distil networks

BOTREPORT

The 2016 Bad Bot Landscape Report

The Rise of Advanced Persistent Bots

(w) www.distilnetworks.com

(e) sales@distilnetworks.com

(p) 1.866.423.0606

Table of Contents

Introduction and Methodology

Now in its third year, the 2016 Distil Networks Bad Bot Landscape Report is the IT security industry's most in-depth analysis about the sources, types, and sophistication levels of last year's bot attacks. There are serious implications for anyone responsible for securing websites and APIs.

The report is the culmination of several months of analysis by Distil Networks' data science and application teams. Its dataset resides in Distil's Hadoop cluster and includes 74 billion bot requests, anonymized data from several hundred customers, and web traffic from our 17 data centers.

Bad bots are unique from many other security threat types in that their manifestations can be as varied as the businesses they target. Bots enable high-speed abuse, misuse, and attacks on websites and APIs. They enable attackers, unsavory competitors, and fraudsters to perform a wide array of malicious activities. This includes web scraping, competitive data mining, personal and financial data harvesting, brute force login and man-in-the-middle attacks, digital ad fraud, spam, transaction fraud, and more.

Bad bots create vast economic and productivity loss. For example, according to an article published in The Register last October, the Dridex Banking botnet breached tens of thousands of organizations across 27 countries around the globe. It has been responsible for losses of over £20M (roughly $30.5M) in the UK, and at least $10M in the United States.

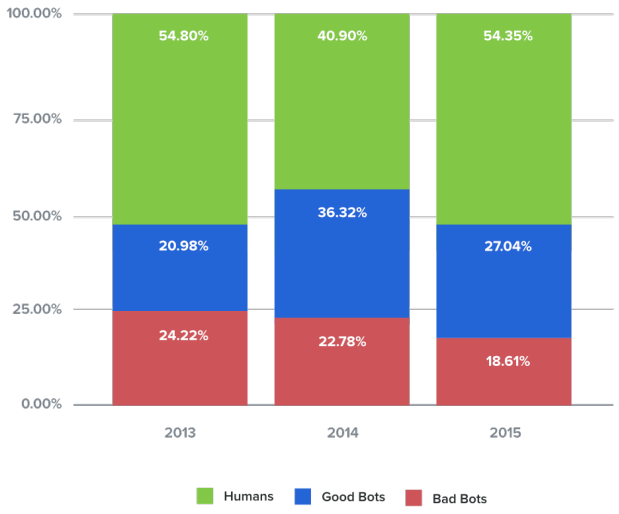
By performing careful analysis on the bots—as well as their origin, behavior, capabilities and evasion techniques, we aim to provide valuable data to all parties looking to make the web more secure.

Key Findings

In 2015, overall bot traffic, as compared to human traffic, decreased slightly from the levels we saw in 2013 and 2014. From 2014 to 2015, good bot traffic decreased from 36.32% to 27.04% of website traffic, and bad bot traffic decreased from 22.78% to 18.61%. The result is that humans now make up 54.4% of all website traffic.

To explain the rise in human traffic, we drew two conclusions, supported with external research and corroborated by the findings herein. First, there has been a significant influx of new internet users, especially from China, India, and Indonesia. Second, bot operators continue to improve their software, creating more advanced persistent bots (APBs). Bad bot operators are opting for quality over quantity.

Bad Bot, Good Bot, and Human Traffic, 2013 to 2015



What is an Advanced Persistent Bot?

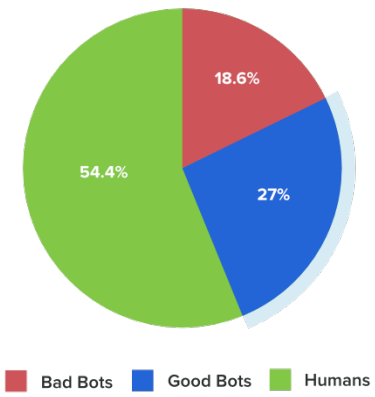
We're seeing ever increasing sophistication in the bad bot threat landscape. So much so that our data warranted the creation of a new term to describe the phenomenon: Advanced Persistent Bots (APBs). They now make up 88% of bad bot traffic, up from 77% in 2014. Meanwhile, simple bots decreased by more than half, from 23% of bad bot traffic in 2014 to 12% in 2015.

APBs have several advanced capabilities. This includes mimicking human behavior, loading JavaScript and external resources, cookie support, browser automation, and spoofing IP addresses and user agents. APBs are much harder to identify and block than simple bots; they fly under the radar of many existing security solutions.

The persistency aspect comes from their ability to evade detection using tactics such as dynamic IP rotation (from huge IP address pools), using Tor networks and peer-to-peer proxies to obfuscate their origin, and distributing attacks over hundreds of thousands of IP addresses. For example, one bot might go through 1,000 IP addresses to make one request apiece, instead of a single IP address to make 1,000 requests.

In fact, bad bots rotating IP addresses is now commonplace. 73% of bad bots rotate or distribute their attacks over multiple IP addresses and of those, a whopping 20 percent surpass 100 IP addresses during the course of their operations. Bad bots are also changing their identities en masse. 36 percent of bad bots disguise themselves using two or more user agents, and the worst APBs change their identities over 100 times.

Bad Bot, Good Bot, and Human Traffic, 2015



Humans finish first in all sized websites

The aforementioned trend of an increase in human traffic holds true across all sizes of surveyed websites. However, there was a noticeable increase of bad bot traffic for medium-sized websites from 2014 to 2015; these increased from 17% to 26% of all traffic.

Traffic Distribution by Size of Website, 2014 vs 2015

	2014			2015		
Site Size (Alexa)	Bad Bots	Good Bots	Humans	Bad Bots	Good Bots	Humans
Large (1 - 10,000)	23%	44%	33%	16%	26%	57%
Medium (10,001 - 50,000)	17%	35%	48%	26%	18%	56%
Small (50,001 - 150,000)	32%	27%	41%	15%	43%	43%

Digital publishers and real estate websites hit hardest by bad bots

Taking both size and industry into account, two groups were hit hardest by bad bots: small digital publishers, serving niche audiences, and large real estate websites. Digital publishing remained on top for having the highest percentage of bad bots. At 31%, this is a slight decrease from 2014.

The more interesting story is real estate, which saw over a 300% increase in bad bot activity. This is likely due to the recent explosion of real estate startups, which may be taking a page out of the travel metasite playbook by scraping and aggregating data to get their businesses off the ground. Why license the data when you can scrape it for free, until your business model proves itself?

Niche publishers and real estate giants crawling in bad bots

An average of 56% of the traffic visiting smaller digital publishers in 2015 was comprised of malicious bots—presumably to scrape the unique and valuable content of these sites. The real estate industry had the opposite problem, where smaller websites faced a comparatively small amount of bad bot traffic, while large, well established sites saw a 48% traffic average originating from bad bots. This further supports our theory that an influx of startups is scraping the established real estate players.

Traffic Distribution by Industry, 2014 vs 2015

	Bad Bots	Good Bots	Humans	Bad Bots	Good Bots	Humans
Digital Publishing	32%	18%	50%	31%	20%	49%
Real Estate	11%	49%	40%	32%	28%	41%
E-Commerce	20%	15%	65%	17%	14%	68%
Directories & Classifieds	21%	48%	31%	16%	49%	34%
Airlines & Travel	28%	3%	69%	7%	6%	87%

Traffic Distribution by Size and Industry, 2014 vs 2015

E-Commerce	2014			2015		
Site Size (Alexa)	**Bad Bots**	**Good Bots**	**Humans**	**Bad Bots**	**Good Bots**	**Humans**
Large	22%	9%	69%	17%	13%	70%
Medium	10%	10%	80%	18%	18%	64%
Small	10%	9%	81%	23%	29%	48%

Real Estate	2014			2015		
Site Size (Alexa)	**Bad Bots**	**Good Bots**	**Humans**	**Bad Bots**	**Good Bots**	**Humans**
Large				48%	22%	30%
Medium	Data Not Available			24%	21%	55%
Small				19%	34%	47%

Airlines & Travel	2014			2015		
Site Size (Alexa)	**Bad Bots**	**Good Bots**	**Humans**	**Bad Bots**	**Good Bots**	**Humans**
Large	16%	2%	82%	7%	6%	87%
Medium	5%	3%	92%	10%	3%	87%
Small	57%	4%	39%	14%	22%	64%

Digital Publishing	2014			2015		
Site Size (Alexa)	**Bad Bots**	**Good Bots**	**Humans**	**Bad Bots**	**Good Bots**	**Humans**
Large	20%	27%	53%	30%	25%	45%
Medium	37%	15%	48%	29%	17%	54%
Small	31%	13%	57%	56%	20%	24%

Directories/Classifieds	2014			2015		
Site Size (Alexa)	**Bad Bots**	**Good Bots**	**Humans**	**Bad Bots**	**Good Bots**	**Humans**
Large	31%	36%	33%	17%	51%	32%
Medium	14%	51%	35%	37%	40%	23%
Small	12%	63%	25%	11%	48%	41%

Chrome the new king and Safari usage on the rise

Bad bots are always looking for ways to evade detection. One way they do this is by changing user agents to report themselves as using a browser popular with human users. In 2015, Chrome took the lead as the most frequently self-reported user agent, making up 26.90% of all user agents leveraged by bad bots. This edged out Firefox, which dropped from 26.62% in 2014 to 17.67% in 2015. With the exception of Internet Explorer usage, user agents selected by their operators for bad bots appear to closely mimic usage trends of real human users (likely to better disguise the bots).

According to Netmarketshare, a site which lists the market share of browsers and operating systems, in 2015 Chrome had an average of 27.88% market share, while Safari averaged 4.91%. Both values correlate very closely to the user agents leveraged for bad bots in our 2015 dataset.

Another interesting explanation for the correlation between self-reported browsers and browser usage comes from the malware world. There are entire categories of bots which solely exist within the infected browsers of their human hosts.

Top Self-Reported Browsers, 2014 vs 2015

	2014		2015	
Rank	User Agent	% of Total	User Agent	% of Total
1	Firefox	26.62%	Chrome	26.90%
2	Chrome	26.01%	Firefox	17.67%
3	Internet Explorer	23.28%	Internet Explorer	11.94%
4	Apache HTTP Client	12.67%	Safari	5.01%
5	Android Webkit Browser	4.87%	Android Webkit Browser	4.37%

Bad Bot Traffic Origins

Chinese ISPs muscle into the most malicious originators list in 2015

The US had the lion's share of bad bot originators in past bad bot reports. This year, however, Chinese sources were on the rise. Out of this year's top twenty ISPs having the highest percentage of bad bot traffic, six came from China. As directed from their servers to our customers, over 72% of the traffic from these ISPs were comprised of bad bots. China Unicom reached a whopping 90% of bad bot traffic.

Verizon Business, Comcast and Time Warner clean up their act

Despite their repeated appearance in the top bad bot originators list in 2013 and 2014, Verizon Business, Comcast and Time Warner fell off the Top 20 Bad Bot Originators for 2015.

Top 20 Bad Bot Originators, 2015

Rank	ISP	Country	Bad Bot Traffic as a % of this SPs traffic	Bad Bot Traffic from this SP as a % of all Bad Bot Traffic Detected
1	China Unicom IP network	China	89.44%	0.3050%
2	Viewqwest Pte Ltd	Singapore	88.90%	0.6441%
3	BSB-Service GmbH	Germany	86.54%	0.2216%
4	Amazon Technologies	United States	84.11%	0.5398%
5	Inflow	United States	83.42%	0.2871%
6	PlusServer AG	Germany	81.57%	0.4620%
7	Daum Communication Co.,LTD	Korea	80.81%	0.4736%
8	Hosting Solutions International	United States	80.59%	0.3238%
9	Henan Mobile Communications Co.,Ltd	China	79.92%	0.1112%
10	China Unicom Jiangsu	China	77.61%	0.2104%
11	BeiJing CloudVsp.Inc	China	76.18%	0.0855%
12	Simpli.fi	United States	75.96%	0.1267%
13	DigitalOcean	United States	75.24%	0.5469%
14	Aliyun Computing Co., LTD	China	73.78%	0.1557%
15	Vietnam Posts and Telecommunications (VNPT)	Vietnam	73.65%	0.1056%
16	SoftLayer Technologies	United States	73.18%	1.0441%
17	One.Tel Ltd	Australia	72.92%	0.0699%
18	Shenzhen Tencent Computer Systems Company Limited	China	72.86%	0.0867%
19	OVH SAS	France	72.28%	1.8101%
20	Oak Point Partners	United States	71.77%	0.1079%

Amazon earns a hat-trick

Amazon has appeared in the top 5 bad bot originators for three years in a row. The same attributes making Amazon EC2 so popular among startups, such as ease of use and its ability to scale on demand, also make it a great choice for bot operators. For example, bots can be loaded onto AMIs, and then spun up and down as needed (based on the demands of a given project, or the need to obtain new IP space in order to avoid detection or evade IP blacklisting).

Honorable mentions

These organizations have made their way into the Top 20 bad Bot Originators list in two of the last three bad bot reports:

- Comcast
- Time Warner
- Verizon Business
- Verizon FioS
- OVH SAS
- Gig Avenue
- Google
- Softlayer Technologies
- Softlayer Dutch Holdings BV
- CIK Telecom

Top 20 Bad Bot Originators, 2013 to 2015

	2013	2014	2015
Rank	ISP	ISP	ISP
1	Verizon Business	Amazon Technologies	China Unicom IP network
2	Level 3	Amazon.com	Viewqwest Pte Ltd
3	Amazon Technologies	Verizon Business	BSB-Service GmbH
4	Las Vegas NV Datacenter	Cogent	Amazon Technologies
5	Amazon.com	Comcast Cable	Inflow
6	Hosting Solutions, Intl.	CIK	PlusServer AG
7	Comcast Cable	Bezeq	Daum Communication Co.,LTD
8	Frontier Comm.	OVH SAS	Hosting Solutions International
9	CIK Telecom	Time Warner Cable	Henan Mobile Communications Co.,Ltd
10	Time Warner Cable	GIG Avenue	China Unicom Jiangsu
11	VNET S.R.O	OVH Hosting	BeiJing CloudVsp.Inc
12	SoftLayer Dutch Holdings BV	Internap Network Services	Simpli.fi
13	ThePlanet.com	SoftLayer Dutch Holdings BV	DigitalOcean
14	Telekom Malaysia	Server Block	Aliyun Computing Co., LTD
15	Telekom Malaysia	Hetzner Online	Vietnam Posts and Telecommunications (VNPT)
16	Verizon FiOS	ColoCrossing	SoftLayer Technologies
17	NOC4Hosts	Telecom Italia	One.Tel Ltd
18	Switch Comms.	Tilla BV	Shenzhen Tencent Computer Systems Company Limited
19	Gig Avenue	T-Mobile USA	OVH SAS
20	SoftLayer Technologies	Google	Oak Point Partners

Top bad bot originating mobile carriers

According to a Comscore report, as of 2015 mobile-only visitors surpassed their desktop-only counterparts. In our investigation, we looked at all mobile traffic producing over 100,000 requests during our sample period. We then tracked the mobile carriers having the highest percentage of bad bot traffic.

While mobile carriers having the most bad bot traffic represent a number of countries, the US (5) and the Netherlands (3) were the only ones with multiple carriers on the list.

Top 20 Countries by Mobile Requests, Ranked by % of Bad Bot Traffic

Rank	Country	% Of All Mobile requests From Bad Bot Traffic
1	Netherlands	12.23%
2	Korea	10.93%
3	United States	10.44%
4	India	9.06%
5	Thailand	8.98%
6	Singapore	6.20%
7	Japan	6.14%
8	Mexico	3.95%
9	France	3.78%
10	Saudi Arabia	3.16%
11	Canada	2.74%
12	Germany	2.63%
13	Russia	2.44%
14	Australia	1.93%
15	United Kingdom	1.50%
16	Spain	1.25%
17	Ireland	1.15%
18	Switzerland	0.96%
19	Italy	0.54%

Top 20 Bad Bot Mobile Originators 2015

Rank	ISP	Country	% of Traffic from Bad Bots
1	PT Excelcomindo Pratama	Indonesia	33.31%
2	Telmex Colombia S.A.	Columbia	18.82%
3	China Telecom Shanghai	China	17.56%
4	T-mobile Netherlands bv.	Netherlands	14.82%
5	Wireless Data Service Provider Corp.	United States	13.09%
6	Global Village Telecom	Brazil	11.53%
7	DTAC	Thailand	10.13%
8	AT&T Wireless	United States	10.06%
9	TM Net (Telecom Malaysia)	Malaysia	9.99%
10	Tele 2 Nederland B.V.	Netherlands	9.06%
11	KPN Mobile	Netherlands	7.86%
12	T-Mobile USA	United States	7.70%
13	MobileOne	Singapore	7.49%
14	Verizon Wireless	United States	7.11%
15	Sprint PCS	United States	6.65%

The Netherlands, Korea, and United States lead mobile-heavy countries in bad bots

When comparing bad bot mobile requests, we decided to zoom in on the top countries sending such traffic through our service in order to gather statistically significant results. The Netherlands, Korea, and the United States take the top three spots. Combining this data with that of the previous chart suggests that in the Netherlands and US, a few key service providers are popular with bot operators. In Korea, meanwhile, bots using mobile user agents are popular, but there isn't a heavily-preferred service provider for bot operators.

Countries originating the most bad bots

Now looking at all traffic types, for the second year in a row the US is the big winner in terms of countries from which the greatest number of bad bots originate. The key contributing factor is the ample supply of cheap cloud computing resources such as Amazon, Google Cloud, and Azure. Additionally, many attackers try to use

origin sources and IP addresses similar to their victims in order to better blend in with legitimate human users.

Top Bad Bot Originating Countries 2014 vs 2015

Rank	2014	2015	Change
1	United States	United States	Equal ↔
2	Germany	India	Up 8 ↑
3	Canada	Israel	Up 11 ↑
4	Italy	Germany	Down 1 ↓
5	France	France	Equal ↔
6	The Netherlands	United Kingdom	Up 3 ↑
7	China	China	Equal ↔
8	Russia	Canada	Down 5 ↓
9	United Kingdom	Russian Federation	Down 1 ↓
10	India	The Netherlands	Down 4 ↓

Contributing to at least one percent of global bad bot traffic

Five new countries contributed more than 1% of global bad bot traffic in 2015. All countries from 2013 and 2014 were still on the list, except for Mexico, which dropped to roughly 0.35% of the world's bot traffic.

Most blocked countries (by customers using geo-fencing rules)

Distil customers appear to be becoming more specific with their geo-fencing blacklists. In the past, many geo-fencing rules centered around China, Russia, and less developed countries.

This may be a sign of "painting with broad strokes," where users were blocking countries they believed to be malicious. However, in 2015 we observed users being more specific in the countries they chose to block. Many industrialized or developed countries made the top blocked list.

Countries Contributing at Least 1 % of the Global Bad Bot Supply

2013	2014	2015
United States	United States	United States
United Kingdom	United Kingdom	United Kingdom
Germany	Germany	Germany
Netherlands	Netherlands	Netherlands
Russia	Russia	Russia
Canada	Canada	Canada
Singapore	Singapore	Singapore
China	China	China
	Italy	Italy
	France	France
	India	India
	Israel	Israel
	Malaysia	Malaysia
	Mexico	
		Spain
		Brazil
		Ukraine
		Switzerland
		Australia

Top 10 Most Blocked Countries, 2014 vs 2015

2014 Rank	Country	2015 Rank	Country	Change
1	China	1	China	Equal ↔
2	Russian Federation	2	Norway	Up 3 ↑
3	Hong Kong	3	Germany	Up 5 ↑
4	Bangladesh	4	Netherlands	New +
5	Norway	5	France	New +
6	India	6	Russian Federation	Down 4 ↓
7	Germany	7	Australia	New +
8	Ukraine	8	Sweden	New +
9	Pakistan	9	Switzerland	New +
10	Brazil	10	Japan	New +

Maldives tops global “bad bot GDP” scale

By comparing the number of bad bots per online user within a certain region, we’re able to find areas which have unusually active bad bot activity. At the top of this year's highest “Bad Bot GDP” is the Maldives, with 526 bad bots per online user in the region. In 2014, we reported that much of that country’s traffic may be due to a Russian hacker named “Track 2.” Despite the fact that, later in 2014, Track 2 was arrested and accused of stealing credit cards, it appears that in 2015 the Maldives was still a bot activity hot bed. This may indicate other hackers have set up their base of operations in the Maldives.

Another interesting finding pertains to the activity level of the countries on this year’s list—which rose dramatically. The 2014 top ten counties having the highest bots per capita averaged 26.1 per online user, while in 2015 that number soared to 99.2.

“Bad Bot GDP”, 2014 vs 2015

2014		2015	
Country	# of Bad Bots Per Online User	Country	# of Bad Bots Per Online User
Singapore	153	Maldives	526
Israel	34	Israel	168
Slovenia	30	Kyrgyzstan	94
Maldives	16	Argentina	44
Ireland	9	Singapore	43
United States	6	Montenegro	34
Malta	5	Myanmar	28
Netherlands	3	Malta	20
Romania	3	United States	19
Denmark	3	Ireland	16

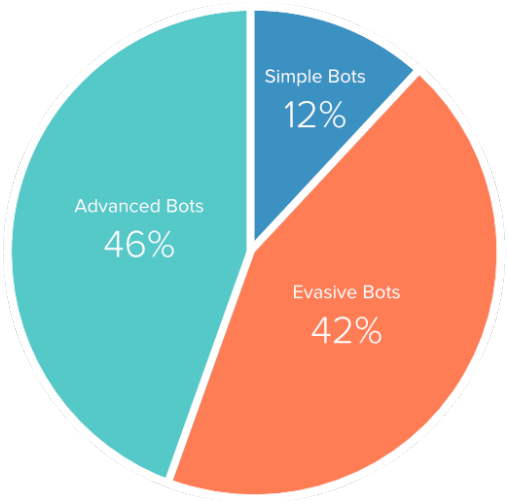
Bad Bot Capabilities and Behavior

Sophisticated bot software more prevalent than ever

Each year we categorize bots based on their level of sophistication. Simple bots have no ability to evade detection, mask their identities, or load JavaScript. The next tier of complexity is what we’ve labeled Evasive bots. These possess a limited ability to disguise themselves and their activities, may rotate IP addresses, change user agents, utilize correct HTTP headers, and more. And then we have Advanced bots, which are able to do things like mimic human behavior, load JavaScript and external assets, tamper with cookies, perform browser automation, and more. In 2015, only 12% of bad bots fell into the simple bots category, 42% were classified as evasive bots, and 46% we deemed to be advanced bots.

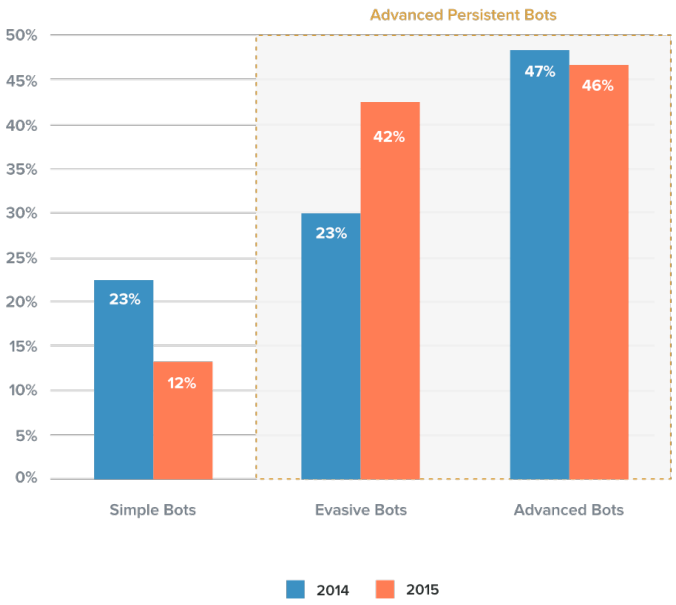
Much education must occur in the market regarding how sophisticated bots have become over the last few years. APBs exhibit one or more characteristics found in either the evasive or advanced bot categories.

Bad Bot Sophistication, 2015



When comparing data between 2014 and 2015, a clear trend emerges—bots have become more sophisticated than in the past. 2015 marked a noticeable shift in bot technology, with roughly 11% of bad bots graduating from the simple category to the evasive category. This supports one of our initial conclusions; that bot operators put some of their 2015 focus on their bot quality, as opposed to quantity.

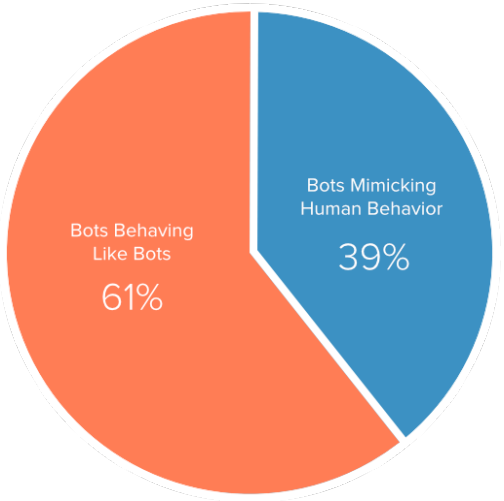
Bad Bot Sophistication, 2014 vs 2015



Many bots mimic human behavior

In 2015, our dataset revealed that roughly 40% of bots are able to mimic human behavior. This makes the case that using tools such as WAFs, web log analysis, or NGFWs—which perform less detailed analysis of clients and their behavior—will likely result in huge amounts of false negatives.

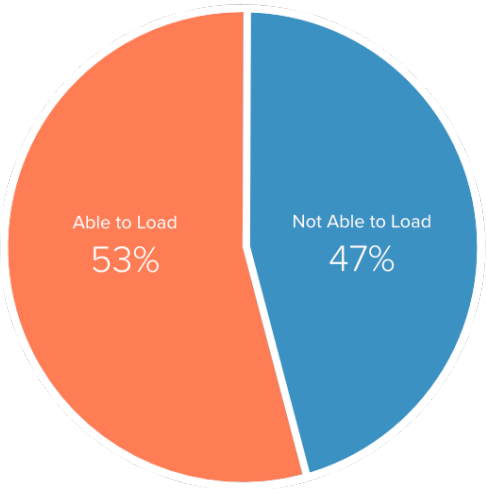
2015, Percentage of Bots Mimicking Humans vs Behaving Like Bots



Bad bot's ability to load external assets such as JavaScript

Many analytic tools, such as Google Analytics, function via a JavaScript code snippet. If bots can load these resources, they'll end up skewing analytic tools and throwing off key business and operational metrics. Based on this year's data, 53% of bad bots will end up falsely attributed as humans in Google Analytics and similar tools.

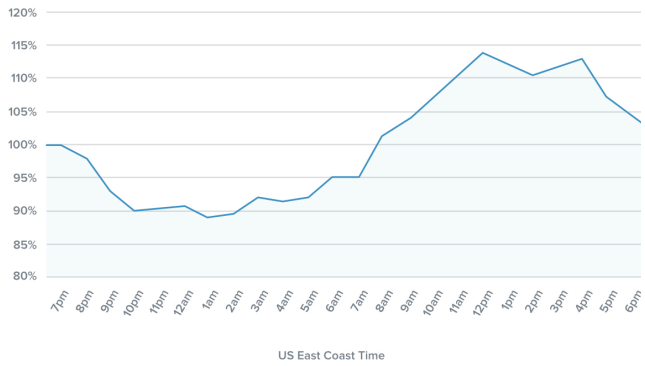
Percentage of Bad Bots Able to Load External Assets, 2015



Bad bots begin to work a “9 to 5”

We normalized time of attack data across all of our data centers and all timezones. An interesting trend emerged; the time of bot attacks heavily correlates with US East Coast daytime (UTC -5) working hours. A potential explanation is that as bots get more sophisticated, they use the time of day as a method of disguise. Bots then visit websites, which already have ample human traffic, at predictable hours, looking to blend in and fly under the radar of detection mechanisms.

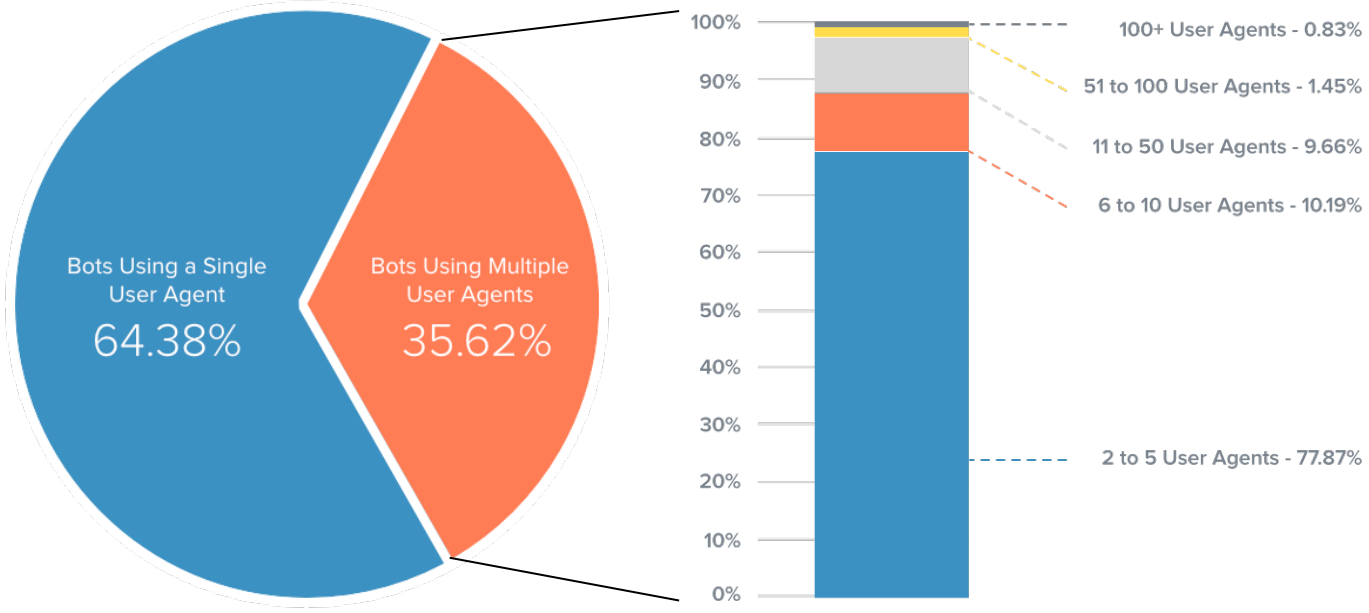
Time of Bad Bot Activity, 2015
(% of Average Bot Activity vs. Time)



Bots rotate user agents en masse

Not only are the bad guys lying about who they say they are, they’re repeatedly changing their identities over and over again. According to this year’s data, around 36% of bad bots disguised themselves using two or more user agents. The worst APBs changed their identities over 100 times.

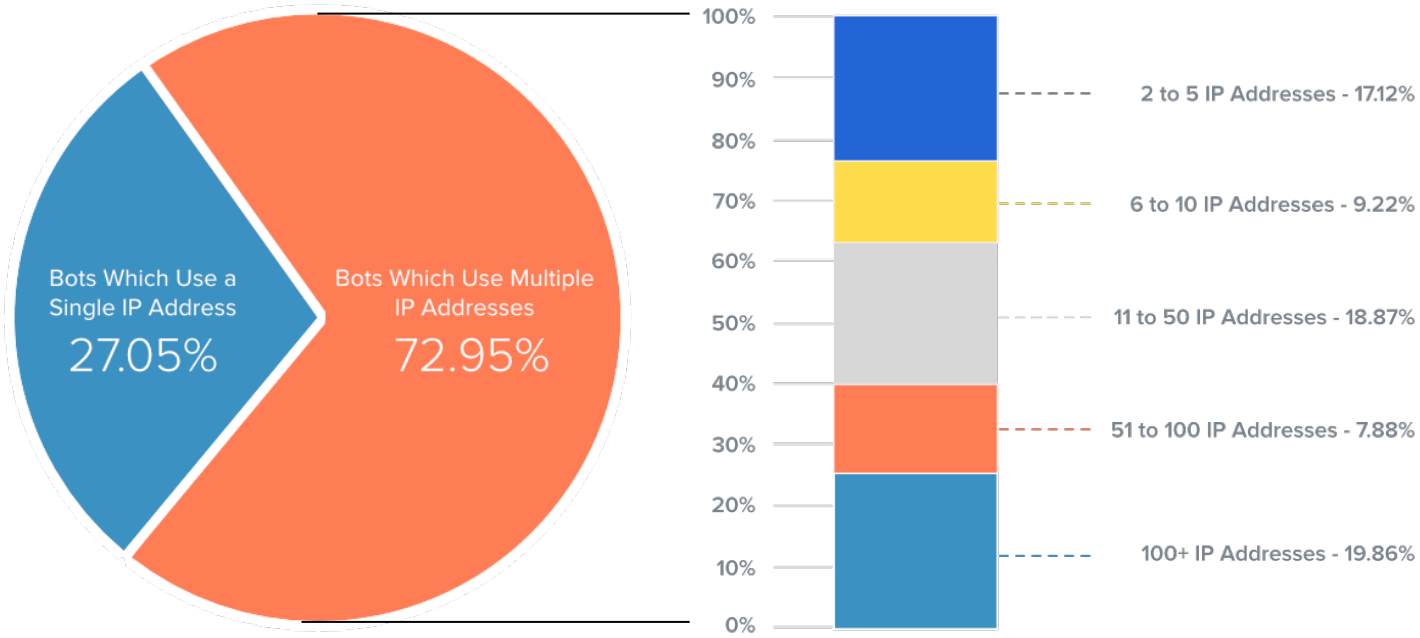
Distribution of Bad Bots Which Use Multiple User Agents, 2015



Bots rotating IP addresses is now a commonplace tactic

Almost 73% of bad bots rotate or distribute their attacks over multiple IP addresses. Of those, 17% utilized between two to five IP addresses, 9% between six and ten, 19% between 11 and 50, 8% up to 100 IP addresses, and a whopping 20% used over 100 IP addresses in their operations.

Distribution of Bad Bots Which Use Multiple IP addresses, 2015



Conclusion

The bad bot landscape continues to evolve rapidly, especially in relation to the sophistication of bot software and the number of bots coming from Chinese service providers. 2015 saw a dramatic increase in APBs, which have sophisticated capabilities. And the advent of cheap or free cloud computing resources lets anyone with basic computer skills download open source software and get into the bot 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 little or no visibility or control over malicious website traffic.

About Distil Networks

Distil Networks is the global leader in bot detection and mitigation. Our service is the first easy and accurate way to identify and police malicious website and API traffic, blocking 99.9% of bad bots without impacting legitimate users. Distil protects against web scraping, brute force attacks, competitive data mining, online fraud, account hijacking, unauthorized vulnerability scans, spam, man-in-the-middle attacks, digital ad fraud, and downtime.

Slash the high tax that bots place on your internal teams and web infrastructure. Make your online applications more secure with API security, real-time threat intelligence, a 24/7 security operations center, and complete visibility and control over human, good bot, and bad bot traffic.

- Harden your website and API 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.