

DIGITAL CONTENT NEXT & WHITE OPS, INC.

DCN

Bot Benchmark Report

What Makes
a Publisher Premium



2015



Table of Contents

DCN Bot Benchmark Report

What Makes a Publisher Premium

Overview.....	4
Major Findings	6
Policies Impact Sophisticated Bot Rates	10
Habits of Highly Successful Publishers.....	12
The Sophisticated Bot Picture for Premium Publishers.....	16
The Rise of Ad Injection Malware	24
Study Methodology	26
Conclusions.....	29
Recommendations: Publisher Action Plan	30
DCN Participants	32

Introduction

In December 2014, the Association of National Advertisers and the digital advertising security firm **White Ops** partnered to release what is now considered the seminal report on advertising fraud. This groundbreaking report made it clear that advertising fraud is the clearest risk to the integrity of the digital media business. There is nothing more important to our member brands than building and keeping the trust of the consumers and the marketers who choose to do business with us. Issues of trust, regardless of where they originate, impact our members in the marketplace. And while we entered last year with viewability as our top advertising issue, fraud is now front and center.

Fraud and viewability are two distinctly different, yet related, challenges. Viewability involves how buyers and sellers measure media — a negotiation if you will. Bot fraud is far more pernicious, because it is criminal behavior driven by financial incentives that takes advantage of everyone in the digital marketplace -- marketers, publishers, agencies, and consumers.

With this in mind, I asked our members how we could best attack the issue. Our goal is to shed more light on the problem and come forward with more solutions and best practices. We decided to conduct our own study with White Ops, as we see our members as “White Hats” who employ policies and business practices that ensure minimal exposure to ad fraud. Nonetheless, we have an opportunity — and a responsibility — to provide leadership for this marketplace.

We recruited a representative sample of our membership to participate in this study. The response was beyond expectations. Thirty-two sites representing close to 90 percent of our membership revenue participated in the study. We asked White Ops and their expert researchers to produce a robust study that would offer insights to both our members and the industry about how we can minimize ad fraud going forward. The study serves as an extremely valuable sales tool for our members in their representation as the most trusted brands on the web. Low ad fraud rates on our sites are a crucial factor in what makes us premium.

While digital media is fraught with challenges, there is also an incredible amount of innovation that is bringing new opportunities to consumers, advertisers and publishers. This research and these open and transparent discussions should continue as they are critical to fostering a vibrant digital media world for the future.

Thank you for the attention to this important topic. As always, we welcome your feedback.



Jason Kint

Chief Executive Officer
Digital Content Next

Overview

The online advertising ecosystem relies on groups of interconnected participants cooperating and competing to deliver content to web and mobile visitors. In many ways, publishers are the last link in a chain designed to deliver advertising to their readers, but too often the viewer is not a human, but a software program.

These programs, or bots, come in many different forms. They can collect information about the site, scrape content from a page, or take advantage of other weaknesses in the ecosystem to make money illegally. Bots that consume ads on web sites without permission or mimic a human click on an advertisement cause advertisers to pay for marketing that was never delivered to an actual human.

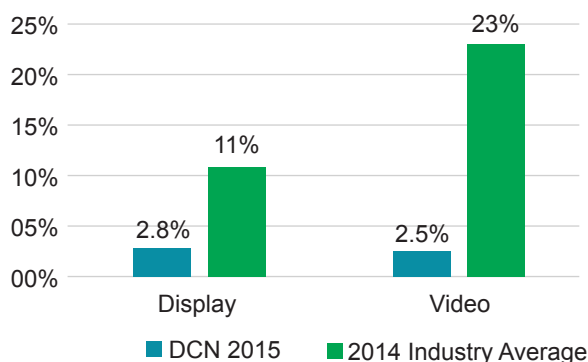
In 2014, White Ops conducted research with the Association of National Advertisers (ANA) and 36 of their members in the automotive, beverage, consumer products, financial, insurance, hospitality, pharmaceuticals, and technology industries to measure more than five billion non-mobile ad impressions over 60 days. *The Bot Baseline: Fraud in Digital Advertising (The Bot Baseline 2014)* found that approximately 11 percent of display advertising impressions were fraudulent. *The Bot Baseline 2014* was broad-based and included both premium and non-premium publishers and inventory.

The *DCN Bot Benchmark Report* measures the bot impact in high-quality inventory. In partnership with Digital Content Next, an association of premium publishers, White Ops collected almost six times the number of impressions as in the *2014 Bot Baseline* study.

- **32 DCN member organizations participated**
- **Study time frame was 53 days, from June 22, 2015 to August 14, 2015**
- **The study analyzed more than 16 billion desktop display and video impressions**
- **Overall traffic to tagged sites and ads contained 2.8 percent covert sophisticated bots**

The study found that sophisticated bots accounted for 1.6 to 6.9 percent of all traffic to the domains studied. The average for display inventory was one-quarter of the level of the Internet average for display inventory as documented in *The Bot Baseline 2014*, and the average bot percentage for video inventory in the domains studied was one-tenth of the Internet average.

Benchmark: Average Sophisticated Bot % by Media Type



Selectively buying direct-placement digital ad inventory from this group of premium publishers, if quality is maintained consistently across inventory, could allow marketers to achieve the valid impression levels of the top quintile of *The Bot Baseline 2014* marketers without instituting any other policy changes or buying decisions.

Methodology Overview

The 2015 *DCN Bot Benchmark Report* and *The Bot Baseline 2014* calculated results using measurable desktop (non-mobile) inventory. For unmeasurable inventory, which accounts for 7.4 percent of total impressions for DCN participants, this study conservatively assumes that the bot percentage is similar to the fraction of bots detected in measurable inventory.

While the *DCN Bot Benchmark Report* involved fewer participants -- 32 companies compared to *The Bot Baseline*'s 36 participants -- the analysis involved a much larger data set: 16 billion impressions compared to 5.5 billion for *The Bot Baseline* data set.

For total bot percentage calculations, the study removes **general bots** (bots known to the industry, such as search engine indexers and known spiders) to match *The Bot Baseline*. The bots detected in this study are **sophisticated bots**, those that will not be detected by using the industry spider and bot list and known browser list. The impact of general bots can be eliminated by removing general bots during billing, while sophisticated bots are not known to the industry and can cause transactions based on fraudulent impressions to occur.

Comparison to the averages seen in *The Bot Baseline 2014*

This study provides benchmark data compared to *The Bot Baseline 2014*. The results from the *DCN Bot Benchmark Report* apply to publishers in the "premium" category who are pro-actively working to maintain valid impression levels in their inventory.

Data in this study and *The Bot Baseline 2014* was supplied by the participants who agreed to participate in the studies to assess the impact of bots on digital advertising inventory.

Significant differences exist between the two studies. The *DCN Bot Benchmark Report* contrasts with *The Bot Baseline* in the following ways:

- *The Bot Baseline* inventory covered several buying platforms including exchanges, programmatic-direct, private exchanges, and programmatic DSPs (demand side platforms), while the *DCN Bot Benchmark Report* shows bot levels in premium direct and programmatic inventory.
- *The Bot Baseline* examined a cross section of media types, while the predominant media type studied in the *DCN Bot Benchmark Report* was premium display inventory.
- *The Bot Baseline* examined a diverse set of quality levels, site volumes, campaign volumes, campaign and placement types, and other parameters. The *DCN Bot Benchmark Report* focused on carefully managed inventory on predominantly high-volume sites owned and operated by nationally recognized publishers.

The Bot Baseline 2014 data set is more representative of typical Internet traffic, because it encompasses data from participants throughout the advertising ecosystem -- including marketers, publishers and advertising networks -- while the DCN data set comes from a single class of sources -- well-known publishing sites. Some comparisons can be made to *The Bot Baseline 2014* study, but it is important to remain aware of the contrasts between the studies and consider the results in context to understand the impact of bots in different types of inventory.

Major Findings

Participants had lower sophisticated bot rates than the 2014 average

The domains studied had lower levels of sophisticated bot traffic overall (2.8%) compared to the average across the sites in The Bot Baseline 2014 (11%)

Using data from more than 16 billion impressions, the study found that sophisticated bots accounted for 2.8 percent of publishers' traffic coming from browsers on desktop systems, laptop computers and gaming consoles. During the study period, an additional 14 billion impressions came from browsers on mobile devices. For accurate comparison to *The Bot Baseline 2014* study, results in this study are confined to desktop (non-mobile) inventory of 16 billion impressions.

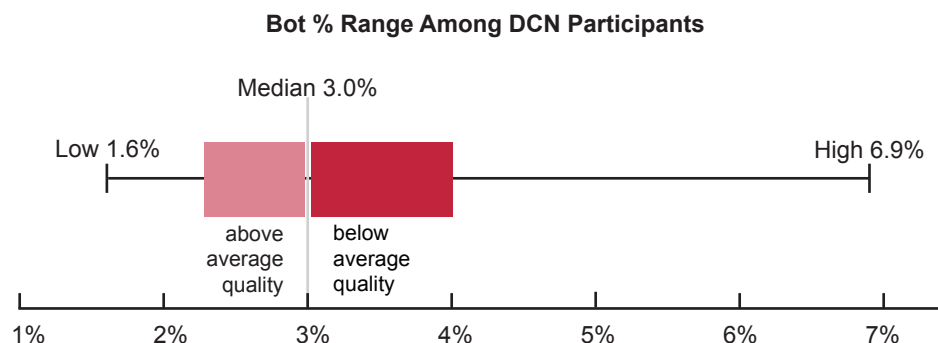
Overall, sophisticated bots accounted for **2.8 percent of all traffic to all DCN publishers**. The overall sophisticated bot percentage is **approximately one quarter of the 11 percent rate of sophisticated bots in display inventory** detected across all ads in the ANA and White Ops industry-wide study *The Bot Baseline: Fraud in Digital Advertising (The Bot Baseline 2014)*.

Among DCN publishers who participated in the *DCN Bot Benchmark Report*, there was a much lower bot percentage, ranging from 1.6 percent to 6.9 percent, with an average of 3.0 percent.

Some DCN publishers had higher bot rates

Even among DCN publishers, some inventory shows high sophisticated bot rates

While the average DCN publisher had a 3.0 percent sophisticated bot rate, two publishers in the group encountered double the average rate — 5.9 and 6.9 percent sophisticated bots. Eight publishers had sophisticated bot rates equal to or exceeding 4.0 percent.





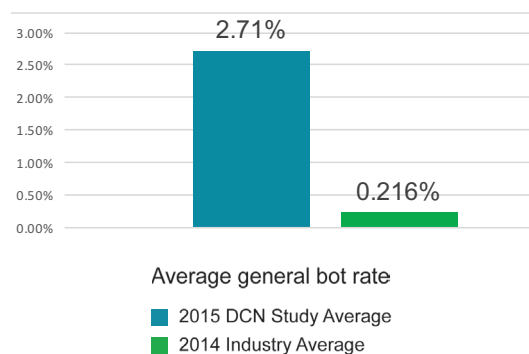
General bots are 10x more common in DCN inventory

Participants in *The Bot Baseline 2014* showed average general bot levels of 0.2 percent, while *DCN Bot Benchmark Report* participants showed an average general bot rate of 2.7 percent, with up to 11.9 percent general bots.

General bots may exist in higher numbers on premium sites as quality control or analysis tools and as content crawlers generated by the publishers or by third parties.

The *DCN Bot Benchmark Report* data filters out impressions matching patterns included in the industry spider and robot list, which can have a dramatic impact on publisher bot rate. This allows the results to be compared to *The Bot Baseline 2014* study. Including these “known” bots would otherwise result in four publishers with double-digit bot percentages.

General bots can be filtered out by the publisher to protect advertisers’ valid impression levels. The ad networks serving ads into these publishers’ sites can also filter for them. Detecting the sophisticated bots that are not revealed by an industry bots and spiders list is much harder, as such bots cannot simply be filtered out when the name matches patterns in the list. Sophisticated bots morph and adapt to defensive measures, so it is imperative for publishers to monitor for these sophisticated bots and then to filter them in order to reduce their exposure to refund requests and make-goods and to demonstrate superior valid impression levels.



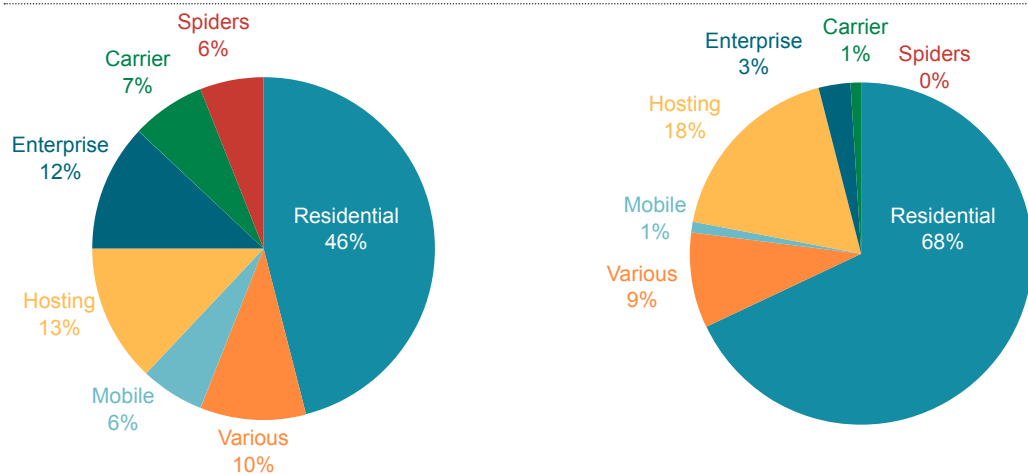
Sophisticated bots are adapting

Bad guys branch out, sending sophisticated bots from a greater variety of Internet addresses

The premium publisher domains studied saw a significantly different distribution of bot sources than industry-wide domains. The *Bot Baseline 2014* found that two-thirds of sophisticated bot traffic came from residential IP addresses and a little less than one-fifth came from hosted data centers. In the *DCN Bot Benchmark Report*, however, nearly a third of sophisticated bot traffic came from enterprises, carrier networks, and mobile networks, compared to only five percent in the previous study. A significant increase in bot traffic labeled as spiders originated from adtech companies running bots that are not registered with the industry bots and spiders list. The two major sources of sophisticated bots in *The Bot Baseline 2014*, residential networks and hosted data centers, accounted for less than 60 percent of DCN traffic. *The Bot Baseline 2014* study showed that 68 percent of sophisticated bots originated from residential IP addresses.

The different distribution of bot sources suggests that the DCN participants encountered a greater use of sophisticated data-center bots, rather than just malware on consumers' home PCs. **The broadening of the base of bots indicates that bot operators are using a wider variety of platforms, possibly adapting to defensive measures, such as malware detection on users' computers and blacklists.** In addition, operators are taking advantage of new ways to exploit platforms other than desktop computers to drive bot traffic.

Bot Sources by IP Type



DCN Benchmark Report

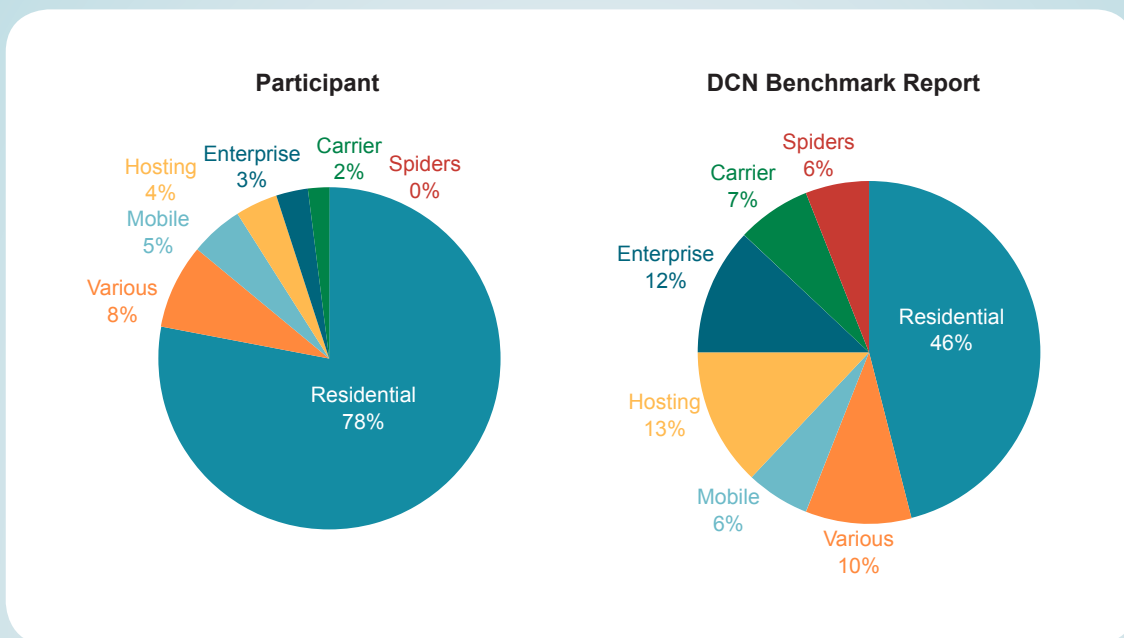
2014 Industry Average

While the greatest portion of sophisticated bots encountered in the DCN web sites studied came from residential Internet addresses, the traffic accounted for only 46 percent of all sophisticated bots. A similar decline was seen in bots coming from servers hosted in data centers, which accounted for 13 percent of all traffic seen by DCN publishers, down from 18 percent in *The Bot Baseline 2014* data set.

Bots originating from mobile networks in the above chart (six percent) came from computers using 3G/4G hotspots tethered to mobile devices.

CASE STUDY

Malware-infected computers drove 78% of one publisher's bot traffic



While the percentage of bots coming from residential IP addresses decreased for the DCN group, at least one publisher appears to have been targeted by bot operators who had compromised large numbers of home computers.

Most of the participant's bots (left side) came from residential sources and fewer bots came from enterprise networks, hosted environments or carrier networks.

Sophisticated bots coming from residential systems accounted for more than three-quarters of this publisher's bot traffic. The publisher also saw elevated levels of long-tail bots (a single bot visit to tens of thousands of pages within the publisher's domains) as well as bot hotspots (specific pages with higher-than-average sophisticated bot percentages).

In this case, elevated traffic from residential botnets (regular home computers with adbot malware infections) overwhelmed all other bot sources. These sophisticated bots highlight the problems that publishers face from advertising traffic coming from infected consumers' PCs and underscore that non-human traffic does not follow a uniform pattern of behavior or distribution — a result that was similar to *The Bot Baseline 2014* study.

Policies Impact Sophisticated Bot Rates

Publishers' policies impact how much sophisticated bot traffic their sites receive. In particular, traffic sourcing appears to be a major factor in the increase of sophisticated bot traffic. The data in the *DCN Bot Benchmark Report* shows additional links between policies impacting sophisticated bot rates.

As such, White Ops attempted to correlate publishers' policies and strategies to the resulting sophisticated bot traffic to their sites. Two broad categories of policies, including those designed to increase traffic to sites, offer alternative methods of selling inventory, improve the quality of traffic, and maintain a flow of content to the site, were reviewed.

DCN Publishers reported their use of the following policies:

1. TRAFFIC SOURCING AND RETARGETING

Uses vendors to drive traffic to outside sources	74%
Uses vendors to drive traffic from outside sources	70%
Retargets site visitors	43%

2. AUDIENCE DEVELOPMENT AND CONTENT POLICIES

Syndicates content	83%
Day-parts some categories	61%
Allows syndicated content on sites	48%
Allows pop-ups	39%
Sells data to 3rd party DMPs	22%
Sells inventory through ad networks	22%
Contains 30%+ user-generated content	4%

The most popular strategies among publishers include syndicating content (sharing content such as a web site, blog, video, articles, etc., to third-party sites), audience extension (generating revenue from web traffic by allowing advertisers to reach audiences beyond the publisher's site) and traffic sourcing (paying other companies to send users to a site).

In addition to measuring policy usage among publishers, publishers revealed whether they take viewability and non-human traffic (NHT) into account for billing purposes. About 59 percent (13 out of 22) of DCN participants who responded take viewability into account for billing purposes, while just 18 percent (4 out of 22) take NHT into account for billing purposes.

Specific impact of policies

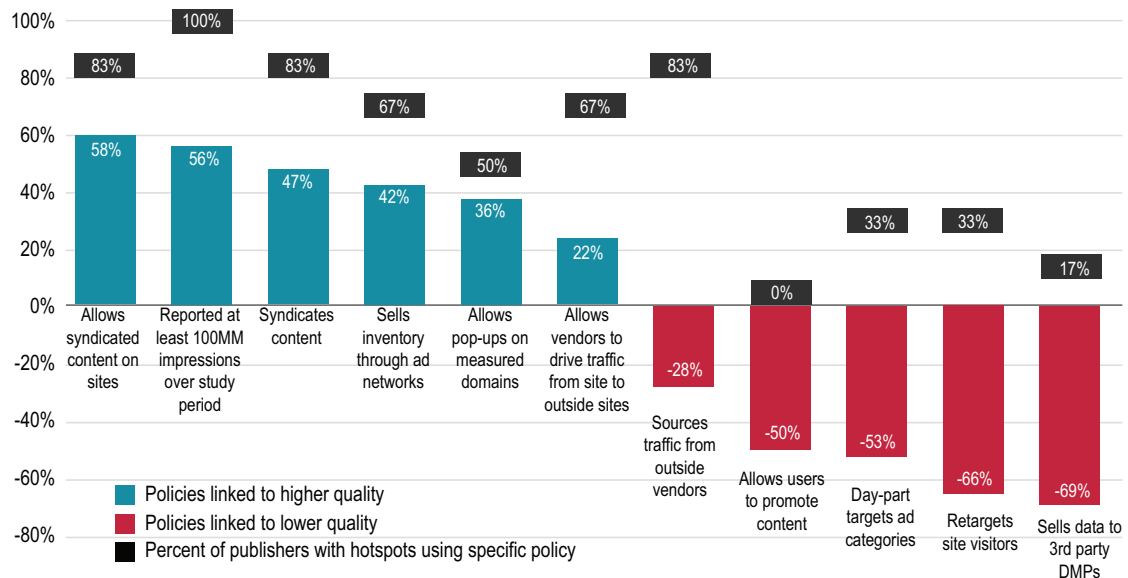
Some policies appear to be correlated with either an increase or a decrease in sophisticated bot traffic to the publishers' sites. Using self-reported survey data on policies, White Ops compared the proportion of companies that had adopted the policy in each of three categories: publishers with low sophisticated bot rates, publishers with high sophisticated bot rates, and publishers with some sophisticated bot hotspots (sites or campaigns with more than five percent sophisticated bot impressions).

Companies who sold data to third parties, retargeted visitors and used day-parting as a strategy to attempt to eliminate bots showed the largest reduction in valid impression levels compared to publishers with the lowest sophisticated bot rates who did not tend to use those strategies. Yet, publishers who used syndicated content, reported at least 100 million impressions over the study period and syndicated their own content were far more likely to be in the low sophisticated bot category than the high bot category.

Publishers who tagged at least 100 million impressions during the study period, bought traffic from third-party sources, and allowed syndicated content on their site or syndicated content elsewhere, all had significant hotspots.

While the combination of the survey and sophisticated bot rates suggests some interesting trends, the survey relied on participants to reveal their own policies and strategies and thus may not be representative of the overall impact of publishers' policies.

Sophisticated bot impact of 22 publishers' self-reported policies



As recommended for advertisers in *The Bot Baseline 2014*, publishers should continuously and actively monitor their own traffic for bot hotspots and bot-heavy referral traffic sources to reduce exposure to bot traffic. This serves to protect both advertisers and site quality.

Habits of Highly Successful Publishers

The DCN data and White Ops analysis showed that the publishers with the cleanest traffic used the following measures:

1. Strictly vetted traffic from third parties

While *The Bot Baseline 2014* showed that sourcing traffic from third parties can result in high bot percentages, publishers that restrict traffic purchases to top-line recommenders can do so without seeing increased bot traffic.

Publishers can safely source traffic while maintaining high valid impression levels in their inventory. Three of the top five publishers with the lowest sophisticated bot percentages sourced traffic from large suppliers and maintained sophisticated bot percentages below two percent with general bot percentages of 0.5 to 1.3 percent.

Another well-known publisher carefully vetted sourced traffic options and found that the available inventory did not meet quality standards. This publisher opted not to source traffic from third parties. This publisher has a sophisticated bot rate of 2.6%, better than its peers among both ANA 2014 marketers and the *DCN Bot Benchmark Report*.

2. Did not use viewability as a sole measure of whether impression was caused by a bot

Two top-volume publishers with bot rates below the DCN average did not bill based solely on viewability. Strategies to maintain high inventory quality for these publishers include managing all factors, including viewability, with a strong foundation of policies that limit the ingress of bots and other inventory defects through all channels.

3. Maintained visibility and control over their inventory

Inventory that was classified for media type had bot percentages that were well below the average bot percentage of unclassified inventory. This may indicate that the inventory that is easier to tag and monitor is also the easiest to manage for quality factors such as humanity percentages.

4. Limited the use of retargeting

The top volume publishers, who did not retarget site visitors, had an average sophisticated bot percentage of 1.8 percent, compared to the higher sophisticated bot average of 2.7 percent for similar publishers who retarget site visitors.

Many publishers in the *DCN Bot Benchmark Report* demonstrate consistently high-quality inventory and maintain quality through best practices to reduce bots in high-value inventory. **Buying direct from these DCN publishers would have the effect of achieving the same low sophisticated bot rates as the inventory for the top six advertisers in *The Bot Baseline 2014* with no additional implementation of policies or procedures.**

CASE STUDY

Some bot rate variations are influenced by current events

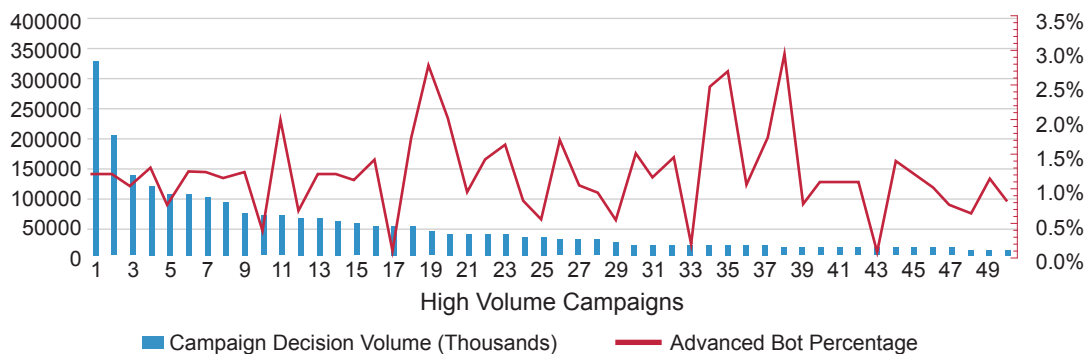
Current events can have an impact on the sophisticated bot rate. In the case of one publisher, recurring events caused the volume of humans interested in visiting the site to increase, resulting in a decline in the percentage of sophisticated bots. For publishers whose human audience levels vary in response to ambient factors including current events, this type of quality variation can be seen independent of the publisher's site policies.

CASE STUDY

Data can pinpoint publishers and campaigns with consistently very low sophisticated bot rates

The DCN data shows that analysis can identify those publishers with very low sophisticated bot rates, both overall and on a per-campaign level. Because marketers can generally expect low bot impact on campaigns from those publishers, direct buys from those publishers will have consistently better inventory quality.

A PUBLISHER'S 50 HIGHEST VOLUME CAMPAIGNS ALL HAD SOPHISTICATED BOT % BELOW 3%, AVG 1.6% SOPHISTICATED BOTS

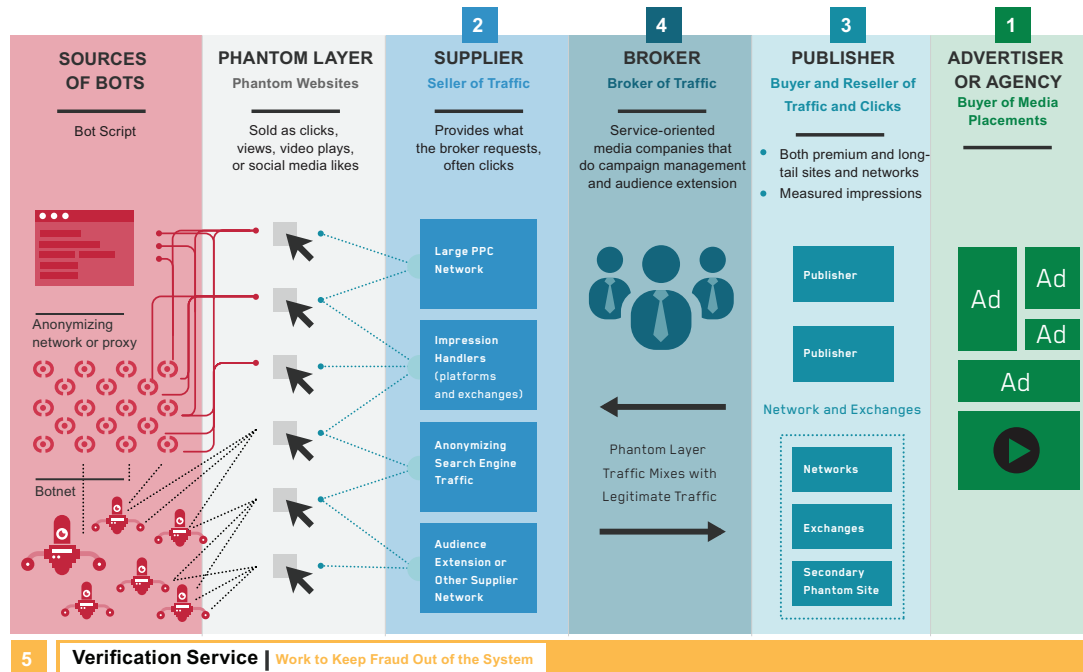


The DCN publisher with the lowest average bot rate, 1.6 percent sophisticated bots as measured by White Ops, had below three percent bot rates on each of the 50 largest ad campaigns.

Definitions

general bots	— bots that can be detected through the industry bots and spiders list and known browser list
sophisticated bots	— bots not listed in the industry bots and spider list and known browser list
sophisticated bot rate	— the percentage of total traffic for which sophisticated bots are responsible, compared to total traffic
impression	— an advertising event, resulting from either a human or a bot causing an advertisement to be accessed or clicked on
valid impression level or impression validity	— the percentage of impression traffic that represents legitimate, human-viewed traffic
mobile impressions	— those ad impressions coming from web pages browsed from a mobile device
non-mobile impressions	— those ad impressions coming from web pages browsed from desktops, laptops and gaming consoles
unmeasurable impressions	— cases where the JavaScript tag was not loaded due to factors such as page abandonment or site configuration
bot hotspot	— campaign or site with more than one thousand impressions and more than five percent sophisticated bots
false positives	— detection results that indicate bot or ad fraud impressions when the impressions are actually legitimate
HREF domain	— the domain where a particular ad impression, page view, video play, or other on-line event appears to occur
referring domain	— domains that send traffic to the site
true domain	— a method of using JavaScript to verify the domain which was loaded

The Online Advertising Ecosystem



The online advertising ecosystem has a significant number of layers and participants that aim to deliver appropriate advertising to online visitors. Many of these groups, such as traffic brokers and advertising networks, deliver legitimate business services. Yet, the complexity of the advertising ecosystem also allows bad actors to hide their activities and allows automated software (bots) to game the system.

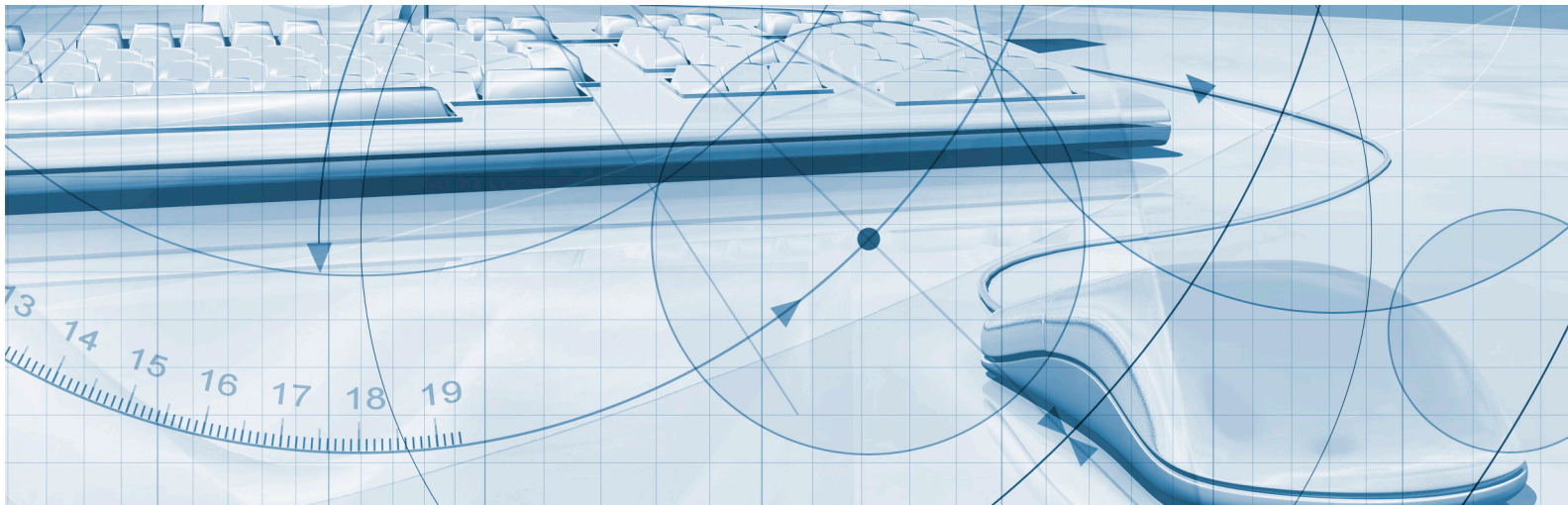
Some links in the digital advertising supply chain are unaware of the bots in their traffic and do not intend to profit illicitly, while others actively encourage and concentrate bot traffic to increase profits.

Stakeholders in the advertising ecosystem:

- 1** Companies buy advertising campaigns when their marketing teams work together with an advertising agency's media buying team.
- 2** Advertising networks sell franchised advertising space on different publishers' web sites.
- 3** Publisher's advertising operations team and sales groups work to create and sell space, impressions and other marketable units.
- 4** Third-party traffic brokers sell audience and clicks to publishers and other groups in the advertising ecosystem.
- 5** Third-party advertising verification services track actual ad clicks and views and work to keep fraud out of the ecosystem, or at least measurable.

Increasingly sophisticated bot operators attempt to fool all stakeholders in the digital advertising supply chain:

- Bots are continuing to get better at masquerading their software as valid impression traffic.
- The ability of bots to automate clicks, impressions, and viewability means that they can scale millions of bot impressions every day.
- By posing as legitimate traffic, bots can sneak through the different layers of the advertising ecosystem.
- An illegitimate layer of the ecosystem (the so-called phantom layer) launders clicks and obfuscates the bot networks that operate behind it.



The Sophisticated Bot Picture for Premium Publishers

The DCN member sites studied are achieving a vastly higher level of inventory quality than the industry average. For the median publisher, **3.0** percent of advertising impressions came from bots, not including search bots, web spiders and other automated traffic caught by the industry spider and bot list. Across all 16 billion impressions studied, only **2.8** percent came from sophisticated bots and **5.9** percent came from bots of all types.

Ad impressions on publishers' sites

White Ops only analyzed bots in traffic from desktop and laptop browsers. To allow comparison between the *DCN Bot Benchmark Report* and *The Bot Baseline 2014*, mobile traffic and unmeasurable traffic were not included in the study. For reference, mobile traffic accounted for 42 percent of all impressions, while unmeasurable traffic accounted for another eight percent.

Publishers with less overall traffic tended to have higher sophisticated bot rates. The half of the DCN participants with the lowest traffic, for example, accounted for seven of the top ten participants with the highest sophisticated bot rates.

Of the bots encountered by DCN publishers, about 53 percent come from sites included on the industry spider and bot list. The other 47 percent are sophisticated bots not caught by the whitelist/blacklist combination. The 53 percent of known bots are mostly search spiders and basic site scraping bots, while more sophisticated bots are used for ad fraud and content theft.

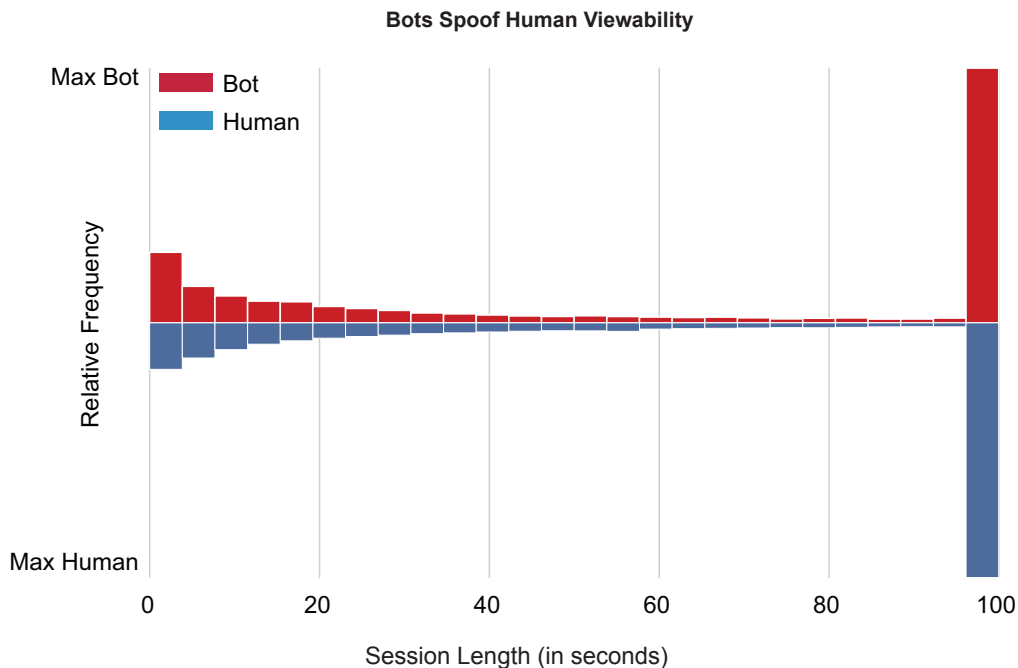
Sophisticated bots mimic viewability

Viewability measurement does not reveal human audience rates

An advertisement's viewability and whether the visitor was human or bot are largely unrelated. The study analyzed the viewability of bot visits (red) and that of human visits (blue). Bot viewability almost perfectly mirrored human viewability apart from a small number of bots with short visits. With this level of similarity, viewability as a measure will not help differentiate a visitor as human or bot. Furthermore, sophisticated bots are known to deliberately spoof viewability and post back detailed viewability data.

White Ops found that bots are now spoofing viewability in nearly three quarters of cases.

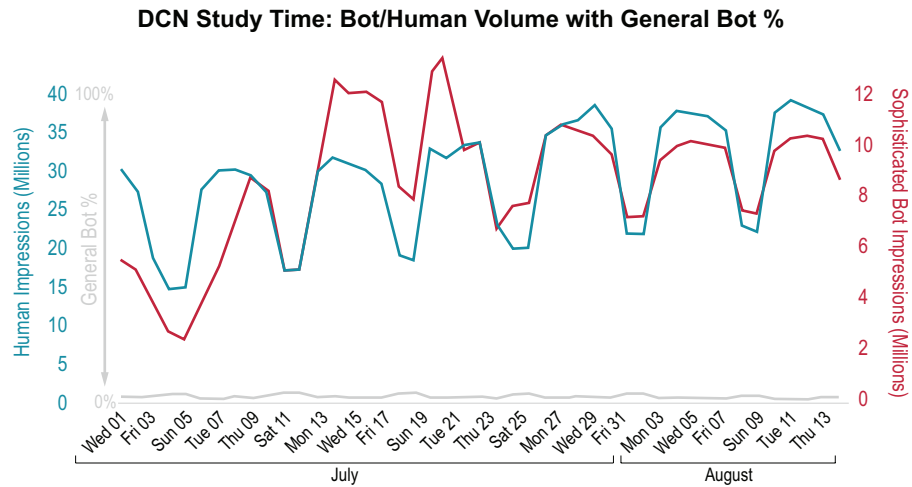
Even bots that come from hosts dedicated to fraud, with no real human traffic have high viewable rates and most frequently post back detailed viewability data.



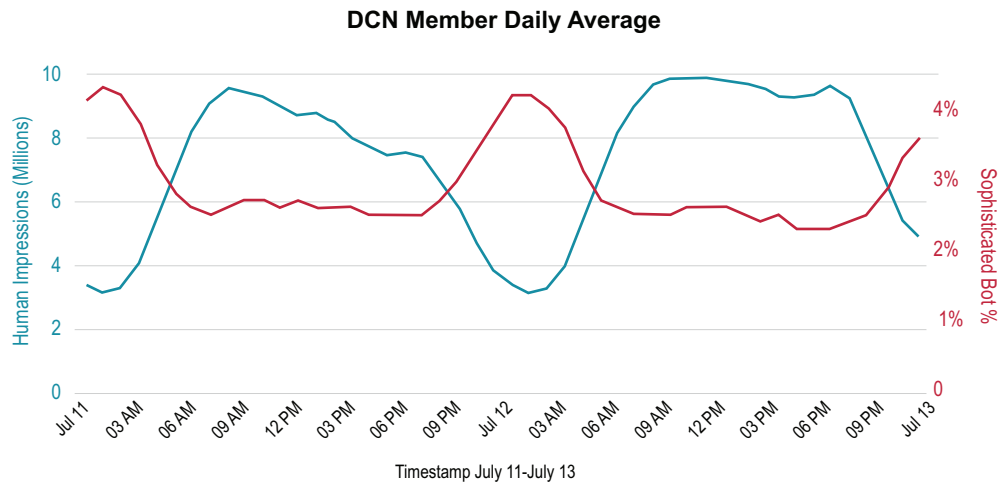
The distribution of session length between bot traffic and human traffic shows that bots are mimicking humans. Although sophisticated bots tend to have slightly shorter session times than humans, they overall do a convincing job of mimicking human patterns, including viewability.

Smart bots mimic human usage patterns

Bot operators are now sending the highest volume of bot traffic during working hours on weekdays in an attempt to foil quality control techniques such as day-parting.

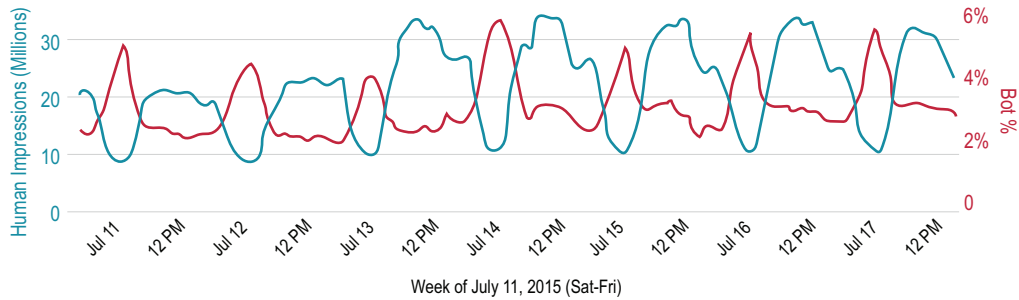


Yet, even with the steep drop in bots sent each night, bot traffic accounts for a larger proportion of night-time traffic to sites, because human traffic to sites drops off even more precipitously at night.

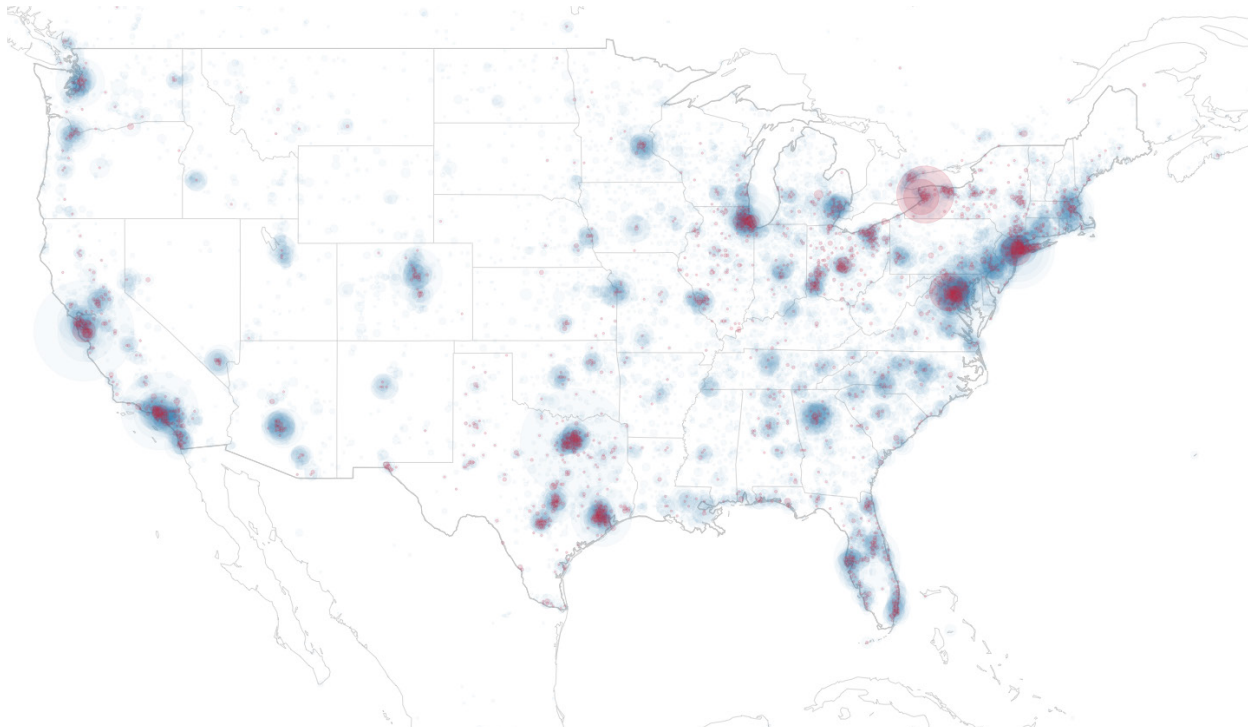


On weekends, the range for both human and bot traffic tends to be smaller as well: Human traffic drops off, and sophisticated bot traffic shows a slight reduction in both maximum and minimum values.

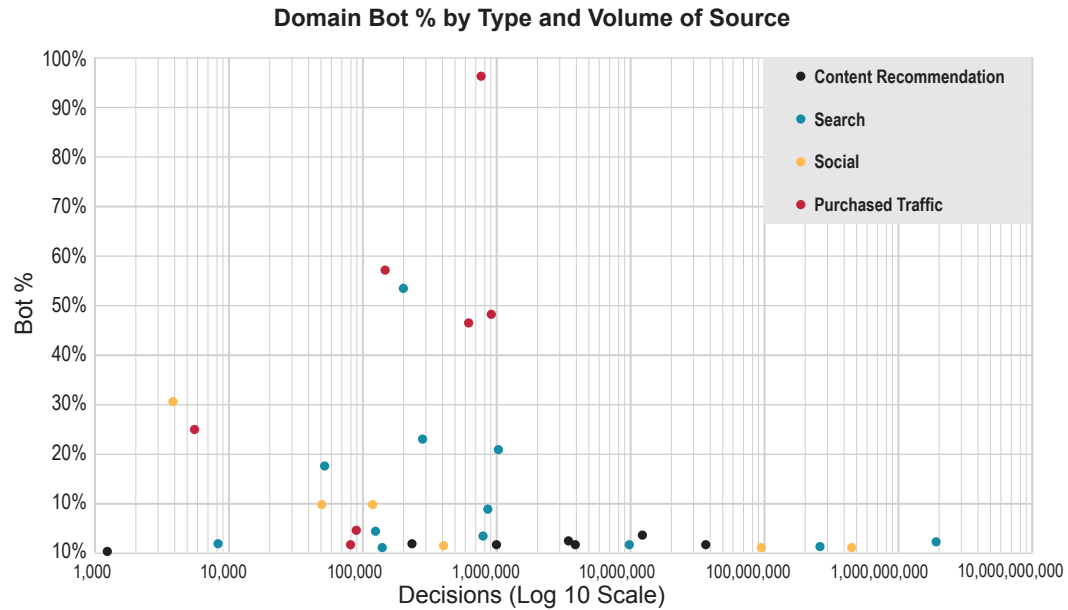
Weekly Bot Rate Increases At Night



Geographic maps created from the source IP addresses of traffic show that sophisticated bots tend to come from areas with large data centers, which in many cases overlaps with large population areas. In some cases, however, such as the large sophisticated bot population coming from Virginia, population alone cannot account for the influx of automated traffic.



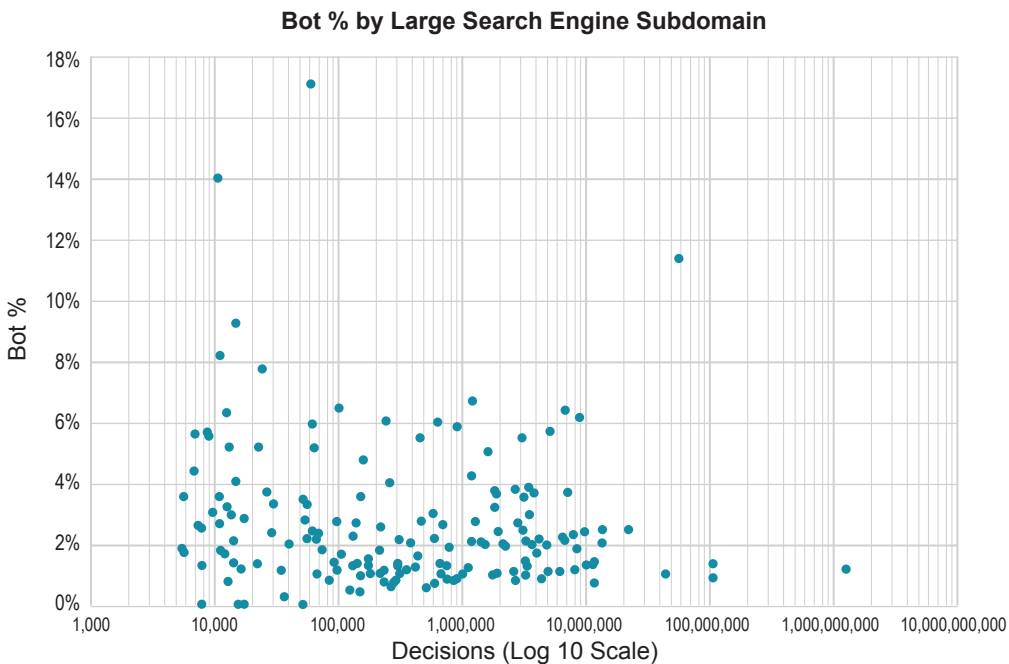
Small traffic recommenders and some search companies have high bot rates



The impact of paying for traffic from third-party sources varies by the type of source. Large content recommenders (black dots) refer very little sophisticated bot traffic.

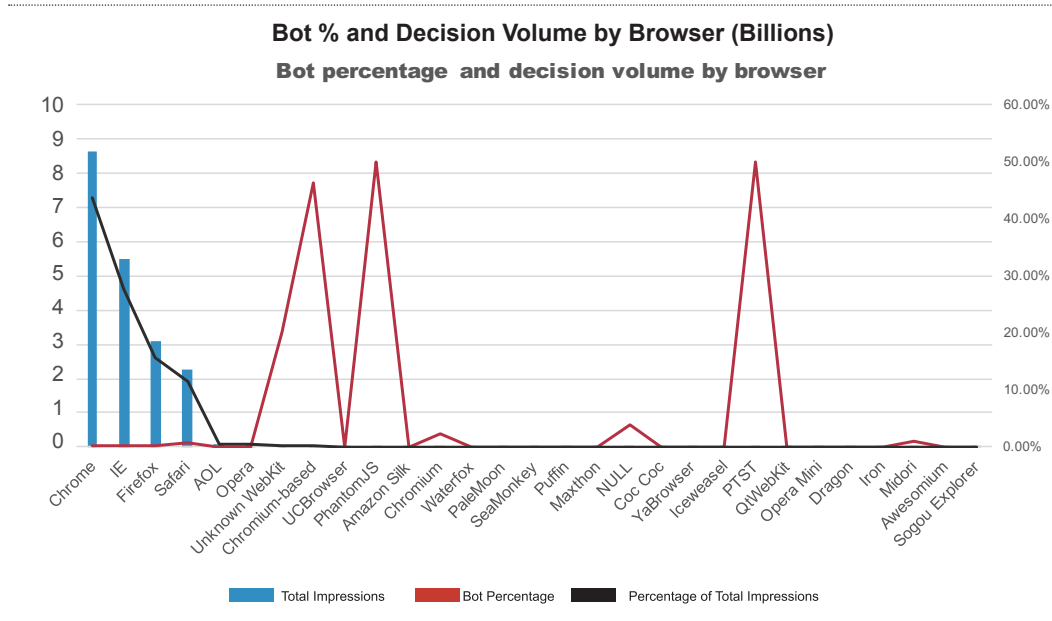
Search traffic is on the whole very clean, but smaller traffic sources (red dots) refer a higher percentage of bots. Traffic from reputable search engines is largely bot-free. Finally, smaller search engines show medium to high levels of bot traffic.

Traffic from social media sites tended to be more clean, with a few exceptions.



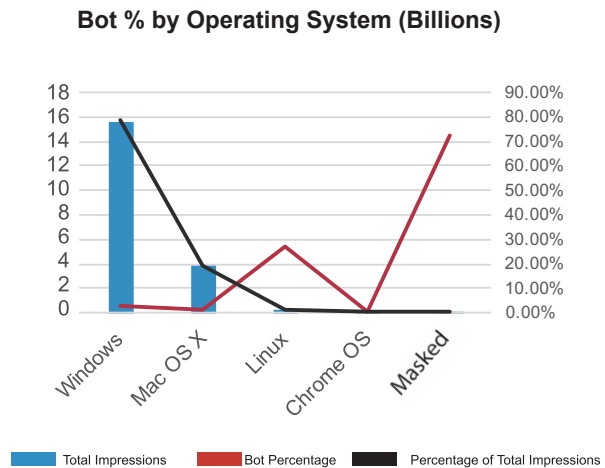
Bot operators focus on a few, minor browsers

The sophisticated bot percentage for popular browsers, as reported in the HTTP request, varied from one percent for Safari browsers to 0.12 percent for Chrome browsers. Among less common browsers, Opera was 0.04 percent sophisticated bot, while “Unkown Webkit” was 20 percent sophisticated bot and “Chromium-based” browsers were 46 percent sophisticated bot.



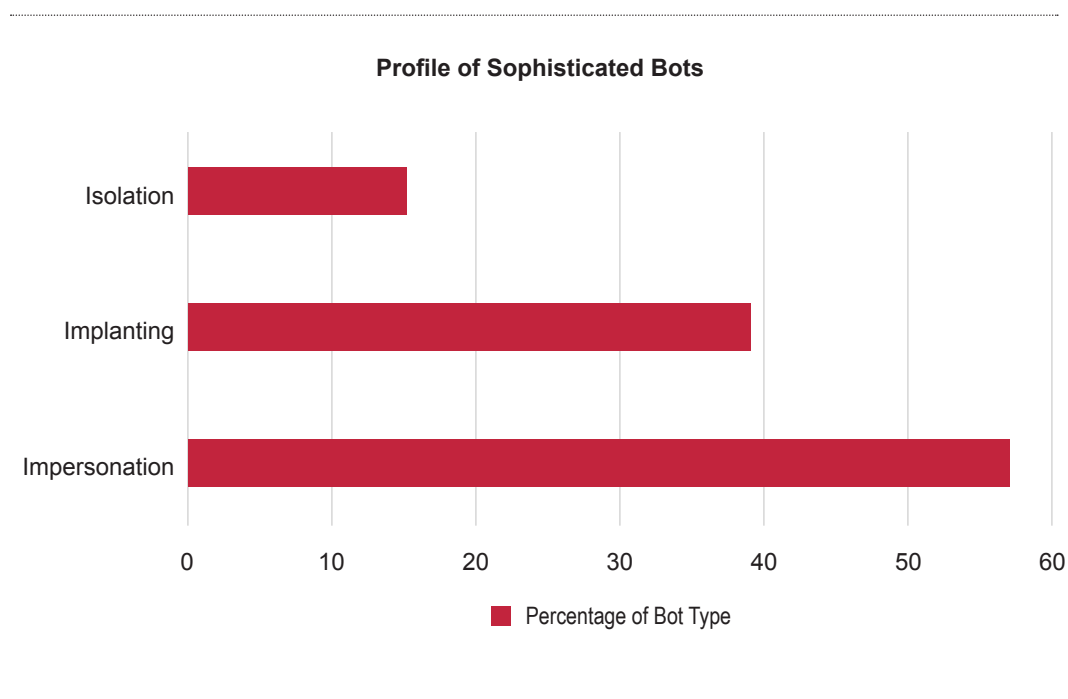
Bot percentage across popular operating systems

Sophisticated bot percentage for self-reported operating system varied from one percent for Mac OS X to 27 percent for Linux. Impressions with unreported operating systems had 73 percent sophisticated bots.



Sophisticated bots use impersonation, hide referrer tag

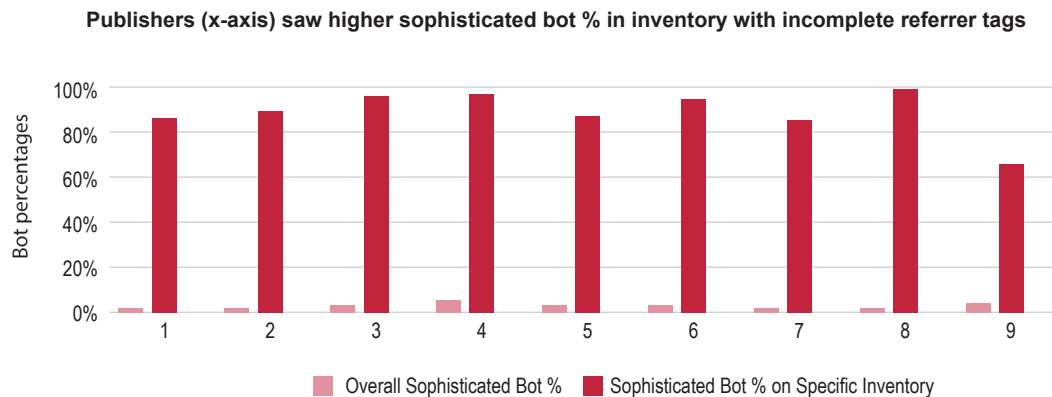
Using the industry bots and spiders list and known browser list, traffic from known bots can easily be filtered from the visitors to publishers' sites. However, bot operators are increasingly using more sophisticated techniques to allow their bots to masquerade as human. Almost 58 percent of sophisticated bots use some sort of impersonation, where the client is pretending to send traffic from a specific browser or from a specific system. Another 40 percent of traffic comes from browsers that have been compromised with covert malware or hidden adware that makes extra web calls unbeknownst to the user. Finally, about one-sixth of all traffic comes from a host dedicated to fraud, with no real human traffic.



Most sophisticated bots impersonate a specific browser in an attempt to be seen as human. Those characteristics are not exclusive. Publishers should consider using additional defense mechanisms: when the impression lacks a referrer tag or has incomplete user information, for example, the publisher may choose to perform additional verification to ensure that it is a human visitor. If the user fails to pass those tests, then the page load and the ad load can be prevented. This targeted defense can maintain impression validity even in potential inventory hotspots and can protect the publisher from coming under attack from sophisticated impersonation bots.

Sometimes, the techniques bots use to escape detection can be used against them

When the HREF tag was obscured in impression data, a cross-section of participants in various industry segments saw elevated bot percentages in those impressions.



Obfuscated HREF tags are highly correlated to sophisticated bot activity, but are not the only indicator. Likewise, five participants saw elevated sophisticated bot activity when the referrer tag was obfuscated in the impression data. In both cases, that data suggests that bot operators are trying to hide their existence or source, or hide some other data in those fields.

BREAKOUT

The benefits of measurement

Publishers who had the most capability to describe their inventory for analysis purposes had the lowest bot rates. Conversely, inventory that was less thoroughly tagged and described for the study had higher bot rates.

These findings may indicate that inventory that is more closely measured is also less vulnerable to bots. Publishers could capitalize on this by updating site technology, reducing the complexity of their web sites, and improving the measurability of high value inventory traits so that they can more easily measure and manage premium inventory.

The Rise of Ad Injection Malware

Some publishers may choose to be more aggressive in defending against **invalid traffic**, or **IVT** (traffic that consists of undesirable human or bot impressions). For example, these publishers may block “outlier” traffic that has malformed, incomplete, or missing information.

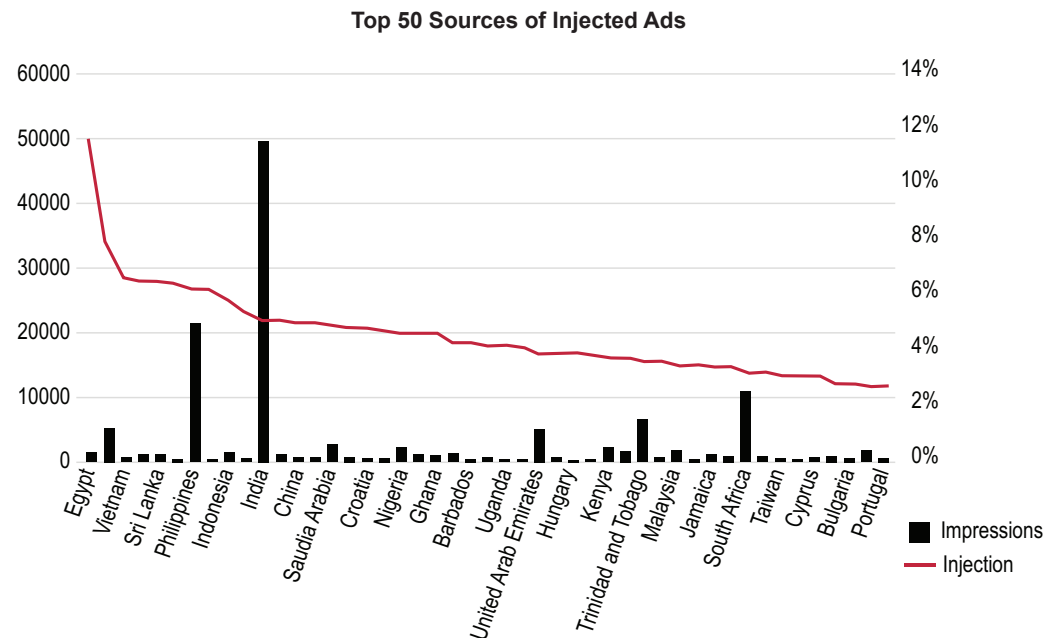
Humans visiting web sites would likely have common browsers, with simple and complete information. Invalid traffic that creates a human impression is more difficult to block or filter or even to identify as invalid. These quality assurance techniques may reduce bot percentages in inventory but cannot counteract invalid human impressions in inventory.

Ad injection is a form of invalid traffic that creates a human impression using a man-in-the-middle malware attack, usually from a user’s infected machine. *The DCN Bot Benchmark Report* found no incentivized traffic or adware but did find evidence of ad injection.

CASE STUDY

A participant’s ad injection rate was higher than their sophisticated bot rate

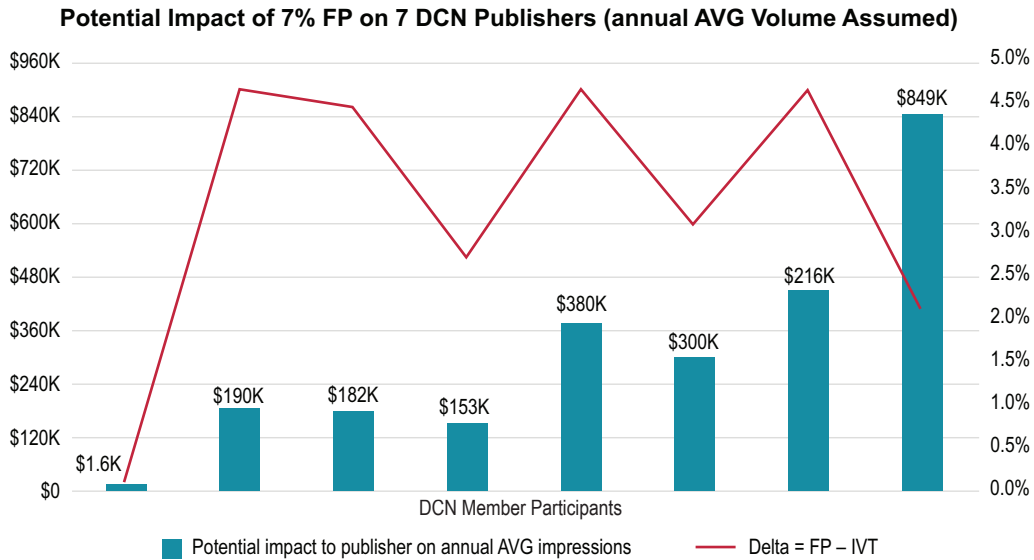
Ad injection continues to cause problems for the advertising ecosystem. Egypt and Vietnam were among the top countries from which ad injection originated, impacting the participant’s sites. While the United States accounted for the largest amount of overall impression traffic, it ranked 54th in terms of ad injection traffic. Overall, however, the top 50 countries responsible for ad injection (red) accounted for only a minority of ad traffic (black).



White Ops tested one participant’s inventory for ad injection and found injected ads comprised greater than three percent of all impressions, while sophisticated bot traffic was less than two percent. Similar levels of injection likely affect other publishers.

Impact of false positives and unverifiable vendor IVT reports

Thirty-nine percent of publishers surveyed said they had been presented with reports of invalid traffic in their data. Most of these reports were not verifiable by anyone outside of the reporting organization because the reports did not include methodology and the vendors did not provide transparency in the methods they used to detect bots.



False positives in billing can affect the publishers' bottom line. Premium publishers can be especially hard hit by unverifiable invalid traffic reports because their bot percentages are generally lower, and the value of the inventory at stake can be higher.

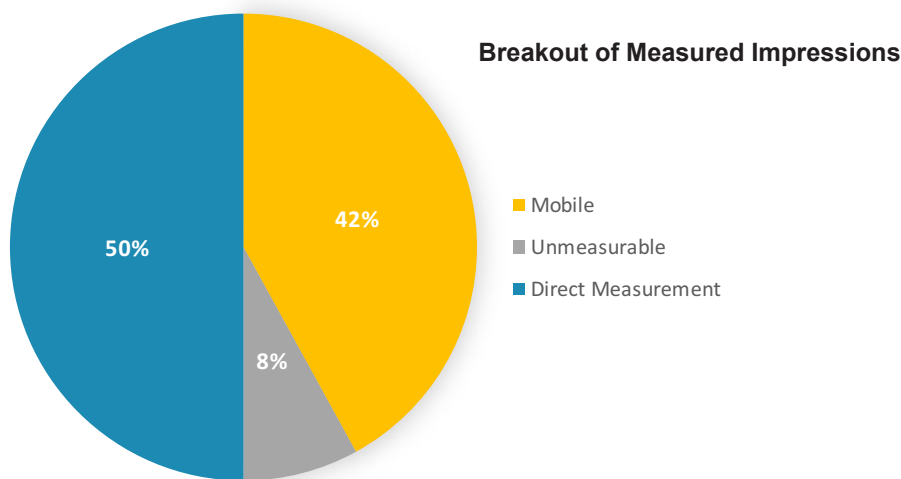
Assuming a seven percent false positive reporting rate, and using a sample of DCN publishers, White Ops calculated the potential annual impact using the detected sophisticated bot percentage with the vendor's self-reported CPM and volume averaged over the group. Losses ranged from less than \$2000 to more than \$849,000 annually.

To protect publishers from unverifiable invalid traffic reports, it is essential for vendors and marketers who report invalid traffic to back up the reports with verifiable methodologies and to be transparent in their detection practices as recommended by nonprofit groups such as the Trustworthy Accountability Group (TAG), Interactive Advertising Bureau (IAB), and Media Rating Council (MRC).

Study Methodology

For the 53 days of the survey, DCN participants deployed White Ops detection tags via their ad server, directly to their web site, or a combination of both. White Ops relied on the participants to comprehensively tag sites and to reveal information on media types, buy types, and their operational policies.

Where possible, the White Ops technology gathered information directly at the time of impression. No data or results were provided to study participants during the data collection period. Because mobile results were not included in *The Bot Baseline 2014*, the *DCN Bot Benchmark Report* focused on non-mobile visitors only. In addition, in cases where the visitor did not completely load the page, the impression was considered “unmeasurable.”



Some publishers choose to sell inventory across web properties that they do not own and operate, commonly referred to as partner networks, syndication networks or audience extension networks. Such inventory falls outside the scope of the results in this study.

Characteristics of the data set

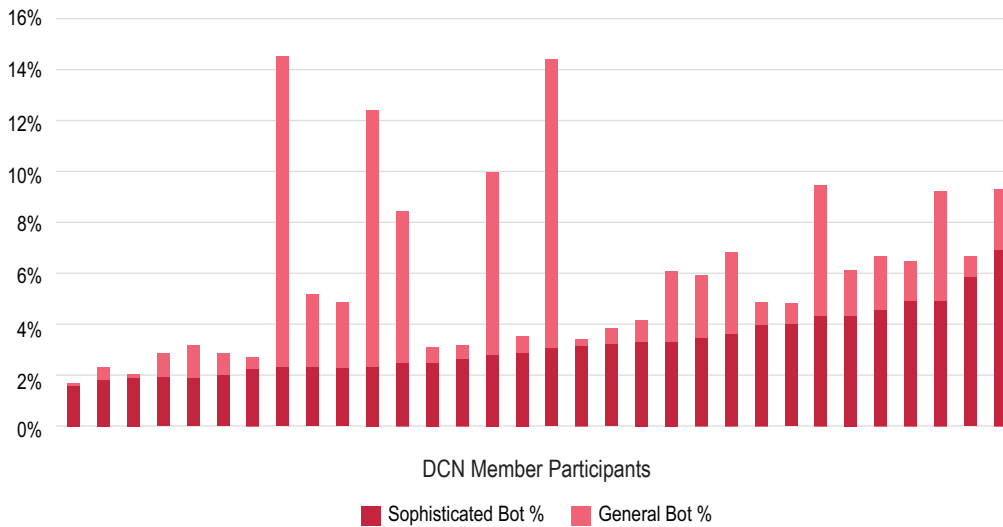
Using White Ops technology, DCN publishers tagged their sites and advertising placements. Tagging was neither uniform nor complete. Ten publishers tagged only their advertising placements. Another nine tagged only their content at the site level. About 40 percent of participants tagged content both at the advertising placement level and at the site level. A total of 32 publishers participated in the study.

To characterize the resulting data set, White Ops used two separate averages throughout the report. The average publisher bot percentage of 3.0 percent was calculated as the median across participants, weighting each publisher equally to create a profile of the sophisticated bot percentage for a “middle-of-the-road” publisher. The average overall bot percentage of 2.8 percent was calculated as the proportion of all measurable, non-mobile traffic that came from sophisticated bots.

Incomplete loads and other non-measurable traffic were not included in the study. The scope of this study looked closely at the quality of traffic running on the owned and operated web properties of the participating DCN publishers.

In addition, White Ops removed general bots (bots detected by the industry spider and bot list), which include legitimate, automated search spiders, as well as known malicious bots. The inclusion of bots from the industry bots and spider list (pink), changes the distribution of bots in the study tremendously.

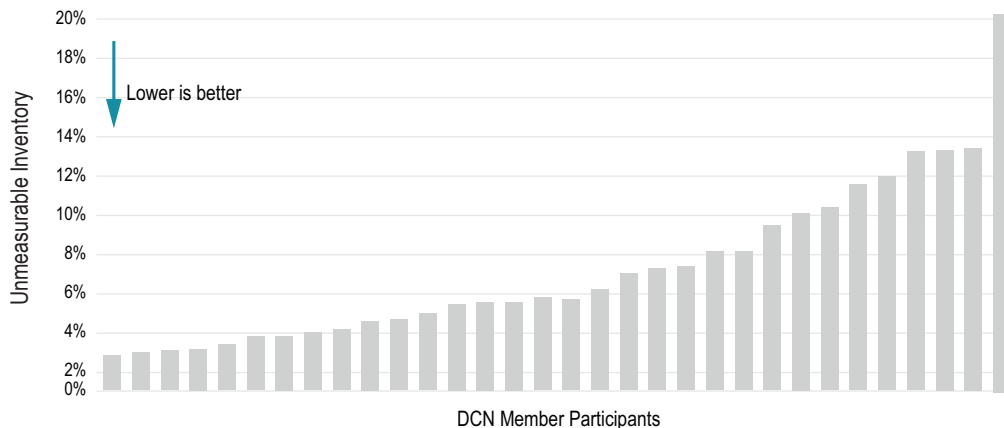
Publisher Sophisticated and General Bot Rates



Some types of inventory were not represented in the study

On average, 92 percent of individual publishers' tagged traffic was measurable. For the 2015 *DCN Bot Benchmark Report* as well as in *The Bot Baseline 2014* study, White Ops calculated results using measurable, non-mobile inventory.

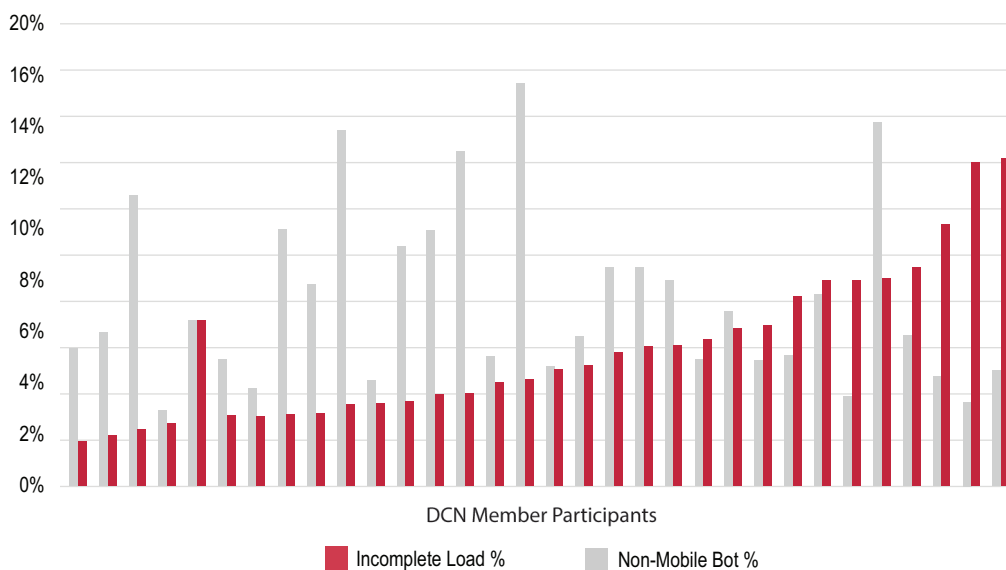
Unmeasurable Inventory by Publisher



White Ops detected several anomalies in the data that indicated that, in some cases, only a portion of some publishers' traffic was available to be analyzed. Some publishers, such as one news organization, were only able to report display ads and were not able to report their video inventory. Others sent very little traffic to White Ops. Seven of the participating publishers did not meet the volume requirement for the study (minimum of one hundred million impressions during the study period). **For some of these publishers, the inventory measured in the study cannot be considered representative of the whole of the publisher's inventory.**

Incomplete loads, or unmeasurable impressions

In many cases, the tagging used by White Ops failed to completely load, and thus the impression data was lost and the validity became unmeasurable for those impressions. There could be a variety of reasons for this behavior, such as a human visitor navigating away from a page before the page completely loaded.



Data consistency

As a requirement for participation in the study, White Ops asked DCN members to tag at least 100 million impressions for the study. The type and categories of the sites being studied varies widely, and some sites could only be tagged at the site level, while others were tagged at the ad level or ad level and site level. This situation created a number of data collection inconsistencies in terms of the member profiles that affected the results of the study. For example, in some cases, a site only sent partial, or no, traffic back to White Ops servers. Some publishers only reported a fraction of their traffic volume. Finally, many publishers were not able to specify media types, such as video.

Sourced traffic

In *The Bot Baseline 2014*, sourced traffic was found to have a high sophisticated bot rate — sourced video traffic, for example, consisted of an average of 52 percent sophisticated bots. However, the incidence of tagged sourced traffic was much lower during the DCN study period due to the reduced volume of sourced traffic among the *DCN Bot Benchmark Report* participants in measured inventory.

Conclusions

Because ad fraud is such a sophisticated form of cybercrime, keeping up with new infections, new rootkits (automated exploit software), and new malware variations requires constant vigilance.

Sophisticated bots will continue to better emulate humans. Already, bot operators vary traffic from their bots according to weekday and weeknight patterns. Bots are evolving and moving in greater numbers to platforms they were not using at scale before. **But ultimately, the solution to this problem is not technical, it's economic.** Stopping this multibillion dollar problem requires shifting spending from the sites that increase their volume with bots to the sites that do not.

In the 2015 *DCN Bot Benchmark Report*, high-quality publishers demonstrated a lower sophisticated bot problem than the broader advertising ecosystem in general, due in large part to their effective policies and strategies.

Premium media must have low bot rates

Any definition of premium quality in the media business should require the kind of quality shown in these study results. Previous studies have shown that visitors who come to web sites as a result of traffic sourcing are actually bots at least 52 percent of the time. Whenever White Ops detects a lot of bots on a well-known site with non-premium inventory quality, it's always for one of two reasons: the site is buying traffic or it has a network of affiliated sites.

During our hunt for bots in advertising traffic and on the Internet, White Ops finds many “cash-out sites” that only a bot could love. These bogus publishing sites are there to trade cash for impressions, but, as soon as they get identified and excluded from buys, new ones arise to take their place. However, our results show that bot traffic is not isolated to low-flying, cash-out sites. One quarter of the bot impressions White Ops examines are on sites in the Alexa Top 1000, the most popular sites on the web. Our results show that popularity is no guarantee of premium quality.

Bots are designed to avoid detection by obscuring factors including HREF, REF, and Campaign ID. Blank values in impression data could indicate higher bot volumes in that inventory. In addition, nearly 60 percent of bots emulate a specific browser. Increasingly, parasitic bots target cookies or implant themselves in legitimate browsers, allowing the automated traffic to don the cloak of legitimacy and appear to be an actual human visiting sites.

Using data-collection systems to analyze the valid impression levels of traffic can help with these issues. Many of the risky policies, such as retargeting and traffic sourcing, can be offset through better traffic analysis and thorough vetting of suppliers. Data-informed policies can help publishers avoid third parties who provide traffic with high sophisticated bot rates.

Direct buys to premium publishers who demonstrate the commitment to maintain low sophisticated bot rates can consistently yield human audience levels of 97% or more.

Publishers can protect their marketers by agreeing to transparency in inventory quality, allowing tracking of inventory quality for validating bot and valid human impression percentages, and by agreeing to bill based on humanity.

Recommendations:

Publisher action plan to combat IVT in inventory

Publishers can reduce the risks and impacts of bots with careful monitoring of at-risk and high-value inventory.

Even premium publishers who do not engage in risky policies or activities will find sophisticated bot activity on their web properties that can expose them to economic risk; publishers must remain diligent in finding and removing sources of bot activity to minimize their risk.

In addition to the successful strategies and policies already in place to maintain inventory quality, the following steps can assist publishers in maintaining high audience humanity levels.

Safely increase traffic through policies such as sourcing

- If sourcing traffic, monitor sourced traffic to protect marketers from increased risk.
- Select suppliers with proven high humanity levels and a commitment to quality.

Defend against unverifiable NHT reports

- Use a monitoring solution to prove low bot levels and protect against lack of vendor accountability.
- When possible, use MRC-accredited monitoring solutions.
- Insist on vendor transparency for detection methodology to improve measurement quality.

Control bot hotspots and problem areas to protect inventory quality

- For inventory with irregular humanity percentages or high CPMs, monitor for bot activity to limit bot spikes.

Improve page measurability to precisely manage inventory quality

- Improve measurability to increase transparency with advertisers and provide the ability to find more fraudulent activity that is actively trying not to be measured.
- Troubleshoot issues with JavaScript measurement that can prevent monitoring of bot activity and avoid becoming low-hanging fruit for botnet administrators.

Monitor for ad injection and other forms of fraud that could damage your brand

- Limit damage done by fraud that is in addition to the advertising dollars siphoned off; monitor for ad injection that can damage publisher brands by negatively affecting the user experience on their web properties.
- Monitor for bots that crawl sites and “look” at content that will give them a specific cookie profile that can be targeted, diminishing the value of publishers’ own real user data.

Use bots’ evasive techniques to identify and expose them

- Monitor for impression data such as HREF, REF, and incomplete campaign ID that tends to mean much higher bots.
- Enforce stringent tagging and tracking measures that ensure that spoofing bots cannot hide in the noise of incompletely-tagged inventory.

About Us



White Ops, Inc.

White Ops is the leading provider of cyber security services for the detection and prevention of advanced bot and malware fraud. White Ops' innovative services help organizations improve their bottom lines and ensure the success of their campaigns, business goals, and the security of their systems and data. Unlike traditional approaches that employ statistical analysis, simple blacklisting or static signatures, White Ops effectively combats criminal activity by actually differentiating between robotic and human interaction within online advertising and publishing, enterprise business networks, e-commerce transactions, financial systems and more, allowing organizations to remove and prevent fraudulent traffic and activity. By working with customers to cut off sources of bad Internet traffic, White Ops makes bot and malware fraud unprofitable and unsustainable for the cyber criminals -- an economic strategy that will eventually eradicate this type of fraud. White Ops was recently appointed to the board of W3C.

More information about White Ops, Inc.
is available at www.whiteops.com



Digital Content Next (DCN)

Digital Content Next (DCN) is the only trade organization dedicated to serving the unique and diverse needs of high-quality digital content companies that manage trusted, direct relationships with consumers and marketers. The organization was founded in June 2001 as the Online Publishers Association (OPA). Comprised of some of the most trusted and well-respected media brands, DCN produces proprietary research for its members and the public, creates public and private forums to explore and advance key issues that impact digital content brands, offers an influential voice that speaks for digital content companies in the press, with advertisers and policy makers, and works to educate the public at large on the importance of quality content brands. Digital Content Next's membership has an unduplicated audience of 233 million unique visitors or 100% reach of the U.S. online population.

More information about Digital Content Next
is available at www.digitalcontentnext.org

SOURCE: comScore, Inc. ®, January 2015



AccuWeather.com™



Bloomberg

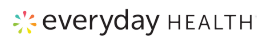
BUSINESS
INSIDER



CONDÉ NAST

THE DAILY CALLER

The Dallas Morning News



Forbes

GANNETT



NBCUniversal



The New York Times

Purch



TheStreet

Time Inc.



The Washington Post

WebMD

Gannett and USA Today participated jointly
CNBC.com and NBC News participated jointly

Thank you to all participants in the 2015 DCN Bot Benchmark Report