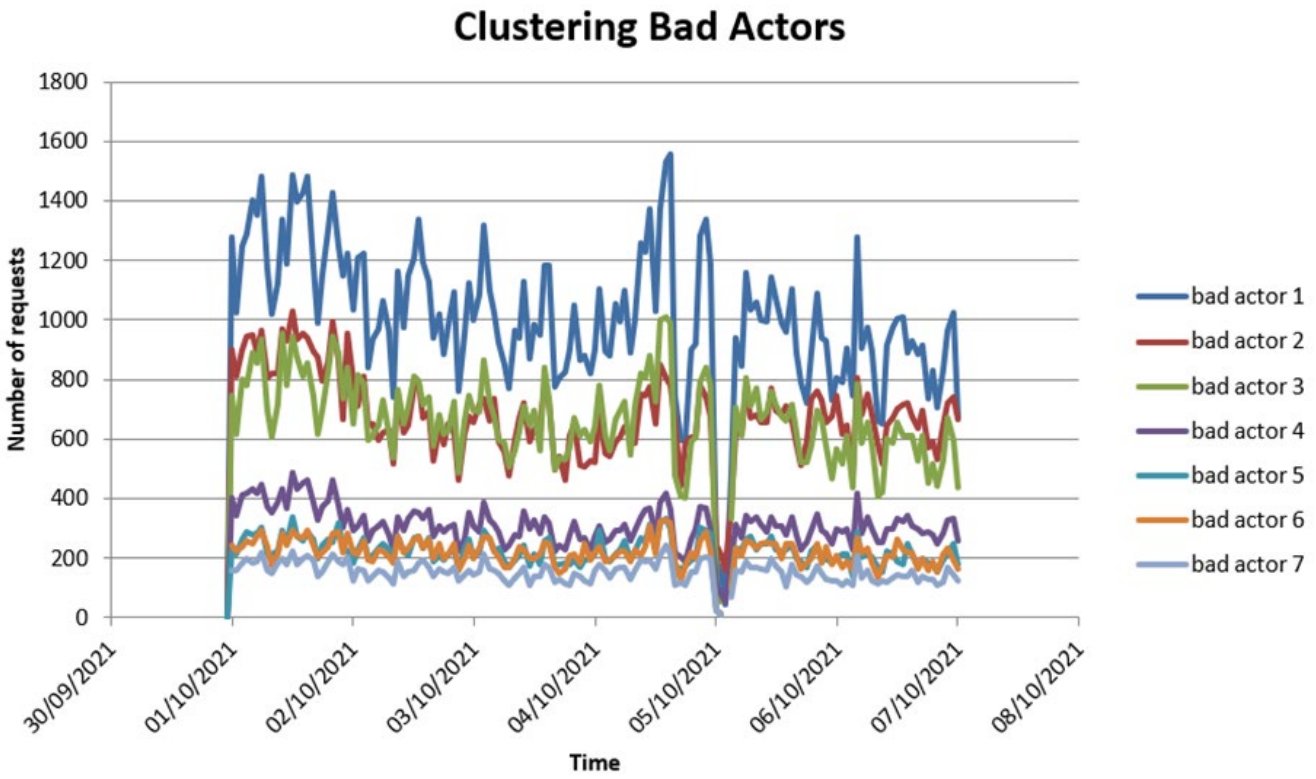
In this chart, 'other' could be humans or an as-yet unidentified bot type:



Traffic Clustering

Each user making requests to a website leaves behind a unique pattern of behavior. By comparing these using an unsupervised machine learning algorithm, we can cluster together visitors behaving in the same way. This allows us to quickly see how similar different actors are from their behavior and classify their intent.

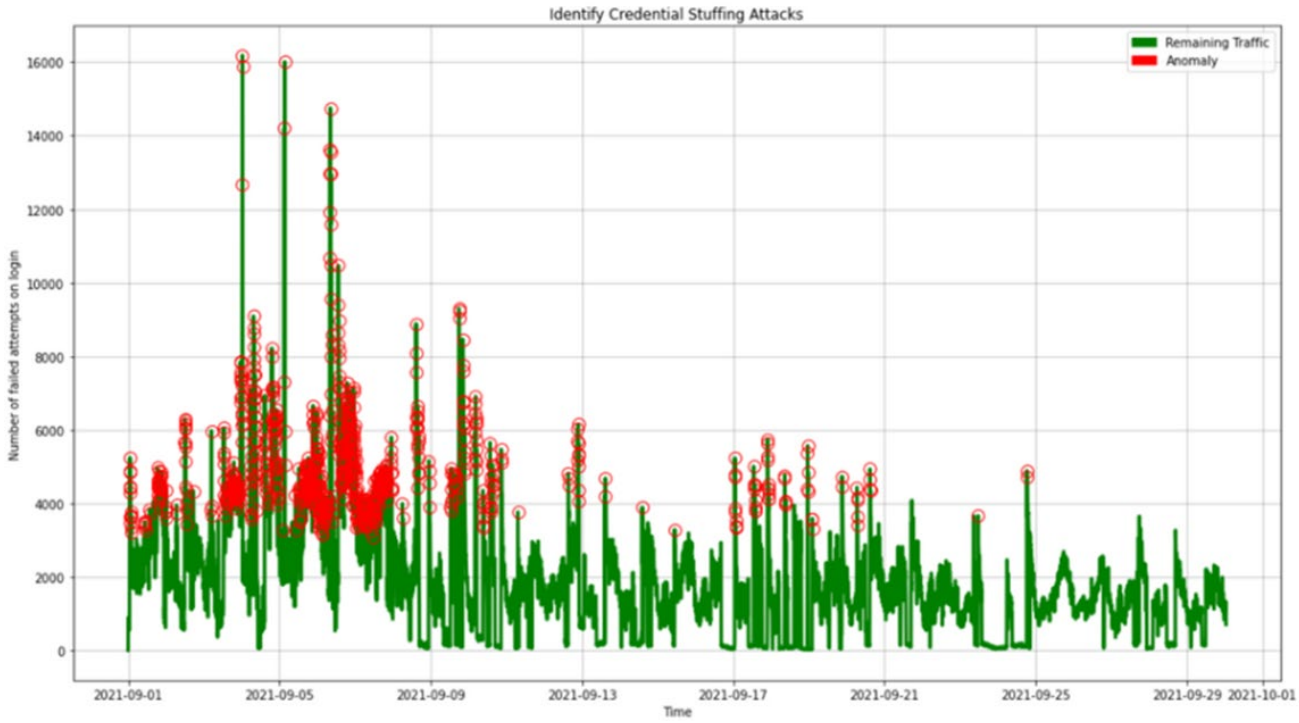
'Unsupervised learning' is a type of machine learning where the algorithm is not provided with any labels to start with, which allows it to self-discover naturally occurring patterns, in our case by clustering similar behaving traffic together.



Anomaly Detection with S-H-ESD Algorithm

Another machine learning algorithm used at Netacea during POCs is the S-H-ESD (Seasonal Hybrid Extreme Studentized Deviate) algorithm. As well as 'global' anomalies that fall outside the average minimum and maximum volume threshold for each time of day, the S-H-ESD algorithm is especially good at identifying 'local' anomalies hiding within seasonal traffic. This makes it a powerful tool to combat bots sophisticated enough to try to conceal themselves amongst usual peaks in traffic. This allows us to spot not just volumetric attacks, but also small, distributed attacks, very quickly.

S-H-ESD and clustering are just two examples of many algorithms we use to detect anomalous and malicious traffic.



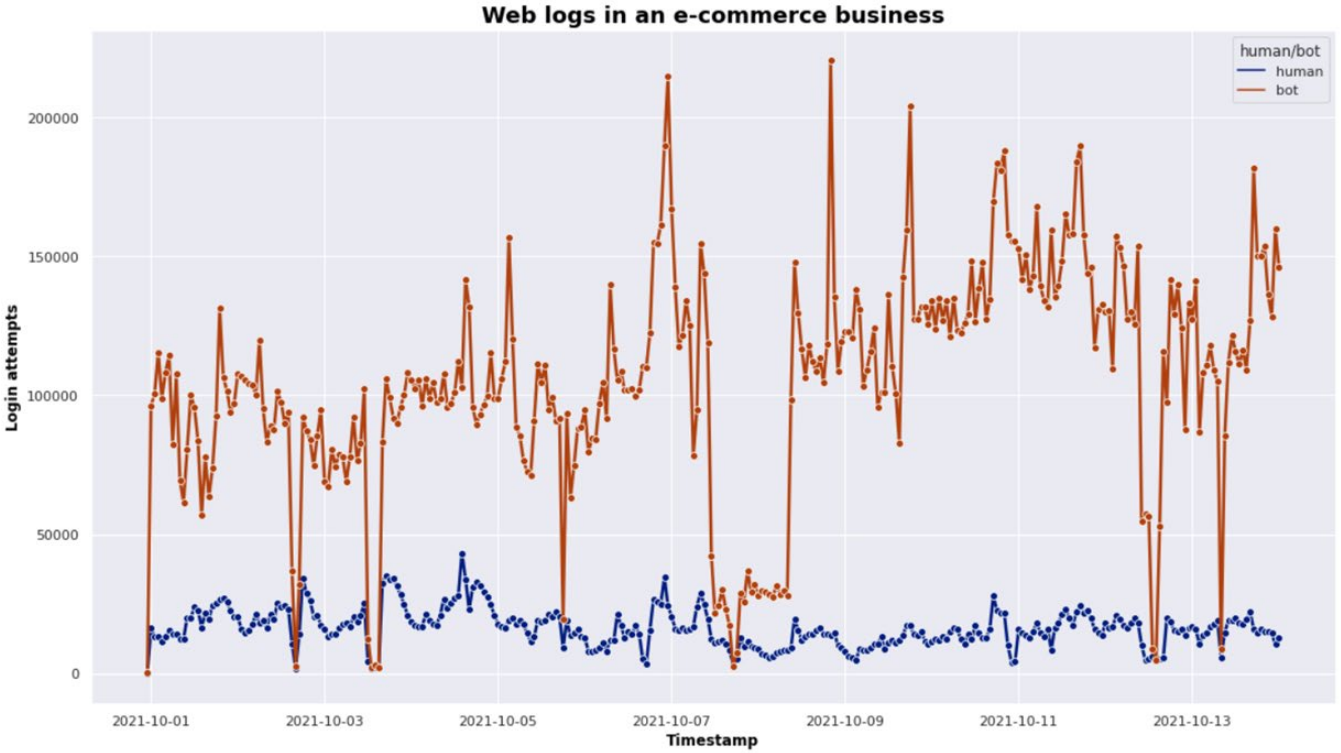
Examples of Bad Bot Behavior in eCommerce

To illustrate typical findings at the POC stage, we can look at eCommerce as an example of an industry heavily targeted by bots. In fact, we have observed an average of 60% of web traffic to eCommerce sites as being malicious bots, most commonly scraping information such as prices, inventory, content, and availability of products, or attempting to scalp products for resale elsewhere.

NETACEA

Another common attack targeting eCommerce businesses is credential stuffing, which focuses on login pages to attempt to validate credentials obtained from the dark web or from a data leak elsewhere. Some attacks are highly volumetric and originate from a small group of IP addresses sent from a single data center, which is easier to spot, but others are highly distributed across multiple points of origin. It is hard to block these attacks if only looking at one point of origin, for example IP addresses, because the traffic is so dynamic and distributed, however when applying multiple levels of blocking (IP, user agent, data center or geolocation) it is straightforward for us to mitigate these kinds of attacks.

We blocked more than 80% of traffic in this very sophisticated credential stuffing attack.



NETACEA

Summary

Netacea Bot Management ingests and processes billions of requests every day. Detecting new attacks allows us to improve our existing machine learning and AI-based detection models and advance the product continually. Each new attack we analyze also teaches our bot experts and data analysts something new as well.

Try Netacea Bot Management For Your Business

Do you want to find out how much more bot traffic you could be blocking? Get in touch to arrange an in-depth proof of concept of Netacea Bot Management for your website, mobile apps and APIs.