

The 2015 Bad Bot Landscape Report



Table of Contents

| | |
|--|----|
| Introduction and Methodology | 3 |
| Key Findings | 3 |
| Bad bots as a percentage of overall traffic | 3 |
| Bad Bot, Good Bot and Human Traffic 2013 and 2014 | 3 |
| Bad Bot, Good Bot and Human Traffic, 2014 | 4 |
| Large sites' staggering bot traffic; smaller sites struggling with bad bots | 4 |
| Traffic by Size of Website (Alexa Score), 2014 | 4 |
| Digital publishing, travel, directory and ecommerce sites hit hardest by bad bots .. | 4 |
| Traffic by Type of Website, 2014 | 4 |
| Traffic to Digital Publishing Websites, 2014 | 5 |
| Traffic to Travel Websites, 2014 | 5 |
| Traffic to Directory Websites, 2014 | 5 |
| Traffic to Ecommerce Websites, 2014 | 5 |
| Mobile bots arrive in droves, no longer "emerging" threat | 5 |
| Bad Bot Self-Reported Browsers, 2014 | 5 |
| Mobile sites easier to scrape | 5 |
| Amazon has a big bot problem; Verizon cleans up its act | 6 |
| Top 20 Bad Bot Originators, 2013 and 2014 | 6 |
| China leads the world in bad bot mobile traffic, T-Mobile USA more of a concern .. | 7 |
| Percent Mobile Bad Bot Traffic for U.S., China and Rest of Word, 2014 | 7 |
| Top 20 Mobile Carriers with Highest Percent of Bad Bot Traffic, 2014 | 7 |
| Percent Mobile Bad Bot Traffic From Largest Chinese and US Mobile Carriers, 2014 .. | 7 |
| USA worst bad bot originator, but Singapore, Israel, Slovenia and Maldives have highest "Bad Bot GDP" | 8 |
| China and Russia most blocked countries | 8 |
| Top 10 Most Blocked Countries, 2014 | 8 |
| Countries Originating the Most Bad Bots, and Most Bad Bots per Online User, 2014 .. | 8 |
| Bad bots mimic human behavior and are very sophisticated | 9 |
| Ways Distil Catches Bots | 9 |
| Bad Bots Sophistication Levels, 2014 | 9 |
| Bad Bots Mimicking Human Behavior, 2014 | 9 |
| Unmasking Human Imposter Bots | 10 |
| Unmasking Sophisticated Bots, 2014 | 10 |
| A More Widely Dispersed Bad Bot Landscape | 10 |
| Percent of Bad Bot Traffic by Hour Hitting the U.S. East Coast, 2013 and 2014 .. | 10 |
| Countries Originating at least 1% of Worldwide Bot Traffic | 10 |
| Conclusion | 11 |
| About Distil Networks | 11 |

Introduction and Methodology

The 2015 Bad Bot Landscape Report is the culmination of months of analysis by the Distil Networks' Data Science Team. The dataset covers the 23 billion bad bot threats we saw in 2014 as well as good bot and human traffic. The dataset resides in Distil's Hadoop Cluster and includes data from 100s of customers as well as Distil's global network of 17 data centers.

Similar to last year's study, the goal of this report is to help identify statistically significant data and insights into the bad bot landscape. Bad bots continue to place a huge tax on IT security and web infrastructure teams across the globe. The variety, volume and sophistication of today's bots wreak havoc across online operations big and small. They're the key culprits behind web scraping, brute force attacks, competitive data mining, brownouts, account hijacking, unauthorized vulnerability scans, spam, man-in-the-middle attacks, and click fraud. By shining a light on their origins, behavior, and obfuscation methods, we hope to fulfill our mission of making the web more secure.

This study does not account for large Distributed Denial of Service (DDoS) attacks. Distil Networks is not a DDoS mitigation service. Rather, our focus is on application-layer attacks. It's also important to note that Distil's customers tend to care more about stopping bots, and thus could be more heavily targeted by bots than an average website.

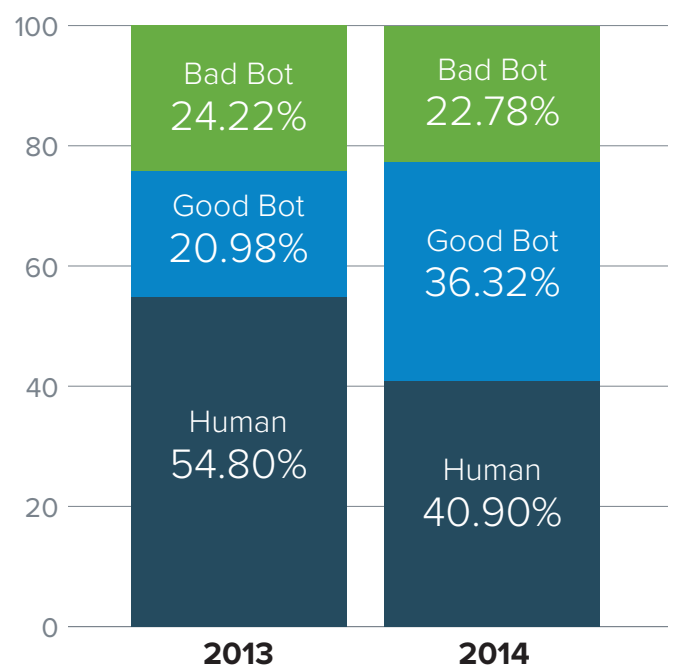
If you have any questions regarding the report or its findings, please contact us today at badbots@distilnetworks.com or on Twitter at @DISTIL.

Key Findings

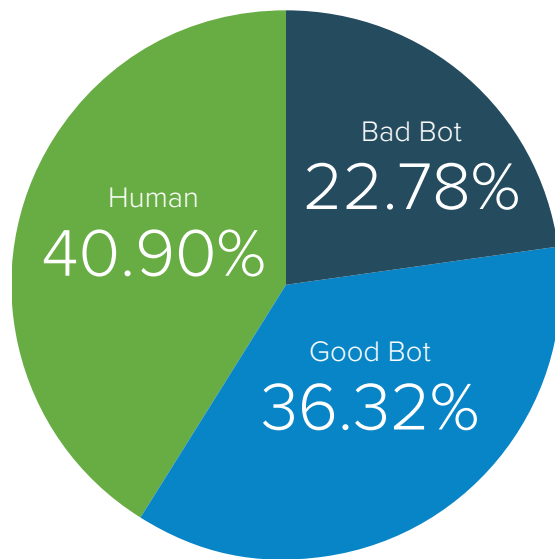
Bad bots as a percentage of overall traffic

The amount of bad bot traffic we saw across our network as a percentage of overall traffic dropped slightly from 24.22% in 2013 to 22.78% in 2014. The bigger move was in Good Bot traffic growing from 20.98% to 36.32%, which might be due in part to more aggressive indexing by Bing and upstart search engines 2014. These results exclude traffic from our single largest customer of 2014 as the weight of that one customer would have distorted the results.

**Bad Bot, Good Bot and Human Traffic
2013 and 2014**



Bad Bot, Good Bot and Human Traffic,



Large sites experience staggering bot traffic; smaller sites struggle with bad bots

Large sites face a staggering amount of good bot traffic. However, good bot traffic can become undesirable and many large sites are governing overly aggressive indexing (e.g., Bingbot). Some bots may be good for some sites, but not for others as we learned from [Twitter's recent earnings fiasco](#).

Traffic by Size of Website (Alexa Score), 2014

| Site Size (Alexa) | Bad Bot % | Good Bot % | Human % |
|--------------------------|-----------|------------|---------|
| Large (1 - 10,000) | 23.07% | 43.65% | 33.28% |
| Medium (10,001 - 50,000) | 17.27% | 34.83% | 47.90% |
| Small (50,001 - 150,000) | 32.04% | 27.02% | 40.94% |

Smaller sites faced a larger percentage of bad bot traffic than their larger brethren. The danger here for smaller sites is that they may not have the infrastructure to support random spikes in bad bot traffic.

Digital publishing, travel, directory and ecommerce sites hit hardest by bad bots

As a percentage of overall traffic, digital publishers, travel, directory and ecommerce sites were hit hardest by bad bots. Travel sites had significantly less good bot traffic than other types of sites at just over 3%. This might be due to the fact that travel site pages are often dynamically generated via third party API calls to booking engines, airlines and other suppliers (versus more SEO-friendly flat HTML files).

Traffic by Type of Website, 2014

| Type of Site | Bad Bot % | Good Bot % | Human % |
|---------------------------|-----------|------------|---------|
| Digital Publishing | 32.16% | 17.72% | 50.12% |
| Travel | 27.78% | 3.06% | 69.16% |
| Directory | 20.91% | 48.42% | 30.67% |
| Ecommerce | 20.15% | 14.44% | 65.41% |
| Real Estate | 11.27% | 48.42% | 40.31% |
| Marketplace & Classifieds | 11.25% | 29.74% | 59.01% |

Traffic to Digital Publishing Websites, 2014

| Digital Publishing Site | Bad Bot % | Good Bot % | Human % |
|-------------------------|-----------|------------|---------|
| Large | 20.19% | 26.90% | 52.91% |
| Medium | 36.95% | 15.23% | 47.82% |
| Small | 30.59% | 12.64% | 56.77% |

Traffic to Travel Websites, 2014

| Travel Site | Bad Bot % | Good Bot % | Human % |
|-------------|-----------|------------|---------|
| Large | 16.65% | 1.65% | 81.70% |
| Medium | 5.42% | 3.19% | 91.39% |
| Small | 56.76% | 4.33% | 38.91% |

Traffic to Directory Websites, 2014

| Directory Site | Bad Bot % | Good Bot % | Human % |
|----------------|-----------|------------|---------|
| Large | 30.87% | 35.85% | 33.28% |
| Medium | 14.27% | 51.23% | 34.49% |
| Small | 12.34% | 62.94% | 24.72% |

Traffic to Ecommerce Websites, 2014

| Ecommerce Site | Bad Bot % | Good Bot % | Human % |
|----------------|-----------|------------|---------|
| Large | 22.13% | 8.81% | 69.06% |
| Medium | 9.68% | 10.16% | 80.16% |
| Small | 10.37% | 8.87% | 80.76% |

Mobile bots arrive in droves, no longer “emerging” threat

2014 is the first year that bots masking themselves as mobile web users arrived in droves. It's also the first time that a mobile carrier (T-Mobile USA) has appeared on the list of top 20 ISPs serving bad bot traffic.

Mobile sites easier to scrape

The same characteristics that make a mobile optimized site easy to quickly navigate for humans also makes them prime targets for bad bots. Mobile sites tend to be easier to scrape because they provide more structured access to website data.

As shown below, the Android Webkit Browser, at 4.87%, entered into the Top 5 list of user agents leveraged by bad bots to hide their identities (a user agent is an application that remotely connects a user — either a bot or a real human — to a server through a network).

Bad Bot Self-Reported Browsers, 2014

| Bad Bot Self-Reported Browser | Percent of Total |
|-------------------------------|------------------|
| Firefox | 26.62% |
| Chrome | 26.01% |
| Internet Explorer | 23.28% |
| Apache HTTP Client | 12.67% |
| Android Webkit Browser | 4.87% |

Amazon has a big bot problem; Verizon cleans up its act

Bad bots made up 78% of Amazon's 2014 traffic, not a huge difference from 2013. Verizon Business really cleaned up its act, cutting its bad bot traffic by 54% in 2014. Amazon EC2 represents one of many players in the "cheap hosting" category — the perfect place from which to launch bad bots.

Many web hosts have little monitoring and few safeguards in place to prevent bad bot origination.

Consumer ISP heavyweights Comcast and Time Warner Cable saw a 300% increase in bad bot activity in 2014, perhaps because of the increased prevalence of botnets infecting home computers.

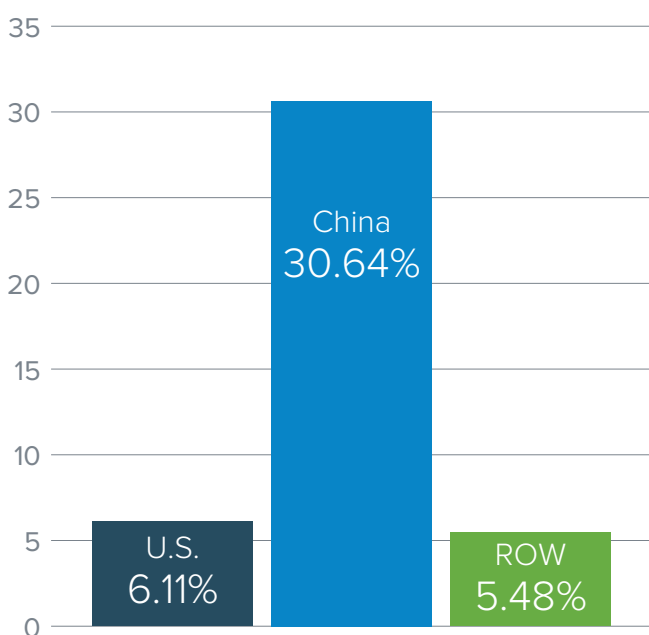
Top 20 Bad Bot Originators, 2013 and 2014

| 2013 | | | 2014 | | |
|-----------------------------|--|--|--------------------------------|--|--|
| Service Provider Name | Bad Bot Traffic as % of this SPs Traffic | Bad Bot Traffic as % of ALL SP Bad Bot Traffic | Service Provider Name | Bad Bot Traffic as % of this SPs Traffic | Bad Bot Traffic as % of all SP Bad Bot Traffic |
| Verizon Business | 68.24% | 10.99% | Amazon Technologies | 77.99% | 15.07% |
| Level 3 | 76.97% | 10.22% | Amazon.com | 69.54% | 3.30% |
| Amazon Technologies | 79.18% | 9.47% | Verizon Business | 36.75% | 2.83% |
| Las Vegas NV Datacenter | 96.67% | 5.58% | Cogent | 18.79% | 2.71% |
| Amazon.com | 74.42% | 4.81% | Comcast Cable | 6.40% | 2.59% |
| Hosting Solutions Int'l | 83.35% | 2.52% | CIK | 84.78% | 2.34% |
| Comcast Cable | 1.76% | 2.20% | Bezeq | 58.99% | 1.76% |
| Frontier Comm. | 24.74% | 2.18% | OVH SAS | 84.08% | 1.54% |
| CIK Telecom | 84.97% | 1.54% | Time Warner Cable | 5.77% | 1.22% |
| Time Warner Cable | 1.81% | 1.39% | Gig Avenue | 98.70% | 1.20% |
| VNET s.r.o. | 92.10% | 1.36% | OVH Hosting | 86.79% | 1.19% |
| SoftLayer Dutch Holdings BV | 97.61% | 1.35% | Internap Network Services Corp | 74.51% | 1.18% |
| ThePlanet.com | 96.87% | 1.12% | SoftLayer Dutch Holdings BV | 95.77% | 1.09% |
| Telekom Malaysia | 14.95% | 1.12% | Server Block | 86.10% | 1.07% |
| Savvis | 26.83% | 1.04% | Hetzner Online AG | 81.29% | 1.03% |
| Verizon FIOS | 2.09% | 1.04% | ColoCrossing | 77.80% | 1.02% |
| NOC4Hosts | 92.44% | 0.95% | Telecom Italia | 18.27% | 1.00% |
| Switch Comms. | 96.41% | 0.84% | Tilaa BV | 99.99% | 1.00% |
| Gig Avenue | 76.86% | 0.75% | T-Mobile USA | 19.17% | 1.00% |
| SoftLayer Technologies | 77.44% | 0.74% | Google | 29.48% | 0.94% |

China leads the world in bad bot mobile traffic, T-Mobile USA more of a concern

Overall, mobile bot traffic from U.S. mobile carriers (as a percentage of their overall traffic) was roughly on par with the rest of the world during 2014.

Percent Mobile Bad Bot Traffic for U.S., China and Rest of World, 2014



Top 20 Mobile Carriers with Highest Percent of Bad Bot Traffic, 2014

| Mobile Carrier | Country of Origin | Bad Bot % of This Carrier's Total Traffic |
|---------------------------|-------------------|---|
| China Mobile | China | 36.80% |
| China Telecom | China | 19.18% |
| China Unicom | China | 16.87% |
| VimpelCom | Norway/Russia | 15.35% |
| MTN | South Africa | 13.01% |
| Axiata | Malaysia | 9.18% |
| Airtel | India | 8.90% |
| America Movil | Mexico | 5.38% |
| TeliaSonera | Finland/Sweden | 4.55% |
| Etisalat | UAE | 4.55% |
| Vodafone | United Kingdom | 4.29% |
| Telenor | Norway | 3.83% |
| Verizon Wireless | United States | 3.72% |
| Orange | France | 3.58% |
| SingTel | Singapore | 3.23% |
| Idea Cellular | India | 3.19% |
| T-Mobile Deutschland GmbH | Germany | 2.32% |
| Telefonica | Spain | 2.31% |
| AT&T Mobility | United States | 2.20% |
| STC | Saudi Arabia | 1.20% |

Percent Mobile Bad Bot Traffic From Largest Chinese and US Mobile Carriers, 2014

| Top 3 China-Based Mobile Carriers | Bad Bot % of Carrier's Total Traffic | Top 3 US-Based Mobile Carriers | Bad Bot % of Carrier's Total Traffic |
|-----------------------------------|--------------------------------------|--------------------------------|--------------------------------------|
| China Mobile | 36.80% | T-Mobile USA | 19.17% |
| China Telecom | 19.18% | Verizon Wireless | 3.72% |
| China Unicom | 16.87% | AT&T Mobility | 2.20% |

USA worst bad bot originator, but Singapore, Israel, Slovenia and Maldives have highest “Bad Bot GDP”

The United States, with thousands of cheap hosts, dominates the rankings in bad bot origination. Taken in isolation, absolute bad bot volume data can be somewhat misleading. Measuring bad bots per online user yields a country’s “Bad Bot GDP.”

China and Russia most blocked countries

For 2014, China and Russia were the most blocked countries. Geo-IP Fencing is an effective website security tactic for those

organizations with well-defined geographical markets.

Top 10 Most Blocked Countries, 2014



Countries Originating the Most Bad Bots, and Most Bad Bots per Online User, 2014

| Countries Originating the Most Bad Bots | | Countries with the Most Bad Bots per Online User | |
|---|-------------------------------|--|---|
| Country | % of Total Volume of Bad Bots | Country | Ratio of Bad Bots to Number of Online Users |
| United States | 52.19% | Singapore | 152.87 |
| Germany | 4.58% | Israel | 34.12 |
| Canada | 4.27% | Slovenia | 29.69 |
| Italy | 3.75% | Maldives | 15.54 |
| France | 3.22% | Ireland | 8.83 |
| Netherlands | 2.52% | United States | 6.24 |
| China | 2.46% | Malta | 4.71 |
| Russia | 2.39% | Netherlands | 3.32 |
| United Kingdom | 2.27% | Romania | 3.31 |
| India | 2.23% | Denmark | 3.00 |

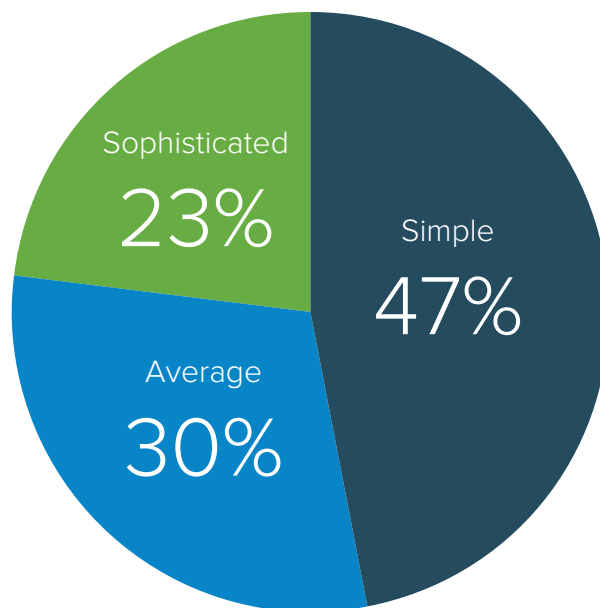
Bad bots mimic human behavior and are very sophisticated

Some bad bots make little or no attempt to obfuscate their identities. “Simple” bad bots show their hand in a number of ways such as leveraging bad user agents or failing basic browser integrity checks. “Average” bad bots can be trapped by forcing them to prove they are using a real web browser, while “Sophisticated” bad bots closely mimic human behavior. For example, browser automation tools like Selenium and PhantomJS can be used to replicate human web browsing, and can be almost indistinguishable from actual human website visitors.

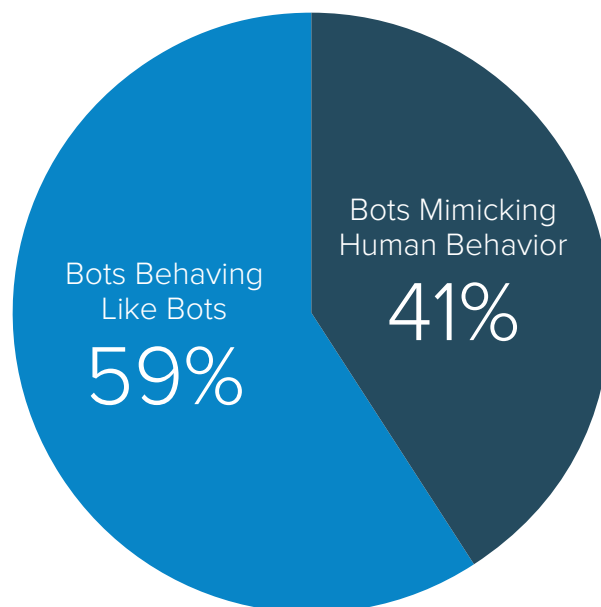
In 2014, a whopping 41% of Bad Bots mimicked human behavior, while only 7% of bad bots disguised themselves as good bots (e.g., Googlebot and Bingbot). Webmasters allow entry of the Googlebot to their website infrastructure for SEO purposes. When a bad bot masked as the Googlebot enters a site, it

can cause a wide range of problems without raising any alarms.

Bad Bots Sophistication Levels, 2014



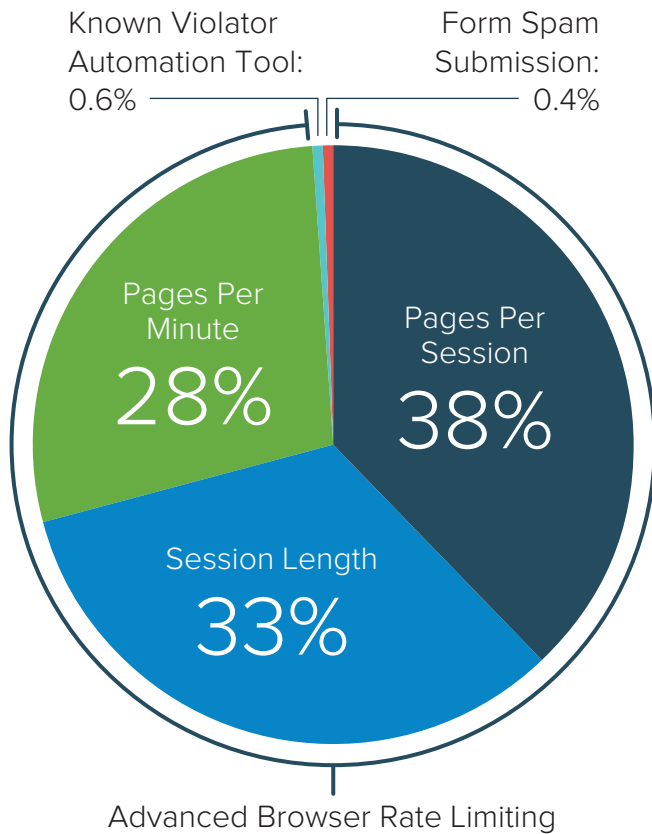
Bad Bots Mimicking Human Behavior, 2014



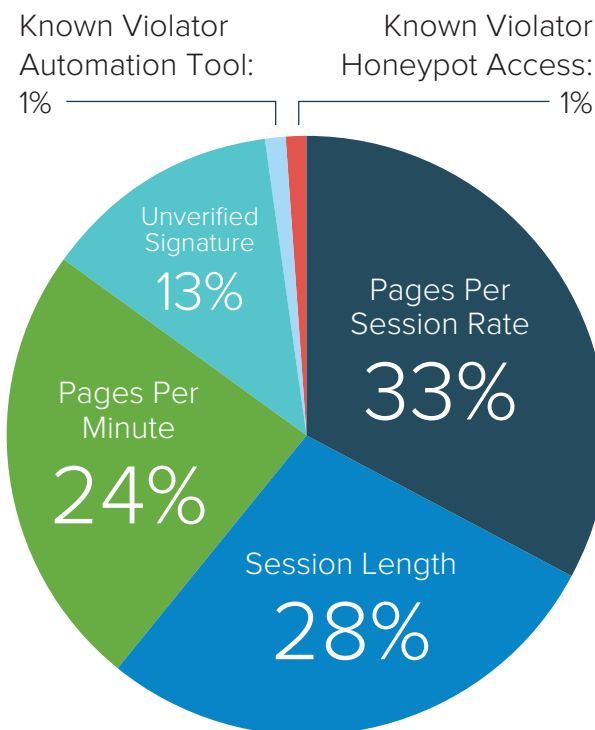
A Few Ways Distil Catches Bots

| Simple Bots | Average Bots | Sophisticated Bots |
|---------------------------|---------------------------------|--------------------------------|
| Known Violator User Agent | Failure to Load External Assets | Cookie Tampering |
| Bad User Agents | Incomplete JavaScript Engine | Known Violator Honeypot Access |
| Browser Integrity Check | | Known Violator Automation Tool |
| Aggregator User Agents | | Pages Per Minute |
| | | Pages Per Session |
| | | Session Length |

Unmasking Human Imposter Bots



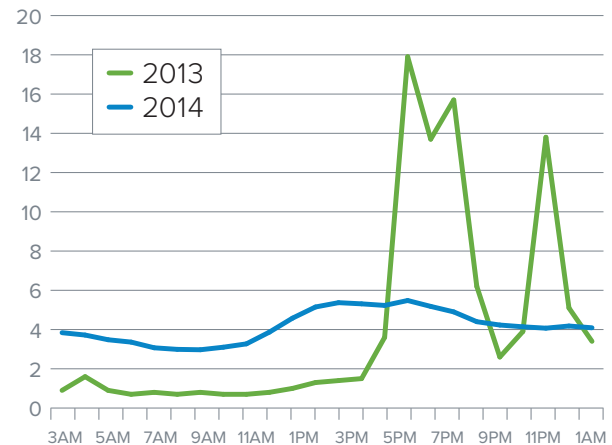
Unmasking Sophisticated Bots, 2014



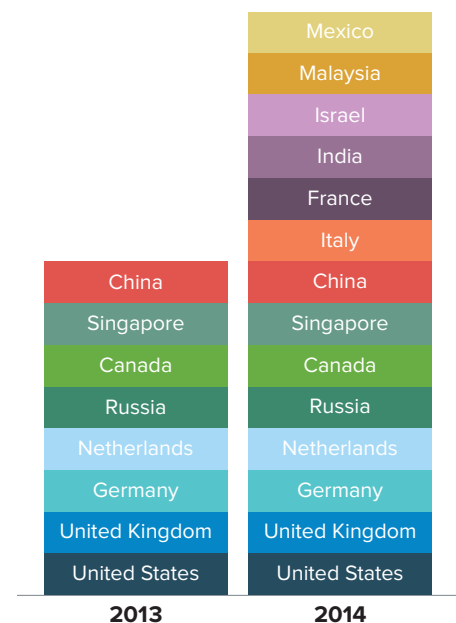
A More Widely Dispersed Bad Bot Landscape

Bad bot threats have taken on less predictable patterns, with bad bots attacking more evenly throughout the day and from a more broadly dispersed set of global points of origin. 14 countries, almost double the number in 2013, originated at least 1% of bad bot traffic volume in 2014.

Percent of Bad Bot Traffic by Hour Hitting



Countries Originating at least 1% of Worldwide Bot Traffic



Conclusion

The bad bot landscape continues to evolve rapidly, particularly with the dramatic increase in mobile bot traffic, and an ever wider range of geographic and ISP points of origin. With the advent of cheap or free cloud computing resources, anyone with basic computer skills can download open source software and get into the game.

Meanwhile, IT infrastructure teams are under increasing pressure to accurately forecast and provision web infrastructure to meet the speed and availability demands of legitimate users. IT security teams must ensure that nefarious actors can't harvest their data or breach their defenses. And marketing teams seek accurate data on website and conversion metrics.

Yet most companies still have no visibility or control over malicious website traffic.

About Distil Networks

Distil Networks, the global leader in bot detection and mitigation, is the first easy and accurate way to identify and police malicious website traffic, blocking 99.9% of bad bots without impacting legitimate users.

Distil protects against web scraping, brute force attacks, competitive data mining, account hijacking, unauthorized vulnerability scans, spam, man-in-the-middle attacks and click fraud.

Slash the high tax that bots place on your internal teams and web infrastructure by outsourcing the problem to the team with a singular focus on blocking malicious bots:

- Harden your website security by eliminating malicious bots
- Increase insight and control over human, good bot and bad bot traffic
- Protect data from web scrapers, unauthorized aggregators and competitors
- Deploy on the Distil Cloud CDN or Distil Appliance (Physical | Virtual | AWS)

For more information on Distil Networks, visit us at <http://www.distilnetworks.com> or follow @DISTIL on twitter.



www.distilnetworks.com

sales@distilnetworks.com

phone: 1.866.423.0606