IMPERVA®

REPORT

Imperva Incapsula 2015 Bot Traffic Report

Humans Take Back the Web, Bad Bots Not Giving Any Ground

Introduction

Imperva Incapsula's annual Bot Traffic Report, now in its fourth year, is a statistical study of the typically imperceptible bot traffic landscape.

Much has changed in the general understanding of bots since we first revealed them to be responsible for the bulk of all website traffic. Today, it is not uncommon to find entire articles (including our own) dedicated to the study of individual bots: their HTTP footprints, points of origin and the nuances of their behavior.

Collectively, however, these non-humans are still discussed in terms of two archetypes: Good Bots and Bad Bots.

Good bots are the worker bees of the Internet that assist its evolution and growth. Their owners are legitimate businesses who use bots to assist with automated tasks, including data collection and website scanning.

Bad bots, on the other hand, are the malicious intruders that swarm the Internet and leave a trail of hacked websites and downed services. Their masters are the bad actors of the cyber-security world, from career hackers to script kiddies. In their hands, bots are used to automate spam campaigns, spy on competitors, launch denial of service (DDoS) attacks or execute vulnerability scans to compromise websites on a large scale.

In the past, good and bad bots have always been responsible for most of the activity on our network. This year, however, we saw a changing of the guard, with humans stepping in to become the Internet's new majority.

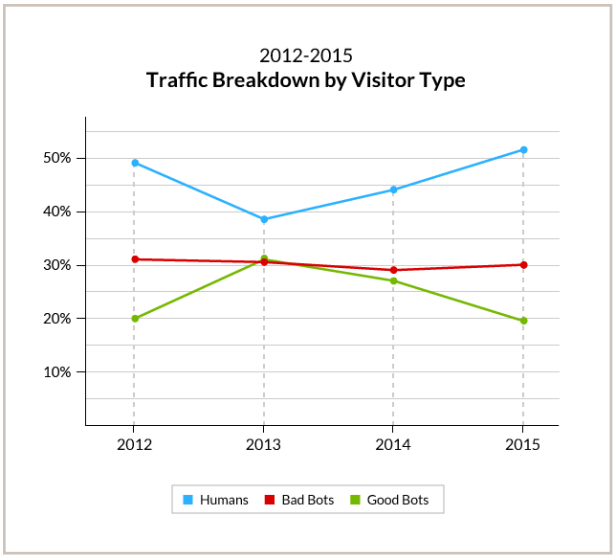
How Humans Are Taking Back the Web

When analyzing our data, we were very surprised to find out that, for the first time, humans were the ones responsible for the majority (51.5 percent) of all online traffic.

These numbers initially appeared to be a complete trend reversal. However, upon closer inspection, and when put in the context of our previous analyses, they actually signaled the continuation of a trend, which consists of:

- An increase in the relative amount of human traffic, from 38.5% in 2013 to 51.5% in 2015.
- A decrease in the relative amount of good bot traffic, from 31% in 2013 to 19.5% in 2015.
- A static amount of bad bot traffic, which fluctuates around 30%.

What's important to note here is that all of these number are relative. This means that each of the groups is only able to grow at the expense of another. It is also important to mention that–in absolute terms–the sheer amount of bot and human traffic on our network is always increasing.



So, the question here is not: Why are good bots becoming less active? But, rather: Why aren't they keeping up with the other two groups?

To answer that, we drilled deeper into our data and this is what we found out:

1. Across-the-Board Decline

Our first assumption was that the downtrend in good bot traffic could be attributed to an individual bot, or group of bots, as was seen with RSS bots last year. To test this theory, we performed a year-over-year comparison of all good bot clients.

To our surprise, the comparison showed that the downtrend occurred across the board, with 442 of the 484 good bot clients displaying negative growth or an extremely low growth of less than 0.01 percent. Individually, none of them declined by more than two percent.

Moreover, out of 42 good bots that displayed increased activity, only one displayed an increase of more than one percent.

Basically, what we had here was a case of good bots collectively not keeping up with growth of human and bad bot traffic.

2. Can't Keep Up on Popular Sites

Our next step was to look more closely at the websites in our sample group.

Last year, we segmented those sites based on the amount of daily visits they received, with the intention of helping website owners better understand their own bot situation. This year, we decided to go back to those stats, using them as a basis for a year-over-year analysis.

Our goal? To see if the decline in bot traffic was especially steep for any of those website groups.

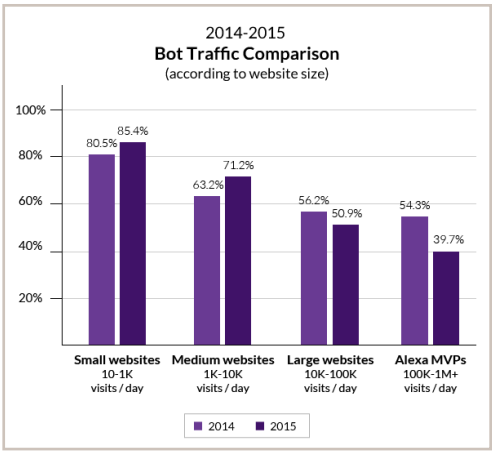
The effort paid off. What we saw was an asymmetric decrease in bot traffic, with traffic to smaller and mid-sized websites actually trending up, while the traffic to larger and more popular websites trending down.

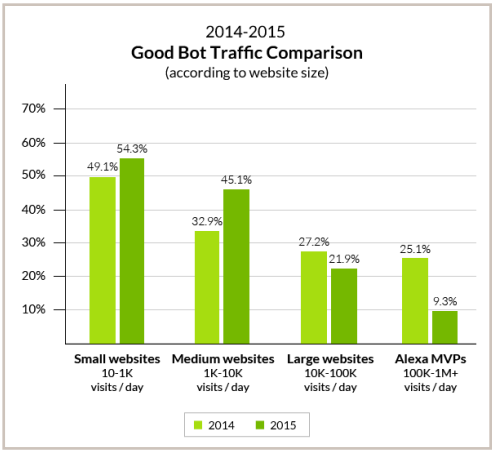
Most prominently, we saw that bot traffic to the most popular websites (those with 100,000 daily visits or more) went down from 54.3 percent to 39.7 percent.

Next we zoomed in and looked at good/bad bot traffic. There, we saw that a year-over-year decline in good bot traffic occurred solely on the high-tier websites.

Most prominently, we noticed that on websites with 100,000 or more daily human visits, good bot activity went down from 21.9 percent to 9.3 percent. This, interestingly enough, was in contrast to activity on low-tier websites, where good bot traffic actually went up.

So, we had our "smoking gun". It appeared that, the more popular a website got, the harder it was for the good bots to keep up with the influx of human and bad bot visits.



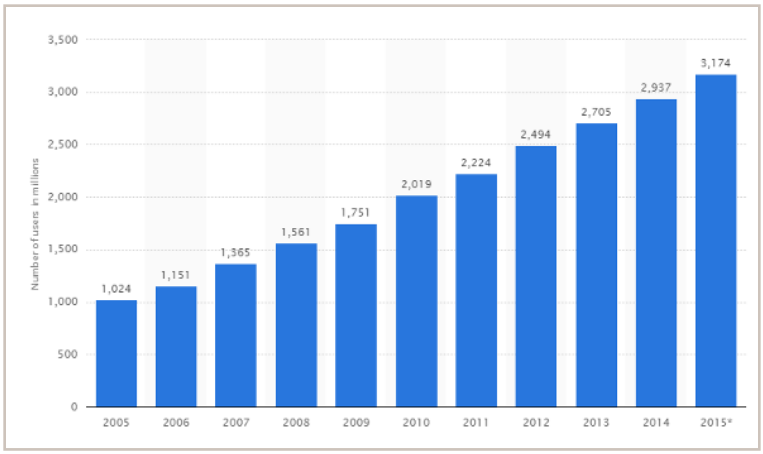


A Question of Motivation

To understand why good bots do not frequent popular sites as often as humans, it is important to consider the motivations that drive good bot traffic.

Broadly speaking, there are two main reasons behind the visit of every good bot:

- **Indiscriminate crawls** – These consist of various crawlers, including search engine bots, marketing research tools (e.g., backlink checkers), legitimate scrapers (e.g., WayBackMachine bot) and so on. These crawlers do not target your website specifically. Your site is one of many domains they keep track of on a regular basis.
- **Targeted scans** – These involve you, or someone interested in your website, using a piece of software to initiate a targeted scan (e.g., uptime monitors, vulnerability scanners, SEO tools, etc.).



Worldwide Internet population, in millions of users (source)

In both cases, high website popularity is unlikely to translate into increased good bot traffic.

For instance, one of our previous studies already showed that there is no correlation between website's popularity and the amount of Google bot visits. This is also the case for the majority of crawlers. And while it's reasonable to think that a popular website will draw more targeted scans, the amount of extra sessions generated are negligible.

But while good bots are agnostic to popular content, humans are not. In fact, the more popular a website gets, the faster its human population tends to grow—its accelerated growth supported by repeated visits and virality, the latter powered by the (also) constantly growing social media.

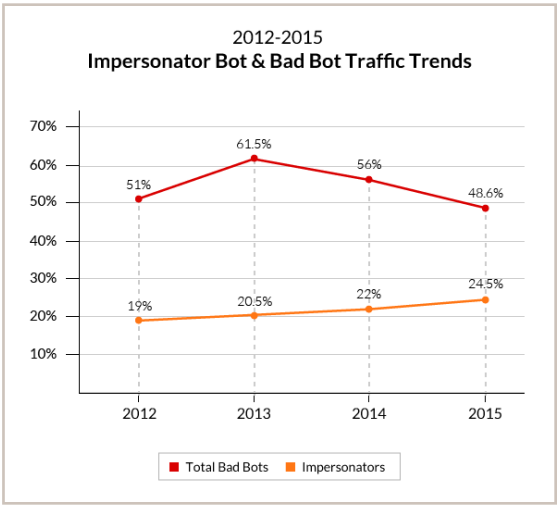
Furthermore, the population of humans on the Internet is consistently growing, as evidenced by the roughly 8.1 percent uptick over the last year in the graph below. Meanwhile, the time we spend online has increased as well.

Bad Bots Are as Active as Ever

In light of the above mentioned trends, it is interesting to note that we saw no decrease in bad bot activity, neither on smaller or more popular websites.

There are several key reasons for this:

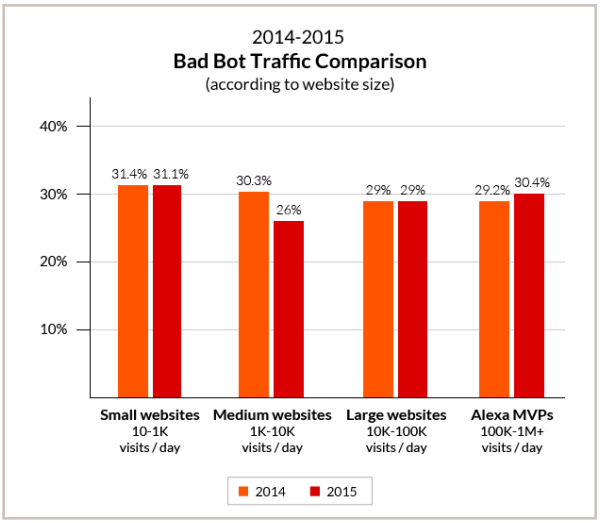
- Unlike good bots, malicious non-humans are often employed by individuals or small criminal groups, not organizations. As a result, their growth is more closely tied to the increase in Internet human population.
- For every Internet user in the developed world, there are two in developing countries. Industry reports, including our own, repeatedly show these countries to be the hubs of botnet activity, indicating a wider availability of under-secured connected devices.



- DDoS tools and DDoS-for-hire services that cause an influx in DDoS traffic are widely available. Recently, we saw this lead to a 121.9 percent quarter-over-quarter increase in the amount of DDoS bots, categorized here as 'Impersonators'. Not surprisingly, this is also the only bot group that has displayed consistent growth over the last four years.

As a result, websites of all sizes are as targeted as ever, and are visited on average by one malicious non-human for every two humans.

These bots continue to pose a serious threat to the Internet ecosystem. The extent of this threat is such that, on any given day, over 90 percent of all security events on our network are the result of bad bot activity. From the looks of it, these are not going away any time soon.

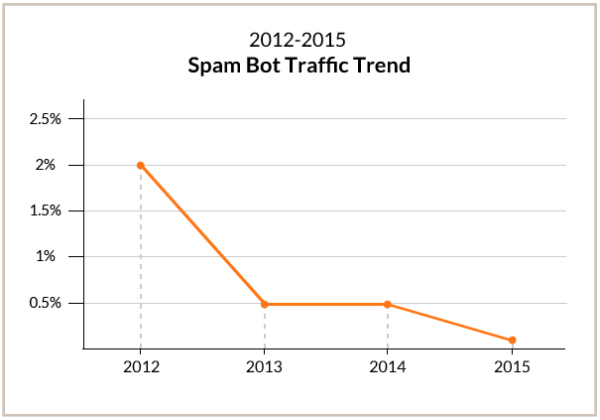


Learning from Spam Bots

So what can be done to minimize the threat posed by malicious bots? Individually, website operators can counter malicious bots with security solutions, like our own, which offer protection from application layer attacks. These include automated hacking attempts, scraping, spamming and DDoS attacks.

On a larger scale, however, security solutions can only treat the symptoms of the actual problem, which is a rapid expansion of botnets, leading to growing amounts of automated attacks.

While it would be naïve to assume that such malicious activity can be completely eradicated, there are ways to limit the motivations that drive its growth. To us, this is the moral of the story of Spam bots; a group of malicious bots that saw a radical decline, from two percent in 2012 to 0.1 percent in 2015.



Once again, it all boils down to motivation.

Prior to 2012, spam bots were commonly employed by Black Hat SEOs who used them to auto-inject links across dozens, hundreds or even thousands of domains at a time. The incentive was the SEO boost the spam links provided, which helped the linked domain climb higher in search engine result pages (SERP).

Things changed in April 2012, when Google stepped up its war against Spam by introducing the first of its Penguin updates. The change in the algorithm penalized all websites involved in spam distribution—not only those who profited from spam links, but also those who hosted them.

With this new proactive approach, Google turned link spam into a high risk/low reward activity. More importantly, it motivated website operators to harden their anti-spam measures to prevent being penalized for allowing spam links on their websites.

This zero-tolerance approach, which promotes security awareness and individual responsibility should be applied to other aspects of online activity.

For example, industry standards could be used to enforce hardened security of connected devices, the negligent handling of which often leads to the creation of botnets that spawn malicious bot traffic. On the other end, legal action and consumer movements could promote a no-negotiation policy with cyber-extortionists, taking away one of the main motivations for hacker activity.

This year, for the first time ever, concerns about botnet activity were voiced by the White House. We also witnessed the prosecution of high-profile botnet operators. While these are unlikely to deter other perpetrators, such steps are important to promote awareness and guide a discussion on the topic of malicious bots, their origin, MO and damage potential.

What's still missing are ground rules that can be adopted by the private and governmental sectors. These rules need to define how the botnet threat should be address and contained, if not resolved.

Methodology

The data presented herein is based on a sample of over 19 billion human and bot visits occurring over a 90-day period, from July 24, 2015 to October 21, 2015. It was collected from 35,000 Incapsula-protected websites having a minimum daily traffic count of at least 10 human visitors.

Year-over-year comparison relies on data collected for the previous Bot Traffic Report, consisting of a sample of over 15 billion human and bot visits occurring over a 90-day period, from August 2, 2014 to October 30, 2014. That sample was collected from 20,000 Incapsula-protected websites having a minimum daily traffic count of at least 10 human visitors.

Geographically, the observed traffic includes all of the world's 249 countries, territories, or areas of geographical interest (per codes provided by an ISO 3166-1 standard).

The analysis was powered by the Incapsula Client Classification engine—a proprietary technology relying on cross-verification of such factors as HTTP fingerprint, IP address origin, and behavioral patterns to identify and classify incoming web traffic. Our Client Classification engine is deployed on all Incapsula-protected websites as part of the company's multi-tier security solution.

© 2016, Imperva, Inc. All rights reserved. Imperva, the Imperva logo, SecureSphere, Incapsula and Skyfence are trademarks of Imperva, Inc. and its subsidiaries. All other brand or product names are trademarks or registered trademarks of their respective holders.