Learn How to Defend Your Online Marketplace from Unwanted Traffic

Speakers

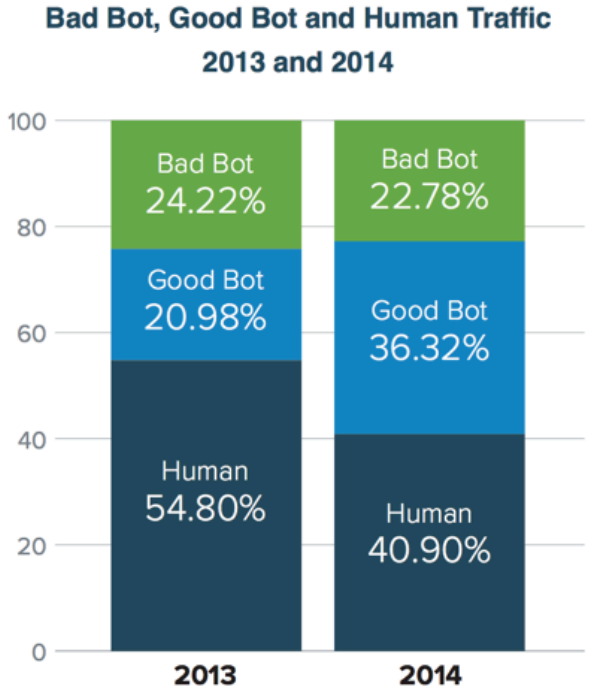
Rami Essaid

CEO & Co-founder

Distil in Classifieds

The Basics of Bots

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



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

distil networks

Bots hurt your KPIs...

Slowdowns, downtime and poor CX

Decline in website traffic

Loss of revenue and customer base

What Is Web Scraping?

distil networks

Web Scraping

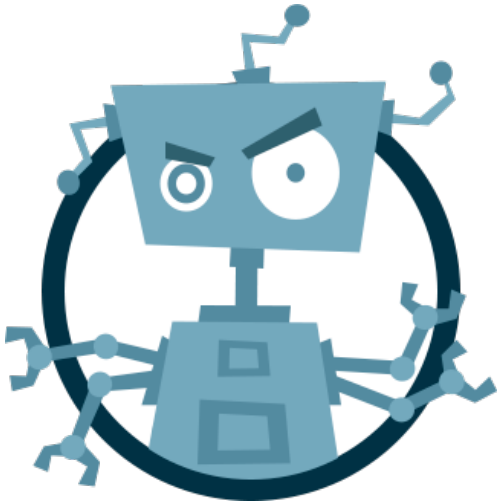
Also known as screen scraping, web scraping is the act of copying large amounts of data from a website – either manually or with an automated program.

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



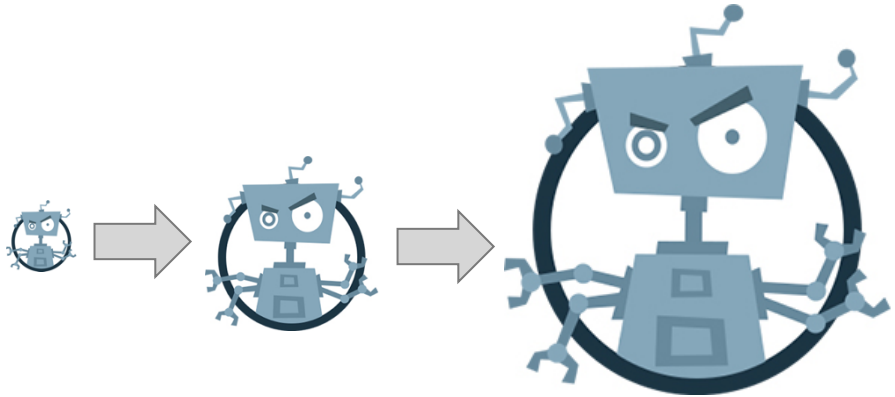
What is Contributing to the Growth in Web Scraping?

distil networks

Cheap scraping software

Inexpensive cloud computing resources

Botnet-as-a-Service



Scraping Online Marketplaces Has Never Been Easier

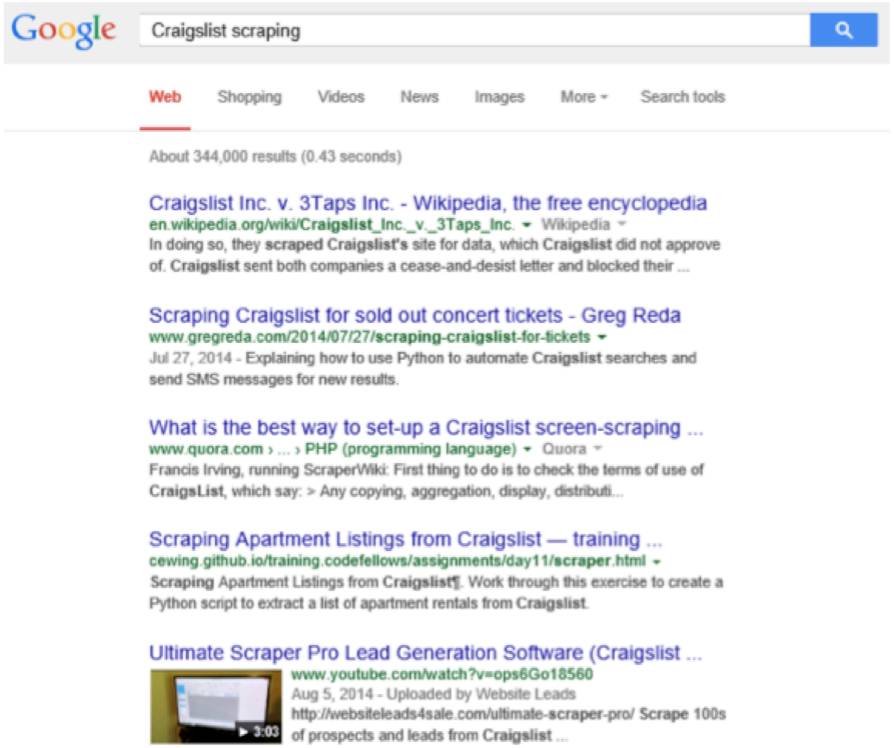
distil networks

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

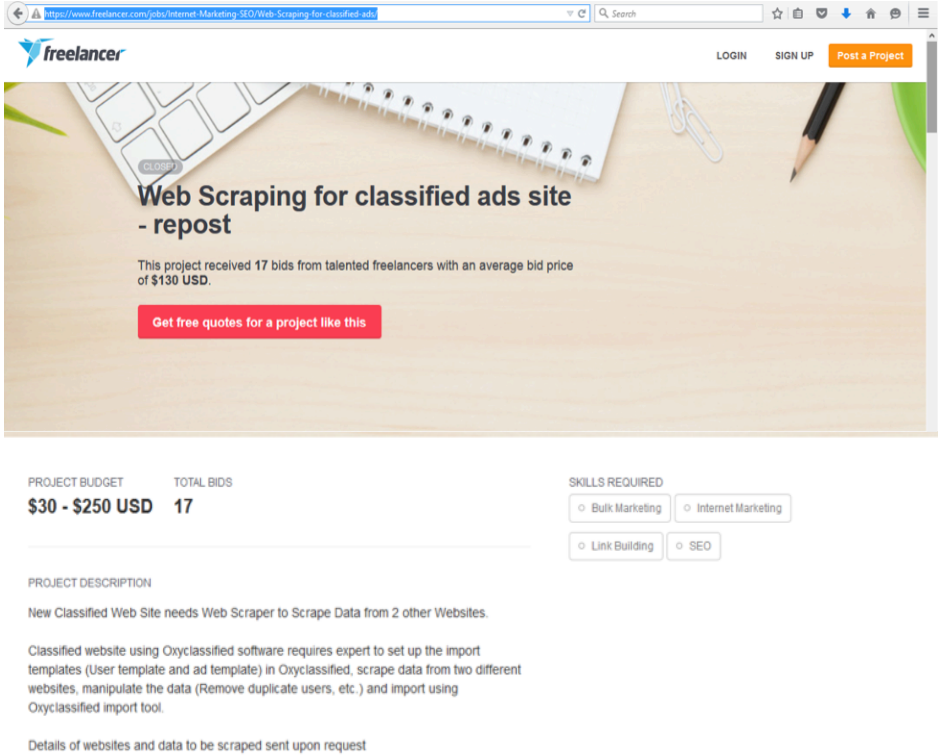
Freelancer.com Rates

Scraping two classified ad websites

Data Manipulation (de-duping, etc.)

Importing into Oxyclassified

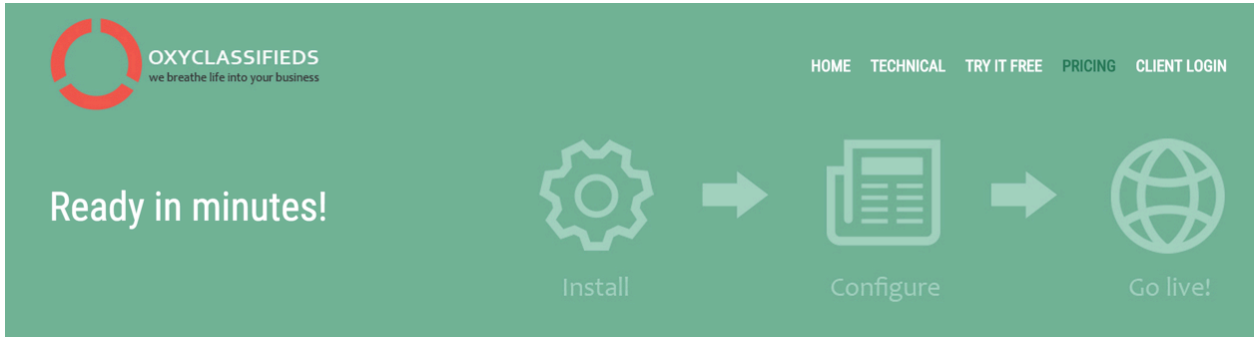
Average Cost - $130 USD



Posting Stolen Data is Quick and Easy due to Turnkey Platforms

distil networks

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



OxyClassifieds Standard Lifetime License €179

Buy OxyClassifieds Standard Version
€179

Click to convert in USD

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

The Cost of Replicating your Website

Scraped Data

$130

Classified Ad Website
$201

$331

Competitive Scraping and Data Mining

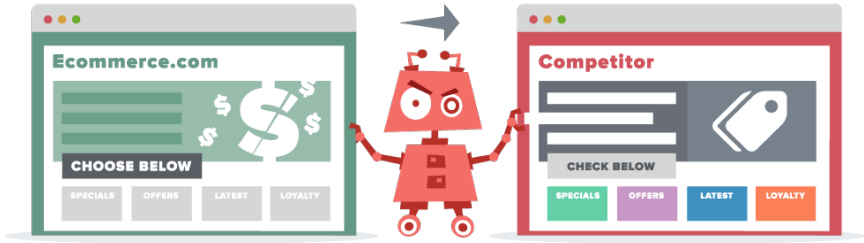
distil networks

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

distil networks

Negative SEO Attacks

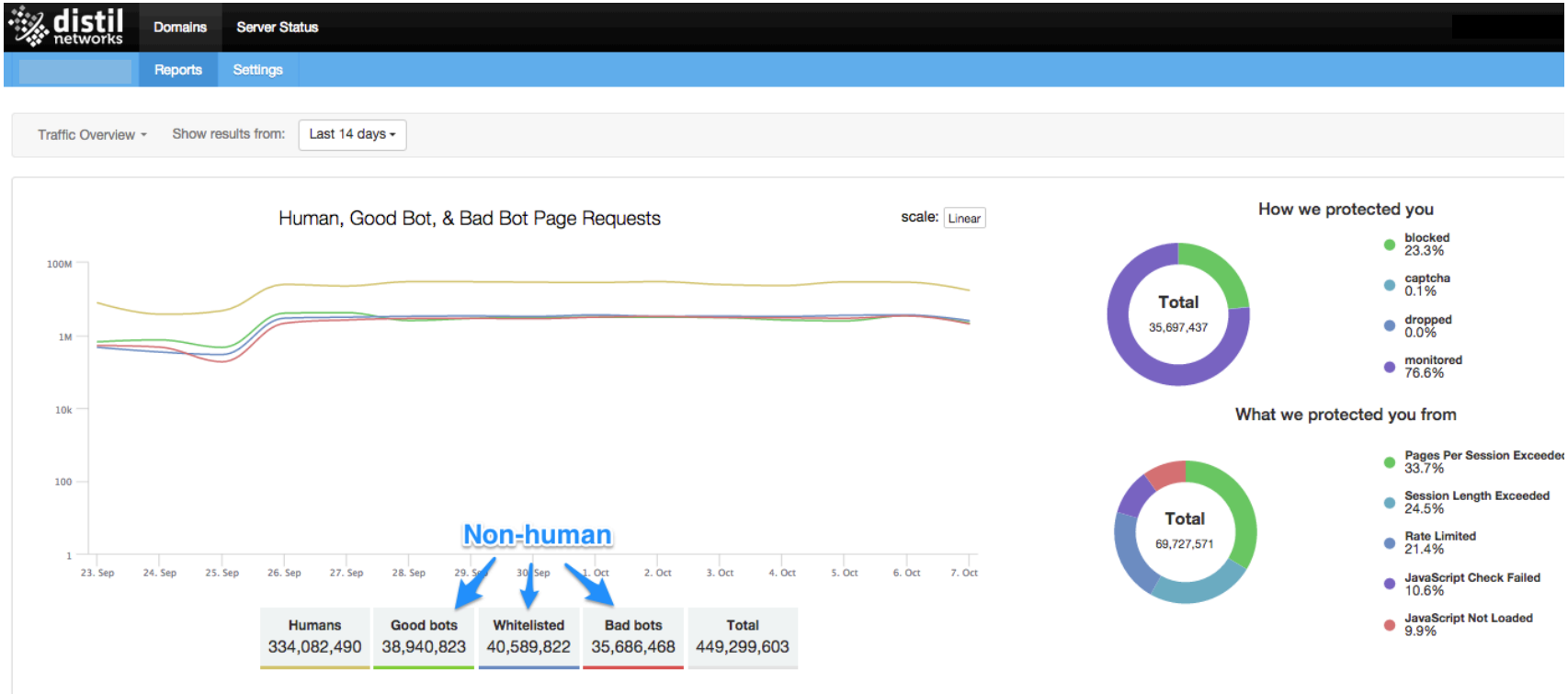
Bots steal content, product lists, and prices for duplication elsewhere on the Internet

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

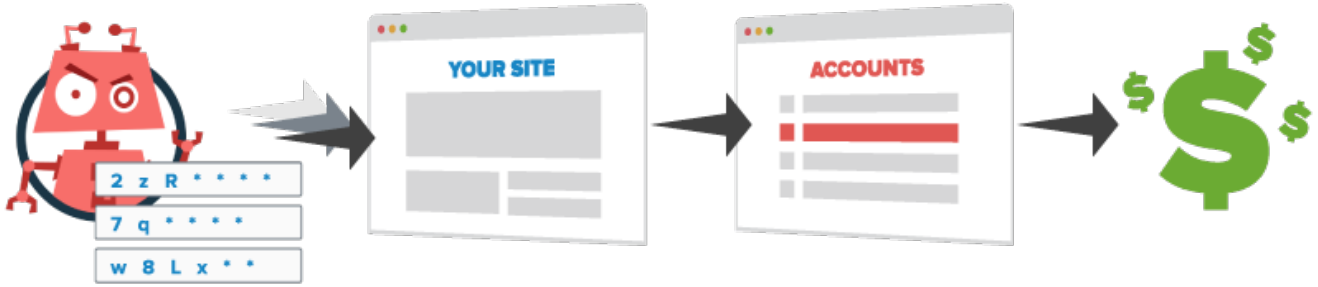


Bots and Security Breaches

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

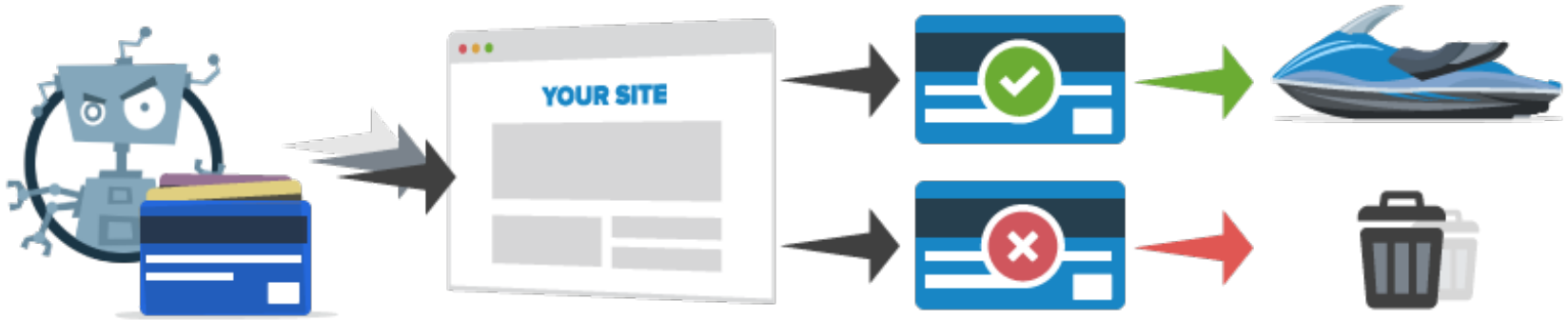


Bots and Transaction Fraud

distil networks

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

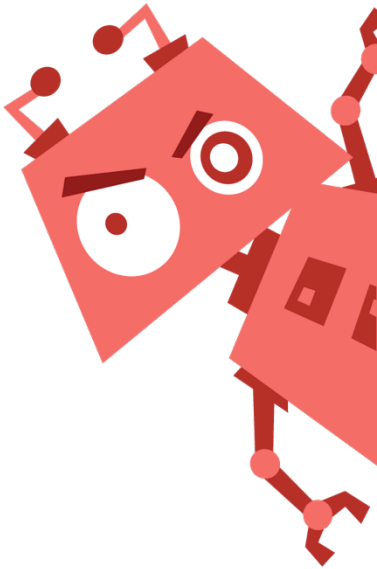
Processes nearly 500 transactions per second

StubHub Bot Challenges

distil networks

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



StubHub Bot Selection Criteria

distil networks

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.)

StubHub Results with Distil Networks

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

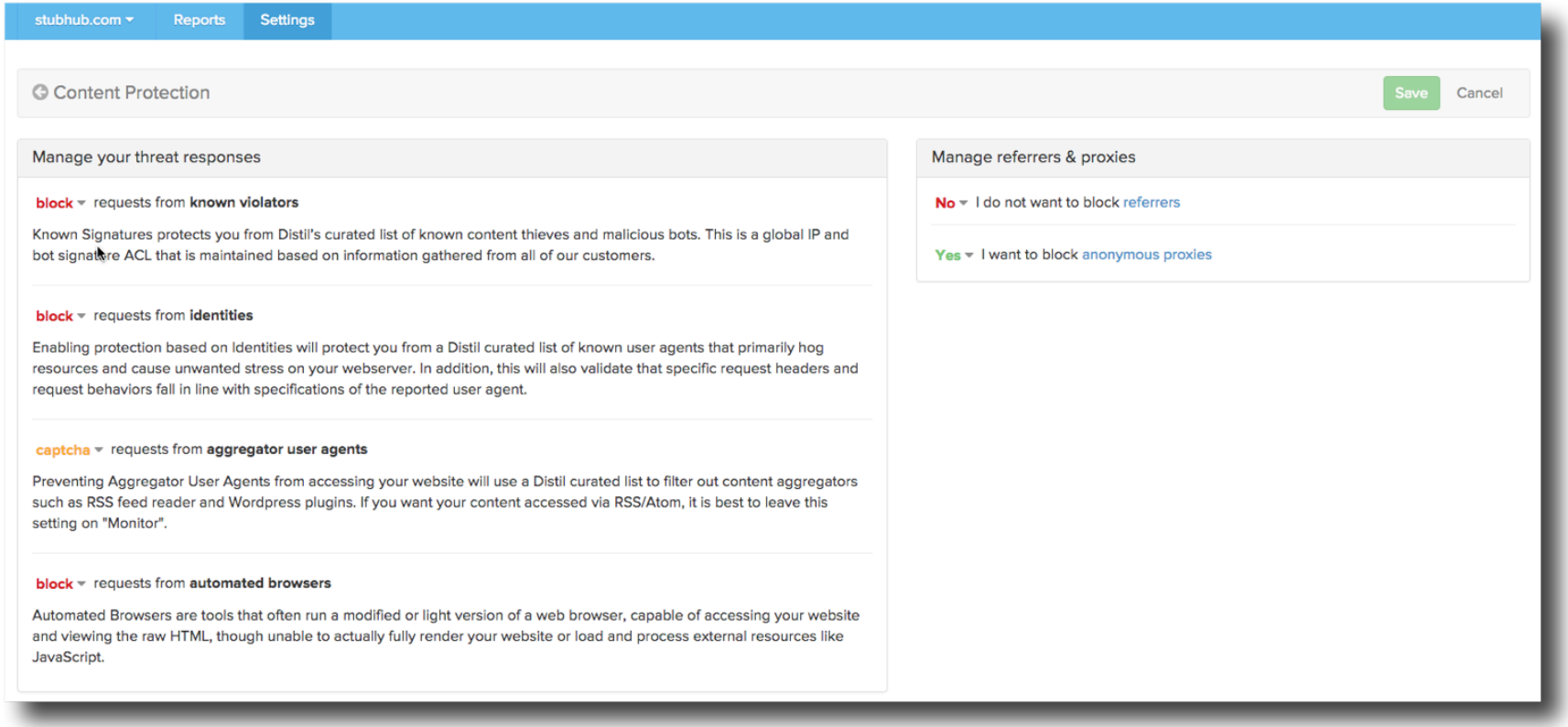
StubHub Results with Distil Networks

Improved traffic quality and enriched analytic data

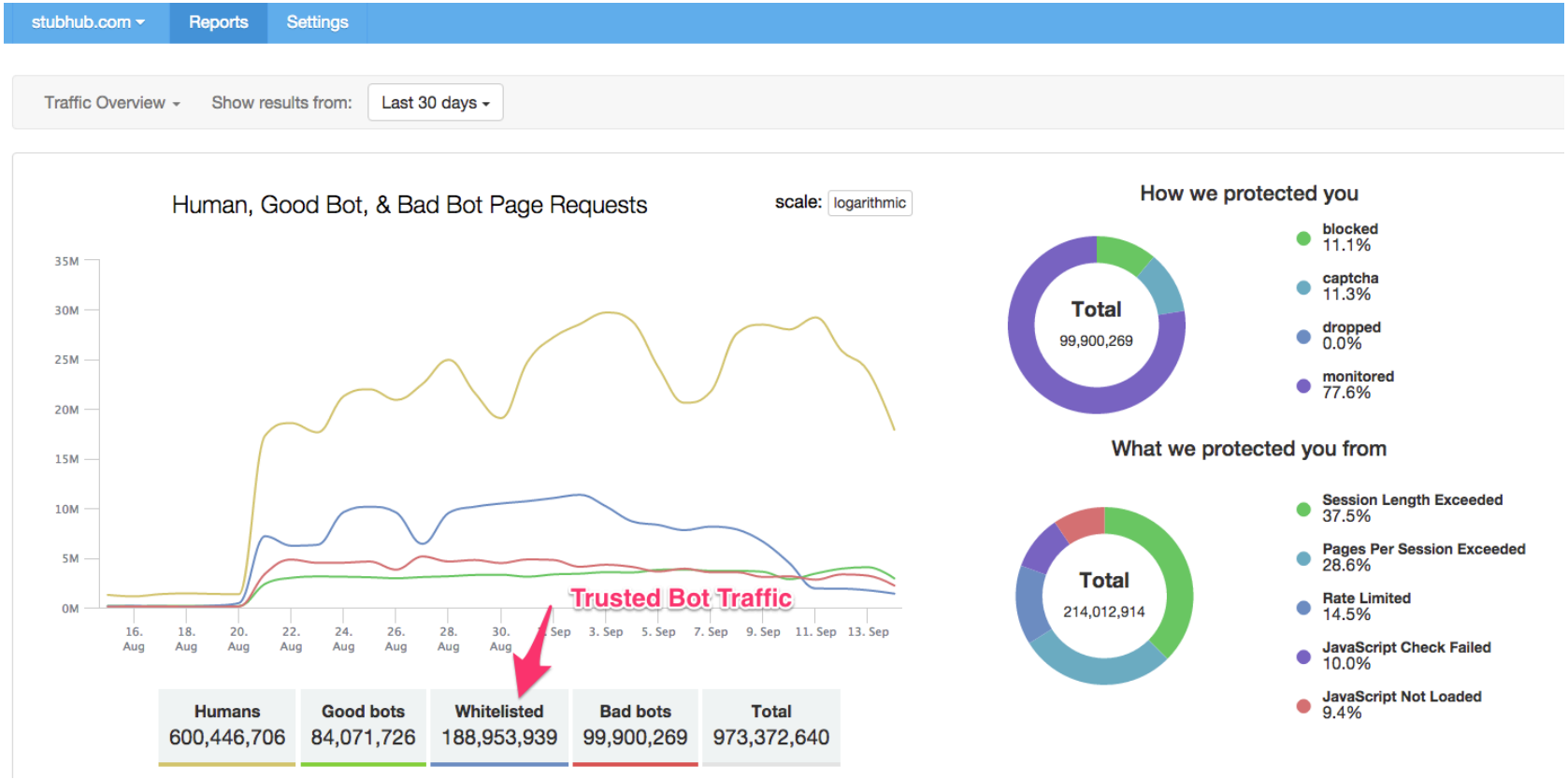
Cut pageviews in half, without impacting human users or ad deliveries

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

Negative Security Model - Blocking Bad Bots



Positive Security Model - Whitelisting Trusted Sources



The Importance of No False Positives / Negative Impact on Humans

distil networks

stubhub.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

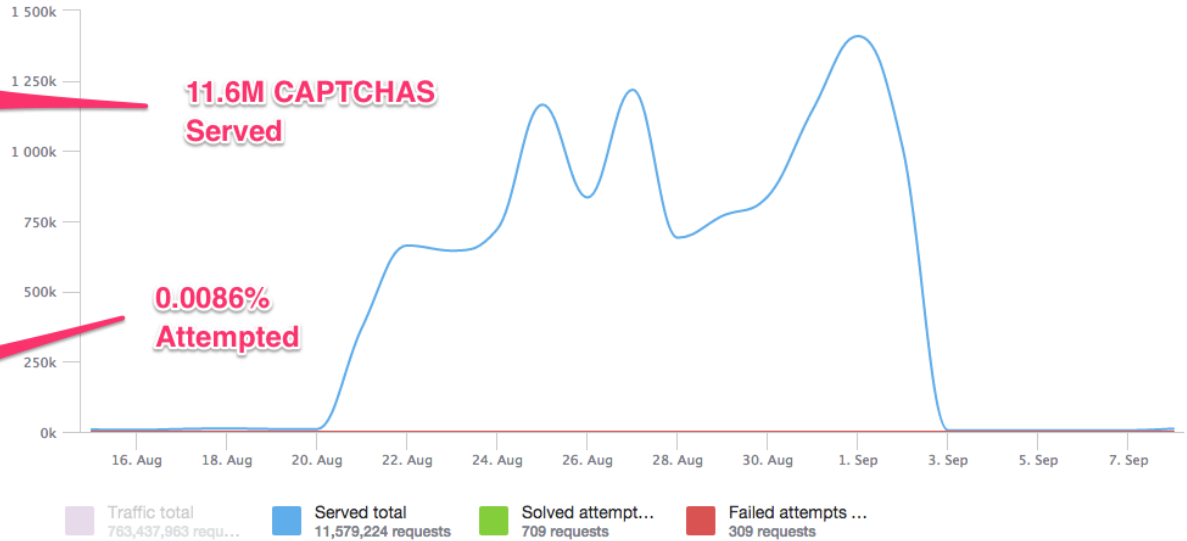
Attempts Solved vs Failed

Solved
711 requests

Failed
311 requests

No Attempt
11,684,857 requests

Daily CAPTCHA Requests

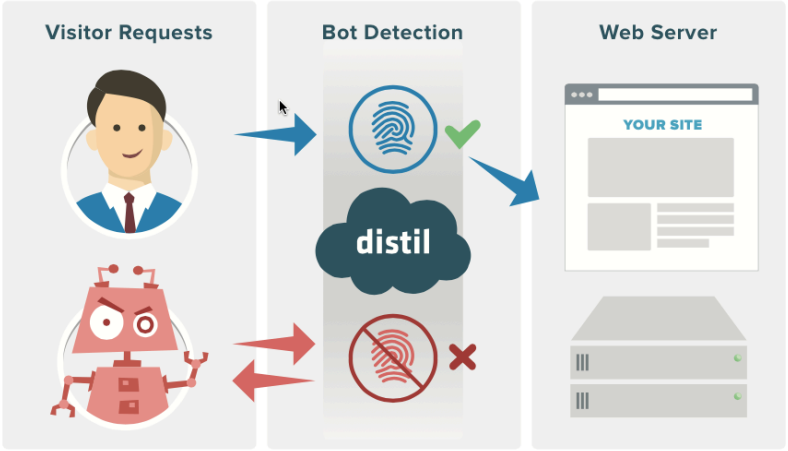


11.6M CAPTCHAS Served

0.0086% Attempted

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

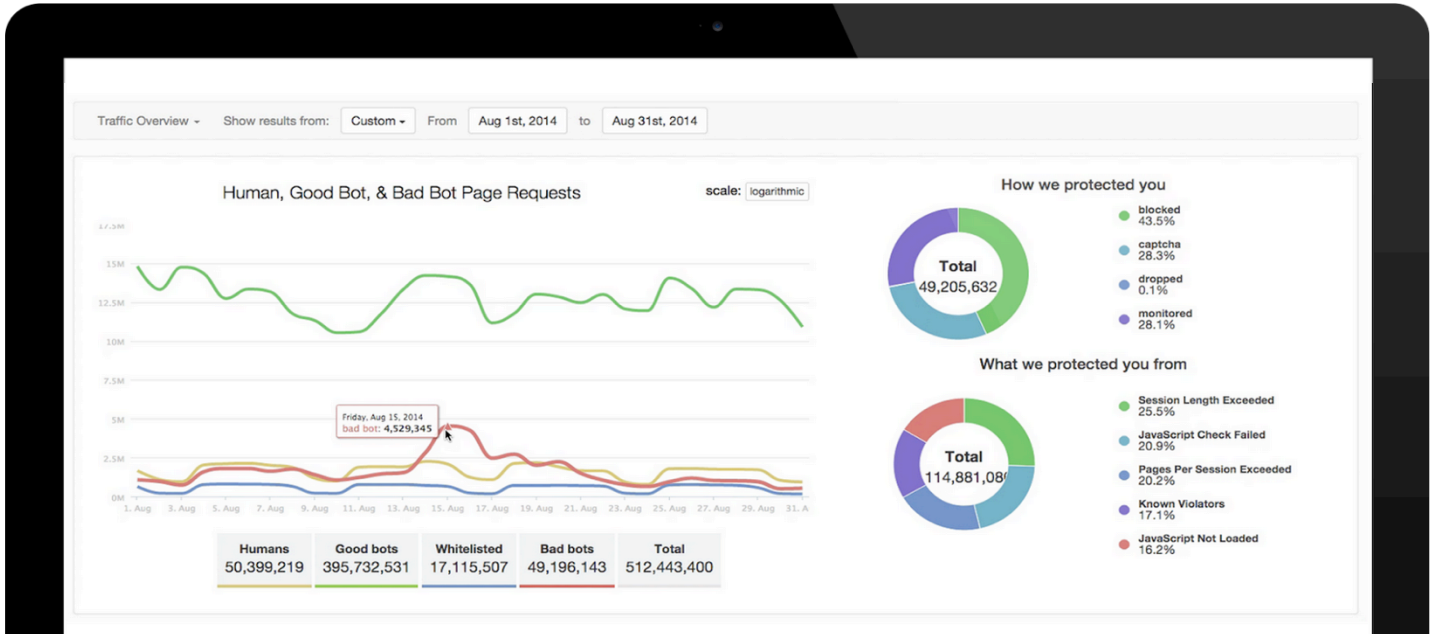
Slash the high tax bots place on internal teams & web infrastructure

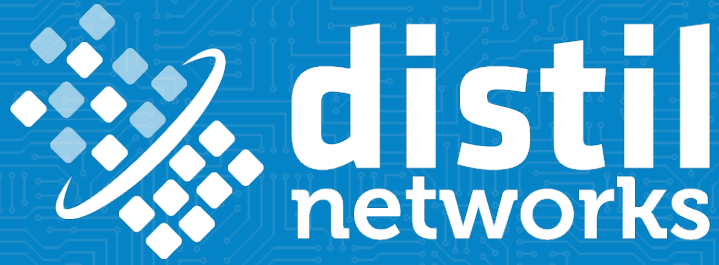
Two Months of Free Service + Traffic Analysis

distil networks

Offer Ends: November 6th

www.distilnetworks.com/trial/





QUESTIONS….COMMENTS ?

INFO@DISTILNETWORKS.COM

OR CALL US ON

1.866.423.0606

White paper w/ source material

https://docs.google.com/document/d/1OnLQojJI_-Dwv3yyEKlrB1SZKUeoJp-xw3TE0ULQEvM/edit

Abstract

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.