



Learn How to Defend Your Online
Marketplace from Unwanted Traffic



Rami Essaid

CEO & Co-founder



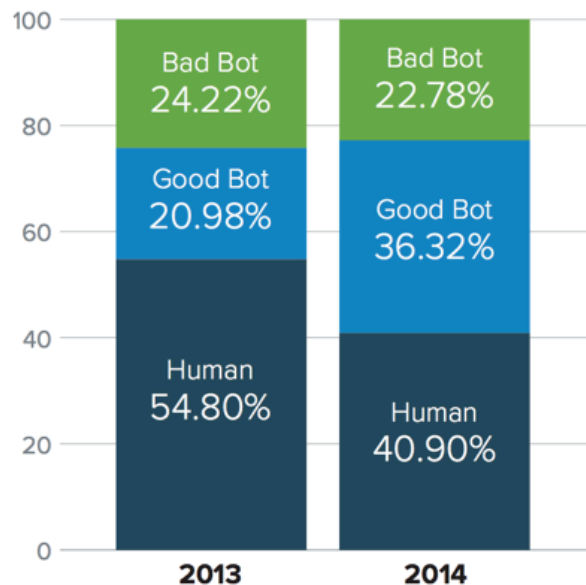
Distil in Classifieds



The Basics of Bots

A “**Bot**” is an automated program that runs on the internet

Bad Bot, Good Bot and Human Traffic
2013 and 2014



Good Bots

- Search Engine Crawling
- Power APIs
- Check system connectivity and status

Bad Bots

- Steal content
- Scan for vulnerabilities
- Perform Fraud
- etc.

Three Layers of Business Damaged Caused by Malicious Bots

Bots hurt your KPIs...

Slowdowns, downtime and poor CX

Decline in website traffic

Loss of revenue and customer base

When Bots Attack...



They Hurt Your KPIs by...

What Is Web Scraping?

Web Scraping

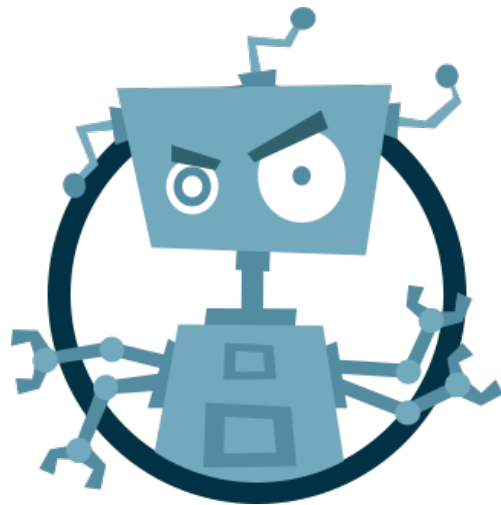
Also known as screen scraping, web scraping is the act of copying large amounts of data from a website – either manually or with an automated program.

Legitimate Scraping

Scraping can sometimes be benevolent and totally acceptable. For example, the search engine bots that index your website

Malicious Scraping

A systematic theft of intellectual property accessible on a website, including pricing, content, images, and proprietary data

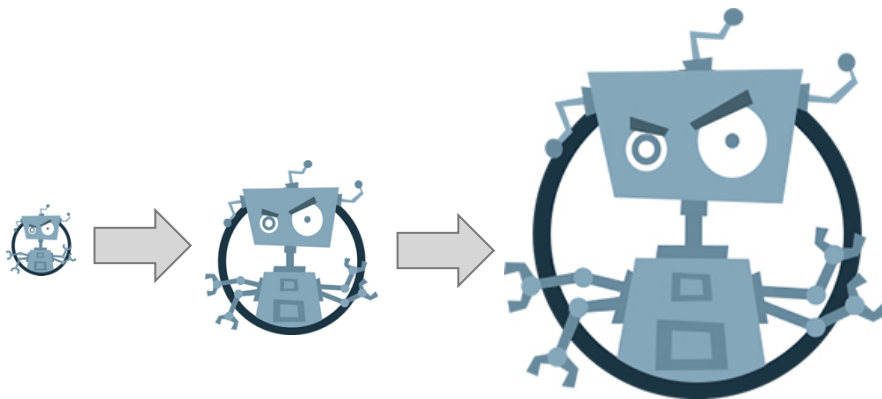


What is Contributing to the Growth in Web Scraping?

Cheap scraping software

Inexpensive cloud computing resources

Botnet-as-a-Service



Scraping Online Marketplaces Has Never Been Easier

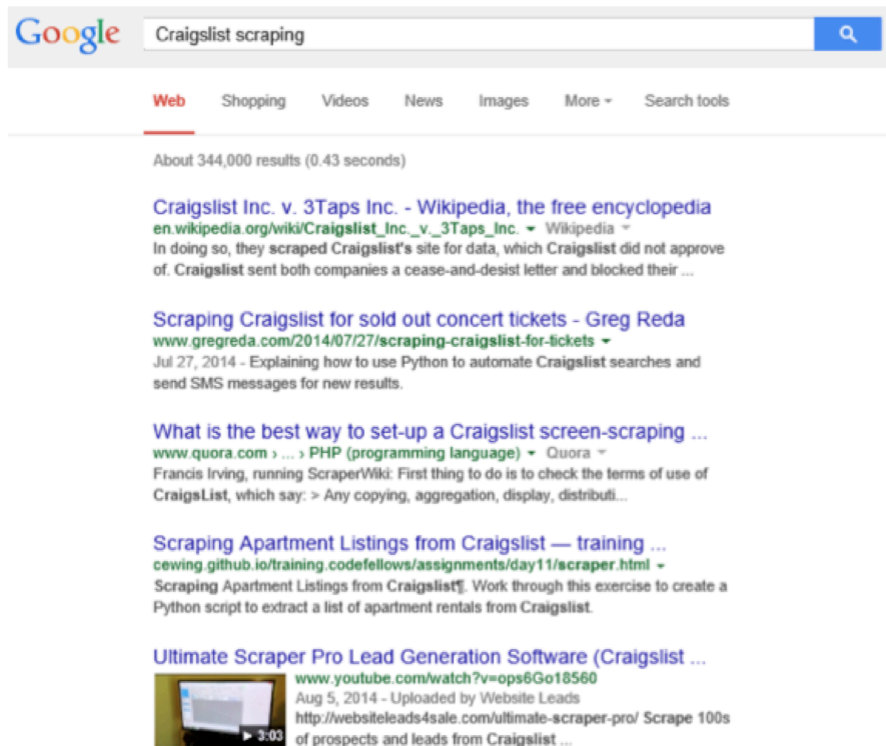


Scraping resources just a click away...

Anyone with basic computer skills can get into the game

Inexpensive relative to the value of the content they steal

Difficult or impossible to prosecute



The Going Rate for Scraping Less than \$130/day

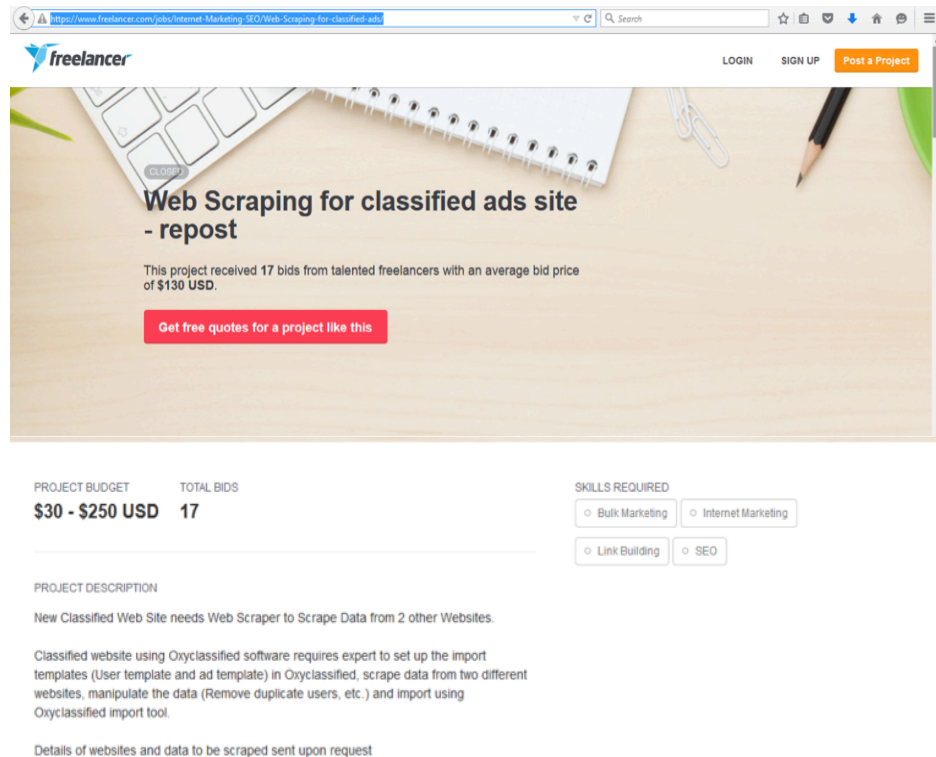
Freelancer.com Rates

Scraping two classified ad websites

Data Manipulation (de-duping, etc.)

Importing into Oxyclassified

Average Cost - \$130 USD



The screenshot shows a Freelancer.com project listing. The title is 'Web Scraping for classified ads site - repost'. Below the title, it states: 'This project received 17 bids from talented freelancers with an average bid price of \$130 USD.' There is a red button that says 'Get free quotes for a project like this'. The project details section shows a budget of '\$30 - \$250 USD' and 17 total bids. The skills required are listed as Bulk Marketing, Internet Marketing, Link Building, and SEO. The project description states: 'New Classified Web Site needs Web Scraper to Scrape Data from 2 other Websites. Classified website using Oxyclassified software requires expert to set up the import templates (User template and ad template) in Oxyclassified, scrape data from two different websites, manipulate the data (Remove duplicate users, etc.) and import using Oxyclassified import tool. Details of websites and data to be scraped sent upon request'.

freelancer

LOGIN SIGN UP Post a Project

Web Scraping for classified ads site - repost

This project received 17 bids from talented freelancers with an average bid price of \$130 USD.

Get free quotes for a project like this

PROJECT BUDGET TOTAL BIDS

\$30 - \$250 USD 17

SKILLS REQUIRED

- Bulk Marketing
- Internet Marketing
- Link Building
- SEO

PROJECT DESCRIPTION

New Classified Web Site needs Web Scraper to Scrape Data from 2 other Websites.

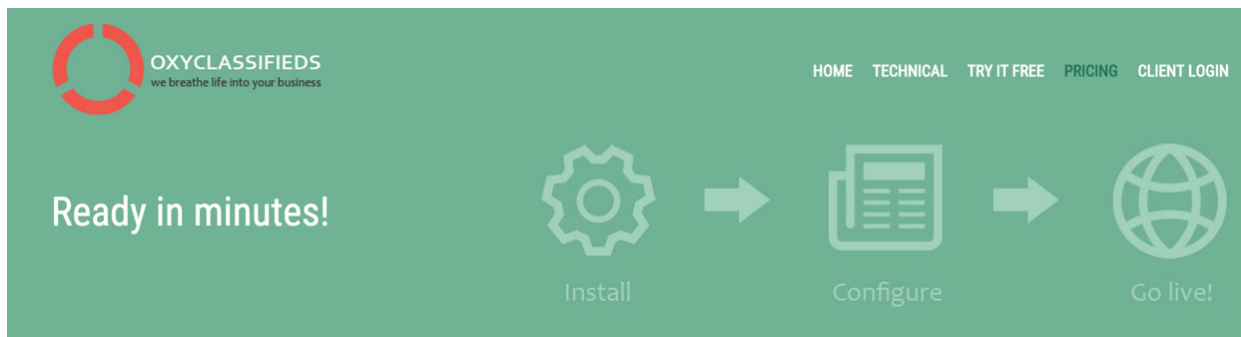
Classified website using Oxyclassified software requires expert to set up the import templates (User template and ad template) in Oxyclassified, scrape data from two different websites, manipulate the data (Remove duplicate users, etc.) and import using Oxyclassified import tool.

Details of websites and data to be scraped sent upon request

Posting Stolen Data is Quick and Easy due to Turnkey Platforms



Classified ad platforms start at \$201 including installation, upgrades, and



OxyClassifieds Standard Lifetime License €179

[Click to convert in USD](#)

- Lifetime license for one domain
- Free upgrades for 6 months
- Free support for 6 months
- Free install service

Buy OxyClassifieds Standard Version

€179

The Cost of Replicating your Website



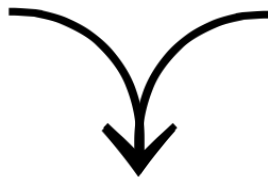
Scraped Data

\$130



Classified Ad Website

\$201



\$331

Competitive Scraping and Data Mining

Duplicating your Product Portfolio

Bots can easily gather product and supplier lists for replication elsewhere

Undermining your Prices

Bots monitor your prices, ensuring competitors can undercut with lower price listings

Availability Tracking

Identifying when your supply has been exhausted provides competitors a unique opportunity to raise the price of their goods



Duplicate Content Results in Diminished SEO

Negative SEO Attacks

Bots steal content, product lists, and prices for duplication elsewhere on the Internet

Duplicated content reduces your company's uniqueness and thus quality score

SEO damage may result, especially if

- Your prices are undercut
- The content is repurposed on a more popular site



Skewed Analytics Present an Inaccurate Picture of Website Traffic



Domains

Server Status

Reports

Settings

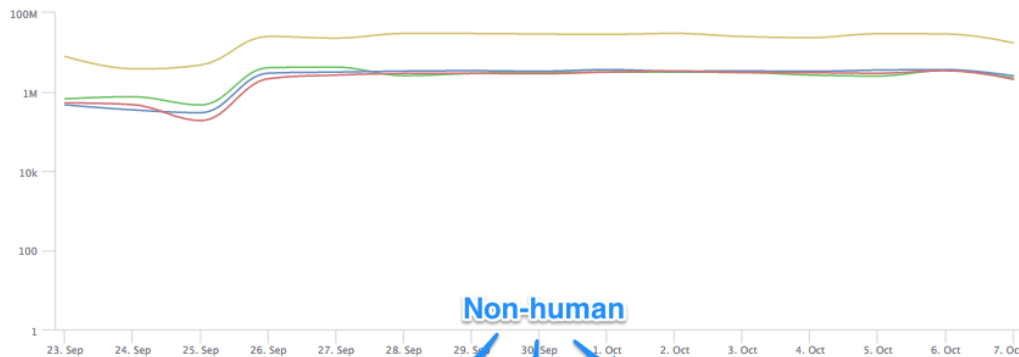
Traffic Overview ▾

Show results from:

Last 14 days ▾

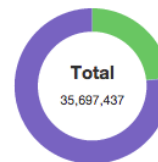
Human, Good Bot, & Bad Bot Page Requests

scale: Linear



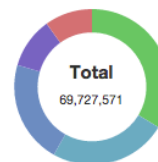
Humans	Good bots	Whitelisted	Bad bots	Total
334,082,490	38,940,823	40,589,822	35,686,468	449,299,603

How we protected you



- blocked 23.3%
- captcha 0.1%
- dropped 0.0%
- monitored 76.6%

What we protected you from

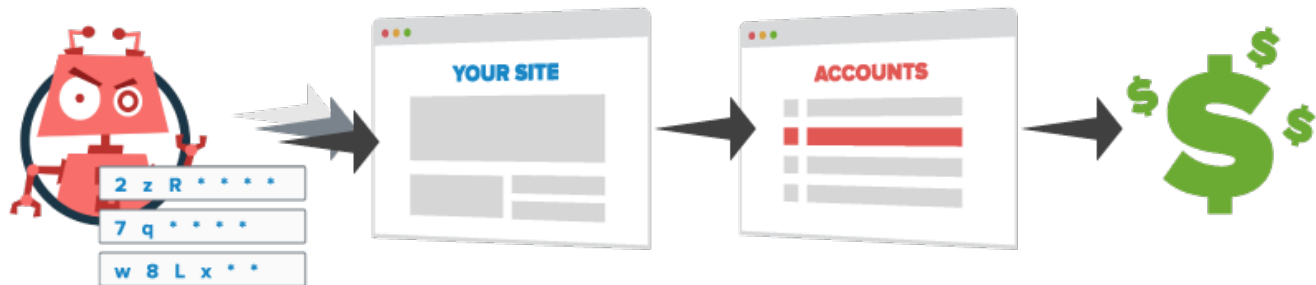


- Pages Per Session Exceeded 33.7%
- Session Length Exceeded 24.5%
- Rate Limited 21.4%
- JavaScript Check Failed 10.6%
- JavaScript Not Loaded 9.9%

Brute Force Account Takeover

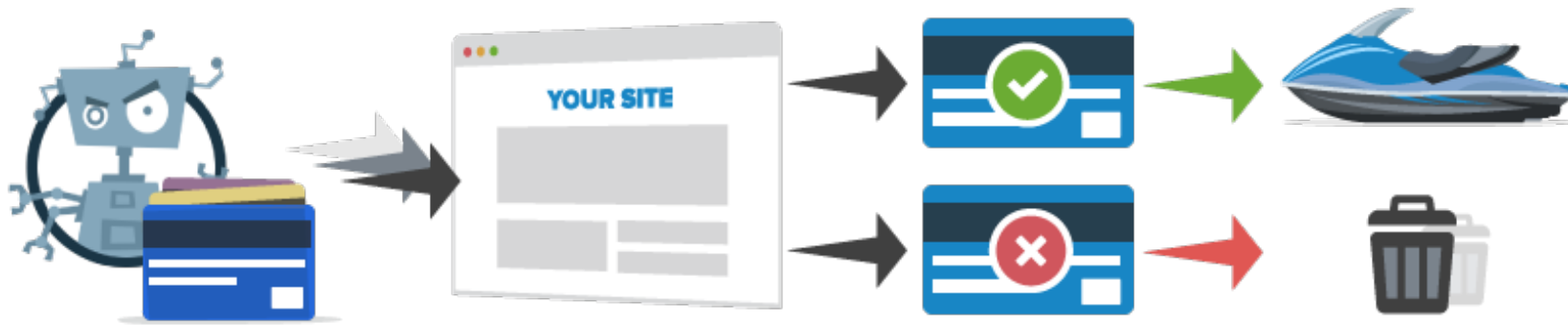
Using a bot to try stolen usernames and passwords from breaches at other websites on your site

Newly compromised accounts are then used for various forms of fraud/theft



Carding

Creating micro-transactions with stolen credit cards against e-commerce sites to test their validity



StubHub Case Study



StubHub is an online marketplace which provides services for buyers and sellers of tickets for sports, concerts, theater and other live entertainment events.

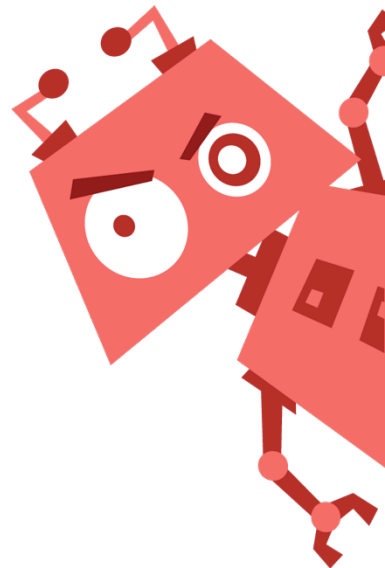
Largest secondary ticket marketplace in the world

An eBay company

Processes nearly 500 transactions per second

Bot Challenges

- Bots were used for brute force account takeovers
- Competitors tried to game the system, scraping prices, and monitoring inventory and customer behavior
- Random spikes in bot traffic were causing increased utilization of resources
- Tested multiple competitor solutions, but they were difficult to configure and in some cases broke our website



Bot Detection and Mitigation Solution Requirements

- Block web scrapers without impacting human visitors
- Accurately identify good bots vs. bad bots
- Cannot solely rely on rule based system
Must include automated learning to “self tune” for defending against emerging and unknown threats
- Needs to include Distil community to improve accuracy of bot detection
- Must seamlessly co-exist with existing solutions (SIEM, CDN, WAF, etc.)



Reduced competitive data mining and fraud

Drastically reduced competitive data mining, increased SEO rankings, and protected our marketplace ecosystem

Distil is a key piece of our fraud detection and prevention suite of tools



StubHub Results with Distil Networks



Improved traffic quality and enriched analytic data

Cut pageviews in half, without impacting human users or ad deliveries

Quality of traffic has greatly improved by stopping unwanted bots and limiting site access for trusted bots



Negative Security Model - Blocking Bad Bots



stubhub.com ▾ Reports Settings

Content Protection Save Cancel

Manage your threat responses

block ▾ requests from **known violators**

Known Signatures protects you from Distil's curated list of known content thieves and malicious bots. This is a global IP and bot signature ACL that is maintained based on information gathered from all of our customers.

block ▾ requests from **identities**

Enabling protection based on Identities will protect you from a Distil curated list of known user agents that primarily hog resources and cause unwanted stress on your webserver. In addition, this will also validate that specific request headers and request behaviors fall in line with specifications of the reported user agent.

captcha ▾ requests from **aggregator user agents**

Preventing Aggregator User Agents from accessing your website will use a Distil curated list to filter out content aggregators such as RSS feed reader and Wordpress plugins. If you want your content accessed via RSS/Atom, it is best to leave this setting on "Monitor".

block ▾ requests from **automated browsers**

Automated Browsers are tools that often run a modified or light version of a web browser, capable of accessing your website and viewing the raw HTML, though unable to actually fully render your website or load and process external resources like JavaScript.

Manage referrers & proxies

No ▾ I do not want to block **referrers**

Yes ▾ I want to block **anonymous proxies**

Positive Security Model - Whitelisting Trusted Sources



stubhub.com ▾

Reports

Settings

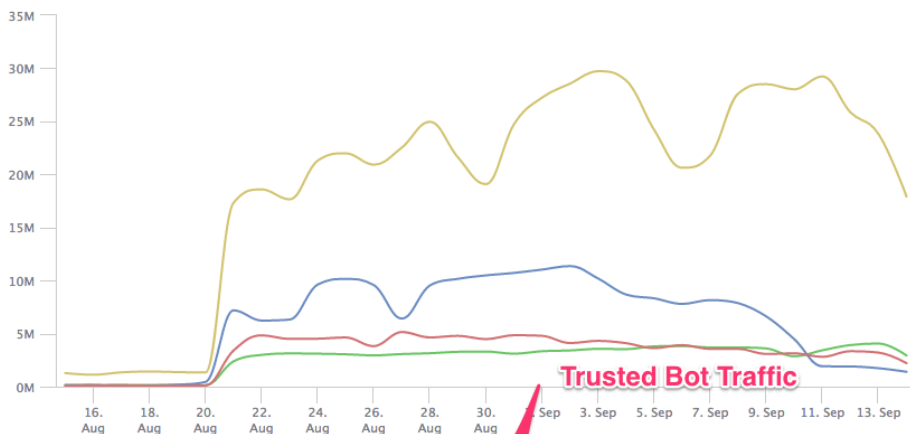
Traffic Overview ▾

Show results from:

Last 30 days ▾

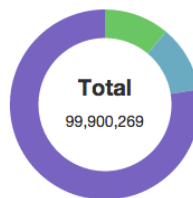
Human, Good Bot, & Bad Bot Page Requests

scale: logarithmic



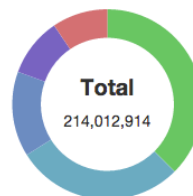
Humans	Good bots	Whitelisted	Bad bots	Total
600,446,706	84,071,726	188,953,939	99,900,269	973,372,640

How we protected you



- blocked 11.1%
- captcha 11.3%
- dropped 0.0%
- monitored 77.6%

What we protected you from



- Session Length Exceeded 37.5%
- Pages Per Session Exceeded 28.6%
- Rate Limited 14.5%
- JavaScript Check Failed 10.0%
- JavaScript Not Loaded 9.4%

The Importance of No False Positives / Negative Impact on Humans

stubbyhub.com ▾

Reports

Settings

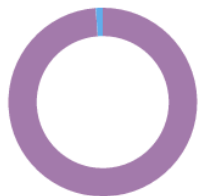
CAPTCHA Requests ▾

Show results from:

Last 30 days ▾

◀ 1 ▶

Total Traffic vs. CAPTCHA Served



Total Traffic
971,534,814 requests

CAPTCHA Served
11,685,879 requests

Attempts Solved vs Failed

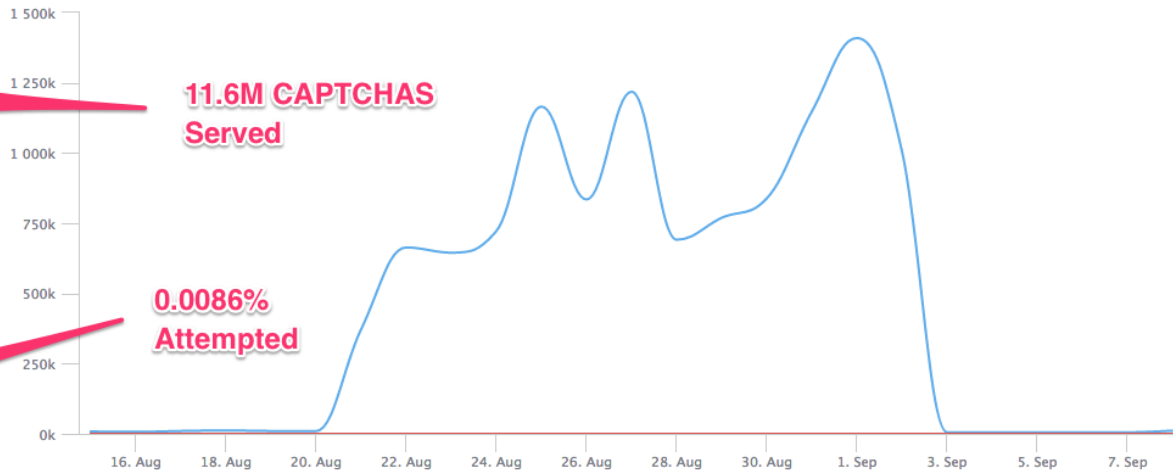


Solved
711 requests

Failed
311 requests

No Attempt
11,684,857 requests

Daily CAPTCHA Requests



Traffic total
763,437,963 requests

Served total
11,579,224 requests

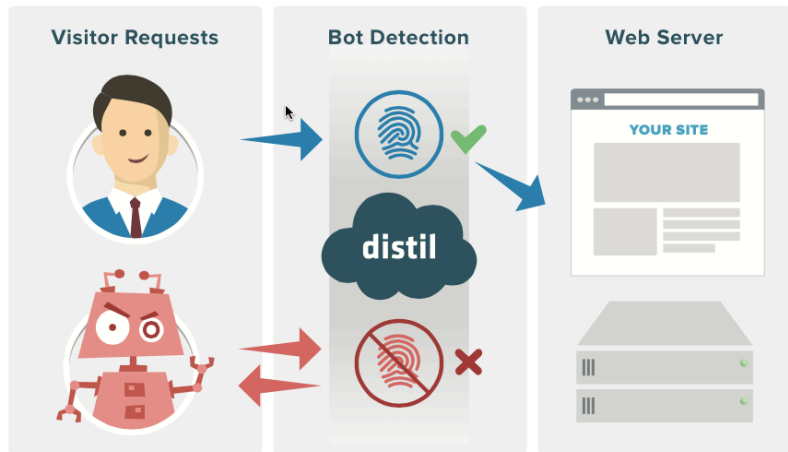
Solved attempt...
709 requests

Failed attempts ...
309 requests



**The First Easy and Accurate Way to Defend
Websites Against Malicious Bots**

The World's Most Accurate Bot Detection System



Inline Fingerprinting

Fingerprints stick to the bot even if it attempts to reconnect from random IP addresses or hide behind an anonymous proxy.

Known Violators Database

Real-time updates from the world's largest Known Violators Database, which is based on the collective intelligence of all Distil-protected sites.

Browser Validation

The first solution to disallow browser spoofing by validating each incoming request as self-reported and detects all known browser automation tools.

Behavioral Modeling and Machine Learning

Machine-learning algorithms pinpoint behavioral anomalies specific to your site's unique traffic patterns.

How Online Marketplaces Benefit from Distil



Increase insight & control
over human, good bot &
bad bot traffic



**Block 99.9% of
malicious bots without
impacting legitimate
users**



Protect data from web
scrapers, unauthorized
aggregators & hackers



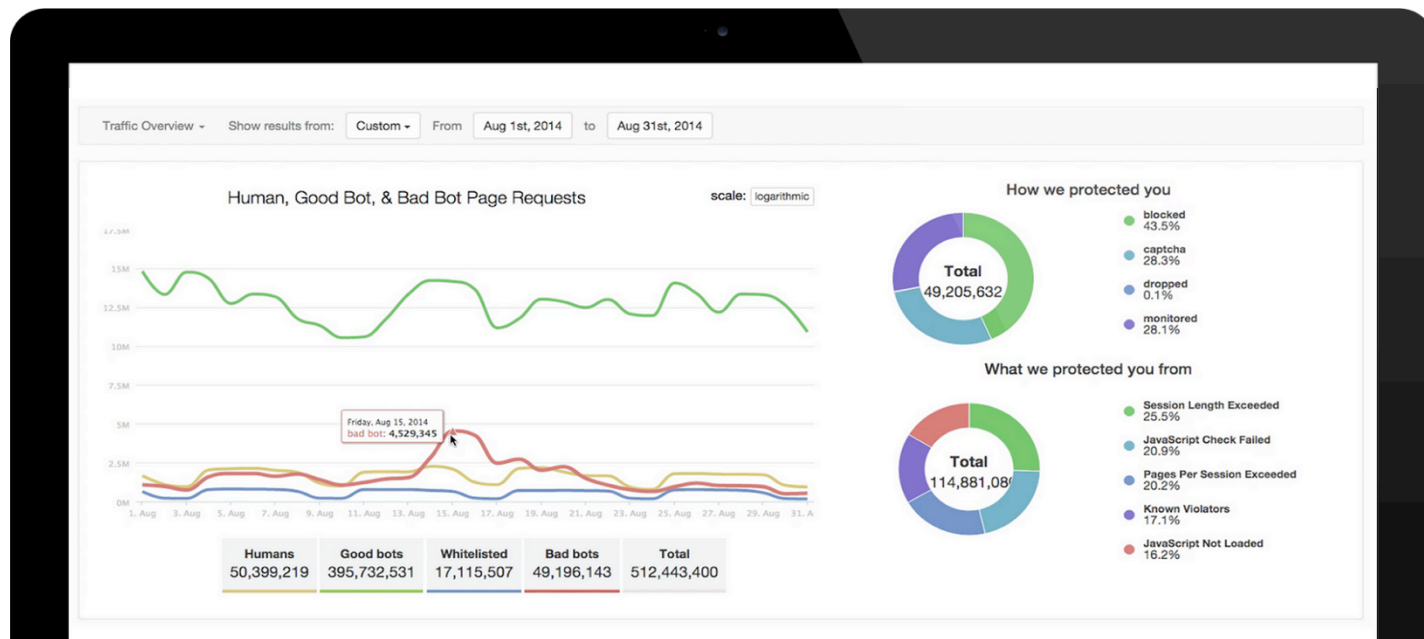
Slash the high tax bots
place on internal teams
& web infrastructure

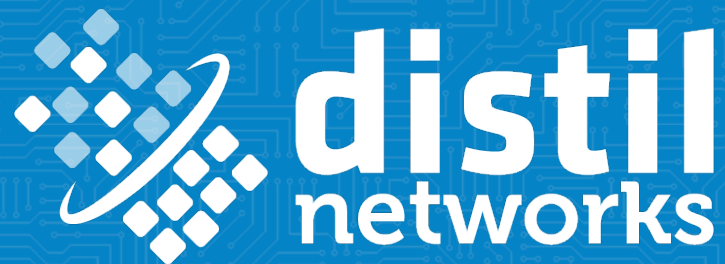
Two Months of Free Service + Traffic Analysis



Offer Ends: November 6th

www.distilnetworks.com/trial/





QUESTIONS....COMMENTS

?

INFO@DISTILNETWORKS.COM

OR CALL US ON

1.866.423.0606

www.distilnetworks.com

White paper w/ source material

[https://docs.google.com/document/d/1OnLQojJI_-
Dwv3yyEKlrB1SZKUeoJp-xw3TE0ULQEvM/edit](https://docs.google.com/document/d/1OnLQojJI_-Dwv3yyEKlrB1SZKUeoJp-xw3TE0ULQEvM/edit)

length: 30 minutes

type: classifieds/marketplaces

Speaker: Rami

Website: <https://www.icmaonline.org/events/conferences/madrid-conference-october-2015>

Title: Changing Threats to Website Security(approved)

Learn How to Defend Your Online Marketplace from Unwanted Traffic

Abstract:

Building an attractive, efficient and popular classifieds and marketplace website is no easy task and the build isn't over when the website goes live. Webmasters need to be aware that threats lurk around every corner, looking to infiltrate a website, alter their stats, change their metrics, and garner important information for another website or company's gain. Bad bots are an evolving problem in addition to other major issues that legitimately ruin your marketplace. In this session, you will learn the most malicious threats that site owners face in ensuring their company and website's security and how to protect yourself from them.

Rami Essaid - CEO & Co-founder of Distil Networks - Thieves and unauthorized aggregators use web scraping bots to steal marketplace listing data, repackage it for sale in the form of lead gen programs, or they place it on unauthorized sites-- all of which hurt SEO, negatively impact your brand, and reduce your ability to monetize your content. Massive bot requests skew your analytics and can even take portions of your site down for extended periods of time. Fake registrations, spammy product listings, and comment spam pollute your site and backend systems. Attend this session to learn key strategies and techniques for defending your site.